43

Thin-Film Solar Cells

43.1 Introduction 497
43.2 Technology Overview 498
Silicon-Based Cells • CdTe-Based Cells • Copper Indium Gallium Selenide–Based Cells • Dye-Sensitized Solar Cells • Organic Solar Cells • Perovskites-Based Cells
43.3 Optical Modeling 503
43.4 Transport Modeling 511
Drift-Diffusion Model • Kinetic Monte Carlo • Example: Transport through 1D Disordered System by kMC
43.5 Specific Modeling Issues 527
Dye Solar Cells • Organic Photovoltaics • Perovskites • Modeling of Morphology and Grain Effects

Matthias Auf der Maur

Tim Albes

and

Alessio Gagliardi

43.1 Introduction

In this chapter, we review simulation models for thin-film solar cells, and especially the specific challenges encountered in modeling of thin-film cells based on emerging technologies. From a technological point of view, the numerous currently available photovoltaic (PV) technologies can be roughly subdivided into two classes. One is based on conventional (mono-)crystalline semiconductor technology, to which belong the first generation of solar cells such as monocrystalline silicon or gallium arsenide p-n junction cells, and the other is based on thin-film technology [1]. The term *thin-film* in this context does not primarily refer to the thickness of the devices, but to the type of production technologies used, although their physical thickness is typically small in the range of few tens of micrometer to well below 1 μm [19,90].

Thin-film technology is characterized by several advantages over conventional technology [4]. First of all, material deposition techniques are often more simple. Depending on the material system, there is a vast diversity in techniques. Many of these are low-temperature methods, when compared to standard crystalline semiconductor technology, and many of them are easily scalable to large-scale production and large device areas. In particular, solar cells based on organic polymers, e.g., can be produced from liquid solutions with processes similar to printing technology, which opens a way to very low-cost large-scale roll-to-roll production [5]. Overall, simpler and lower cost production methods and often cheaper, more abundant, or environmentally unproblematic materials are the main driving factors for development and optimization of thin-film PV concepts. Still, these technologies have to compete with costs of traditional PV, other renewable energy sources, and most of all with that of fossil energy sources in order to penetrate the energy market. In 2014, thin-film technologies accounted for roughly 9% of the global PV module production [6].

The "classical" thin-film solar cell technologies include the second generation cells like amorphous and microcrystalline silicon (a-Si, μc-Si), the chalcogenide alloy copper indium gallium diselenide (CIGS), and cadmium telluride (CdTe), which together count up for almost all of the current thin-film solar cell market.

Although the record efficiency map regularly published by the National Renewable Energy Laboratory (NREL) [7] groups only these technologies under the term "thin-film," we assume a broader interpretation by including and actually focusing on emerging concepts. These concepts are organic-based (bulk) heterojunction cells (organic photovoltaics [OPV]), dye solar cells (DSCs), and Perovskite-based cells. Recent reviews on the state of the art of these emerging technologies, also known as third-generation cells [8], can be found in [9–11].

Modeling of thin-film solar cells requires three essential ingredients, namely a model for the optical absorption, i.e., carrier or charge generation, for the charge separation, and for electronic transport. While the details of the latter depend mostly on the material system and technology of a given cell, the optical absorption is particularly critical for thin-film devices compared to devices employing thick absorbers. This is because they typically consist of several layers of different materials with thicknesses comparable to or smaller than the optical wavelength. Therefore, the optical absorption profile can be significantly determined by interference effects, so that the calculation of the optical absorption requires basically the explicit solution of the full Maxwell equations. Even if the optical absorption in some planar devices with thicker absorber layers, like DSCs, may be calculated with simplified approaches, the inclusion of light management–improving measures like scattering elements or back contact gratings usually requires a scheme based on the solution of Maxwell's equations, too.

Most of the models described in literature are using for the optical modeling constant (but material and wavelength dependent) bulk absorption coefficients, such that the optical absorption can be calculated once, providing the carrier generation profile for the electronic transport simulation. Analytical models for carrier transport can be found under suitable simplifying assumptions, as described e.g., in [1], and such models are very useful to understand basic features of solar cell behavior. Here, however, we are interested in physics-based simulation beyond analytical approximations, which provides more detailed insight into device operation and which, in particular, can be used for device optimization. Moreover, analytical models can typically be formulated only for one-dimensional (1D) problems, while numerical simulations can treat higher dimensions. This allows for simulation-based optimization of the device geometry, e.g., related to contacting schemes, which is hardly possible using only 1D simulations.

The chapter is organized as follows. First, we give a short overview on the basic features of the different thin-film technologies. Then we review the basic physics-based modeling approaches typically used for thin-film solar cells, both for optical absorption and electronic transport. Then we will describe critical details and modeling issues for the different third-generation thin-film concepts. More advanced concepts based on low-dimensional structures, hot carriers or using plasmonic effects, which generally require more complex simulation models, are not covered in this chapter.

43.2 Technology Overview

In the following sections, we shortly review the main characteristics of different thin-film technologies. For more in-depth discussions and further references, we refer, e.g., to [2,12].

43.2.1 Silicon-Based Cells

The currently most important silicon thin-film technologies are based on amorphous (a-Si) and microcrystalline (μc-Si) form. In contrast to crystalline silicon, which forms a regular tetrahedral configuration in a diamond lattice such that every silicon atom has four neighboring atoms, amorphous silicon does not show the same regular structure. The tetrahedral coordination of the atoms is mostly maintained, but random variations in bond angles and random appearance of dangling bonds, i.e., atoms having less than four neighbors, destroys the regular lattice and the associated periodicity.

The dangling bonds in amorphous silicon lead to a large number of defect states inside the bandgap, which adversely influence its electronic properties. In particular, the material can hardly be doped.

However, passivation of at least part of these dangling bonds by hydrogen atoms, known as *hydrogenation*, has been found to be necessary and sufficient to improve considerably material quality and allow it to be doped [13]. The hydrogenated form of a-Si is known as a-Si:H.

μc-Si (or more correctly μc-Si:H) is a more complex material, consisting of amorphous phases and conglomerates of silicon nanocrystals. Their microstructure depends largely on deposition methods and underlying layers, and the presence of grain boundaries between crystalline domains will influence further electronic properties.

As expected, the electronic properties of such disordered materials are worse than those of crystalline silicon. In particular, carrier mobilities are much lower. However, both materials are good enough to build working solar cells.

Both a-Si and μc-Si can be produced by a plasma-enhanced chemical vapor deposition (PECVD) technique introduced around 1970 [13,14]. The advantages of this technique over crystalline silicon deposition methods is that the required substrate temperature is significantly lower (<600 °C) and that it can be employed for large-area deposition. The former allows the use of cheap substrate materials such as plastic, while the latter allows to process large modules at once. Moreover, a-Si and μc-Si have two material intrinsic advantages. First, the loss of long-range order effectively leads to a transition from an indirect bandgap to a direct one, resulting in higher optical absorption coefficient. This allows to use thinner absorber layers without loss of absorption and thus to a cost reduction. Second, because of the different electronic transport properties these cells actually perform better at higher temperatures, contrary to crystalline silicon solar cells.

Typically, a-Si and μc-Si solar cells are built as p-i-n structures as shown in Figure 43.1a rather than pn-junctions, i.e., with an intrinsic layer between the n- and p-doped contact layers, since the carrier diffusion length in doped material are rather short due to the relatively poor transport properties. The transparent conductive oxide (TCO) layer in Figure 43.1a is e.g., indium tin oxide (ITO) or SnO_x. This layer is common to all thin-film solar cell technologies and serves to achieve a sufficiently low front contact resistivity without optical absorption. Cell thickness is usually rather small (around 400 nm for a-Si, few μm for μc-Si) in order to reduce recombination losses. Current best research cell efficiency is around 10% for a-Si:H and 12% for μc-Si single junction cells, compared to 25% of single-crystal silicon cells [15].

One of the most important problems in hydrogenated silicon cells at current is a light-induced performance degradation by prolonged illumination, known as the Wronski–Staebler effect, which is believed to be related to light-induced diffusion of the hydrogen atoms [16].

A recent review on thin-film Si solar technology can be found, e.g., in [17].

43.2.2 CdTe-Based Cells

CdTe is a very suitable material for PV applications. It has a direct bandgap of $E_g \approx 1.45$ eV, which is in the optimum range of values, and it has a high optical absorption coefficient $>10^4$ cm^{-1} over the whole solar spectrum, allowing for thin absorbers.

Both monocrystalline and polycrystalline CdTe cells can be produced. Polycrystalline CdTe can be deposited by a vast range of techniques, including vacuum techniques such as sublimation, evaporation, atomic layer deposition, chemical vapor deposition, but also nonvacuum techniques based on solution processes. Low-temperature processes are especially interesting since they allow for deposition on cheap and flexible substrates and for large-scale roll-to-roll process technology. Deposition is followed by a $CdCl_2$ treatment to improve material quality.

Some disadvantages of CdTe are the inability to tune the bandgap by alloying, and the presence of native defects to an amount that makes it difficult to achieve high-doping levels. The latter is amended by smart contacting schemes. At the front contact, an n-doped CdS buffer layer is introduced before the p-type CdTe absorber layer, which also acts as a buffer layer for CdTe deposition. The back contact is challenging because of the impossibility to achieve high p-doping density and since CdTe has a high electron affinity leading to high Schottky barriers for usual contact metals. Several solutions exist, e.g., suitable combinations of

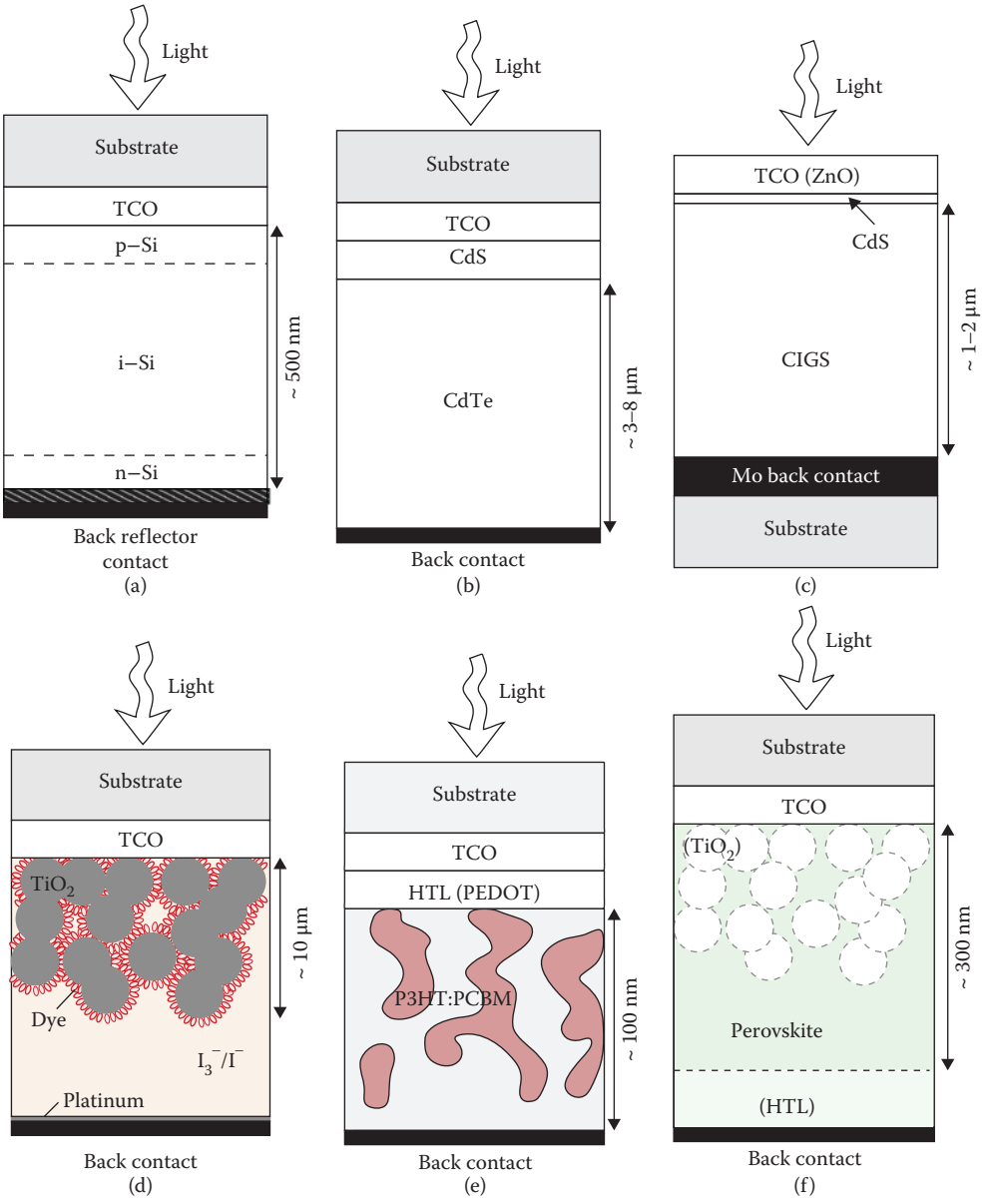


FIGURE 43.1 Typical schematic structures of different thin-film solar cells. (a) Amorphous silicon, with intrinsic absorber layer (i-Si) and doped contact layers (n/p-Si); (b) cadmium telluride (CdTe); (c) copper indium gallium selenide (CIGS); (d) dye solar cell (DSCs): TiO_2 nanoparticles are in the order of 10 nm diameter; (e) organic solar cell: the characteristic dimension of the absorber blend is in the order of 10 nm; and (f) perovskite cell, in some versions including a mesoporous TiO_2 scaffold as used in DSCs. Typical absorber thicknesses are indicated in the figures. TCO is the transparent conducting oxide, like indium tin oxide (ITO) or zinc oxide (ZnO), HTL is the hole transport layer.

buffer layers with metallization or particular surface treatments. Transport and recombination properties in CdTe films depend much on deposition technology and on the properties of the polycrystallinity of the film. The best research cell efficiency reported so far is 21% [15]. A typical CdTe solar cell scheme is shown in Figure 43.1b.

CdTe thin-film solar cells have been commercialized successfully and in 2014 they covered roughly 44% of the market share of all thin-film technologies [6]. It has to be noted that the presence of cadmium in CdTe has induced some doubts about the environmental compatibility of this technology. Recent reviews can be found in [18,19].

43.2.3 Copper Indium Gallium Selenide–Based Cells

Chalcopyrite-based solar cells is the second most important thin-film technology nowadays with a market share of 39% [6]. The most relevant compounds of this material class for PV applications are $CuInSe_2$, $CuInS_2$, and $CuGaSe_2$ having a bandgap of 1.0, 1.5, and 1.7 eV, respectively. These compounds can be easily alloyed, providing an extended range of bandgaps and lattice constants. Typically, the absorber consists of a $Cu(In,Ga)(S,Se)_2$ alloy, giving the name "CIGS," or of $CInS_2$ which has a gap of 1.5 eV. The highest cell efficiency reported in 2015 was around 21%, higher than multicrystalline Si cells [15]. Beside significantly higher efficiency than a-Si thin-film cells, CIGS cells show better stability and have no light-induced degradation.

A typical CIGS cell is shown in Figure 43.1c [20]. The nontransparent back contact is typically molybdenum sputtered onto a suitable substrate. Then follows a polycrystalline p-type CIGS absorber layer of few micrometer thickness. A heterointerface with a very thin n-type CdS is subsequently formed, followed by a TCO-like heavily doped ZnO as front contact and window layer.

Like CdTe and a-Si, CIGS cells can be produced on flexible substrates, allowing for specialized application fields. Reviews can be found, e.g., in [21,22].

43.2.4 Dye-Sensitized Solar Cells

A dye-sensitized solar cell (DSC) is an electrochemical system where the absorption and charge transport components are spatially separated [23]. The standard device is made of a mesoporous oxide, usually titanium dioxide (TiO_2), composed of nanoparticles of 15–20 nm in diameter. The surface of the mesoporous oxide is covered by a monolayer of molecules, the dye, conventionally organometallic complexes like the N719 [24] based on ruthenium. Finally, the oxide is sourrounded by a liquid electrolyte containing a redox pair. The most common one is triiodide/iodide pair (I_3^-/I^-), but many other redox pairs have been tried. In particular, cobalt-based redox pairs [25] have shown interesting properties. The contacts also are made of a glass with a transparent conducting oxide layer like ITO or fluorine-doped tin oxide (FTO) on the TiO_2 side, the anode, while the cathode is a thin layer of platinum deposited on a metallic substrate. In the device, the dye is the active material which absorbs light, while the redox pair in the electrolyte and the TiO_2 are the hole and electron transporting materials, respectively. The mesoporous structure of the TiO_2 is necessary in order to increase the surface for the dye and thus enhance dramatically the absorption efficiency, since necessarily only a monolayer of dye molecules can be used in order to efficiently extract the electrons and regerenate the dye.

The photoconversion in a DSC is far more complex than in conventional PV devices. After the absorption of the photon by the dye, the electron is transferred to the TiO_2 by an extremely fast charge transfer. Then the electron percolates in the mesoporous TiO_2 until it is collected by the ITO or FTO at the anode. The oxidized dye is regenerated by an electrochemical process with the redox pair. In a standard DSC this means that iodide is consumed and triiodide is generated nearby of the dye. The process generates an excess of triiodide with respect to the equilibrium concentration of the I_3^-/I^- couple in the bulk electrolyte. This concentration gradient moves triiodide toward the cathode where it is transformed back into iodide by taking two electrons from the catalyst, the platinum, closing the circuit. The platinum takes electrons directly from the external circuit.

There are several advantages of a DSC: it is a thin-film device, with a thickness in the range of tens of micrometer; it is quite cheap as the most expensive components (platinum and dye) are used in tiny

amounts (few nanometer in thickness and a single molecular layer, respectively); the decoupling of the absorption and transport components strongly reduces recombination processes as electron and hole pairs are immediately separated on different materials at the very beginning of the photoconversion process; and fabrication processes are very cheap.

The main drawback of the device is the liquid component that can easily leak out from microcracks in the encapsulation material or be polluted by the external environment. Moreover, the most common redox pair, triiodide/iodide, is highly corrosive leading to problems in the metals used to make the external contacts and degradation of the dye layer due to desorption of molecules from the surface. All these effects are obviously detrimental for the life span of the device. In order to correct these problems, an alternative DSC has been proposed [26], where the liquid electrolyte for hole transport is substituted by an organic semiconductor (e.g., Spiro-OMeTAD or P3HT [27]). The advantage is that the new hole transporter is more stable and noncorrosive; however, the charge transport in these organic semiconductors is worse leading to higher recombinations at the interface. Moreover, the intermixing between the organic material and the mesoporous TiO_2 is less efficient. DSCs have reached 12% of confirmed efficiency [15]. Recent reviews can be found in [28,29].

43.2.5 Organic Solar Cells

OPV cells are classified as organic because their active layer consists entirely of carbon-based materials, such as polymers or small organic molecules. A major advantage of OPVs is their compatibility with techniques suitable for large-scale and low-cost fabrication, such as roll-to-roll processing. The ability to chemically engineer organic materials allows to create a variety of compounds suitable for OPV applications. Especially due to the high absorption coefficient of specific polymers, very thin layers of 50–300 nm provide sufficient absorption and make OPVs a favorable choice for thin-film solar cells. Additional features, such as mechanical flexibility and low weight, open up the possibility for a variety of applications.

A peculiarity of the organic materials used for OPVs is their low permittivity ($\varepsilon_r \approx 3-4$), which leads to a weak screening of electrostatic interactions. On absorption of a photon, an electron–hole pair is created. Due to the strong Coulomb interaction between the electron and the hole, the pair forms an exciton with binding energies of several hundred millielectron volts (meV). The thermal energy of 25 meV at room temperature is not large enough to split the exciton and separate the charges. Therefore, the concept of a heterojunction composed of a (electron) donor, usually a polymer or small molecule, and an (electron) acceptor was introduced. The most prominent acceptor material is a derivative of the Buckminsterfullerene, [6,6]-phenyl-C61 butyric acid methyl ester (PCBM), because its large electronegativity has shown major improvement in exciton dissociation [30].

The first organic solar cell with efficiency of >1% was developed by Tang [31] and consisted of a planar heterojunction between donor and acceptor materials. Planar architectures have the disadvantage of a low exciton splitting efficiency, since excitons typically have a very small diffusion length and thus only excitons created nearby the junction can diffuse to the interface and become separated before recombination. A boost in device performance was achieved by introducing the concept of a bulk heterojunction (BHJ) [32], which is an intermixed morphology between donor and acceptor materials, as indicated schematically in Figure 43.1c. It provides a trade-off between efficient exciton splitting and good charge transport. The spatial structure of the BHJ is crucial for device performance and is still object of intense study.

The pioneering organic devices nowadays are based on the polymer/fullerene bulk-heterojunction concept with efficiencies of up to 10.8% [15,33]. Although much progress has been achieved in the last decade, higher efficiencies are needed in order for organic solar cells to be commercially competitive with established (inorganic) technologies. Since the experimental investigation of OPVs during operation is difficult, simulations can give valuable insights into the still not fully understood charge generation and transport processes.

More details and reviews on OPV can be found, e.g., in [9,34,35].

43.2.6 Perovskites-Based Cells

Since the first attempts in 2009 and a breakthrough in 2012 regarding efficiency and stability [36], the organometal lead halide perovskite $CH_3NH_3PbX_3$ (with X = Cl, Br, or I) has attracted great attention as absorber material in PV devices. Originally, perovskites have been used as alternative dyes in DSCs, but then perovskite-based solar cells started forming a PV technology on their own. Structurally, these perovskites consist of an ionic cage built from PbX in a tetragonal structure, with an organic methylammonium (CH_3NH_3) molecule at the center [37].

One of the most astonishing properties of perovskite-based solar cells is the fact that perovskite layers can be very easily deposited with simple solution processes, and at the same time easily reaching high photoconversion efficiencies. In fact, recent perovskite cells show efficiencies comparable to the best CdTe and CIGS cells, and thus much higher than all other emerging PV concepts [15]. Such a good performance is due to good absorption properties and good, balanced electron and hole transport properties.

A further useful property of methylammonium lead halide perovskites ($MAPbX_3$) is the ability to tune the bandgap by alloying $MAPbI_3$ with $MAPbBr_3$, which allows to extend the absorption spectra up to 800 nm.

Since perovskite solar cells originally emerged from the DSC technology, their structure often resembles a DSC. Usually, a perovskite layer of few hundred nanometer thickness is deposited on a TiO_2 layer, with or without mesoporous phase, followed by a hole transport layer (HTL), as shown in Figure 43.1f. However, the latter may even be omitted [38].

Substantially, perovskite technology is "dirty," cheap but performing at the same time, reaching efficiencies as high as 20% at cell level [15]. However, it struggles with several open issues. Most importantly, stability is still very poor [39], which is mostly due to the sensitivity of the $MAPbI_3$ to humidity [40]. Second, often the measured current–voltage characteristics show pronounced hysteresis, which is nowadays believed to be due to migration of charged defects, especially lead vacancies, in the perovskite. This poses difficulties for the accurate experimental evaluation of cell efficiencies. Finally, toxicity of lead halide perovskites may be an issue. In fact, lead compounds are toxic, and therefore often banned by law and certainly unpopular in industry. Identification of lead-free perovskites, however, did not yet lead to satisfactory candidates. It seems, that lead-based halides provide best performance.

Recent reviews on perovskite-based solar cells can be found in [37,38,40,41].

43.3 Optical Modeling

The calculation of the optical field distribution in a solar cell is a critical modeling step for any solar cell, as it provides the carrier generation profile, which is the most important factor determining the photocurrent. The generation of carriers is related to the optical absorption of a material for a given wavelength λ and can be written as [42,43]

$$G(\lambda; x) = \alpha(\lambda; x) I(\lambda; x) \tag{43.1}$$

where G is the carrier generation rate per wavelength interval at a given incident free-space wavelength λ at position x, usually given in units of $cm^{-3}s^{-1}nm^{-1}$; α is the optical absorption coefficient at wavelength λ, with units cm^{-1}; and $I(\lambda; x)$ is the local light intensity given as photon flux in $cm^{-2}s^{-1}nm^{-1}$. In Equation 43.1, we assume that every absorbed photon produces a carrier, which may be an unbound electron–hole pair or an exciton. The total generation profile is then obtained by summing the contributions of all relevant wavelengths present in the incident light. The knowledge of the local absorption requires the knowledge of the local wavelength-dependent absorption coefficient, which is usually taken as a material constant, and of the local intensity spectrum. The latter is related to the local strength of the

electromagnetic field, which can be written as [44]

$$I(\lambda; x) = \frac{\lambda \varepsilon_0 n}{2h} |E(\lambda; x)|^2 \tag{43.2}$$

where ϵ_0 is the vacuum permittivity, n the refractive index, h Planck's constant, E the amplitute of the electric field, and λ the free space wavelength. Therefore, the knowledge of the electromagnetic field distribution in a solar cell device allows to calculate the local generation rate G.

While the local electromagnetic field is accessible only via simulation, it is possible to both calculate and measure the absorption spectrum α. The usual way to calculate α is by using Fermi's golden rule and evaluating the optical transition matrix elements between the involved electronic states [1,45]. In this approximation, the transition probability between two states i and f with different energies E_i and E_f is given by

$$P_{i \to f} = \frac{2\pi}{\hbar} |\langle i|H'|f\rangle|^2 \delta(E_f - E_i \pm \hbar\omega), \tag{43.3}$$

where H' is the Hamiltonian describing the light-matter interaction, treated as a perturbation, and $\hbar\omega$ is the photon energy. $\langle i|H'|f\rangle$ in Equation 43.2 is the matrix element coupling initial and final states i and f, and its calculation requires an explicit representation of these states in some suitable basis. The delta function is an expression of the conservation of energy, and the minus sign describes photon absorption, while a positive sign describes photon emission, i.e., radiative recombination. The net rate of photogeneration is then given by the product of transition probability with the number of occupied initial and available final states, and by subtracting emission from absorption. The occupation of the states is given by the carriers' Fermi–Dirac distribution functions f_i and f_f. Dividing the net rate by the incident photon flux leads to the absorption coefficient [45]

$$\alpha(\hbar\omega) = \frac{\pi e^2}{n_r c \varepsilon_0 m_0^2 \omega V} \sum_{\substack{\mathbf{k}_i \mathbf{k}_f \\ \sigma_i, \sigma_f}} |\hat{e} \cdot \mathbf{p}_{fi}|^2 \delta(E_f - E_i - \hbar\omega)(f_i - f_f) \tag{43.4}$$

Here, e is the elementary charge, n_r is the relative dielectric constant, ε_0 is the permittivity, m_0 is the electron mass, V is the normalization volume, and the optical matrix element has been substituted by the momentum matrix element applying the dipole approximation [45]. The summation has to be done over the crystal momentum $\mathbf{k}$ and all quantum numbers σ of both initial and final states. For absorption and at not too high optical excitation, one can usually assume all initial states as occupied and all final states as empty, i.e., $f_i \approx 1$ and $f_f \approx 0$.

Radiative recombination is obtained in a similar way; however, in that case only transitions between occupied electron and hole states are relevant. Therefore, the summation can be restricted to states with energies near the band extrema. For absorption, however, all transitions are relevant, and therefore the electronic states have to be known with sufficient accuracy for all energy bands and in the whole reciprocal k-space (or more apropriately in the full first Brillouin zone). This makes calculation of absorption spectra more difficult, so that usually measured spectra are used.

In principle, the local electric field strength at all relevant wavelengths can be obtained by solving Maxwell's equations using one of the standard methods such as finite difference time domain (FDTD) [46], finite elements (FEM) [47], or rigorous coupled-wave analysis (RCWA) [48], but this is not always necessary. For a homogeneous thick absorber, in the absence of coherent effects, the Beer–Lambert approximation can be used [1]. It allows to calculate the local optical intensity simply considering the absorption along the optical path, starting from Equation 43.1. Assume a monochromatic light beam with wavelength λ with intensity I traveling in a material with absorption coefficient α. When it travels an infinitesimal

distance dx, the intensity of the beam reduces by $\alpha I \mathrm{d}x$, since a part of the photons are absorbed. We can therefore deduce a continuity law for the light intensity given by

$$\frac{\mathrm{d}I}{\mathrm{d}x} = -\alpha I \tag{43.5}$$

The solution to this is an exponential function, describing exponential decay of the intensity known as Beer–Lambert law, $I(x) = I_0 \mathrm{e}^{-\alpha x}$, where I_0 is the intensity of the incident beam (after subtraction of reflection losses, i.e., immediately inside the device) and x the distance from the point of incidence. In the case of spatially inhomogeneous absorption, the result is easily generalized to

$$I(\lambda; x) = I_0(\lambda) \mathrm{e}^{-\int_0^x \alpha(\lambda; x)\, \mathrm{d}x} \tag{43.6}$$

Equation 43.6 together with Equation 43.1 are used succesfully for the simulation of solar cells where layer thicknesses are larger than the optical wavelength. However, since this model is based on a ray optics description of light, it will break down whenever effects like interference or diffraction are important. In that case a more rigorous approach is needed to calculate the optical intensity profile.

Under the assumption of a flat, planar layered structure, a more precise model can be formulated, without increasing much the computational efforts. This model is based on a 1D approximation along the optical path of a plane wave, including sequential reflections at material interfaces and transmissions in homogeneous portions of the device, and is known as the *Transfer Matrix Method* (TMM) [49,50]. Today, this is probably the most used model for one-dimensional optical modeling of thin-film devices. The model has been first applied to organic thin-film solar cells in the late 1990s [51–53], and since then it has become a standard tool [54–60].

The basic principle of the TMM is shown schematically in Figure 43.2. To apply the method, we assume that the device can be sliced into layers of homogeneous properties with flat, parallel interfaces. We do not include a possible thick substrate that should be treated with an incoherent method (details on this can be found, e.g., in [44]). Further, we assume a monochromatic optical field written as $E(x) \exp(i\omega t)$, where $E(x)$ is a complex local field amplitude, where x is the coordinate along the stack, with $x = 0$ being the point of incidence. For simplicity, we assume a transverse electromagnetic (TEM) field with orthogonal incidence here, and in the following we drop the harmonic time dependence.

The amplitudes of the electric and magnetic field at each position can be found by using two relations derived from Maxwell's equations. The first relates the fields at the left interface of a layer to that at the right interface by means of the free propagation in a homogeneous medium, while the second relates the fields at the left and the right of each interface by means of continuity conditions. Maxwell's equations in

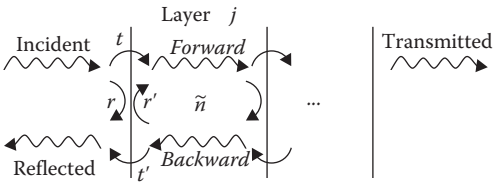


FIGURE 43.2 Scheme of the transfer matrix method. The device is assumed to be sliced in homogeneous layers, and monochromatic light is incident from the left. We decompose the optical field in a forward and backward traveling component. At each interface between two layers, these two components undergo reflection and transmission, r and t, while inside the layer the waves propagate freely with propagation constant according to the complex refractive index $\tilde{n}$.

homogeneous form (without source terms) are given by [49,61]

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \cdot \mathbf{D} = 0 \tag{43.7}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \quad \nabla \cdot \mathbf{B} = 0 \tag{43.8}$$

$$\mathbf{B} = \mu \mathbf{H}, \quad \mathbf{D} = \varepsilon \mathbf{E} \tag{43.9}$$

Here, **E**, **H**, **D**, and **B** are the electric field, the magnetic field (or magnetizing field, magnetic field strength), the electric displacement, and the magnetic flux density (or magnetic induction, magnetic field), respectively. The two electric fields and magnetic fields are related by the constitutive relations (Equations 43.7 through 43.9), where ε and μ are the electric permittivity and the magnetic permeability, respectively. The latter is usually equal to the vacuum permeability μ_0 for the materials of interest in this context. Since we are interested in time harmonic fields, Equations 43.7 through 43.9 are conveniently transformed to Fourier space, leading to

$$\nabla \times \mathbf{E} = -i\omega \mathbf{B} \tag{43.10}$$

$$\nabla \times \mathbf{H} = i\omega \mathbf{D} \tag{43.11}$$

while the other expressions remain the same. Note that the fields are now complex quantities. Moreover, the permittivity, which we write as product of vacuum permittivity ε_0 and relative permittivity $\tilde{\varepsilon}$, is frequency dependent (dispersive) and complex in case of lossy, i.e., absorbing materials.

In a homogeneous medium, the solution of Equations 43.7 through 43.9, or equivalently Equations 43.10 and 43.11, is found to be a propagating wave $E(x) = E_0 \exp(\pm ikx)$, where k is the propagation constant (in our case the component along x). The latter is related to the relative permittivity $\tilde{\varepsilon}$, or, equivalently, to the refractive index $\tilde{n} = \sqrt{\tilde{\varepsilon}}$ by [43]

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c} = \frac{\omega\sqrt{\tilde{\varepsilon}}}{c_0} = \frac{2\pi}{\lambda_0}\tilde{n} \tag{43.12}$$

where c_0 and λ_0 are the vacuum light speed and the free space wavelength, respectively; $\omega = 2\pi f$ is the angular frequency, and we have used the expression for the light speed $c = 1/\sqrt{\varepsilon\mu}$. Note that the refractive index $\tilde{n} = n_0 - i\kappa$ is a complex quantity, where the imaginary part is related to absorption.

It is convenient to decompose the electric field into *forward* and *backward* travelling components E^+ and E^-, as indicated in the figure. This is written as

$$E(x) = E^+(x) + E^-(x) \tag{43.13}$$

where once again all amplitudes are complex quantities.

We can now relate the field amplitudes at the left and right interface of the layer j by

$$E^+_{j;\mathrm{R}} = E^+_{j;\mathrm{L}} \mathrm{e}^{-ik_j t_j} \tag{43.14}$$

$$E^-_{j;\mathrm{R}} = E^-_{j;\mathrm{L}} \mathrm{e}^{ik_j t_j} \tag{43.15}$$

where k_j and t_j are the propagation constant and thickness of layer j, respectively. This can be put into matrix form as

$$\begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{j;\mathrm{R}} = \mathbf{M}_j \begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{j;\mathrm{L}}, \quad \mathbf{M}_j = \begin{pmatrix} \mathrm{e}^{-ik_j t_j} & 0 \\ 0 & \mathrm{e}^{ik_j t_j} \end{pmatrix} \tag{43.16}$$

where we have defined the transmission matrix of layer j as $\mathbf{M}_j$. Comparing Equation 43.16 with Equation 43.6 and considering Equation 43.2, the relation between $\mathrm{Im}(k)$, commonly called *extinction coefficient*, and the absorption coefficient α can be immediately found as

$$\alpha = \frac{4\pi}{\lambda}\kappa \tag{43.17}$$

Based on Maxwell's equations, one can derive the boundary conditions for the in-plane and orthogonal components of the electric and magnetic fields at planar material interfaces (orthogonal to the x-axis) as shown in Figure 43.3, which are well known and given by

$$\mathbf{n}_{lm} \times (\mathbf{E}_m - \mathbf{E}_l) = 0 \tag{43.18}$$

$$\mathbf{n}_{lm} \times (\mathbf{H}_m - \mathbf{H}_l) = 0 \tag{43.19}$$

i.e., the tangential components of **E** and **H** need to be continuous. In the above equations, $\mathbf{n}_{lm}$ denotes the normal on the interface between layers l and m. The orthogonal components of **D** and **B** have to be continuous, which is not relevant in our case of normal incidence.

Furthermore, using Equation 43.10 we can express the magnetic field in terms of the electric field as

$$B^{\pm} = \pm\frac{\tilde{n}}{c_0}E^{\pm} \tag{43.20}$$

which is obtained by evaluating the curl of the electric field and using Equation 43.12. With this we can rewrite Equations 43.18 and 43.19 in terms of the forward and backward traveling electric field components as

$$E_l^+ + E_l^- = E_m^+ + E_m^- \tag{43.21}$$

$$\tilde{n}_l(E_l^+ - E_l^-) = \tilde{n}_m(E_m^+ - E_m^-) \tag{43.22}$$

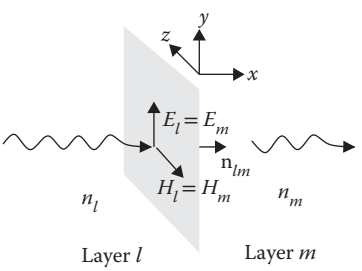


FIGURE 43.3 A material interface and the respective boundary conditions for the in-plane field components.

which can be rearranged such as to provide the transfer matrix of the material interface

$$\begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{l,\mathrm{R}} = \underbrace{\frac{\tilde{n}_l + \tilde{n}_m}{2\tilde{n}_l} \begin{pmatrix} 1 & \frac{\tilde{n}_l - \tilde{n}_m}{\tilde{n}_l + \tilde{n}_m} \\ \frac{\tilde{n}_l - \tilde{n}_m}{\tilde{n}_l + \tilde{n}_m} & 1 \end{pmatrix}}_{\mathbf{I}_{lm}} \begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{m,\mathrm{L}} \tag{43.23}$$

In the expressions for the elements of the transfer matrix $\mathbf{I}_{lm}$, we can identify the reflection and transmission coefficients of the interface between adjacent layers l and $m = l + 1$ with respect to the amplitudes

$$r = r_{lm} = \frac{E^-_{l;\mathrm{R}}}{E^+_{l;\mathrm{R}}} = \frac{n_l - n_m}{n_l + n_m} \tag{43.24}$$

$$t = t_{lm} = \frac{E^+_{m;\mathrm{L}}}{E^+_{l;\mathrm{R}}} = \frac{2n_l}{n_l + n_m} = 1 + r_{lm} \tag{43.25}$$

$$r' = r_{ml} = \frac{E^+_{m;\mathrm{L}}}{E^-_{m;\mathrm{L}}} = -r_{lm} \tag{43.26}$$

$$t' = t_{ml} = \frac{E^-_{l;\mathrm{R}}}{E^-_{m;\mathrm{L}}} = \frac{1 - r^2_{lm}}{t_{lm}} \tag{43.27}$$

so that we can write

$$\mathbf{I}_{lm} = \frac{1}{t_{lm}} \begin{pmatrix} 1 & r_{lm} \\ r_{lm} & 1 \end{pmatrix} \tag{43.28}$$

Combining Equations 43.16 and 43.28 we obtain the transfer matrix for the fields at the left interface between layers l and $m = l + 1$ as

$$\begin{pmatrix} E^+ \\ E^- \end{pmatrix}_l = \mathbf{M}_l \mathbf{I}_{lm} \begin{pmatrix} E^+ \\ E^- \end{pmatrix}_m \tag{43.29}$$

and finally the transfer matrix $\mathbf{T}$ relating the fields to the left and to the right of stack as a product

$$\mathbf{T} = \prod_l \mathbf{M}_l \mathbf{I}_{l,l+1} \tag{43.30}$$

Note that TMM can be implemented in several ways, and that it is inherently numerically unstable for larger layer thicknesses [62]. Stabilization is possible, e.g., by resorting to alternative representations of the transfer matrix [63].

As an example for the importance of interference effects in layered structures, we present a comparison of the local field strength and generation rate between results of a TMM calculation and those of an incoherent calcuation. For the latter, we considered the reflectance and transmittance at each material interface and summed up the intensities of the forward and backward traveling waves. The test structure, similar to the one used in [64], is shown in Figure 43.4a. For simplicity, we assumed total reflection at the aluminum back contact, and we did not consider the glass substrate. The results have been obtained for a wavelength of 500 nm. Figure 43.4b and c presents the field and generation profiles for two different

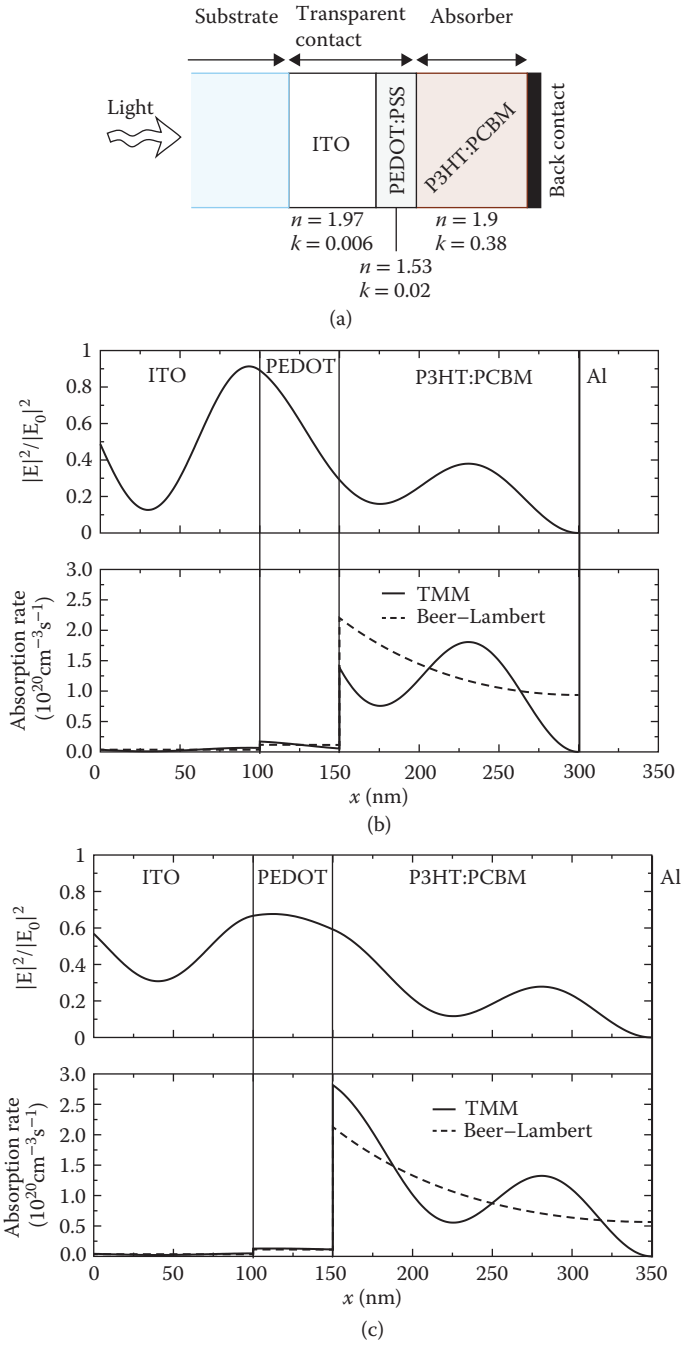


FIGURE 43.4 Structure for comparison of coherent and incoherent optical field calculation. The parameters used in the simulation are given in (a). In (b) and (c) the optical field intensity and absorption profiles for two different absorber thicknesses is shown. The local absorption rate considerably deviates from the Beer–Lambert prediction, in particular for thin layers. For thick layers Beer–Lambert behavior is recovered. (ITO, indium tin oxide; PCBM, [6,6]-phenyl-C61 butyric acid methyl ester;TMM, transfer matrix method.)

absorber thicknesses, showing that the profiles differ significantly from what would be expected from the Beer–Lambert model, and that it depends critically on layer thicknesses. In particular, absorption maxima can be located deeply inside the layer instead of near the interface. When the layers are becoming optically thick, the local absorption rate mediated over a wavelength converges to the result obtained by the Beer–Lambert law.

Figure 43.5 shows another example, borrowed from Reference [53], where the obtained results have been compared with experimental data. In this case, the donor–acceptor heterojunction is formed by the

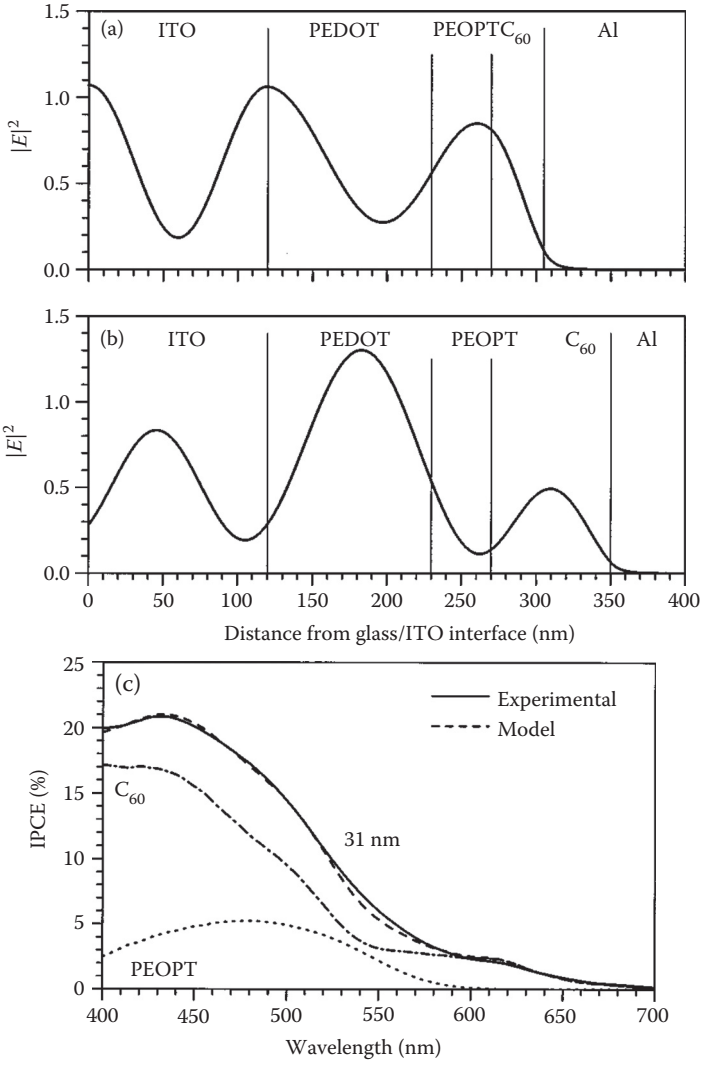


FIGURE 43.5 Field intensity profiles in an organic heterojunction solar cell for two different thicknesses of the C_{60} layer (a, b). The quality and accuracy of the optical simulation is proven by the comparison of experimental and simulated quantum efficiency in (c). Note that the simulation provides the information on where absorption happens, which cannot be obtained from the measurement. For example, in (c) the wavelength resolved contributions to the total absorption in the C_{60} and PEDOT layers is reported for an optimized C_{60} layer thickness. (Reproduced with permission from Leif A.A. Pettersson et al. Modeling photocurrent action spectra of photovoltaic devices based on organic thin films. *Journal of Applied Physics*, 86(1):487—496, 1999 Copyright 1999, American Institute of Physics Publishing LLC.)

PEOPT/C_{60} interface. To achieve maximum efficiency, a peak in field intensity should be obtained near this interface. Figure 43.3 presents the intensity profile at a wavelength of 460 nm for a C_{60} layer thickness of 35 and 80 nm, respectively, indicating that device optimization can benefit substantially from optical simulations. Note that the field profile is optimized in this way for a single wavelength only. Although the field profile cannot be measured directly experimentally, the accuracy of the optical simulation can be assessed by measuring the incident photocurrent efficiency (IPCE) (also denoted as external quantum efficiency, EQE), which is the product of quantum absorptivity and the current collection efficiency at each wavelength and gives the ratio of the number of collected electrons to the number of incident photons [43]. This quantity has been simulated in [53] by combining optical and current transport simulations, and the comparison between measurement and simulation is presented in Figure 43.3 for 31-nm thick C_{60} layer, showing excellent agreement. This example also shows how a simulation can give accurate insight into internal device behavior that is inaccessible by measurement alone, opening the way to efficient device optimization.

While 1D TMM-based modeling is a powerful tool to describe the optical field in a planar layered structure, it is less suitable to treat devices where light management techniques are employed, such as (possibly random) front or back surface texturing. Such approaches are becoming more and more important in order to approach the theoretical limits of light-to-electricity conversion, especially for technologies that are already optimized from an electrical point of view [65]. For organic solar cells, e.g., there is interest in grating structures on the back contact in order to modulate the electric field profile and thus to optimize optical generation and efficiency. Plasmonic effects are also studied for improving solar cell performance [66]. To accurately model texturing, gratings, or plasmonic effects, usually full 3D Maxwell solvers are employed [67,68]. Unfortunately, such calculations are computationally quite expensive, so that approximate models formulated in one dimension are also of interest [69]. Special approaches have been devised also for other cases, like thick layers with incoherent coupling [70].

Also note that from Figure 43.1 it might seem that TMM is not suited for devices based on mesoporous or blended materials, since the materials are not really piecewise homogeneous and there are apparently no planar interfaces. However, since the characteristic dimensions are in the order of at most tens of nanometers, these composite materials can be described with sufficient accuracy as effective materials so that TMM can be applied without particular restrictions.

43.4 Transport Modeling

While optical calculations are important for the simulation of solar cells, taken alone they can only provide upper limits for cell efficiency. For a comprehensive view on cell operation, the transport of the generated carriers and their collection at the electrical contacts has to be taken into account. This section reviews the two carrier transport models that are used most often for solar cell simulations, namely the drift-diffusion model and kinetic Monte Carlo (kMC).

43.4.1 Drift-Diffusion Model

Modeling of electronic transport in thin-film solar cells is mostly based on the drift-diffusion model, which has been the work-horse of physics-based electronic device simulation since decades. A thorough introduction, derivation of the model and implementation hints can be found in different textbooks such as [71], and a discussion of theoretical and mathematical aspects can be found, e.g., in [72]. A more detailed description is provided in Chapter 50.

Since the drift-diffusion model is well known and is described in numerous books and articles, we will provide only a short generic description and concentrate more on the features important for thin-film solar cell simulation.

For a "normal" semiconductor where only electrons and holes are mobile carriers, the drift-diffusion model consists of a set of three continuity equations for the electric displacement $D = -\varepsilon\nabla\varphi$ and the two particle currents J_n and J_p:

$$\nabla(\varepsilon\nabla\varphi) = q(n - p - C) \tag{43.31}$$

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla J_n = -R, \qquad J_n = q(D_n\nabla n - \mu_n n\nabla\varphi) \tag{43.32}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla J_p = \quad R, \qquad J_p = -q(D_p\nabla p + \mu_p p\nabla\varphi) \tag{43.33}$$

This nonlinear system of partial differential equations is also known as the *basic semiconductor device equations* or the van Roosbroeck system, who was the first to formulate this set of equations in the framework of electronic transport in semiconductors [73]. The first of the above equations (Equation 43.31) is the Poisson equation, relating the electrostatic potential φ with the total charge density $\rho = q(p - n + C)$ where n and p are the electron and hole densities, respectively, while ε is the permittivity. C denotes the net density of fixed charges such as dopants or traps. The other two equations are the continuity equations for the electron and hole currents, where the constitutive equations for the currents are given on the right (which are also known as *drift-diffusion equations*) where q is the electron's charge, $D_{n,p}$ are the electron and hole diffusion coefficients, and $\mu_{n,p}$ are the electron and hole mobilities. R is the net recombination rate, including all processes which lead to destruction or generation of electron–hole pairs.

The three most important recombination models are the trap-assisted (Shockley–Read–Hall [SRH]) recombination, bimolecular (direct, radiative) recombination, and Auger recombination, given by [71,74]

$$R_{SRH} = \frac{np - n_i^2}{\tau_p(n + n_t) + \tau_n(p + p_t)} \qquad \text{SRH} \tag{43.34}$$

$$R_{rad} = B(np - n_i^2) \qquad \text{Radiative} \tag{43.35}$$

$$R_{auger} = (C_n n + C_p p)(np - n_i^2) \qquad \text{Auger} \tag{43.36}$$

Here, n_i^2 is the square of the intrinsic carrier density, given by the product of the equilibrium densities n_0 and p_0. n_t and p_t are parameters depending on the energetic position inside the gap of the trap. Note that Equation 43.34 is valid for a single discrete trap level and in the absence of an electric field, however, it can be generalized to other cases [75,76]. The other parameters in the expressions are mostly material related. The relative importance of the different recombination mechanisms depends on the operating conditions and on material quality. SRH recombination is important at low carrier densities and in presence of defects, while Auger recombination usually becomes dominant only at high carrier densities.

Written in standard form (Equations 43.31 through 43.33), the particle currents are seen to be decomposed into a diffusion term proportional to the gradient of the carrier density and a drift component proportional to the conductivity times the electric field. This is an intuitive description of particle transport, whenever the system is dominated by scattering, that is when the particles undergo Brownian motion with some thermal velocity, and the external perturbation (electric field or density gradient) induces only a small nonzero mean carrier velocity leading to a net current flow.

The system of Equations 43.31 through 43.33, supplemented with suitable boundary conditions, is usually solved numerically, although under certain circumstances analytic solutions exist in one dimensions. For implementation details and numerical methods, we refer to [71] and the vast existing literature on the topic.

The device equations (Equations 43.31 through 43.33) are still the most important model for transport in semiconductor devices, since they can be solved much faster than any other transport model and

therefore represent a good compromise between computational efficiency and physical accuracy. While the equations allow easy interpretation, and can be solved relatively easily, they are based on a number of assumptions that are important when applying them to systems other than crystalline semiconductors. The other important point to note is that they give a good description of device behavior only when good models for the quantities appearing as parameters in Equations 43.31 through 43.33 can be provided. In particular, the models for the recombination and generation processes and for the carrier mobilities, in combination with accurate material-dependent parameters, are important in order to reach satisfactory results with the drift-diffusion model. Most often this implies careful parameter fitting against experimental data, careful checking whether the fitted parameters provide a predictive model or not.

In the following, we discuss some of the basic assumptions underlying the semiconductor device equations (Equation 43.31 through 43.33) and possible consequences when applying them to the materials of interest in this chapter. A discussion of technology-specific issues regarding the basic equations or the parameter models is deferred to Section 43.5.

The standard way to derive the drift-diffusion equations is to start from the semiclassical Boltzmann transport equation [72]. These derivations include implicitly or explicitly assumptions on the distribution function such that the carrier density takes the form of the well-known Boltzmann approximation, i.e.,

$$n = N_c e^{\frac{E_{F,n}-E_c}{k_B T}}, \quad p = N_v e^{-\frac{E_{F,p}-E_v}{k_B T}} \tag{43.37}$$

Here, N_c and N_v are the effective density of states (DOS) in the conduction and valence band, $E_{F,n}$ and $E_{F,p}$ are the electron and hole quasi Fermi levels, and E_c and E_v are the conduction and valence band edges. The latter can be expressed as $E_{c,v} = E_{c,v}^{(0)} - q\varphi$, where $E_{c,v}^{(0)}$ are material band edges in the absence of an electric field, identifiable with electron affinity and ionization energy, k_B is the Boltzmann constant, and T the temperature.

The diffusion constants and mobilities in Equations 43.31 through 43.33 are not independent, which can be easily seen when substituting Equation 43.27 into Equations 43.31 through 43.33 and solving for equilibrium, i.e., zero current under no external bias nor carrier generation. It is found that the two quantities are related by a so-called Einstein relation [74]:

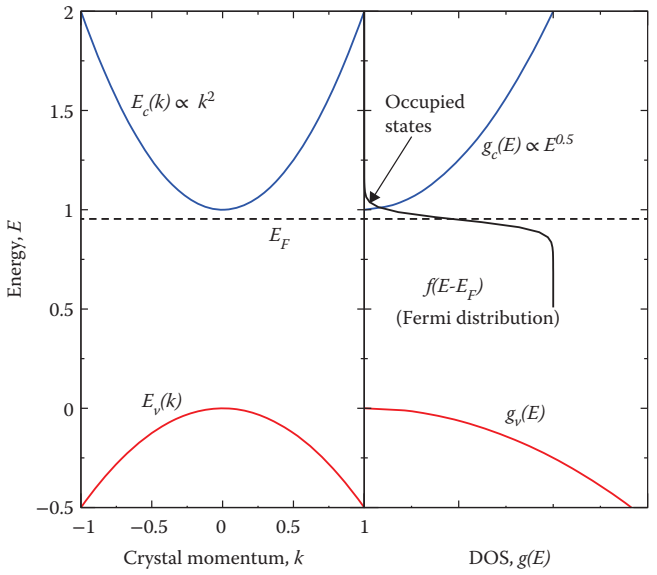
$$D_n = \frac{k_B T}{q}\mu_n, \quad D_p = \frac{k_B T}{q}\mu_p \tag{43.38}$$

which in actual implementations of the drift-diffusion model is usually used to obtain the diffusion coefficients from the mobilities.

The expressions (Equation 43.37) provide an adequate description of the carrier densities for most nondegenerate crystalline semiconductors, i.e., at not too high carrier densities. This is because the same expressions are obtained when calculating the carrier density from the DOS and the Fermi–Dirac distribution, and making the Boltzmann approximation for small densities. Generally, the density, e.g., of electrons is found under the assumption of a local equilibrium distribution by occupying the available states, given by the DOS, with a local quasi Fermi level and integrating over all energies, i.e.,

$$n(\mathbf{r}) = \int g(E) f\left(\frac{E - E_{F,n}(\mathbf{r})}{k_B T}\right) \mathrm{d}E \tag{43.39}$$

where $f(x)$ is the Fermi–Dirac distribution and $g(E)$ the DOS. The DOS of a crystalline semiconductor at the relevant energies, i.e., near the band extrema, is obtained from the energy dispersion, which is approximately parabolic in this range, as shown schematically in Figure 43.6. Details on the calculation of the carrier densities can be found, e.g., in [42,71,74,77].



$$E(k) = E_c^{(0)} + \frac{\hbar^2 k^2}{2m^*}, \quad f(x) = \frac{1}{1+\mathrm{e}^x}$$

$$g_c(E) = \frac{1}{\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\sqrt{E}$$

$$n = \frac{1}{(2\pi)^3}\int f\left(\frac{E_c(k)-E_F}{k_\mathrm{B}T}\right)\mathrm{d}k^3$$

$$= \int_{E_c}^{\infty} g_c(E) f\left(\frac{E-E_F}{k_\mathrm{B}T}\right)\mathrm{d}E$$

$$= \underbrace{2\left(\frac{m^* kT}{2\pi\hbar^2}\right)^{3/2}}_{N_c} F_{1/2}(E_F/k_\mathrm{B}T)$$

$$\approx N_c \cdot \mathrm{e}^{(E_F-E_c)/k_\mathrm{B}T}$$

FIGURE 43.6 Parabolic approximation to the energy dispersion $E(k)$ of a simple semiconductor and the DOS resulting from this dispersion, obtained by setting $1/(2\pi)^3 \int \mathrm{d}k^3 = \int g(E)\mathrm{d}E$. On the right, some of the expressions are given, and the calculation of the density is sketched ($m^* = \hbar^2(\partial^2 E(k)/\partial k^2)^{-1}$ is the effective mass, E_F is the quasi Fermi level).

The key point is that the derivations are based on the assumption that the carriers are moving freely in a periodic potential of the underlying crystal lattice, which allows to define the momentum k of the carriers and which leads to the parabolic dispersion valid at least in the vicinity of the band extrema. While this is a good approximation for crystalline semiconductors, it does not hold in many materials of interest in thin-film PV technology. In particular, whenever the periodicity of the crystal structure is lost, the DOS will become substantially different. In an amorphous semiconductor such as a-Si, e.g., the crystalline lattice structure is maintained only locally and only approximately: a-Si is mostly tetrahedrally bonded, but with a statistical distribution of bond angles and with certain number of silicon atoms with three or five neighbors. This modifies the DOS by washing out the band edge and appearance of approximately exponential band tails, known as Urbach tails [78], and appearance of localized states in the gap even in absence of impurity atoms, as shown schematically in Figure 43.7b. Similarly, organic semiconductors are usually characterized by a certain energetic disorder in the molecular energy levels, and the definition of carrier momentum is largely meaningless. In this case, the DOS is completely different from the "standard" semiconductor bulk DOS, and is often well described by a Gaussian distribution with a certain width σ characterizing the statistical scattering of the molecular levels [79], shown in Figure 43.7c.

A consequence of a DOS, which differs from that resulting from a parabolic dispersion, is that the Einstein relation relating the diffusion constants and the mobilities needs to be generalized. This general form can be easily found from Equations 43.31 through 43.33 as

$$D_\mathrm{n} = n\left(\frac{\partial n}{\partial \varphi}\right)^{-1}\mu_\mathrm{n}, \quad D_\mathrm{p} = -p\left(\frac{\partial p}{\partial \varphi}\right)^{-1}\mu_\mathrm{p} \tag{43.40}$$

An accurate description of semiconductor transport therefore requires to take the functional form of the DOS into account, although it does not appear explicitly in the basic semiconductor equations. Beside the formal modification of the Einstein relation, the DOS also influences some directly measurable quantities.

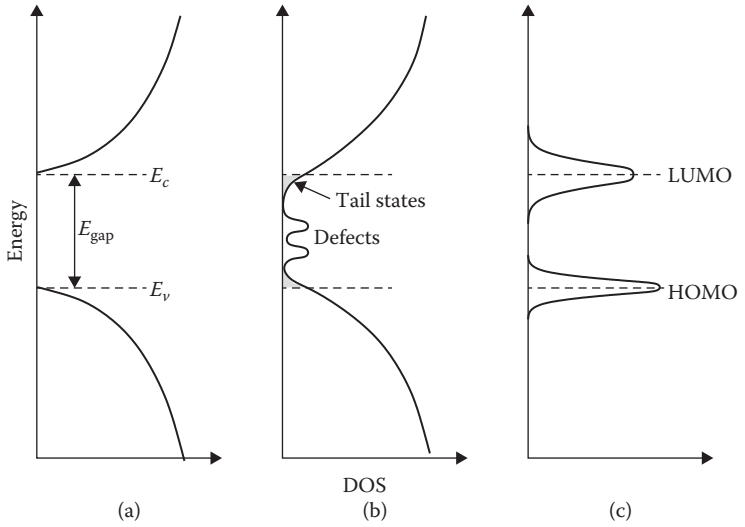


FIGURE 43.7 Sketch of different density of states (DOS): (a) crystalline semiconductor with parabolic band dispersion, (b) amorphous semiconductor with band tails and defect states, (c) disordered semiconductor, e.g., organic, with the DOS formed from energetic and spatial disorder of molecular orbital energies. In the last case, the bands are usually denoted as HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital).

Assume, e.g., a bulk absorber in flat band condition subject to homogeneous optical carrier generation, so that there is no current flow, and assume that only radiative recombination is relevant. In this case, all generated electron–hole pairs have to recombine radiatively according to Equation 43.35, i.e., $G = Bnp$. The splitting of the electron and hole quasi Fermi levels induced by the illumination is a measure for the open-circuit voltage V_{oc} that can be expected. Using Equation 43.37, an analytic formula can be easily found as [80]

$$V_{oc} = \frac{1}{q}(E_{F,n} - E_{F,p}) = \frac{1}{q}E_g + \frac{k_B T}{q}\ln\left(\frac{G}{BN_c N_v}\right) \tag{43.41}$$

and apparently the V_{oc} should vary by a constant rate of $2.3k_B T/q$ per decade in intensity. A different result is obtained if the same quantity is calculated using a different expression for the DOS, e.g., using a Gaussian DOS (GDOS) as typical for organic semiconductors and given by

$$g(E) = \frac{N_0}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(E-E_0)^2}{2\sigma^2}\right) \tag{43.42}$$

Here N_0 is the total density, σ the width (variance) of the Gaussian, and E_0 the center of Gaussian. Using approximations of the formulas for the densities [81], and assuming the same parameters for the electron and hole DOS, we can obtain the open-circuit voltage for sufficiently large carrier densities in this case as

$$V_{oc} = \frac{1}{q}E_g - 2\frac{\sigma\sqrt{2}}{qH(\sigma)}\mathrm{erfc}^{-1}\left(2\sqrt{\frac{G}{BN_0^2}}\right) \tag{43.43}$$

Here, E_g is defined as the difference of the center energies of the electron and hole DOS, and $H(\sigma) = k_B T\sqrt{2}/\sigma \cdot \mathrm{erfc}^{-1}(\exp(-\frac{1}{2}\sigma^2/(k_B T)^2))$. Figure 43.8 compares the derivative of V_{oc} with respect to the logarithm of the generation rate for a parabolic DOS and a GDOS, showing that for typical Gaussian widths

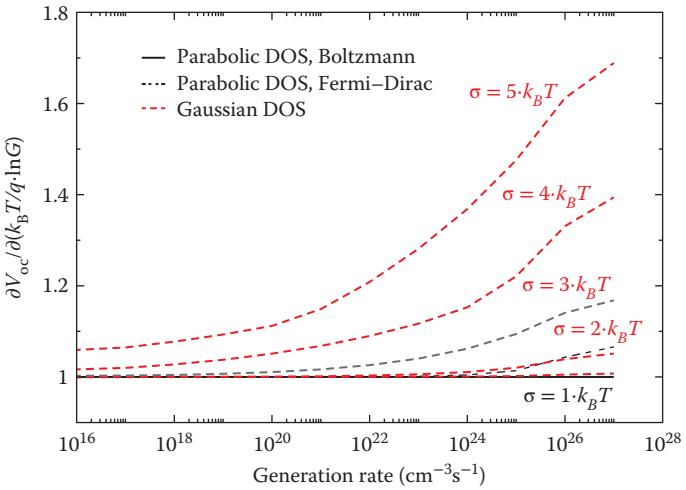


FIGURE 43.8 Variation of the bulk quasi Fermi level splitting due to a homogeneous generation rate. Deviation from the exponential form of the carrier density leads to considerable deviation from the expected rate of change. For the Gaussian density of states (DOS), $N_0 = 10^{21}$ cm^{-3} has been used, for the parabolic case $N_c = N_v = 10^{19}$ cm^{-3}. The recombination rate constant was $B = 10^{-10}$ cm^3·s^{-1}.

the DOS influences the rate even at moderate generation rates. The effect becomes stronger with increasing intensity. The measurable effect of energetic disorder is described, e.g., in [82].

Note that a further modification of the form (Equations 43.31 through 43.33) of the drift-diffusion equations is needed whenever the material parameters in the formula for the carrier densities (i.e., effective DOS or the band edges, see Figure 43.6), or generally the DOS, change spatially. This can be seen easily when evaluating the gradient of the densities, which then leads to additional terms that can be interpreted as effective potentials. These corrections are simple in the case of the standard exponential form (Equation 43.37) of the carrier densities, but they become more complicated in the degenerate case and for other DOS. It is therefore sometimes preferred to write the drift-diffusion model in a different but equivalent form, derived from thermodynamics [61,83]:

$$\nabla(\varepsilon\nabla\varphi) = q(n - p - C) \tag{43.44}$$

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla J_n = -R, \qquad J_n = -q\mu_n n\nabla\phi_n \tag{43.45}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla J_p = -R, \qquad J_p = -q\mu_p p\nabla\phi_p \tag{43.46}$$

Here, $\phi_{\{n,p\}}$ are the electrochemical potentials of the electrons and holes, respectively, related to the quasi Fermi levels by $E_{F,\{n,p\}} = -q\phi_{\{n,p\}}$. While in Equations 43.31 through 43.33, the dependent variables are the electrostatic potential and the carrier densities, in (Equations 43.44 through 43.46) they are the electrostatic and the electrochemical potentials, and the particle currents are driven by the gradient of the electrochemical potential. The advantage of this formulation with respect to the one in Equation 43.31 through 43.33 is that it is independent on the functional form of the density, and the Einstein relation is not needed explicitly. For the rest, it has the same limitations of applicability as Equations 43.31 through 43.33.

From a technical point of view, note that the decoupled carrier transport equations in Equations 43.44 through 43.46 are quasilinear, with the differential operator being proportional to the carrier density, while in Equations 43.31 through 43.33 the transport equations are linear. This makes the numerical solution procedure more involved.

If we are using the thermodynamic formulation (Equations 43.44 through 43.46) for simulation of solar cells, then we need an adequate model for the DOS or, respectively, the carrier density for all the materials involved, and we need models for the carrier mobilities and the recombination rates. Especially, the mobility is a critical parameter. In a crystalline semiconductor, the mobility is related to the scattering the carriers are subject to. So it depends strongly on temperature and defect or doping densities, and usually mobility increases with decreasing temperature and decreasing defect densities, since the associated scattering rates decrease. Mobility also depends on the electric field, typically by decreasing with increasing field due to a saturation of the mean drift velocity. In fact, the proportionality between drift velocity and electric field, given by $v = \mu_n E$, holds only at low field strength [84]. In noncrystalline materials such as amorphous or organic semiconductors, these dependencies can be radically different. For example, mobility may increase with temperature because carriers can escape more easily local energy minima or overcome small energy barriers. Or it may increase with electric field because the field reduces the effective energy barriers hindering carrier flow. In some cases, such behavior is a manifestation of an underlying microscopic transport that deviates from a diffusive one.

In disordered organic materials, e.g., the deviation from a periodic crystalline structure makes the concept of band-transport inapplicable. Instead, charges are spatially localized and charge transport occurs by transitions between localized states, a process termed as *hopping*. In reality, disordered systems show both disordered and (semi-)crystalline regions so that charge transport is always an interplay between hopping and band conduction. Higher contributions of hopping transport usually lead to lower charge mobilities. Localized states can be imagined as local energy valleys in which charges get confined. Transport can either take place by surmounting the energy barrier to a neighboring localized state by absorbing phonons, or by tunneling through the barrier. The concept of phonon-assisted hopping was introduced by Mott [85]. In 1960, Miller and Abrahams described a hopping model between two localized states consisting of both a contribution due to thermal activation and tunneling. After their model, the hopping process between the states i and j with energy difference $\Delta E = E_j - E_i$ and distance r_{ij} is of the form [86]

$$a_{i\to j} = a_0 \exp\left(-2\gamma r_{ij}\right) \begin{cases} \exp\left(-\frac{\Delta E}{k_B T}\right) & \text{if } \Delta E > 0 \\ 1 & \text{if } \Delta E \leq 0 \end{cases} \tag{43.47}$$

where a_0 is given by a typical phonon frequency in the system and γ is the inverse localization constant. The formula combines the contribution of two different processes to hopping: the first exponential describes a tunneling contribution through a barrier given by the distance r_{ij} of the localized states, and the latter term stands for the thermally activated contribution with the relation between barrier height ΔE_{ij} and the thermal energy $k_B T$.

Another model that takes into account strong coupling between charges and lattice is the polaron model. In organic materials with low permittivities, a charge on a molecule can considerably deform the structure of the molecule. The complex of charge and the distortion of the lattice it induces is known as a quasiparticle called *polaron*. Marcus [87] developed a model for the transition of polarons:

$$a_{i\to j} = \frac{\left|J_{ij}\right|^2}{\hbar}\sqrt{\frac{\pi}{\lambda k_B T}}\exp\left(-\frac{(\Delta G_{ij} - \lambda)^2}{4\lambda k_B T}\right) \tag{43.48}$$

where J_{ij} is the transfer integral for electron and hole transfer, ΔG_{ij} is the free energy difference between state, i and j and λ the reorganization energy, which accounts for the reconfiguration of the molecule due to the deformation by the charge.

In order to derive a mobility model for disordered materials, a description for the distribution of available states must be available. The spatial irregularity of these systems leads to fluctuations in the intermolecular interactions. A successful model to represent these energetic fluctuations is the GDOS as given in

Equation 43.42, in which the energy distribution of the localized states is spread by a Gaussian distribution function around the HOMO/LUMO energy levels of the organic material. The GDOS in combination with the Miller–Abrahams hopping was applied by Bässler [79] by Monte Carlo simulations to study the effect of temperature and electric field on the charge mobility. He found a non-Arrhenius temperature dependence $\mu(T) = \mu_0 \exp\left(-(t\hat{\sigma})^2\right)$ with $\hat{\sigma} = \sigma/k_B T$, and, Poole–Frenkel type behavior $\mu(F) = \mu_0 \exp\left(\gamma\sqrt{F}\right)$ for the dependence on the electric field F for a limited range of electric fields. To include the influence of charge carrier densities on the mobility, Pasveer et al. [88] parametrized the results based on a direct solution of the Master equation in a system with a GDOS. They found as expression for the hopping mobility:

$$\mu(T, \rho, F) = \mu(T, \rho) f(T, F) \text{ with} \tag{43.49}$$

$$\mu(T, \rho) = \mu_0 c_1 \exp\left[-c_2 \left(\frac{\sigma}{k_B T}\right)^2\right] \exp\left[\frac{1}{2k_B T}\left(\frac{\sigma^2}{k_B T} - \sigma\right)(2\rho a^3)^\delta\right] \tag{43.50}$$

$$f(T, F) = \exp\left\{0.44\left[\left(\frac{\sigma}{k_B T}\right)^{\frac{3}{2}} - 2.2\right]\left[\sqrt{1 + 0.8\left(\frac{eFa}{\sigma}\right)^2} - 1\right]\right\} \tag{43.51}$$

where ρ is the charge carrier density, either electrons or holes, F is the electric field, $k_B T$ the thermal energy, $a = N_0^{-\frac{1}{3}}$ the distance between localized states, σ the energetic disorder in the GDOS, $c_1 = 1.8 \cdot 10^{-9}$, $c_2 = 0.42$, $\mu_0 = a^2 \nu_0 e/\sigma$, with ν_0 the prefactor in the Miller–Abrahams formula, and

$$\delta = 2\frac{\log(s^2 - s) - \log(\log(4))}{s^2} \text{ with } s = \frac{\sigma}{k_B T} \tag{43.52}$$

This parametrization can be used in drift-diffusion models to account for the hopping nature of transport.

43.4.2 Kinetic Monte Carlo

In this section, we give an overview of the kinetic Monte Carlo (kMC) method [89,90] and how it can be applied to set up a model for organic solar cells.

Monte Carlo methods offer a simple yet powerful tool to solve real world problems by the use of random numbers. In the hierarchy of transport models shown in Figure 43.9, kMC bridges the gap between macroscopic models such as drift-diffusion and molecular dynamics (MD) and quantum mechanical models.

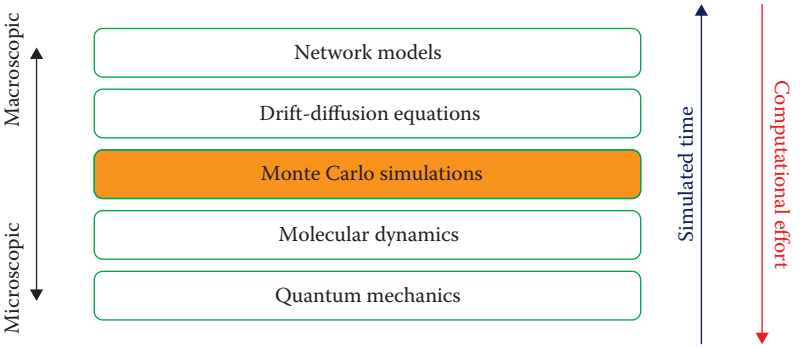


FIGURE 43.9 Hierarchy of transport models. Kinetic Monte Carlo simulations bridge the gap between macroscopic and microscopic modeling techniques.

The idea for the use of random numbers in an algorithm came from Fermi [91] and was further pursued by Ulam and Metropolis in the 1940s studying neutron diffusion in fissionable materials. They developed the most used method of this kind, the Metropolis algorithm [92], during a study about the calculation of state equations for a system of interacting particles described by a hard sphere cutoff potential.

In the following decades, Monte Carlo methods have proven to be a versatile tool and much progress in the development of calculations for a variety of areas, ranging from surface adsorption processes, radiation damage annealing, and statistical physics [93–95], has been achieved. Especially, the development of a Monte Carlo algorithm class that models physical systems as dynamically evolving from state to state by a certain set of transition rates obtained large interest as it allowed to study time-dependent properties of physical systems. This class of algorithms is today commonly known as *kinetic Monte Carlo*. It has received great benefit in the growing computational power of modern processors and provides a flexible tool to simulate dynamical behavior of physical systems, e.g., transport processes in inorganic semiconductor devices [96]. Major contributions to the modern kMC algorithm, as it is described here, have been made by Bortz et al. in 1975 [95] and Gillespie in 1976 [89] and 1977 [97].

Compared to MD simulations, which include the interactions of single atoms with all dynamics of the molecules like vibrations, occurring on very small time scales (fs), kMC characterizes the state of a system by a set of more macroscopic variables which do not include underlying dynamics like fast vibrational modes. These motions on smaller time scales can be ignored as long as the current state can still be assigned to a long-time state. The dynamical evolution of the system is only seen by transitions between such long-time states, called events. Such a system is called an infrequent-event system [90]. Figure 43.10 illustrates this concept. By ignoring the underlying fast motions and only considering the transitions between long-time states, a sufficient amount of processes can be neglected. Thus, much larger overall simulation times of the order of milliseconds to seconds can be reached.

The above mentioned concept can be applied in particular to organic solids, where particles such as electrons, holes and excitons reside in localized states and transport is determined by hopping from one localized state to another. The time scale of these hopping events is much longer than other processes, which do not change the spatial localization of the carriers.

Compared to drift-diffusion, kMC has the advantages of being able to capture the discrete nature of the problem by modeling individual particles at their molecular positions, and of providing a physically more accurate description of hopping transport without having to resort to effective continuous media models. Moreover, it can be much easier and physically more sound to treat blended materials with nanoscale morphology in kMC.

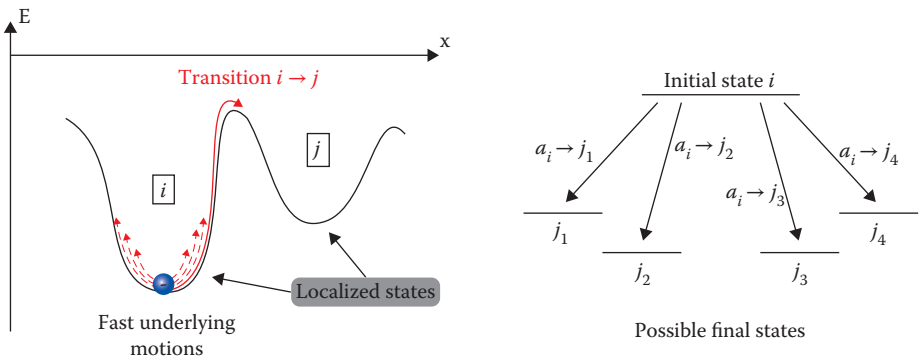


FIGURE 43.10 Transitions from long-time state to long-time state with underlying fast modes in an infrequent-event system (left). State-to-state transitions reduced to the needs for kinetic Monte Carlo (right).

43.4.2.1 kMC Algorithm

The essential concept of the kMC method is the characterization of the considered system by a set of system-specific long-time states. A state is, e.g., defined by the position of all particles in the system at a certain time. Under dynamical evolution, it is possible for the system to change its state by certain system-specific transitions, thus motion of particles from one state to another. The task for an accurate model is to identify all the important processes that take place in the system. By considering processes that happen on long time scales and neglecting those which act on very small time scales, one is able to accomplish a tradeoff between accuracy and sufficiently long simulation times. It is of importance to carefully choose the implemented processes in order not to lose essential parts of the system dynamics.

Let the system be in a certain state i. Depending on the current system setup, a set of several different states j exists to which a transition can take place. This is indicated in Figure 43.10 (right). The transitions from beginning states i to possible end states j can be characterized by rate constants a_{ij} given in units of s^{-1}. Rate constants are a measure for the probability of the corresponding transition from $i \rightarrow j$. An important property must be that the rates are memoryless, meaning that they are only dependent on the setup of the current state and are not dependent on how the system evolved into it. This property is known from the theory of Markov chains [98] and allows, if the set of rate constants is known, to determine the dynamical evolution of the system.

As known from statistical physics [99], stochastical processes characterized by discrete transitions from state to state under temporal evolution can phenomenologically be described by the master equation

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = -\sum_j a_{ij}\, p_i(t) + \sum_j a_{ji}\, p_j(t) \tag{43.53}$$

The probability $p_i(t)$ to find the system in state i at a certain time t changes in a way that is determined by the rate constants a_{ij} and the probabilities of only the initial state $p_i(t)$ and the possible final states $p_j(t)$. Finding solutions of the master equation, be it by analytical or numerical methods, is often complex and not feasible. The kMC method is based on a stochastical framework and provides a numerical approach to obtain the dynamic system evolution based on the time-dependent propagations from state to state. By choosing a transition pathway through a chain of subsequent states, one possible dynamic system evolution is obtained. Such a walk through a pathway of transitions forms a Markov chain. Averaging over a large number of such Markov chains provides an equivalent system behavior as described by the master equation. Therefore, the kMC algorithm is essentially emulating the master equation.

For a detailed derivation of the associated mathematical framework, the reader is referred to the original work of Gillespie [89]. In the following, we present only the equations necessary for a possible implementation of the algorithm. The algorithm as proposed by Gillespie relies on the ability to characterize the system events by a set of transitions

$$\{\mu\} \text{ with } \mu = 1, 2, \ldots, M \tag{43.54}$$

where M is the total number of transitions. μ is equal to a transition from state i to one of the possible states j in that configuration: $i \rightarrow j_\mu$. Then a_μ is the rate value describing process μ and $\{ a_\mu \}$ is the corresponding set of rates. All transitions must describe distinct processes and their number M can vary over time, depending on which events are enabled at a certain step in the system evolution. It is essential to consider all major processes in order to be able to accurately reproduce the behavior of the system. The main task is now to develop a procedure to simulate the time-dependent evolution of the system under consideration of the given transitions. This makes it necessary to derive an algorithm that is able to choose which transition

will be carried out and (for the time dependency) at what time span, all based on the drawing of random numbers.

At any given time, the set of $\{a_\mu\}$ characterizes the possible transitions. They are functions of the current system setup and will change over time, as described later.

Instead of finding a solution to the master equation, the kMC method is based on a function called the probability density function (PDF). The PDF is the suitable function to select a transition and its corresponding time step based on random numbers. The PDF is defined as

$$\begin{aligned} P(\tau,\mu)\,d\tau = {} & \text{probability at time } t \text{ for the next transition} \\ & \text{to occur in time interval } (t+\tau\,,\ t+\tau+d\tau) \\ & \textit{and} \text{ being a transition of type } R_\mu \end{aligned} \tag{43.55}$$

where $P(\tau,\mu)$ is the PDF. This function can be interpreted as follows: it is a joint PDF of the continuous time variable τ ($0 \leq \tau < \infty$) and the integer variable μ ($\mu = 1, 2, \ldots, M$) characterizing the transition. At a certain time t, the PDF represents the probability for transition μ to occur within an infinitesimal time span of $d\tau$ after no other event has taken place for a time interval τ. Hence, $P(\tau,\mu)$ can be seen as the independent probabilities to have no transition in the time interval $(t\,,\ t+\tau)$ *and* to have a transition of type R_μ immediately afterward in the infinitesimal interval $(t+\tau\,,\ t+\tau+d\tau)$. In common terms, the transition μ needs time τ before the next transition occurs.

Using the abbreviation

$$a = \sum_{\nu=1}^{M} a_\nu \tag{43.56}$$

as the total transition probability per unit time, the total expression for the PDF is given by

$$P(\tau,\mu) = a_\mu \cdot \exp\left(-\sum_{\nu=1}^{M} a_\nu \tau\right) = a_\mu \exp(-a\tau) \tag{43.57}$$

This function, only dependent on the set of events a_μ, is the starting point to choose a random number pair (τ,μ) that characterizes the next transition and time step in the dynamic evolution of the system. The PDF is properly normalized, meaning that every pair (τ,μ) with $\tau \in (0 \leq \tau < \infty)$ and $\mu = 1, 2, \ldots, M$ has its weighted contribution to the PDF and that it is guaranteed to obtain an associated value for the PDF by randomly choosing one pair:

$$\int_{t=0}^{\infty} \mathrm{d}\tau \sum_{\nu=1}^{M} P(\tau,\nu) = \sum_{\nu=1}^{M} a_\nu \int_{t=0}^{\infty} \mathrm{d}\tau \exp(-a\tau) = 1 \tag{43.58}$$

Now, the fundamental Monte Carlo step can be defined: by choosing a random pair of the variables (τ,μ) according to the PDF $P(\tau,\mu)$ after Equation 43.57, a weighting based on the magnitude of the various transitions and the exponential decrease in time is performed. For the execution of this step, two approaches were suggested by Gillespie: the *direct method* and the *first reaction method*. In this context, the former method is used. The direct method is most efficient in the way it utilizes random numbers to determine a pair of transition and a time step, while the first reaction method can save computational power at the cost of accuracy.

In the direct method, the PDF is divided into two separate PDFs, each of which is itself normalized and only dependent on either the time τ or the transition μ:

$$P(\tau, \mu) = P_1(\tau) \cdot P_2(\mu|\tau) \tag{43.59}$$

Here, the first factor represents the probability density that any of the possible transitions takes place in the time interval $(t+\tau, t+\tau+d\tau)$, irrespective of which one. Thus, $P_1(\tau)$ is the sum over all transitions. With Equation 43.57 follows

$$P_1(\tau) = \sum_{\nu=1}^{M} P(\tau, \nu) = \sum_{\nu=1}^{M} a_\nu \exp(-a\tau) = a \cdot \exp(-a\tau) \tag{43.60}$$

From this and Equation 43.59, $P_2(\mu|\tau)$ is easily derived as

$$P_2(\mu|\tau) = \frac{P(\tau, \mu)}{P_1(\tau)} = \frac{a_\mu}{a} \tag{43.61}$$

and $P_2(\mu|\tau)\, d\tau$ represents the probability that a reaction μ occurs, given the condition that a time step of τ has been chosen before. Both P_1 and P_2 are properly normalized PDFs on their own.

A pair of random numbers (τ, μ) according to the distributions (Equations 43.60 and 43.61) can be produced as follows, using (pseudo-)random number generator with uniform distribution available in all modern programming languages. Every normalized PDF $P(x)$ has an associated cumulative distribution function, defined by

$$F_1(x) = \int_{-\infty}^{x} P_1(x')\, dx' \qquad \text{for a continuous } P_1(x) \tag{43.62}$$

$$F_2(i) = \sum_{i'=1}^{i} P_2(i') \qquad \text{for a discrete } P_2(i) \tag{43.63}$$

Because $\lim_{x \to -\infty} F_1(x) = 0$ and $\lim_{x \to \infty} F_1(x) = 1$ for the continuous case, and $F_2(0) = 0$ and $F_2(M) = 1$ for the discrete case, both functions are monotone and limited. Thus, there exists an inverse function F^{-1} for each of them. It is made sure that every value from the output range has a corresponding value x or i in the domain range. The concept of the inversion method is to draw two uniform random numbers $r_1, r_2 \in (0, 1)$, and calculate $x = F_1^{-1}(r)$ and i from $F_2^{-1}(i)$. This way the obtained values for x and i are distributed according to the PDF from which F is derived.

Using this inversion method and the PDF $P_1(\tau) = a \cdot \exp(-a\tau)$, the corresponding cumulative probability function is[†]

$$F_1(\tau) = \int_0^{\tau} a \cdot \exp(-a\tau')\, d\tau' = 1 - \exp(-a\tau) \tag{43.64}$$

The continuous and exponentially distributed time variable τ, which describes the time passed before one transition (or, for a more vivid interpretation: the time "needed" for one transition), is then

[†] Note that $P_1(t)$ is only defined for $t \geq 0$.

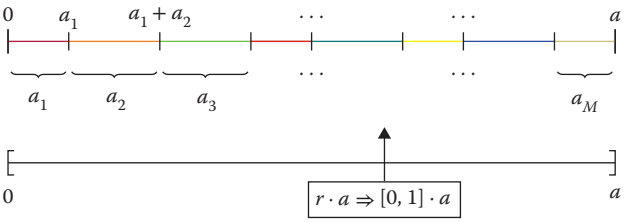


FIGURE 43.11 Drawing a discrete transition by a continuous random variable $r \in [0, 1]$. All enabled rate constants are added to yield the total rate a. Multiplication of a random number in the unit interval with the total rate corresponds to one specific rate process, weighted by its magnitude. Rates are added until the next addition will exceed $r \cdot a$. The last rate in the addition is chosen.

given by the inverse function of F_1. Hence,[†]

$$\tau = F_1^{-1}(r_1) = \frac{1}{a} \ln\left(\frac{1}{r_1}\right) \qquad \text{with } a = \sum_{\nu=1}^{M} a_\nu \text{ and } r_1 \in (0, 1) \tag{43.65}$$

To choose a transition μ from the set of events $\{a_\mu\}$ ($\mu = 1, 2, ..., M$) by a random number r_2, the cumulative probability must satisfy

$$F_2(\mu - 1) < r_2 \leq F_2(\mu) \tag{43.66}$$

That is, $r_2 \in (0, 1)$ corresponds to one plateau in the discrete histogram of F_2. Substituting Equation 43.61 in Equation 43.63 yields an expression for $F_2(\tau|\mu)$:

$$F_2(\tau|\mu) = \sum_{\nu=1}^{\mu} \frac{a_\nu}{a} \tag{43.67}$$

With this, the abovementioned condition reads

$$\sum_{\nu=1}^{\mu-1} \frac{a_\nu}{a} < r_2 \leq \sum_{\nu=1}^{\mu} \frac{a_\nu}{a} \tag{43.68}$$

or, multiplying by a,

$$\sum_{\nu=1}^{\mu-1} a_\nu < r_2 \cdot a \leq \sum_{\nu=1}^{\mu} a_\nu \quad \text{with } a = \sum_{\nu=1}^{M} a_\nu \text{ and } r_2 \in (0, 1) \tag{43.69}$$

This represents a sum of all transition probabilities, up to an index for which the sum is larger than the random variable r_2. This index is set to be the chosen transition μ. The process to derive μ computationally from a random number is illustrated in Figure 43.11.

The essence of the Monte Carlo method is condensed in the two equations (Equations 43.65 and 43.69) to calculate a pair of random numbers (τ, μ), while the rest of the method is fairly straightforward and can be reduced to only the following few, simple steps, shown in a flowchart in Figure 43.12:

[‡] The term $1 - r$ has been replaced by r. Since we are dealing with probabilities, this is valid for uniform random numbers in the unit interval.

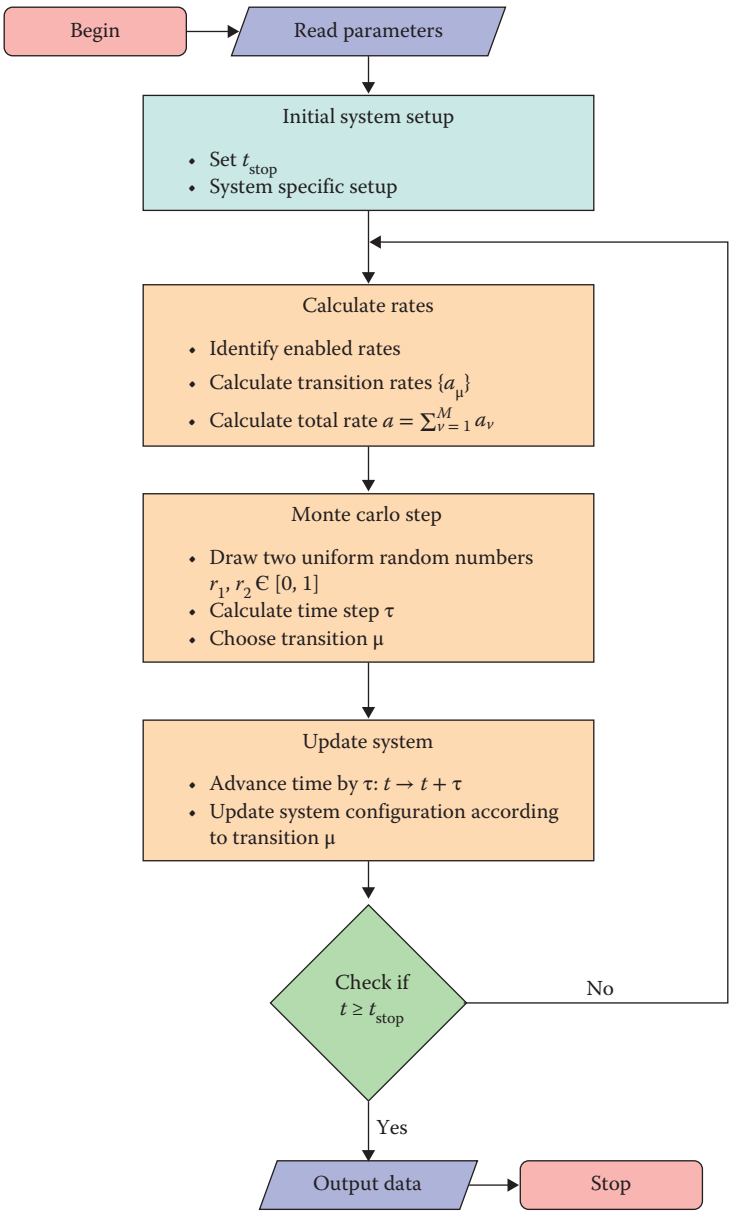


FIGURE 43.12 Flowchart of a general kinetic Monte Carlo algorithm.

0 – Initialization

In the beginning, system-specific parameters are defined in the program. This includes all parameters that must be known in order to set up the system and to calculate the transition rates. Variables to store the occurrence of relevant processes during the simulation are set up here. The simulation time t_{stop} is defined. It determines how long the dynamic evolution should be simulated. The starting time is set to $t = 0$.

1 – Calculation of transition rates

The kMC method is reliant on physical models for all processes that can occur in the system. All physics enters the model in terms of the event rates that are calculated in this step. Depending on the

current state, only some transitions may be activated. Conditions must be met in order to make sure that rates for processes that cannot occur in a certain system state are deactivated. This is realized by setting the value for a deactivated rate $a_\mu = 0$ so that it is not considered in the selection a transition after Equation 43.69. After all rates have been calculated and stored, they are summed up to provide the total rate a.

2 – Monte Carlo step

As described in Section 43.4.2.1, a random pair (τ, μ) is generated according to the PDF $P(\tau, \mu)$ from the list of available (and activated) transitions $\{a_\mu\}$. This accounts for the stochastically correct time step τ for the selection of a transition μ.

3 – Update system

To update the system the time t is advanced by τ. Based on which transition μ has been chosen, the system configuration needs to be updated accordingly. How every update is specifically depends on the physical process.

After updating the system, a check is performed whether the simulation time has reached the specified stopping time t_{stop}. If $t < t_{\text{stop}}$, the simulation will continue by jumping back to step 1 and recalculate the event rates. This way, the system update on the behavior of other processes is reflected. The steps 1–3 represent the so called *Monte Carlo loop*. They will be executed as long as termination condition is not yet fulfilled.

4 – Termination and data output

Through the successive time advance t eventually becomes larger than the stopping time and the Monte Carlo loop is terminated. At last the variables tracking selected events can be evaluated.

Overall, the kMC method offers a straightforward algorithm to simulate the time-dependent evolution of a system. It is only reliant on (1) the assumption that all physical processes can be determined by rate expressions, and (2) that a (pseudo-)random number generator with a sufficiently small correlation between two subsequent numbers and a large period is provided.

43.4.3 Example: Transport through 1D Disordered System by kMC

To illustrate an actual implementation of the kMC algorithm, we provide a short manual on how to model charge transport in a 1D, energetically disordered system, e.g., an organic polymer. We will go through the steps presented in the previous section and in Figure 43.12: the initial system setup, the simulation loop, and the evaluation of quantities.

43.4.3.1 System Setup and Starting Conditions

In the beginning, the available states that the system can occupy must be defined. The setup is illustrated in Figure 43.14. Here, we consider a linear chain of N equidistant nodes indexed by $i = 1, \ldots, N$. The position of each node is quantified by $x_i = i \cdot L$ where L is the fixed distance between all nodes. Then, energy values are assigned to each node accounting for the influence of an applied electric field F, which is assumed to be constant leading to a linear energy drop across the system, and an energetic fluctuation $r(\mu, \sigma)$ drawn from a GDOS distribution with mean μ and standard deviation σ as given in Equation 43.42. This makes up for an energy landscape of the form $E_i = qFx_i + r_{\text{GDOS}}(\mu = 0, \sigma)$. The disorder is picked independenly of the position. A typical configuration for E_i is depicted in Figure 43.14.

Next, initial conditions must be set. The simulation time is set to an initial time $t = 0$. It will be updated during each simulation step. The termination condition is defined by the particle position: if the particle reaches the last node N, the simulation loop stops. We allow one electron of charge $q = -e$ at a time to be in the system. In the beginning it is set to the first node x_1.

A model for the transition rates needs to be defined. We allow the particle to move only to its immediately neighboring sites to the left or to the right. The transitions are described by hopping rates according to the Miller–Abrahams formula (Equation 43.49). There are then only two distinct rates: a_l for hopping to the left and a_r for hopping to the right (Figure 43.13). They depend on the relative energy difference $\Delta E = E_j - E_i$ of the current node i of the particle and the final node $i \pm 1$. Typical values are $a_0 = 10^{11}\ \mathrm{s}^{-1}$, $\gamma = 2\mathrm{nm}$, and $r_{ij} = L = 1\ \mathrm{nm}$.

43.4.3.2 Simulation Loop

After system setup, the simulation loop is executed until the termination condition is reached. It consists of the following steps:

1. Check the energy landscape to the right and to the left of the current particle position $E_{i\pm1}$.
2. Calculate the hopping rates $a_\mathrm{l}(E_{i-1} - E_i)$ and $a_\mathrm{r}(E_{i+1} - E_i)$.
3. Pick two distinct random numbers r_1, r_2 uniformly from the interval $(0, 1)$.
4. Calculate the total rate $a = a_\mathrm{l} + a_\mathrm{r}$. Then, determine the transition (hopping left or right): If $r_1 \cdot a \leq a_l$, pick a_l. Otherwise, if $a_\mathrm{l} < r_1 \cdot a \leq a_\mathrm{l} + a_\mathrm{r} = a$, pick a_r.
5. Calculate timestep: $\tau = -\frac{r_2}{a}$.
6. Update system according to the transition picked, i.e., change the particle position to the node on the right or on the left.
7. Advance time $t \to t + \tau$.
8. Check termination condition. If still true, go to step 1. Otherwise, exit loop.

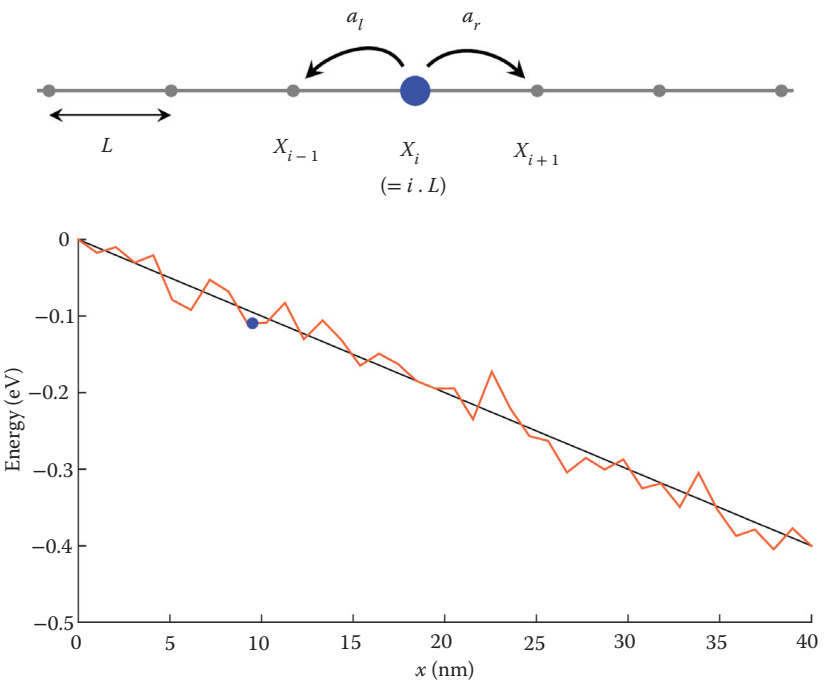


FIGURE 43.13 Setup to model hopping in a 1D system. Top: Spatial discretization of the problem in equidistant nodes. Bottom: Energetic configuration of the nodes. The setup shows the energetic landscape for $N = 40$ nodes comprised of a linear drop of the electric field $F = 10^5$ V/cm and a Gaussian energetic disorder with $\sigma = 30$ meV. The particle is allowed to hop to neighboring nodes under the influence of the energetic configuration.

43.4.3.3 Evaluation

After termination, desired quantities can be evaluated. In this case, since the time that the particle needed to hop through the disordered landscape t_{sim} is now known, the effective velocity $v = \frac{N \cdot L}{t_{sim}}$ can be calculated. From v and the electric field F, the effective transport mobility can be extracted as $\mu = \frac{v}{F}$.

43.4.3.4 Further Info

The general system setup and transition rates can be extended to 2D or 3D (see Figure 43.15). More particle types (holes, excitons) and more effects than hopping, such as injection, recombination, photogeneration of excitons can be introduced to obtain a more realistic model. The principle of choosing transition and timestep stays the same.

43.5 Specific Modeling Issues

In this section, we provide further details on the implementation of models specific to some of the emerging PV technologies, and we cover generic modeling issues.

43.5.1 Dye Solar Cells

DSCs have been extensively simulated using analytical models or numerical ones at the level of drift-diffusion. Most simulations have been performed using 1D models [100–102] and an effective material approximation (see Section 43.5.4 for an extensive description of this approximation). Only few simulations going beyond the effective medium model and trying to include the morphology [103] or to go toward 2D and 3D simulations [104,105] have been reported so far.

The main issues concerning simulations of a DSC are for sure related to the complex morphology of the mesoporous material and the chemical reactions at the interfaces. The latter include both the dye regeneration processes as well as all the recombination events where electrons in the TiO_2 are transferred to reduce ionized dyes or triiodide ions. The main issue in modeling is that all these processes are in general complex electrochemical rates where the charge transfer occurs in several steps. They depend on the effect of the electrolyte, the reorganization energy of the environment, temperature, and voltage [106]. Moreover, different dyes or electrolytes will lead to different kinetics and thus different recombination rates. This leads to recombination rates that formally are completely different from the standard formulas given in Equations 43.34 through 43.36.

Beside that, boundary conditions at the contacts also are very uncommon with respect to more conventional PV devices. While the anode with the ITO or FTO can be approximated by a standard ohmic or Schottky contact, the cathode at the platinum is an electrochemical contact which requires to take into account the formation of an ion accumulation layer. The standard boundary condition is a Butler–Volmer (BV) equation, which is generally a Robin-type boundary condition, since it connects the electron transfer rate with the electrochemical potential at the contact. A BV equation [107] describes the dependency of the electrical current at an electrode j on the electrode potential when both anodic and cathodic reactions are involved:

$$j = j_0 \cdot \left\{ \exp\left[\frac{\alpha_A z e}{k_B T}\eta\right] - \exp\left[\frac{\alpha_C z e}{k_B T}\eta\right] \right\} \tag{43.70}$$

where j_0 is the exchange current density prefactor, $\eta = E - E_{eq}$ the activation overpotential given by the difference in electrode potential and equilibrium potential, α_A and α_C are charge transfer coefficients for anode and cathode, respectively, z is the number of electrons participating in the reaction at the electrode, e

is the elementary charge and $k_B T$ the thermal energy. A Robin-type boundary condition in general is a combination of Dirichlet boundary conditions and Neumann boundary conditions. Imposed on a differential equation of function f, it reads in a general form

$$af + b\frac{\partial f}{\partial n} = g \tag{43.71}$$

where a and b are nonzero constants and g is a function defined on the boundary of the problem. Depending on whether the limiting process is the ionic diffusion of the redox pair or the charge transfer between platinum and triiodide ions, we can have different shapes of the BV equation. Finally, electron transport in the TiO_2 is mediated by trapping and detrapping events, a physical situation captured by the multitrapping model, where the mobility becomes density dependent [108,109]. For solid state DSCs, the simulation issues are more similar to OPVs as the electrolyte is replaced by an organic semiconductor.

43.5.2 Organic Photovoltaics

Modeling of OPV devices incurs several important differences when compared to inorganic devices. First of all, as mentioned already in section 43.4.1, the DOS is fundamentally different as it does not result from electronic states in a periodic potential, but from a both energetically and spatially disordered energy landscape of molecular levels as shown schematically in Figure 43.14. This has to be considered when using or implementing a drift-diffusion based simulation tool, since the standard drift-diffusion model has been derived for carriers with a parabolic DOS. As a further consequence of this energy disorder, transport cannot be described adequately by carriers moving freely in conduction and valence bands, but rather by hopping between energy levels on different molecules [86,110–112], as described in Section 43.4.1. Fortunately, on a macroscopic level, i.e. on a scale much larger than the molecular structure, transport can still be described in terms of drift and diffusion, if a suitable model for the carrier mobilities can be found. Such a model is described in [88], where the mobility has been parametrized based on numerical results of the Master equation.

Due to the different DOSs, the models for transport across interfaces and in particular for injection at the contacts also are different from their counterparts employed in standard drift-diffusion simulations of inorganic semiconductors. The properties of a metal-semiconductor contact, known as Schottky contact in the inorganic case, are controlled by the enetrgetic alignment of the metal work function with the semiconductor band edges, by the semiconductor's Fermi level, presence of interface states, and by the semiconductor DOS. The boundary condition on the particle current e.g. for the electrons at the metal-semiconductor contact can be written as $J_n = qv_R(n - n_0)$, where n_0 is the electron density at the contacts in equilibrium and v_R is a recombination velocity, with units cm/s. The latter quantity depends on the exact

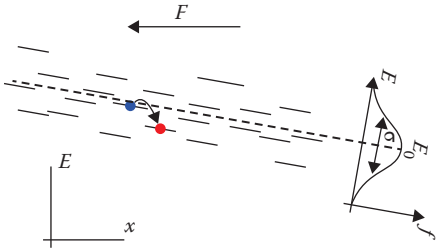


FIGURE 43.14 Schematic representation of the energetic disorder in an organic semiconductor. Each level represents a molecular energy level, which are disordered in energy and in space, following approximately a Gaussian distribution. Transport is driven by hopping between levels, described by energy difference and distance-dependent coefficients that depend also on the external electric field F.

transport mechanism and on the semiconductor DOS, and is therefore very different between inorganic and organic semiconductors [113].

A further difference of OPVs is their low permittivity of $3 \sim 4$ compared to inorganic materials with $\varepsilon_r \approx 10$. This leads to strong local Coulomb interaction which induces the formation of excitons even at room temperature. Since only electrons and holes can be collected at the contacts, the photo-generated excitons need to dissociate first. For this they need to reach the donor/acceptor interface, which is however possible only within their diffusion length of approximately 10 nm. For this reason, OPVs employ a blended donor/acceptor absorber, forming a BHJ. In combination with the hopping transport of charges in spatially and energetically disordered materials, these factors are the key differences to inorganics and make up for most of the challgenges associated with the modeling of OPVs. Since excitonic effects are so important in OPVs, several authors have proposed drift-diffusion-based models including explicitly exciton generation and dissociation [114–116].

The main disadvantage of state-of-the-art drift-diffusion models is the simplification of the BHJ morphology to a 1D effective medium. More details on the background and modeling techniques of the morphology can be found in Section 43.5.4. Since no interfaces exist in the effective medium, all interface effects between the donor/acceptor boundaries, such as exciton splitting and electron–hole recombination, are assumed to be bulk processes and evenly distributed across the photoactive layer. An absorbed photon directly creates and electron–hole pair and exciton diffusion and loss processes cannot be considered. Also, no charge accumulation effects at the donor/acceptor interface can be observed due to the lack of the blend morphology.

Most of these limitations do not apply to kMC simulations. A real 3D blend morphology, explicit treatment of exciton diffusion and splitting, and other interface processes, can easily be implemented. Also, the hopping transport in energetically and spatially disordered materials is naturally representable on a discrete lattice. However, since kMC is a statistical method, it often has to deal with unfeasible simulation times in order to achieve a reasonable accuracy for the desired quantities. This is especially critical when one has to sample very slow processes in presence of very fast events. Specifically, the evaluation of Coulomb interaction is expensive, and since particles are modeled explicitly and not by continuous quantities, the kMC algorithm is very demanding, in particular for a large number of particles. Large particle numbers are induced by a high energetic disorder, small injection barriers ("small barrier problem" in kMC [117]), or large device dimensions/thicknesses and limit the choice of parameters where kMC can be applied to certain ranges. Procedures to parallelize the algorithm and the Coulomb interaction need to be applied [118] to model larger systems and/or get better statistics in the results.

43.5.3 Perovskites

Perovskite, based devices pose two types of issues related to simulation. On the one hand, the fundamental working principle has not been completely unveiled yet, in particular the role of the perovskite in carrier transport. Therefore, it is not clear yet if the perovskite should be used merely as sensitizing absorber similar to the dye in a DSC, or if it should be used in a configuration more similar to an inorganic cell where the absorber acts also as electron and hole transporter. Partly, such questions arise because the development of perovskite PV is driven by the DSC and OPV comunities, with different views on device structure and functioning, and so also the point of view for modeling varies. From a fundamental point of view, being a crystalline material, perovskite seems to be more similar to an inorganic material with well-defined band structure [119], and it would seem that standard simulation models used for inorganic semiconductors should apply.

There are several issues encountered experimentally, where device simulation could give a strong contribution for a better understanding. One open question is whether ferroelectricity and ferroelectric domains exist in perovskite layers and whether they play a role on device functioning. While a fundamental answer to such questions based on density functional theory or MD is not easy, device-level simulation can at least provide inside into the qualitative effects that could be expected in presence of spatially varying

polarization fields. This requires, however, simulation models that can include such effects. In particular, realistic domains have to be generated numerically, and the drift-diffusion model has to include the variable polarization field. Attempts in this direction have been made [120].

Another challenge for modeling is the hysteresis typically observed experimentally when measuring perovskite solar cell current–voltage characteristics [40,121]. Currently, the most probable cause of such hysteresis is believed to be migration of halide ion vacancies [122,123]. However, modeling of this effect in a standard device simulator is not easy, since the simulation model assumes transport of electrons and holes for a given fixed spatial distribution of defects. Therefore, self-consistent modeling of electronic transport and ion migration requires an extension of the standard drift-diffusion model with further mobile species, which is currently not available in standard device simulation software.

43.5.4 Modeling of Morphology and Grain Effects

Many device architectures for OPVs, (solid-state-) DSSCs and perovskite solar cells have regions with complex intermixed morphologies between two different materials where important processes are assumed to be controlled by the shape of the intermixing, especially mobilities through the morphologies and interface processes such as charge recombination and charge generation. For the sake of simplicity and low computational effort, the 3D blend is often simplified and treated as an effective medium, hiding the interface processes and 3D morphology effects in effective models for charge generation and recombination, and the mobility. An effective medium can be defined such that its band edges are given by the lower and higher values of the conduction and valence band edges of the two blended materials, respectively. Parameters like recombination rate constants and mobilities are then usually fitted to reproduce experimental data like current–voltage curves, or experimentally measured mobilities. It is yet to be investigated under which conditions an effective medium treatment is valid and which models are accurate for this description. Alternatively to the effective medium, the mobility can be derived from microscopic simulations considering the 3D blend morphology, like kMC. A comparison between the effective medium approach and a real blend kMC simulation in BHJ OPVs shows, however, that it might be difficult to include all interface effects in a blend into mobility and recombination models within an effective medium [124], especially in low-ε_r materials. A common basis in as many parameters as possible needs to be established for a comparison: the MO levels, the generation profile, and the injection processes, to name a few. By extracting electron and hole mobilities from the 3D simulations, the effect of the blend mobility can effectively be plugged into the effective medium model. It is then tried to obtain an agreement between the two models by tuning the recombination ratio. No general agreement could be obtained, which might indicate that it is not possible to merge all parameters in effective models accurately. It has been observed that, for instance, the charge density distribution within the 3D blend cannot be recovered by an effective model, since local accumulations at an interface are not present in the effective medium where both phases are assumed to be everywhere.

Thus, performing simulations on a realistic 3D blend is important to get a detailed picture about the processes in the blend. For a drift-diffusion model, complex 3D meshes are not easy to generate, increase considerably computation time and are prone to cause problems related to convergence. To avoid these problems for 3D blends, the kMC method is a suitable choice. The kMC algorithm does not have issues with convergence, even for complex morphologies. However, due to the popularity and establishment of drift-diffusion models, more effort should be made to include morphology effects in drift-diffusion-based simulation models. An example in this direction is given, e.g., in [125].

In kMC models, a blend can be created using a spin-exchange algorithm based on the work of Watkins et al. [111] which allows to control the grade of intermixing between two phases. However, the structure is based on a discrete lattice and is not suitable as basis for a DD simulation for which a high-quality finite element mesh is needed. A smoothing algorithm can be applied in order to achieve a mesh suitable for DD simulations. An illustration of a discrete BHJ lattice and a successively smoothed mesh is

depicted in Figure 43.15. Further tests with an implementation of the 3D blend morphology as a mesh for DD simulations need to be performed to increase the common basis of the two models and get more information about where the two models differ and when the effective medium is appropriate.

Apart from nanometer-scale morphology effects, grain boundaries in multicrystalline materials can also affect optoelectronic device properties. Grain dimensions in such materials can range from nanometers to micrometers. A correct description of the effects of the grain boundaries, which can be related to both interface recombination and transport, may require to include the grain interfaces explicitly in the simulation model, especially if there is a limited number of interfaces along the carrier transport path such that an effective medium approximation may become questionable. Grain boundaries in organic materials have been studied related mostly to organic thin-film transistors [126,127]. Similar studies have been performed on polycrystalline silicon [128], and the effect of grain boundaries on carrier collection and recombination has been in CIGS devices has been studied in [129] and [130].

Note that also kMC does not automatically include the effect of grain boundaries, which should be included ad hoc with parameters based on *ab initio* or MD simulations [131].

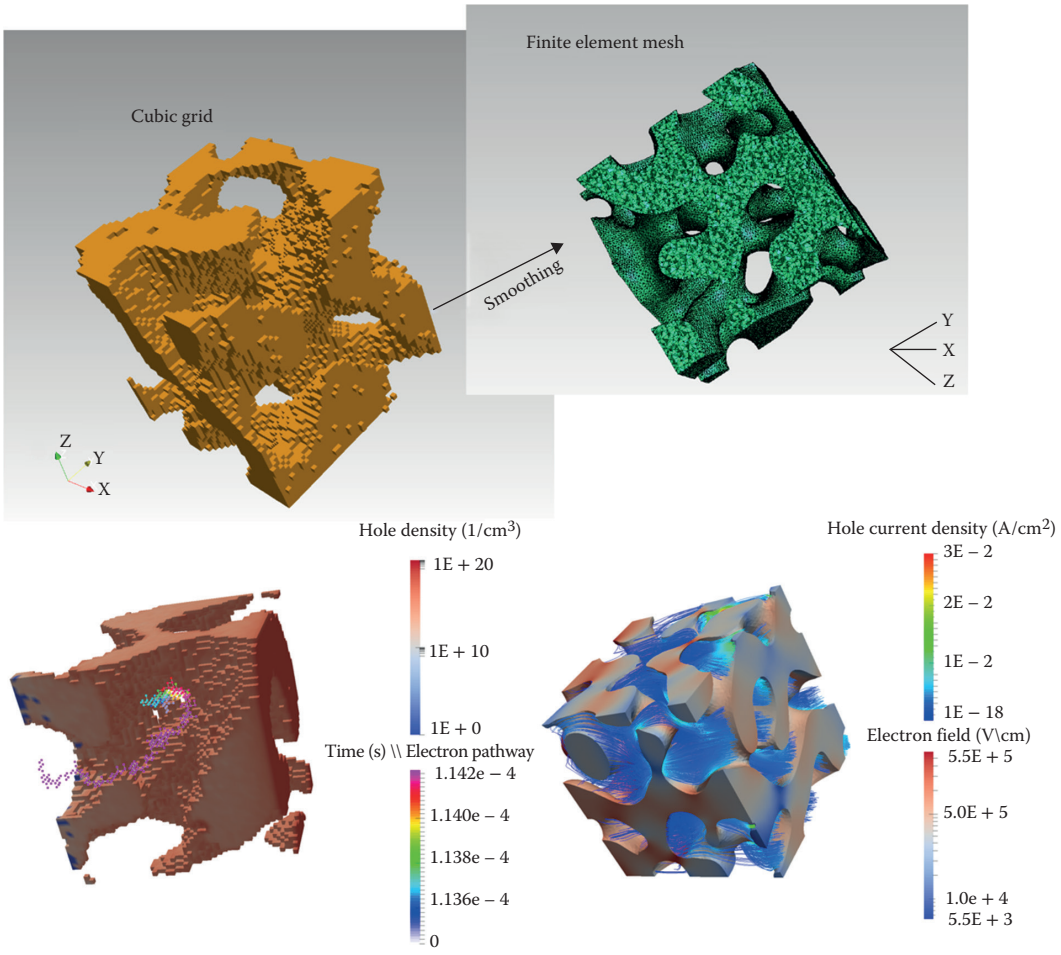


FIGURE 43.15 Smoothing a discrete lattice used for kMC simulations (top left) to obtain a suitable mesh for DD simulations (top right). Exemplary results of the kMC method (bottom left) highlighting the advantage to track single charge carrier trajectories and the DD model to extract, e.g., directed streams such as current densities through the morphology (bottom right).

References

1. J. Nelson. *The Physics of Solar Cells*. London: Imperial College Press, 2003.
2. J. Poortmans and V. Arkhipov, editors. *Thin Film Solar Cells—Fabrication, Characterization and Applications*. Hoboken, NJ: John Wiley & Sons, 2006.
3. K.L. Chopra, P.D. Paulson, and V. Dutta. Thin-film solar cells: An overview. *Progress in Photovoltaics: Research and Applications*, 12(2–3):69–92, 2004.
4. A. Jäger-Waldau. Status and perspectives of thin film photovoltaics. In A. Bosio and A. Romeo, editors, *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
5. T.M. Brown, A. Reale, and A. Di Carlo. Organic and hybrid solar cells. In A. Bosio and A. Romeo, editors, *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
6. Fraunhofer Institute for Solar Energy (ISE). *Photovoltaics Report*, 2015. Available at http://www.ise.fraunhofer.de/.
7. http://www.nrel.gov/ncpv.
8. M.A. Green. Consolidation of thin-film photovoltaic technology: The coming decade of opportunity. *Progress Photovoltaics: Research and Applications*, 14:383–392, 2006.
9. L. Dou, J. You, Z. Hong, Z. Xu, G. Li, R.A. Street, Y. Yang, L. Dou, J. You, Z. Hong, Z. Xu, G. Li, and Y. Yang. 25th Anniversary Article: A decade of organic/polymeric photovoltaic research. *Advanced Materials*, 25:6642–6671, 2013.
10. W. Cao and J. Xue. Recent progress in organic photovoltaics: Device architecture and optical design. *Energy & Environmental Science*, 7:2123–2144, 2014.
11. S.D. Stranks and H.J. Snaith. Metal-halide perovskites for photovoltaic and lightemitting devices. *Nature Nanotechnology*, 10(5):391–402, 2015.
12. A. Bosio and A. Romeo, editors. *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
13. P. Delli Veneri and L.V. Mercaldo. Thin film silicon solar cells. In A. Bosio and A. Romeo, editors, *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
14. M. Zeman. Advanced amorphous silicon solar cell technologies. In J. Poortmans and V. Arkhipov, editors, *Thin Film Solar Cells—Fabrication, Characterization and Applications*. Hoboken, NJ: John Wiley & Sons, 2006.
15. M.A. Green, K. Emery, Y. Hishikawa, W. Warta, and E.D. Dunlop. Solar cell efficiency tables (version 46). *Progress in Photovoltaics: Research and Applications*, 23(7):805–812, 2015.
16. H. Fritzsche. Development in understanding and controlling the Staebler-Wronski effect in a-Si:H. *Annual Review of Materials Science*, 31:47–79, 2001.
17. E.A. Schiff, S. Hegedus, and X. Deng. Amorphous silicon-based solar cells. In A. Luque and S. Hegedus, editors, *Handbook of Photovoltaic Science and Engineering*, pp. 487–545. Hoboken, NJ: John Wiley and Sons, 2011.
18. A. Bosio, A. Romeo, D. Menossi, S. Mazzamuto, and N. Romeo. Review: The secondgeneration of CdTe and CuInGaSe2 thin film PV modules. *Crystal Research and Technology*, 46(8):857–864, 2011.
19. S. Girish Kumar and K.S.R. Koteswara Rao. Physics and chemistry of CdTe/CdS thin film heterojunction photovoltaic devices: Fundamental and critical aspects. *Energy and Environmental Science*, 7:45–102, 2014.
20. R. Klenk and M.C. Lux-Steiner. Chalcopyrite based solar cells. In J. Poortmans and V. Arkhipov, editors, *Thin Film Solar Cells—Fabrication, Characterization and Applications*. Hoboken, NJ: John Wiley & Sons, 2006.

21. P. Reinhard, A. Chirila, P. Blosch, F. Pianezzi, S. Nishiwaki, S. Buechelers, and A.N. Tiwari. Review of progress toward 20% efficiency exible cigs solar cells and manufacturing issues of solar modules. *IEEE Journal of Photovoltaics,* 3(1):572–580, 2013.
22. T. Unold and C.A. Kaufmann. *Chalcopyrite Thin-Film Materials and Solar Cells*, Volume 1. Amsterdam: Elsevier, 2012.
23. B. O'Regan and M. Grätzel. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO2 films. *Nature*, 353:737–740, 1991.
24. K. Kalyanasundaram, editor. *Dye Sensitized Solar Cells*. Lausanne: EPFL Press, 2010.
25. S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B.F.E. Curchod, N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M.K. Nazeeruddin, and M. Grtzel. Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers. *Nature Chemistry*, 6(3):242–247, 2014.
26. U. Bach, D. Lupo, P. Comte, J.E. Moser, F. Weissrtel, J. Salbeck, H. Spreitzer, and M. Grtzel. Solid-state dye-sensitized mesoporous TiO_2 solar cells with high photon-to-electron conversion efficiencies. *Nature*, 395(6702):583–585, 1998.
27. F. Matteocci, G. Mincuzzi, F. Giordano, A. Capasso, E. Artuso, C. Barolo, G. Viscardi, T.M. Brown, A. Reale, and A. Di Carlo. Blocking layer optimisation of poly(3-hexylthiopene) based solid state dye sensitized solar cells. *Organic Electronics*, 14(7):1882–1890, 2013.
28. A. Hagfeldt, G. Boschloo, L. Sun, L. Kloo, and H. Pettersson. Dye-sensitized solar cells. *Chemical Reviews*, 110:6595–6663, 2010.
29. A. Hinsch. Status of the dye solar cell technology (DSC) as a guideline for further research. In *28th European PV Solar Energy Conference and Exhibition*, Paris, France, 2013.
30. N.S. Sariciftci, L. Smilowitz, A.J. Heeger, and F. Wudl. Photoinduced electron transfer from a conducting polymer to buckminsterfullerene. *Science*, 258(5087):1474–1476, 1992.
31. C.W. Tang. Two-layer organic photovoltaic cell. *Applied Physics Letters*, 48(2):183–185, 1986.
32. G. Yu, J. Gao, J.C. Hummelen, F. Wudl, and A.J. Heeger. Polymer photovoltaic cells: Enhanced efficiencies via a network of internal donor-acceptor heterojunctions. *Science*, 270(5243):1789–1790, 1995.
33. Y. Liu, J. Zhao, Z. Li, C. Mu, W. Ma, H. Hu, K. Jiang, H. Lin, H. Ade, and H. Yan. Aggregation and morphology control enables multiple cases of high-efficiency polymer solar cells. *Nature Communications*, 5:5293, 2014.
34. C. Deibel and V. Dyakonov. Polymer-fullerene bulk heterojunction solar cells. *Reports on Progress in Physics*, 73(9):096401, 2010.
35. G. Li, R. Zhu, and Y. Yang. Polymer solar cells. *Nature Photonics*, 6(3):153–161, 2012.
36. H.-S. Kim, C.-R. Lee, J.-H. Im, K.-B. Lee, T. Moehl, A. Marchioro, S.-J. Moon, R. Humphry-Baker, J.-H. Yum, J.E. Moser, M. Grtzel, and N.-G. Park. Lead iodide perovskite sensitized all-solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9%. *Scientific Reports*, 2:591, 2012.
37. H.S. Jung and N.-G. Park. to Perovskite solar cells: From materials devices. *Small*, 11(1):10–25, 2015.
38. P. Gao, M. Grätzel, and M.K. Nazeeruddin. Organohalide lead perovskites for photovoltaic applications. *Energy and Environmental Science*, 7(8):2448–2463, 2014.
39. D. Wang, M. Wright, N.K. Elumalai, and A. Uddin. Stability of perovskite solar cells. *Solar Energy Materials and Solar Cells*, 147:255–275, 2016.
40. M. Grätzel. The light and shade of perovskite solar cells. *Nature Materials*, 13(9):838–842, 2014.
41. P.P. Boix, K. Nonomura, N. Mathews, and S.G. Mhaisalkar. Current progress and future perspectives for organic/inorganic perovskite solar cells. *Materials Today*, 17(1):16–23, 2014.
42. P. Würfel. *Physics of Solar Cells*. Weinheim: Wiley-VCH Verlag, 2009.
43. G.P. Smestad. *Optoelectronics of Solar Cells*. SPIE Press, 2002.

44. N.-K. Persson and O. Inganäs. Simulations of optical processes in organic photovoltaic devices. In S.-S. Sung and N.S. Sariciftci, editors, *Organic Photovoltaics—Mechanisms, Materials and Devices*. Hoboken, NJ: Taylor & Francis, 2005.
45. S.L. Chuang. *Physics of Optoelectronic Devices*. *Wiley Series in Pure and Applied Optics*, 1st edition. Hoboken, NJ: John Wiley & Sons, 1995.
46. A. Taove and S.C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Norwood, MA: Artech House Publishers, 2005.
47. J. Jin. *The Finite Element Method in Electromagnetics*. Hoboken, NJ: Wiley, 2002.
48. I. Semenikhin, M. Zanuccoli, M. Benzi, V. Vyurkov, E. Sangiorgi, and C. Fiegna. Computational efficient RCWA method for simulation of thin film solar cells. *Optical and Quantum Electronics*, 44(3):149–154, 2012.
49. M. Born and E. Wolf. *Principles of Optics*. New York, NY: Cambridge University Press, 1999.
50. B. Harbecke. Coherent and incoherent reflection and transmission of multilayer structures. *Applied Physics B Photophysics and Laser Chemistry*, 39(3):165–170, 1986.
51. J.J.M. Halls, K. Pichler, R.H. Friend, S.C. Moratti, and A.B. Holmes. Exciton diffusion and dissociation in a poly(*p*-phenylenevinylene)/C_{60} heterojunction photovoltaic cell. *Applied Physics Letters*, 68(22):3120–3122, 1996.
52. J.J.M. Halls. *Photoconductive Properties of Conjugated Polymers*. PhD dissertation, Cambridge: St. John's College, 1997.
53. L.A.A. Pettersson, L.S. Roman, and O. Inganäs. Modeling photocurrent action spectra of photovoltaic devices based on organic thin films. *Journal of Applied Physics*, 86(1):487–496, 1999.
54. L.A.A. Pettersson, L.S. Roman, and O. Inganäs. Quantum efficiency of exciton-to-charge generation in organic photovoltaic devices. *Journal of Applied Physics*, 89(10):5564–5569, 2001.
55. N.-K. Persson, M. Schubert, and O. Inganäs. Optical modelling of a layered photovoltaic device with a polyuorene derivative/fullerene as the active layer. *Solar Energy Materials and Solar Cells*, 83(23):169–186, 2004.
56. H. Hoppe, N. Arnold, N.S. Sariciftci, and D. Meissner. Modeling the optical absorption within conjugated polymer/fullerene-based bulk-heterojunction organic solar cells. *Solar Energy Materials and Solar Cells*, 80(1):105–113, 2003.
57. D.P. Gruber, G. Meinhardt, and W. Papousek. Spatial distribution of light absorption in organic photovoltaic devices. *Solar Energy*, 79(6):697–704, 2005.
58. F. Monestier, J.-J. Simon, P. Torchio, L. Escoubas, F Flory, S. Bailly, R. de Bettignies, S. Guillerez, and C. Defranoux. Modeling the short-circuit current density of polymer solar cells based on p3HT:PCBM blend. *Solar Energy Materials and Solar Cells*, 91(5):405–410, 2007.
59. R. Husermann, E. Knapp, M. Moos, N.A. Reinke, T. Flatz, and B. Ruhstaller. Coupled optoelectronic simulation of organic bulk-heterojunction solar cells: Parameter extraction and sensitivity analysis. *Journal of Applied Physics*, 106(10):104507, 2009.
60. G.F. Burkhard, E.T. Hoke, and M.D. McGehee. Accounting for interference, scattering, and electrode absorption to make accurate internal quantum efficiency measurements in organic and other thin solar cells. *Advanced Materials*, 22(30):3293–3297, 2010.
61. L.D. Landau, E.M. Lifshitz, and L.P. Pitaevskii. *Electrodynamics of Continuous Media*. Amsterdam: Elsevier, 1993.
62. H.-Y.D. Yang. A spectral recursive transformation method for electromagnetic waves in generalized anisotropic layered media. *IEEE Transactions on Antennas and Propagation,* 45(3):520–526, 1997.
63. J. Ning and E.L. Tan. Hybrid matrix method for stable analysis of electromagnetic waves in stratified bianisotropic media. *IEEE Microwave and Wireless Components Letters,* 18(10):653–655, 2008.
64. A.H. Fallahpour, A. Gagliardi, F. Santoni, D. Gentilini, A. Zampetti, M. Auf der Maur, and A. Di Carlo. Modeling and simulation of energetically disordered organic solar cells. *Journal of Applied Physics*, 116(18), 2014.

65. M. Zeman, O. Isabella, K. Jger, R. Santbergen, S. Solntsev, M. Topic, and J. Krc. Advanced light management approaches for thin-film silicon solar cells. *Energy Procedia*, 15:189–199, 2012.
66. S. In, D.R. Mason, H. Lee, M. Jung, C. Lee, and N. Park. Enhanced light trapping and power conversion efficiency in ultrathin plasmonic organic solar cells: A coupled optical-electrical multiphysics study on the effect of nanoparticle geometry. *ACS Photonics*, 2(1):78–85, 2015.
67. M. Peters, B. Bläsi, S.W. Glunz, A.G. Aberle, J. Luther, and C. Battaglia. Optical simulation of silicon thin-film solar cells. *Energy Procedia*, 15:212–219, 2012.
68. A.H. Fallahpour, G. Ulisse, M. Auf der Maur, A. Di Carlo, and F. Brunetti. 3-D simulation and optimization of organic solar cell with periodic back contact grating electrode. *IEEE Journal of Photovoltaics*, 5(2):591–596, 2015.
69. M. Zeman, O. Isabella, S. Solntsev, and K. Jäger. Modelling of thin-film silicon solar cells. *Solar Energy Materials and Solar Cells*, 119:94–111, 2013.
70. N. Tucher, J. Eisenlohr, P. Kiefel, O. Höhn, H. Hauser, M. Peters, C. Müller, J.C. Goldschmidt, and B. Bläsi. 3D optical simulation formalism optos for textured silicon solar cells. *Optics Express*, 23(24):A1720–A1734, 2015.
71. S. Selberherr. *Analysis and Simulation of Semiconductor Devices*, 1st edition. New York, NY: Springer-Verlag Wien, 1984.
72. P.A. Markowich, C.A. Ringhofer, and C. Schmeiser. *Semiconductor Equations*, 1st edition. New York, NY: Springer-Verlag Wien, 1990.
73. W.V. van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell System Technical Journal*, 29:560–607, 1955.
74. S.M. Sze. *Semiconductor Devices: Physics and Technology*. New York, NY: John Wiley & Sons, 1985.
75. A. Schenk. An improved approach to the Shockley-Read-Hall recombination in inhomogeneous fields of spacecharge regions. *Journal of Applied Physics*, 71(7):3339–3349, 1992.
76. T. Goudon, V. Miljanovic, and C. Schmeiser. On the Shockley-Read-Hall model: Generation-recombination in semiconductors. *SIAM Journal on Applied Mathematics*, 67(4):1183–1201, 2007.
77. E. O'Reilly. *Quantum Theory of Solids*. London: Taylor & Francis, 2002.
78. S. John, C. Soukoulis, M.H. Cohen, and E.N. Economou. Theory of electron band tails and the urbach optical-absorption edge. *Physical Review Letters*, 57:1777–1780, 1986.
79. H. Bässler. Charge transport in disordered organic photoconductors: A Monte Carlo simulation study. *Physica Status Solidi (B)*, 175(1):15–56, 1993.
80. L.J.A. Koster, V.D. Mihailetchi, R. Ramaker, and P.W.M. Blom. Light intensity dependence of open-circuit voltage of polymer: Fullerene solar cells. *Applied Physics Letters*, 86(12):123509, 2005.
81. G. Paasch and S. Scheinert. Charge carrier density of organics with Gaussian density of states: Analytical approximation for the Gauss-Fermi integral. *Journal of Applied Physics*, 107(10):104501, 2010.
82. A. Zampetti, A.H. Fallahpour, M. Dianetti, L. Salamandra, F. Santoni, A. Gagliardi, M. Auf der Maur, F. Brunetti, A. Reale, T.M. Brown, and A. Di Carlo. Influence of the interface material layers and semiconductor energetic disorder on the open circuit voltage in polymer solar cells. *Journal of Polymer Science Part B: Polymer Physics*, 53(10):690–699, 2015.
83. G.K. Wachutka. Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *IEEE Transactions on Computer-Aided Design*, 11:1141–1149, 1990.
84. N.A. Zakhleniuk. Nonequilibrium drift-diffusion transport in semiconductors in presence of strong inhomogeneous electric fields. *Applied Physics Letters*, 89(25):252112, 2006.
85. N.F. Mott. On the transition to metallic conduction in semiconductors. *Canadian Journal of Physics*, 34(12A):1356–1368, 1956.
86. A. Miller and E. Abrahams. Impurity conduction at low concentrations. *Physical Review*, 120(3):745, 1960.
87. R.A. Marcus. Chemical and electrochemical electron-transfer theory. *Annual Review of Physical Chemistry*, 15(1):155–196, 1964.

88. W.F. Pasveer, J. Cottaar, C. Tanase, R. Coehoorn, P.A. Bobbert, P.W.M. Blom, D.M. De Leeuw, and M.A.J. Michels. Unified description of charge-carrier mobilities in disordered semiconducting polymers. *Physical Review Letters*, 94(20):206601, 2005.
89. D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.
90. A.F. Voter. Introduction to the kinetic Monte Carlo Method. *Radiation Effects*, 235:1–23, 2005.
91. H.L. Anderson. Scientific uses of the maniac. *Journal of Statistical Physics*, 43(5–6):731–748, 1986.
92. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087, 1953.
93. J.R. Beeler (Jr). Displacement spikes in cubic metals. i. α-Iron, copper, and tungsten. *Physical Review*, 150(2):470, 1966.
94. E.S. Hood, B.H. Toby, and W.H. Weinberg. Precursor-mediated molecular chemisorption and thermal desorption: The interrelationships among energetics, kinetics, and adsorbate lattice structure. *Physical Review Letters*, 55(22):2437, 1985.
95. A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. A new algorithm for Monte Carlo simulation of Ising spin systems. *Journal of Computational Physics*, 17(1):10–18, 1975.
96. C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation*. *Computational Microelectronics*. Vienna: Springer, 1989.
97. D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
98. J.R. Norris. *Markov Chains*. Number 2008. Cambridge: Cambridge University Press, 1998.
99. F. Schwabl. *Statistical Mechanics*. *Advanced Texts in Physics*. Berlin: Springer, 2006.
100. J. Ferber, R. Stangl, and J. Luther. An electrical model of the dye sensitized solar cell. *Solar Energy Materials and Solar Cells*, 53:29–54, 1998.
101. P.R.F. Barnes and B.C. O'Regan. Electron recombination kinetics and the analysis of collection efficiency and diffusion length measurements in dye sensitized solar cells. *The Journal of Physical Chemistry C*, 114(44):19134–19140, 2010.
102. A. Gagliardi, S. Mastroianni, D. Gentilini, F. Giordano, A. Reale, T.M. Brown, and A.D. Carlo. Multiscale modeling of dye solar cells and comparison with experimental data. *IEEE Journal of Selected Topics in Quantum Electronics,* 16(6):1611–1618, 2010.
103. A. Gagliardi, M. Auf der Maur, D. Gentilini, F. Di Fonzo, A. Abrusci, H.J. Snaith, G. Divitini, C. Ducati, and A. Di Carlo. The real TiO_2/HTM interface of solid-state dye solar cells: Role of trapped states from a multiscale modelling perspective. *Nanoscale*, 7(3):1136–1144, 2015.
104. A. Gagliardi, M. Auf der Maur, D. Gentilini, and A. Di Carlo. Simulation of dye solar cells: Through and beyond one dimension. *Journal of Computational Electronics*, 10(4):424–436, 2011.
105. A. Gagliardi, M. Auf der Maur, and A. Di Carlo. Theoretical investigation of a dye solar cell wrapped around an optical fiber. *IEEE Journal of Quantum Electronics*, 47(9):1214–1221, 2011.
106. J. Nissfolk, K. Fredin, A. Hagfeldt, and G. Boschloo. Recombination and transport processes in dye-sensitized solar cells investigated under working conditions. *The Journal of Physical Chemistry B*, 110(36):17715–17718, 2006.
107. U. Landau. Current distribution in electrochemical cells: Analytical and numerical modeling. In *Modern Aspects of Electrochemistry No. 44: Modelling and Numerical Simulations II*, pp. 451–501. New York, NY: Springer New York, 2009.
108. J. Bisquert and V.S. Vikhrenko. Interpretation of the time constants measured by kinetic techniques in nanostructured semiconductor electrodes and dye-sensitized solar cells. *Journal of Physical Chemistry B*, 108(7):2313–2322, 2004.
109. D. Gentilini, A. Gagliardi, M. Auf der Maur, L. Vesce, D. D'Ercole, T.M. Brown, A. Reale, and A. Di Carlo. Correlation between cell performance and physical transport parameters in dye solar cells. *Journal of Physical Chemistry C*, 116(1):1151–1157, 2012.

110. S.D. Baranovskii. Theoretical description of charge transport in disordered organic semiconductors. *Physica Status Solidi (B)*, 251(3):487–525, 2014.
111. P.K. Watkins, A.B. Walker, and G.L.B. Verschoor. Dynamical Monte Carlo modelling of organic solar cells: The dependence of internal quantum efficiency on morphology. *Nano Letters*, 5(9):1814–1818, 2005.
112. M. Casalegno, G. Raos, and R. Po. Methodological assessment of kinetic Monte Carlo simulations of organic photovoltaic devices: The treatment of electrostatic interactions. *Journal of Chemical Physics*, 132(9):094705, 2010.
113. F. Santoni, A. Gagliardi, M. Auf der Maur, and A. Di Carlo. The relevance of correct injection model to simulate electrical properties of organic semiconductors. *Organic Electronics*, 15(7):1557–1570, 2014.
114. I. Kamohara, M. Townsend, and B. Cottle. Simulation of heterojunction organic thin film devices and exciton diffusion analysis in stacked-hetero device. *Journal of Applied Physics*, 97(1):014501, 2005.
115. Il-S. Park, S.-R. Park, D.-Y. Shin, J.-S. Oh, W.-J. Song, and J.-H. Yoon. Modeling and simulation of electronic and excitonic emission properties in organic host/guest systems. *Organic Electronics*, 11(2):218–226, 2010.
116. B. Ruhstaller, T. Beierlein, H. Riel, S. Karg, J.C. Scott, and W. Riess. Simulating electronic and optical processes in multilayer organic light-emitting devices. *IEEE Journal of Selected Topics in Quantum Electronics*, 9(3):723–731, 2003.
117. D. Kipp and V. Ganesan. A kinetic Monte Carlo model with improved charge injection model for the photocurrent characteristics of organic solar cells. *Journal of Applied Physics*, 113(23):234502, 2013.
118. N.J. van der Kaap and L.J.A. Koster. Massively parallel kinetic monte carlo simulations of charge carrier transport in organic semiconductors. *Journal of Computational Physics*, 307:321–332, 2016.
119. L. Pedesseau, J.-M. Jancu, A. Rolland, E. Deleporte, C. Katan, and J. Even. Electronic properties of 2D and 3D hybrid organic/inorganic perovskites for optoelectronic and photovoltaic applications. *Optical and Quantum Electronics*, 46(10):1225–1232, 2014.
120. A. Pecchia, D. Gentilini, D. Rossi, M. Auf der Maur, and A. Di Carlo. Role of ferroelectric nanodomains in the transport properties of perovskite solar cells. *Nano Letters*, 16(2):988–992, 2016.
121. D. Liu, M.K. Gangishetty, and T.L. Kelly. Effect of CH3NH3PbI3 thickness on device efficiency in planar heterojunction perovskite solar cells. *Journal of Materials Chemistry A*, 2:19873–19881, 2014.
122. S. Meloni, T. Moehl, W. Tress, M. Franckeviius, M. Saliba, Y.H. Lee, P. Gao, M.K. Nazeeruddin, S.M. Zakeeruddin, U. Rothlisberger, and M. Graetzel. Ionic polarization-induced current-voltage hysteresis in $CH_3NH_3PbX_3$ perovskite solar cells. *Nature Communications*, 7, 2016, Article number: 10334. doi:10.1038/ncomms10334.
123. J.M. Azpiroz, E. Mosconi, J. Bisquert, and F. De Angelis. Defect migration in methylammonium lead iodide and its role in perovskite solar cell operation. *Energy and Environmental Science*, 8(7):2118–2127, 2015.
124. T. Albes, P. Lugli, and A. Gagliardi. Investigation of the blend morphology in bulk-heterojunction organic solar cells. *IEEE Transactions on Nanotechnology*, 15(2):281–288, 2016.
125. H.K. Kodali and B. Ganapathysubramanian. A computational framework to investigate charge transport in heterogeneous organic photovoltaic devices. *Computer Methods in Applied Mechanics and Engineering*, 247:113–129, 2012.
126. A. Bolognesi, M. Berliocchi, M. Manenti, A. Di Carlo, P. Lugli, K. Lmimouni, and C. Dufour. Effects of grain boundaries, field-dependent mobility, and interface trap states on the electrical characteristics of pentacene TFT. *IEEE Transactions on Electron Devices*, 51(12):1997–2003, 2004.
127. Y. Hu, L. Wang, Q. Qi, D. Li, and C. Jiang. Charge transport model based on single-layered grains and grain boundaries for polycrystalline pentacene thin-film transistors. *Journal of Physical Chemistry C*, 115(47):23568–23573, 2011.

128. M. Kimura, S. Inoue, T. Shimoda, and T. Sameshima. Device simulation of carrier transport through grain boundaries in lightly doped polysilicon films and dependence on dopant density. *Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers*, 40(9A):5237–5243, 2001.
129. W.K. Metzger and M. Gloeckler. The impact of charged grain boundaries on thin-film solar cells and characterization. *Journal of Applied Physics*, 98(6):063701, 2005.
130. K. Taretto and U. Rau. Numerical simulation of carrier collection and recombination at grain boundaries in Cu(In,Ga)Se_2 solar cells. *Journal of Applied Physics*, 103(9), 2008.
131. H. Kobayashi and Y. Tokita. Modeling of hole transport across grain boundaries in organic semiconductors for mesoscale simulations. *Applied Physics Express*, 8(5):051602, 2015.

IX

Novel Applications

44 Electroluminescent Refrigerators *Kuan-Chen Lee and Shun-Tung Yen* 541
Introduction • Theoretical Methods • Results and Discussion • Summary

45 Photonic Crystal Laser Diodes *Maciej Dems* 561
Photonic Crystals Basics • Fully Vectorial Method for Modeling VCSELs with Subwavelength Structures • VCSELs with Photonic Crystals for Light Confinement • Subwavelength Gratings in VCSELs • Summary

46 Single-Photon Sources *Niels Gregersen, Dara P. S. McCutcheon, and Jesper Mørk* 585
Introduction • Modeling of Spontaneous Emission • Cavity QED Approach • Waveguide QED Approach • Planar Dielectric Antenna Approach • Decoherence Effects • Summary

47 Nanoplasmonic Lasers and Spasers *A. Freddie Page and Ortwin Hess* 609
Introduction • Models for Metals and Gain • Transfer Matrix Method • Evolutionary Algorithm • Finite-Difference Time-Domain Method • Loss Compensation • Lasing Dynamics • Stopped-Light Lasing • Summary

48 Quantum-Dot Nanolasers *Christopher Gies, Michael Lorke, Frank Jahnke, and Weng W. Chow* 627
Going Nano: A New Era of Small Lasers • Connecting to Earlier Work: Laser Rate Equations and the β Factor • Key Aspects of Nanolasers • Essential Physics I: Quantum-Optical Semiconductor Laser Theory • Essential Physics II: Carrier Kinetics in QD Systems • Connecting to Experiments: Transition to Lasing and the β Factor • Interemitter Coupling Effects and Superradiance in Nanolasers • Nonresonant Coupling and Lasing from Multi-Exciton States in Few-QD Systems • Outlook

49 Nonlinear Dynamics in Quantum Photonic Structures *Gabriela Slavcheva and Mirella Koleva* 661
Introduction • Maxwell-Pseudospin Method • Simulation of Nonlinear Dynamics in Optical Waveguides and Semiconductor Microcavities • Optical Orientation in Quantum Wells and QDs • Modeling Coherent Spin Generation, Manipulation, and Readout in Charged QDs • Quantum Stochastic Formalism for Modeling Cavity-Emitter Systems • Modeling Optical Rotation in Chiral Carbon Nanotubes with Master Maxwell-Pseudospin Equations • Conclusion

44

Electroluminescent Refrigerators*

Kuan-Chen Lee
and
Shun-Tung Yen

44.1 Introduction 541
44.2 Theoretical Methods 542
Fundamentals of EL Refrigeration • Device Structure and Main Equations • Recombination and Generation Processes • Boundary Conditions • Output Quantities and Models for Steady-State Analysis
44.3 Results and Discussion 549
Influence of Device Structures on EL Refrigeration • Influence of Optical Processes on EL Refrigeration
44.4 Summary 557

44.1 Introduction

Solid-state optical refrigeration, which is featured by carrying the internal thermal energy away through light emission, has progressed rapidly after the first demonstration of photoluminescent (PL) refrigeration in 1995 (Epstein et al., 1995). To date, it has been shown that an Yb-doped $LiYF_4$ cooler can reach a temperature of 114 K and give a cooling power of 750 mW at room temperature (Melgaard et al., 2014). Further improving this kind of rare-earth-doped coolers to an even lower cooling temperature will be a difficult task. This is because the cooling process involves the thermalization of dopants and becomes inefficient when the thermal energy is comparable to the energy difference between discrete dopant levels in the ground-state manifold (Sheik-Bahae and Epstein, 2007). Different to the doped coolers, a semiconductor luminescent refrigerator is not restricted by such limitation and is expected to have lower operation temperature and higher cooling power. Moreover, it can be directly integrated with other semiconductor devices. These attractive features have stimulated extensive research on semiconductor luminescent refrigeration (Sheik-Bahae and Epstein, 2007, 2009). However, the required external efficiency for cooling is as high as nearly unity such that the realization of semiconductor PL refrigeration was not achieved until recently by Zhang et al. (2013) on CdS mircorod.

Semiconductor luminescent refrigeration using electrical pumping scheme, on the other hand, receives growing attention because of its looser requirement on the external efficiency compared to semiconductor PL refrigeration. Despite several works on the feasibility of electroluminescent (EL) refrigeration several decades ago (Berdahl, 1985; Dousmanis et al., 1964), the experimental verification of EL cooling has been reported just recently. In 2012, Santhanam et al. (2012) biased a light-emitting diode (LED) under several tens of microvolts and showed that the output emission power can exceed twice of the applied electrical

* This chapter is reproduced with permission from Yen and Lee (2010), and Lee and Yen (2012).

input power. This positive difference of the output over the input powers implies that the LED is capable of extracting thermal energy from the environments, as a refrigerator, to keep heat flow balanced. However, the cooling power is smaller than nanowatts so that the cooling phenomenon is not directly observed. It has been known for a while that the cooling power density of an EL cooler can be significantly high if the cooler is biased at a voltage of $\sim E_g/q$, where E_g is the bandgap energy of the active layer and q the elementary charge. Extensive studies have been made to investigate the cooling feasibility and capability of EL refrigeration in this bias region. Mal'shukov and Chao studied the influence of the Auger recombination on the cooling capability of a GaAs-based refrigerator. They showed that a net cooling power density of several watts per square centimeter is achievable (Mal'shukov and Chao, 2001). Wang et al. performed a detailed self-consistent calculation with various recombination mechanisms accounted. Their analysis showed an impressive cooling efficiency of 35% for a cooler of p-GaAs active layer (Wang et al., 2005). Oksanen and Tulkki (2010) proposed a thermal heat pump, which consists of an EL cooler integrated directly with an absorbing diode, to deal with the problem of low light extraction efficiency encountered in semiconductor luminescent refrigeration. Other theoretical works, such as on heat transfer and cooling temperature (Han et al., 2007; Heikkilä et al., 2010; Lin et al., 2012; Piprek and Li, 2016), the design of refrigerator structure (Oksanen and Tulkki, 2010; Yen and Lee, 2010), the influence of the active materials (Yu et al., 2007), light extraction efficiency (Yen and Lee, 2010), and the photon recycling effect (Heikkilä et al., 2009; Lee and Yen, 2012), has also been conducted to provide a deep insight into EL refrigeration. Experimental works on the feasibility (Xue et al., 2015) and the application (Liu et al., 2013) of EL refrigeration are also being conducted. Although the high-cooling-power EL refrigeration has not been achieved, a thorough understanding of its operation mechanism still plays a necessary role in its realization.

In the following sections, we begin with the fundamental concept of EL refrigeration. The self-consistent calculation generally used for analyzing EL refrigeration and the operation principle of EL refrigeration is introduced in Section 44.2. Calculated results together with the analysis of various structural and optical effects on EL refrigeration are given in Section 44.3. Finally, a conclusion is given in Section 44.4.

44.2 Theoretical Methods

44.2.1 Fundamentals of EL Refrigeration

In EL refrigeration, electrons and holes are mediated as refrigerants and are intended to be injected into the active region for radiative recombination, which generates photons with thermal energy carried away. Recently, studies indicate that the extracted thermal energy comes from the thermalization of carriers during transport, similar to the Peltier effect in thermoelectric (TE) cooling (Santhanam et al., 2012; Oksanen and Tulkki, 2015; Xue et al., 2015). This process can be seen during the cooling cycle of EL refrigeration as illustrated in Figure 44.1a. At terminal contacts, carriers are injected into the device under an external bias V and flow all the way to the active layer. In the process of carrier transport, local cooling occurs in the place where the external bias is not large enough for carriers to overcome the potential barriers. This situation can be seen, for instance, in the cladding layer near the active layer. In such regions, only carriers of sufficiently high energy can surmount the barrier and move forward as shown in the inset circle. This causes high-energy empty states that are then occupied by carriers through thermalization, i.e., drawing phonons from the lattice matrix. The redistribution of carriers progresses rapidly and hence the quasi-thermal equilibrium is achieved as a result of much short phonon relaxation time. In the active layer, the injected electron–hole pairs recombine through various mechanisms. A heating energy of qV is generated if the electron–hole pair recombines nonradiatively. Contrary, for each photon generated from radiative recombination escaping out of the device, the internal energy of the whole system is reduced by $\hbar\omega - qV$. Here, $\hbar\omega$ is the photon energy, which comes from the combination of the thermal energy from the lattice matrix and the electric potential energy from the external bias. These three processes, i.e., electrical excitation from external bias, thermalization during transport, and relaxation through radiative recombination, complete the cooling cycle of EL refrigeration. Figure 44.1b shows the system view of an ideal EL cooler

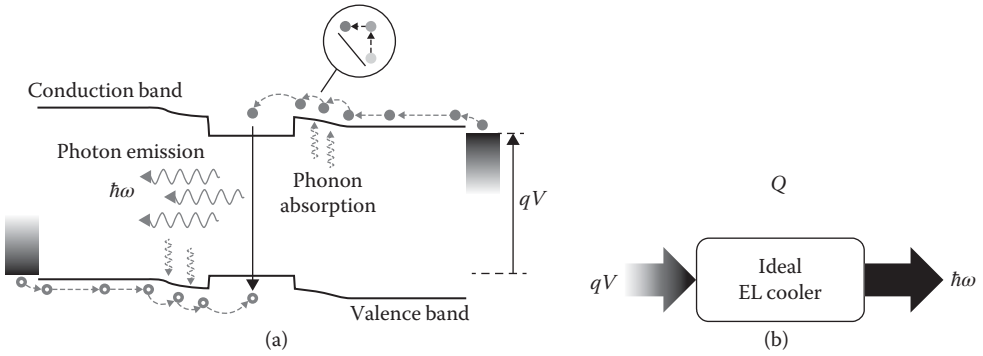


FIGURE 44.1 (a) The illustration of cooling cycling for EL refrigeration. The inset circle shows the thermalization process of carriers, which happens in the cladding layer near the active region. This is similar to the Peltier effect in thermoelectric cooling and attributed to EL refrigeration. (b) The system view of an ideal EL cooler. Net cooling occurs when $Q > 0$; otherwise, the device generates heat.

for a single cooling cycle, in which the parasitic effects are omitted. In this figure, an energy flow per cycle $Q = \hbar\omega - qV$ is required to be drawn from the ambient environment for compensating the internal energy loss of the system. If $Q > 0$, heat flows into the device and net cooling occurs; otherwise, the device generates heat. In this chapter, the cooling power, one of the most important parameters for coolers, is obtained in a similar way as Q described earlier, for which the input and the output powers are calculated from a self-consistent calculation based on one-dimensional drift-diffusion model with various recombination mechanisms accounted. In such a picture, the cooler is assumed to be in direct contact with a heat reservoir and the thermal resistance inside the cooler is neglected so that the temperature of the whole device is kept at a constant value of 300 K. In reality, a temperature gradient between the cooler and its environment exists due to a finite thermal conductance. The temperature distribution may also be nonuniform inside the cooler because of a nonuniform distribution of TE heat exchange (Yu et al., 2007). To account these effects, the heat balance equation between the cooler and its environment and the heat transfer equation inside the cooler have to be considered. Interested readers can refer to the works of Ashcroft and Mermin (1976), Han et al. (2007), Heikkilä et al. (2010), Lin et al. (2012), and Piprek and Li (2016) for more information. The geometric effects, such as current spreading and the detailed optical processes, are not accounted either. The consideration of these effects relies on the detailed information about the device geometry and a full three-dimensional simulation, which is beyond the scope of this chapter. The goal of this chapter is to provide a comprehensive view on how various structural and optical components affect the performance of an EL cooler without losing accuracy.

44.2.2 Device Structure and Main Equations

The device structure used in this work is composed of seven layers, as shown in Figure 44.2. They are, in order, a 100-nm p-doped $Al_{0.25}Ga_{0.75}As$ cladding layer (with an acceptor concentration of $N_a = 1 \times 10^{18}$ cm^{-3}), a 50-nm p-doped $Al_xGa_{1-x}As$ carrier blocking layer ($N_a = 1 \times 10^{18}$ cm^{-3}), a 50-nm undoped $Al_{0.25}Ga_{0.75}As$ spacer, an undoped GaAs active layer of thickness $L = 100$ nm, a 50-nm undoped $Al_{0.25}Ga_{0.75}As$ spacer layer, a 50-nm n-doped $Al_xGa_{1-x}As$ carrier blocking layer (with a donor concentration of $N_d = 1 \times 10^{18}$ cm^{-3}), and a 100-nm n-doped $Al_{0.25}Ga_{0.75}As$ cladding layer ($N_d = 1 \times 10^{18}$ cm^{-3}). Figure 44.2 also demonstrates the calculated energy band profiles, including the band edges of the conduction band E_c and the valence band E_v, the electron quasi-Fermi level E_{Fc}, and the hole quasi-Fermi level E_{Fv}, along the growth direction of z under a forward bias of $V = 1.4$ V. The Al content x is 0.4 in the carrier blocking layer. For the drift-diffusion model, the central problem is to find out the electrostatic potential Φ

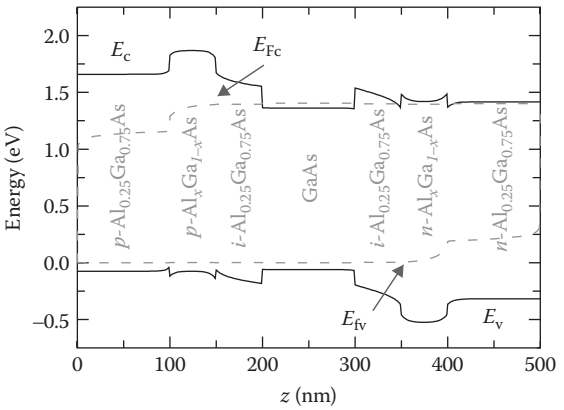


FIGURE 44.2 The energy band profiles, E_c and E_v, and the quasi-Fermi level profiles, E_{Fc} and E_{Fv}, along the growth direction z of the conduction and the valence bands, respectively, for the heterostructure under consideration. The profiles are obtained under a forward bias of 1.4 V.

and the quasi-Fermi levels for electrons E_{Fc}, and for holes E_{Fv} by solving three coupled equations (Horio and Yanai, 1990; Yen and Lee, 2010). These are the one-dimensional Poisson equation in the steady state:

$$\frac{\mathrm{d}}{\mathrm{d}z}\epsilon\frac{\mathrm{d}}{\mathrm{d}z}\Phi = -q\left(p - n + N_\mathrm{d}^+ - N_\mathrm{a}^-\right), \tag{44.1}$$

and the one-dimensional continuity equations for the electron and the hole currents:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}J_\mathrm{n} &= q\,(R-G),\\ \frac{\mathrm{d}}{\mathrm{d}z}J_\mathrm{p} &= -q\,(R-G),\end{aligned} \tag{44.2}$$

where ϵ is the electric permittivity; p, n, N_d^+, and N_a^- are the concentrations of the holes, electrons, ionized donors, and ionized acceptors, respectively; J_n (J_p) is the current densities due to the electron (hole) flow; R is the electron–hole recombination rate minus the intrinsic generation rate due to thermal excitation; and G is the electron–hole generation rate due to the photon recycling effect to be described. It is noted that all of the aforementioned $p, n, N_\mathrm{d}^+, N_\mathrm{a}^-, J_\mathrm{n}, J_\mathrm{p}, R$, and G are functions of Φ, E_{Fc}, and E_{Fv}. The first six ones can be expressed by the following formulas:

$$n = N_\mathrm{c}F_{1/2}\left(\eta_\mathrm{n}\right), \eta_\mathrm{n} \equiv \left(E_{Fc} - E_\mathrm{c}\right)/k_\mathrm{B}T, \tag{44.3}$$

$$p = N_\mathrm{v}F_{1/2}\left(\eta_\mathrm{p}\right), \eta_\mathrm{p} \equiv \left(E_\mathrm{v} - E_{Fv}\right)/k_\mathrm{B}T, \tag{44.4}$$

$$N_\mathrm{d}^+ = \frac{N_\mathrm{d}}{\left\{1 + 2\exp\left[\left(E_{Fc} - E_\mathrm{d}\right)/k_\mathrm{B}T\right]\right\}}, \tag{44.5}$$

$$N_\mathrm{a}^- = \frac{N_\mathrm{a}}{\left\{1 + 4\exp\left[\left(E_\mathrm{a} - E_{Fv}\right)/k_\mathrm{B}T\right]\right\}}, \tag{44.6}$$

$$J_\mathrm{n} = -q\mu_\mathrm{n}n\mathrm{d}\Phi/\mathrm{d}z + qD_\mathrm{n}\mathrm{d}n/\mathrm{d}z, \tag{44.7}$$

$$J_\mathrm{p} = -q\mu_\mathrm{p}p\mathrm{d}\Phi/\mathrm{d}z - qD_\mathrm{p}\mathrm{d}p/\mathrm{d}z, \tag{44.8}$$

where N_c (N_v) is the effective density of states for the conduction (valence) band, F_j the Fermi–Dirac integral of order j, k_B the Boltzmann constant, T the temperature, E_d (E_a) the donor (acceptor) level, μ_n (μ_p) the electron (hole) mobility, and D_n (D_p) the diffusivity for the electrons and the holes.

44.2.3 Recombination and Generation Processes

For the recombination processes in the device, the Shockley–Read–Hall (SRH), the radiative, and the Auger recombinations are considered in the high-quality GaAs active layer, while only the SRH recombination process is assumed to be dominant in the other layers containing aluminum. Therefore, the recombination rate is

$$R = \begin{cases} R_{SRH} + R_{rad} + R_{Aug} & z \text{ in the active layer,} \\ R_{SRH} & \text{otherwise,} \end{cases} \tag{44.9}$$

where R_{SRH}, R_{rad}, and R_{Aug} are the recombination rates due to the SRH, the radiative, and the Auger recombination processes, respectively. In the calculation of R_{rad}, we consider the spontaneous emission contributed from the transition from the conduction band (c) to both the heavy-hole (hh) and the light-hole (lh) bands and express R_{rad} as

$$R_{rad} = \int r(\hbar\omega)\, d\hbar\omega, \; r = r_{hh} + r_{lh}, \tag{44.10}$$

where $\hbar$ is the reduced Planck constant, ω the optical angular frequency, and r the total spontaneous emission rate that consists of r_{hh} and r_{lh} attributed to the transitions of c→hh and c→lh, respectively. With the parabolic band approximation, r_{hh} can be expressed as (Chuang, 1995)

$$r_{hh}(\hbar\omega) = C_{hh} \frac{\hbar\omega\sqrt{\hbar\omega - E_g}\,\Theta(\hbar\omega - E_g)}{\left[1 + e^{-\eta_n} \exp\left(\frac{m_r}{m_c}\frac{\hbar\omega - E_g}{k_B T}\right)\right]\left[1 + e^{-\eta_P} \exp\left(\frac{m_r}{m_{hh}}\frac{\hbar\omega - E_g}{k_B T}\right)\right]}, \tag{44.11}$$

where

$$C_{hh} = \frac{q^2 n_r m_r^{3/2} E_P}{3\sqrt{2}\pi^3 \epsilon_0 m_0 \hbar^5 c^3}. \tag{44.12}$$

ϵ_0, m_0, and c are the electric permittivity, the electron mass, and the light speed, *in vacuo*, respectively; n_r is the refractive index; E_P is the optical matrix parameter; $m_r = m_c m_{hh}/(m_c + m_{hh})$ is the reduced mass; and Θ is the unit step function. For r_{lh}, it is simply obtained by replacing the role of m_{hh} in Equations 44.11 and 44.12 with the effective mass of the light hole band, m_{lh}.

For the Auger recombination rate, the CHHS and CHCC processes are considered. Therefore, R_{Aug} is

$$R_{Aug} = \left(C_p p + C_n n\right)\left(np - n_i^2\right), \tag{44.13}$$

where C_p (C_n) is the Auger coefficient corresponding to the CHHS (CHCC) process and n_i the intrinsic carrier density. For the SRH recombination, we use

$$R_{SRH} = \frac{pn - n_i^2}{\tau_n\left(p + n_i\right) + \tau_p\left(n + n_i\right)}, \tag{44.14}$$

where τ_n (τ_p) is the minority carrier lifetime when $p \gg n, n_i$ ($n \gg p, n_i$).

For the generation rate G, the carriers generated by photon recycling are accounted in the active layer. Photon recycling is a process involving the absorption of spontaneously emitted photons by the active material itself and hence strongly related to the distribution of optical field in the device. A detailed consideration of the recycling process relies on the information about the device geometry, which is out of the scope for the present work. In this work, the generation rate G is directly connected to R_{rad} by

$$G = \frac{1}{L}\eta_{\mathrm{pr}}\left(1-\eta_{\mathrm{xp}}\right) \int\limits_{z\in \text{ active layer}} R_{\mathrm{rad}}\,dz, \tag{44.15}$$

where η_{xp} is the light extraction efficiency, the probability for the spontaneously generated photons emitting out of the device, and η_{pr} is the photon recycling efficiency, the ratio of trapped photons eventually being absorbed by the active layer. Therefore, Equation 44.15 implies that the photon recycling process generates carriers uniformly in the active layer, which results from the absorption of a portion of spontaneously generated photons trapped in the device. The assumption of a uniform carrier generation rate is justifiable for an active layer thickness no longer than the inverse of the absorption coefficient.

44.2.4 Boundary Conditions

To obtain an unambiguous distribution of Φ, E_{Fc}, and E_{Fv}, the boundary conditions for Equations 44.1 and 44.2 have to be specified. For the boundary condition of Φ, it is simply assumed that the electrostatic potential Φ and the displacement field $-\epsilon \mathrm{d}\Phi/\mathrm{d}z$ are continuous at interfaces. The boundary conditions of E_{Fc} and E_{Fv} are, however, complicated and classified into two cases according to the magnitude of band offsets. For the case of a band offset smaller than $2k_{\mathrm{B}}T$, the quasi-Fermi levels are determined from the continuity of drift-diffusion currents at the both sides of the interface under the assumption of a continuous quasi-Fermi level. For the other case of a larger band offset, the drift-diffusion model is no longer applicable for carriers' transport between layers. In this case, the thermionic process has to be taken into account, which allows quasi-Fermi levels to split at the interface. The thermionic emission current density due to the electron flow at the interface z_{i} with Fermi–Dirac distribution accounted can be expressed as[†]

$$J_{\mathrm{n}}\left(z_{\mathrm{i}}\right) = A_{\mathrm{n}}\left(z_{\mathrm{i}}^{+}\right)T^{2}F_{1}\left[\xi_{\mathrm{n}}\left(z_{\mathrm{i}}^{+}\right)\right] - A_{\mathrm{n}}\left(z_{\mathrm{i}}^{-}\right)T^{2}F_{1}\left[\xi_{\mathrm{n}}\left(z_{\mathrm{i}}^{-}\right)\right], \tag{44.16}$$

where $z_{\mathrm{i}}^{\pm} = \lim_{\varepsilon\to 0}\left(z_{\mathrm{i}} \pm \varepsilon\right)$, $A_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right)$ is the effective Richardson constant of electrons at $z_{\mathrm{i}}^{\pm}$, and $\xi_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right)$ is the energy difference, normalized by $k_{\mathrm{B}}T$, from the quasi-Fermi levels at $z_{\mathrm{i}}^{\pm}$ to the larger one of the conduction band edges at the interface. By matching the drift-diffusion current to the thermionic current at interfaces, we can write the boundary condition for E_{Fc} as

$$\begin{aligned} &J_{\mathrm{n}}\left(z_{\mathrm{i}}\right) = J_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right), && \text{for } \left|E_{\mathrm{c}}\left(z_{\mathrm{i}}^{+}\right) - E_{\mathrm{c}}\left(z_{\mathrm{i}}^{-}\right)\right| \geq 2k_{\mathrm{B}}T, \\ &J_{\mathrm{n}}\left(z_{\mathrm{i}}^{+}\right) = J_{\mathrm{n}}\left(z_{\mathrm{i}}^{-}\right), && \text{otherwise,} \end{aligned} \tag{44.17}$$

with $J_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right)$ being the drift-diffusion current at $z_{\mathrm{i}}^{\pm}$. The boundary condition for E_{Fv} is determined by the continuity of J_{p} at the interface, similar to that for E_{Fc} just described above.

At the end of surfaces on which the metal contacts are to be made, the Fermi levels are assumed to be pinned at a fixed position from the band edges such that the carrier densities equal to the values of those

[†] It is noted that, at a low temperature, the contribution of the tunneling current through an abrupt heterojunction may not be neglected. In such a situation, Equation 44.16 has to be modified. Interested readers can refer to Yang et al. (1993) for detailed expression.

in thermal equilibrium. The Fermi level positions at the surface are thus determined by the charge neutral condition, $p + N_d^+ = n + N_A^-$, with $E_{Fc} = E_{Fv}$.

With the boundary conditions being specified, the differential Equations 44.1 and 44.2 can be solved steadily by the Newton–Raphson iteration scheme for each applied voltage V. The external electrical work qV done on each of injected carriers gives the Fermi level difference between two surfaces $E_{Fc}\big|_{z \text{ at cathode}} - E_{Fc}\big|_{z \text{ at anode}} = qV$. Parameters used in this work are listed in Table 44.1, except for mobility, which is obtained from an empirical model (Sotoodeh et al., 2000). The band offset ratio ($\Delta E_c/\Delta E_v$) is set as 0.6:0.4 for GaAs/$Al_xGa_{1-x}As$ interfaces. The binding energies of donors and acceptors are set at $k_B T$ for simplicity.

44.2.5 Output Quantities and Models for Steady-State Analysis

Once the solutions Φ, E_{Fc}, and E_{Fv} are obtained, the quantities of interest, including various current components, power densities, and efficiencies, can be calculated and used for further analysis. Figure 44.3 shows the working principle of an EL refrigerator. In the steady state, electrons and holes are injected

TABLE 44.1 Material Parameters of GaAs and AlGaAs at 300 K Used for Calculation

Parameter	Symbol	Unit	GaAs	$Al_xGa_{1-x}As$ ($x < 0.45$)
Bandgap	E_g	eV	$1.424 + 1.247x$	
Effective mass	m_c	m_0	$0.063 + 0.083x$	
	m_{hh}		$0.51 + 0.25x$	
	m_{lh}		$0.082 + 0.068x$	
Effective density of states	N_c	cm^{-3}	$2.5 \times 10^{19}(m_c/m_0)^{3/2}$	
	N_v		$2.5 \times 10^{19}[(m_{hh}/m_0)^{3/2} + (m_{lh}/m_0)^{3/2}]$	
Effective Richardson constant	A_n	$A/cm^2/K^2$	$120 \times (m_c/m_0)$	
	A_p		$120 \times (m_{hh} + m_{lh})/m_0$	
Electric permittivity	ϵ	ϵ_0	$12.9 - 2.84x$	
Refractive index	n_r	–	3.6	–
Optical matrix parameter	E_p	eV	25.7	–
SRH recombination lifetime	τ_n	s	1.3×10^{-6}	1×10^{-8}
	τ_p		1.2×10^{-6}	1×10^{-8}
Auger coefficient	C_n	cm^6/s	1.0×10^{-31}	–
	C_p		1.2×10^{-30}	–

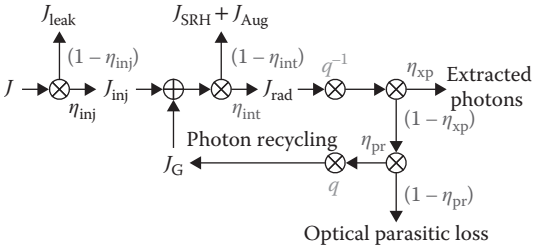


FIGURE 44.3 A diagram illustrating of the operation of an electroluminescent refrigerator with photon recycling as a feedback. It indicates the relationships among various currents. Symbol + stands for an addition operation that sums its inputs, indicated by inward arrows with the tips pointing at the symbol, and gives its output, indicated by an outward arrow with the tip pointing away from the symbol. Symbol × stands for multiplication operations, which give their outputs, indicated by outward arrows, by multiplying their inputs, indicated by inward arrows, by the factors along with the individual outward arrows.

from the electrodes under a forward bias, leading to a total current J in the device. Most of the electrons and holes are intended to be injected into the active region, recombine therein, and result in a major current component, called the injection current J_{inj}. The injection efficiency η_{inj} is defined as the ratio J_{inj}/J. The other part of the small current, called the leakage current J_{leak}, arises from the SRH recombination outside the active region $J'_{\text{SRH}} = \int_{z\notin\text{active layer}} R_{\text{SRH}}\text{d}z$ and the recombination current at the device surfaces $J_{\text{sr}} = J_{\text{n}}\big|_{z\text{ at anode}} + J_{\text{p}}\Big|_{z\text{ at cathode}}$. In this figure, we see that there are two means of injecting carriers into the active region. One is by the external electric injection and the other is by the photon recycling effect, which result in two effective current components J_{inj} and J_{G}, respectively. As a result, the sum $J_{\text{inj}} + J_{\text{G}}$ accounts for the total effective injection rate of electrons and holes to the active region. In the steady-state condition, the injection rate is balanced by the recombination rate in the active layer. Hence,

$$J_{\text{inj}} + J_{\text{G}} = J_{\text{SRH}} + J_{\text{rad}} + J_{\text{Aug}}, \tag{44.18}$$

where J_{SRH}, J_{rad}, and J_{Aug} are the current components in the active region due to the SRH, the radiative, and the Auger recombinations, respectively, with $J_i = q\int_{z\in\text{active region}} R_i\text{d}z$, for i = SRH, rad, and Aug. The internal quantum efficiency η_{int} is naturally defined as the probability for electron–hole pairs recombining through the radiative recombination:

$$\eta_{\text{int}} = \frac{J_{\text{rad}}}{J_{\text{SRH}} + J_{\text{rad}} + J_{\text{Aug}}}. \tag{44.19}$$

Owing to the radiative recombination, photons are generated with a rate J_{rad}/q. Since only a ratio η_{xp} of generated photons can escape out of the device, the emission flux is $\eta_{\text{xp}}J_{\text{rad}}/q$ and the corresponding emission power is:

$$P_{\text{rad}} = \eta_{\text{xp}}\int\int r\hbar\omega\text{d}\hbar\omega\text{d}z \equiv \eta_{\text{xp}}\langle\hbar\omega\rangle J_{\text{rad}}/q, \tag{44.20}$$

where $\langle\hbar\omega\rangle$ is the average photon energy. In addition to these emitted photons, the others are trapped in the device, either being parasitically absorbed, causing heat, or being recycled, reincarnating into electron–hole pairs in the active region. Figure 44.3 shows that the photon recycling effect is a feedback process, supplying a portion of electron–hole pairs required for recombinations. From Figure 44.3, we can simply express J_{G} as $J_{\text{G}} = \eta_{\text{pr}}\left(1-\eta_{\text{xp}}\right)J_{\text{rad}}$, which is equivalent to Equation 44.15 with the equality $J_{\text{G}} = qGL$.

In conventional optoelectronic devices, the external quantum efficiency η_{ext} is defined as the ratio of the photon flux emitted out of the devices to the carrier flux injected into the devices, i.e., $\eta_{\text{ext}} = \eta_{\text{xp}}J_{\text{rad}}/J$. Using Equation 44.18 and the definitions of various efficiencies, η_{ext} with photon recycling effect accounted can be expressed in terms of the other efficiencies:

$$\eta_{\text{ext}} = \frac{\eta_{\text{xp}}\eta_{\text{int}}\eta_{\text{inj}}}{1-\eta_{\text{int}}\eta_{\text{pr}}\left(1-\eta_{\text{xp}}\right)}. \tag{44.21}$$

In the absence of photon recycling ($\eta_{\text{pr}} = 0$), η_{ext} returns to the conventional form of $\eta_{\text{ext}} = \eta_{\text{xp}}\eta_{\text{int}}\eta_{\text{inj}}$.

For the purpose of refrigeration, the cooling power and the cooling efficiency are two important parameters of interests. Following the concepts shown in Figure 44.1b, we define the cooling power as the power difference of the output and the input powers, which is required to keep the cooler operating steadily at 300 K:

$$P_{\text{c}} = P_{\text{rad}} - JV = \eta_{\text{c}}JV, \tag{44.22}$$

where $\eta_c \equiv P_c/JV$ is known as the cooling efficiency. The notion of EL refrigeration is to set the devices in the cooling mode, which requires a positive cooling power or cooling efficiency. Using Equation 44.20 and the definition of η_{ext}, one can rewrite the cooling efficiency as

$$\eta_c = \frac{\eta_{ext}}{\eta_{ext,cr}} - 1, \tag{44.23}$$

where

$$\eta_{ext,cr} \equiv \frac{qV}{\langle \hbar\omega \rangle}, \tag{44.24}$$

is the critical external efficiency for the devices to act in the cooling mode, which is consistent with the result from Figure 44.1b if the external efficiency of the cooler is considered. From Equation 44.21, the critical extraction efficiency for cooling is:

$$\eta_{xp,cr} = \frac{1 - \eta_{pr}\eta_{int}}{\eta_{int}\left(\eta_{inj}/\eta_{ext,cr} - \eta_{pr}\right)}, \tag{44.25}$$

for $\eta_{inj}/\eta_{ext,cr} > \eta_{pr}$.

44.3 Results and Discussion

In this section, we present the results simulated by the aforementioned self-consistent calculation to investigate various structural and optical effects on the cooling feasibility and capability of GaAs/AlGaAs EL refrigerators. In the discussion of the structural effects, we have assumed an unity light extraction efficiency ($\eta_{xp} = 1$).

44.3.1 Influence of Device Structures on EL Refrigeration

44.3.1.1 Leakage Current and Carrier Blocking Layer

We first study the cooling feasibility of a simple double heterostructure (DH) EL cooler by plotting the internal, the external, and the critical efficiencies shown in Figure 44.4 as functions of the applied voltage

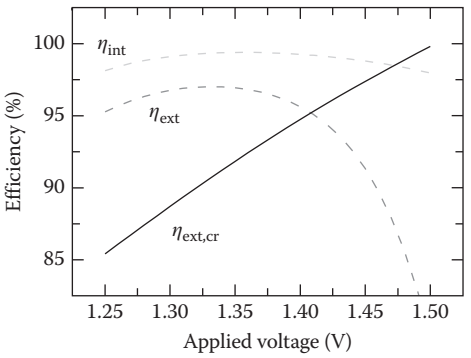


FIGURE 44.4 The internal efficiency η_{int}, the external efficiency η_{ext}, and the critical external efficiency $\eta_{ext,cr}$ as functions of the applied voltage V for the heterostructure with an active layer of width $L = 0.1$ μm and carrier blocking layers of Al composition $x = 0.25$, which has no effect on blocking carriers.

V for $\eta_{xp} = 1$. The Al content x in the blocking layer is 0.25, which is identical with that in the cladding and the spacer layers so that there is no blocking effect. As can be seen, while the critical external efficiency $\eta_{ext,cr}$ increases linearly with the applied voltage V, the internal efficiency η_{int} first increases to a maximum at $V = 1.36$ V and then decreases. The variation of η_{int} on V reveals a more important role played by the SRH and the Auger recombinations at a lower and a higher voltage, respectively. Similar to η_{int}, the external efficiency η_{ext} reaches to a maximum at $V = 1.33$ V and then droops more severely than η_{int} as V further increases, crossing with $\eta_{ext,cr}$ at the critical voltage $V_{cr} = 1.41$ V, where $\eta_{ext} = \eta_{ext,cr}$. We see that the DH EL cooler can act in the cooling mode for $V < 1.41$ V.

It is not difficult to find that the deterioration of η_{ext} is caused by a rapid reduction in the injection efficiency η_{inj} since $\eta_{inj} = \eta_{ext}/\eta_{int}$ for $\eta_{xp} = 1$. This implies that V_{cr} can be improved if the leakage current can be alleviated. In our calculation, it is shown that the strong surface recombination dominates in the leakage current as a result of a high minority carrier density near the terminal contacts. This situation becomes worse in particular for the devices of shorter cladding layers as in this work. The high minority carrier density can be reduced significantly by carrier blocking layers. In Figure 44.5a, we show the electron density in the center of the p-cladding layer as a function of V for $x = 0.25$, 0.3, 0.35, and 0.4. As can be seen, the electron density decreases monotonically with increasing x. This indicates that the energy barrier provided by the blocking layer can effectively block the minority carriers, preventing them from flowing over the carrier blocking layer. This leads to a lower minority carrier density in the cladding layer and hence a lower surface recombination current. The low surface recombination current is confirmed by the injection efficiency as plotted in Figure 44.5b. Evidently, the injection efficiency is improved significantly by increasing the barrier height when $x \leq 0.3$. The improvement however becomes insignificant as x increases from 0.35 to 0.4, in contrast to the result from Figure 44.5a This is because the leakage current across the blocking layer is almost eliminated when $x = 0.35$. In this case, $\eta_{inj} \approx 1$ and the residue of the leakage current is mostly due to the SRH recombination in the space layer. Contrarily, for $x \leq 0.3$, the injection efficiency decreases rapidly at high bias as a result of J_{sr}.

In Figure 44.6, we show the cooling power as a function of the applied voltage for EL cooler of various Al content in the blocking layer. As expected, the critical voltage increases from 1.41 V for $x = 0.25$ to 1.47 V for $x = 0.35$ and 0.4. Moreover, the cooling power also increases from 7.5 to 20 W/cm^2 for $x = 0.25$ to 0.35. The improvement of cooling power comes from an improved η_{inj} and hence η_{ext} as indicated by Equations 44.22 and 44.23, i.e., $P_c = JV\left(\eta_{ext}/\eta_{ext,cr} - 1\right)$. Further increasing the Al content to 0.4 does not help in improving the cooling power, which is consistent with the variation of the injection efficiency as shown in Figure 44.5b.

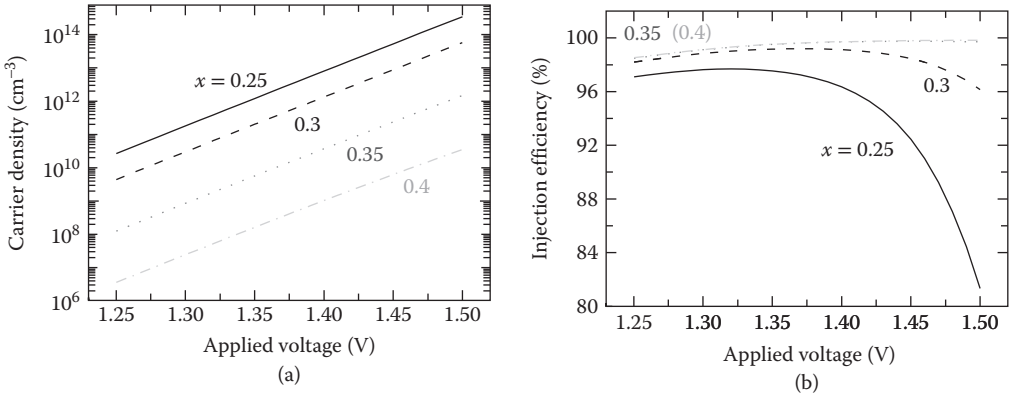


FIGURE 44.5 (a) The minority carrier (electron) density in the center of the p-cladding layer and (b) the injection efficiency as functions of the applied voltage V for various aluminum content $x = 0.25$, 0.3, 0.35, and 0.4.

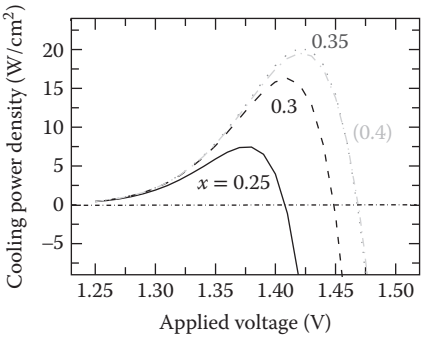


FIGURE 44.6 The ideal cooling power density P_c^0 as a function of the applied voltage V for various Al compositions of $x = 0.25$, 0.3, 0.35, and 0.4 in the carrier blocking layer. The extraction efficiency is set at unity.

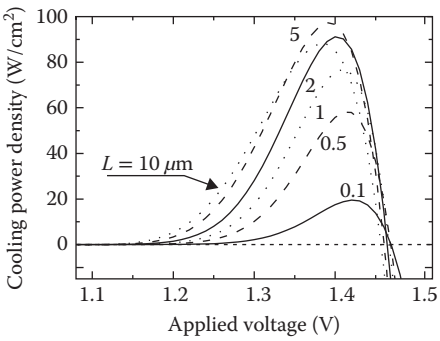


FIGURE 44.7 The ideal cooling power density P_c^0 as a function of V for $L = 0.1$, 0.5, 1, 2, 5, and 10 μm.

44.3.1.2 Active Layer Thickness

The cooling power density of 20 W/cm^2 seems to be the limiting value for the device of active layer thickness $L = 0.1$ μm. To further enhance it, one can increase the active layer thickness. Intuitively, a wider active layer can contain more electron–hole pairs to recombine therein, leading to an enhanced J_{rad} and hence more effective cooling cycles. Figure 44.7 shows the cooling power density P_c as a function of V for the structures with the same Al content ($x = 0.4$) in the carrier blocking layer but various active layer thickness $L = 0.1$, 0.5, 1, 2, 5, and 10 μm; here, $\eta_{xp} = 1$. As can be seen, the maximum cooling power $P_{c,max}$ can be improved about five times up to as high as 97 W/cm^2 by increasing L from 0.1 to 5 μm. However, this improvement is only prominent for $L \leq 5$ μm. Further increasing the active layer thickness does not enhance but degrades the maximum cooling power due to the remarked nonuniformity of carrier distribution, as L is comparable to or greater than the carrier diffusion length. From our calculation, we find that the average carrier density in the active layer is lower for a larger L; this difference in the average carrier density between different Ls is particularly remarkable when V is high. This renders the cooling power for $L = 10$ μm larger than that for $L = 5$ μm at a low V but smaller at a high V.

In Figure 44.8, we show the dependences of various efficiencies on the active layer thickness by plotting the injection efficiency η_{inj} and the internal efficiency η_{int} in Figure 44.8a, the product $\eta_{inj}\eta_{int}$ and the critical external efficiency $\eta_{ext,cr}$ in Figure 44.8b, and the cooling efficiency η_c in Figure 44.8c as functions of the applied voltage V for the active layer thickness $L = 0.1$, 1, and 10 μm and $\eta_{xp} = 1$. The Al content of the blocking layer is $x = 0.4$. Since J_{inj} is enhanced by increasing L, the η_{inj} is also enhanced. For $L \geq 1$ μm, the injection efficiency is nearly 100%, almost independent of the applied voltage. As been described, the internal efficiency η_{int} reaches to a maximum at about 1.37 V. It declines more profoundly as V decreases

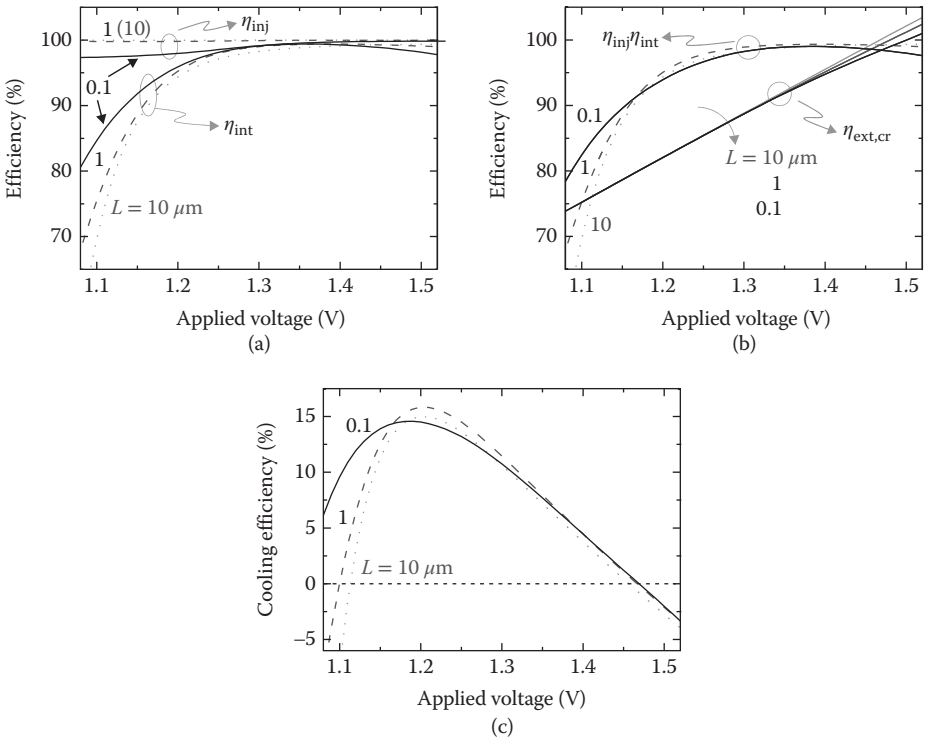


FIGURE 44.8 (a) The injection efficiency η_{inj} and the internal efficiency η_{int}, (b) the production of $\eta_{inj}\eta_{int}$ and the critical extraction efficiency $\eta_{ext,cr}$, and (c) the ideal cooling efficiency η_c^0 as functions of the applied voltage for $L = 0.1$, 1, and 10 μm.

than increases, implying a more important SRH recombination at low V than the Auger recombination at high V. Such a lowering of η_{int} at low V is more prominent for a larger L because of a lower carrier density in the active layer. On the other side, at high V, the Auger recombination is less important for a larger L due also to a lower carrier density. The ratio of $\eta_{ext,cr}/(\eta_{inj}\eta_{int})$ is the critical extraction efficiency $\eta_{xp,cr}$, which gives the design margin for η_{xp}. For instance, at $V = 1.2$ V, $\eta_{inj}\eta_{int} = 94\%$ and $\eta_{ext,cr} = 82\%$, as can be found in Figure 44.8b, give $\eta_{xp,cr} = 87\%$. Evidently, the margin becomes tighter when bias V deviates from 1.2 V. If an EL cooler is improperly biased at a voltage where $\eta_{xp,cr} \geq 1$, the ideal cooling efficiency will become negative (Figure 44.8c). In this case, it is impossible for the device to act in the cooling mode. At $V = 1.39$ V where the maximum cooling power occurs (see Figure 44.7), $\eta_{xp,cr} = 96\%$ for $L = 10$ μm. In such a bias condition, the corresponding cooling efficiency is only about 4%. The cooling efficiency can be improved by biasing the cooler at $V = 1.2$ V at which it can be as high as 15% at the cost of the cooling power density.

44.3.1.3 Doping in Active Layer

In Figure 44.9, we show the influences of the doping types and concentrations on EL refrigeration by plotting the ideal cooling power as a function of the applied voltage V for the active regions which are undoped, doped with $N_d = 1 \times 10^{18}$ cm^{-3}, with $N_a = 1 \times 10^{18}$ cm^{-3}, and with $N_a = 5 \times 10^{18}$ cm^{-3}. Here, the thickness of the active layer is $L = 1$ μm and the Al content of the blocking layers is $x = 0.4$. We find that intentionally doping into the active region does not improve the cooling power. The degradation caused by p-type doping is mainly due to the enhanced Auger recombination. The n-type doping suppresses the radiative recombination and also the Auger recombination, leading to a slightly enhanced cooling power at high V but a reduced cooling power at low V.

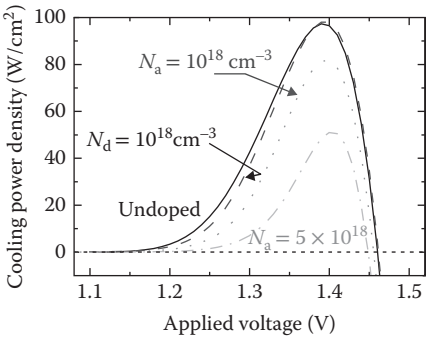


FIGURE 44.9 The ideal cooling power density P_c^0 as a function of the applied voltage for active layers with different doping types and concentrations, including undoped, doped with $N_d = 1 \times 10^{18}$ cm^{-3}, with $N_a = 1 \times 10^{18}$ cm^{-3}, and with $N_a = 5 \times 10^{18}$ cm^{-3}. Here, the active layer thickness is $L = 1$ μm, the Al content of the blocking layer is $x = 0.4$, and the light extraction efficiency is set at unity.

44.3.2 Influence of Optical Processes on EL Refrigeration

44.3.2.1 Influence of Light Extraction

Light extraction efficiency influences significantly the performance of an EL refrigerator in the absence of photon recycling. This is because the photons trapped within the device turn out to be totally absorbed and transformed into thermal energy. In this subsection, we study the influence of the light extraction efficiency on EL refrigeration for the worst case of zero photon recycling efficiency ($\eta_{pr} = 0$). Here, the active layer thickness L is set at 1 μm and the Al content $x = 0.4$ is used for the carrier blocking layer. Figure 44.10a shows the curves of the cooling power P_c as a function of the applied voltage V for five values of extraction efficiency $\eta_{xp} = 0.8$, 0.85, 0.9, 0.95, and 1. The curves exhibit a peak for $\eta_{xp} = 0.9$, 0.95, and 1, but no peak for $\eta_{xp} = 0.8$ and 0.85. As expected, the height of the peak diminishes drastically from 79 to 2 W/cm^2 as η_{xp} reduces slightly from unity to 0.9. The voltage at which the curve peaks moves from 1.41 V for $\eta_{xp} = 1$ to 1.28 V for $\eta_{xp} = 0.9$. The cooling power falls rapidly from the peak value to negative as the voltage increases.

These behaviors can be understood from Equation 44.23, i.e., $\eta_c = \eta_{ext}/\eta_{ext,cr} - 1$, with the help of the relationships between the curves of the corresponding external efficiencies η_{ext} and the critical external efficiencies $\eta_{ext,cr}$ shown in Figure 44.10b. The curves of $\eta_{ext,cr}$ for the different η_{xp} values merge into a single line, since $\eta_{ext,cr} = qV/\langle\hbar\omega\rangle$ is nearly proportional to V and independent of η_{xp}. The curves of η_{ext} for the different η_{xp} values behave similarly and separate almost equidistantly from their neighbors. As V increases, the η_{ext} value increases and then almost reaches saturation at a level $\eta_{ext} \approx \eta_{xp}$, with a slight decline in the high-voltage region. These features can be explained by the simple relation $\eta_{ext} = \eta_{xp}\eta_{int}\eta_{inj}$, where the injection efficiency η_{inj} is almost unity for the device containing carrier blocking layers. Therefore, the variation of η_{ext} with V for various η_{xp} simply follows the variation of the internal efficiency η_{int} multiplied by a scaling factor of η_{xp}. In Figure 44.10b, we see that there exist operation regions of refrigeration for $\eta_{xp} = 0.9$, 0.95, and 1, but no operation region for $\eta_{xp} = 0.8$ and 0.85, where the operation region of refrigeration is a range of voltage in which the device operates in the cooling modes ($\eta_c > 0$ or $\eta_{ext} > \eta_{ext,cr}$). This explains the feature shown in Figure 44.10a that only the three curves for $\eta_{xp} = 0.9$, 0.95, and 1 exhibit a peak of positive P_c. Figure 44.10c shows the cooling efficiency η_c as a function of V for $\eta_{xp} = 0.85$, 0.9, 0.95, and 1. We see that the peaks for η_c-V curves fall at $V = 1.2$ V. The cooling efficiencies decline when the operation voltage deviates from 1.2 V. Different to the peak voltage for η_c-V curves, the peaks of P_c-V curves in Figure 44.10a occur at voltages near the critical voltages, for which the cooling efficiency is small. This is why a slight increase of V may cause the drastic reduction of P_c from the peak value and turn the operation of the device from the cooling mode to the heating mode.

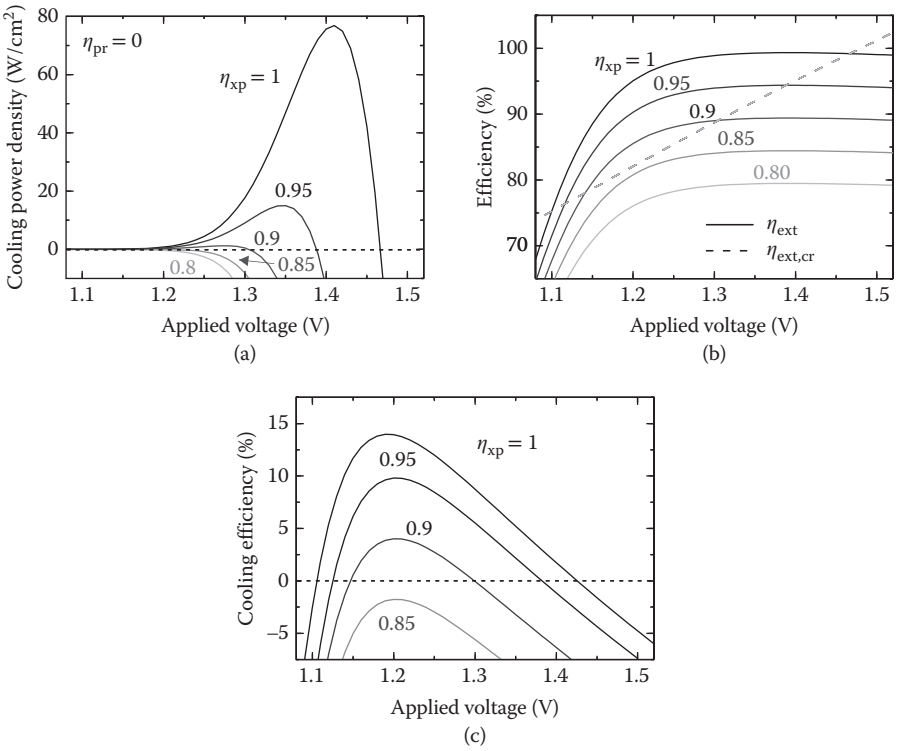


FIGURE 44.10 (a) The cooling power density P_c, (b) the external quantum efficiency η_{ext} and the critical extraction efficiency $\eta_{ext,cr}$, and (c) the cooling efficiency η_c as functions of the applied voltage for various extraction efficiency $\eta_{xp} = 0.8, 0.85, 0.9, 0.95$, and 1. The photon recycling efficiency is $\eta_{pr} = 0$.

44.3.2.2 Influence of Photon Recycling

Actually, most of the trapped photons can be reabsorbed by the high-quality GaAs active region and reincarnated into electron–hole pairs rather than into the thermal energy if the device structure is well designed. The photon recycling, as will be described, alleviates considerably the requirement of extraction efficiency for EL refrigeration. In this subsection, we demonstrate the photon recycling effect by examining another extreme case of unity recycling efficiency $\eta_{pr} = 1$, i.e., all of the trapped photons are recycled, turning into electron–hole pairs in the active region.

Figure 44.11a shows the cooling power density P_c as a function of V for five values of the extraction efficiency $\eta_{xp} = 0.2, 0.4, 0.6, 0.8$, and 1. All the curves in the figure behave as an asymmetrical peak of positive P_c, even for the low extraction efficiency of 0.2. The height of the peak now diminishes gently as η_{xp} reduces. These peaks occur almost at the same voltage ($V = 1.41$ V) except for $\eta_{xp} = 0.2$, for which the curve peaks at 1.39 V. Similarly, these behaviors can also be understood from Equation 44.23 and the relations between the corresponding η_{ext} and $\eta_{ext,cr}$ shown in Figure 44.11b. Comparing with those in Figure 44.10b, the external efficiencies η_{ext} shown in Figure 44.11b still vary with V in the manner that they first rise, reach a maximum, and then slightly decline. The difference is that η_{ext} is now much less sensitive to η_{xp}, especially in the voltage range where the internal efficiency η_{int} is near unity, as plotted in Figure 44.11c together with η_{inj} for various η_{xp}. In the voltage range between 1.3 and 1.5 V, where η_{int} is close to unity, the external efficiency η_{ext} can be approximated by

$$\eta_{ext} \approx \eta_{inj}\left[1 - \frac{1}{\eta_{xp}}\left(1 - \eta_{int}\right)\right]. \tag{44.26}$$

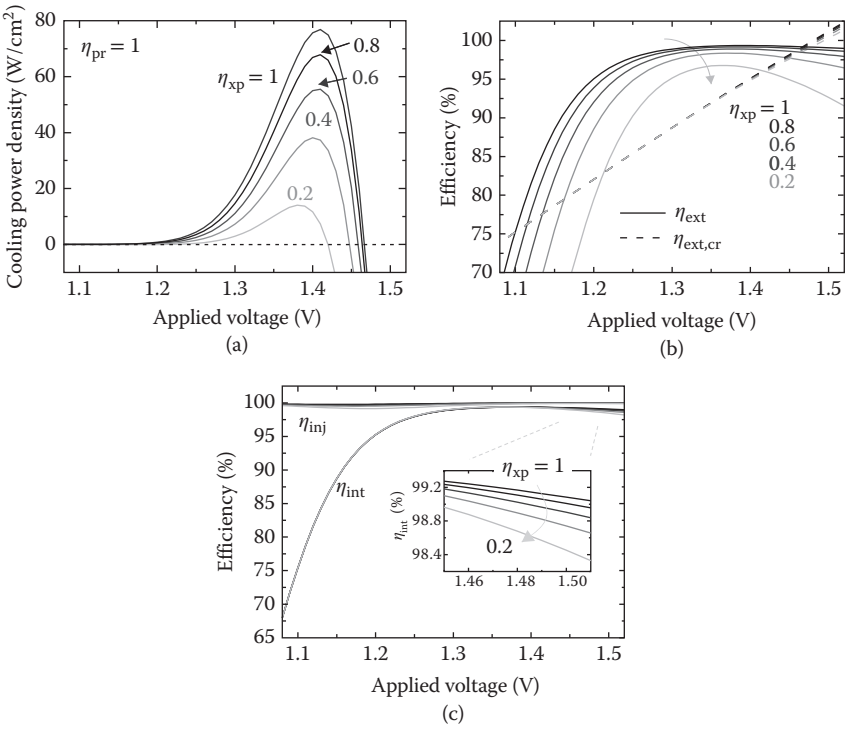


FIGURE 44.11 (a) The cooling power density P_c, (b) the external quantum efficiency η_{ext} and the critical external efficiency $\eta_{ext,cr}$, and (c) the injection efficiency η_{inj} and the internal efficiency η_{int} as functions of V for various extraction efficiency $\eta_{xp} = 0.2$, 0.4, 0.6, 0.8, and 1. The inset of (c) shows η_{int} at high V. The photon recycling efficiency is set at unity.

Equation 44.26 explains that, as $1 - \eta_{int}$ is small, η_{ext} is insensitive to η_{xp} and has a value close to η_{inj}, which is nearly unity.

In principle, the variation of η_{ext} basically follows the variation of η_{int} modified by photon recycling. This is because the only important loss is the nonradiative recombination in the active region when we set $\eta_{pr} = 1$ and $\eta_{inj} \approx 1$. For $\eta_{xp} = 1$, all the generated photons are emitted out of the device and there is no photon recycling, regardless of the value of η_{pr}, as can also be seen from Equation 44.21. The photon recycling effect manifests when η_{xp} deviates from unity. It causes an effective feedback current J_G and boosts the carrier concentration in the active region. The increasing carrier concentration degrades η_{int} and hence η_{ext} in the high-voltage region. This effect is particularly important for a low η_{xp} (see Figure 44.11b and the inset in Figure 44.11c). The lines of $\eta_{ext,cr}$ shown in Figure 44.11b do not merge as completely as those for the case of $\eta_{pr} = 0$ shown in Figure 44.10b. This is a consequence of the photon recycling effect, which causes a higher average photon energy $\langle \hbar\omega \rangle$ and, hence, a lower $\eta_{ext,cr}$ for a lower η_{xp}, due to a higher carrier concentration. The condition that $P_c = 0$ occurs at a critical voltage of about 1.47 V for $\eta_{xp} = 0.4$, 0.6, 0.8, and 1. The voltages are close to each other because of the crowding of the η_{ext}-V curves in the high-voltage region, which cross the $\eta_{ext,cr}$-V curves in the vicinity of 1.47 V. This implies that the P_c-V curves peak almost at the same voltage. The voltages for the peaks of the η_c-V curves, which can be estimated from Figure 44.11b, move from 1.2 to 1.3 V as η_{xp} decreases from 1 to 0.2.

Under the condition that $\eta_{pr} = 1$ and $\eta_{inj} \approx 1$, the critical extraction efficiency $\eta_{xp,cr}$ for the device to operate in the cooling mode can be written as

$$\eta_{xp,cr} \approx \frac{(1/\eta_{int}) - 1}{(1/\eta_{ext,cr}) - 1}. \tag{44.27}$$

The above expression implies that $\eta_{xp,cr}$ can be very small when η_{int} approaches unity, while $\eta_{ext,cr}$ does not. An EL refrigerator having a very low η_{xp} could operate in the cooling mode if there were no optical parasitic loss ($\eta_{pr} = 1$) and no current leakage ($\eta_{inj} = 1$).

The aforementioned discussion reveals that a high internal efficiency is essential to the photon recycling process. This implies that the quality of the materials making up the device is critical for refrigeration. If the material quality is not as high as in the study and gives $\eta_{int} < \eta_{ext,cr}$, it is impossible for the device to operate in the cooling mode as a result of the inequality $\eta_{ext} < \eta_{int}$ (see Equation 44.21), regardless of the values of the other efficiencies.

We now know that the internal efficiency η_{int} is an important factor in determining η_{ext} and, hence, to the performance of the device. It is influenced by photon recycling through the feedback generation of electron–hole pairs in the active region. To further gain insight into this effect, we set $\eta_{pr} = 1$ and investigate the average carrier density in the active region (Figure 44.12) as a function of η_{xp} when the device is biased at $V = 1.2$ and 1.4 V. As shown in Figure 44.12a, the average carrier density is low ($\approx 3 \times 10^{16}$ cm^{-3}) and nearly independent of η_{xp} for the device biased at 1.2 V, but is more than 5×10^{17} cm^{-3} and increases as η_{xp} reduces for the device biased at 1.4 V. This difference arises from the difference in the electric potential energy of the electrons (holes) in the active region relative to that in the *n*-type (*p*-type) cladding layer, from which the electrons (holes) are injected. This situation can be seen in Figure 44.12b, which shows the conduction band profiles at the *n*-doping side as functions of *z* for $V = 1.2$ and 1.4 V. When the device is biased at a low voltage, such as $V = 1.2$ V, the electric potential energy of the electrons is higher in the active layer than in the *n*-type cladding layer. This drop in electric potential energy causes the excess carriers, due to optical generation, to leak out of the active region, leading to an effective current J_G opposite the injection current J_{inj}. Consequently, accumulation of carriers in the active region is negligible, and the carrier density is basically independent of the photon recycling (and, hence, η_{int}). The drop in electric potential energy diminishes as the bias increases and, eventually, the situation reverses. At $V = 1.4$ V, the electric potential energy of the electrons is lower in the active region than in the n-type cladding layer. The potential profile now forms a potential well for the carriers and prohibits most of the excess carriers in the active region from leaking out. As a result, the optically generated carriers are mostly accumulated within the active region, recombined therein, and contribute an additional current, known as J_G. The carrier accumulation depends on the photon recycling and becomes particularly important for low η_{xp}, as shown in Figure 44.12a.

The variation of recombination current densities (J_{SRH}, J_{rad}, and J_{Aug}) on the extraction efficiency also follows the variation of carrier density such that they are invariant on η_{xp} at $V = 1.2$ V but increase considerably at $V = 1.4$ V for a decreasing η_{xp} (Lee and Yen, 2012). The feedback current J_G, which compensates J_{inj} for the total recombination in the active region, increases with decreasing η_{xp} for both

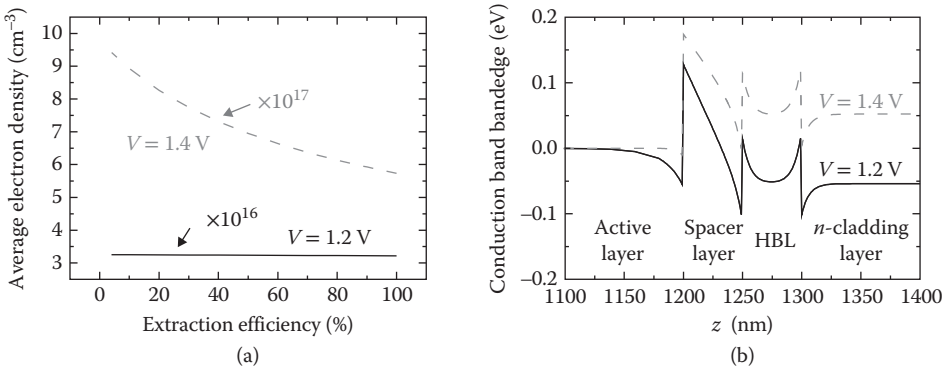


FIGURE 44.12 (a) The average electron densities in the active region for $V = 1.2$ and 1.4 V as functions of the extraction efficiency and (b) the conduction band profiles near the *n*-doping side as functions of *z* for $V = 1.2$ and 1.4 V. (HBL, hole-blocking layer.)

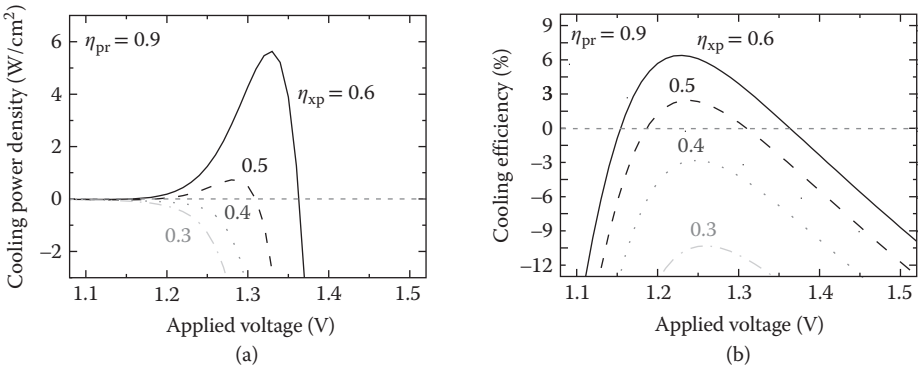


FIGURE 44.13 (a) The cooling power density P_c and (b) the cooling efficiency η_c as functions of applied voltage for various light extraction efficiencies $\eta_{xp} = 0.3$, 0.4, 0.5, and 0.6. The recycling efficiency is set at a practical value of 0.9.

applied voltages because $J_G = \eta_{pr}(1 - \eta_{xp})J_{rad}$. This leads to a significant reduction in J_{inj} and hence improves external efficiency for a low η_{xp}.

For a more practical case, we show the cooling power P_c in Figure 44.13a and the cooling efficiency η_c in Figure 44.13b as functions of the applied voltage V. Since $L = 1$ μm, η_{xp} is at most 60%. We therefore consider $\eta_{xp} = 0.3$, 0.4, 0.5, and 0.6 and set $\eta_{pr} = 0.9$, which is potentially achievable for a high-quality GaAs material. The resulting P_c-V relationship changes with η_{xp} to a degree milder than for $\eta_{pr} = 0$ (Figure 44.10a) but more severe than for $\eta_{pr} = 1$ (Figure 44.11a). The peak cooling power is now 5.6 W/cm^2 at $V = 1.33$ V for $\eta_{xp} = 0.6$, which is only 10% for the case of $\eta_{pr} = 1$ (Figure 44.11a). At the voltage of 1.33 V, the cooling efficiency η_c is small (2%) for $\eta_{xp} = 0.6$ and falls negative as η_{xp} reduces to 0.5, associated with the severe degradation of P_c from 5.6 to −3 W/cm^2. The situation can be cured by biasing the device at a lower voltage for an improved η_c. For example, at $V = 1.24$ V, η_c is 6.4% for $\eta_{xp} = 0.6$ and remains positive (about 2%) for $\eta_{xp} = 0.5$. The device can operate in the cooling mode for $\eta_{xp} > 0.45$. However, the cooling power is sacrificed. It is now only 0.9 W/cm^2 for $\eta_{xp} = 0.6$ and 0.3 W/cm^2 for $\eta_{xp} = 0.5$.

In the end of this chapter, we remark on the choices of active materials and operation temperatures for EL refrigeration. From Equation 44.24, the critical external efficiency of an EL cooler operating at E_g/q for maximum cooling power can be approximated by $\eta_{ext,cr} \approx E_g/\left(E_g + 2k_BT\right)$. Evidently, the external efficiency required for cooling is alleviated with decreasing bandgap energy and an increasing operating temperature. Indeed, recent research has shown that operating a 450-nm InGaN LED at 615 K is capable of producing a negative Peltier heating power inside the device (Xue et al., 2015). Moreover, theoretical study also confirms that elevating the operating temperature can improve the cooling power and efficiency for a GaAs EL cooler (Heikkilä et al., 2010). These improvements, as can be seen from the approximation of $\eta_{ext,cr}$, are strongly related to the ratio $E_g/2k_BT$. A smaller value of $E_g/2k_BT$ implies a fewer number of effective cooling cycles required for compensating a single parasitic loss. This means that the EL refrigeration will be more likely achieved for a cooler with an active material of smaller bandgap energy and operating in a higher temperature if the internal efficiency is high.

44.4 Summary

We discussed the influences of device structures and optical processes on the feasibility and capability of GaAs-based EL refrigerators by self-consistent calculation. For EL devices with short cladding layers, the leakage current, which arises from the minority carriers flowing toward the opposite electrodes, deteriorates the injection efficiency significantly at high bias. The degradation of injection efficiency can be cured

by employing carrier blocking layers of $Al_xGa_{1-x}As$ ($x \geq 0.35$), leading to an injection efficiency of nearly unity. The maximum cooling power density generally increases with the active layer thickness because of an increasing total radiative recombination rate. A GaAs active layer with thickness of 5 μm can give a high internal efficiency of nearly 100% and a limiting cooling power density of 97 W/cm^2. Further increasing the active layer thickness causes a nonuniform distribution of carriers, reducing the radiative power and also the maximum cooling power. Doping in the active layer does not help significantly in improving the cooling power density. Contrarily, a heavily *p*-doped active layer has much lower cooling power density than an undoped one owing to the enhanced Auger recombination.

The cooling capability depends strongly on the external efficiency, and hence the extraction efficiency, due to a small ratio of the extracted thermal energy to the work done on each injected electron–hole pair. For a GaAs EL cooler to act in the cooling mode, an external efficiency of 82% is needed at 1.2 V. This corresponds to an extraction efficiency of 87% if the trapped photons are all absorbed parasitically. This harsh requirement of extraction efficiency can be alleviated by photon recycling process provided that the internal efficiency is sufficiently high. The photon recycling behaves as an additional means of generating carriers, improving the external efficiency by reducing the driving currents. Therefore, for an ideal case of unity recycling efficiency, an EL device can act in the cooling mode even if the extraction efficiency is only 20%. For a more practical case of 90% recycling efficiency, our calculation shows that a maximum cooling power of 5.6 W/cm^2 and a maximum cooling efficiency of 6.4% can be achieved for an EL cooler of 1-μm GaAs active layer. The limiting extraction efficiency for this cooler is found to be 45%, which is achievable for the state-of-the-art technology. This reveals a good possibility of experimentally realizing EL refrigeration for high-cooling-power application.

References

Ashcroft NW, Mermin ND (1976) *Solid State Physics*. Fort Worth, TX: Harcourt Brace College.

Berdahl P (1985) Radiant refrigeration by semiconductor diodes. *J. Appl. Phys.* 58:1369–1374.

Chuang SL (1995) *Physics of Optoelectronic Devices*. New York, NY: Wiley.

Dousmanis GC, Mueller CW, Nelson H, Petzinger KG (1964) Evidence of refrigerating action by means of photon emission in semiconductor diodes. *Phys. Rev.* 133:A316–A318.

Epstein RI, Buchwald MI, Edwards BC, Gosnell TR, Mungan CE (1995) Observation of laser-induced fluorescent cooling of a solid. *Nature* 377:500–503.

Han P, Jin KJ, Ren SR, Zhou YL, Lu HB (2007) Numerical analysis of optothermionic refrigeration in semiconductor triple-well structure. *J. Appl. Phys.* 102:114501.

Heikkilä O, Oksanen J, Tulkki J (2009) Ultimate limit and temperature dependency of light-emitting diode efficiency. *J. Appl. Phys.* 105:093119.

Heikkilä O, Oksanen J, Tulkki J (2010) The challenge of unity wall plug efficiency: The effects of internal heating on the efficiency of light emitting diodes. *J. Appl. Phys.* 107:033105.

Horio K, Yanai H (1990) Numeral modeling of heterojunctions including the thermionic emission mechanism at heterojunction interface. *IEEE Tran. Electron. Devices* 37:1093–1098.

Lee KC, Yen ST (2012) Photon recycling effect on electroluminescent refrigeration. *J. Appl. Phys.* 111:014511.

Lin FR, Lee KC, Yen ST (2012) The ultimate cooling temperature of semiconductor electroluminescent refrigeration. In *the 12th International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD 2012)*, Shanghai, China, ThPD2.

Liu X, Zhao G, Zhang Y, Deppe DG (2013) Semiconductor laser monolithically pumped with a light emitting diode operating in the thermoelectrophotonic regime. *Appl. Phys. Lett.* 102:081116.

Mal'shukov AG, Chao KA (2001) Opto-thermionic refrigeration in semiconductor heterostructures. *Phys. Rev. Lett.* 86:5570–5573.

Melgaard S, Seletskiy D, Polyak V, Asmerom Y, Shiek-Bahae M (2014) Identification of parasitic loss in Yb:YLF and prospects for optical refrigeration down to 80 K. *Opt. Express* 22:7756–7764.

Oksanen J, Tulkki J (2010) Thermophotonic heat pump—A theoretical model and numerical simulations. *J. Appl. Phys.* 107:093106.

Oksanen J, Tulkki J (2015) LEDs feed on waste heat. *Nat. Photonics* 9:782–783.

Piprek J, Li Z-M (2016) Electroluminescent cooling mechanisms in InGaN/GaN light-emitting diodes. *Opt. Quant. Electron*. 48:472.

Santhanam P, Gray DJ Jr., Ram RJ (2012) Thermoelectrically pumped light-emitting diodes operating above unity efficiency. *Phys. Rev. Lett.* 108:097403.

Sheik-Bahae M, Epstein RI (2007) Optical refrigeration. *Nat. Photonics* 4:693–699.

Sheik-Bahae M, Epstein RI (2009) Laser cooling of solids. *Laser Photon. Rev.* 3:67–83.

Sotoodeh M, Khalid AH, Rezazadeh AA (2000) Empirical low-field mobility model for III-V compounds applicable in device simulation codes. *J. Appl. Phys.* 87:2890–2900.

Wang JB, Ding D, Yu SQ, Johnson SR, Zhang YH (2005) Electroluminescence cooling in semiconductors. In 2005 *Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science and Photonic Applications Systems Technologies, QELS'05*, Baltimore, MD, 655–657.

Xue J, Zhao Y, Oh SH, Herrington WF, Speck JS, Denbaars ST, Nakamura S, Ram RJ (2015) Thermally enhanced blue light-emitting diode. *Appl. Phys. Lett*. 107:121109.

Yang K, East JR, Haddad GI (1993) Numerical modeling of abrupt heterojunctions using a thermionic-field emission boundary condition. *Solid State Electron.* 36:321–330

Yen ST, Lee KC (2010) Analysis of heterostructures for electroluminescent refrigeration and light emitting without heat generation. *J. Appl. Phys.* 107:054513.

Yu SQ, Wang JB, Ding D, Johnson SR, Valsileska D, Zhang YH (2007) Impact of electronic density of states on electroluminescence refrigeration. *Solid State Electron.* 51:1387–1390.

Zhang J, Li D, Chen R, Xiong Q (2013) Laser cooling of a semiconductor by 40 Kelvin. *Nature* 493:504–508.

45

Photonic Crystal Laser Diodes

45.1 Photonic Crystals Basics 561
Optical Periodic Structures • Photonic Bandgap • Photonic Crystals in Lasers
45.2 Fully Vectorial Method for Modeling VCSELs with Subwavelength Structures 565
Plane-Wave Admittance Method • Plane-Wave Reflection Transfer
45.3 VCSELs with Photonic Crystals for Light Confinement..... 570
45.4 Subwavelength Gratings in VCSELs..... 574
High-Contrast Subwavelength Grating Theory • Modeling of Subwavelength Grating Reflectivity • Cavities with Subwavelength Grating Mirrors • Manufacturing Sensitivity of High-Contrast Grating Mirrors
45.5 Summary..... 582

Maciej Dems

45.1 Photonic Crystals Basics

45.1.1 Optical Periodic Structures

Photonic crystals are periodic structures designed to affect electromagnetic waves in a similar manner to how solid-state crystals affect electrons. The simplest, one-dimensional form of a photonic crystal is a periodic multilayer film (Bragg mirror). Such films were already being studied 120 years ago, in 1887, by Rayleigh [1]. They continued to be the subject of intensive research, resulting in the development of thin-film optics: the science and technology of fabricating dielectric multilayer mirrors, filters, polarizers, antireflection coatings, and so on. However, it took a whole century before the technology was expanded into two and three dimensions by Yablonovitch [2] and John [3] in 1987. Since then, the term *photonic crystal* has become widely used and the subject is studied all over the world.

The physics of photonic crystals has much in common with solid-state physics, as both concern the solution of wave equations in a periodic medium. Hence, concepts known from solid-state physics can also be used in photonics. In particular, for any regular lattice, it is possible to define a *reciprocal lattice* in the Fourier space, i.e., the space of the wavevectors. Such reciprocal lattices are an important tool for the analysis of light propagation in photonic crystals. Because of the Bloch theorem, any eigenmode in periodic medium can be represented as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_{\mathbf{k}}(\mathbf{r}) \exp(-i\,\mathbf{k}\cdot\mathbf{r}), \tag{45.1}$$

where $\mathbf{k}$ is a wavevector and $\mathbf{E}_{\mathbf{k}}$ is a periodic function with the same periodicity as the refractive index, i.e.,

$$n_{\mathrm{R}}(\mathbf{r} + \mathbf{a}_i) = n_{\mathrm{R}}(\mathbf{r}), \tag{45.2}$$

$$\mathbf{E}_{\mathbf{k}}(\mathbf{r} + \mathbf{a}_i) = \mathbf{E}_{\mathbf{k}}(\mathbf{r}). \tag{45.3}$$

Vectors $\mathbf{a}_i$ are referred to as the base vectors of the periodic lattice and their number is equal to the number of dimensions (in Equations 45.2 and 45.3).

A reciprocal lattice has vectors $\mathbf{b}_j$ defined in the space of wavevectors. They fulfill the condition

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij},$$

where δ_{ij} is Kronecker's delta. For three 3D lattices, $\mathbf{b}_j$ can be expressed directly as

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{a_1 a_2 a_3}, \qquad \mathbf{b}_2 = 2\pi\frac{\mathbf{a}_3 \times \mathbf{a}_1}{a_1 a_2 a_3}, \text{ and} \qquad \mathbf{b}_3 = 2\pi\frac{\mathbf{a}_1 \times \mathbf{a}_2}{a_1 a_2 a_3},$$

and for 2D lattices as

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \hat{\mathbf{n}}}{a_1 a_2}, \qquad \mathbf{b}_2 = 2\pi\frac{\hat{\mathbf{n}} \times \mathbf{a}_1}{a_1 a_2},$$

where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the lattice plane. These vectors define the periodicity of the wavevector $\mathbf{k}$ from Equation 45.1. This periodicity enables the modal analysis of PCs with the vectors $\mathbf{k}$ located within a limited area, called the first Brillouin zone [4].

In the case of two-dimensional photonic crystals, two lattice configurations are particularly common. These are the square lattice and the hexagonal lattice, shown in Figure 45.1a and c. Their reciprocal lattices have the same properties as corresponding physical lattices and are depicted in Figure 45.1b and d. Each of these lattices possesses several high symmetry points, which are important for determination of the photonic band structure. The Γ point is always defined for wavevector $\mathbf{k} = \mathbf{0}$. Others are located at the edges of the first Brillouin zone. For the square lattice, these are X for $\mathbf{k} = \frac{\pi}{a}\hat{\mathbf{x}}$ and M for $\mathbf{k} = \frac{\pi}{a}\hat{\mathbf{x}} + \frac{\pi}{a}\hat{\mathbf{y}}$†. In the case of the hexagonal lattice, the high symmetry points are M ($\mathbf{k} = \frac{\pi}{a}\hat{\mathbf{x}} + \frac{\pi}{a\sqrt{3}}\hat{\mathbf{y}}$) and K ($\mathbf{k} = \frac{4\pi}{3a}\hat{\mathbf{x}}$). These positions are given for the points marked in Figure 45.1, although there are other points of each kind on the other edges and corners. In practice, this changes little, as the eigenmodes may be subjected to any symmetry transformation that does not change the distribution of the refractive index. Hence, the triangles made by the adjacent symmetry points are the *irreducible zone*, into which all the possible wavevectors can

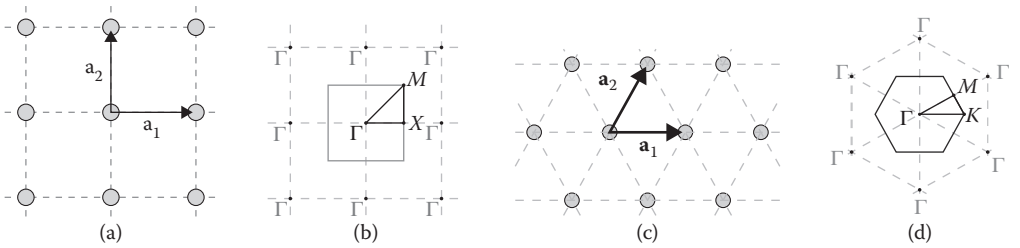


FIGURE 45.1 Arrangements of periodic cells and base vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ of two-dimensional photonic crystals for (a) rectangular and (c) triangular lattices and reciprocal lattices of (b) rectangular and (d) triangular two-dimensional photonic crystals. The important high-symmetry points in the middle and at the edges of the first Brillouin zone (marked with solid lines) are shown. The periodicity of this lattice allows each point outside of the first Brillouin zone to be mapped within it. All the lattice nodes correspond to the Γ point. The triangles formed by the high-symmetry points and shown in the figures are irreducible zones. Each point outside this zone can be mapped to one inside it by periodicity and symmetry transformations.

† $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit tensors of the Cartesian coordinate system.

be transformed using periodicity and symmetry transformations. As a consequence, analysis of the Bloch modes of a photonic crystal can be limited to this irreducible zone. In order to determine the bandgap, it is even sufficient to compute the frequencies for the wavevectors at the edges of that zone.

45.1.2 Photonic Bandgap

One of the most important properties of photonic crystals is the *photonic bandgap* (PBG), i.e., a range of frequencies in which light cannot propagate through the crystal [5]. If a light wave of such frequency were to propagate into the photonic crystal, it would be almost 100% reflected. This effect is well known in one-dimensional periodic stacks and is commonly used in distributed Bragg reflectors (DBRs, see Chapter 34). Figure 45.2a shows the reflectivity spectrum of a typical quarter-wavelength GaAs/AlAs DBR. The clearly visible high-reflectivity range between 927 nm and 1039 nm is the bandgap. The same effect in two dimensions can be seen in Figure 45.2b, which shows the optical bands as a horizontal wavevector to normalized frequency relation and the density of optical states within an infinite crystal. The bandgap is the region with zero density of states. In this frequency range, there is no band for any wavevector.

Two- and three-dimensional photonic crystals can be considered a multi-dimensional equivalent of DBRs. Hence, one of their major applications is to confine light. This is what they are most commonly used for in vertical-cavity surface-emitting lasers (VCSELs).

In order to confine light using a photonic crystal, one must create a *defect*: a localized perturbation of the periodic structure. Such perturbations can be either linear, in which case we have a waveguide, or isolated, in which case it makes a cavity. Examples of two-dimensional photonic crystal waveguides are shown in Figure 45.3. Because of the reflective properties of the photonic crystal within the bandgap, the waveguides can be bended with sharp angles and still effectively transmit light. In lasers, such waveguides can be coupled with edge-emitting lasers. In the case of surface-emitting lasers, the light is confined within a two-dimensional photonic crystal cavity. The most common configurations of such cavities are shown in Figure 45.4.

45.1.3 Photonic Crystals in Lasers

Photonic crystals are seldom used in edge-emitting lasers. One design of such a laser is presented in [6]. It is a W1 or W3 linear waveguide filled with an active material. Such lasers are reported to be intrinsically

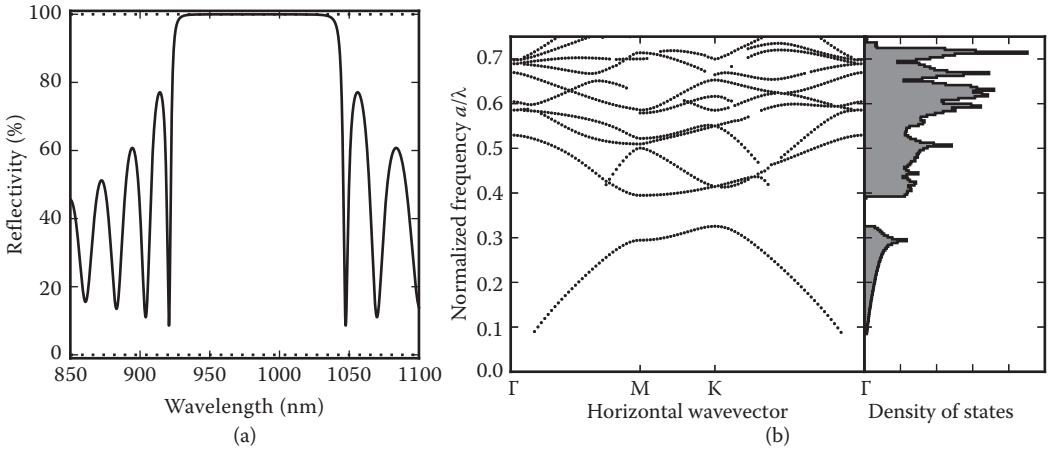


FIGURE 45.2 (a) Reflectivity of a quarter-wavelength one-dimensional stack and (b) band diagram and density of states (right histogram) of a two-dimensional photonic crystal. A photonic bandgap is visible in both cases as the frequency/wavelength region with very high reflectivity or vanishing density of states.

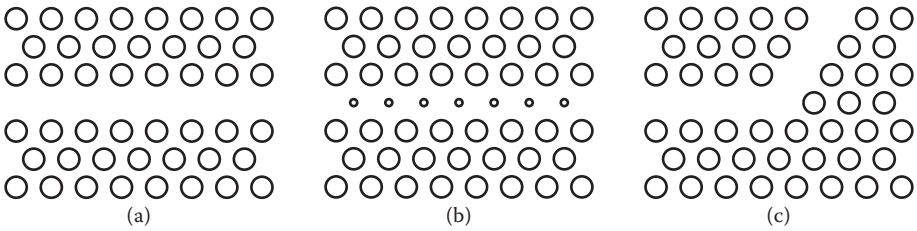


FIGURE 45.3 Examples of photonic crystal waveguides: (a) simple W1 waveguide with one row of missing holes, (b) linear waveguide with smaller holes, and (c) bent waveguide.

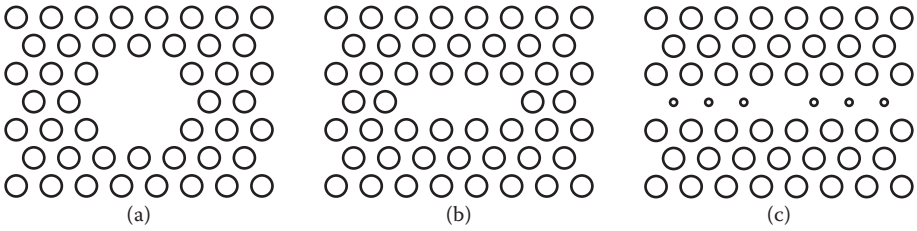


FIGURE 45.4 Examples of photonic crystal cavities: (a) cavity formed by removing seven holes, (b) L3 cavity with shifted edge holes, and (c) birefringent cavity formed by one missing hole and a row of smaller holes.

single-mode, in contrast to classical first-order DFB lasers. Depending on the configuration, the emitting mode has its wavevector either at the Γ or *K* point of the hexagonal photonic crystal lattice. The emission spectrum has two peaks, corresponding to the edges of the so-called mini-stop-band, which appears when two modes cross (as is the case at both Γ and *K* points). If emission takes place at frequencies that do not fall within the bandgap of the photonic crystal, only one of these modes is confined to the waveguide. The other spreads infinitely into the photonic crystal, and therefore suffers significantly larger losses. This is the main cause of single-mode operation.

In contrast to edge-emitting lasers, photonic crystals are much more commonly used in surface-emitting lasers. The first experimental designs were published at the end of the nineties. Similarly to photonic-crystal edge-emitting lasers (PC-EELs), such lasers were at first pumped with an optical beam and their design was not very complicated. The simplest surface-emitting photonic crystal laser is a PC membrane with a defect inside. The greatest advantage of such a design is its simplicity and the ease of fabricating a photonic crystal with a small lattice constant. For a typical hexagonal lattice of air holes, etched in a semiconductor material, the pitch must be around $0.3\,\lambda$ if the frequency of the lasing mode c/λ is to be inside the bandgap. Because of the Purcell effect, the small size of the cavity helps to achieve a high spontaneous emission rate and a high Q-factor that can reach up to 10^4 [7].

In most designs of photonic crystal surface-emitting lasers [8–12], the light is not confined by the bandgap, but through the effective index of the defect area, which is higher than that of the surrounding photonic crystal. The role of the photonic crystal is to provide single-mode emission on either a higher order [9] or a fundamental mode [12]. Single-mode emission is one of the most promising applications of photonic crystals in the area of VCSEL design. This is especially true for high-power lasers [13], where classical designs fail because of the necessity of using broad area cavities.

The origin of a single-mode emission in PC-VCSELs can be attributed to one of the two factors. The first is a significant increase of modal losses for the higher-order modes, caused by light scattering. By carefully choosing the etching depth of the holes in the top DBR and their diameter, one can maximize this difference while keeping optical losses for the fundamental mode reasonably low [14,15]. The other factor is a cutoff similar to that found in so-called *endlessly single-mode* fibers [16], i.e., fibers that remain

single-mode for any particular wavelength. The origin of this phenomena lies in the dependence of the PC effective index on the wavelength. The number of modes in a classical fiber depends on the parameter V_{eff}, which is defined as

$$V_{\mathrm{eff}} = \frac{2\pi r}{\lambda}\left(n_{\mathrm{core}}^2 - n_{\mathrm{clad}}^2\right)^{1/2}, \tag{45.4}$$

where r is the radius of the core, λ is the wavelength, and n_{core} and n_{clad} are the effective indexes of the fiber core and the cladding, respectively. The fiber is single-mode if $V_{\mathrm{eff}} < 2.405$. It is possible to design a photonic crystal fiber in which the V_{eff} does not exceed 2.405 even for $\lambda \to 0$. Because the VCSEL cavity can be approximated as a waveguide (although many phenomena are disregarded in such an approximation), Equation 45.4 can be used in this case. In practice, this means that single-mode operation is achieved if the diameters and depths of the holes in the photonic crystal are sufficiently small.

45.2 Fully Vectorial Method for Modeling VCSELs with Subwavelength Structures

Unlike classical oxide-confined VCSELs, discussed in Chapter 34, VCSELs incorporating subwavelength structures cannot be modeled using either the effective frequency method or any other approximate scalar method. The reason for this is the fact that such methods do not adequately consider light scattering from the subwavelength structures. Hence, in order to correctly model the optical properties of such devices, one needs to apply more exact vector methods, such as the finite-difference time-domain method [17], plane-wave expansion [18], the finite element method (FEM) [19], the coupled mode model (CMM) [20], the plane-wave admittance method (PWRT) [21], or plane-wave reflection transfer [22]. The greatest drawback of all these models is the increased computation time and computer memory usage. This cost is, however, necessary to bear, as only the rigorous solution of Maxwell equations provided by these methods allows light scattering from subwavelength structures to be included in the model.

In this chapter, we illustrate the fundamental optical properties of surface-emitting lasers incorporating subwavelength periodic structures, called photonic crystals. First, we need to sketch the principles of the rigorous vector methods used to obtain the results presented later in the chapter. Instead of describing the details of all popular vector models (some of which are discussed elsewhere in this book), we focus on the two that have proved to be the most useful for our purposes: the plane-wave admittance method and PWRT.

45.2.1 Plane-Wave Admittance Method

The plane-wave admittance method is an extension of the vector approach presented in Chapter 34. It uses the same procedure—admittance transfer—to find the eigenmodes of a resonant cavity. However, in this method, the electromagnetic field is expanded in the plane-wave Fourier basis. This basis is particularly useful for periodic structures, as it has natural periodic boundary conditions. Hence, modeling of infinite two-dimensional photonic crystals is straightforward and requires little effort. However, to analyze localized modes in lasers and correctly consider emitted radiation, one needs to introduce absorbing perfectly matched layers at the edges of the computational domain.

As the exact computation procedure was described in Chapter 34, it need not be repeated here. Instead, this section presents a derivation of the matrix form of Maxwell equations using the plane-wave expansion approach. Such matrices can be entered straightforwardly into computer memory for effective computations.

To start the derivation, consider an anisotropic linear dielectric or semiconductor material without any free charges ($\rho = 0$) or currents ($\mathbf{j} = 0$). In such a material, the time-independent Maxwell equations read

$$\nabla \times \mathbf{E} = -i\omega\mu\mu_0\mathbf{H}, \tag{45.5}$$

$$\nabla \times \mathbf{H} = \quad i\omega\varepsilon\varepsilon_0\mathbf{E}, \tag{45.6}$$

where $i^2 = -1$, ω is the angular frequency of light, and physical observables are the real parts of $\mathbf{E}$ and $\mathbf{H}$.

Now consider an anisotropic material in Cartesian coordinates. Assuming that both permittivity and permeability of the material can be represented as diagonal tensors, the Equations 45.5 and 45.6 can be expanded as

$$-\partial_z E_y + \partial_y E_z = -i\omega\mu_x\mu_0 H_x, \tag{45.7}$$

$$\partial_x E_z - \partial_z E_x = -i\omega\mu_y\mu_0 H_y, \tag{45.8}$$

$$-\partial_y E_x + \partial_x E_y = -i\omega\mu_z\mu_0 H_z, \tag{45.9}$$

$$-\partial_z H_y + \partial_y H_z = i\omega\varepsilon_x\varepsilon_0 E_x, \tag{45.10}$$

$$\partial_x H_z - \partial_z H_x = i\omega\varepsilon_y\varepsilon_0 E_y, \tag{45.11}$$

$$-\partial_y H_x + \partial_x H_y = i\omega\varepsilon_z\varepsilon_0 E_z, \tag{45.12}$$

where ∂_ξ means a partial derivative in the ξ direction. These equations can be transformed into the form already introduced in Chapter 34:

$$\partial_z \begin{bmatrix} -E_y \\ E_x \end{bmatrix} = -i\frac{\eta_0}{k_0} \begin{bmatrix} \partial_y\varepsilon_z^{-1}\partial_y + \mu_x k_0^2 & -\partial_y\varepsilon_z^{-1}\partial_x \\ -\partial_x\varepsilon_z^{-1}\partial_y & \partial_x\varepsilon_z^{-1}\partial_x + \mu_y k_0^2 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix}, \tag{45.13}$$

$$\partial_z \begin{bmatrix} H_x \\ H_y \end{bmatrix} = -i\frac{1}{\eta_0 k_0} \begin{bmatrix} \partial_x\mu_z^{-1}\partial_x + \varepsilon_y k_0^2 & \partial_x\mu_z^{-1}\partial_y \\ \partial_y\mu_z^{-1}\partial_x & \partial_y\mu_z^{-1}\partial_y + \varepsilon_x k_0^2 \end{bmatrix} \begin{bmatrix} -E_y \\ E_x \end{bmatrix}, \tag{45.14}$$

where $k_0 = \omega/c = \omega\left(\mu_0\varepsilon_0\right)^{1/2}$.

Now, consider an electromagnetic field in a structure exhibiting two-dimensional periodicity, i.e.,

$$\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} + \mathbf{a}_i) \qquad (i = 1, 2), \tag{45.15}$$

$$\mu(\mathbf{r}) = \mu(\mathbf{r} + \mathbf{a}_i) \qquad (i = 1, 2), \tag{45.16}$$

with $\mathbf{a}_1$ and $\mathbf{a}_2$ being elementary lattice vectors in the *xy* plane. Because of their periodicity, the solutions of the Maxwell equations obey the Bloch theorem, i.e., both the electric and magnetic fields can be represented as

$$\mathbf{E}(\mathbf{r}) = \bar{\mathbf{E}}(\mathbf{r})\exp(-i\,\mathbf{k}_{xy}\cdot\mathbf{r}), \tag{45.17}$$

$$\mathbf{H}(\mathbf{r}) = \bar{\mathbf{H}}(\mathbf{r})\exp(-i\,\mathbf{k}_{xy}\cdot\mathbf{r}), \tag{45.18}$$

where $\bar{\mathbf{E}}(\mathbf{r})$, $\bar{\mathbf{H}}(\mathbf{r})$ are periodic functions and $\mathbf{k}_{xy}$ is a projection of the mode wavevector into the *xy* plane. Furthermore, the functions $\bar{\mathbf{E}}(\mathbf{r})$ and $\bar{\mathbf{H}}(\mathbf{r})$ can be expanded into the two-dimensional

Fourier (plane-wave) series as

$$\bar{\mathbf{E}}(\mathbf{r}) = \mathbf{E}^{\mathbf{g}} \exp(i\,\mathbf{g}\cdot\mathbf{r}) = \mathbf{E}^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle, \tag{45.19}$$

$$\bar{\mathbf{H}}(\mathbf{r}) = \mathbf{H}^{\mathbf{g}} \exp(i\,\mathbf{g}\cdot\mathbf{r}) = \mathbf{H}^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle, \tag{45.20}$$

where the Einstein summation convention[†] is used and $\mathbf{g}$ is the reciprocal lattice vector:

$$\mathbf{g} = l_1\mathbf{b}_1 + l_2\mathbf{b}_2, \tag{45.21}$$

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}, \tag{45.22}$$

with l_1 and l_2 being arbitrary integers and δ_{ij} the Kronecker's delta. As we perform Fourier expansion only in the xy plane, the coefficients $\mathbf{E}^{\mathbf{g}}$ and $\mathbf{H}^{\mathbf{g}}$ are functions of z.

Using the Bloch theorem together with Equations 45.19 and 45.20, we can represent both fields as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{\mathbf{g}} \left|\varphi_{\mathbf{g}-\mathbf{k}_{xy}}\right\rangle, \tag{45.23}$$

$$\mathbf{H}(\mathbf{r}) = \mathbf{H}^{\mathbf{g}} \left|\varphi_{\mathbf{g}-\mathbf{k}_{xy}}\right\rangle. \tag{45.24}$$

Now consider Equations 45.13 and 45.14. As the Fourier basis is orthonormal, i.e.,

$$\langle\varphi_{\mathbf{g}}|\varphi_{\mathbf{g}'}\rangle = \delta_{\mathbf{g}\mathbf{g}'}, \tag{45.25}$$

we can introduce Equations 45.23 and 45.24 into Equations 45.13 and 45.14 and left-multiply them by $\langle\varphi_{\mathbf{g}-\mathbf{k}_{xy}}|$ to obtain

$$\partial_z \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix} = -i\frac{\eta_0}{k_0} \left\langle \varphi_{\mathbf{g}-\mathbf{k}_{xy}} \right| \begin{matrix} \partial_y\varepsilon_z^{-1}\partial_y + \mu_x k_0^2 & -\partial_y\varepsilon_z^{-1}\partial_x \\ -\partial_x\varepsilon_z^{-1}\partial_y & \partial_x\varepsilon_z^{-1}\partial_x + \mu_y k_0^2 \end{matrix} \left| \varphi_{\mathbf{g}'-\mathbf{k}_{xy}} \right\rangle \begin{bmatrix} H_x^{\mathbf{g}'} \\ H_y^{\mathbf{g}'} \end{bmatrix}, \tag{45.26}$$

$$\partial_z \begin{bmatrix} H_x^{\mathbf{g}'} \\ H_y^{\mathbf{g}'} \end{bmatrix} = -\frac{i}{\eta_0 k_0} \left\langle \varphi_{\mathbf{g}-\mathbf{k}_{xy}} \right| \begin{matrix} \partial_x\mu_z^{-1}\partial_x + \varepsilon_y k_0^2 & \partial_x\mu_z^{-1}\partial_y \\ \partial_y\mu_z^{-1}\partial_x & \partial_y\mu_z^{-1}\partial_y + \varepsilon_x k_0^2 \end{matrix} \left| \varphi_{\mathbf{g}'-\mathbf{k}_{xy}} \right\rangle \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix}, \tag{45.27}$$

where, in order to simplify notation, we represent $E_y(\mathbf{r})$ as

$$E_y(\mathbf{r}) = -E_y^{\mathbf{g}} \left|\varphi_{\mathbf{g}-\mathbf{k}_{xy}}\right\rangle \tag{45.28}$$

instead of directly using Equation 45.23.

The matrices in the above equations can be computed easily when both permittivity and permeability are expanded in the Fourier basis. For this purpose, the analyzed layer is assumed to be invariant in the

[†] By this convention, if an index is repeated twice in any product, we perform summation over this index. We use this convention in this whole chapter, unless explicitly stated otherwise.

z-direction. Then, we can write

$$\varepsilon_i(\mathbf{r}) = \varepsilon_i^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle \quad (i = x, y), \tag{45.29}$$

$$\mu_i(\mathbf{r}) = \mu_i^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle \quad (i = x, y), \tag{45.30}$$

$$\varepsilon_z^{-1}(\mathbf{r}) = \kappa^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle, \tag{45.31}$$

$$\mu_z^{-1}(\mathbf{r}) = \gamma^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle. \tag{45.32}$$

Substituting this into Equations 45.26 and 45.27, we finally obtain the matrix form of the equations in the plane-wave basis:

$$\partial_z \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix} = -i\frac{\eta_0}{k_0} \begin{bmatrix} -(g_y - k_y)(g'_y - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_x^{\mathbf{g}-\mathbf{g}'} & (g_y - k_y)(g'_x - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} \\ (g_x - k_x)(g'_y - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} & -(g_x - k_x)(g'_x - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_y^{\mathbf{g}-\mathbf{g}'} \end{bmatrix} \begin{bmatrix} H_x^{\mathbf{g}'} \\ H_y^{\mathbf{g}'} \end{bmatrix}, \tag{45.33}$$

$$\partial_z \begin{bmatrix} H_x^{\mathbf{g}} \\ H_y^{\mathbf{g}} \end{bmatrix} = -\frac{i}{\eta_0 k_0} \begin{bmatrix} -(g_x - k_x)(g'_x - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_y^{\mathbf{g}-\mathbf{g}'} & -(g'_x - k_x)(g'_y - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} \\ -(g_y - k_y)(g'_x - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} & -(g_y - k_y)(g'_y - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_x^{\mathbf{g}-\mathbf{g}'} \end{bmatrix} \begin{bmatrix} E_y^{\mathbf{g}'} \\ E_x^{\mathbf{g}'} \end{bmatrix}, \tag{45.34}$$

where k_x, k_y, g_x, and g_y are the corresponding components of the mode wavevector and the reciprocal lattice vector $\mathbf{g}$, respectively. These equations can be introduced into computer memory for further computations by truncating the Fourier basis $\{\mathbf{g}\}$ at some point. The introduced error will depend on two factors—the distribution of the electromagnetic field and the distribution of the material parameters. The former, as a result of simulation, cannot be influenced directly, although the latter can be convoluted with a Gaussian window to improve convergence.

The vertical components of electric and magnetic fields can also be expanded in the Fourier basis as

$$E_z^{\mathbf{g}} = \frac{\eta_0}{k_0}\kappa^{\mathbf{g}-\mathbf{g}'}\left[-(g'_y - k_y)H_x^{\mathbf{g}'} + (g'_x - k_x)H_y^{\mathbf{g}'}\right], \tag{45.35}$$

$$H_z^{\mathbf{g}} = \frac{1}{k_0\eta_0}\gamma^{\mathbf{g}-\mathbf{g}'}\left[(g'_x - k_x)E_y^{\mathbf{g}'} + (g'_y - k_y)E_x^{\mathbf{g}'}\right]. \tag{45.36}$$

Once Equations 45.33 and 45.34 have been derived and represented in their numerical forms, we can use them to determine the eigenmodes of the analyzed structure. For this purpose, we first show their solution for a single z-invariant layer and then present a numerically stable technique for combining fields in all layers and deriving an eigenmode condition for the whole structure. The procedure described below was originally used with the method of lines [23] and is also applied to the plane-wave admittance method.

Further in this section, we will use the short representation of the vectors and matrices introduced in Equations 45.33 and 45.34. In particular, we will name

$$\hat{\mathbf{E}} = \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix},$$

$$\hat{\mathbf{H}} = \begin{bmatrix} H_x^{\mathbf{g}} \\ H_y^{\mathbf{g}} \end{bmatrix},$$

$$\mathbf{R_E} = \frac{1}{\eta_0 k_0}\begin{bmatrix} -(g_x - k_x)(g'_x - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_y^{\mathbf{g}-\mathbf{g}'} & -(g'_x - k_x)(g'_y - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} \\ -(g_y - k_y)(g'_x - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} & -(g_y - k_y)(g'_y - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_x^{\mathbf{g}-\mathbf{g}'} \end{bmatrix},$$

$$\mathbf{R_H} = \frac{\eta_0}{k_0}\begin{bmatrix} -(g_y - k_y)(g'_y - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_x^{\mathbf{g}-\mathbf{g}'} & (g_y - k_y)(g'_x - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} \\ (g_x - k_x)(g'_y - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} & -(g_x - k_x)(g'_x - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_y^{\mathbf{g}-\mathbf{g}'} \end{bmatrix}.$$

It is worth noting that the whole procedure described below can be applied with any correct representation of the vectors and matrices above. Thus, it is applicable not only to plane-wave expansion, but also to any orthogonal or non-orthogonal basis. In particular, the representation of $\hat{\mathbf{E}}$, $\hat{\mathbf{H}}$, $\mathbf{R_E}$, and $\mathbf{R_H}$ by finite-difference approximation is used in the method of lines.

Take Equations 45.33 and 45.34 and represent them in their short form:

$$\partial_z\hat{\mathbf{E}} = -i\mathbf{R_H}\hat{\mathbf{H}}, \tag{45.37}$$

$$\partial_z\hat{\mathbf{H}} = -i\mathbf{R_E}\hat{\mathbf{E}}. \tag{45.38}$$

By taking the z-derivative of one of these equations and substituting the other into it, one can write two second-order equations with the fields decoupled:

$$\partial_z^2\hat{\mathbf{E}} = -\mathbf{Q_E}\hat{\mathbf{E}}, \tag{45.39}$$

$$\partial_z^2\hat{\mathbf{H}} = -\mathbf{Q_H}\hat{\mathbf{H}}, \tag{45.40}$$

where $\mathbf{Q_E} = \mathbf{R_H}\mathbf{R_E}$ and $\mathbf{Q_H} = \mathbf{R_E}\mathbf{R_H}$.

The eigenmodes can be found using the procedure described in Chapter 34 by constructing and solving equation $\mathbf{Y}\bar{\mathbf{E}} = 0$.

45.2.2 Plane-Wave Reflection Transfer

PWRT differs from the plane-wave admittance method in the way the diagonalized Equations 45.33 and 45.34 are solved. In this method, the electric field in each layer uniform in z-direction is chosen to have the form

$$\tilde{\mathbf{E}} = \exp(-i\boldsymbol{\Gamma}z)\tilde{\mathbf{F}} + \exp(i\boldsymbol{\Gamma}z)\tilde{\mathbf{B}}, \tag{45.41}$$

where $\tilde{\mathbf{F}}$ denotes the amplitude of the forward propagating wave and $\tilde{\mathbf{B}}$ the amplitude of the backward propagating wave. The reflection transfer means an iterative determination of the reflection matrix $\mathbf{R}$, which is defined in each layer as a relation between $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{B}}$ such that

$$\tilde{\mathbf{F}} = \mathbf{R}\,\tilde{\mathbf{B}}. \tag{45.42}$$

We assume that in the first layer, which has infinite thickness, there is no in-going field (we will introduce any optional incident field in the last layer). Thus, $\tilde{\mathbf{F}} \equiv 0$ and so $\mathbf{R} = \mathbf{0}$. The reflection matrices in the subsequent layers can be determined in an iterative process using the requirement that the electric and magnetic fields are continuous at layer interfaces.

By careful selection of units in the magnetic field, it is possible to show that in the diagonalized domain it takes the form of

$$\tilde{\mathbf{H}} = \exp(-i\boldsymbol{\Gamma}z)\tilde{\mathbf{F}} - \exp(i\boldsymbol{\Gamma}z)\tilde{\mathbf{B}}, \tag{45.43}$$

which can be derived from Equation 45.41. The requirement of field continuity between the n-th and $(n+1)$th layers can be expressed as

$$\mathbf{T_E}^n \left\{ \boldsymbol{\varphi}^n \tilde{\mathbf{F}}^n + \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n \right\} = \mathbf{T_E}^{n+1} \left\{ \tilde{\mathbf{F}}^{n+1} + \tilde{\mathbf{B}}^{n+1} \right\}, \tag{45.44}$$

$$\mathbf{T_H}^n \left\{ \boldsymbol{\varphi}^n \tilde{\mathbf{F}}^n - \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n \right\} = \mathbf{T_H}^{n+1} \left\{ \tilde{\mathbf{F}}^{n+1} - \tilde{\mathbf{B}}^{n+1} \right\}, \tag{45.45}$$

where $\boldsymbol{\varphi}^n = \exp\left(-i\boldsymbol{\Gamma} d_n\right)$ and d_n is the thickness of nth layer. Using the relation (Equation 45.42), we obtain

$$\tilde{\mathbf{F}}^{n+1} + \tilde{\mathbf{B}}^{n+1} = \mathbb{E}^n \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n, \tag{45.46}$$

$$\tilde{\mathbf{F}}^{n+1} - \tilde{\mathbf{B}}^{n+1} = \mathbb{H}^n \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n, \tag{45.47}$$

where

$$\mathbb{E}^n = \left(\mathbf{T_E}^{n+1}\right)^{-1} \mathbf{T_E}^n \left\{ \boldsymbol{\varphi}^n \mathbf{R}^n \boldsymbol{\varphi}^n + \mathbf{I} \right\}, \tag{45.48}$$

$$\mathbb{H}^n = \left(\mathbf{T_H}^{n+1}\right)^{-1} \mathbf{T_H}^n \left\{ \boldsymbol{\varphi}^n \mathbf{R}^n \boldsymbol{\varphi}^n - \mathbf{I} \right\}. \tag{45.49}$$

This, after simple algebra, gives the reflection matrix for the $(n+1)$th layer in the form

$$\mathbf{R}^{n+1} = \left[\mathbb{E}^n + \mathbb{H}^n\right] \left[\mathbb{E}^n - \mathbb{H}^n\right]^{-1}. \tag{45.50}$$

In contrast to plane-wave admittance transfer, we need not assume the electric field at the boundaries is equal to zero and, hence, there is no need to use absorbing perfectly matched layers as vertical boundary conditions. Instead, at the edge of the structure, we can assume the $\mathbf{R}^0$ matrix to be equal to 0. This means that in the outermost layer, there is no forward propagating field. This field corresponds to the wave coming into the structure. As the analyzed device is a laser resonant cavity, it only emits light; hence, $\tilde{\mathbf{F}}^0 = \mathbf{0}$ and $\tilde{\mathbf{B}}^0 \neq \mathbf{0}$. Then, similarly to PWAM, we can perform iterative progressions of the reflection matrix from both top and bottom boundaries, up to a selected layer. In this, the following relations hold:

$$\tilde{\mathbf{F}}^{\text{top}} = \tilde{\mathbf{B}}^{\text{bottom}}, \tag{45.51}$$

$$\tilde{\mathbf{B}}^{\text{top}} = \tilde{\mathbf{F}}^{\text{bottom}}. \tag{45.52}$$

Using this condition, it is possible to construct a matching matrix $\mathbf{M}$ such that

$$\mathbf{M}\tilde{\mathbf{F}}^{\text{bottom}} = \left[\mathbf{R}^{\text{top}} - \left(\mathbf{R}^{\text{bottom}}\right)^{-1}\right] \tilde{\mathbf{F}}^{\text{bottom}} = \mathbf{0}.$$

In other words, this matrix must be singular, which is the condition for finding the eigenmodes.

45.3 VCSELs with Photonic Crystals for Light Confinement

In VCSELs, photonic crystals are used to provide effective light confinement. By etching a regular lattice of holes in either the DBRs or the VCSEL cavity (or both), together with intentional lattice defects, it is possible to mold the optical mode distribution. In particular, one can laterally confine the fundamental laser mode within a photonic crystal defect without adding any more lowindex oxidation layers. There are several advantages of this approach: First, it allows the fabrication of VCSELs with well-defined cavities in

material systems where selective oxidation is either impossible or difficult to achieve. Second, nonuniform patterns in the periodic lattice can be tuned to interact differently with fundamental and high-order modes, providing increased mode selectivity and allowing single-mode VCSEL operation at higher currents (see Section 45.1.3).

To illustrate the exclusive optical properties of PC-VCSELs, consider a basic optical-only VCSEL structure made of GaAs, as shown in Figure 45.5a. It contains 24 GaAs/AlAs pairs of top DBRs, 29 pairs of GaAs/AlAs bottom DBRs, and a very short single-wavelength GaAs cavity. The exact layer structure is listed in Table 45.1. The photonic crystal is based on a triangular lattice (of pitch L) and takes the form of three hexagons surrounding a small cavity in the form of a missing PC hole. Holes with diameter d are etched either in the top DBRs (with varying etching depths), throughout the whole structure, or only inside the cavity.

We now investigate the impact of photonic crystal parameters on the properties of the fundamental mode in the VCSEL in question. Due to its Gaussian-like shape, this mode is the most preferred in most applications. Furthermore, VCSELs with incorporated photonic crystals are known for their high mode selectivity [24], which makes it possible to tune them into a single-mode source with high power emission.

Our results are depicted in Figures 45.6 through 45.9. They express typical tendencies observed in PC-VCSELs. Figure 45.6 shows that increasing the size of the holes results in a decrease in the resonant wavelength, which can be explained by two factors. The first is the decrease in the area of the photonic crystal defect, which forms the cavity (corresponding to a decrease in the aperture of a classical VCSEL).

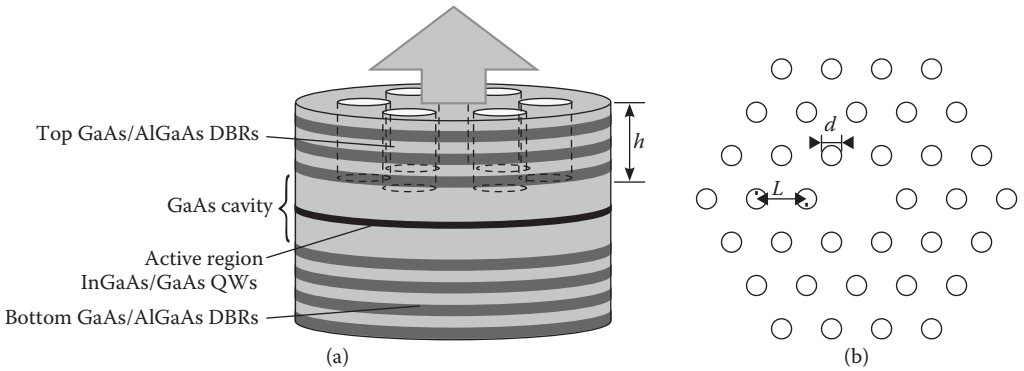


FIGURE 45.5 Analyzed VCSEL structure: (a) schematic three-dimensional view, (b) PC structure. (DBRs, distributed Bragg reflectors; QWs, quantum wells.)

TABLE 45.1 Details of the Layer Structure of the Analyzed VCSEL.

	Thickness [nm]	Material	Refractive index
		Air	1.00
Top DBR	69.40	GaAs	3.53
24 pairs	79.55	AlGaAs	3.08
	121.71	GaAs	3.53
Cavity	$3 \times 8.00 + 2 \times 5.00$	QW/GaAs	$\begin{cases} 3.56 + j\,n_g & \text{for } r < a \\ 3.56 - 0.01j & \text{for } r \geq a \end{cases} \Big/ 3.53$
	121.71	GaAs	3.53
Bottom DBR	79.55	AlGaAs	3.08
29 pairs	69.40	GaAs	3.53
Substrate		GaAs	3.53

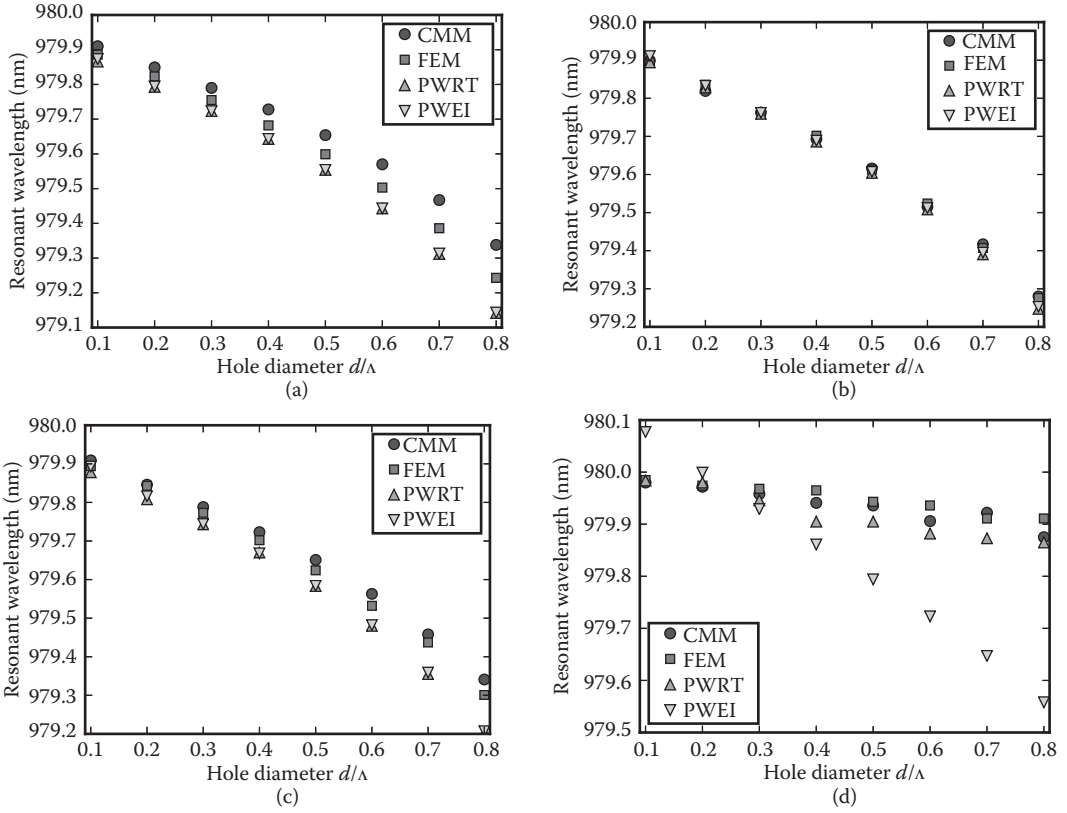


FIGURE 45.6 Resonant wavelengths of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of hole diameters d. The photonic crystal pitch Λ is 4 μm in all the plots.

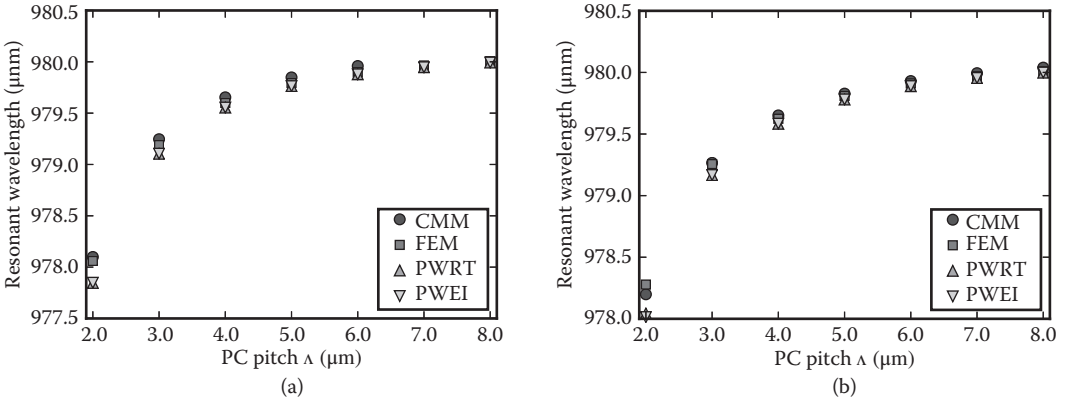


FIGURE 45.7 Resonant wavelengths of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of the photonic crystal pitch Λ. The diameter of the etched holes d is proportional to the pitch and equals 0.5 Λ.

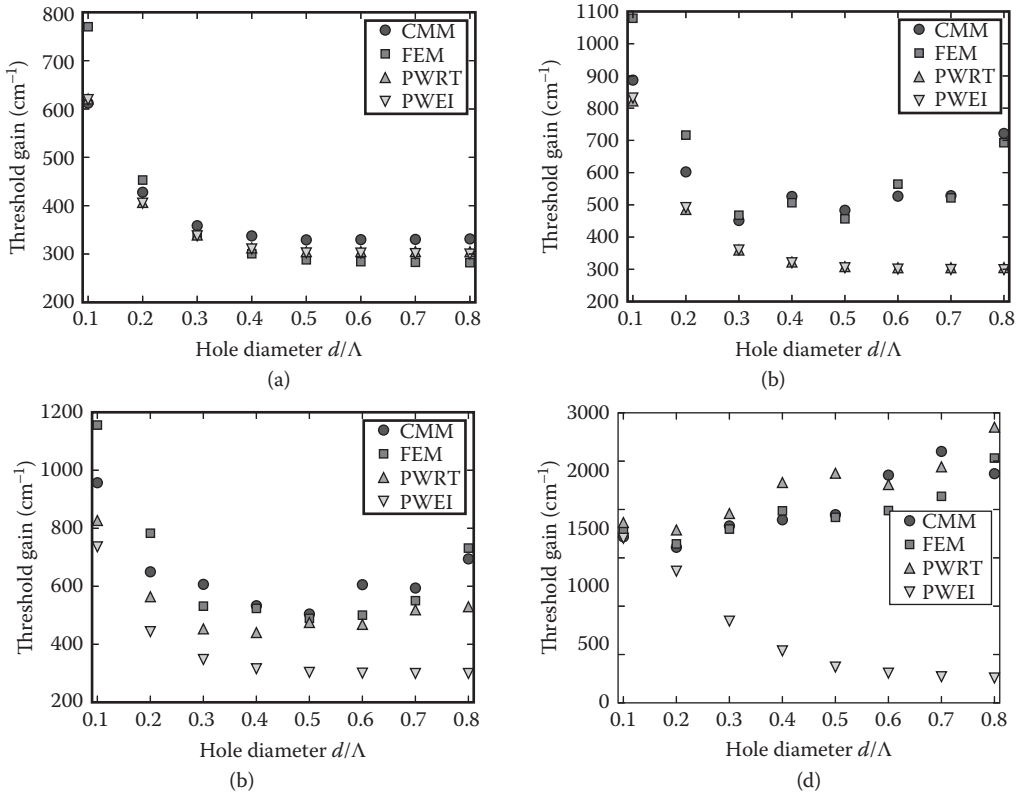


FIGURE 45.8 Threshold gains of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of the hole diameters d. The photonic crystal pitch Λ is 4 μm in all the plots.

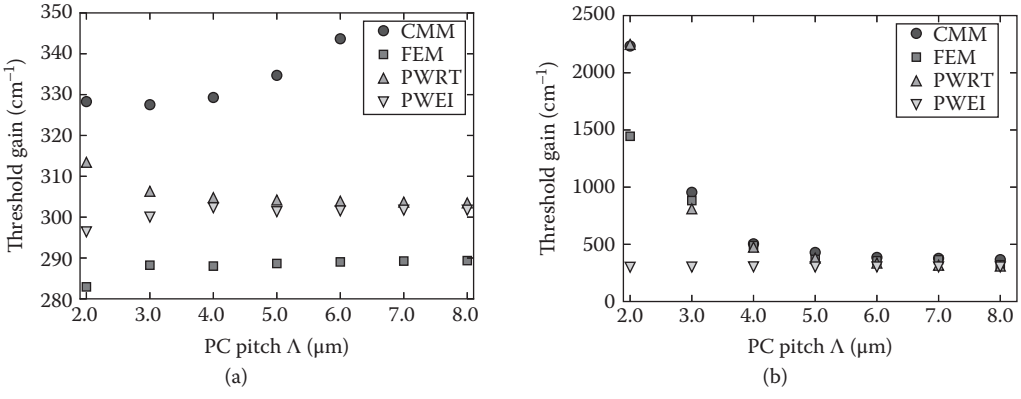


FIGURE 45.9 Threshold gains of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of the photonic crystal pitch Λ. The diameter of the etched holes d is proportional to the pitch and equals 0.5 Λ.

The other factor is the increase of the effective refractive index contrast between the cavity and the photonic crystal surrounding it. Larger holes mean a larger average air-to-semiconductor ratio, which results in a lower effective index outside the cavity.

The opposite tendency is shown in Figure 45.7. Increasing the photonic crystal pitch results in a rapid increase in the resonant wavelength, as the cavity area increases in proportion to Λ. Due to the constant d/Λ ratio, the effective index contrast remains unchanged. The change is fastest for smaller pitches, where the light confinement is high. With increasing Λ, it increases asymptotically to 980 nm, which is the resonant wavelength of the presented VCSEL with no lateral confinement at all.

Much more interesting is the behavior of the threshold gain. In each case, it is large for small holes (Figure 45.8) because of the weak confinement and large overlap of the mode with the absorbing areas of the active region outside the gain aperture. As the size of the holes increases, the mode is increasingly better confined and therefore the threshold decreases. However, with shallow etching (Figure 45.8b through d), scattering starts to play an important role, and with larger holes the threshold starts increasing again. On the other hand, the photonic crystal pitch does not influence the threshold gain significantly (Figure 45.9), except when it drops too strongly (below 4 μm) in structures with holes etched only in the DBR (Figure 45.9b), in which case light scattering starts to play an important role.

In each case, the same optical properties were computed using four different numerical methods: PWRT presented above, the CMM [20], the FEM [25], and the plane-wave effective index (PWEI) [26]. These methods have already been rigorously compared [27]. Here, we show them to emphasize that the three-dimensional structure of a PC-VCSEL is complex to analyze numerically, even nowadays. Practically all of the methods have shown behavior qualitatively different to the others in at least one case. This shows how important it is to calibrate numerical models and to choose the right one, which will yield correct results for a particular case.

In most circumstances, however, numerical analysis allows some conclusions to be made regarding the optical properties of PC-VCSELs. The results depicted in Figures 45.6 and 45.8 show computed resonant wavelengths and threshold gains as functions of the hole diameters when the photonic crystal pitch Λ equals 4 μm. A pitch of 4 μm is a typical value and has been kept fixed for all etchings. Figures 45.7 and 45.9 show the results obtained from different sizes of the photonic crystal lattice constant, with holes etched through the whole cavity and all layers of the top DBR.

In most cases, the PC-VCSEL resonant wavelength is predicted consistently by all numerical methods. The only exception is with large, shallow etched holes, as can be seen in Figure 45.6d. The differences between the vector methods (CMM, FEM, and PWRT) and the scalar method (PWEI), which gives a too small wavelength, are clearly noticeable. The origin of these discrepancies is the significant light scattering at the bottoms of the holes, which is correctly considered by the vector methods, but neglected by the scalar method.

Unsurprisingly, greater differences are observable in the threshold gain data. The methods show good agreement only in case of the $\Lambda = 4$ μm wholly etched structure. All other situations require more detailed analysis. Figure 45.8c shows the threshold gains for holes etched only in the top DBR. Due to their complicated arrangement, significant scattering occurs at their edges, which is reflected in an increase in the threshold gain for $d/\Lambda > 0.4$. This effect is not predicted by PWEI, which—as a variation of the effective index method—is incapable of properly considering scattering losses. PWEI even more significantly underestimates the threshold gain for shallow holes (Figure 45.6d), where scattering is much stronger due to the weaker mode confinement.

45.4 Subwavelength Gratings in VCSELs

In the first part of this chapter, we described the two-dimensional photonic crystals used to provide light confinement of the optical mode. However, periodic subwavelength structures can offer much more. In what follows, we will look at specific one-dimensional photonic crystals: subwavelength gratings with high

refractive index contrast (referred to as high-contrast gratings or HCGs). Such structures may effectively replace DBR mirrors in VCSELs, as they can provide very high reflectivity, while being significantly thinner than a multilayer DBR stack.

Because the light emitted by most VCSELs is linearly polarized, one-dimensional linear gratings can be efficiently applied. Furthermore, they ensure stable polarization of the emitted light, which otherwise might suddenly change at higher output powers [28]. For the same reason, circular gratings are much less suitable as mirror replacements for VCSELs, as they cannot provide uniform reflectivity for linearly polarized light.

In order to investigate the properties of VCSELs with subwavelength gratings, we first need to understand the behavior of grating mirrors. For this reason, in the next subsection, we will outline the basic physical principles behind the high reflectivity of HCGs. Next, we will study their reflectivity in various configurations. After that, we will move on to the determination of resonant modes in VCSEL cavities with subwavelength grating.

45.4.1 High-Contrast Subwavelength Grating Theory

Despite their very slight thicknesses—usually only a few hundreds of nanometers—HCGs can be effective mirrors with reflectivity comparable or exceeding that of thick multilayered DBRs. This behavior can be explained in two ways: either as an effect of the Fano resonance between the incident wave and the excited mode propagating inside the grating [29] or as an effect of destructive interference between transmitted and leaky modes [30]. Below, we briefly summarize the latter explanation.

Consider a grating suspended in air. It can be seen as a periodic array of waveguides, propagating light in a direction perpendicular to the grating plane (Figure 45.10). Assume an incident plane wave propagating in the z-direction and polarized parallel to the grating bars (with the nonzero E_y component). At the incidence side of the grating (region 1 in Figure 45.10), the electric field can be described as the single zero-order incident plane wave and multiple orders of the reflected wave:

$$E_1(x,z) = E_I \exp\left(-ik_0 z\right) + \sum_n E_R^n \cos\left(2n\pi x/L\right) \exp\left(+i\sqrt{k_0^2 - (2n\pi/L)^2}\, z\right), \tag{45.53}$$

where E_I is the amplitude of the incident field, E_R^n is the horizontal distribution of the reflected modes of different orders, $k_0 = \omega/c$ is the normalized frequency, and L is the grating period. Inside the grating, the field is the sum of forward and backward propagating waveguide array modes:

$$E_2(x,z) = \sum_m \left[E_F^m(x) \exp\left(-i\beta_m z\right) + E_B^m(x) \exp\left(+i\beta_m z\right)\right], \tag{45.54}$$

where β_m is the propagation constant of the mth mode and E_F^m and E_B^m are amplitudes of the forward and backward propagating modes, respectively. Finally, in the last section, the field consists only of the

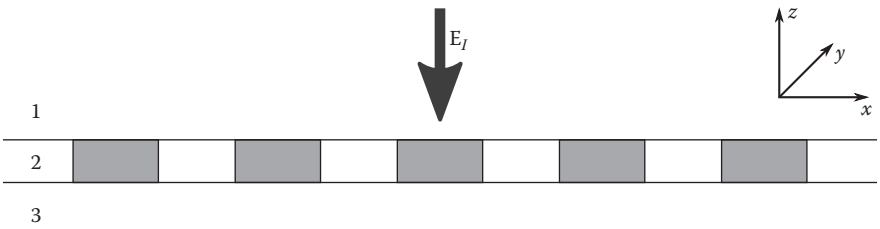


FIGURE 45.10 Grating as a periodic waveguide array.

transmitted modes with amplitudes E_T^n:

$$E_3(x,z) = \sum_n E_T^n \cos(2n\pi x/L) \exp\left(-i\sqrt{k_0^2 - (2n\pi/L)^2}\, z\right).$$

For subwavelength gratings $L < \lambda = 2\pi/k_0$, only the zero orders of both the reflected and transmitted modes propagate and the other modes are evanescent. Although they do not transport any energy, they are very important for ensuring proper boundary conditions for Maxwell equations. By applying these boundary conditions to both grating boundaries, it is possible to determine the amplitudes of both the reflected and transmitted modes (E_R^n and E_T^n) and the modes inside the grating (E_F^m and E_B^m). As only one transmitted mode (E_T^0) propagates energy, it is possible to tune the grating parameters in such a way that $E_T^0 = 0$. This corresponds to the zero energy transmission equivalent, as there is no gain inside the grating, up to 100% reflectivity.

For gratings operating in a near-subwavelength regime, the periodic waveguide array supports only two even modes with propagation constants β_0 and β_2 [31]. In such a case, E_T^0 can be derived analytically as

$$E_T^0 = \frac{1}{L}\int_0^L \left\{E_F^0(x) + E_B^0(x)\right\} dx + \frac{1}{L}\int_0^L \left\{E_F^2(x) + E_B^2(x)\right\} dx.$$

In other words, E_T^0 is a sum of the spatial averages of both waveguide array modes. If these averages are antimatched, i.e.,

$$\int_0^L \left\{E_F^0(x) + E_B^0(x)\right\} dx = -\int_0^L \left\{E_F^2(x) + E_B^2(x)\right\} dx,$$

there is no energy transmission through the grating. This corresponds to the destructive interference of these modes at the far edge of the grating.

For a fuller derivation and discussion of the modal properties of subwavelength gratings, we refer the reader to [31], where the interaction between gratings and light waves is described in detail.

45.4.2 Modeling of Subwavelength Grating Reflectivity

The reflectivity of subwavelength HCGs can be studied using the PWRT method presented at the beginning of this chapter. This method is mathematically equivalent to the popular rigorous coupled-wave analysis (RCWA) [32]. Both methods rely on expansion of the electromagnetic field into a complex Fourier series and both require the analyzed structure to be split into a set of layers, each of which is uniform in the surface normal direction (see Figure 45.11). PWRT, which was used to obtain all the results presented in this chapter, is applied in the same way as for cavity analysis, as described in Section 45.2.2 with the last step omitted. Instead, after determining the reflection matrix $\bar{\mathbf{R}}$, this is multiplied by a predefined incidence vector $\mathbf{F}$ to calculate the vector of reflected diffraction orders $\mathbf{B}$:

$$\mathbf{B} = \bar{\mathbf{R}}\,\mathbf{F}. \tag{45.55}$$

In most cases, the incident vector has only one nonzero component, which corresponds to the incident plane wave of the zero-order. It is possible to construct more complex incidence vectors, e.g., one forming a Gaussian beam. However, for subwavelength gratings the reflection of the zero-order plane wave is dominant and the results for a Gaussian beam do not differ significantly from those for a plane wave.

Figure 45.11 shows a typical single-period cell of a subwavelength grating. It is comprised of parallel bars made of a high-refractive-index material (usually Si or GaAs), surrounded on both sides by low-index

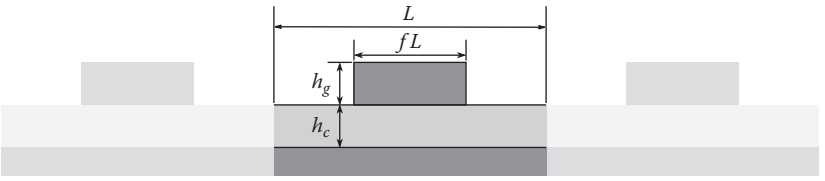


FIGURE 45.11 Schematic diagram of the computational domain for grating reflection computation. The structure is separated into a number of layers uniform in the perpendicular direction and only one grating period is analyzed.

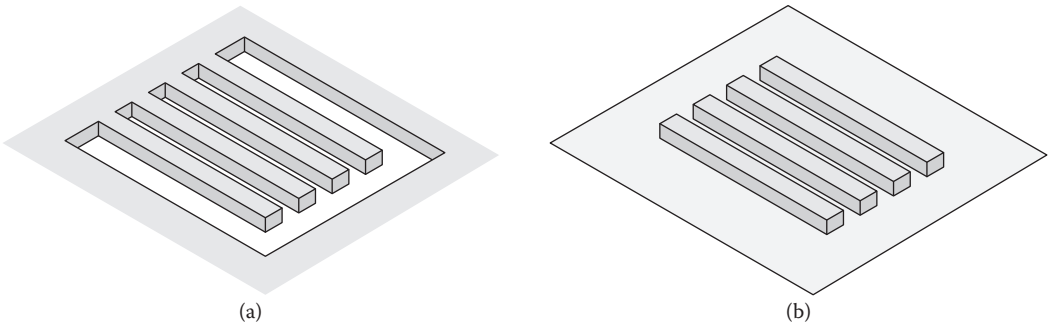


FIGURE 45.12 (a) High-contrast grating membrane suspended in air and (b) similar grating located on low-index cladding (e.g., SiO_2). The former structure is more difficult to fabricate; however, it allows the membrane to be shifted using piezoelectric forces for VCSEL cavity tuning.

layers. These layers can be either the air on both sides (in which case we speak of grating membranes; see Figure 45.12a) or one of the layers may be some kind of low-index solid (usually an oxide, such as amorphous SiO_2). The former are significantly more difficult to fabricate, as the process involves etching a sacrificial layer first and then dissolving it. The resulting grating is suspended in the air. A common problem is then buckling of the stripes, so some additional stress-relaxing etching may be necessary. The advantage of this solution is the fact that the membrane can be placed on a piezoelectric element and then moved freely, allowing for cavity tuning. On the other hand, the latter (Figure 45.12b) is usually much easier to fabricate and the bars are permanently attached to the cladding. The lack of tunability is compensated by the durable mount and no problems with buckling. If the refractive index of the cladding is low enough, there are no significant differences in the optical properties of either type of grating.

Although the gratings are very thin compared to multilayered DBRs, they can provide very high reflectivity, which in an ideal situation can even reach 100%. Due to their strongly anisotropic geometry, the gratings' reflective properties are qualitatively different, depending on the incident light polarization. For an ideal uniform grating, we can distinguish two distinct polarizations: one with an electrical field vector parallel to the grating bars (transverse electric, TE) and the other perpendicular to it (transverse magnetic, TM). Depending on the grating design, any of these polarizations can be strongly reflected, while the other is reflected only moderately. This effect can be used to design VCSELs with strong anisotropy and fixed polarization of the emitted light.

Sample reflectivities computed for a Si membrane suspended in air are shown in Figure 45.13a and b. These figures show reflectivity for both light polarizations with two different gratings. Their exact dimensions are summarized in Table 45.2. Both version were designed to provide high reflectivity at a wavelength of 1500 nm. As can be seen in Figure 45.13, both TE and TM polarization can be strongly reflected, with almost zero transmission. If two such peaks coincide, we observe a wide reflectivity bandwidth. This is clearly visible in Figure 45.13a, where the wide reflectivity area has a small dip in the middle, indicating that it consists of two independent peaks.

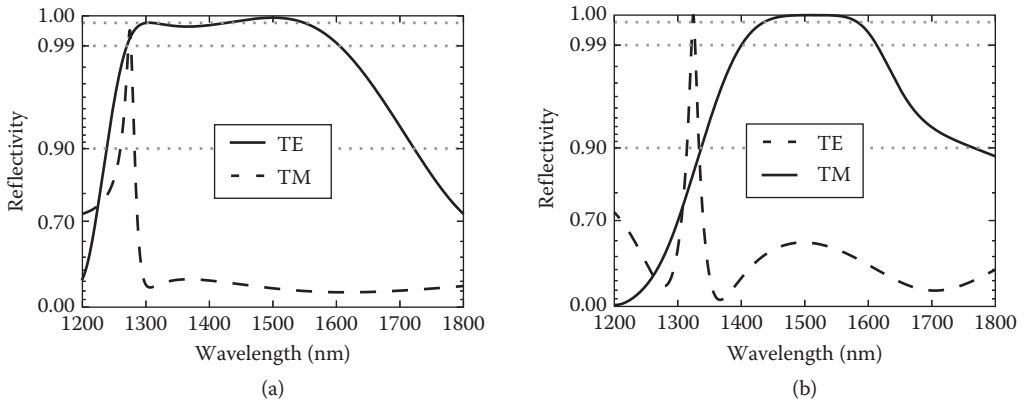


FIGURE 45.13 Silicon-on-silica grating reflectivity for TE and TM polarization. The dimensions of the gratings are optimized to provide high reflectivity around 1500 nm for (a) TE and (b) TM polarization. The exact grating dimensions are summarized in Table 45.2. The dotted lines show reflectivity levels important for laser operation: 90%, 99%, and 99.8%. Notice the non-linear scale in the plots.

TABLE 45.2 Dimensions for Gratings with Reflectivities Shown in Figure 45.13a and b.

	(a)	(b)
Grating period L [μm]	1.00	0.64
Grating fill-factor f	0.30	0.64
Grating thickness h_g [μm]	0.22	0.42
Air-gap/cladding thickness h_c [μm]	0.83	0.83

Note: The meaning of the parameters is explained in Figure 45.11. The refractive index of grating bars and substrate is 3.48 (silicon) and of the cladding is 1.47 (silica).

Due to the finite thickness of the grating, it can rather be considered as an array of cavities for the two waveguide array modes. The negative interference described in Section 45.4.1 corresponds to the resonance thickness of these cavities. If we plot grating reflectivity as a function of wavelength and grating thickness (Figure 45.14a and b), we can notice two overlapping periodic patterns, corresponding to the resonance conditions of both waveguide array modes (with propagation constants β_0 and β_2), giving a characteristic chessboard pattern.

Similar calculations can be performed for an HCG located on cladding (Figure 45.12b). With carefully chosen dimensions, it can reflect both TE and TM polarizations equally well. A sample reflectivity spectrum and reflectivity map for TM light polarization is shown in Figure 45.15. The grating parameters are as follows: $L = 0.70\,\mu\text{m}$, $h_g = 0.46\,\mu\text{m}$, $f = 0.75$, and $h_c = 0.83\,\mu\text{m}$. The cladding material is silica with a refractive index of 1.47. Such gratings are usually easier to manufacture and show the same patterns as air membranes.

45.4.3 Cavities with Subwavelength Grating Mirrors

The grating mirrors presented in Section 45.4.2 can be used to replace multi-layered Bragg mirrors as reflectors in VCSELs. Typical structures for such lasers—both with an HCG membrane suspended in air and an HCG grating located on a low-index cladding—are presented in Figure 45.16. For manufacturing feasibility, the bottom mirrors are classical DBRs and the top ones are replaced with HCGs. In the case of the

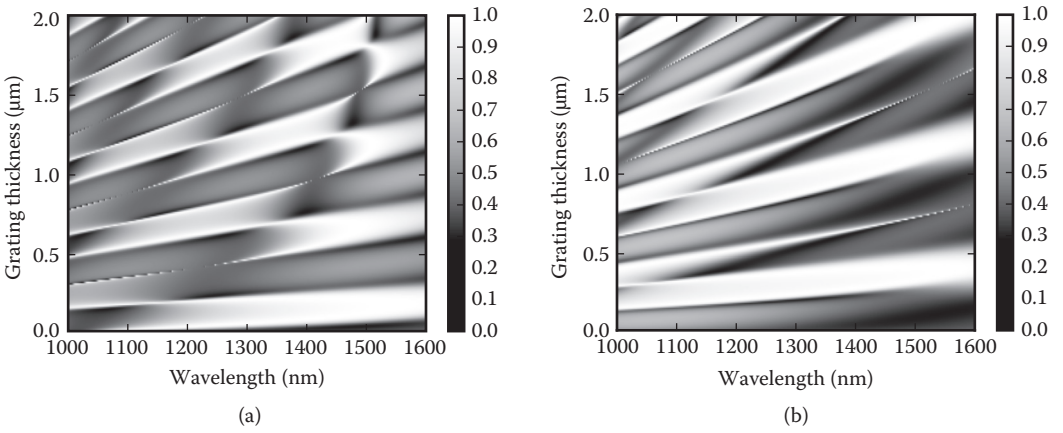


FIGURE 45.14 Reflectivity maps for gratings optimized for (a) TE and (b) TM polarizations as a function of the wavelength and grating thickness.

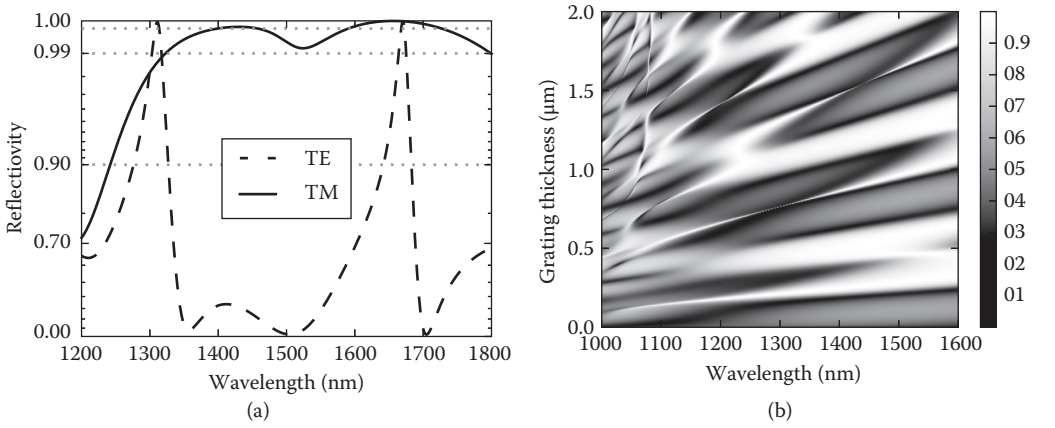


FIGURE 45.15 (a) Reflectivity spectrum and (b) TM reflectivity map for various thicknesses of a grating located on dielectric cladding (Figure 45.12b).

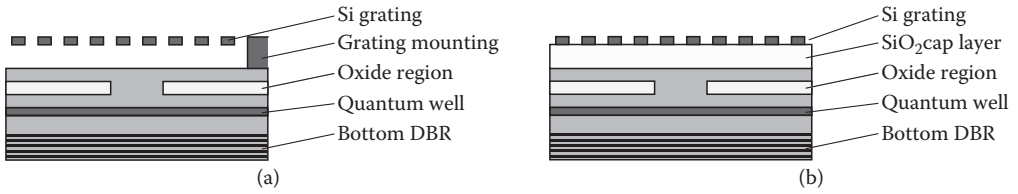


FIGURE 45.16 Typical structure of an HCG-VCSEL with (a) grating membrane suspended in air and (b) grating layer located on a low-index cladding.

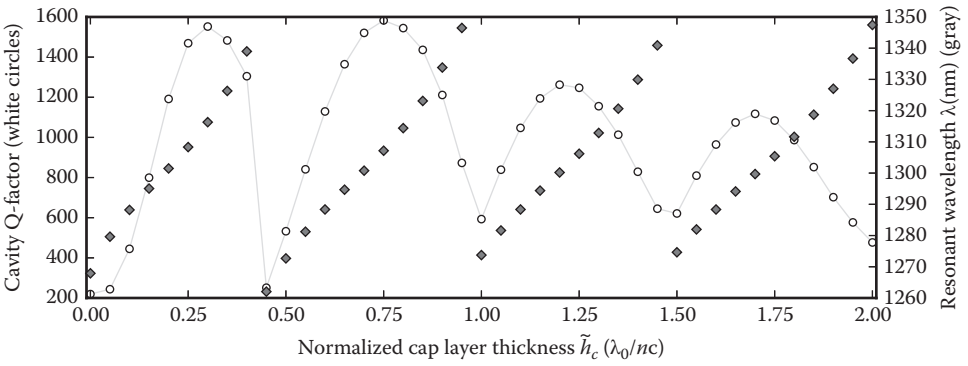


FIGURE 45.17 Cavity Q-factor (light circles) and resonant wavelength (gray diamonds) as functions of the normalized cap layer thickness.

design shown in Figure 45.16a, the membrane can be shifted by piezoelectric forces, effectively changing the cavity length and enabling the emitted wavelength to be tuned. Such a tunable HCG VCSEL has been demonstrated by [33].

In this section, we will analyze an HCG-VCSEL with the grating located on a low-index cladding layer (Figure 45.16b). It consists of a 2-λ GaAs/AlGaAs cavity† with a quantum well in one anti-node and an oxide aperture in the second. The bottom mirror is constructed from 29 pairs of GaAs/AlGaAs quarter-wavelength DBRs. At the top, there is an Si HCG separated by a SiO_2 cap layer that provides low index on both sides of the HCG. The cavity and the bottom DBR are designed for a wavelength of 1.3 μm and the HCG parameters are as follows (see Figure 45.11 for the meaning of the symbols): $L = 0.567\,\mu m$, $h_g = 0.373\,\mu m$, $f = 0.65$, and h_c varies. We are considering the cold-cavity properties of the laser, thus there is no gain in the quantum well. The oxide aperture is 4 μm and there are a total of seven grating periods within the aperture window. As the computations have to be fully three-dimensional, 891 plane waves have been used to represent the lateral electromagnetic field distribution in each layer: 11 in the direction parallel to the grating and 81 in the direction perpendicular to the grating.

Figure 45.17 shows the computed resonant wavelength of the emitted light and the cavity Q-factor‡ as functions of cap layer thickness normalized to the target cavity wavelength λ_0 of 1.3 μm. This normalized thickness is computed as $\tilde{h}_c = h_c\, n_c/\lambda_0$. As can be seen, both the resonant wavelength and the cavity Q-factor depend periodically on $\tilde{h}_c$, with a period of $\lambda/2$ and the highest Q-factor corresponding to the quarter-wavelength cap thickness (i.e., $\tilde{h}_c = 1/4 + k/2$ with k being an arbitrary integer positive number). However, as the cap thickness increases over $\tilde{h}_c = 1.0$, the consequent maxima correspond to a lower and lower Q-factor. This can be attributed to the fact that the finite size of the grating causes an increase in scattering losses, as the divergent beam is not fully back-reflected to the active region.

45.4.4 Manufacturing Sensitivity of High-Contrast Grating Mirrors

The results in Figure 45.17 show that HCGs can be effectively used to replace DBR mirrors in VCSELs. However, the question arises whether they are similarly robust. The need to etch a precise lithographic pattern can make them prone to manufacturing errors. We therefore investigated the reflectivity of a grating mirror with an imperfect periodic grating. We performed simulations of multiple grating periods in which grating bar thicknesses and positions were randomly varied with a specified standard deviation σ. Due

† 2-λ means that the optical cavity length is exactly two wavelengths.

‡ Q-factor is a dimensionless rate of energy loss relative to the stored energy of the resonator. It can be determined from the equation $Q = \frac{1}{2}\mathrm{Re}(\tilde{\omega})/\mathrm{Im}(\tilde{\omega})$, where $\tilde{\omega}$ is the complex eigen-frequency of the resonant mode in the cavity.

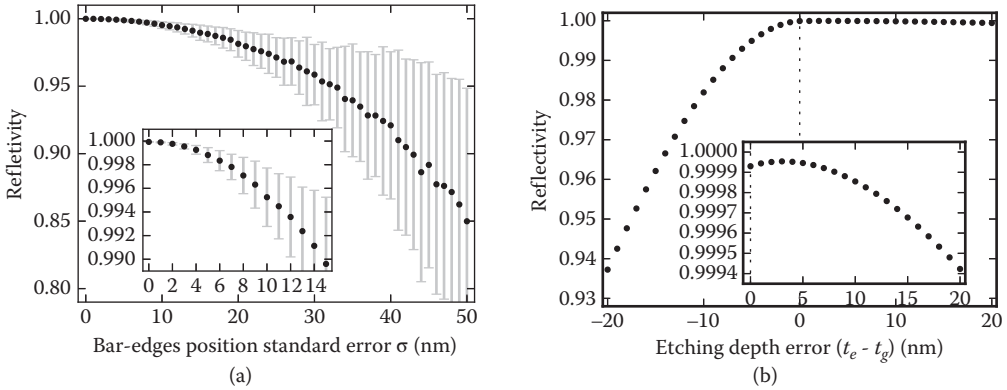


FIGURE 45.18 (a) Computed reflectivity for an HCG mirror with the position of the bar edges varied randomly using different standard deviations. The vertical bars show the uncertainties of the results. The inset shows the magnified part of the full figure. (b) Dependence of reflectivity on the difference between etching depth (h_e) and the thickness of the Si layer (h_g, see Figure 45.11). Negative values mean too shallow holes while positive values designate holes that are too deep. The latter are shown magnified in the inset.

to the randomness introduced, we repeated our calculations 1000 times for each value of σ. The results are presented in Figure 45.18a. Their uncertainties (shown with gray vertical lines) are computed using Student's *t*-distribution for a confidence interval of 90%.

As can be seen, even a small variation causes a significant drop in reflectivity. As the precision of typical photo-lithography is around 10 nm, mirror reflectivity can drop by 0.5% for $\sigma = 10$ nm. Larger manufacturing errors can make the mirrors unusable in VCSELs, as reflectivity can drop well below 90%. The largest allowable error, for which the mirrors reflect more than 99% of the incident light in at least half of the cases, is $\sigma = 15$ nm. The stronger requirement of reflectivity exceeding 99.8% is fulfilled if the error is not greater than $\sigma = 8$ nm. This clearly shows that the main physical phenomenon behind the properties of HCG mirrors is Bragg reflection, in which perfect periodicity plays a crucial role. This contrasts with most modern designs for photonic crystal VCSELs, in which the most important factor is the reduced effective index of the photonic crystal area [8].

Another test we have performed is an analysis of the precision of etching depth. In the ideal case, the Si bars should rest fully separated on a SiO_2 layer. In other words, the spaces between the bars should be etched to a depth exactly equal to the thickness of the Si layer. In reality, this may not always be the case, as the etching can be slightly deeper or slightly shallower. However, locally one can assume uniform etching conditions, whereby the depths of adjacent etched areas should be the same.

In order to investigate the impact of etching depth precision on the performance of HCG mirrors, we computed their reflectivity for various etching depths h_e (h_e is the depth of etching spaces between the bars; in an ideal situation it should be equal to the thickness of the high-index grating layer h_g, see Figure 45.11). Incident light is the same as in the previous case, i.e., the polarization is perpendicular to the bars and the wavelength is 1.3 μm. The results are shown in Figure 45.18b. As one can see, too shallow etching causes a strong decrease in reflectivity. On the other hand, too deep etching does not influence the results significantly. Furthermore, etching 3 nm of the SiO_2 layer can increase reflectivity slightly, although this effect is so weak that it is negligible in practice. The main conclusion is that it is better to etch the space around the bars a little more than desired in order to provide optimum mirror reflectivity even in the case of manufacturing imprecisions.

Similar analysis can be performed for HCG VCSELs, as described in Section 45.4.3. As the reflectivity of HCGs drops with manufacturing imprecisions, so should the VCSEL cavity Q-factor. A numerical analysis of this effect is shown in Figure 45.19. The computation procedure is identical to that used for generating Figure 45.18. However, each point was averaged over 100 simulations instead of 1000 (due to the larger

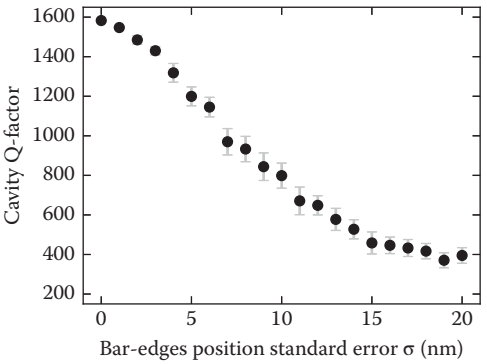


FIGURE 45.19 HCG-VCSEL cavity Q-factor with an imperfect mirror. The edges of the grating bars are varied randomly using different standard deviations. The vertical bars show the uncertainties of the results.

numerical effort necessary to determine the proper VCSEL modes). The resonant wavelength is obviously unaffected, as it is determined by the cavity itself and not the quality of the mirror. We can notice that even a 10 nm error can lead to a twofold reduction in the Q-factor. This must be taken into consideration in designs for devices that are to be produced on an industrial scale. Furthermore, the analysis presented does not consider the effects of the surface roughness of the bars, which to some extent is inevitable in the lithography process.

45.5 Summary

Periodic structures have been used in laser diodes for a long time. Their most basic form are one-dimensional distributed Bragg reflectors, commonly used in VCSELs. However, two- or three-dimensional photonic crystals allow further tuning of laser properties, e.g., to achieve a single-mode light emission. One-dimensional subwavelength high-contrast gratings can make efficient broadband laser mirrors with reduced size and wavelength tuning ability.

The downside of such structures is the increased laser complexity, which requires more advanced numerical methods for the device optical modeling. In particular, only vectorial methods like the finite-difference time-domain method, the plane-wave expansion, the finite element method, the coupled mode model, the plane-wave admittance method, or the plane-wave reflection transfer can yield proper optical characteristics of PC-VCSELs.

References

1. Strutt, J. W. (1887). On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure. *Phil. Mag*. 27, 145–159.
2. Yablonovitch, E. (1987). Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett*. 58, 2059–2062.
3. John, S. (1987). Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett*. 58, 2486–2489.
4. Kittel, C. (2004). *Introduction to Solid State Physics* (8th ed.). Hoboken, NJ: Wiley.
5. Joannopoulos, J. D., R. D. Meade, and J. N. Winn (1995). *Photonic Crystals: Molding the Flow of Light*. Princeton, NY: Princeton University Press.
6. Checoury, X., P. Boucaud, J.-M. Lourtioz, F. Pommereau, C. Cuisin, E. Derouin, O. Drisse, L. Legouezigou, O. L. Legouezigou, F. Lelarge, F. Poingt, G.-H. Duan, S. Bonnefont, D. Mulin, J.

Valentin, O. Gauthier-Lafaye, F. Lozes-Dupuy, and A. Talneau (2005). Distributed feedback-like laser emission in photonic crystal waveguides on InP substrate. *IEEE J. Sel. Top. Quant. Electron*. 11, 1180–1186.

7. Painter, O., J. Vuckovic, and A. Scherer (1999). Defect modes of a two-dimensional photonic crystal in an optically thin dielectric slab. *J. Opt. Soc. Am. B*. 16, 275–285.
8. Danner, A. J., J. J. Raftery Jr., L. P. O., and K. D. Choquette (2006). Single mode photonic crystal vertical cavity lasers. *Appl. Phys. Lett*. 88, 091114.
9. Danner, A. J., J. J. Raftery Jr., N. Yokouchi, and K. D. Choquette (2004). Transverse modes of photonic crystal vertical-cavity lasers. *Appl. Phys. Lett*. 84, 1031–1033.
10. Leisher, P. O., A. J. Danner, and K. D. Choquette (2006). Single-mode 1.3 μm photonic crystal vertical-cavity surface-emitting laser. *IEEE Photon. Technol. Lett*. 18, 2156–2158.
11. Song, D. S., S. H. Kim, H. G. Park, C. K. Kim, and Y. H. Lee (2002). Single-fundamental mode photonic-crystal vertical-cavity surface-emitting lasers. *Appl. Phys. Lett*. 80, 3901–3903.
12. Yokouchi, N., A. J. Danner, and K. D. Choquette (2003). Two-dimensional photonic crystal confined vertical-cavity surface-emitting lasers. *IEEE J. Sel. Top. Quant. Electron*. 9, 1439–1445.
13. Liu, H., M. Yan, P. Shum, H. Ghafouri-Shiraz, and D. Liu (2004). Design and analysis of anti-resonant reflecting photonic crystal VCSEL lasers. *Opt. Express*. 12, 4269–4274.
14. Czyszanowski, T., M. Dems, and K. Panajotov (2007a). Impact of the hole depth on the modal behaviour of long wavelength photonic crystal VCSELs. *J. Phys. D: Appl. Phys*. 40, 2732–2735.
15. Czyszanowski, T., M. Dems, R. P. Sarzała, W. Nakwaski, and K. Panajotov (2011). Precise lateral mode control in photonic crystal vertical-cavity surface-emitting lasers. *IEEE J. Quant. Electron*. 47, 1291–1296.
16. Birks, T. A., J. C. Knight, and P. S. J. Russell (1997). Endlessly single-mode photonic crystal fiber. *Opt. Lett*. 22, 961–963.
17. Taflove, A. and S. C. Hagness (2000). *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (2nd ed.). Boston, MA: Artec House.
18. Johnson, S. G. and J. D. Joannopoulos (2001). Block iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express*. 8, 173–190.
19. Nyakas, P. (2007). Full-vectorial three-dimensional finite element optical simulation of vertical-cavity surface-emitting lasers. *IEEE J. Lightwave Techn*. 25, 2427–2434.
20. Bava, G. P., P. Debernardi, and L. Fratta (2001). Three-dimensional model for vectorial fields in vertical-cavity surface-emitting lasers. *Phys Rev. A*. 63, 23816.
21. Dems, M., R. Kotynski, and K. Panajotov (2005). Plane-wave admittance method—A novel approach for determining the electromagnetic modes in photonic structures. *Opt. Express*. 13, 3196–3207.
22. Dems, M. (2011). Modelling of high-contrast grating mirrors. The impact of imperfections on their performance in VCSELs. *Opto-Electr. Rev*. 19, 340–345.
23. Conradi, O., S. F. Helfert, and R. Pregla (2001). Comprehensive modeling of vertical-cavity laser-diodes by the method of lines. *IEEE J. Quant. Electron*. 37, 928–935.
24. Czyszanowski, T., M. Dems, and K. Panajotov (2007b). Single mode condition and modes discrimination in photonic-crystal 1.3 μm AlInGaAs/InP VCSEL. *Opt. Express*. 15, 5604–5609.
25. Gedney, S. D. (1996). An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices. *IEEE T. Antenn. Propag*. 44, 1630–1639.
26. Dems, M. (2009). Semi-vectorial method based on effective index for VCSEL analysis. *J. Opt. Soc. Am. B*. 26, 792–796.
27. Dems, M., I.-S. Chung, P. Nyakas, S. Bischoff, and K. Panajotov (2010). Numerical methods for modeling photonic-crystal VCSELs. *Opt. Express*. 18, 16042–16054.
28. Panajotov, K., B. Ryvkin, J. Danckaert, M. Peeters, H. Thienpont, and I. Veretennicoff (1998). Polarization switching in VCSEL's due to thermal lensing. *IEEE Photon. Technol. Lett*. 10, 6–8.
29. Fan, S., W. Suh, and J. Joannopoulos (2003). Temporal coupled-mode theory for the Fano resonance in optical resonators. *J. Opt. Soc. Am. A*. 20 (3), 569–572.

30. Karagodsky, V., F. G. Sedgwick, and C. J. Chang-Hasnain (2010). Theoretical analysis of subwavelength high contrast grating reflectors. *Opt. Express*. 18, 16973–16988.
31. Chang-Hasnain, C. J. and W. Yang (2012). High-contrast gratings for integrated optoelectronics. *Adv. Opt. Photon.* 4, 379–440.
32. Rosenblatt, D., A. Sharon, and A. A. Friesem (1997). Resonant grating waveguide structures. *IEEE J. Quant. Electron*. 33, 2038–2059.
33. Huang, M. C. Y., Y. Zhou, and C. J. Chang-Hasnain (2008). A nanoelectromechanical tunable laser. *Nat. Photon.* 2, 180–184.

46

Single-Photon Sources

Niels Gregersen
Dara P. S. McCutcheon
and
Jesper Mørk

46.1 Introduction 585
Characteristics • Single-Photon Generation Schemes • Photon Extraction Strategies
46.2 Modeling of Spontaneous Emission 590
Classical Dipole • Spontaneous Emission • Simulation Techniques
46.3 Cavity QED Approach 594
Micropillar
46.4 Waveguide QED Approach 596
Photonic Nanowire • Photonic Crystal Waveguide
46.5 Planar Dielectric Antenna Approach 600
46.6 Decoherence Effects 601
Quantum Emitter Master Equation • Emitter-Cavity Master Equation • Spectral Wandering • Coupling to Longitudinal Acoustic Phonons
46.7 Summary 604

46.1 Introduction

The single-photon source (SPS) is a solid-state device capable of emitting single photons on demand. It is a key component within the emerging field of quantum information processing, where the information is encoded on single photons. In these applications, the quantum mechanical properties of the single photon are exploited, e.g., to enable secure communication where undetected eavesdropping is impossible or to perform quantum calculations taking advantage of massive parallelization.

This tutorial chapter reviews the characteristic figure of merits of the deterministic SPS based on a quantum emitter in a semiconductor solid-state system. The theory and the modeling methods used to describe single-photon emission are discussed. Photon extraction strategies based on cavity and waveguide quantum electrodynamics (QED) are presented, and finally we discuss decoherence mechanisms due to the solid-state environment.

46.1.1 Characteristics

For use in quantum information technology, the SPS must meet three basic performance requirements: pure single-photon emission, indistinguishability, and high efficiency, as discussed separately in the three following subsections.

46.1.1.1 Pure Single-Photon Emission

The most fundamental characteristic of the SPS is that it should emit only one photon at a time. This ability is tested in the so-called Hanbury, Brown, and Twiss (HBT) experiment illustrated in Figure 46.1.

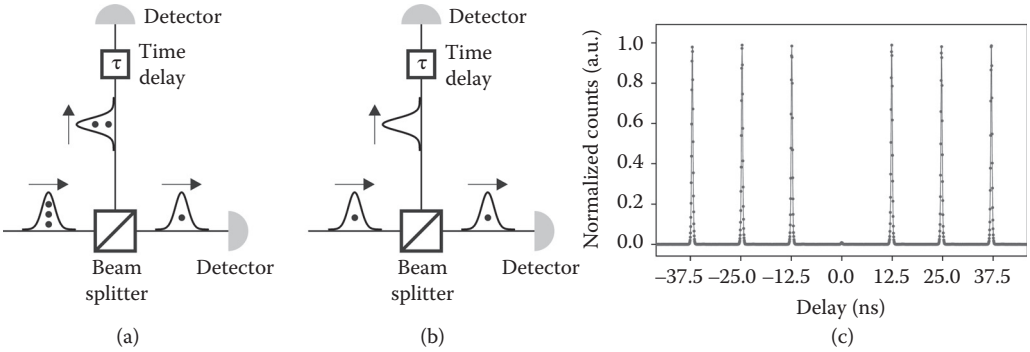


FIGURE 46.1 The Hanbury, Brown, and Twiss experiment illustrated for input pulses containing (a) multiple photons and (b) single photons. (c) The $g^{(2)}(\tau)$ autocorrelation function for a micropillar SPS. (Reprinted with permission from Ding, X. et al., *Physical Review Letters*, 116, 020401, 2016. Copyright 2016 by the American Physical Society.)

Here an optical pulse ideally containing only one photon is directed toward a beam splitter and is either reflected or transmitted toward two detectors. A variable time delay τ is implemented in one of the two output paths, and it is adjusted such that the photons from either of the two ports of the beam splitter arrive simultaneously at the two detectors for zero time delay. The single-photon emission ability is now characterized by measuring the autocorrelation function $g^{(2)}(\tau)$. The experiment is carried out by sending in light pulses from the SPS and by measuring the number of simultaneous detection events in the two detectors as function of time delay.

If the input pulse contains multiple photons as depicted in Figure 46.1a, some photons may be reflected while others are transmitted leading to coincidence clicks in the two detectors at zero time delay such that $g^{(2)}(\tau = 0) \neq 0$. However, if the pulse contains only a single photon as illustrated in Figure 46.1b, this photon will be either transmitted or reflected and no coincidence events are measured. An example of such a HBT measurement is given in Figure 46.1c (Ding et al., 2016), where we observe coincidence counts as function of delay for a SPS generating pulses every 12 ns. The peaks in the measurement correspond to integer values of the pulse time separation, where photons from different pulses may lead to coincidence clicks. However, at zero time delay, the correlation counts are strongly suppressed corresponding to an ideal single-photon emission with $g^{(2)}(\tau = 0) = 0$.

46.1.1.2 Indistinguishability

A second requirement for the SPS is that the emitted photons should be quantum mechanically indistinguishable, meaning that the emitted photons should have the same wavelength, phase variation across the pulse, polarization, optical mode profile, etc. The single-photon indistinguishability is characterized in the Hong–Ou–Mandel (HOM) experiment sketched in Figure 46.2 (Hong et al., 1987).

In this experiment, two single-photon pulses are inbound on a 50–50 beam splitter, where the two pulses are either transmitted or reflected into the output paths. A variable time delay τ is implemented in one of the two output paths adjusted such that the photons reach the detectors simultaneously at zero time delay. The experiment then takes place by sending in single-photon pulses along the input paths and by measuring the correlation function $g^{(2)}(\tau)$ describing the simultaneous detection events for this setup as function of time delay.

The impinging photons may leave the beam splitter either in separate output paths or together as illustrated in Figure 46.2a and b, respectively. In the case where the input photons are distinguishable, the output photons may leave the beam splitter along separate paths leading to $g^{(2)}(\tau = 0) \neq 0$. However, when the input photons are completely indistinguishable and arrive simultaneously at the beam splitter, they always exit the beam splitter along the same output path as depicted in Figure 46.2b, resulting in a $g^{(2)}(\tau = 0) = 0$. This phenomenon is known as the HOM effect and can be understood from a quantum-optical consideration of the probabilities of the four possible output states sketched in Figure 46.2c. The

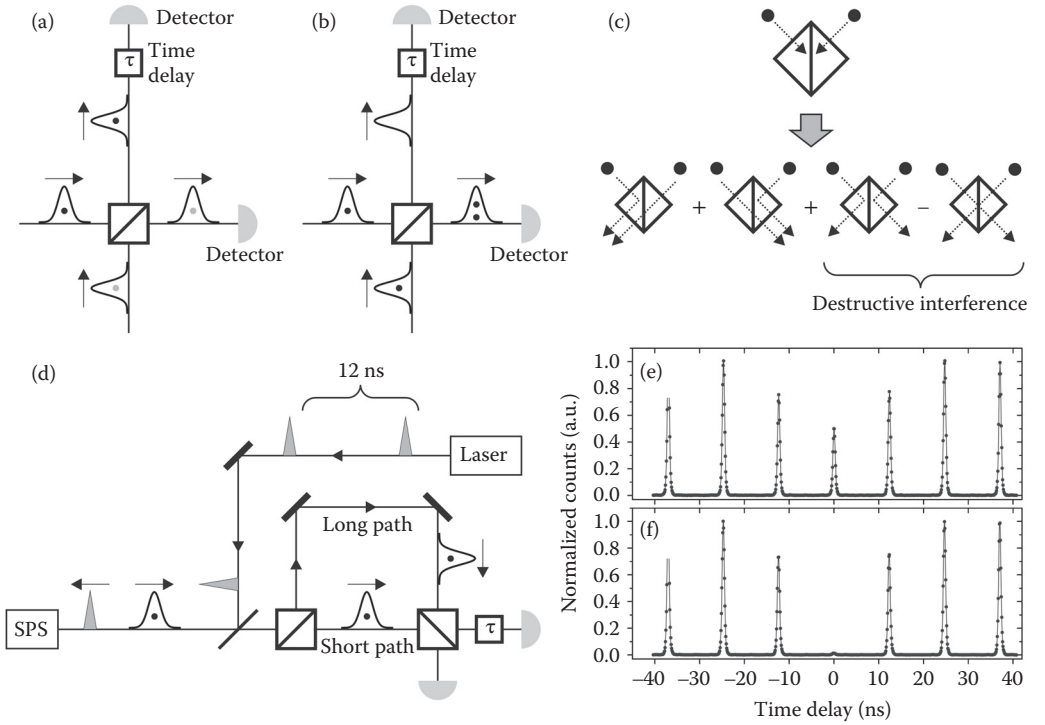


FIGURE 46.2 Outcome of the Hong–Ou–Mandel experiment for (a) distinguishable and (b) indistinguishable photons. (c) Various possible output states. Sketch of a typical experimental setup. Measured coincidence counts in (e) cross and (f) parallel polarization. (Reprinted figure with permission from Ding, X. et al., *Physical Review Letters*, 116, 020401, 2016. Copyright 2016 by the American Physical Society.)

probability amplitude of the photons leaving the beam splitter along separate output paths consists of two contributions corresponding to the photons being both either reflected or transmitted at the beam splitter, and for indistinguishable photons these two contributions add up destructively (Gerry and Knight, 2008).

In most HOM experiments, the impinging photons are generated by the same source using the setup sketched in Figure 46.2d. A laser generates classical pulses at a typical repetition rate of 80 MHz. The pulses are transmitted to the SPS, and each laser pulse triggers a single-photon emission event leading to a train of single photons with 12.5-ns time separation inbound on the first beam splitter. Two subsequent photons in the train can propagate along the short and the long paths in various configurations. The long and short path lengths are adjusted such that a first photon taking the long path and a second photon taking the short path will arrive simultaneously at the beam splitter. Typical HOM measurements for indistinguishable photons having orthogonal or parallel polarization are illustrated in Figure 46.2e and f, respectively. For orthogonal polarizations, the photons are distinguishable and $g^{(2)}(\tau = 0) \neq 0$, whereas for parallel polarizations the HOM effect with $g^{(2)}(\tau = 0) = 0$ is observed. Again, $g^{(2)}(\tau)$ has peaks for time delays equal to an integer times the pulse separation. These peaks correspond to coincidence events for photon pairs that do not arrive simultaneously at the beam splitter and are thus not subject to the HOM effect. The indistinguishability is then quantified as one minus the normalized area of the $\tau = 0$ peak and should ideally equal unity.

46.1.1.3 Efficiency

The third important characteristic of the SPS is its efficiency or brightness (Gregersen et al., 2013). The ideal SPS should be deterministic and provide single photons on demand. It should act as a "single-photon gun" such that every trigger releases exactly one photon into the collection optics. The efficiency ε is then

defined as the number of photons detected by the collection optics per trigger. Within the QED design approaches discussed in Sections 46.3 and 46.4, the efficiency is characterized using two governing parameters as illustrated in Figure 46.3. The first is the spontaneous emission factor β describing the fraction of light coupled from the emitter into the optical mode of interest. The β factor can be written as

$$\beta = \frac{\Gamma_M}{\Gamma_M + \Gamma_{Rad}}, \tag{46.1}$$

where Γ_M is the spontaneous emission rate into the optical mode and Γ_{Rad} is the emission rate into background radiation modes. One may also add a term Γ_{NonRad} to the denominator in the definition (46.1) to take into account nonradiative recombination channels of the quantum emitter. Whereas Γ_M and Γ_{Rad} can be modified by tailoring the photonic environment, the parameter Γ_{NonRad} represents inherent properties of the emitter that generally cannot easily be controlled.

The second governing parameter is the outcoupling efficiency γ describing the fraction of light transmitted from the optical mode to the first objective lens. In this simple picture, the efficiency is given by $\varepsilon = \beta\gamma$. In a multiphoton interference experiment involving j photons, the total success probability of the experiment will scale as ε^j. For scalable optical quantum information technologies involving a large number of photons, a single-photon generation efficiency close to unity is thus required.

46.1.2 Single-Photon Generation Schemes

A major workhorse for generating single photons has been the spontaneous parametric down-conversion (SPDC) process (Brendel et al., 1999). Let us recall that in a nonlinear isotropic material, the polarization **P** is related to the electric field **E** as

$$\mathbf{P} = \varepsilon_0 \left(\chi^{(1)}\mathbf{E} + \chi^{(2)}\mathbf{E}^2 + \chi^{(3)}\mathbf{E}^3 + \ldots \right), \tag{46.2}$$

where $\chi^{(n)}$ is the nth-order susceptibility. In the SPDC scheme, a crystal featuring a $\chi^{(2)}$ nonlinearity such as lithium niobate is illuminated with a strong laser pump pulse with an energy $\hbar\omega_p$ as sketched in Figure 46.4. Provided that proper phase-matching conditions are met, the $\chi^{(2)}$ nonlinearity then enables conversion of a pump photon to two single photons, the signal and the idler photons of energies $\hbar\omega_s$ and $\hbar\omega_i$, respectively, such that $\hbar\omega_p = \hbar\omega_s + \hbar\omega_i$. The advantage of the SPDC source is its simplicity and the high

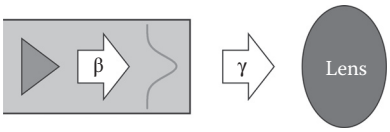


FIGURE 46.3 Efficiency model for the quantum electrodynamics approach. The single-photon emitter is represented by the triangle.

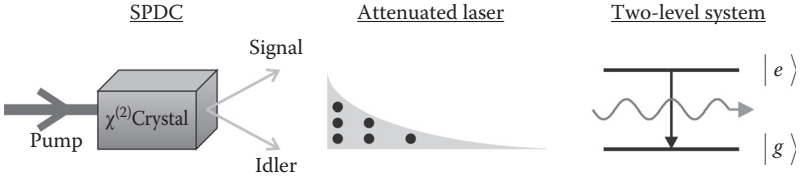


FIGURE 46.4 Major single-photon generation schemes.

degree of indistinguishability of the signal and idler photons. Its main drawback is that the downconversion process is inherently probabilistic, leading to efficiencies below 0.01. While the efficiency may be improved by cranking up the pump power, this occurs at the cost of significant increase of the probability of multiphoton emission. The probabilistic nature thus limits SPDC to experiments involving only a few photons.

Another commercially important source is the attenuated laser SPS (Buttler et al., 2000). A pulse generated using a classical laser propagates through an absorbing material and is attenuated to the single-photon level. The advantage of the attenuated laser is its simplicity, and similar to the SPDC process its main drawback is the probabilistic nature of the attenuation leading to either low single-photon generation efficiency or high probability of multiphoton events. Additionally, the generated photons are not indistinguishable, preventing their use in many quantum optics and quantum information experiments. However, the attenuated laser remains an important device in commercial products for quantum key distribution, where clean single-photon emission and indistinguishability are not required.

The quantum-dot emitter (Shields, 2007) has recently emerged as a solution, overcoming the probabilistic limitations of the SPDC and attenuated laser strategies. It consists of a two-level system as shown in Figure 46.4, where an electron can occupy either a ground state $|g\rangle$ or an excited state $|e\rangle$ characterized by energies E_g and E_e, respectively. The system is initially in the ground state and is then excited using an optical or electrical pulse, after which the electron returns to the ground state by the spontaneous emission of a single photon of energy $\hbar\omega_0 = E_e - E_g$. A main advantage of the two-level emitter is that the upper level can contain only one excitation at a time, thereby ensuring the single-photon nature of the emitted light, combined with the deterministic nature of the spontaneous emission event and the possibility of structuring the medium to realize highly directional emission, potentially allowing for unity efficiency.

Any solid-state system featuring discretized energy levels can in principle be employed as a SPS. The InAs quantum dot (QD) grown in a GaAs host material as shown in Figure 46.5a represents a mature platform for a deterministic SPS. The QD consists of a pyramidal-shaped island of InAs atoms surrounded by GaAs. Due to the lower bandgap of InAs, the electron wavefunction will be fully confined in all three dimensions leading to discretized energy levels in the conduction and valence bands depicted in Figure 46.5b. The emission of InAs QD ensembles features strong inhomogeneous broadening with typical emission wavelength between 900 and 950 nm. The lowest energy *s* state in the conduction band is typically used as the excited state. Due to the spin degree of freedom, the *s* state supports a biexciton leading to two-photon emission. However, the InAs QD features a fine-structure splitting arising from the exciton binding energy of a few nanometers allowing for the selection of the excitonic line using spectral filtering thus enabling single-photon emission at cryogenic temperatures. The indistinguishability of the emitted photons depends greatly on the charge environment and the excitation scheme employed, as discussed in detail in Section 46.6.

Other two-level systems including diamond nitrogen-vacancy centers (Babinec et al., 2010) and organic molecules (Siyushev et al., 2014) have been pursued for single-photon generation. These systems feature the advantages of room temperature operation and very long spin coherence times. However, the maturity

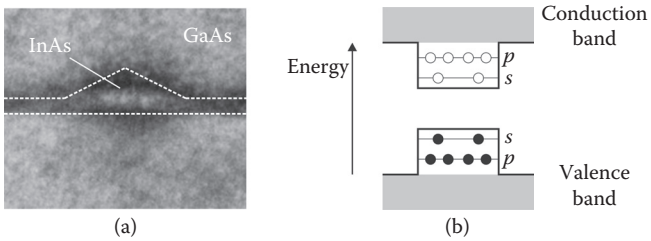


FIGURE 46.5 Scanning electron micrograph (a) and energy level diagram (b) of an InAs quantum dot.

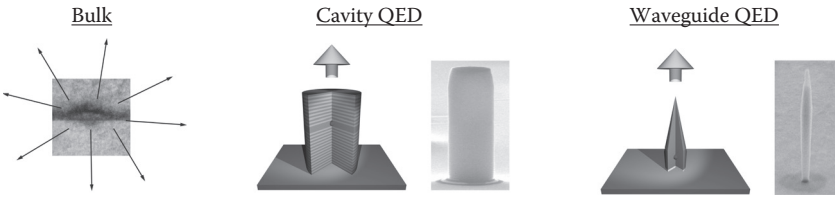


FIGURE 46.6 Major photon extraction strategies. (Reprinted from Gregersen, N et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2013 IEEE and Reprinted by permission from Macmillan Publishers Ltd: Claudon, J et al., *Nature Photonics* 4(3): 174–177, copyright 2010.)

of III–V semiconductor cleanroom fabrication has made the InAs QD the leading platform for highly efficient SPSs, and for this reason we focus in this tutorial on the InAs QD.

46.1.3 Photon Extraction Strategies

While the self-assembled InAs QD features an internal quantum efficiency η of almost unity, a major challenge in single-photon engineering lies in the efficient extraction of the emitted photon (Gregersen et al., 2013). Figure 46.6 summarizes the main SPS photon extraction strategies. The simplest approach consists of a planar sample with a QD in a bulk material. The main advantages of this approach are its simplicity as no additional structuring is needed after QD growth as well as the absence of fluctuating surface charges near the QD as discussed further in Section 46.6. The disadvantage is the poor extraction efficiency. The symmetry of the system leads to light emission in all directions limiting the efficiency to only a few percent.

The first strategy proposed to improve the efficiency and control the light emission was based on cavity QED, where the QD was placed in a microcavity (Gérard, 2003). Here the Purcell effect ensures a preferential single-photon emission from the QD into the cavity mode. This preferential coupling requires a careful spectral alignment between the QD emission line and the cavity line as discussed in detail in Section 46.3. This approach has led to efficiencies of 0.8 (Gazzano et al., 2013) as well as an indistinguishability of 0.99 from the micropillar cavity SPS (Ding et al., 2016; Somaschi et al., 2016), representing the current state of the art in highly efficient sources of indistinguishable photons.

A second design strategy is based on waveguide QED, where the QD is positioned inside a waveguide (Claudon et al., 2013). The preferential coupling to the waveguide mode is ensured by a dielectric screening effect (Bleuse et al., 2011). The main advantage of this nonresonant approach is that no spectral alignment is required. While an efficiency of 0.75 has been demonstrated for the photonic nanowire SPS, the demonstration of highly indistinguishable photons from a waveguide-based SPS is still lacking.

46.2 Modeling of Spontaneous Emission

A major challenge in engineering deterministic SPSs lies in optimizing the photon extraction efficiency and the indistinguishability to achieve values close to unity. Proper modeling of the light emission from a two-level system generally requires a QED description of the spontaneous emission process. However, SPSs generally operate in the weak coupling regime, and here it turns out that the spontaneous emission rate can be calculated in a purely classical manner (Novotny and Hecht, 2012).

46.2.1 Classical Dipole

We first consider the electromagnetic field generated by a classical current distribution $\mathbf{j}$ emitting light at the angular frequency ω_0. We thus work in the frequency domain with harmonic time dependence such that $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega_0 t}$ and $\mathbf{j}(\mathbf{r}, t) = \mathbf{j}(\mathbf{r})e^{-i\omega_0 t}$. Furthermore, we assume nonmagnetic materials and the

absence of free charges for simplicity. The classical Maxwell's equations are then combined to yield the second-order wave equation for the electric field:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \varepsilon_r(\mathbf{r}) k_0^2 \mathbf{E}(\mathbf{r}) = i\omega_0 \mu_0 \mathbf{j}(\mathbf{r}), \tag{46.3}$$

where $k_0 = \omega_0/c$ is the free-space wavenumber and ε_r is the dielectric constant.

Equation 46.3 can be solved in an elegant manner by introducing the dyadic Green's function $\overline{\overline{G}}$ (Novotny and Hecht, 2012). It is defined as the solution to

$$\nabla \times \nabla \times \overline{\overline{G}}(\mathbf{r}, \mathbf{r}') - \varepsilon_r(\mathbf{r}) k_0^2 \overline{\overline{G}}(\mathbf{r}, \mathbf{r}') = \overline{\overline{I}} \delta\left(\mathbf{r} - \mathbf{r}'\right), \tag{46.4}$$

where $\overline{\overline{I}}$ is the unit dyad. $\overline{\overline{G}}$ and $\overline{\overline{I}}$ are both 3×3 tensors and we stress that $\overline{\overline{G}}(\mathbf{r}, \mathbf{r}')$ depends on both $\mathbf{r}$ and $\mathbf{r}'$. Now, using this definition of the dyadic Green's function, we note that

$$\mathbf{E}(\mathbf{r}) = i\omega_0 \mu_0 \int_V \overline{\overline{G}}(\mathbf{r}, \mathbf{r}') \mathbf{j}(\mathbf{r}') \, \mathrm{d}\mathbf{r}' \tag{46.5}$$

is a solution to Equation 46.3. The Green's function $\overline{\overline{G}}(\mathbf{r}, \mathbf{r}')$ can thus be interpreted physically as the field generated at the position $\mathbf{r}$ by a current at the position $\mathbf{r}'$. Once the Green's function has been calculated, the entire field profile can be determined from Equation 46.5.

The standard procedure for computing the power P emitted by the current distribution is to integrate the time average of the normal component of the Poynting vector over the surface S of volume V enclosing the current distribution as

$$P = \frac{1}{2} \int_S Re\left(\mathbf{E} \times \mathbf{H}^*\right) \cdot \mathbf{n}_S \mathrm{d}S, \tag{46.6}$$

where $\mathbf{n}_S$ is a unit vector normal to the surface S. Now, the average energy dissipation rate W inside the volume is given by

$$W = -\frac{1}{2} \int_V Re\left(\mathbf{j}^* \cdot \mathbf{E}\right) \mathrm{d}\mathbf{r}, \tag{46.7}$$

and according to Poynting's theorem, we have $P = W$.

We now consider the current distribution generated by a point dipole $\mathbf{d}$ positioned at $\mathbf{r}_0$, as illustrated in Figure 46.7, such that

$$\mathbf{j}(\mathbf{r}) = -i\omega_0 \mathbf{d} \delta\left(\mathbf{r} - \mathbf{r}_0\right). \tag{46.8}$$

By inserting Equation 46.8 into Equation 46.5 and subsequently Equation 46.5 into Equation 46.7, we obtain an expression for the power P in terms of the dipole moment given by

$$P = \frac{\omega_0^3 \mu_0 \left|\mathbf{d}\right|^2}{2} \mathrm{Im}\left(\mathbf{n}_\mathrm{d} \cdot \overline{\overline{G}}(\mathbf{r}_0, \mathbf{r}_0) \cdot \mathbf{n}_\mathrm{d}\right), \tag{46.9}$$

where $\mathbf{n}_\mathrm{d} = \mathbf{d}/\left|\mathbf{d}\right|$ is the orientation of the dipole moment. The power P emitted by the point dipole can thus be obtained by evaluating the imaginary part of the Green's function at the position of the dipole.

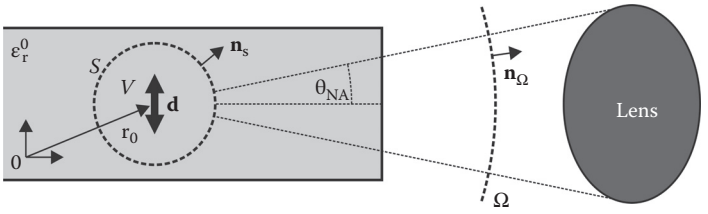


FIGURE 46.7 Light emission from a classical dipole **d** placed at the position $\mathbf{r}_0$ inside a volume V in a material of dielectric constant ε_r^0. The volume is limited by the surface S having normal unit vector $\mathbf{n}_S$. The entire structure is enclosed in a sphere Ω with normal unit vector $\mathbf{n}_\Omega$.

46.2.2 Spontaneous Emission

In the weak coupling regime, the total spontaneous emission rate Γ of a two-level system at the position $\mathbf{r}_0$, with transition energy $\hbar\omega_0$ and with dipole moment **d,** is derived from Fermi's golden rule as

$$\Gamma = \frac{\pi\omega_0}{\hbar\varepsilon_0} |\mathbf{d}|^2 \rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right). \tag{46.10}$$

Here, $\rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right)$ is the local density of states (LDOS) defined as

$$\rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right) = \int \left|\mathbf{n}_\mathrm{d} \cdot \widetilde{\mathbf{E}}_\alpha(\mathbf{r}_0)\right|^2 \delta(\omega_0 - \omega_\alpha) d\alpha, \tag{46.11}$$

where $\widetilde{\mathbf{E}}_\alpha$ are the so-called normal modes solutions to the wave equation in the absence of free currents and charges:

$$\nabla \times \nabla \times \widetilde{\mathbf{E}}_\alpha(\mathbf{r}) - \varepsilon(\mathbf{r}) \left(\frac{\omega_\alpha}{c}\right)^2 \widetilde{\mathbf{E}}_\alpha(\mathbf{r}) = 0, \tag{46.12}$$

and the index α is a short notation for all appropriate discrete and continuous parameters describing a mode. The normal modes are normalized such that

$$\int \varepsilon_r(\mathbf{r}) \widetilde{\mathbf{E}}^*_\alpha(\mathbf{r}) \cdot \widetilde{\mathbf{E}}_{\alpha'}(\mathbf{r}) \mathrm{d}\mathbf{r} = \delta(\alpha - \alpha'). \tag{46.13}$$

We now write the dyadic Green's function as an expansion in normal functions. Referring the reader interested in the derivation to Novotny and Hecht (2012), we simply state the result:

$$Im \left(\overline{\overline{G}}(\mathbf{r}_0, \mathbf{r}_0, \omega_0)\right) = \frac{\pi c^2}{2\omega_0} \int \widetilde{\mathbf{E}}^*_\alpha(\mathbf{r}_0) \widetilde{\mathbf{E}}_\alpha(\mathbf{r}_0) \delta(\omega_0 - \omega_\alpha) d\alpha, \tag{46.14}$$

where the integrand contains the outer product of the normal modes. Comparing this expression to Equation 46.11, we observe that the imaginary part of the Green's function is closely related to the LDOS. Indeed, outer multiplications of the dyad with the dipole unit vector and using the definition (46.11) lead to

$$\rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right) = \frac{2\omega_0}{\pi c^2} \mathrm{Im} \left(\mathbf{n}_\mathrm{d} \cdot \overline{\overline{G}}(\mathbf{r}_0, \mathbf{r}_0, \omega_0) \cdot \mathbf{n}_\mathrm{d}\right), \tag{46.15}$$

and we observe that, except for physical constants, the LDOS is identical to the average power emitted by a classical dipole as given in Equation 46.9. This means that we obtain the very important relationship

$$\frac{\Gamma}{\Gamma_{\text{Bulk}}} = \frac{P}{P_{\text{Bulk}}}, \tag{46.16}$$

which states that the spontaneous emission rate, Γ, of a two-level system embedded in a medium with dielectric constant ε_r^0 normalized to its value, Γ_{Bulk}, in a bulk material of dielectric constant ε_r^0 is equal to the power dissipation, P, from a classical dipole normalized to the rate, P_{Bulk}, in the bulk material.

46.2.3 Simulation Techniques

Analytic expressions for the Green's function are only available in a limited number of geometries and generally we rely on numerical simulation techniques to compute the electric field. When using spatial discretization methods such as the finite-difference time-domain technique and the finite-elements method (Lavrinenko et al., 2014), the field is expanded on a grid as sketched in Figure 46.8a and either Maxwell's equations or a vectorial wave equation are directly solved to give the optical field.

Alternatively, one may use a modal method (Lavrinenko et al., 2014). Here, the geometry is sliced into layers uniform along a propagation axis as illustrated in Figure 46.8b and the eigenmodes of each layer are then computed. The optical field is then expanded on the eigenmodes, and the fields at each side of a layer interface are connected using the scattering matrix formalism. This allows for direct access to scattering coefficients and the possibility of studying the various parts of the SPS individually.

As we have seen, the QD can be modeled as a classical point dipole placed inside the geometry under study. The electric field profile generated by the dipole is then computed, and the classical power P can then be obtained from Equation 46.7. We note that, for the current source defined by Equation 46.8, the real part of the electric field diverges at the position of the dipole, and an evaluation of the power P by integrating over a small sphere using Equation 46.6 is thus often preferred for numerical stability. The power collected by the lens of numerical aperture NA can be evaluated by computing the electromagnetic field on the surface of a large sphere Ω using a standard near-field to far-field transformation (Balanis, 1989) and integrating the Poynting vector over the unit solid angle subject to the condition $\theta < \theta_{\text{NA}}$, where $\theta_{\text{NA}} = \text{asin}(\text{NA})$. The efficiency ε is then given as

$$\varepsilon = \frac{\frac{1}{2}\int\limits_{\theta<\theta_{\text{NA}}} \text{Re}\,(\mathbf{E}\times\mathbf{H}^*)\cdot\mathbf{n}_\Omega d\Omega}{P}. \tag{46.17}$$

While Equation 46.17 is exact, this direct procedure does not give complete insight into the governing physics. For this reason, the efficiency in the QED SPS designs is instead treated using an elements-splitting

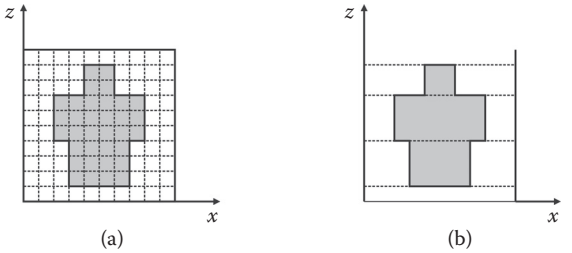


FIGURE 46.8 Discretization schemes for (a) the finite-difference time-domain technique and for (b) the modal method.

approach where β, Γ_M, $\Gamma_{Rad,}$ and γ are analyzed and optimized separately and the efficiency is computed using $\varepsilon = \beta\gamma$. The rates Γ_M and Γ_{Rad} are available from Equation 46.11 as

$$\Gamma_M = \frac{\pi\omega_0}{\hbar\varepsilon_0} |\mathbf{d}|^2 \left|\mathbf{n}_d \cdot \widetilde{\mathbf{E}}_M(\mathbf{r}_0)\right|^2 \tag{46.18}$$

$$\Gamma_{Rad} = \frac{\pi\omega_0}{\hbar\varepsilon_0} |\mathbf{d}|^2 \int\limits_{\alpha \neq M} \left|\mathbf{n}_d \cdot \widetilde{\mathbf{E}}_\alpha(\mathbf{r}_0)\right|^2 \delta(\omega_0 - \omega_\alpha) d\alpha, \tag{46.19}$$

and we observe that their evaluation requires access to the electric field profiles of the optical mode of interest as well as the radiation modes.

When performing numerical simulations of SPS designs, the advantages of the spatial discretization methods are their maturity, their availability in commercial user-friendly software packages (Lumerical, n.d.; JCMWave, n.d.), and their ability to exploit parallelization as required in the treatment of large geometries. Their major weakness is that they do not give direct access to the mode profiles $\widetilde{\mathbf{E}}_\alpha(\mathbf{r})$. While the total power P can be obtained from Equation 46.6 or Equation 46.7, the individual rates Γ_M and Γ_{Rad} are not directly available and β thus cannot be directly evaluated. On the other hand, in the modal expansion methods, the mode profiles $\widetilde{\mathbf{E}}_\alpha(\mathbf{r})$ are directly available, and this represents a major advantage. However, the drawbacks of the modal method are the lack of user-friendly software packages as well as its difficulty in handling large 3-D geometries due to the lack of eigenvalue solvers capable of exploiting parallelization.

46.3 Cavity QED Approach

Inspection of Equation 46.1 reveals two options for optimizing β. The first is to increase Γ_M using cavity QED effects in the weak coupling regime. Here, the emitter is placed inside a cavity with quality factor Q, mode volume V, and resonance wavelength $\lambda_r = 2\pi c/\omega_r$ as illustrated in Figure 46.9a. If the emitter is placed at a maximum of the optical field profile and the emitter and cavity lines are spectrally aligned such that $\omega_r = \omega_0$, the spontaneous emission rate Γ_M into the cavity mode is enhanced by the factor F_P known as the Purcell factor (Purcell, 1946). It describes the rate Γ_M relative to the rate Γ_{Bulk} for an emitter in a bulk medium and is given by (Gérard, 2003)

$$F_P = \frac{\Gamma_M}{\Gamma_{Bulk}} = \frac{3}{4\pi^2} \frac{Q}{V} \left(\frac{\lambda_r}{n}\right)^3, \tag{46.20}$$

where n is the refractive index of the material enclosing the emitter. The mode volume is computed as†

$$V = \frac{\int \varepsilon_r(\mathbf{r}) \left|\widetilde{\mathbf{E}}_M(\mathbf{r})\right|^2 d\mathbf{r}}{\varepsilon_r(\mathbf{r}_{max}) \left|\widetilde{\mathbf{E}}_M(\mathbf{r}_{max})\right|^2}, \tag{46.21}$$

where the denominator is evaluated at the position $\mathbf{r}_{max}$ of a maximum of the electric field intensity profile.

The expression (Equation 46.20) holds provided the emitter line is narrow compared to the cavity linewidth. If this is not the case, an integration over the electronic density-of-states should be performed,

† When integrated over all space, the integral in Equation 46.21 diverges. To remedy this, it is possible to define a generalized mode volume (Kristensen et al., 2012) that converges in an unambiguous way. However, in high-Q cavities, one can usually employ Equation 46.21 and limit the integration to the device geometry with negligible error.

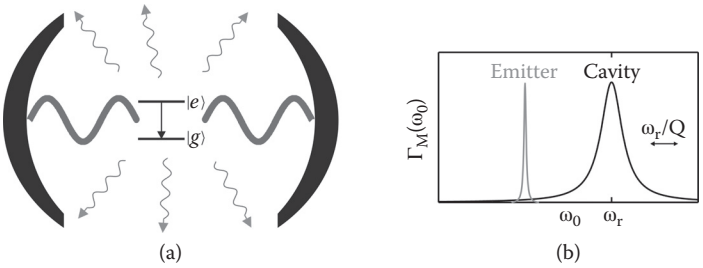


FIGURE 46.9 (a) Cavity-enhanced spontaneous emission. (b) Cavity emission rate Γ_M as a function of frequency.

which generally leads to a reduction in the enhancement. The β factor can be written in terms of the Purcell factor as

$$\beta = \frac{F_P}{F_P + (\Gamma_{Rad}/\Gamma_{Bulk})} \approx \frac{F_P}{F_P + 1}. \tag{46.22}$$

For micropillars with diameters in the few micrometer range, the spontaneous emission into radiation modes is approximately equal to the rate in a bulk material, and we can make the approximation $\Gamma_{Rad} = \Gamma_{Bulk}$ leading to the second equality in Equation 46.22.

We observe from Equations 46.20 through 46.22 that a high-efficiency cavity-based SPS requires a large Q factor combined with a low mode volume, and the optical design challenge for the cavity-based SPS is to maximize the ratio Q/V while maintaining a high outcoupling coefficient γ.

The expression (46.20) assumes perfect spectral alignment between the emitter and the cavity line. If they are detuned, a correction factor given by a Lorentzian lineshape function is introduced, such that (Lorke et al., 2013)

$$\frac{\Gamma_M(\omega_0)}{\Gamma_{Bulk}} = \frac{3}{4\pi^2}\frac{Q}{V}\left(\frac{\lambda_r}{n}\right)^3 \frac{(\delta\omega_r)^2}{(\delta\omega_r)^2 + (\omega_0 - \omega_r)^2}, \tag{46.23}$$

where $\delta\omega_r = \omega_r/Q$ is the cavity linewidth. The cavity emission rate Γ_M is plotted in Figure 46.9b as function of frequency and we observe that for an emitter-cavity misalignment of only a few linewidths $\delta\omega_r$, the enhancement is lost. For the cavity design approach to work, it is thus absolutely necessary to control the spectral alignment.

46.3.1 Micropillar

The AlGaAs microcavity pillar geometry (Reitzenstein and Forchel, 2010) illustrated in Figure 46.10a is one of the most popular platforms for constructing a high-efficiency SPS. It is a vertical structure consisting of a GaAs cavity with an embedded InAs QD surrounded by GaAs/AlAs distributed Bragg reflectors (DBRs). The optical cavity mode profile is confined vertically by the DBRs and laterally by total internal reflection. The emitted light is collected by an objective lens from above.

The thicknesses of the cavity and the DBR layers should be chosen as λ_d/n_{eff} and $\lambda_d/(4n_{eff})$, respectively, where λ_d is the design wavelength and n_{eff} is the effective index of the fundamental mode. These thicknesses ensure that the resonance is centered on the DBR stopband for maximum reflection (Coldren et al., 2012).

In the ideal structure, light escapes only through the bottom and top DBRs. The light emitted into the substrate is lost and the reflectivity of the bottom DBR should thus be significantly higher than that of the top DBR to obtain a high efficiency. The transmission γ to the lens for a micropillar of infinite diameter is presented in Figure 46.10b, and we observe that $\gamma > 0.99$ requires approximately 18 extra layer pairs in the bottom DBR.

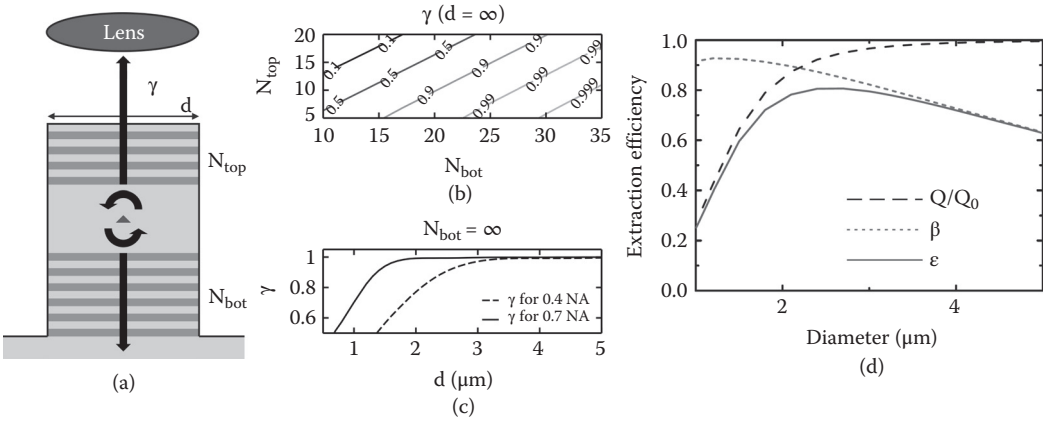


FIGURE 46.10 (a) The micropillar geometry. Transmission γ as function of DBR layer pairs (b) and as function of the pillar diameter (c). (d) Extraction efficiency as function of the pillar diameter. (Reprinted from Gregersen, N. et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2013 IEEE and reprinted by permission from Macmillan Publishers Ltd: Gazzano, O. et al., *Nature Communications* 4: 1425, copyright 2013.)

After the numbers of layer pairs are fixed, the pillar diameter should be chosen. The transmission γ as function of the pillar diameter is shown in Figure 46.10c for a perfectly reflecting bottom DBR. The far-field emission pattern is related to the Fourier transform of the near-field profile. For small diameters, the far-field emission pattern is broad and the transmission is low. For larger diameters, the beam divergence is reduced and the transmission improves. We observe that a pillar diameter above ∼2 μm leads to near-unity transmission for a 0.7 NA lens.

In the optimization of the extraction efficiency, two additional mechanisms should be considered. While the ideal structure emits light only along the vertical axis, fabricated pillar structures suffer from sidewall imperfections leading to in-plane scattering of light. The amount of light lost can be estimated by measuring the Q factor and comparing with the quality factor Q_0 of a planar sample without microstructuring. The light lost due to imperfections is given by $1 - Q/Q_0$ and the predicted efficiency should then be corrected by the fraction Q/Q_0 such that $\varepsilon = \beta\gamma Q/Q_0$. The dashed curve in Figure 46.10d shows a fit to experimentally measured Q/Q_0 ratios for the pillar geometry described in Gazzano et al. (2013). As the diameter increases, the relative overlap of the mode profile with the imperfections decreases, and Q thus increases with diameter toward the planar value.

While a large pillar diameter optimizes γ and Q/Q_0 toward unity, it also increases the mode volume V leading to a reduced F_P. When choosing the diameter, it is thus necessary to consider the magnitude of β as function of diameter as given by Equation 46.22. The calculated ratio $\beta = F_P/(F_P + 1)$ is displayed in Figure 46.10d as the dotted curve and it decreases with the diameter due to the increase in V. The product $\varepsilon = \beta Q/Q_0$ is shown in the figure as the full curve, and we observe that a maximum of ∼ 0.8 is obtained for $d \sim 2.5$ μm. This estimation of the efficiency assumes a transmission γ of unity, which is a good approximation when using a 0.7 NA lens.

This micropillar SPS reported in Gazzano et al. (2013) indeed features an efficiency of 0.79. The simultaneous spectral and spatial alignment was achieved using an *in situ* fabrication method pioneered by the group.

46.4 Waveguide QED Approach

The second option for optimizing β as defined by Equation 46.1 is to decrease Γ_{Rad} and this represents the waveguide QED approach. Two main geometries taking advantage of this approach are the photonic nanowire presented in Section 46.4.1 and the photonic crystal waveguide discussed in Section 46.4.2.

Assuming that the QD is positioned at a maximum of the guided mode, the spontaneous emission rate to the guided mode in an infinite uniform waveguide is given by (Claudon et al., 2013)

$$\frac{\Gamma_{\mathrm{M}}}{\Gamma_{\mathrm{Bulk}}} = \frac{3}{4\pi A}\left(\frac{\lambda}{n}\right)^2 \frac{n_{\mathrm{g}}}{n}. \tag{46.24}$$

Here, n_{g} is the group index and

$$A = \frac{\int \varepsilon_r(\mathbf{r}_\perp)\left|\mathbf{E}(\mathbf{r}_\perp)\right|^2 \mathrm{d}\mathbf{r}_\perp}{\varepsilon_r(\mathbf{r}_{\max})\left|\mathbf{E}(\mathbf{r}_{\max})\right|^2}, \tag{46.25}$$

where the integration is over the lateral plane normal to the waveguide axis. In the case of the periodic waveguide of periodicity a, the rate is given by

$$\frac{\Gamma_{\mathrm{M}}}{\Gamma_{\mathrm{Bulk}}} = \frac{3a}{4\pi V}\left(\frac{\lambda}{n}\right)^2 \frac{n_{\mathrm{g}}}{n}, \tag{46.26}$$

where the integration (46.21) of the mode volume V is limited to a single periodic cell.

The waveguide QED approach relies on significant suppression of the spontaneous emission into radiation modes, and this rate Γ_{Rad} should be quantified. The suppression occurs for waveguide diameters on the order of λ/n, and Γ_{Rad} can no longer be approximated to its value in a bulk material. Instead, Γ_{Rad} should be computed using full 3-D numerical simulations. Whereas the calculation of Γ_{M} is usually straightforward, a precise computation of Γ_{Rad} is generally quite demanding because the radiation modes propagate away from the device and perfect absorbing boundary conditions are required to avoid artificial reflections from the boundary walls. It is often determined indirectly by first computing the total spontaneous emission rate Γ_{T}, which for most methods is more accessible, and then subtracting Γ_{M} such that $\Gamma_{\mathrm{Rad}} = \Gamma_{\mathrm{T}} - \Gamma_{\mathrm{M}}$.

46.4.1 Photonic Nanowire

The photonic nanowire (Claudon et al., 2013) is a vertical GaAs cylinder with an embedded QD sitting on a substrate. The geometry is illustrated in Figure 46.11a. Light is guided in the nanowire by total internal reflection at the nanowire-air interface. Single photons emitted by the QD are predominantly coupled to the fundamental HE_{11} waveguide mode as will be discussed in the following. The emission is again collected from an objective lens above the nanowire.

Similar to the micropillar SPS, a bottom mirror is required to avoid loss of light into the substrate. Furthermore, the small nanowire diameter leads to a highly divergent far-field emission pattern, and a top taper strategy is required to reduce this divergence. The reader is referred to (Claudon et al., 2013; Gregersen et al., 2013) for more details on the mirror and the taper strategies. In this tutorial, we will simply study spontaneous emission in the infinite nanowire of Figure 46.11b.

The physical origin of the strong suppression of the spontaneous emission into radiation modes can be understood from Figure 46.11c and d. We first consider a radiation mode defined from the illumination of an infinite photonic nanowire by a plane wave $E_x(\mathbf{r}) = E_0 \exp(ik_0 y)$ propagating along the y axis. The corresponding field intensity profile is shown in Figure 46.11c for a nanowire with diameter $d << \lambda$. Outside the nanowire we observe some field enhancement; however, inside the nanowire the field is strongly suppressed. This field suppression is due to a dielectric screening effect (Bleuse et al., 2011) illustrated in Figure 46.11d, which displays the refractive index profile $n(x)$ and the normalized electric field component $E_x(x)/E_0$ in along the $y = 0$ axis. Now, at the nanowire-air interface, continuity is dictated by Maxwell's equations for the component $D_x(x) = n(x)^2 E_x(x)$ of the displacement field normal to the boundary.

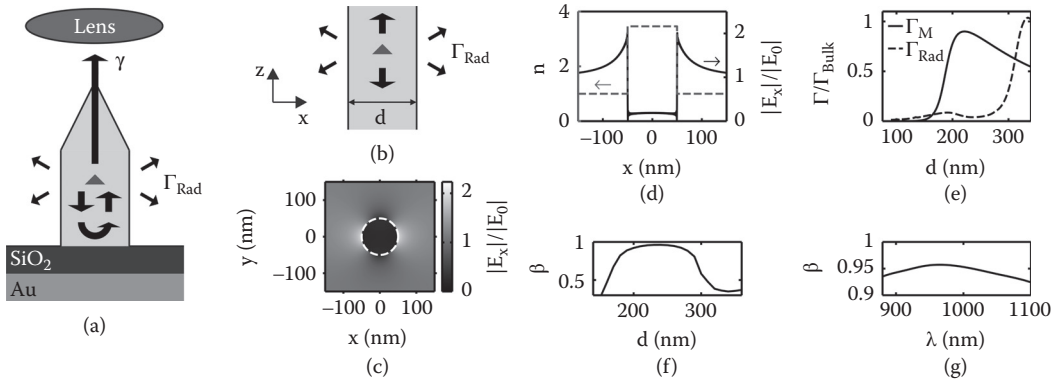


FIGURE 46.11 (a) The full photonic nanowire SPS geometry. (b) The infinite nanowire geometry. (c) The normalized E_x field amplitude of a plane wave propagating along the y axis and scattered by a nanowire of diameter $d = 100$ nm. The nanowire boundary is represented by the dashed white line. (d) Normalized E_x field amplitude (solid) and refractive index profile (dashed) at the position $y = 0$. (e) Spontaneous emission rates into the fundamental waveguide mode (solid) and into radiation modes (dashed) as function of nanowire diameter d. Spontaneous emission β factor (f) as function of d and (g) as function of λ for $d = 240$ nm. Other parameters: $\lambda = 950$ nm and $n_{GaAs} = 3.46$ unless otherwise noted. (Reprinted from Gregersen, N. et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2013 IEEE.)

Since the refractive index of GaAs is large, around 3.5, the E_x component inside the nanowire must be correspondingly suppressed. Indeed, an analytic expression for the spontaneous emission inside the nanowire is available in the electrostatic limit, where $d << \lambda$, and is given by

$$\frac{\Gamma}{\Gamma_{Bulk}} = \frac{\Gamma_{Rad}}{\Gamma_{Bulk}} = \frac{4}{n\left(n^2+1\right)^2}. \tag{46.27}$$

We observe that for GaAs the suppression is on the order of 150. The expression (46.27) is valid for a dipole oriented in the plane normal to the axis, which is the case for InAs QDs. However, for a dipole oriented along the z axis, the relevant boundary condition for the radiation modes at the nanowire-air interface is simply continuity of the E_z component. The mechanism ensuring suppression of the electric field is no longer present, and Γ_{Rad} thus remains substantial even when $d << \lambda$.

The normalized spontaneous emission rates Γ_M and Γ_{Rad} are shown in Figure 46.11e for finite nanowire diameters for an on-axis radially oriented dipole. For $d < 150$ nm, both rates are low. As d increases toward 200 nm, the nanowire becomes large enough to support a well-confined fundamental mode and Γ_M increases toward unity. However, for $d < 300$ nm, Γ_{Rad} remains below 0.1 Γ_{Bulk} due to the dielectric screening effect discussed above. The corresponding β factor is presented in Figure 46.11f, and we observe a maximum β of ~0.95 for a diameter of ~240 nm.

Importantly, this screening effect is not a resonant effect. The β factor as a function of wavelength is presented in Figure 46.11g, and we observe that $\beta > 0.93$ is obtained in a 100-nm range. The waveguide QED is thus broadband, and no spectral alignment between an emitter line and a narrow cavity is required. This represents a significant advantage in the fabrication and a major asset of the waveguide QED approach.

46.4.2 Photonic Crystal Waveguide

For integrated quantum photonics, in-plane emission of the emitted light into a planar waveguide is desired. A platform allowing for the efficient in-plane emission of single photons is the photonic crystal (PhC) membrane with an embedded QD (Yao, Manga Rao, and Hughes, 2010) as illustrated in Figure 46.12a. By

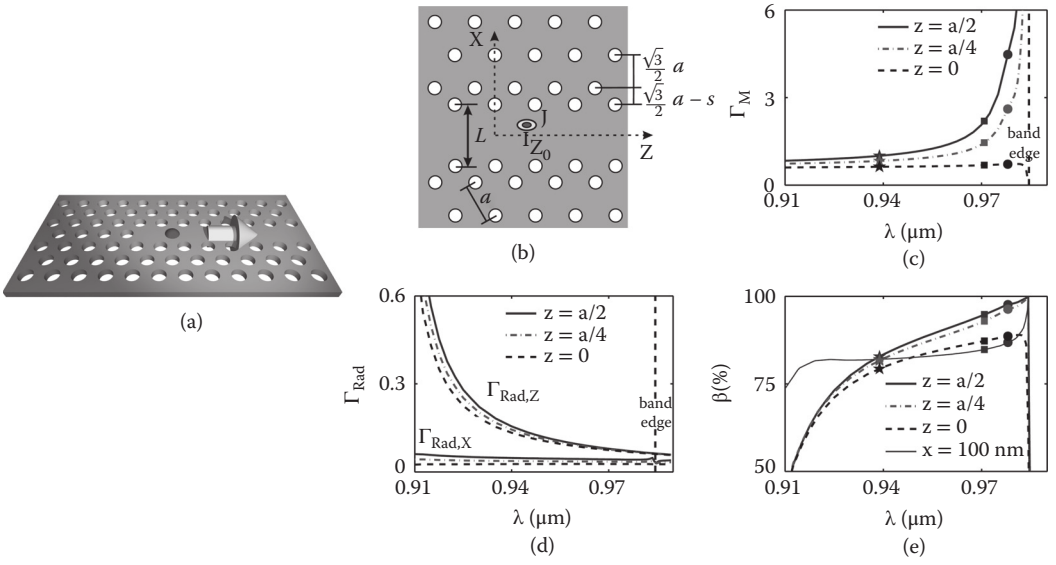


FIGURE 46.12 (a) The PhC waveguide SPS. (b) Geometry of the infinite periodic PhC waveguide. (c) Normalized SE decay rates Γ_M as function of λ. Normalized SE decay rates (d) into the radiation modes for the two in-plane dipole moment orientations X and Z. (e) Associated β factors. The thin solid curve is obtained for an off-axis dipole location ($x = 100$ nm and $z = a/2$). Circles, squares, and stars are reference marks locating the fundamental Bloch-mode group velocities equal to $c/20$, $c/10$, and $c/5$, respectively, where c is the speed of light. (Reprinted with permission from Lecamp, G. et al., *Physical Review Letters*, 99, 023902, 2007. Copyright 2007 by the American Physical Society.)

removing a row of holes, a waveguide can be formed. The waveguide confines light in the plane using the PhC bandgap and vertically using total internal reflection.

In the following, we study the spontaneous emission of the infinite periodic PhC waveguide illustrated in Figure 46.12b. In this tutorial, let us simply consider a dipole positioned in the center of the waveguide $x = 0$ at the position $z = a/2$ with hole displacement $s = 0$. The corresponding results are displayed in Figure 46.12c through 46.12e as full curves. The spontaneous emission rate Γ_M into the fundamental PhC waveguide mode for a dipole oriented along the x axis is shown in Figure 46.12c. For short λ, the rate is similar to that in a bulk material; however, near the band edge Γ_M diverges. This strong enhancement is due to the slow-light effect occurring near the band edge, where the group refractive index diverges (Baba, 2008).

The emission into radiation modes is illustrated in Figure 46.12d for dipoles aligned along the x and the z axes. $\Gamma_{Rad,X}$ is strongly suppressed in the entire wavelength regime due to the same dielectric screening effect discussed for the photonic nanowire. As for the photonic nanowire, this suppression does not take place for the dipole oriented along the z axis and $\Gamma_{Rad,Z}$ is small only near the band edge.

The average spontaneous emission β factor for dipoles oriented along the x and z axes can be written as

$$\beta = \frac{\Gamma_M}{\Gamma_M + \Gamma_{Rad,X} + \Gamma_{Rad,Z}} \tag{46.28}$$

and is depicted in Figure 46.12e. We observe that β approaches unity near the band edge thanks to the suppression of Γ_{Rad} combined with the enhanced Γ_M. The β factor remains larger than 0.75 over a 50 nm wavelength range, an effect predominantly due to the broadband suppression of the spontaneous emission into radiation modes. By displacing the QD x position, the range can be increased as illustrated by the thin solid curve for $x = 100$ nm in Figure 46.12e.

46.5 Planar Dielectric Antenna Approach

A final SPS design discussed in this tutorial is the planar dielectric antenna (PDA) illustrated in Figure 46.13a. This design approach (Chen, Götzinger, and Sandoghdar, 2011) differs conceptually from the previous QED strategies. Whereas the quantum emitter in the QED approaches couples to a specific optical cavity or waveguide mode with the coupling coefficient β, the quantum emitter in the PDA design instead couples to a continuous range of radiation modes and the collection efficiency ε is computed by integrating the contributions from this entire range.

We consider the geometry illustrated in the inset of Figure 46.13b, where a quantum emitter with horizontal electric dipole (HED) or vertical electric dipole (VED) moment orientation is placed in a material of refractive index n_2. In this geometry, light will be collected by an objective lens placed below the emitter.

Without any precautions, light is emitted in all directions due to symmetry, and it is necessary to implement a top mirror to reflect light propagating upward back toward the bottom. While a gold mirror can provide a high broadband reflectivity, coupling to surface plasmons reduces the reflectivity for certain angles of incidence (Chen, Götzinger, and Sandoghdar, 2011). However, the high reflectivity can be restored by introducing an intermediate low-index layer of index $n_3 < n_2$.

With the introduction of the metal mirror, the light is now propagating downward with a highly divergent far-field pattern. To reduce this divergence, a solid immersion lens of index $n_1 > n_2$ is introduced and the downward propagating field is now limited to a cone of polar angle $\theta = \sin^{-1}(n_2/n_1)$. The power density and the collection efficiency as function of the polar angle θ is shown in Figure 46.13b for the two dipole orientations. We observe that a collection efficiency above 0.99 is obtained for an angle $\theta = 55°$. The wavelength dependence on the efficiency is presented in Figure 46.13c, and we observe that $\varepsilon > 0.99$ is maintained over a wavelength range of several hundred nanometers for both emitter orientations. The inset of Figure 46.13c shows the influence of the spacer thickness s on the collection efficiency. As s decreases, the increased coupling of the vertical components of the electric field generated by a VED with surface plasmons reduces the efficiency. However, the field generated by the HED has much weaker vertical components; the coupling to surface plasmons is reduced and the efficiency for this dipole orientation thus remains high.

Whereas the PDA design features record-high broadband efficiency, the condition $n_1 > n_2$ imposes constraints on the choice of material for the medium 1 if the emitter is, e.g., a QD in GaAs with index $n_2 = 3.5$. This can be remedied by introducing an intermediate low-index buffer layer between media 1 and 2. However, an important currently unresolved issue for the PDA design remains, namely the highly non-Gaussian far-field emission pattern illustrated in Figure 46.13b, which makes efficient coupling of the emitted photons to the Gaussian mode profile of a single-mode fiber a challenge.

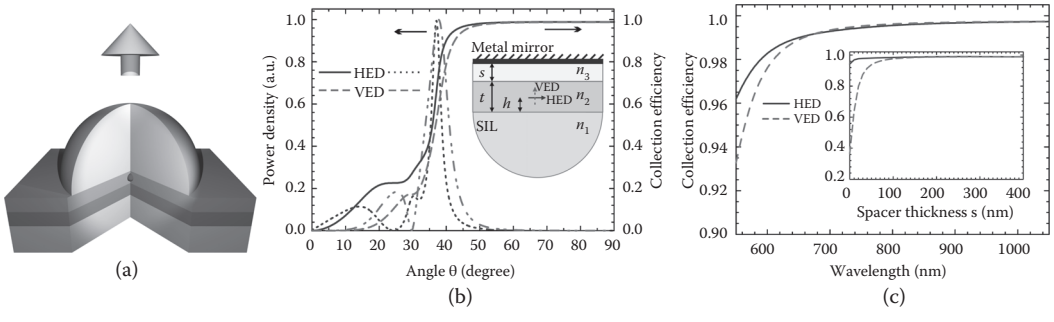


FIGURE 46.13 (a) Planar dielectric antenna. (b) Power density (left axis) and collection efficiency (right axis) versus emission/collection angle θ at a wavelength of 637 nm. The inset shows the schematics of the device with vertically and horizontally oriented dipole emitters. (c) Collection efficiency as a function of wavelength. The inset plots the collection efficiency versus spacer thickness s at 637 nm. (Reprinted from Chen, X,-W. et al., *Optics Letters* 36, 3545–3547, 2011. With permission of Optical Society of America.)

46.6 Decoherence Effects

As discussed above, as well as producing single photons and having a high extraction efficiency, an ideal SPS would produce photons that are quantum mechanically indistinguishable. With reference to Figure 46.2, when two indistinguishable photons are incident on a 50–50 beam splitter they emerge in the same output arm, which provides the fundamental quantum interference necessary for quantum information applications. For photons that are not completely indistinguishable, there remains a nonzero probability that a single photon emerges in both output arms. Based on these arguments, the indistinguishability of a SPS can be quantified by Kaer et al. (2013):

$$I = \frac{\int_0^\infty \int_0^\infty \left| \left\langle \hat{A}^\dagger(t+\tau)\hat{A}(t) \right\rangle \right|^2 dt d\tau}{\int_0^\infty \int_0^\infty \left\langle \hat{A}^\dagger(t)\hat{A}(t) \right\rangle \left\langle \hat{A}^\dagger(t+\tau)\hat{A}(t+\tau) \right\rangle dt d\tau}, \tag{46.29}$$

where $I = 1$ for perfectly indistinguishable photons and $I = 0$ for distinguishable photons. The operators $\hat{A}^\dagger$ and $\hat{A}$ are the creation and annihilation operators for the electric field incident on the beam splitter. Further details of the derivation leading to Equation 46.29 can be found in Unsleber et al. (2015). The challenge in modeling the indistinguishability of photons emitted from a semiconductor-based SPS lies in (a) relating the correlation function and expectation values in Equation 46.29 to operators pertaining to the quantum emitter, and (b) ensuring that the model of the emitter itself captures all important processes that affect the photon indistinguishability.

46.6.1 Quantum Emitter Master Equation

To gain some insight, we first consider a QD emitting in a bulk material or a very low-Q cavity. In this case, provided the emitter is sufficiently far from the beam splitter optics, the field operators in Equation 46.29 can be replaced with operators describing the emitter. We write

$$\hat{A}(t) \sim \hat{\sigma}(t), \tag{46.30}$$

where $\hat{\sigma} = |g\rangle \langle e|$ annihilates an excitation of the emitter (Kiraz et al., 2004). The proportionality constants that we have omitted are independent of time, and with reference to Equation 46.29 we see that they will cancel when the indistinguishability is calculated.

Our task now is to calculate the correlation function in the numerator of Equation 46.29, which can be achieved with help from the quantum regression theorem. To begin, we must first write down a master equation describing the dynamics of the quantum emitter, which takes the form of an equation of motion for the density operator $\hat{\rho}$ (Carmichael, 1999). This can be done to various levels of rigor, but in order to capture the most important qualitative features it suffices to write

$$\frac{d}{dt}\hat{\rho} = \Gamma_{\text{Bulk}} L_{\hat{\sigma}}(\hat{\rho}) + 2\gamma^* L_{\hat{\sigma}^\dagger\hat{\sigma}}(\hat{\rho}), \tag{46.31}$$

where the Lindblad operators are defined to satisfy $L_{\hat{O}}(\hat{\rho}) = \hat{O}\hat{\rho}\hat{O}^\dagger - \left(\hat{O}^\dagger\hat{O}\hat{\rho} + \hat{\rho}\hat{O}^\dagger\hat{O}\right)/2$ and Γ_{Bulk} is the bulk spontaneous emission rate. The first term on the right side of Equation 46.31 describes the decay of the excited state while the second term describes the decoherence of the two-level system. The quantity γ^* is the dephasing or decoherence rate, which we will discuss further in the following sections. Equation 46.31

allows us to establish equations of the motion for expectation values of the emitter. The dipole expectation value, e.g., obeys

$$\frac{d}{dt}\left\langle \hat{\sigma} \right\rangle = -\frac{1}{2}\left(\Gamma_{\text{Bulk}} + 2\gamma^*\right)\left\langle \hat{\sigma} \right\rangle. \tag{46.32}$$

Armed with Equation 46.31, we can now invoke the quantum regression theorem, which states that the correlation function in Equation 46.29 obeys the same equation of motion as the dipole operator in Equation 46.32 but with respect to τ, namely $\frac{d}{d\tau}\left\langle \hat{\sigma}^\dagger(t+\tau)\hat{\sigma}(t) \right\rangle = -\frac{1}{2}\left(\Gamma_{\text{Bulk}} + 2\gamma^*\right)\left\langle \hat{\sigma}^\dagger(t+\tau)\hat{\sigma}(t) \right\rangle$. The initial condition is found by setting $\tau = 0$, from which we find $\left\langle \hat{\sigma}^\dagger(t)\hat{\sigma}(t) \right\rangle = e^{-\Gamma_{\text{Bulk}}t}$ if we assume the emitter is initially in its excited state at time $t = 0$. This allows us to find the correlation function, and inserting this solution into Equation 46.29 we find that the indistinguishability is given by Kaer et al. (2013):

$$I = \frac{\Gamma_{\text{Bulk}}}{\Gamma_{\text{Bulk}} + 2\gamma^*}. \tag{46.33}$$

This expression demonstrates that for our simplistic model, the indistinguishability depends solely on the ratio, $\Gamma_{\text{Bulk}}/\gamma^*$ that we should seek to maximize.

46.6.2 Emitter-Cavity Master Equation

So far we have considered a QD emitting into a bulk medium. However, using similar techniques to those above it is also possible to calculate the indistinguishability of photons emitted from a coupled emitter-cavity system. The calculation is performed by replacing the master equation in Equation 46.31 with the more complicated expression (Kaer et al., 2013)

$$\frac{d}{dt}\hat{\rho} = -i\left[\hat{H}, \hat{\rho}\right] + \Gamma_{\text{Bulk}} L_{\hat{\sigma}}(\hat{\rho}) + 2\gamma^* L_{\hat{\sigma}^\dagger\hat{\sigma}}(\hat{\rho}) + \kappa L_{\hat{a}}(\hat{\rho}). \tag{46.34}$$

Here, $\hat{H} = g\left(\hat{\sigma}\hat{a}^\dagger + \hat{a}\hat{\sigma}^\dagger\right)$ describes the interaction between the emitter and the cavity with g the emitter-cavity coupling strength and $\hat{a}$ the annihilation operator of the cavity mode. The last term in Equation 46.34 describes the outcoupling of light from the cavity with κ the cavity mode decay rate. Broadly speaking, provided the losses in the system are weaker than the emitter-cavity coupling strength, the cavity mode can be adiabatically eliminated from this equation of motion. For details of this procedure, we refer the reader to Kaer et al. (2013). This results in a master equation describing the emitter only, which is equal to Equation 46.31 but with Γ_{Bulk} replaced with $\Gamma = \Gamma_{\text{Bulk}} + \Gamma_{\text{M}}$ with $\Gamma_{\text{M}} = \Gamma_{\text{Bulk}} F_{\text{P}}$ the emission rate into the cavity, and the Purcell factor now given by

$$F_{\text{P}} = \frac{4g^2}{\kappa\,\Gamma_{\text{Bulk}}}. \tag{46.35}$$

We note that the Purcell factor given here is in fact equivalent to that given in Equation 46.20. The apparent difference is a matter of notation, with Equation 46.35 being preferred in the present context since it makes explicit the dependence on the emitter-cavity coupling strength g. With this modification, the indistinguishability measure for a weakly coupled emitter-cavity system becomes

$$I = \frac{\left(F_{\text{P}} + 1\right)\Gamma_{\text{Bulk}}}{\left(F_{\text{P}} + 1\right)\Gamma_{\text{Bulk}} + 2\gamma^*}. \tag{46.36}$$

This expression encompasses the two main strategies used to increase photon indistinguishability. First, as one might expect, eliminating sources of decoherence, and thus reducing γ^*, serves to increase I. Methods to achieve this are briefly discussed in the following section. Second, placing the emitter in a photonic structure such as a micropillar can give rise to a Purcell enhancement of the spontaneous emission rate,

which will also increase the indistinguishability. In either case, the strategy amounts to extracting the single photon on a timescale faster than the decoherence rate. We note, however, that these simple arguments only apply as long as the dephasing rate γ^* and the Purcell factor are independent parameters. As we will see below, dephasing caused by coupling to phonons can increase as the Purcell factor is increased, meaning that increasing photon indistinguishability by increasing the spontaneous emission rate may not always be advisable.

46.6.3 Spectral Wandering

Having introduced the basic theory behind photon indistinguishability, we shall now discuss two main mechanisms of dephasing in semiconductor emitters. We begin with spectral wandering, which refers to the process by which the spectral emission line of a QD moves (wanders) over time (Berthelot et al., 2006). In such a case, successively emitted photons may have different frequencies and therefore cannot be considered indistinguishable. Many processes in semiconductor systems can give rise to spectral wandering, with the amplitude and frequency distribution being dependent on the nature and timescale of the process. The most common process in charge-neutral QDs is spectral diffusion caused by the temporary trapping of charges in nearby sites, which induces a change of the electrostatic environment of the QD.

In order to model spectral diffusion, it is typically sufficient to include a phenomenological dephasing term in the master equation describing the QD degrees of freedom as in Equation 46.31 above. As we have seen, dephasing of this form can be combated by increasing the Purcell effect. Recalling also that the efficiency of a source increases with the Purcell effect, we conclude that in the case of simple phenomenological dephasing, such as that caused by spectral wandering, an increased Purcell factor is always advantageous.

46.6.4 Coupling to Longitudinal Acoustic Phonons

Coupling to phonons, present in any semiconductor sample, represents an additional source of decoherence. It has been demonstrated that the dominant phonon-mediated dephasing mechanism is due to coupling of the QD exciton to longitudinal acoustic phonons via the deformation potential (Ramsay et al., 2010). Phonons deform the QD structure and give rise to fluctuations in the exciton energy, which in turn cause a dephasing of superpositions of the ground and excited state. As such phonons are sensitive to the QD dressed state energy difference, which itself depends on the strength of any QD-cavity coupling g. Therefore, in contrast to spectral wandering, phonons cannot be accurately included by the addition of a simple phenomenological dephasing rate that is independent of other parameters.

For these reasons, correctly incorporating phonon coupling into the description of a QD-based SPS is particularly challenging. One approach is to include the exciton-phonon coupling at a Hamiltonian level, where one essentially takes a large but finite number of harmonic oscillators to model the phonon environment. The dynamics describing the complete system can then be solved numerically, giving results that are exact to within a set numerical tolerance (Kaer et al., 2012, 2013). This method, however, requires a large amount of computing power and can obscure the underlying physical processes. Another less numerically demanding method involves deriving a time-local master equation describing the emitter-cavity system using projection operators. This results in more tractable expression from which the indistinguishability can be calculated, though is valid only to second order in the exciton–phonon coupling strength. A comparison of these methods and a simple phenomenological approach can be found in Kaer et al. (2013).

Here we will adopt a simple model that manages to capture the salient qualitative features while being simple to implement. We replace the constant dephasing rate γ^* in Equation 46.34 with the expression (Ramsay et al., 2010)

$$\gamma_{\mathrm{ph}} = \frac{\pi}{4} J(2g) \coth\left(\frac{g}{k_{\mathrm{B}} T}\right), \tag{46.37}$$

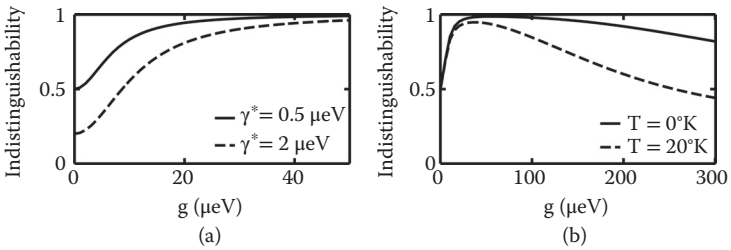


FIGURE 46.14 (a) Photon indistinguishability as a function of emitter-cavity coupling strength for a simple constant dephasing rate of $\gamma^* = 0.5\,\mu eV$ (solid) and $\gamma^* = 2\,\mu eV$ (dashed). (b) A similar plot where the dephasing rate is composed of a constant value of $\gamma^* = 0.5\,\mu eV$ plus a phonon-induced contribution given by Equation 46.37 for temperatures of $T = 0$ K (solid) and $T = 20$ K (dashed). Other parameters are $\Gamma_{Bulk} = 1\,\mu eV$, $\kappa = 100\,\mu eV$, $\eta = 0.032$ ps^2 and $\omega_c = 0.95\,\mu eV$.

where T is the sample temperature and the function $J(\omega)$ is the exciton-phonon spectral density. For typical GaAs QDs, the spectral density takes the form

$$J(\omega) = \eta\omega^3 e^{-\left(\frac{\omega}{\omega_c}\right)^2}, \tag{46.38}$$

where η is a quantity describing the overall strength of the interaction and can be determined from material parameters, while ω_c is a characteristic cutoff frequency that is inversely proportional to the linear dimension of the QD size (Ramsay et al., 2010). Using Equation 46.37 in Equation 46.34 and again invoking the quantum regression theorem, Equation 46.29 can be evaluated to explore the indistinguishability of photons emitted from a coupled QD-cavity system subject to dephasing caused by phonons.

One final consideration is the correct operator to insert into Equation 46.29. Since in the majority of this chapter we are interested in the emission via the cavity mode, if this mode cannot be adiabatically eliminated it is more appropriate to use the cavity operator $\hat{a}$ rather than the emitter operator in Equation 46.29. In doing so, we are explicitly considering the indistinguishability of photons emitted into the cavity mode of interest.

In Figure 46.14a, we show the photon indistinguishability as a function of the emitter-cavity coupling strength g for a constant dephasing rate of $\gamma^* = 0.5\,\mu eV$ (solid) and $\gamma^* = 2\,\mu eV$ (dashed). For $g = 0$, the indistinguishability is given by its value in bulk (Equation 46.33). As g is increased, the coupling of the emitter to the cavity gives rise to an enhancement of the Purcell effect that accelerates the spontaneous emission, thus increasing the indistinguishability. Figure 46.14b shows a similar plot where we now include a constant dephasing rate $\gamma^* = 0.5\,\mu eV$ and also a phonon contribution given by Equation 46.37. For small values of g, we again observe an increase in I due to the Purcell enhancement. However, as g is increased further, the phonon contribution to the dephasing increases, which serves to reduce the indistinguishability. These plots demonstrate the subtle interplay between the beneficial effects of the Purcell enhancement and the detrimental effects of increased coupling to phonons. The reader is referred to (Kaer et al., 2012, 2013; Kaer and Mørk, 2014) for a more extensive discussion of these effects.

46.7 Summary

In this chapter, we have presented the key figures of merit for the single-photon source. For applications in quantum information processing, clean single-photon emission, efficient photon extraction, and indistinguishable photon emission are required. Sources based on spontaneous parametric down-conversion or attenuation are simple to implement but suffer inherently from low efficiency. We have presented the two-level system with focus on the semiconductor quantum dot as an attractive platform for realizing

a deterministic highly efficient single-photon source. Design strategies for controlling the light emission and ensuring a near-unity coupling efficiency based on cavity and waveguide quantum electrodynamics approaches as well as the planar dielectric antenna have been presented. We have discussed the theory and major numerical simulation techniques for modeling the light emission. Finally, we have discussed the influence of several decoherence processes arising from the solid-state environment including simple pure dephasing mechanisms such as charge fluctuations and the more advanced phonon-induced decoherence process due to interaction with longitudinal acoustic phonons.

While, in recent years, tremendous progress has been witnessed towards achieving efficient and pure emission of indistinguishable photons, the simultaneous realization of near-unity efficiency and indistinguishability remains elusive. Fundamental challenges persist both on the design level and on the fabrication level making single-photon sources an interesting research topic most likely for many years to come.

References

Baba, T. 2008. Slow light in photonic crystals. *Nature Photonics* 2 (8): 465–473. doi:10.1038/nphoton.2008.146.

Babinec, T. M., B. J. M. Hausmann, M. Khan, Y. Zhang, J. R. Maze, P. R. Hemmer, and M. Lončar. 2010. A diamond nanowire single-photon source. *Nature Nanotechnology* 5 (3): 195–199. doi:10.1038/nnano.2010.6.

Balanis, C. A. 1989. *Advanced Engineering Electromagnetics*. 1st ed. Hoboken, NJ: Wiley.

Berthelot, A., I. Favero, G. Cassabois, C. Voisin, C. Delalande, P. Roussignol, R. Ferreira, and J-M. Gérard. 2006. Unconventional motional narrowing in the optical spectrum of a semiconductor quantum dot. *Nature Physics* 2 (11): 759–764. doi:10.1038/nphys433. http://arxiv.org/abs/cond-mat/0610346/nhttp://www.nature.com/doifinder/10.1038/nphys433.

Bleuse, J., J. Claudon, M. Creasey, N. S. Malik, J-M. Gérard, I. Maksymov, J-P. Hugonin, and P. Lalanne. 2011. Inhibition, enhancement, and control of spontaneous emission in photonic nanowires. *Physical Review Letters* 106 (10): 103601. doi:10.1103/PhysRevLett.106.103601.

Brendel, J., N. Gisin, W. Tittel, and H. Zbinden. 1999. Pulsed energy-time entangled twin-photon source for quantum communication. *Physical Review Letters* 82 (12): 2594–2597. doi:10.1103/PhysRevLett.82.2594. http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.82.2594.

Buttler, W. T., R. J. Hughes, S. K. Lamoreaux, G. L. Morgan, J. E. Nordholt, and C. G. Peterson. 2000. Daylight quantum key distribution over 1.6 Km. *Physical Review Letters* 84 (24): 5652–5655.

Carmichael, H. J. 1999. *Statistical Methods in Quantum Optics*. 1st ed. Berlin-Heidelberg: Springer.

Chen, X-W., S. Götzinger, and V. Sandoghdar. 2011. 99% efficiency in collecting photons from a single emitter. *Optics Letters* 36 (18): 3545–3547. doi:10.1364/OL.36.003545. http://ol.osa.org/abstract.cfm?URI=ol-36-18-3545.

Claudon, J., J. Bleuse, N. S. Malik, M. Bazin, P. Jaffrennou, N. Gregersen, C. Sauvan, P. Lalanne, and J-M. Gérard. 2010. A highly efficient single-photon source based on a quantum dot in a photonic nanowire. *Nature Photonics* 4 (3): 174–177. doi:10.1038/nphoton.2009.287. http://dx.doi.org/10.1038/nphoton.2009.287.

Claudon, J., N. Gregersen, P. Lalanne, and J-M. Gérard. 2013. Harnessing light with photonic nanowires: Fundamentals and applications to quantum optics. *ChemPhysChem* 14 (11): 2393–2402. doi:10.1002/cphc.201300033.

Coldren, L. A., S. W. Corzine, and M. L. Mashanovitch. 2012. *Diode Lasers and Photonic Integrated Circuits*. 2nd ed. Hoboken, NJ: Wiley.

Ding, X., Y. He, Z.-C. Duan, N. Gregersen, M.-C. Chen, S. Unsleber, S. Maier, et al. 2016. On-demand single photons with high extraction efficiency and near-unity indistinguishability from a resonantly driven quantum dot in a micropillar. *Physical Review Letters* 116 (2): 020401. doi:10.1103/PhysRevLett.116.020401. http://arxiv.org/abs/1601.00284.

Gazzano, O., S. Michaelis de Vasconcellos, C. Arnold, A. Nowak, E. Galopin, I. Sagnes, L. Lanco, A. Lemaître, and P. Senellart. 2013. Bright solid-state sources of indistinguishable single photons. *Nature Communications* 4: 1425. doi:10.1038/ncomms2434. http://www.ncbi.nlm.nih.gov/pubmed/23385570.

Gérard, J.-M. 2003. Solid-state cavity-quantum electrodynamics with self-assembled quantum dots. In *Single Quantum Dots, Topics in Applied Physics*, vol. 90, edited by P. Michler, 269–315. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-39180-77. http://www.springerlink.com/index/4nn94hfbm1gpmyxq.pdf/nhttp://www.springerlink.com/content/4NN94HFBM1GPMYXQ.

Gerry, C. C. and P. L. Knight. 2008. *Introductory Quantum Optics*. New York, NY: Cambridge University Press.

Gregersen, N., P. Kaer, and J. Mørk. 2013. Modeling and design of high-efficiency single-photon sources. *IEEE Journal of Selected Topics in Quantum Electronics* 19 (5): 9000516. doi:10.1109/JSTQE.2013.2255265. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6493378.

Hong, C. K., Z. Y. Ou, and L. Mandel. 1987. Measurement of subpicosecond time intervals between two photons by interference. *Physical Review Letters* 59 (18): 2044–2046. doi:10.1103/PhysRevLett.59.2044. http://journals.aps.org.globalproxy.cvt.dk/prl/abstract/10.1103/PhysRevLett.59.2044.

JCMWave. n.d. JCMSuite. http://www.jcmwave.com.

Kaer, P., N. Gregersen, and J. Mørk. 2013. The role of phonon scattering in the indistinguishability of photons emitted from semiconductor cavity QED systems. *New Journal of Physics* 15: 035027. doi:10.1088/1367-2630/15/3/035027.

Kaer, P. and J. Mørk. 2014. Decoherence in semiconductor cavity QED systems due to phonon couplings. *Physical Review B* 90 (3): 035312. doi:10.1103/PhysRevB.90.035312. http://journals.aps.org/prb/abstract/10.1103/PhysRevB.90.035312.

Kaer, P., T. R. Nielsen, P. Lodahl, A.-P. Jauho, and J. Mørk. 2012. Microscopic theory of phonon-induced effects on semiconductor quantum dot decay dynamics in cavity QED. *Physical Review B* 86 (8): 085302. doi:10.1103/PhysRevB.86.085302.

Kiraz, A., M. Atatüre, and A. Imamoğlu. 2004. Quantum-dot single-photon sources: Prospects for applications in linear optics quantum-information processing. *Physical Review A* 69 (3): 032305. doi:10.1103/PhysRevA.69.032305. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.69.032305.

Kristensen, P. T., C. Van Vlack, and S. Hughes. 2012. Generalized effective mode volume for leaky optical cavities. *Optics Letters* 37 (10): 1649–1651. doi:10.1364/OL.37.001649.

Lavrinenko, A. V., J. Lægsgaard, N. Gregersen, F. Schmidt, and T. Søndergaard. 2014. *Numerical Methods in Photonics*. Boca Raton, FL: CRC Press. http://www.crcpress.com/product/isbn/9781466563889.

Lecamp, G., P. Lalanne, and J. P. Hugonin. 2007. Very large spontaneous-emission β factors in photonic-crystal waveguides. *Physical Review Letters* 99 (2): 023902. doi:10.1103/PhysRevLett.99.023902.

Lorke, M., T. Suhr, N. Gregersen, and J. Mørk. 2013. Theory of nanolaser devices: Rate equation analysis versus microscopic theory. *Physical Review B* 87 (20): 205310. doi:10.1103/PhysRevB.87.205310.

Lumerical. n.d. FDTD Solutions. http://www.lumerical.com/.

Novotny, L. and B. Hecht. 2012. *Principles of Nano-Optics*. 2nd ed. New York, NY: Cambridge University Press.

Purcell, E. M. 1946. Spontaneous emission probabilities at radio frequencies. *Physical Review* 69: 681.

Ramsay, A. J., A. V. Gopal, E. M. Gauger, A. Nazir, B. W. Lovett, A. M. Fox, and M. S. Skolnick. 2010. Damping of exciton rabi rotations by acoustic phonons in optically excited InGaAs/-GaAs quantum dots. *Physical Review Letters* 104 (1): 017402. doi:10.1103/PhysRevLett.104.017402. http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.104.017402.

Reitzenstein, S. and A. Forchel. 2010. Quantum dot micropillars. *Journal of Physics D: Applied Physics* 43 (3): 033001. doi:10.1088/0022-3727/43/3/033001.

Shields, A. J. 2007. Semiconductor quantum light sources. *Nature Photonics* 1 (4): 215–223. doi:10.1038/nphoton.2007.46. http://www.nature.com/nphoton/journal/v1/n4/full/nphoton.2007.46.html.

Siyushev, P., G. Stein, J. Wrachtrup, and I. Gerhardt. 2014. Molecular photons interfaced with alkali atoms. *Nature* 509 (7498): 66–70. doi:10.1038/nature13191. http://www.ncbi.nlm.nih.gov/pubmed/24784217.

Somaschi, N., V. Giesz, L. De Santis, J. C. Loredo, M. P. Almeida, G. Hornecker, S. L. Portalupi, et al. 2016. Near-optimal single-photon sources in the solid state. *Nature Photonics* 10 (5): 340–345. doi:10.1038/nphoton.2016.23. http://dx.doi.org/10.1038/nphoton.2016.23.

Unsleber, S., D. P. S. McCutcheon, M. Dambach, M. Lermer, N. Gregersen, S. Höfling, J. Mørk, C. Schneider, and M. Kamp. 2015. Two-photon interference from a quantum dot microcavity: Persistent pure dephasing and suppression of time jitter. *Physical Review B* 91 (7): 075413. doi:10.1103/PhysRevB.91.075413.

Yao, P., V. S. C. Manga Rao, and S. Hughes. 2010. On-chip single photon sources using planar photonic crystals and single quantum dots. *Laser & Photonics Reviews* 4 (4): 499–516. doi:10.1002/lpor.200810081. http://doi.wiley.com/10.1002/lpor.200810081.

47

Nanoplasmonic Lasers and Spasers

47.1 Introduction 609
47.2 Models for Metals and Gain 610
Drude–Lorentz Model • Multilevel Maxwell–Bloch Theory
47.3 Transfer Matrix Method 613
47.4 Evolutionary Algorithm 616
47.5 Finite-Difference Time-Domain Method 618
47.6 Loss Compensation 620
47.7 Lasing Dynamics 623
47.8 Stopped-Light Lasing 623
47.9 Summary 624

A Freddie Page
and
Ortwin Hess

47.1 Introduction

Semiconductor-based electronics has, in recent years, enjoyed a miniaturization trend down to structure sizes on nanometer scales. However, corresponding photonic devices, such as, in particular, lasers, have not been able to follow suit with a similarly dramatic pace [1]. This is because the generation of photons in small volumes appeared challenging due to the diffraction limit, which apparently seemed to set the minimum wavelength of light to be approximately inversely proportional to its frequency. Considering typical semiconductor band gaps with associated frequencies, this would imply that we would have a lower limit of optical devices and lasers of typically around 1 μm. Yet the availability of nanoscale lasers would be a true starting point for on-chip optical circuits, with applications in telecommunications, data storage, optical computing and sensing. So, can we shrink lasers to be much smaller than the wavelength of the light they emit, particularly down to the nanoscale?

Let us recall that lasers operate on the basis of two features: feedback and gain. In general, a cavity provides feedback by supporting resonant lasing modes (closed loops where light constructively interferes with itself on the round trip). On each loop, the gain medium amplifies the light, and when the amplification is greater than the threshold of energy lost (to dissipation in the material or by escaping the cavity), the system can enter a lasing regime. The physical size of a laser is thus dependent on the cavity supporting lasing modes and this, in turn, is limited by diffraction. Metallic cavities, on the other hand, allow for the circumvention of this diffraction limit by confining light in the subwavelength surface plasmon modes [2,3]. On a metal-insulator interface, light couples to the oscillations of surface charges as *surface plasmon polariton* (SPP) modes. But how does this help? There are two primary features of SPPs that make them useful in nano-optics applications: First, SPPs break the diffraction limit, i.e., they have wavevectors larger than what is generally allowed for a photon of a given frequency. Second, SPPs have or are associated with large field enhancements at the material interface. This combination of shorter wavelengths and higher field densities then significantly enhances the interaction of light with the electronic systems of active media

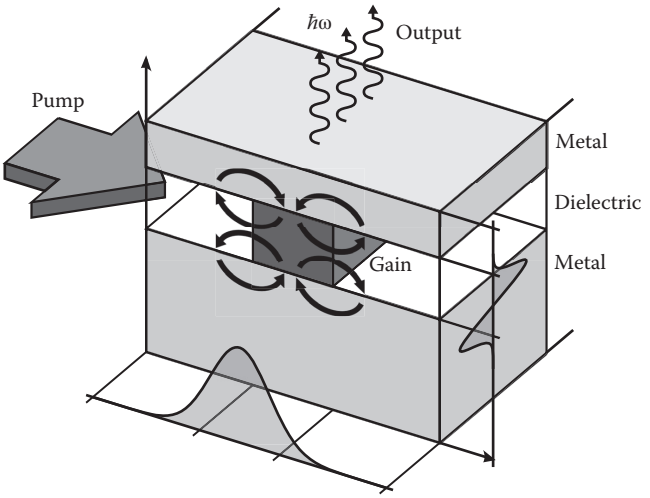


FIGURE 47.1 Nanoplasmonic laser illustration. Light is trapped in a metallic structure such that it forms a closed path (black arrows) that forms in overlap with a gain medium (dark box). In this setup, the gain is pumped from the side and emits from the top. A field strength profile is also shown superimposed on the structure.

[4,5], particularly those interactions that are usually weak. Having subwavelength confinement is, however, frequently traded off to a certain degree with not insignificant dissipative material losses in the metallic components, which requires a corresponding round-trip gain in order to be compensated [6].

Lasers that operate below the diffraction limit on nanometer scales by using metallic cavities are hence called nanoplasmonic lasers. A schematic illustration is given in Figure 47.1. Here there are two options for the light emission: Generated light can be channeled into photonic modes, as in a conventional laser, but can also generate surface plasmons, in which case the laser is more aptly described as a *spaser* for "surface plasmon amplification by stimulated emission of radiation" [7]. Indeed, the possibility to make nanoplasmonic lasers has been explored in various configurations such as, nanorod [8], double-fishnet [9], whispering gallery [10], and stopped light lasers [11].

This chapter will outline a range of techniques that enable the design, optimization, and simulation of nanoplasmonic lasers while tracking the example of a stopped light laser to outline concepts.

47.2 Models for Metals and Gain

47.2.1 Drude–Lorentz Model

In the following, we will chart our theoretical basis for describing the optical response of both metals (which provide subwavelength confinement) and semiconductor gain media (which provides the amplification in a nanoplasmonic laser), involving a combination of a Drude–Lorentz model and a multilevel Maxwell–Bloch theory, respectively, in the harmonic oscillator form. The Drude–Lorentz model describes the oscillations of charges in response to an applied electric field, while a Maxwell–Bloch approach is used to describe the behavior of dipole transitions such as in a four-level system model of a semiconductor.

In its most general terms, the Drude–Lorentz model relates the time evolution of the polarization of a material, **P**, to a driving electric field, **E**, by a second-order differential equation:

$$\frac{\partial^2 \mathbf{P}}{\partial t^2} + \gamma_L \frac{\partial \mathbf{P}}{\partial t} + \omega_L^2 \mathbf{P} = \omega_p^2 \varepsilon_0 \mathbf{E}\,, \tag{47.1}$$

where γ_L is a damping term, ω_L is the restoring force frequency, and ω_p is the plasma frequency—the coupling strength to the driving **E** field. It is a causal response that captures the behavior of most resonance phenomena. In the frequency domain, it is expressed as

$$\tilde{\mathbf{P}} = \varepsilon_0 \frac{\omega_p^2}{\omega_L^2 - \omega(\omega + i\gamma_L)} \tilde{\mathbf{E}} \,, \tag{47.2}$$

where the Drude–Lorentz susceptibility can be extracted:

$$\chi_L(\omega) = \frac{\omega_p^2}{\omega_L^2 - \omega(\omega + i\gamma_L)} \,. \tag{47.3}$$

For $\gamma_L > 0$, this obeys the Kramers–Kronig relations relating real and imaginary parts as required by causality.

The Drude–Lorentz model describes noble metals well, and one may explain the origins of each of the terms with a simple model for a free electron gas. For each electron, let the displacement from its nucleus be given as $\mathbf{x}(t)$. A Newtonian force equation can be built assuming the electron acts as a free particle under the Lorentz force law of an external field **E**, i.e.,

$$m^* \frac{\partial^2 \mathbf{x}}{\partial t^2} = -e\mathbf{E} \,, \tag{47.4}$$

where m^* is the effective mass of an electron and e is its charge. Additional forces can be included, such as a collision term proportional to the velocity and inversely to an average collision time $F_{\text{col}} = -\frac{\partial \mathbf{x}}{\partial t}/\tau$. In this free electron picture, there is no restoring force as for metals, electrons are not bound to their lattice sites. The macroscopic polarization can be related to these electron displacements, as each displacement induces an electronic dipole moment, and these can be averaged to become the polarization by multiplication by the electron number density, $\mathbf{P} = -en\mathbf{x}$. A comparison of terms allows for the associating:

$$\omega_p^2 \to \frac{ne^2}{m^*} \,,\ \gamma \to \tau^{-1} \,,\ \omega_L \to 0 \,. \tag{47.5}$$

This is the Drude model, and with an additional static background permittivity ε_{bg}, it describes metals well.

Real materials typically have multiple resonances, which can be incorporated by adding multiple Drude–Lorentz resonances:

$$\chi = \sum_i \frac{\omega_{p_i}^2}{\omega_{L_i}^2 - \omega(\omega + i\gamma_{L_i})} \,. \tag{47.6}$$

These can have a variety of origins, i.e., in addition to the plasma mode discussed, rotational modes, electronic transitions, etc., and indeed the magnetic response, χ_m, can also be described as a sum of Drude–Lorentz resonances.

47.2.2 Multilevel Maxwell–Bloch Theory

A Drude–Lorentz (harmonic oscillator) type system can also model the electronic transitions between two levels of a semiconductor system, with an energy difference $\hbar\omega_e$. In this picture, there is, however, a field-dependent occupation density of upper and lower levels. Light interacts with this *two-level system* by stimulated emission and absorption processes. When the lower level is more occupied than the upper

level, photons are absorbed and electromagnetic energy is dissipated within the system; however, when the reverse is true, photons can be emitted, stimulated by other photons, adding gain to the system. This dissipative/amplifying response is accompanied by a corresponding refractive response. The strength of the resonance is proportional to the *inversion density* of the two levels, i.e., how much more the higher level is occupied than the lower one, $\Delta N = N_2 - N_1$, such that a layer with emitters embedded may be represented by

$$\varepsilon = \varepsilon_{bg} + \frac{-\Delta N \omega_{pe}^2}{\omega_e^2 - \omega(\omega + i\gamma_e)}, \tag{47.7}$$

$$\omega_{pe}^2 = \sqrt{\varepsilon_{bg}}\gamma_e \sigma_e c, \tag{47.8}$$

where ε_{bg} is the permittivity of the layer hosting the resonance, σ_e is the emission cross-section, and γ_e is the width of the resonance [6]. Note that for negative inversion, the emitter becomes an absorber as there is a higher density of emitters in the lower state. In time, the inversion varies as light is emitted or absorbed. The auxiliary equations to take this into account are

$$\frac{\partial N_1}{\partial t} = -\frac{1}{\hbar\omega_e}\left(\frac{\partial \mathbf{P}_e}{\partial t} + \frac{\gamma_e}{2}\mathbf{P}_e\right) \cdot \mathbf{E} + \frac{N_2}{\tau_{21}}, \tag{47.9}$$

$$\frac{\partial N_2}{\partial t} = +\frac{1}{\hbar\omega_e}\left(\frac{\partial \mathbf{P}_e}{\partial t} + \frac{\gamma_e}{2}\mathbf{P}_e\right) \cdot \mathbf{E} - \frac{N_2}{\tau_{21}}, \tag{47.10}$$

with the first terms representing stimulated emission, and the final ones to account for non-radiative recombination at a rate τ_{21}^{-1}.

The two-level system is appropriate for modeling a single electronic transition; however, there is no way in which a two-level system can reach a state of inversion by relying solely on the processes of spontaneous emission and absorption—the best that can be achieved is equal occupation of the levels. A *four-level system*, on the other hand, can be constructed such that there can be a dynamically maintained population inversion between two of its levels, from where stimulated emission can occur. The construction contains two two-level systems, labeled e and a for emission and absorption; e with energy levels 1 and 2 and energy gap $\hbar\omega_e$, in-between a with energy levels 0 and 3 and greater energy gap $\hbar\omega_a$, as depicted in Figure 47.2. The two upper levels, 2 and 3, are coupled by a fast non-radiative relaxation channel, with rate τ_{32}^{-1}, as are the two lower levels, 0 and 1 (τ_{10}^{-1}). This has the effect of rapidly depleting the first and third levels shortly

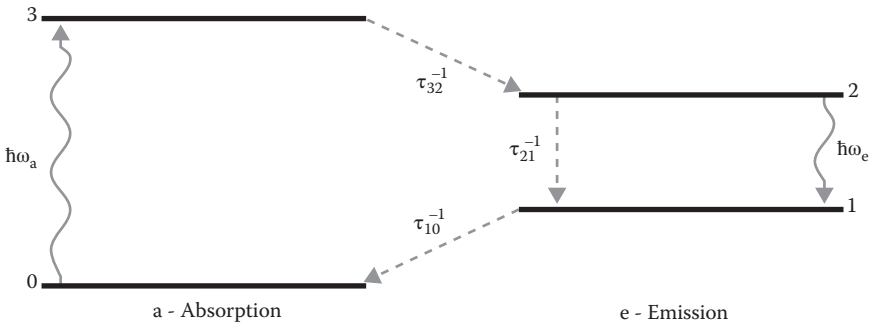


FIGURE 47.2 Schematic of the four-level system. A four-level system comprises two two-level systems, a and e, with active transitions ($0 \leftrightarrow 3$) and ($1 \leftrightarrow 2$), respectively, which are coupled by nonradioactive transition rates τ_{32}^{-1} and τ_{10}^{-1}. In this scheme, the emission subsystem permits radiative transitions at an energy of $\hbar\omega$ and has a slow nonradiative recombination rate τ_{21}^{-1}.

after they become occupied. Between the levels of the emission two-level subsystem, there is additionally a (slow) nonradiative channel (τ_{21}^{-1}).

The key point of a four-level system is that electrons that are pumped from levels 0 and 3, will quickly decay to level 2, leading to an inversion density of level 2 over level 1, which is available for stimulated emission. This allows the optical pumping between levels 0 and 3 to generate inversion between levels 1 and 2.

Four-level systems are incorporated into time-domain simulations using the differential equation for the polarization given in Equation 47.1 for each transition:

$$\frac{\partial^2 \mathbf{P}_\mathrm{a}}{\partial t^2} + \gamma_\mathrm{a} \frac{\partial \mathbf{P}_\mathrm{a}}{\partial t} + \omega_\mathrm{a}^2 \mathbf{P}_\mathrm{a} = -\Delta N \omega_\mathrm{pa}^2 \varepsilon_0 \mathbf{E}, \tag{47.11}$$

$$\frac{\partial^2 \mathbf{P}_\mathrm{e}}{\partial t^2} + \gamma_\mathrm{e} \frac{\partial \mathbf{P}_\mathrm{e}}{\partial t} + \omega_\mathrm{e}^2 \mathbf{P}_\mathrm{e} = -\Delta N \omega_\mathrm{pe}^2 \varepsilon_0 \mathbf{E}\,. \tag{47.12}$$

Here the polarization has been split off into a two parts $\mathbf{P}_\mathrm{a}$ and $\mathbf{P}_\mathrm{e}$ that are connected with the radiative resonances, which are added to the total polarization. The corresponding level occupation densities update with the auxiliary equations [6]

$$\frac{\partial N_3}{\partial t} = \left(\frac{\partial \mathbf{P}_\mathrm{a}}{\partial t} + \frac{\gamma_\mathrm{a}}{2} \mathbf{P}_\mathrm{a} \right) \cdot \mathbf{E} - \frac{N_3}{\tau_{32}}, \tag{47.13}$$

$$\frac{\partial N_2}{\partial t} = \frac{N_3}{\tau_{32}} + \frac{1}{\hbar \omega_\mathrm{e}} \left(\frac{\partial \mathbf{P}_\mathrm{e}}{\partial t} + \frac{\gamma_\mathrm{e}}{2} \mathbf{P}_\mathrm{e} \right) \cdot \mathbf{E} - \frac{N_2}{\tau_{21}}, \tag{47.14}$$

$$\frac{\partial N_1}{\partial t} = \frac{N_2}{\tau_{21}} - \frac{1}{\hbar \omega_\mathrm{e}} \left(\frac{\partial \mathbf{P}_\mathrm{e}}{\partial t} + \frac{\gamma_\mathrm{e}}{2} \mathbf{P}_\mathrm{e} \right) \cdot \mathbf{E} - \frac{N_1}{\tau_{10}}, \tag{47.15}$$

$$\frac{\partial N_0}{\partial t} = \frac{N_1}{\tau_{10}} - \left(\frac{\partial \mathbf{P}_\mathrm{a}}{\partial t} + \frac{\gamma_\mathrm{a}}{2} \mathbf{P}_\mathrm{a} \right) \cdot \mathbf{E}\,, \tag{47.16}$$

which encode both the stimulated terms and nonradiative recombinations.

47.3 Transfer Matrix Method

The geometry of the perhaps simplest types of nanoplasmonic lasers is planar waveguide structures. Here, planar layers of heterogeneous materials are stacked to form a waveguide in the center. The modes of such structures depend strongly on the width and material of each layer and can be calculated using the transfer matrix method (TMM), a linear, frequency-domain analysis that shall be used to design and optimize structures before simulation in time domain.

Isotropic homogeneous materials support transverse plane wave solutions in bulk. These waves have the dispersion relation

$$\varepsilon \mu \frac{\omega^2}{c^2} - k^2 = 0\,, \tag{47.17}$$

and the relative amplitudes of the electric and magnetic fields set by

$$E = \sqrt{\frac{\mu}{\varepsilon}} Z_0 H\,, \tag{47.18}$$

where Z_0 is the impedance of free space. When two materials meet at a planar interface, the solutions in each space must be matched together, and the prescription here is that the E and H field components parallel to the interface are continuous. Each interface couples a forward traveling and a backward traveling wave in each layer together.

Let us consider a stack with an axis parallel to the z-direction and waves traveling in a plane perpendicular to the y-direction. There are two independent polarization solutions, *transverse magnetic* (TM) where H is perpendicular to the plane of incidence (parallel to y) with E parallel to the plane, and *transverse electric* (TE) where E is perpendicular to the plane of incidence. TM equations shall be considered herein as this is the character of the SPP solutions. The corresponding TE equations can be recovered by making the substitutions: $\varepsilon \to \mu$, $\mu \to \varepsilon$, $\mathbf{E} \to -Z_0\mathbf{H}$, $Z_0\mathbf{H} \to \mathbf{E}$.

The system can be solved by performing a change of basis between forward and backward propagating waves and electric and magnetic field components, i.e.,

$$\begin{pmatrix} E_{xi} \\ Z_0 H_{yi} \end{pmatrix} = \underbrace{\begin{pmatrix} u_i & -u_i \\ 1 & 1 \end{pmatrix}}_{\mathbf{U}} \begin{pmatrix} A_i^+ \\ A_i^- \end{pmatrix}, \tag{47.19}$$

and translating the forward and backward field components within a layer:

$$\left.\begin{pmatrix} A_i^+ \\ A_i^- \end{pmatrix}\right|_z = \underbrace{\begin{pmatrix} e^{ik_{zi}(z-z')} & 0 \\ 0 & e^{-ik_{zi}(z-z')} \end{pmatrix}}_{\mathbf{T}} \left.\begin{pmatrix} A_i^+ \\ A_i^- \end{pmatrix}\right|_{z'}, \tag{47.20}$$

with $u_i = ck_{zi}/(\varepsilon_{xi}\omega)$. These are the matrices of the TMM. The matrices, $\mathbf{U}$ and $\mathbf{T}$, are assigned such that $\mathbf{U}$ converts from the basis of forward and backward traveling waves to electric and magnetic field components, and $\mathbf{T}$ propagates forward and backward traveling waves from a point z' to a point z.

For a stack of N planar slabs on a substrate as in Figure 47.3, each with its own material properties, we demonstrate how the TMM can determine the bound modes. Let the top interface of the ith slab be at z_i, vector components with that label, e.g., $A_i^{\pm}$, be measured from there, i.e., just below the top interface in that

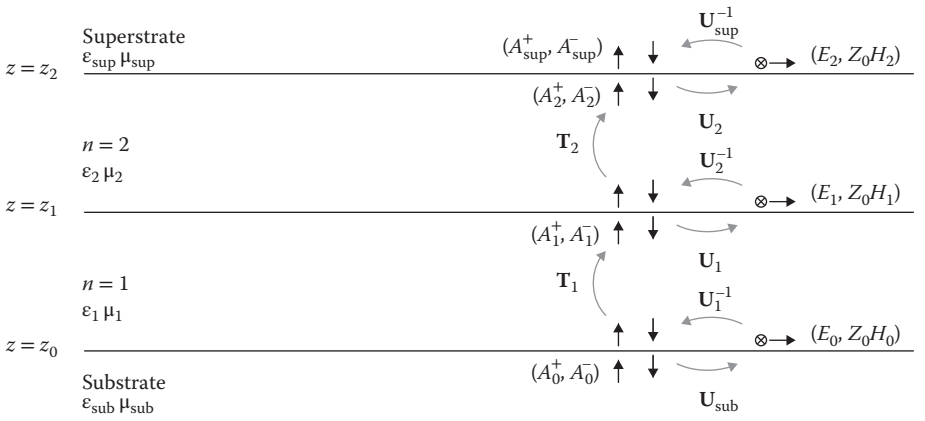


FIGURE 47.3 Schematic of the transfer matrix method. Fields (black arrows) are operated on by matrix transformations (gray arrows). The forward and backward components at the top of the substrate layer (A_0^+, A_0^-) are transferred through the stack to forward and backward components at the bottom of the superstrate layer (A_{sup}^+, A_{sup}^-). Fields move between bases of forward and backward waves ($\uparrow\downarrow$) and electric and magnetic components ($\otimes\rightarrow$).

layer. Starting in the forward and backward basis, one may use the transfer matrices to propagate through the stack from the one layer to the next:

$$\begin{pmatrix} A_{i+1}^{+} \\ A_{i+1}^{-} \end{pmatrix} = \mathbf{T}_{i+1}\mathbf{U}_{i+1}^{-1}\mathbf{U}_i \begin{pmatrix} A_i^{+} \\ A_i^{-} \end{pmatrix} . \tag{47.21}$$

The exception to this being the superstrate, the $(N+1)$th layer, which by definition does not have a top interface. As such, $E_{\text{sup}}^{\pm}$ is defined specially as the forward and backward components just above the top interface of the Nth layer, i.e.,

$$\begin{pmatrix} A_{\text{sup}}^{+} \\ A_{\text{sup}}^{-} \end{pmatrix} = \mathbf{U}_{\text{sup}}^{-1}\mathbf{U}_N \begin{pmatrix} A_N^{+} \\ A_N^{-} \end{pmatrix} . \tag{47.22}$$

The full transfer matrix $\mathcal{M}$ will map forward and backward traveling waves in the substrate to the superstrate:

$$\begin{pmatrix} A_{\text{sup}}^{+} \\ A_{\text{sup}}^{-} \end{pmatrix} = \mathcal{M} \begin{pmatrix} A_{\text{sub}}^{+} \\ A_{\text{sub}}^{-} \end{pmatrix} , \tag{47.23}$$

and is given as the product of each layer hop:

$$\mathcal{M} = \mathbf{U}_{\text{sup}}^{-1}\mathbf{U}_N \cdots \mathbf{T}_1\mathbf{U}_1^{-1}\mathbf{U}_{\text{sub}} . \tag{47.24}$$

Terms with similar indices can be grouped to give a matrix for each layer:

$$\mathbf{M}_i = \mathbf{U}_i\mathbf{T}_i\mathbf{U}_i^{-1} , \tag{47.25}$$

which maps electric and magnetic fields at the bottom of the layer through to the top. As such, the transfer matrix for the whole stack is

$$\mathcal{M} = \mathbf{U}_{\text{sup}}^{-1}\left(\prod_{i=N}^{1}\mathbf{M}_i\right)\mathbf{U}_{\text{sub}} , \tag{47.26}$$

noting the reverse order on the product.

The eigenmodes of the system that have non-zero fields for a zero driving field, i.e.,

$$\begin{pmatrix} A_{\text{sup}}^{+} \\ 0 \end{pmatrix} = \mathcal{M} \begin{pmatrix} 0 \\ A_{\text{sub}}^{-} \end{pmatrix} , \tag{47.27}$$

with the prescription that $\operatorname{Im} k_z > 0$ in both the superstrate and substrate such that fields on both sides decay away to infinity. This yields the condition that $\mathcal{M}_{22} = 0$ for bound modes. For dispersive media, this is satisfied for either one of the frequency ω or wavevector q being a complex number, with the other remaining real. In general, analytical dispersion relation solutions do not exist, but rather solutions for $\mathcal{M}_{22}(\mathbf{q}, \omega) = 0$ are found implicitly, using a complex root-finding algorithm.

To illustrate the method, let's consider the SPP solution. Consider a metal substrate with a planar interface to a dielectric layer. In this case, the transfer matrix is given as

$$\mathcal{M} = \mathbf{U}_{\text{sup}}^{-1}\mathbf{U}_{\text{sub}} \tag{47.28}$$

$$= \frac{1}{2}\begin{pmatrix} \frac{\varepsilon_{\text{bg}}\omega}{ck_{z,\text{air}}} & 1 \\ -\frac{\varepsilon_{\text{bg}}\omega}{ck_{z,\text{air}}} & 1 \end{pmatrix}\begin{pmatrix} \frac{ck_{z,\text{metal}}}{\varepsilon(\omega)\omega} & -\frac{ck_{z,\text{air}}}{\varepsilon(\omega)\omega} \\ 1 & 1 \end{pmatrix}, \tag{47.29}$$

which gives the SPP condition:

$$\mathcal{M}_{22} = \frac{1}{2}\left(\frac{\varepsilon_{\text{bg}}\omega}{ck_{z,\text{air}}}\frac{ck_{z,\text{metal}}}{\varepsilon(\omega)\omega} + 1\right) = 0, \tag{47.30}$$

$$\frac{\varepsilon_{\text{bg}}}{\varepsilon(\omega)} + \frac{k_{z,\text{air}}}{k_{z,\text{metal}}} = 0\,. \tag{47.31}$$

One can see that for this equation to have solutions, given $\varepsilon_{\text{bg}} > 0$ and $\operatorname{Im} k_z > 0$, it must hold that $\operatorname{Re}\varepsilon(\omega) < 0$, which is characteristic of metals for frequencies below their plasma frequency.

For a lossless Drude metal, i.e., $\varepsilon(\omega) = \varepsilon_\infty - (\omega_{\text{p}}/\omega)^2$, an analytical solution exists:

$$q^2 = \varepsilon_{\text{bg}}\left(\frac{\omega}{c}\right)^2 \frac{\varepsilon_\infty\omega^2 - (\varepsilon_{\text{bg}} + \varepsilon_\infty)\omega_{\text{sp}}^2}{(\varepsilon_{\text{bg}} + \varepsilon_\infty)(\omega^2 - \omega_{\text{sp}}^2)}\,, \tag{47.32}$$

where $\omega_{\text{sp}} = \omega_{\text{p}}/\sqrt{\varepsilon_{\text{bg}} + \varepsilon_\infty}$ is the surface plasmon frequency.

47.4 Evolutionary Algorithm

For a planar stratified structure whose properties can be determined by a TMM, one may wish to optimize some aspect of its response to electromagnetic radiation. This may be in the dispersion relation of bound modes or in the transmission and reflection spectra. For example, to prescribe the frequency of a bound mode, or a reflection peak; or set whether there is only a single mode or many modes within a frequency range; or control the amplitude and phase response to an incoming wavepacket. These properties are entirely determined by the composition of the structure; however, it is often not obvious how changes in the structure lead to changes in the response.

The structures to be considered can be described by a finite number of numeric or discrete parameters, i.e., for each layer: the thickness of the layer, the material model to use, and the parameters of that model. This is well suited to optimization using an *evolutionary algorithm* (EA) [12].

In brief, an EA operates by storing multiple copies of these parameter sets, perturbing the values of the parameters randomly, ordering the sets by how the system they describe performs against a fitness metric, and keeping only the best variants to the next round.

More fully, in the language of an EA, this would be to say that a *population* of *chromosomes* is kept, and on each iteration of the algorithm, new chromosomes are generated from *mutations* of the current ones; the next-generation population is filled by *selection* of these chromosomes by a *fitness function*. The goal of the EA is to find a chromosome that extremizes the fitness function globally, and as such an EA is most suitable when the evaluation of the fitness function can be performed relatively quickly as a large number of members of the state space must be evaluated and compared.

A chromosome is simply the set of parameters that describes a system, i.e., a list of numbers or discrete option values. They are atomic entities, independent of other chromosomes. Each chromosome may have its fitness function evaluated to determine how suitable its corresponding physical system is to a problem.

Each iteration of the EA holds a set of chromosomes, ordered by their fitness function, known as a *population*. The first-generation population is either seeded with randomly generated structures, or structures that have been tuned by hand, perhaps by approximative theoretical methods. Populations of subsequent generations are constructed by first taking a small quantile of the fittest chromosomes from the previous iteration, with the remaining places filled with *mutations* of chromosomes selected from the previous set. The progress of the algorithm can be tracked by examining the fittest chromosome at the end of each iteration. At the end of each iteration, a *terminating condition* is evaluated, which determines if the algorithm should stop, returning the fittest chromosome, e.g., if the best structure in the population is within a defined tolerance of the target output, or if a set number of iterations have elapsed. If the terminating condition is not met, then the algorithm continues for another iteration. The operation of the EA is described as a flowchart in Figure 47.4.

The mutation process generates a new chromosome by randomly perturbing the values of parameters of a parent chromosome. It is by mutation that the state space of chromosomes is mapped.

For the layered structures considered here, the mutation process will perturb both the properties of the layers themselves and the relationship between layers. Numeric parameters, such as layer thicknesses and those used in material models, are perturbed by a Gaussian random variable, and the discrete parameters, such as the layer model, can be randomly selected from a list. Typically, only one or few parameters will be changed in each mutation. The parameters of the layers themselves do not sit in isolation. The number of layers in these structures is variable and mutation allows for adding or removing of layers or indeed swapping the order of layers and shifting the boundary between adjacent layers.

Whereas the EA is widely applicable to problems in nano-optics, the fitness function is not general and has to be constructed for each problem based on the type and kind of characteristics that are to be selected for. As an example, the EA can be applied to the case of stopped-light lasing. Here structures are desired that support a bound mode with two points of zero group velocity (ZGV) in their dispersion. An SL structure is deemed better if the ratio of the separations of the wavevectors and frequencies of the ZGV points, the band index $|c/v_b| = (k_2 - k_1)/(\omega_2 - \omega_1)$, is higher. This ensures that a wide range of wavevectors can be held at a single frequency, which reduces the group velocity for the whole band and the wave dispersion that will occur for a wavepacket made from Fourier components in the band. Figure 47.5 shows an example of

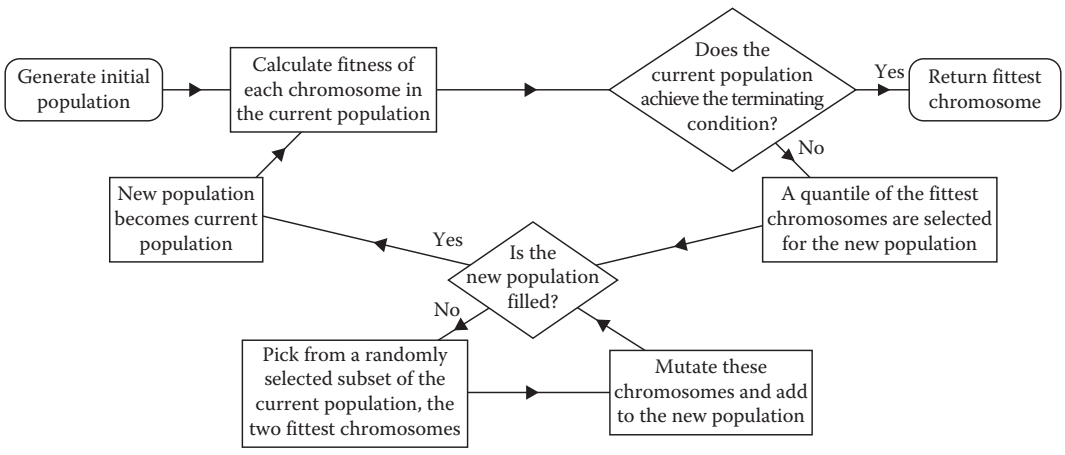


FIGURE 47.4 Evolutionary algorithm flowchart.

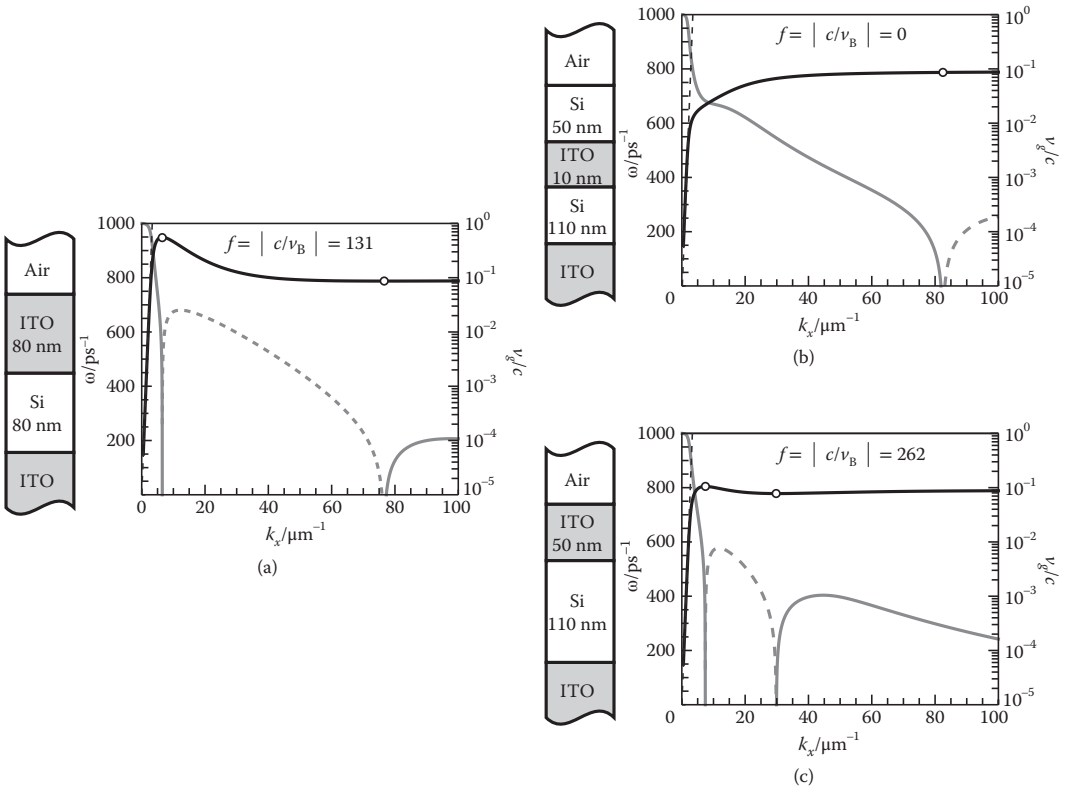


FIGURE 47.5 Diagrams of planar plasmonic nanostructures and corresponding dispersion relations (black solid curve). The group velocity is also shown (gray solid, positive; gray dashed, negative) as well as zero group velocity points (circles). An evolutionary algorithm, using *band index* as a metric, given a. as a starting chromosome may produce b. and c. as child chromosomes. b. has a metric of zero and is thrown out, whereas c. may stay to the next generation as it has a higher metric than its parent.

potential SL structures, their dispersion relations, and group velocity. Each structure is labeled by its band index, and for a parent structure, two possible child structures are presented.

47.5 Finite-Difference Time-Domain Method

When describing processes such as gain, a frequency-domain analysis can only bring insight to a certain level as it has limitations to describe nonlinearities. When plasmons or phonons are emitted, electrons are demoted from higher energetic states to lower ones. This depletes the population inversion, reducing the available gain over time, leading to a nonlinear dependence in the field strength. In addition, spatially resolved effects, such as spatial hole burning, cannot be considered in TMM that assumes uniformity in the direction of propagation, whereas the level of inversion can vary both in this direction and perpendicular to the stacking. Depending on the mode formed, some regions may host higher field densities that can deplete the local gain. Thus a dynamic, spatially resolved, time-domain simulation is required to capture all the aspects of a nanoplasmonic lasing device. Here we will discuss and use the *finite-difference time-domain* (FDTD) method [13,14].

In FDTD, the goal is to solve Maxwell's equations for successive instants in time on a discretized lattice. This is approached by focusing on the curl equations and making the time derivatives of the fields the

subject. Let us assume that there are no external sources:

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_0}\left(\nabla \times \mathbf{H} - \frac{\partial \mathbf{P}}{\partial t}\right) \tag{47.33}$$

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu_0}\left(\nabla \times \mathbf{E} - \frac{\partial \mathbf{M}}{\partial t}\right), \tag{47.34}$$

where **P** and **M** are the polarization and magnetization, respectively. Now, space and time are discretized as $(t, x, y, z) \to (n\Delta_t, i\Delta_x, j\Delta_x, k\Delta_x)$ and derivatives are turned into finite differences, i.e.,

$$\frac{\partial \mathbf{E}}{\partial t}(\mathbf{x}, t) \to \frac{\mathbf{E}^{n+1}_{i,j,k} - \mathbf{E}^{n}_{i,j,k}}{\Delta_t}. \tag{47.35}$$

Strictly, the above is a central difference and approximates the time derivative at a time $(n+1/2)\Delta_t$. Therefore, the corresponding right-hand side terms must be evaluated at this time. The curl operator is handled similarly:

$$\nabla \times \mathbf{E}(\mathbf{x}, t)|_x \to \frac{E_z|^n_{i,j+1,k} - E_z|^n_{i,j,k}}{\Delta_x} - \frac{E_y|^n_{i,j,k+1} - E_y|^n_{i,j,k}}{\Delta_x}. \tag{47.36}$$

In order to make best use of this central difference, the lattice is arranged as a Yee grid [13], where fields are stored separately by component and **E** and **P** fields are specified on integer time coordinates, whereas **H** and **M** are specified on half-integer time coordinates. The lattice coordinates are more complicated; for **E**, components are stored at half-integer coordinates in their own direction and at integer coordinates otherwise. This is the opposite for the **H** field, as illustrated in Figure 47.6. From here, one can write the *update equations*:

$$\begin{aligned} E_x|^{n+1}_{i+\frac{1}{2},j,k} &= E_x|^{n}_{i+\frac{1}{2},j,k} - \varepsilon_0^{-1} P_x|^{n+1}_{i+\frac{1}{2},j,k} + \varepsilon_0^{-1} P_x|^{n}_{i+\frac{1}{2},j,k} \\ &+ \frac{c\Delta_t}{\Delta_x}\left(Z_0 H_z|^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2},k} - Z_0 H_z|^{n+\frac{1}{2}}_{i+\frac{1}{2},j-\frac{1}{2},k}\right) \\ &- \frac{c\Delta_t}{\Delta_x}\left(Z_0 H_y|^{n+\frac{1}{2}}_{i+\frac{1}{2},j,k+\frac{1}{2}} - Z_0 H_y|^{n+\frac{1}{2}}_{i+\frac{1}{2},j,k-\frac{1}{2}}\right) \end{aligned} \tag{47.37}$$

$$\begin{aligned} Z_0 H_x|^{n+\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} &= Z_0 H_x|^{n-\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} + cM_x|^{n+\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} - cM_x|^{n-\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} \\ &+ \frac{c\Delta_t}{\Delta_x}\left(E_z|^{n}_{i,j+1,k+\frac{1}{2}} - E_z|^{n}_{i,j,k+\frac{1}{2}}\right) \\ &- \frac{c\Delta_t}{\Delta_x}\left(E_y|^{n}_{i,j+\frac{1}{2},k+1} - E_y|^{n}_{i,j+\frac{1}{2},k}\right), \end{aligned} \tag{47.38}$$

with corresponding equations for the omitted components.

With the updated equations of the primary fields determined, it now remains to be described how the auxiliary fields are updated. In general, the constitutive equations are functionals of all past values of all fields, though this would prove tricky to implement directly. In practice, material responses can be cast in terms of a finite difference if they can be expressed in terms of a temporal differential equation. This lends the Drude–Lorentz model well to inclusion within FDTD by converting its auxiliary equations into finite differences.

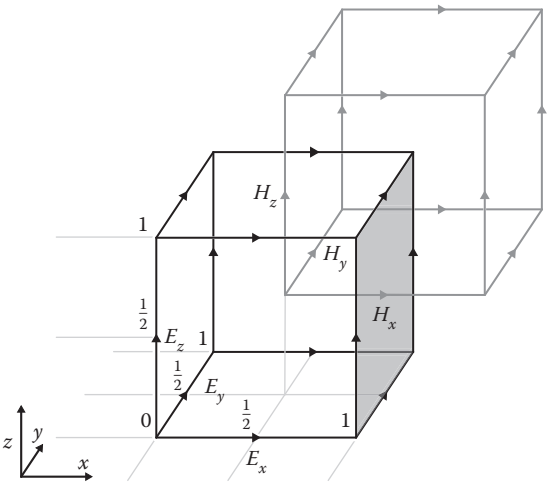


FIGURE 47.6 Schematic of FDTD Yee cell. Electric (black) and magnetic (gray) field grids. They are shifted half a step from each other, forming an interlocking link. Each field line passes through the center of a square formed from lines of the other field, e.g., H_x here passes through the center of the highlighted *E*-field face, where the finite difference $\nabla \times \mathbf{E}$ is defined.

The FDTD algorithm proceeds by iterating over, and updating in order, the **P** field, **E** field, **M** field, and **H** field over all space for each timestep.

Of course, there are limitations to FDTD. Its staggered grid structure means that material interfaces are not sharply defined, and any curvature on an interface will also experience a staircasing effect that can lead to spurious hotspots especially in plasmonic structures where geometric effects play a key role. The grid is also regular, which is to say it cannot be adaptively refined in areas of interest, rather all points must be simulated to the same resolution. This adds to the numerical load, as the smallest wavelength component of the fields, as well as the smallest geometrical features must be provided with sufficient resolution. This is again especially relevant in plasmonics, where field confinement is a feature.

Because of its nature as a full-wave time-domain solver, FDTD can often be the final say on a theoretical investigation. Short of doing a real-world experiment, FDTD numerical simulations find their utility as a heavy-duty tool. That is, the Maxwell's equations are solved at every discretized point in space, so it is often slower and more cumbersome, but can as a result of the inherent thoroughness produce results for questions that other techniques are unable to.

47.6 Loss Compensation

The high field enhancements afforded by plasmon modes supported by negative permittivity in metals must be traded off against Ohmic losses in the optical frequencies. Loss compensation is the incorporation of gain elements into a system in order to offset the loss in a particular frequency range while retaining desired properties from the lossy structure.

The properties one might wish to retain, e.g., negative effective index [6], stopped light points [11], or hyperbolic dispersion [15], etc., are frequency dependent. Causality will influence the position and width of the frequency range where losses can be compensated and desirable properties are to be found simultaneously [16,17], and adding gain to a system will necessarily change the refractive properties in accordance with the Kramers–Kronig relations.

The induced change in the mode structure by the addition of gain can be analyzed in a frequency-domain analysis, modeling gain as a four-level system in frequency domain with constant electron occupation

densities. Using the TMM, this would mean replacing a part or whole of a dielectric layer with an equivalent Drude–Lorentz model with properties that map to the dielectric layer for equilibrium carrier occupation densities.

To illustrate this, the SL structure as optimized previously is modified by replacing the dielectric layer with a quantum emitter. The emission frequency is set to match either one of the ZGV points, and the other parameters are set representing the inclusion of realistic laser dye molecules [18]. The resulting dispersion and loss relations are shown in Figure 47.7 for inversion densities ΔN varying between 0 and N, with a fixed emitter density, and excitation about ZGV1 and ZGV2. The first point to note is that, even on full inversion, the addition of gain does not significantly change the dispersion, with the maximum shift in frequency being around 0.6%. The presence of ZGV points is preserved, though they may drift slightly, i.e., ZGV2 moves slightly right with emission about ZGV1. There is no change at the frequency that is being excited because the permittivity change of a Lorentzian is zero at the resonant frequency. Adding gain has the side effect of making the structure a slightly better SL structure as the band velocity is marginally reduced.

The key change though is in the loss. As the inversion density increases, the loss decreases for wavevectors where the dispersion is within the gain width. Initially the plasmonic loss is reduced as the inversion increases, and then for an inversion of around $\Delta N/N \approx 0.4$, some wavevectors become undamped, and even eventually experience gain. Plasmons sitting in these modes will grow exponentially in amplitude while small enough to remain in the small-signal gain regime. As ZGV2 is flatter than ZGV1, when the

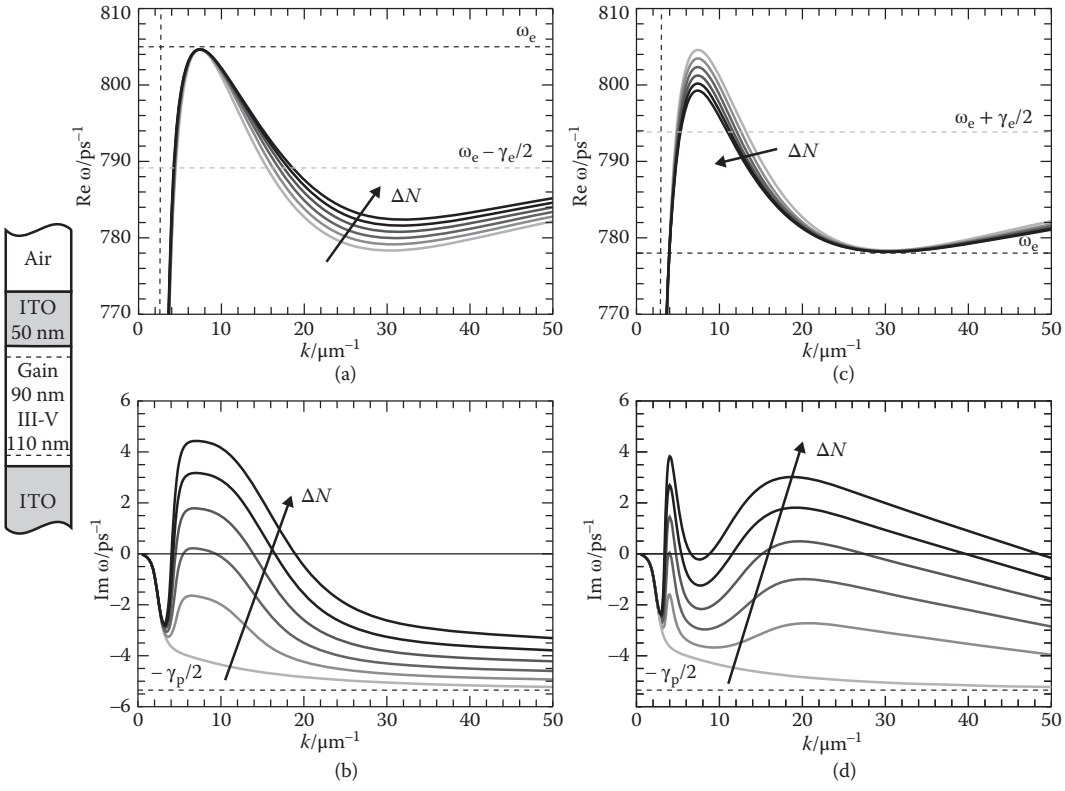


FIGURE 47.7 Perturbation of dispersion relation and loss with Lorentzian gain. The change of mode shape plotted as Lorentzian gain is introduced to the system. For subfigures on the left, emission is about ZGV1, and on the right about ZGV2. a and b show a zoom in of how the dispersion relation changes with increasing inversion density, while c and d show the corresponding loss/gain of the mode.

emission is about this point, a wider range of frequencies fall within the gain width, leading to a larger range of k values that can become undamped.

An analysis of loss compensation can also be performed in the time domain. This has been done in Reference [6] where the structure was a negative refractive index double-fishnet metamaterial. Pump-probe spectroscopy was performed in FDTD in order to retrieve effective parameters using realistic parameters of rhodamine 800 dye molecules as a four-level system. It was shown that the imaginary part of the effective refractive index, which encodes the loss, transitioned from positive to zero as pumping increased while retaining a negative real part of the effective index between wavelengths of 706 and 716 nm. The results are shown in Figure 47.8.

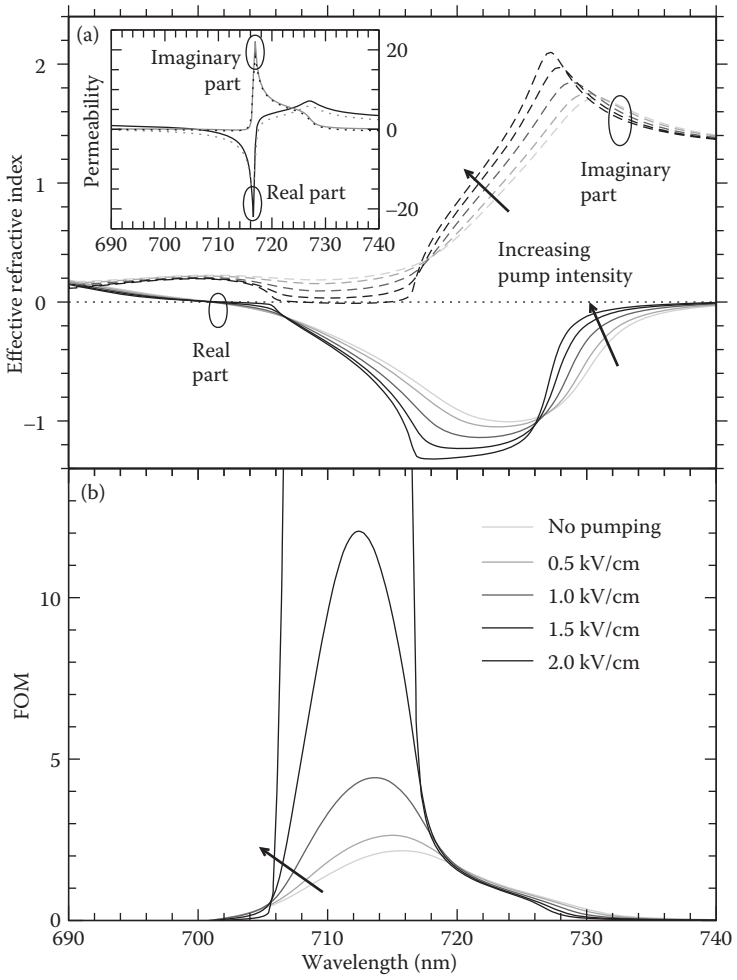


FIGURE 47.8 (a) Real and imaginary part of the retrieved effective refractive indices of the double-fishnet structure for different pump amplitudes. The peak electric field amplitude of the pump increases in steps of 0.5 kV/cm from no pumping (lightest) to a maximum of 2.0 kV/cm (darkest). The inset shows the real and imaginary part of the effective permeability (black and gray line, respectively) and the result of the Kramers–Kronig relation (dotted lines) for the highest peak electric field amplitude of 2.0 kV/cm . (b) The figures of merit (FOM) for the same pumping amplitudes. (Reproduced from Sebastian Wuestner et al., *Phys. Rev. Lett.*, 105,127401, 2010.)

47.7 Lasing Dynamics

In order to transition from loss compensation to lasing, photons emitted by the gain medium must feedback and stimulate on average more emission than photons that are lost on the round trip. Photons can be lost by escaping the cavity to radiation modes, absorbed incoherently by dissipative channels, or absorption by the gain medium.

Initially, for a cold cavity with no lasing mode field built up, an inverted gain system may spontaneously emit into cavity modes. This spontaneous emission may trigger further stimulated emission in a process known as *amplified spontaneous emission* (ASE), which can be added to the semiclassical FDTD description by including spatially resolved dynamic Langevin noise to the system, which accounts for the dissipative reservoirs feeding back stochastically on the system. The noise couples to the four-level system and induces incoherent transitions, which then become amplified, allowing for the triggering of the lasing regime [8].

As a device enters a lasing regime, characteristic relaxation oscillations can be seen where inversion initially builds up and spontaneous emission events are induced. The fields then are amplified as they stimulate further emission, growing the field in the mode coherently. From here on the emitted fields start to grow exponentially, and when they are of sufficient strength will deplete the inversion density. This reduction of inversion feeds back by decreasing the available gain, leading to a decrease in field energy as the energy in the field is lost to dissipative processes in the metal layers. The decrease in energy density allows for the inversion to rebuild. This continues in an oscillatory manner with the energy density lagging behind the inversion by 90°. The amplitude of the inversion and energy oscillations decrease with each cycle until a stable steady value for both is reached. In contrast, the control structure, albeit entering a regime of relaxation oscillations, is more erratic and the oscillations do not settle to a steady state. Instead, the oscillations continue with a factor of 4 between the peak energy density and the trough. An example of these relaxation oscillations in shown for the SL laser set in Figure 47.9 IN contrast to a "control structure" where the waveguide is not of high enough quality for the device to enter the lasing regime. Relaxation oscillations are accompanied by a characteristic decrease of the width of the power spectrum, from a wide peak in ASE to a sharp one in the lasing regime. This observation is usually taken as a crucial evidence for lasing in experimental setups.

47.8 Stopped-Light Lasing

Stopped-light lasing differs from traditional nanoplasmonic lasers in that there is no cavity with predefined modes. Whereas in the structures presented, light is trapped in a waveguide mode, it is free to propagate in one direction, albeit at ZGV. This means that there is a continuum of planewave modes for the gain to emit into, which get selected on the basis of the overlap with the gain media. The feedback in the system can be understood by considering the spatially resolved, cycle-averaged Poynting vector, energy density, and inversion that are plotted in Figure 47.10 for a structure with a gain width of 1000 nm. The energy is concentrated on metal-dielectric interfaces, strongest on the lower interface, and is localized around the gain medium despite there being no cavity along the horizontal direction.

The Poynting vector shows how energy circulates in the structure. There are four energy flux vortices, one in each corner of the structure; as described in the figure caption, these have energy moving out of the gain region into the metal layers and into it in dielectric layers. Considering the x component of the Poynting flux Figure 47.10a, the forward and backward flows are in exact balance. The balanced counter-propagation of energy in the negative-permittivity (metal) layers against the dielectric layer is the basis of the (subwavelength) feedback in the system.

The inversion is shown in Figure 47.10d; there are areas of spatial hole burning where the energy density is highest. The spatial modulation seen is explained when considering the field profile and its formation.

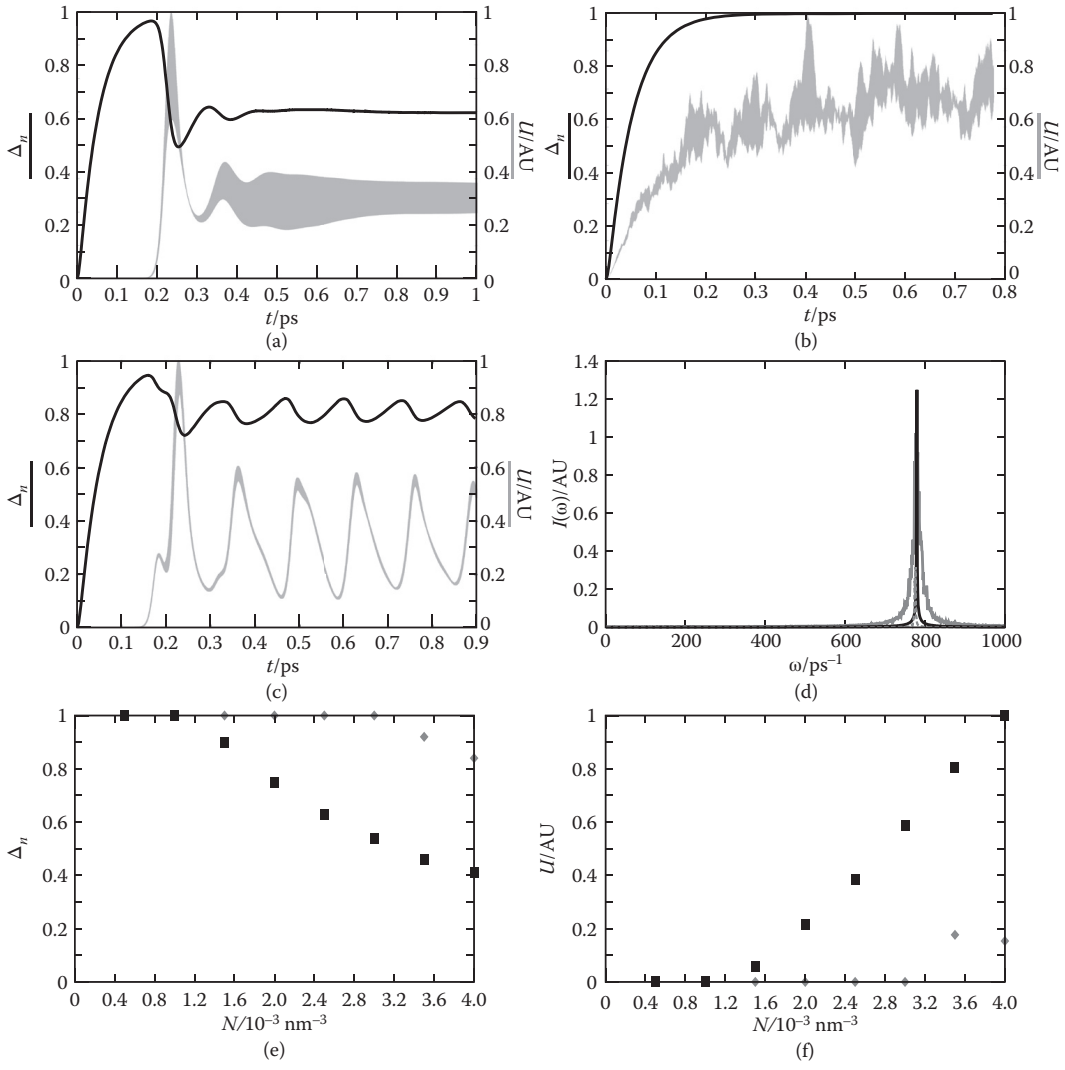


FIGURE 47.9 Lasing onset with increasing gain density. a, b, c. Energy density and mean inversion for (a) the stopped light structure with a gain density $N = 0.002\,\mathrm{nm}^{-3}$, (b) the stopped light structure with a gain density $N = 0.001\,\mathrm{nm}^{-3}$, and (c) the control structure with a gain density $N = 0.004\,\mathrm{nm}^{-3}$. (d) Power spectra for SL structure at $N = 0.002\,\mathrm{nm}^{-3}$ (black), SL structure at $N = 0.001\,\mathrm{nm}^{-3}$ (×5000 dashed gray), and control structure at $N = 0.004\,\mathrm{nm}^{-3}$ (solid gray). (e) Steady-state inversion of SL (black) and control (gray) structures with varying gain density. (f) Steady-state cycle-averaged energy density with varying gain density.

47.9 Summary

In this chapter, we have explored techniques to design, optimize, and simulate a nanoplasmonic laser. We started by defining harmonic oscillator type Drude–Lorentz and Maxwell–Bloch models, a universal type of models for metals, semiconductors, and gain media. From a simple model of forces in time domain, a causal frequency-domain response function was derived. From here, the four-level system was introduced as two dipole transitions in the frequency domain and as a dynamic system with self-consistently evolving

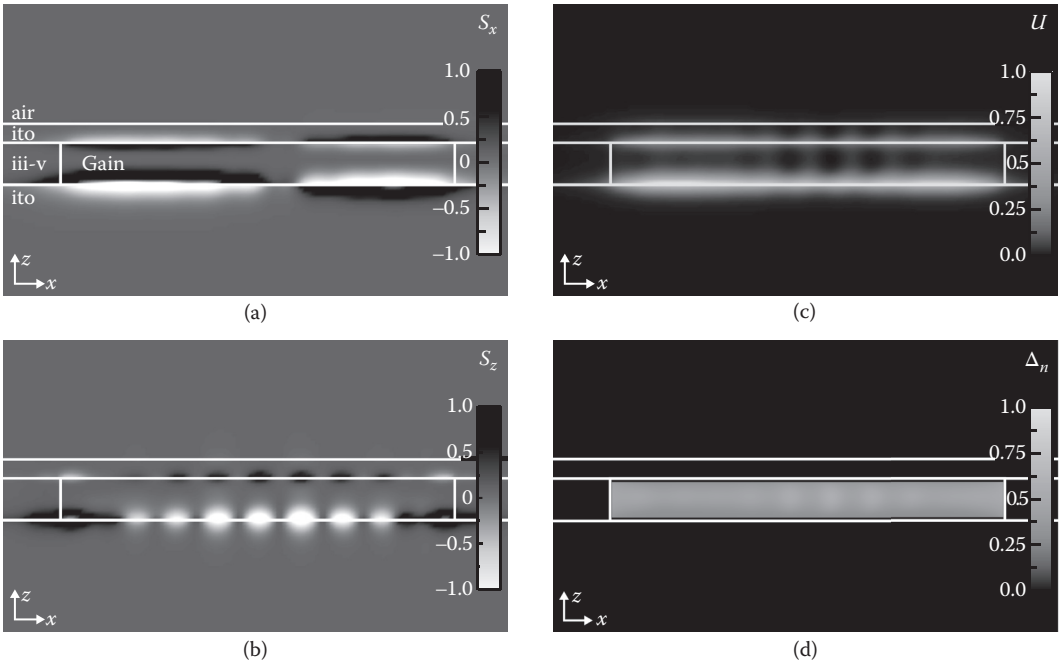


FIGURE 47.10 Energy flux and density of the lasing mode. The lasing mode of a stopped-light laser structure with a width 1000 nm. (a,b) The x and z components of the cycle-averaged Poynting vector, respectively, with positive values in black and negative in white. Focusing on the bottom right lobe of the flux in panel a, energy flows into the gain region in the dielectric layers and out of it in the metal layer. For the same lobe in panel b, energy moves downward out of the gain region toward the center, and upward into the region at the edge. The counter-clockwise energy vortex in this corner is mirrored in the other corners. (c) The cycle-averaged energy density and (d) inversion plots. The inversion is the complement of the energy density, being highest in areas with low energy density and depleted where the mode sits.

electron occupation levels in the time domain. The TMM was next introduced as an exact frequency-domain tool to analyze planar structures. EA are suited to optimize systems that can be characterized by a range of numeric or discrete parameters and perform characterized by a fitness metric. The EA approach was introduced to and demonstrated in the case of the TMM and applied to designing stopped light laser structures. Next, FDTD, the time-domain numerical method, was discussed and used to simulate active hybrid metallo-dielectric structures with nonlinearities and time-dependant dispersion such as a lasing system. This allowed for the simulation of lasing dynamics where loss compensation, ASE, and relaxation oscillations could be captured. Finally, some specifics of the recently established stopped light laser principle were explored as an example system of a designed and optimized nanoplasmonic laser.

References

1. Jennifer A. Dionne, Luke A. Sweatlock, Matthew T. Sheldon, A. Paul Alivisatos, and Harry A. Atwater. Silicon-based plasmonics for on-chip photonics. *IEEE Journal of Selected Topics in Quantum Electronics*, 16(1):295–306, 2010.
2. Anatoly V. Zayats and Igor I. Smolyaninov. Near-field photonics: Surface plasmon polaritons and localized surface plasmons. *Journal of Optics A: Pure and Applied Optics*, 5(4):S16, 2003.
3. Dmitri K. Gramotnev and Sergey I. Bozhevolnyi. Plasmonics beyond the diffraction limit. *Nature Photonics*, 4(2):83–91, 2010.

4. Kevin F. MacDonald, Zsolt L. Samson, Mark I. Stockman, and Nikolay I. Zheludev. Ultrafast active plasmonics. *Nature Photonics*, 3(1):55–58, 2009.
5. Ortwin Hess, John B. Pendry, Stefan A. Maier, Rupert Oulton, Joachim M. Hamm, and Kosmas L. Tsakmakidis. Active nanoplasmonic metamaterials. *Nature Materials*, 11(7):573–584, 2012.
6. Sebastian Wuestner, Andreas Pusch, Kosmas L. Tsakmakidis, Joachim M. Hamm, and Ortwin Hess. Overcoming losses with gain in a negative refractive index meta-material. *Physical Review Letters*, 105(12):127401, 2010.
7. David J. Bergman and Mark I. Stockman. Surface plasmon amplification by stimulated emission of radiation: Quantum generation of coherent surface plasmons in nanosystems. *Physical Review Letters*, 90:27402, 2003.
8. Andreas Pusch, Sebastian Wuestner, Joachim M. Hamm, Kosmas L. Tsakmakidis, and Ortwin Hess. Coherent amplification and noise in gain-enhanced nanoplasmonic metamaterials: A Maxwell-Bloch Langevin approach. *ACS Nano*, 6(3):2420–2431, 2012.
9. Sebastian Wuestner, Joachim M. Hamm, Andreas Pusch, Fabian Renn, Kosmas L. Tsakmakidis, and Ortwin Hess. Control and dynamic competition of bright and dark lasing states in active nanoplasmonic metamaterials. *Physical Review B*, 85(20):201406, 2012.
10. Bumki Min, Eric Ostby, Volker Sorger, Erick Ulin-Avila, Lan Yang, Xiang Zhang, and Kerry Vahala. High-Q surface-plasmon-polariton whispering-gallery microcavity. *Nature*, 457(7228):455–458, 2009.
11. Tim W. Pickering, Joachim M. Hamm, A. Freddie Page, Sebastian Wuestner, and Ortwin Hess. Cavity-free plasmonic nanolasing enabled by dispersionless stopped light. *Nature Communications*, 5:4972, 2014.
12. William M. Spears. *Evolutionary Algorithms: The Role of Mutation and Recombination. Natural Computing Series*. Heidelberg: Springer Berlin Heidelberg, 2000.
13. Kane Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14:302–307, 1966.
14. Allen Taflove. *Computational Electrodynamics: The Finite-Difference Time-Domain Method. Antennas and Propagation Library*. Norwood, MA: Artech House, 1995.
15. Xingjie Ni, Satoshi Ishii, Mark D. Thoreson, Vladimir M. Shalaev, Seunghoon Han, Sangyoon Lee, and Alexander V. Kildishev. Loss-compensated and active hyperbolic metamaterials. *Optics Express*, 19(25):25242–25254, 2011.
16. Johannes Skaar. Fresnel equations and the refractive index of active media. *Physical Review E*, 73(2):26605, 2006.
17. Paul Kinsler and Martin McCall. Causality-based criteria for a negative refractive index must be used with care. *Physical Review Letters*, 101(16):1–4, 2008.
18. P. Sperber, W. Spangler, Bernd Meier, and Alfons Penzkofer. Experimental and theoretical investigation of tunable picosecond pulse generation in longitudinally pumped dye laser generators and amplifiers. *Optical and Quantum Electronics*, 20(5):395–431, 1988.

48

Quantum-Dot Nanolasers

Christopher Gies
Michael Lorke
Frank Jahnke
and
Weng W. Chow

48.1 Going Nano: A New Era of Small Lasers 627
48.2 Connecting to Earlier Work: Laser Rate Equations and the β Factor 628
48.3 Key Aspects of Nanolasers 631
48.4 Essential Physics I: Quantum-Optical Semiconductor Laser Theory 635
Dynamical Laser Equations • Pump Process for Continuous-Wave-Excited Lasers
48.5 Essential Physics II: Carrier Kinetics in QD Systems 639
Carrier–Phonon Interaction • Coulomb Interaction
48.6 Connecting to Experiments: Transition to Lasing and the β Factor 646
48.7 Interemitter Coupling Effects and Superradiance in Nanolasers 647
48.8 Nonresonant Coupling and Lasing from Multiexciton States in Few-QD Systems 651
48.9 Outlook 653

48.1 Going Nano: A New Era of Small Lasers

The impact of semiconductor lasers on our daily lives is immense. They perform work every time someone gathers information for the internet, makes a telephone call, prints an article or pays for an item at a store. There are over 2.5 billion VCSELs[†] in use at our homes or offices, and roughly 100 million are produced each year to keep up with demand. VCSEL technology is now considered conventional, and research has progressed to smaller devices, modulated at higher speeds and producing output with greater spectral stability. This chapter focuses on two of the advances. One is the reduction of optical cavity volume by as much as two orders of magnitude, making a transition from microlasers (such as VCSELs) to nanolasers. Two is a change in active medium from quantum wells (QWs) to quantum dots (QDs), bringing about, e.g., devices operating with very few (tens) of emitters. These developments ushered in a new era for semiconductor device physics, one where quantum optical and many-body electron interaction effects dominate. The underlying motivation is the control of spontaneous emission [1]. The typical laser mitigates the randomness (noise) caused by spontaneous emission by overwhelming it with stimulated emission. Here, we are speaking of actually quieting the spontaneous emission noise, both spatially and temporally. In the former, we use nanocavities to inhibit spontaneous emission in undesirable directions. With the latter, we use very

[†] Vertical cavity surface emitting lasers, cf. Chapter 34.

few QDs (ideally only one) to control the timing of photon emission, thereby improving photon statistics beyond the limit described by the Poisson distribution. Devices incorporating the two advances are being fabricated and experiments are being performed. The results are both promising and exciting from the device and physics aspects, respectively. They also lead to many new questions and renewed interest in some old ones: What is lasing and where is the threshold? Why is there not a phonon bottleneck? What is the homogeneous width of a QD transition [2–4]? What is the inhomogeneous broadening in my samples? Is there really thresholdless lasing [5–7]? Are QD lasers better than QW lasers [8]? Why do QD lasers not show modulation speeds as advertised? This chapter describes a theoretical framework capable of addressing all the abovementioned questions. The building blocks come from quantum electrodynamics (QED), which is the quantum theory of the interaction of light with matter, and carrier interactions described by many-body theory.

The fundamental concepts of lasing are often discussed in terms of rate equations for material excitations and cavity photons that may well be familiar to the reader. We begin this chapter with an excursion to show profit and limitations of laser rate equations for understanding the behavior of nanolasers, and at the same time establish a connection to more sophisticated semiconductor laser models. In Section 48.3, we give a short overview of the key aspects in modeling nanolasers and the characterization of their emission properties. In Section 48.4, we introduce a microscopic laser model for solid-state emitters that goes beyond the rate equations and that provides access to the quantum-statistical properties of the emission. A central aspect of device performance is determined by the efficiency of carrier scattering processes following excitation. Section 48.5 provides an overview of the underlying mechanisms of carrier–carrier and carrier-phonon interaction and explains methods to incorporate the dynamics into semiconductor laser models within different levels of approximation. The underlying formalism allows one to address a wide variety of effects present in semiconductor nanolasers, and we provide a few examples in Sections 48.6 and 48.7, where QD lasing and radiative coupling between individual emitters are discussed. Section 48.8 gives an outlook toward systems with only a few QDs inside the cavity, where the individual electronic many-particle properties of each emitter matter, and where device properties are dominated by quantum and correlation effects.

48.2 Connecting to Earlier Work: Laser Rate Equations and the β Factor

Lasing operation takes place when pumping of the active material provides sufficient gain to compensate the losses. It is thus determined by the interplay of the excitation and emission dynamics of the gain material, taking into account the feedback provided by the radiation field inside the resonator. A basic understanding of lasing action can be obtained from laser rate equations in the form of two coupled equations for the gain (in this case, the number of excitons N in the QD ensemble) and the intracavity photon number n:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} N &= P - \gamma_l (n+1) N - \gamma_{\mathrm{nl}} N, \\ \frac{\mathrm{d}}{\mathrm{d}t} n &= -\kappa n + \gamma_l (n+1) N. \end{aligned} \tag{48.1}$$

Here, P and κ are the pump and photon loss rates, γ_l the spontaneous emission rate into the laser mode, γ_{nl} the rate of radiative recombination into other modes, and $\gamma_{\mathrm{sp}} = \gamma_l + \gamma_{\mathrm{nl}}$ the total spontaneous emission rate. The bracket $(n+1)$ contains both the stimulated and spontaneous emission contributions. The ratio

$$\beta = \frac{\gamma_l}{\gamma_{\mathrm{sp}}} \tag{48.2}$$

defines the β-factor illustrated in Figure 48.1. It is a key characteristic quantity of any laser device as it defines device performance in the spontaneous-emission regime (i.e., below threshold) and, more importantly, the threshold current. The rate equations (Equation 48.1) are often expressed in terms of β:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} N &= P - \beta\gamma_{\mathrm{sp}}\,(n+1)N - (1-\beta)\gamma_{\mathrm{sp}}\,N, \\ \frac{\mathrm{d}}{\mathrm{d}t} n &= -\kappa n + \beta\gamma_{\mathrm{sp}}\,(n+1)N. \end{aligned} \tag{48.3}$$

In conventional lasers, $\beta \approx 10^{-5}$ [9], so that the majority of spontaneously emitted photons are lost. Once the system starts lasing at sufficiently high excitation, stimulated emission completely overwhelms spontaneous emission, so that the fraction of photons lost into nonlasing modes becomes insignificant. This behavior can be directly traced by accessing both emission channels separately. In micropillar lasers, the fundamental mode often used as laser mode has a far-field emission in vertical direction of the pillar, whereas losses occur laterally through the pillar side walls. The input–output curve (photon number versus pump rate) collected separately from both directions, together with a schematic of the experiment, is shown in Figure 48.2 [10]. As long as spontaneous emission is the dominating emission process, both emission intensities increase linearly with pump with a fixed ratio determined by the β factor. At higher excitation powers, stimulated emission from the laser mode sets in and photon emission from the laser mode increases nonlinearly and more strongly than the emission into loss channels. The resulting increase of the normalized ratio r of both emission channels therefore gives direct account of the onset of stimulated emission. Note that the S shape of the intensity curve for the lateral emission is due to stray light from the laser mode that is collected in lateral direction. Ideally, a clamped intensity is expected.

When continuous excitation is considered, the rate equations are solved in the steady state. Input–output curves obtained from the rate equations (Equations 48.1) for different values of β are shown in Figure 48.3. For $\beta \ll 1$, a sudden increase in emission intensity is visible that allows one to identify the lasing threshold. With increasing coupling efficiency β, the threshold shifts to lower pump rates, offering the prospect of devices with reduced threshold currents. At the same time, the intensity jump decreases until the "thresholdless" case is obtained for $\beta = 1$. In this regime an identification of the transition into lasing is impossible from the emission intensity alone. A characterization of the threshold requires going beyond the rate equation approximation to study statistical and coherence properties of the emission. One of the attractiveness of nanolasers is that they are capable of operating in a regime close to $\beta = 1$ due to their small effective mode volumes and large cavity lifetimes that cause strong funneling of spontaneously emitted photons into a single resonator mode [1]. At the same time, the photonic density of states is reduced away from the cavity-mode frequency, which further suppresses emission into nonlasing modes and thereby also

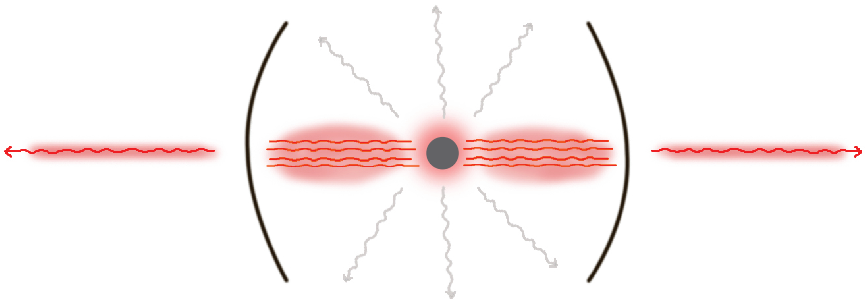


FIGURE 48.1 Spontaneous emission of the emitter into the laser mode (black lines) and into nonlasing modes (gray lines). The ratio of the spontaneous emission into the laser mode and the total spontaneous emission is described by the β factor.

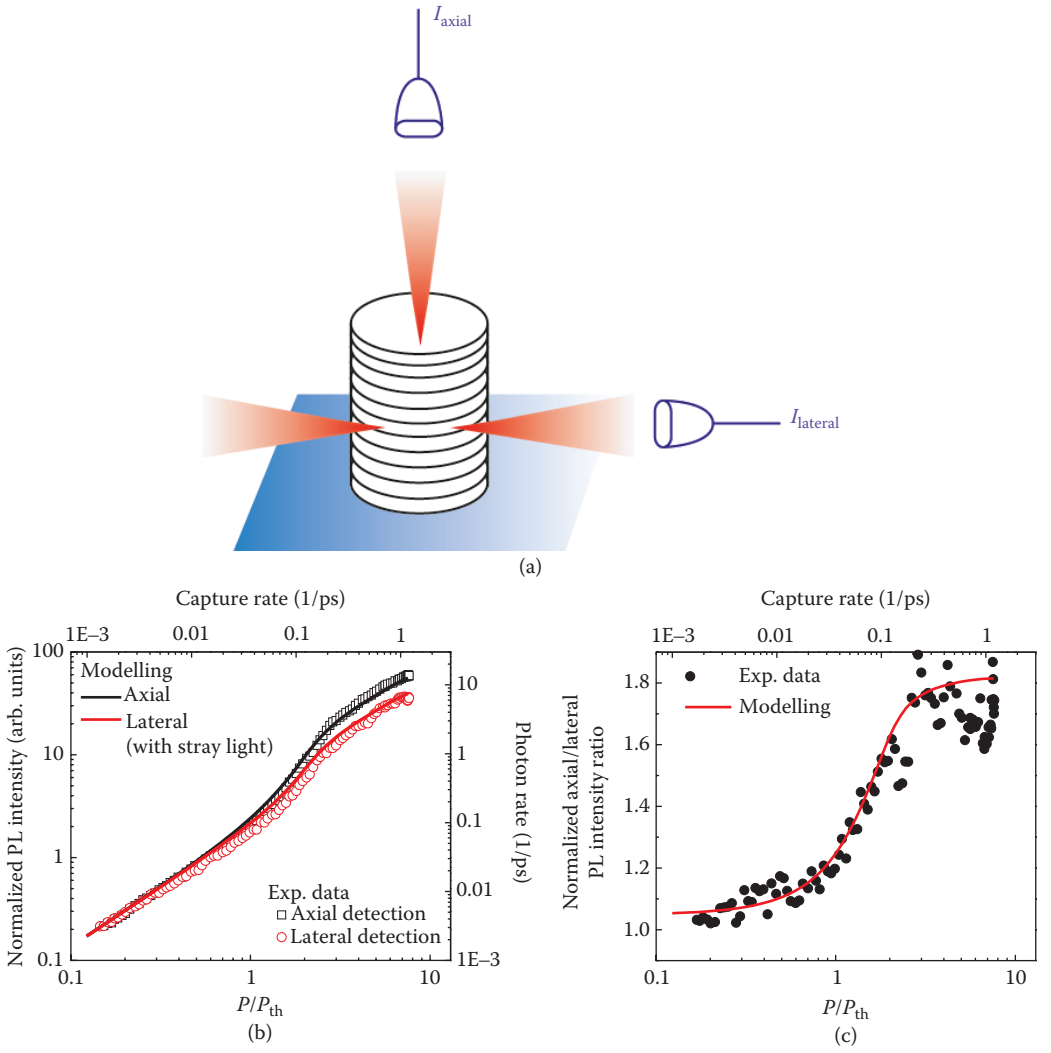


FIGURE 48.2 (a) Sketch of an experiment for detecting laser-mode emission in axial direction, and leaky mode emission in lateral direction separately. (b) Input–output characteristics of a quantum dot (QD) micropillar laser. Emission intensity into the laser mode (axial direction) and into nonlasing modes (lateral direction) are shown separately. (c) Their ratio r increases at the onset of stimulated emission, providing a direct visualization of the transition into the lasing regime. The results have been obtained from the semiconductor laser theory introduced in Section 48.4. Note that in lateral emission some scattered light from the axial emission is collected, which is responsible for the unclamped intensity behavior. [PL, photoluminescence. (Figures b and c adapted from Musiał, A. et al., *Physical Review B*, 91, 205310, 2015.)]

increases the β factor. While cavities operate at a specific narrow frequency window, structures that cause a broadband enhancement of spontaneous emission without using a cavity, such as photonic trumpets [11], are based on the very same concept.

The rate equations (Equation 48.1) have further drawbacks for modeling QD-based nanolasers. Most importantly, they do not provide information on laser field coherence and photon correlations. The rate equations are obtained by neglecting all correlations between light and matter degrees of freedom. Due to the enhanced interaction inside the microresonator, photon, and carrier-photon correlations have been demonstrated to play an important role in nanolasers, and this is where new approaches are required

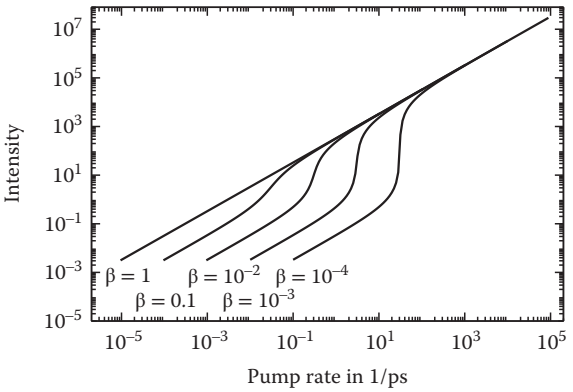


FIGURE 48.3 Laser input–output characteristics as obtained from rate equations. Shown are steady-state results at each pump value, parameterized by the β factor and the cavity loss rate. With increasing β, the threshold shifts to lower pumping and the laser threshold becomes increasingly hard to identify, and additional indicators beyond the reach of the rate-equation description are needed.

to understand and control these properties for new devices. In Section 48.4, we introduce a systematic approach to include semiconductor and correlation effects in extended laser models and discuss effects in nanolasers that set these systems apart from conventional laser devices.

48.3 Key Aspects of Nanolasers

Nanolasers use resonators that confine light in a volume of the light wavelength cube or smaller. The free spectral range, i.e., the energetic distance between resonator modes, grows with decreasing size of the cavity mode volume. Consequently, there is increased potential that the material gain overlaps only with one, or at most, a very few modes [12] of the resonator, as illustrated in the top right panel of Figure 48.4. This sets them apart from other lasers, such as edge-emitting lasers and the typical VCSELs [13,14], where optical modes lie more densely and emission is typically multimode (left panel). A nano-resonator may be a disk or pillar, employing distributed-Bragg reflectors [15], total internal reflection at semiconductor-air interfaces [16], or photonic crystal cavities [17].

It is a great advantage of small resonators that spontaneous emission enhancement $\propto Q/V$ can be used to strongly enhance the light–matter interaction and to favor emission into the single mode and thereby attain high values of the β factor that can approach unity. Here, V is the effective mode volume, and $Q = E/\Delta E$, with $E = \hbar\omega$ the mode energy and ΔE the mode linewidth, is the quality factor of the mode. The spontaneous-emission enhancement is direly needed to make up for the limited amount of gain material that can be placed inside the cavity. This so-called Purcell enhancement [18] can be directly observed in the temporal decay of the emission signal following short excitation. This is shown in Figure 48.5 [19] for micropillar nanolasers of different diameter and, therefore, different effective mode volume V.

As mentioned in Section 48.2, additional indicators are required to identify lasing in high-β devices due to the vanishing kink in the input–output curve. The most established method is to identify the statistical fluctuations of the emitted light. In the thermal regime below threshold, photons prefer to come in bunches. Lasing, on the other hand, originates from a coherent state of the light field, and photons arrive randomly in time. A transition from bunching to close to Poissonian emission is therefore a strong indicator for crossing the threshold to lasing. State-of-the-art streak-camera measurements give a direct visual account of this behavior: Figure 48.6 [20] shows temporally resolved single photon detection events for a system that has been excited by a short laser pulse. Counting statistics can be used to identify single photons, photon pairs, or bunches of more photons as indicated by the boxes. In practice, measurements are repeated thousands

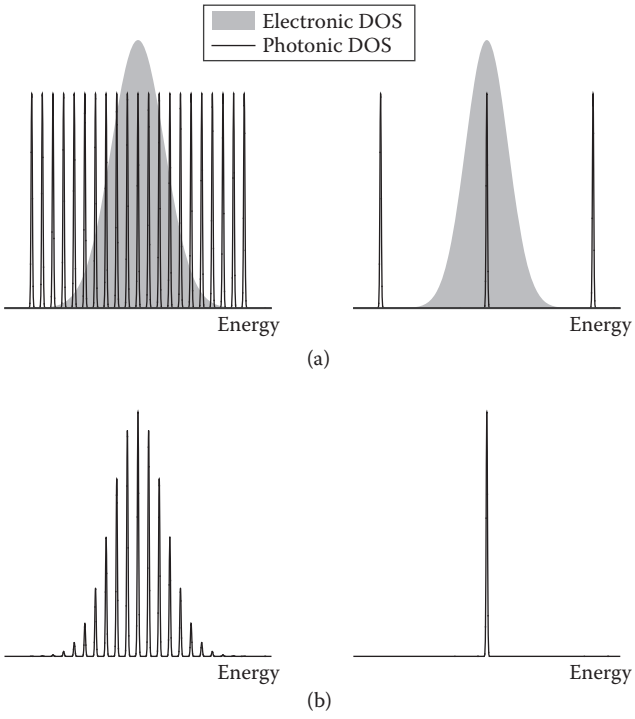


FIGURE 48.4 (a) Illustration of the photonic (black) and electronic (shaded) density of states of a typical laser (left) and a nanolaser (right) with the latter having a much larger free spectral range (energetic spacing between photon modes). The QD gain in the right panel is inhomogeneously broadened due to variations in composition or dimension, see Section 48.8. (b) The resulting spontaneous emission spectra are determined by the overlap of the photonic and electronic density of states. On the other hand, with the proper pillar, disk or photonic crystal design, a nanolaser can be single mode both below and above the lasing threshold. (DOS, density of states.)

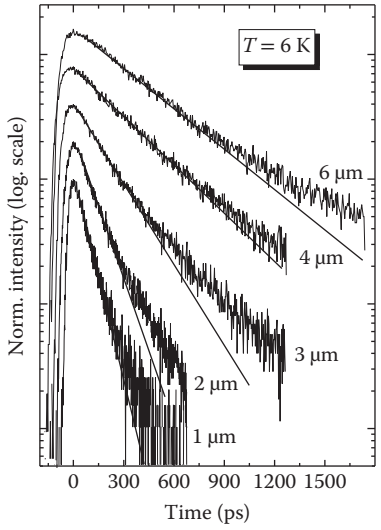


FIGURE 48.5 Time-resolved photoluminescence spectra following a *weak* excitation pulse. A decrease of the spontaneous emission time is observed with decreasing diameter of the micropillar cavities, reflecting the Purcell-enhancement $\propto Q/V$. The weak excitation ensures that the effect is not related to stimulated emission. (Reprinted from Schwab, M. et al., *Physical Review B*, 74, 045323, 2006. With permission.)

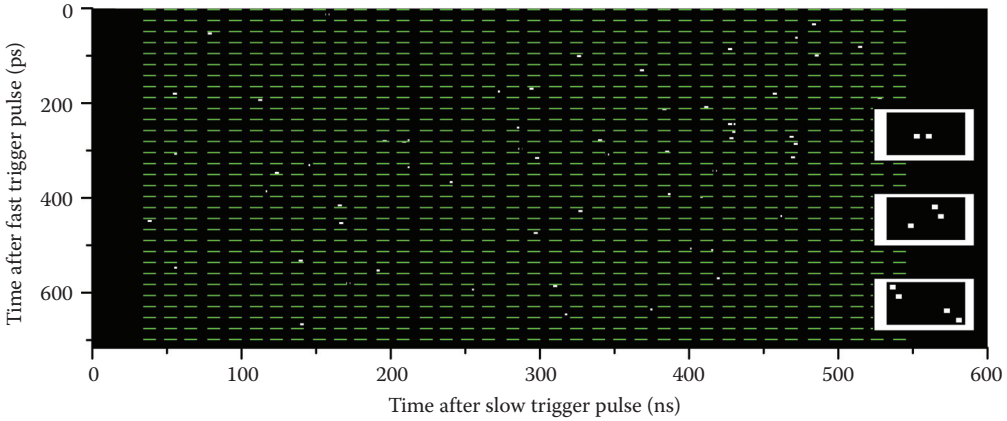


FIGURE 48.6 Streak-camera measurements provide impressive account of statistical fluctuations of the emitted light on the level of single photons. Single- and multiphoton emission events can be detected within a time-bin size of 2 ps (see magnification in the boxes) and provide insight into the quantum-mechanical state of the light field. To perform counting statistics in order to construct correlation functions like the ones shown in Figure 48.7, ten thousands of such measurements are performed. (Reprinted from Aßmann, M. et al., *Science* 325, 297–300, 2009. With permission).

of times. From the collected data, photon correlation functions $g^{(n)}$ can be obtained. They are of particular relevance, as they allow one to characterize the state of the quantum-mechanical light field as given by the distribution function p_n, which describes the *photon statistics* in terms of the probability of finding the system in the photon number state $|n\rangle$. Photon correlation functions are connected to the moments of the photon statistics:

$$\langle n^i \rangle = \sum_{k=0}^{\infty} k^i p_k. \tag{48.4}$$

As an example, the second-order photon-correlation function (sometimes referred to as photon autocorrelation function) is given by

$$g^{(2)} = \frac{\langle n^2 \rangle - \langle n \rangle}{\langle n \rangle^2}. \tag{48.5}$$

Correlation functions of order n take on a value of $n!$ if the light field is in a thermal state, and at all orders they possess a value of 1 for a perfectly coherent state [21]. For a micropillar QD nanolaser, this has been experimentally observed for $g^{(2)}, g^{(3)}$ and $g^{(4)}$ as function of pump rate, shown in Figure 48.7.

More commonly used than a streak camera is a Hanbury Brown and Twiss type measurement [22] schematically shown in Figure 48.8, where the bunching behavior of photons is quantified by the simultaneous registration of two photons at the two detector arms, providing access to the second-order correlation function $g^{(2)}$. For detectors, avalanche photodiodes (APDs) are used. Their time resolution is typically in the range of hundreds of picoseconds. Whether this resolution is sufficient depends on the type of measurement performed and the coherence time τ_{coh} of the emission signal. On the timescale of the latter, $g^{(2)}(\tau)$ goes to unity with respect to the delay time τ between two detection events, and the actual $g^{(2)}(0)$ at $\tau = 0$ can no longer be resolved if the detector's time resolution exceeds the coherence time, leading to the effect that $g^{(2)}(0)$ seems to take on a value of 1 even for thermal light [23].

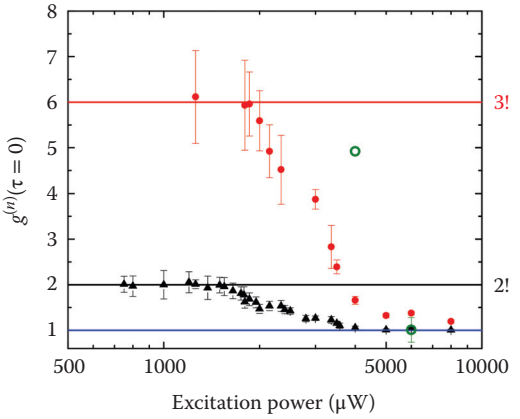


FIGURE 48.7 Photon autocorrelation functions $g^{(2)}(0)$ (triangles), $g^{(3)}(0)$ (dots), and $g^{(4)}(0)$ (open circles) as function of excitation power obtained from streak-camera measurements like the ones shown in the previous figure. For a typical LED to laser transition, all correlation functions show a transition from their thermal values $g^{(n)} = n!$ to that of a coherent field, $g^{(n)} = 1$. (Courtesy of Marc Aßmann, TU Dortmund.)

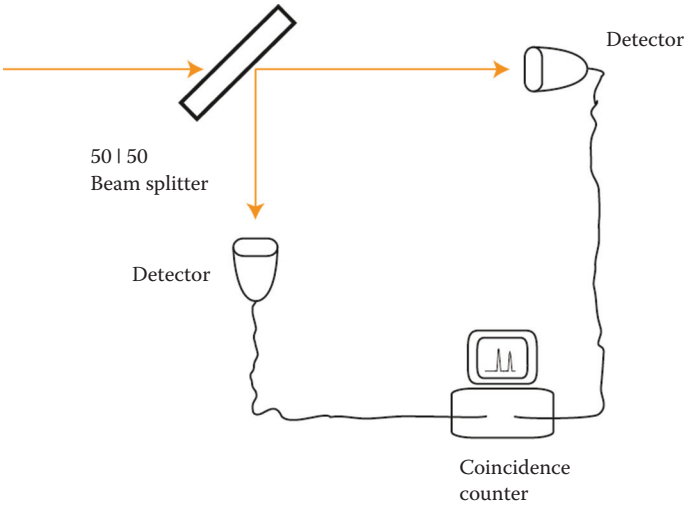


FIGURE 48.8 Illustration of a Hanbury Brown and Twiss measurement. A beam splitter distributes the emission between two detectors. Equal-time detection of two or more photons provides access to the second-order photon autocorrelation function $g^{(2)}(\tau = 0)$. A delay line in one arm can be used to measure $g^{(2)}(\tau \neq 0)$.

Another accessible indicator for coherent emission is a linewidth narrowing, or equivalently, an increase of the coherence time. Experimentally, the linewidth can be determined from spectra or Michelson interferometry measurements [21]. In the steady state $t = t_{ss}$, the coherence time τ_{coh} is related to the first-order coherence of the emission

$$\tau_{coh} = \int_{-\infty}^{\infty} d\tau \, |g^{(1)}(t_{ss}, \tau)|^2, \tag{48.6}$$

which requires the calculation of the two-time quantity in the integral by means of the quantum-regression theorem [24,25]. We will not go into more detail here and refer the reader to [26].

The criteria for lasing in nanolaser devices are a subject of much discussion [5,27]. Ideally, a combination of several criteria is used to identify the lasing regime and the threshold. These include the following:

1. Coherent emission approaching Poissonian photon statistics p_n.
2. A value close to 1 in the second-order photon correlation function $g^{(2)}(0)$. Note that this is a weaker criterion than 1., as $g^{(2)}(0)$ is only linked to the first two moments of the full photon distribution function p_n, c.f. Equation 48.5.
3. An increase in the coherence time at the onset of lasing, when phase-coherent stimulated emission begins to dominate over spontaneous emission.
4. Cavity-photon-induced feedback of the gain material providing gain instead of absorption.
5. A mean intracavity photon number above 1.
6. An S-shaped nonlinear region in the input–output characteristics at threshold.
7. Carrier dynamics developing a hole-burning effect in the carrier-population functions at the energy of the laser transition, cf. Figure 48.13.

All of these criteria have their justifications, but on their own, neither may suffice to conclusively identify lasing. For example, coherent emission (1) is a prerequisite of lasing, but it may be realized without amplification (4 and 5) or originate from an external coherent drive. An S-shaped jump (6) in the input–output characteristic can originate from a variety of effects, such as a transition between multiexciton configurations feeding the laser mode [28], background contributions [29], or spectral wandering. On its own it is not a clear indicator for lasing. Moreover, lasing can take place in the absence of a nonlinear emission regime, e.g., if spontaneous emission losses are small ($\beta \to 1$), or saturation effects set in before the S shape can fully develop, which is particularly relevant under pulsed excitation [30].

48.4 Essential Physics I: Quantum-Optical Semiconductor Laser Theory

Different approaches can be taken in formulating a laser theory, including semiclassical methods, where the radiation field is classical and the system dynamics under the influence of the light–matter interaction is described by Maxwell–Bloch equations [31,32]. In the following, we introduce an extended laser model that accounts for semiconductor-specific effects, such as carrier scattering and dephasing, and that provides access to the statistical properties of the emission. The model is obtained in the framework of a correlation-expansion technique for the quantized light field interacting with carriers occupying the discrete QD conduction- and valence states. The method is versatile, as it allows one to include various aspects of semiconductor and correlation effects on a required level in a systematic and consistent manner. At the same time, the rate equations are contained as limiting case, as has been shown in [33].

In contrast to semiclassical approaches, using a quantized light field introduces the possibility to investigate emission properties beyond the classical regime. In fact, nonclassical effects, such as antibunched light emission around the threshold region, have been demonstrated in QD nanolasers [34]. In semiclassical models, the electromagnetic field drives a coherent polarization $\langle v_\nu^\dagger c_\nu \rangle$ across the bandgap of the semiconductor medium, as illustrated in Figure 48.9. The operators c_ν ($c_\nu^\dagger$) annihilate (create) a conduction-band carrier in the state $|\nu\rangle$, the operators v_ν ($v_\nu^\dagger$) are the equivalent for valence carriers. The resulting dynamics is familiar from atomic two-level models, where a resonant field drives Rabi oscillations. In a fully quantum-mechanical approach, the electromagnetic field is described by modes that are occupied by field quanta, the photons. If the modes are energetically well separated, like in a nanolaser, where the cavity possesses a large free spectral range, it suffices to consider the interaction of the active medium with photons in a single mode q. The coherent polarization $\langle v_\nu^\dagger c_\nu \rangle$ is then replaced by a photon-assisted polarization $\langle b_q^\dagger v_\nu^\dagger c_\nu \rangle$ that describes a carrier deexcitation (or excitation via $\langle b_q c_\nu^\dagger v_\nu \rangle$) by simultaneous emission (absorption)

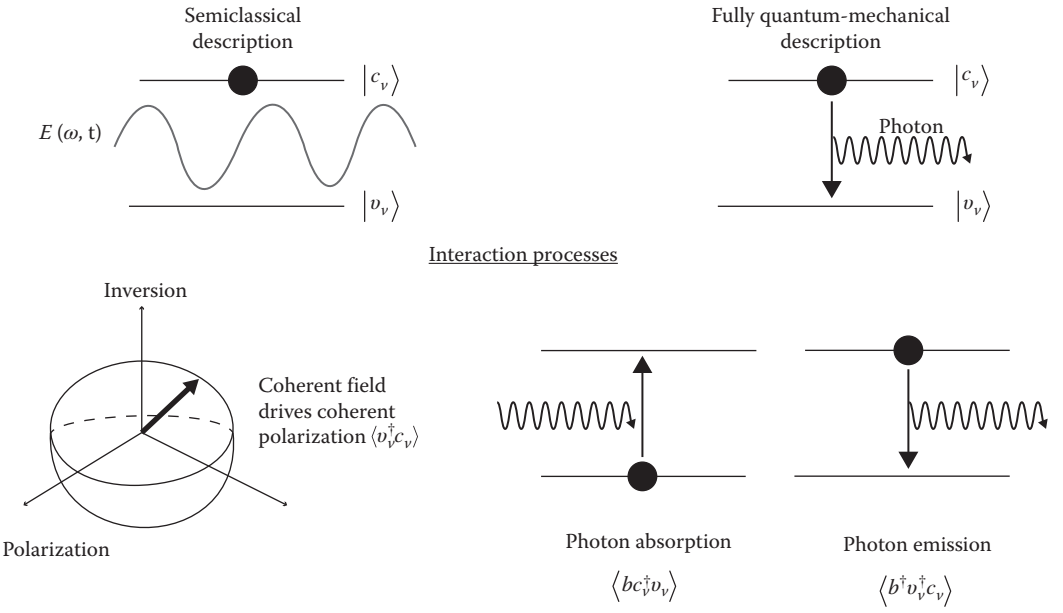


FIGURE 48.9 Illustration of the fundamental light–matter interaction in a semiclassical (left) and fully quantum-mechanical description (right). Semiclassically, matter degrees of freedom are described quantum-mechanically (e.g., by the states $|c_\nu\rangle$ and $|v_\nu\rangle$). The interaction with a classical coherent field $E(\omega, t)$ drives a coherent polarization $\langle v_\nu^\dagger c_\nu\rangle$. The polarization and population dynamics of often described in terms of the rotation of a vector on a unit sphere (Bloch sphere), as illustrated in the bottom left. By quantizing the light field, field quanta (photons) are introduced, and the fundamental interaction processes become the electronic (de)excitations via the simultaneous absorption or emission of a photon (bottom right). The semiclassical coherent polarization is replaced by the photon-assisted polarization $\langle b^\dagger v_\nu^\dagger c_\nu\rangle$.

of a photon into the mode q; the Bose operators b_q ($b_q^\dagger$) annihilate (create) a photon in mode q. The quantum-mechanical and semiclassical pictures are schematically compared in Figure 48.9.

The fully quantum-mechanical laser model is obtained by calculating the dynamical equations for carrier populations $f_\nu^c = \langle c_\nu^\dagger c_\nu\rangle$, $f_\nu^v = \langle v_\nu^\dagger v_\nu\rangle$ in the conduction and valence states of the QD, and the expectation value of the photon number operator in the laser mode q, for which we drop the index q in the following. These equations are obtained from Heisenberg's equations of motion for operators $A(t)$, $\dot{A} = i[H, A]/\hbar$. In the Hamiltonian, we consider the free contributions of QD single-particle states (first and second term in Equation 48.8) and the electromagnetic field (third term in Equation (48.8)), and the light–matter interaction,

$$H = H^{\text{free}} + H^{\text{LM}}, \tag{48.7}$$

with

$$H^{\text{free}} = \sum_\nu \varepsilon_\nu^c\, c_\nu^\dagger c_\nu + \sum_\nu \varepsilon_\nu^v\, v_\nu^\dagger v_\nu \; + \hbar\omega\left(b^\dagger b + \frac{1}{2}\right) \tag{48.8}$$

and

$$H_{\text{LM}} = -i\sum_{\alpha\nu}\left(g_{\alpha\nu}\, c_\alpha^\dagger v_\nu b + g_{\alpha\nu}\, v_\alpha^\dagger c_\nu b\right) + \text{h.c.} \tag{48.9}$$

Here, $\hbar\omega$ is the mode energy, and $g_{\alpha\nu}$ is the light–matter coupling strength that is determined by the overlap of the wave functions of the QD single-particle states $|\alpha\rangle$ and $|\nu\rangle$ and the mode function of the confined electromagnetic field [33]. The sums run over the single-particle states of the individual QD emitters, and the index is to be understood to contain both, QD label and single-particle state. The interaction Hamiltonian contains the elementary processes of electron–hole recombination and generation via emission and absorption of a photon, including the ones schematically shown in the bottom right of Figure 48.9. Note that the counter-rotating terms in Equation (48.9) are often neglected in the rotating wave approximation.

48.4.1 Dynamical Laser Equations

For the dynamical evolution of the photon number $\langle b^\dagger b\rangle$ in the given mode and the electron and hole populations $f_\nu^e = \langle c_\nu^\dagger c_\nu\rangle$, $f_\nu^h = 1 - \langle v_\nu^\dagger v_\nu\rangle$, the contribution of the light–matter interaction H_{LM} in the Heisenberg equations of motion leads to [33]

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t} + \kappa\right)\langle b^\dagger b\rangle = 2\,\mathrm{Re}\sum_{\nu'}|g_{\nu'}|^2\,\langle b^\dagger v_{\nu'}^\dagger c_{\nu'}\rangle, \tag{48.10}$$

$$\hbar\frac{\mathrm{d}}{\mathrm{d}t}f_\nu^{e,h} = -2\,\mathrm{Re}\,|g_\nu|^2\,\langle b^\dagger v_\nu^\dagger c_\nu\rangle. \tag{48.11}$$

Here, we have scaled $\langle b^\dagger v_\nu^\dagger c_\nu\rangle \to g_{q\nu}\langle b^\dagger v_\nu^\dagger c_\nu\rangle$ to have the modulus of the coupling matrix elements appear and have further assumed that g is diagonal in the equal-envelope approximation. The finite lifetime of the cavity mode is introduced by considering a complex mode energy, $\tilde{\omega} = \omega - i\kappa$, where κ is directly connected to the Q-factor of the laser mode, $Q = \hbar\omega/2\kappa$. The dynamics of the photon number is determined by the abovementioned photon-assisted polarization $\langle b^\dagger v_\nu^\dagger c_\nu\rangle$ that describes the expectation value for a correlated event, where a photon in the mode q is created in connection with an interband transition of an electron from a conduction to a valence state. The sum over ν involves all possible interband transitions from various QDs.

An important component of the carrier dynamics is the recombination into nonlasing modes giving rise to the β factor discussed in Section 48.2. For the fraction $1-\beta$ of carriers emitting into nonlasing modes, a term proportional to the spontaneous emission rate can be added to the carrier population equations to account for the losses [33]:

$$\frac{\mathrm{d}}{\mathrm{d}t}f_\nu^{e,h}\Big|_{\mathrm{nl}} = \gamma_{\mathrm{sp}}(1-\beta)f_\nu^e f_\nu^h. \tag{48.12}$$

In contrast to the rate equations (Equation 48.3), where this contribution is typically accounted for by the β coefficient, the spontaneous emission losses in a semiconductor are determined by both electron and hole populations. The dynamical equation for the cavity-photon-assisted polarization is given by

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t} + \kappa + \Gamma + i\left(\tilde{\varepsilon}_\nu^e + \tilde{\varepsilon}_\nu^h - \hbar\omega\right)\right)\langle b^\dagger v_\nu^\dagger c_\nu\rangle =$$

$$f_\nu^e f_\nu^h - \left(1 - f_\nu^e - f_\nu^h\right)\langle b^\dagger b\rangle + \frac{1}{g_\nu}\sum_\alpha g_\alpha C^{\mathrm{x}}_{\alpha\nu\nu\alpha} + \delta\langle b^\dagger b c_\nu^\dagger c_\nu\rangle - \delta\langle b^\dagger b v_\nu^\dagger v_\nu\rangle. \tag{48.13}$$

The free evolution of $\langle b^\dagger v_\nu^\dagger c_\nu\rangle$ is determined by the detuning of the QD transitions from the cavity laser mode. This oscillatory term drops out if the transition is in perfect resonance with the mode. Only accounting for identical resonant QDs is a common and well-justified practice in ensembles of many (>100 emitters), as the summation over all QDs becomes a mere prefactor of the QD number, which is a significant simplification. In a semiconductor, the source term of spontaneous emission is described by an expectation value of four carrier operators $\langle c_\alpha^\dagger v_\alpha v_\nu^\dagger c_\nu\rangle$, see [35]. The Hartree–Fock factorization of

this source term leads to $f_\nu^e f_\nu^h$, which appears as the first term on the right hand side of Equation 48.13. Corrections to this factorization are included in $C^{\mathrm{x}}_{\alpha'\nu\nu'\alpha} = \delta\langle c^\dagger_{\alpha'} v^\dagger_\nu c_{\nu'} v_\alpha\rangle$. Dephasing is what destroys a coherent polarization. In semiconductors, dephasing arises from carrier scattering processes, excitation of carriers, and from the interaction with lattice vibrations that are responsible for the homogeneous linewidth. These contributions are here summarized in a phenomenological constant Γ.

The correlation functions $\delta\langle b^\dagger_q b_q c^\dagger_\nu c_\nu\rangle$ and $\delta\langle b^\dagger_q b_q v^\dagger_\nu v_\nu\rangle$ introduce carrier-photon correlations that are neglected on the level of rate equations. Their time evolution is given by [33]

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t} + 2\kappa\right)\delta\langle b^\dagger b c^\dagger_\nu c_\nu\rangle = -2\left|g_\nu\right|^2 \mathrm{Re}\Big[\delta\langle b^\dagger b^\dagger b v^\dagger_\nu c_\nu\rangle + \left(\langle b^\dagger b\rangle + f^e_\nu\right)\langle b^\dagger v^\dagger_\nu c_\nu\rangle\Big], \tag{48.14}$$

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t} + 2\kappa\right)\delta\langle b^\dagger b v^\dagger_\nu v_\nu\rangle = 2\left|g_\nu\right|^2 \mathrm{Re}\Big[\delta\langle b^\dagger b^\dagger b v^\dagger_\nu c_\nu\rangle - +\left(\langle b^\dagger b\rangle + f^h_\nu\right)\langle b^\dagger v^\dagger_\nu c_\nu\rangle\Big] \tag{48.15}$$

which couples to

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t} + 3\kappa + \Gamma + i\left(\tilde{\varepsilon}^e_\nu + \tilde{\varepsilon}^h_\nu - \hbar\omega\right)\right)\delta\langle b^\dagger b^\dagger b v^\dagger_\nu c_\nu\rangle = -2\left|g_\nu\right|^2\langle b^\dagger v^\dagger_\nu c_\nu\rangle^2 + 2\left[\langle b^\dagger b\rangle + f^h_\nu - f^e_\nu\right]\times$$
$$\left[\delta\langle b^\dagger b c^\dagger_\nu c_\nu\rangle - \delta\langle b^\dagger b v^\dagger_\nu v_\nu\rangle\right] - (1 - f^e_\nu - f^h_\nu)\,\delta\langle b^\dagger b^\dagger b b\rangle \tag{48.16}$$

and

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t} + 4\kappa\right)\delta\langle b^\dagger b^\dagger b b\rangle = 4\left|g_\nu\right|^2 \sum_{\nu'}\delta\langle b^\dagger b^\dagger b v^\dagger_{\nu'} c_{\nu'}\rangle. \tag{48.17}$$

On this level of approximation, we obtain a closed system of coupled equations that go beyond the rate equations by containing carrier-photon and photon–photon correlation functions that provide access to the second-order photon-correlation function

$$g^{(2)}(\tau = 0) = 2 + \frac{\delta\langle b^\dagger b^\dagger b b\rangle}{\langle b^\dagger b\rangle^2}. \tag{48.18}$$

48.4.2 Pump Process for Continuous-Wave-Excited Lasers

The earlier equations describe the light–matter interaction of a semiconductor system in terms of carrier populations and the polarization-analogue for the quantized light field, i.e., the photon-assisted polarization. An important component is yet missing, namely the interaction of charge carriers with the environment that allows for some sort of excitation mechanism. Pumping can be either done by current injection or by optical excitation. In both cases, electrons and holes are typically created in the continuum states of the wetting layer (WL) or barrier material, from where they are captured into the discrete QD states. This incoherent capture is one example of carrier scattering processes that, at this stage, are still missing from our description. While the quantitative modeling of these many-body processes is discussed in detail in Section 48.5, for continuous-wave (CW) excitation it is often sufficient to consider only the dynamics of carriers in the localized QD states. In its most simple form, two localized states are used for electrons and holes each, which closely resembles a four-level laser scheme. This choice is physically motivated by the electronic structure of typical QDs, which, depending on their size, consists of several confined states. We refer to the lowest (highest) and second-lowest (second highest) electron (hole) state as *s*- and *p*-states, respectively, as shown in Figure 48.10. In this reduced level scheme, the *p*-states are used for carrier pumping at rate P, which approximates the combined effect of carrier excitation in higher

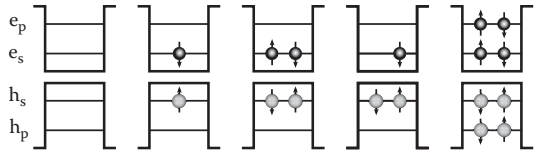


FIGURE 48.10 In a QD with several confined single-particle states for electrons and holes (the s- and p-states are shown), a multitude of many-particle configurations can be realized by occupying these states with electrons and holes. These many-particle configurations contain the neutral, charged, and multiexciton states, and dark states that differ in energy due to the Coulomb interaction between carriers.

states and the successive capture into the QD states:

$$\frac{\mathrm{d}}{\mathrm{d}t} f_{\mathrm{p}}^{\mathrm{e}}\Big|_{\mathrm{Pump}} = P(1 - f_{\mathrm{p}}^{\mathrm{e}} - f_{\mathrm{p}}^{\mathrm{h}}) \tag{48.19}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} f_{\mathrm{p}}^{\mathrm{h}}\Big|_{\mathrm{Pump}} = P(1 - f_{p}^{\mathrm{e}} - f_{\mathrm{p}}^{\mathrm{h}}). \tag{48.20}$$

The blocking factors in the brackets ensure that the occupation remains bounded by unity. From their generation in the QD *p*-states, a phenomenological set of equations distributes carriers into the QD s-states:

$$\frac{\mathrm{d}}{\mathrm{d}t} f_{\mathrm{s}}^{\mathrm{e}}\Big|_{\mathrm{scatt}} = \gamma_{\mathrm{p}\to\mathrm{s}}^{r,e} f_{\mathrm{p}}^{\mathrm{e}}(1 - f_{\mathrm{s}}^{\mathrm{e}}) - \gamma_{s\to p}^{r,e} f_{\mathrm{s}}^{\mathrm{e}}(1 - f_{\mathrm{p}}^{\mathrm{e}}) \tag{48.21}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} f_{\mathrm{s}}^{\mathrm{h}}\Big|_{\mathrm{scatt}} = \gamma_{\mathrm{p}\to\mathrm{s}}^{r,h} f_{\mathrm{p}}^{\mathrm{h}}(1 - f_{\mathrm{s}}^{\mathrm{h}}) - \gamma_{s\to p}^{r,h} f_{\mathrm{s}}^{\mathrm{h}}(1 - f_{\mathrm{p}}^{\mathrm{h}}) \tag{48.22}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} f_{\mathrm{p}}^{\mathrm{e}}\Big|_{\mathrm{scatt}} = -\gamma_{\mathrm{p}\to\mathrm{s}}^{\mathrm{r,e}} f_{\mathrm{p}}^{\mathrm{e}}(1 - f_{\mathrm{s}}^{\mathrm{e}}) + \gamma_{s\to p}^{\mathrm{r,e}} f_{\mathrm{s}}^{\mathrm{e}}(1 - f_{\mathrm{p}}^{\mathrm{e}}) \tag{48.23}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} f_{\mathrm{p}}^{\mathrm{h}}\Big|_{\mathrm{scatt}} = -\gamma_{\mathrm{p}\to\mathrm{s}}^{\mathrm{r,h}} f_{\mathrm{p}}^{\mathrm{h}}(1 - f_{\mathrm{s}}^{\mathrm{h}}) + \gamma_{\mathrm{s}\to\mathrm{p}}^{\mathrm{r,h}} f_{\mathrm{s}}^{\mathrm{h}}(1 - f_{\mathrm{p}}^{\mathrm{h}}). \tag{48.24}$$

Equations 48.21 through 48.24 are motivated in the more general concepts of carrier kinetics in the following section. Together, Equations 48.10–48.24 serve as an accessible and complete model for QD-based lasers, and we discuss exemplary results in Section 48.6.

48.5 Essential Physics II: Carrier Kinetics in QD Systems

As we overviewed in Section 48.2, rate equations can provide important insights for the theoretical understanding of lasers. They can be used to describe input–output curves [36–39], and, combined with Langevin approaches [40,41] even the photon statistics and intensity noise.

An important reason why carrier dynamics is essential for the understanding of laser operation is the excitation mechanism. To see this, let us assume that carrier populations in lasers are given by Fermi–Dirac distribution functions (often a reasonable approximation). This implies that states with below-average kinetic energy are Pauli-blocked for the excitation mechanism. This is illustrated in Figure 48.11, where states with above-average kinetic energy are available for the excitation, while highly occupied states (shaded region, applied to 77K) are not. Therefore, it is important to understand how carriers, injected at the green-shaded area, end up in the spectral window from where they recombine. Accountable for this are two mechanisms that we discuss in this section. The carrier–carrier Coulomb interaction redistributes carriers between different states and leads to an equilibration towards a Fermi–Dirac distribution [42–46]. But as every carrier scattered downward in energy is partnered by a scattering partner that is scattered upward (right sketch in Figure 48.12), the Coulomb scattering is not able to dissipate energy.

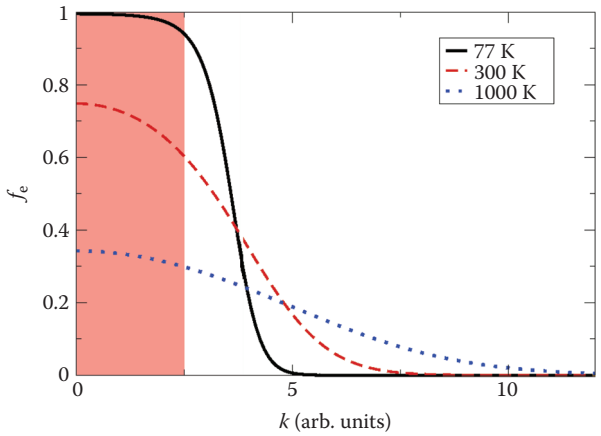


FIGURE 48.11 Fermi–Dirac distribution for electrons in a QW at different temperatures and a carrier density of $5 \times 10^{11}\text{cm}^{-2}$, the shaded background indicates the availability for carrier injection for the low-T distribution.



FIGURE 48.12 Conceptual sketch of carrier scattering processes. Left: carrier–phonon interaction, right: carrier–carrier Coulomb interaction.

This leads to a hot Fermi–Dirac function such as depicted in Figure 48.11, that has significantly different properties regarding inversion and gain. The cooling toward the lattice temperature is provided by the carrier-phonon interaction [47–51], that provides for the cooling by taking up the energy of the hot carriers and distributing it toward crystal lattice (left sketch in Figure 48.12). In a laser under CW-excitation conditions the energetic difference from excitation to emission energy also leads to heating and often the carrier-phonon interaction is not strong enough to provide complete cooling towards the lattice temperature. Therefore, a dynamical equilibrium is reached that may consist of Fermi–Dirac distributions with several 100K above the lattice temperature [52,53].

As already outlined in Section 48.4, it is generally necessary to set up equations of motion not for the total carrier density N but for individual carrier populations of different single particle states $f_\alpha^c = \langle c_\alpha^\dagger c_\alpha \rangle$ to describe these effects. Such so-called "kinetic equations" together with the laser equations, Equations 48.10 through 48.20, enable us to treat excitation kinetics and turn-on dynamics of semiconductor QD-based laser systems and will be the topic of this section.

The question of what timescales and by which processes the carriers approach (quasi-) equilibrium conditions (if at all) influences the design of laser devices, as the answer to these questions determines, e.g., the dynamical response and turn-on delay [53–57]. While rate equations can also describe the turn-on dynamics for conventional laser devices, where carrier dynamics takes place on a fs to ps timescale, whereas the laser turn-on happens on the order of 10 ns, the situation is very much different in nanocavities with large Purcell enhancement, i.e., the enhancement of the spontaneous emission rate due to the presence of the cavity, where the turn-on delay can be on the order of 10 ps [53,58] and thus on the same timescale as the carrier scattering. The use of population function both for QD and WL states also allows one to describe effects such as spectral hole burning, or the difference between optical and injection pumping, that are unavailable to a theoretical model that tracks only the dynamics of the total carrier density N.

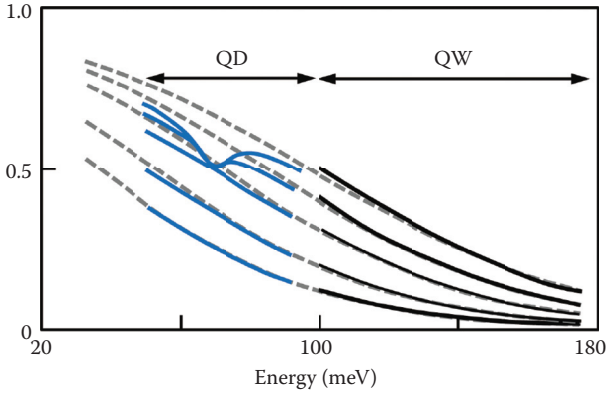


FIGURE 48.13 Carrier populations in a QD+WL system under laser operation for increasing pump intensity. The gray lines describe the population of QD carriers (subject to inhomogeneous broadening), the black lines show the populations of the WL states. For high pump rates the "hole" burnt into the population function by laser operation is clearly visible. The shaded-dashed lines are corresponding Fermi–Dirac functions. (From Chow, W. W. et al., *Light: Science & Applications*, 3, e201, 2014. With permission.)

An example of such carrier population functions are shown in Figure 48.13. The seemingly continuous features of the discrete QD single-particle state populations originate from the large number of emitters in the inhomogeneously broadened ensemble, cf. Section 48.8.

The simplest equation for such state-dependent carrier dynamics is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} f_\alpha = (1 - f_\alpha)\Gamma_\alpha^{\mathrm{in}} - f_\alpha \Gamma_\alpha^{\mathrm{out}}, \tag{48.25}$$

where $\Gamma_\alpha^{\mathrm{in}}$ and $\Gamma_\alpha^{\mathrm{out}}$ give the in- (out-) scattering rates into (out of) the state α. A more detailed version of Eq. (48.25) has the structure

$$\frac{\mathrm{d}}{\mathrm{d}t} f_\alpha = \sum_\beta f_\beta (1 - f_\alpha)\gamma_{\beta\to\alpha} - f_\alpha (1 - f_\beta)\gamma_{\alpha\to\beta} \tag{48.26}$$

and describes that the population of state α is enlarged by scattering from β to α and reduced by scattering from α to β. The Pauli-blocking terms $f_\beta(1 - f_\alpha)$ and $f_\alpha(1 - f_\beta)$ occur as scattering has to be proportional to the occupancy f of the initial and to the non-occupancy $(1 - f)$ of the final state. Under continuous excitation (CW laser), it may suffice to restrict the description to the s-shell and p-shell states of the QD and to use phenomenological scattering rates $\gamma^{e,h}_{\beta\to\alpha}$, in which case we recover Equation 48.21 through 48.24 in Section 48.4.

An often used approximation that is generally of relevance if not only QD states are discussed but also the continuous states of the WL or barrier are taken into consideration is the so-called relaxation time approximation

$$\frac{d}{dt} f_\alpha = \frac{f_\alpha - F_\alpha(N, T)}{\tau}, \tag{48.27}$$

which describes the relaxation of a nonequilibrium carrier population f_α toward a Fermi function $F_\alpha(N, T)$ at temperature T and carrier density N on a time-scale τ. While this approximation often works well, it also necessitates to determine the time-scale τ and the carrier temperature T. The last point is not trivial as in laser devices the carriers typically posses a higher temperature than the lattice [52,53]. To answer these

questions and also to investigate situations where the relaxation time approximation (Equation 48.27) is not applicable we have to take the microscopic origin of the scattering process into account. In the following, we discuss the dominant physical mechanisms behind the carrier dynamics, which are carrier-phonon scattering and carrier–carrier Coulomb scattering.

48.5.1 Carrier–Phonon Interaction

As discussed earlier, carrier temperature is an important factor for laser operation as it controls the population inversion. Hence the optical gain depends strongly on the temperature of the carrier system [59,60]. As the carrier–carrier Coulomb interaction cannot dissipate energy, the source for thermalization of the carriers toward the lattice temperature is the interaction with phonons. To investigate this interaction mechanism, we first have to determine which phonon modes are responsible for the carrier dynamics. A main difference between different phonon branches is whether they are longitudinal or transversal in nature. One can show [61] that due to their ability to form polarizations, mainly longitudinal phonons couple to the carrier system and that for most situations the interaction with transversal phonons is negligible. This leaves us with two phonon branches that are most important for the carrier-phonon interaction. These are the longitudinal optical (LO) phonons and longitudinal acoustic (LA) phonons. The speed of propagation of an acoustic phonon, which is also the speed of sound in the lattice, is given by the slope of the acoustic dispersion relation, $\frac{\partial\omega_k}{\partial k}$. At low values of k, the dispersion relation is almost linear, and the speed of sound is approximately independent of the phonon wavenumber.

Even though we are interested in laser structures, where the active material is composed of nanostructures (QDs), it is a good approximation to take the phonon modes and interaction potentials of the surrounding bulk material. This is the case, as we are mostly interested in long-wavelength phonons where the phonon wavelength is much larger than the size of the nanostructure we are investigating.

The interaction of carriers with LO phonons can be described by the Fröhlich coupling [61]. As this coupling mechanism is Coulombic in nature, the corresponding interaction matrix elements drop off rapidly with increasing momentum transfer $\mathbf{q}$. Therefore a often used approximation consists of treating the LO phonons as dispersionless with constant frequency $\hbar\omega_{\mathbf{q}} = \hbar\omega_{\mathrm{LO}}$ [62]. For acoustic phonons, the main interaction mechanisms are the deformation potential coupling and the piezoelectric coupling [62]. While these play an important role in the line broadening of QD lines in the low temperature and low carrier-density regime, they do not provide an efficient source for carrier scattering.

The interaction matrix elements for the interaction of carriers with LO-phonons are [60,63]

$$|M_{\alpha\beta}|^2 = \frac{M_{\mathrm{LO}}^2}{e^2/\varepsilon_0} V_{\alpha\beta\alpha\beta}, \tag{48.28}$$

containing Coulomb matrix elements $V_{\alpha\beta\alpha\beta}$ (see Equation 48.33 and Reference [64]), the elementary charge e and the vacuum permittivity ε_0. The prefactor $M_{\mathrm{LO}} = 4\pi\alpha\frac{\hbar}{\sqrt{2m}}(\hbar\omega_{\mathrm{LO}})^{\frac{3}{2}}$ includes the polar coupling strength α and the reduced mass m. The matrix elements are given here in a way that explicitly shows the Coulombic nature of the carrier-LO-phonon coupling.

A first approach toward carrier scattering is given by the Boltzmann equation [40,61]. For carrier-phonon scattering, this is given by Equation 48.26 with

$$\gamma_{\beta\to\alpha} = \sum_q M_{\beta\alpha}(q)\Big((1+n_q)\delta(\epsilon_\beta - \epsilon_\alpha - \hbar\omega_{\mathrm{LO}}) + n_q\delta(\epsilon_\beta - \epsilon_\alpha + \hbar\omega_{\mathrm{LO}})\Big). \tag{48.29}$$

The terms proportional to n_q and $(1 + n_q)$ describe phonon absorption and emission processes. In QD systems, additional effects exist that are not described within the framework of a Boltzmann equation, but are closely connected to carrier scattering, as they stem from the same microscopic interaction mechanism. These are quasi-particles and non-Markovian effects. The main reason is that the carrier-phonon interaction leads to the formation of a new quasi-particle. The quasi-particle obtained by dressing the carriers with the carrier-phonon interaction—the polaron—describes the lattice distortion accompanying the electron in its motion. In the framework of Green's functions (GFs), polarons are described using the retarded GF Kadanoff–Baym equation [63]:

$$\left[i\hbar\frac{\partial}{\partial\tau} - e_\alpha^a\right] G_\alpha^{a,R}(\tau) = \delta(\tau) + \int d\tau' \; \Sigma_\alpha^{a,R}(\tau - \tau')\, G_\alpha^{a,R}(\tau'). \tag{48.30}$$

The corresponding retarded selfenergy in random-phase approximation (RPA) [62] is given by

$$\Sigma_\alpha^{a,R}(\tau) \;=\; i\hbar \sum_\beta \; |M_{\alpha\beta}|^2 G_\beta^{a,R}(\tau)\, d^<(-\tau), \tag{48.31}$$

where the phonon propagators $d^{\gtrless}$ contain the phonon frequency and the phonon population [63]. The spectral function $\hat{G}_\alpha(\hbar\omega) = 2\mathrm{Im}(G_\alpha^{a,R}(\hbar\omega))$, which follows from the retarded GF, can be seen as a generalization of the single particle energy, i.e., a free particle has a spectral function of the form $\hat{G}_\alpha(\hbar\omega) = \delta(\epsilon_\alpha - \hbar\omega)$. In contrast, the Polaronic spectral functions but also posses a finite spectral width that is connected to the lifetime of the respective quasi-particle.

Using the quasi-particles as described by the polaron retarded GF, we can formulate a quantum-kinetic equation for polaron scattering:

$$\frac{\partial f_\alpha^a(t)}{\partial t} = 2\,\mathrm{Re}\sum_\beta \int_{-\infty}^{t} dt' \; |M_{\alpha\beta}|^2 G_\beta^{a,R}(t,t') \left[G_\alpha^{a,R}(t,t')\right]^* \\ * \left\{ \left[f_\beta^a(t')(1 - f_\alpha^a(t'))\right] d^>(t',t) - \left[f_\alpha^a(t')(1 - f_\beta^a(t'))\right] d^<(t',t) \right\}, \tag{48.32}$$

which contains scattering by phonon emission and absorption processes just as in the Boltzmann equation. Effects included here that go beyond the Boltzmann treatment are a weakening of the exact energy conservation due to the polaron properties (finite spectral width) and due to non-Markovian effects described by the explicit dependence of the time change of f_α^a at time t on all earlier times.

Results from such a quantum-kinetic treatment are shown in Figure 48.14. The QD level spacing is chosen to be 10% higher than the LO-phonon energy of 36meV. Therefore, a Boltzmann scattering equation predicts a complete inhibition of scattering due to the δ-functions in Equation (48.29). In contrast, the quantum-kinetic treatment shows that, due to quasi-particle effects (dotted line), scattering channels open up. The solid line shows the combined influence of quasi-particles and non-Markovian effects, that are caused by the time integrals in Equation (48.32). The inhibition of scattering as given by a Boltzmann equation, was predicted as a "phonon-bottleneck" early on [65,66], however, fast-carrier scattering was observed e.g., in Reference [67], showing the relevance of polaronic effects. For details see References [63,68,69].

48.5.2 Coulomb Interaction

The most important carrier scattering mechanism in the regime of high excitation densities, relevant for laser operation, is the carrier–carrier Coulomb interaction. It leads to a fast redistribution of carriers from

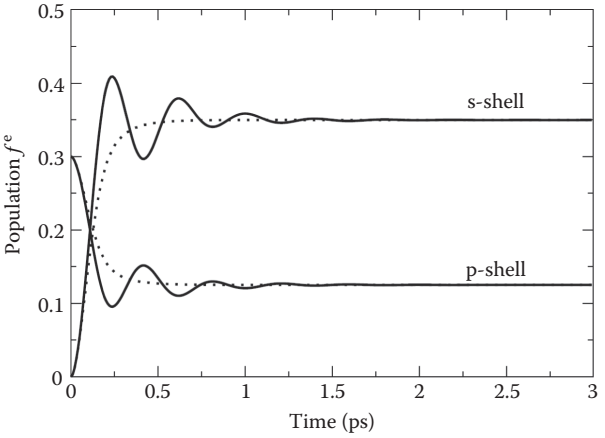


FIGURE 48.14 Redistribution of carrier occupations due to scattering of electrons from p-shell to s-shell in a QD via the carrier phonon interaction. In a Boltzmann-equation approach the occupations would remain constant due to a complete lack of scattering processes as described in the text (Reprinted from Seebeck, J. et al., *Physical Review B*, 71, 125327, 2005. With permission).

the spectral window where excitation takes place toward the spectral window from where they recombine. While Coulomb scattering cannot dissipate energy and thermalize the system (see Figure 48.12), it leads to equilibration toward a hot Fermi–Dirac function.

To determine the strength of the Coulomb interaction in nanostructures, we need to determine matrix elements of the Coulomb interaction in the single-particle basis given by, e.g., tight-binding descriptions of QDs. Using such single-particle wavefunctions, we can construct the Coulomb matrix elements

$$V_{\alpha\beta\gamma\delta} = \frac{1}{A} \sum_{\mathbf{q}} V_{\mathbf{q}} \left\langle \alpha \left| e^{-i\mathbf{q}\cdot\mathbf{r}} \right| \delta \right\rangle \left\langle \beta \left| e^{+i\mathbf{q}\cdot\mathbf{r}} \right| \gamma \right\rangle, \tag{48.33}$$

consisting of overlap integrals between single-particle wavefunctions and the Coulomb potential $V_{\mathbf{q}}$ [64]. A closer analysis of these Matrix elements also points towards a source for the efficiency of Coulomb scattering in nanostructures. Even though the available phase space of scattering partners is reduced significantly in QDs, compared to QW or bulk systems, this is partially balanced by enhanced interaction matrix elements due to the strong localization of the QD wave functions.

Like for the carrier-phonon scattering a first description of carrier–carrier scattering can be given by a Boltzmann-type equation

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f_\alpha(t) = & \frac{2\pi}{\hbar} \sum_{\beta\gamma\delta} \left(\left| W_{\alpha\gamma\delta\beta} \right|^2 - W_{\alpha\gamma\delta\beta} W^*_{\alpha\gamma\beta\delta} \right) \delta(\epsilon_\alpha - \epsilon_\beta + \epsilon_\gamma - \epsilon_\delta) \\ & \Big[f_\beta(t)(1 - f_\alpha(t)) f_\delta(t)(1 - f_\gamma(t)) - f_\alpha(t)(1 - f_\beta(t)) f_\gamma(t)(1 - f_\delta(t)) \Big], \end{aligned} \tag{48.34}$$

which contains screened interaction matrix elements $W_{\alpha\gamma\delta\beta}$ and an energy conserving δ-function. A feature of the carrier–carrier Coulomb scattering in QDs is that, while most scattering channels become extremely inefficient at low carrier density due to the population factors in Equation (48.34), some some scattering channels like electron–hole scattering are possible at very low carrier densities [70].

Results for the carrier–carrier Coulomb scattering in a QD laser under optical cw excitation are presented in Figure 48.15, where the carrier population functions is shown as a function of energy for different times. Discrete symbols denote the QD populations for s-shell and p-shell and the lines the continuous

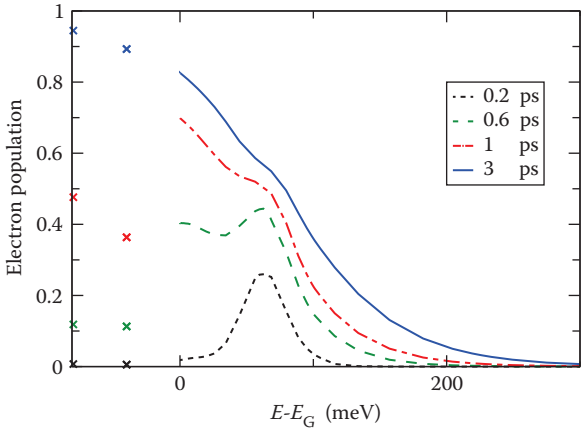


FIGURE 48.15 Carrier populations as a function of energy for electrons in a QD system under laser operation for different times. Symbols denote QD populations, lines WL populations. (Reprinted from Lorke, M. *Applied Physics Letters*, 99, 151110, 2011. With permission.)

WL states that are optically excited. After the initial Gaussian population profile has been excited by the pump, an ultra-fast redistribution is observed, that immediately starts to populate the QD states. It should be noted that, contrary to an often assumed situation [56,71], the timescales between the relaxation dynamics in the WL and the dynamics of the QD populations are not decoupled. On one hand, this is due to the efficiency of the initial in-scattering into the QD, as the QD states are initially empty. On the other hand, the relaxation within the WL is significantly slower than in a pure QW, as the capture into the QD states is most efficient for the WL states with low quasi-momenta. Therefore, these states are constantly depleted during the early stage of the kinetics, which slows down the relaxation of the WL distribution itself toward a quasi-equilibrium distribution [53] Furthermore, after the initial relaxation that lasts about 3 ps, a heating of the carrier population is observed on a timescale of 100 ps, as the pumping into the WL injects carriers at a higher energy than the emission energy.

These results also show (compared to those of the preceding section) that Coulomb and phonon scattering act on the same timescale in QD systems.

As for the phonons, quantum-kinetic effects lead to quasi-particle formation and non-Markovian effects. The corresponding quantum-kinetic equations can be written as

$$\frac{\partial f_\alpha^a(t)}{\partial t} = \frac{2}{\hbar}\mathrm{Re}\sum_{\beta\gamma\delta}\int_{-\infty}^{t} \mathrm{d}t' \Big(\left|W_{\alpha\gamma\delta\beta}\right|^2 - W_{\alpha\gamma\delta\beta}W^*_{\alpha\gamma\beta\delta}\Big) G_\beta^{a,\mathrm{R}}(t,t') \left[G_\alpha^{a,\mathrm{R}}(t,t')\right]^* * G_\delta^{a,\mathrm{R}}(t,t') \left[G_\gamma^{a,\mathrm{R}}(t,t')\right]^*$$
$$* \Big[f_\beta(t')(1-f_\alpha(t'))f_\delta(t')(1-f_\gamma(t')) - f_\alpha(t')(1-f_\beta(t'))f_\gamma(t')(1-f_\delta(t')) \Big]. \quad (48.35)$$

As for the Boltzmann-type equation for carrier scattering, this equation contains direct and exchange scattering, proportional to $\left|W_{\alpha\gamma\delta\beta}\right|^2$ and $W_{\alpha\gamma\delta\beta}W^*_{\alpha\gamma\beta\delta}$, respectively. The energy conserving δ-function is generalized by the product of retarded GFs just as for the carrier-phonon interaction and via the explicit time dependence on earlier times memory effects are included. While these equations have not been evaluated in detail due to their numerical complexity, it is expected that the included quasi-particle and non-Markovian effects would influence several properties of nanolasers. On one hand, large signal modulation and turn-on delay are expected to be influenced by non-Markovian effects. On the other hand, quasi-particle effects can, as for the carrier-phonon interaction, alter the scattering rates

influencing also the modulation response and threshold current. In the photon-assisted polarization, Equation 48.13 exist analogous expressions to the scattering contributions given in Equation 48.35 that lead to excitation-enhanced dephasing

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle b^{\dagger}v_{\alpha}^{\dagger}c_{\alpha}\rangle(t) = \int_{-\infty}^{t} \mathrm{d}t'(-\Gamma_{\alpha}^{\mathrm{DD}}(t,t'))\langle b^{\dagger}v_{\alpha}^{\dagger}c_{\alpha}\rangle(t') + \sum_{\beta}\Gamma_{\alpha\beta}^{\mathrm{OD}}(t,t')\langle b^{\dagger}v_{\beta}^{\dagger}c_{\beta}\rangle(t'), \tag{48.36}$$

where Γ^{DD} and Γ^{OD} consist of interaction matrix elements W, retarded GFs $G^{a,\mathrm{R}}$ and population functions f_{α} in a similar way as they appear in Equation 48.35 for the carrier scattering. We have investigated the influence of quasi-particle and non-Markovian effects on the quasi-classical analogues of these quantities in detail to study optical gain spectra of QD systems [8,60,72–75] and we expect an influence of these contributions on the turn-on behavior and the modulation dynamics [38].

48.6 Connecting to Experiments: Transition to Lasing and the β Factor

As first example, we discuss characteristic input–output curves and statistical properties of a Purcell-enhanced QD nanolaser device. In Figure 48.16 results from the semiconductor laser model are shown for various values of the β factor and fixed number of QD emitters. For $\beta = 1$, all spontaneous emission is

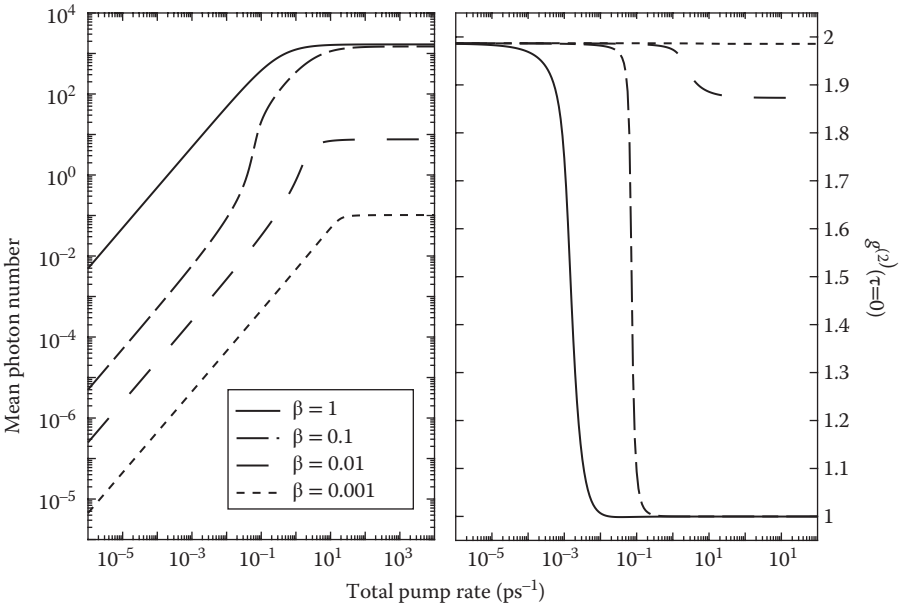


FIGURE 48.16 Input–output curves (left) and autocorrelation function $g^{(2)}(0)$ (right) for a nanolaser with 300 QD emitters. Curves are shown for different values of the β factor. For $\beta = 1$, a visible threshold is absent in the input–output curve, whereas $g^{(2)}(0)$ shows a clear transition from thermal to coherent emission. Note that the jump from below to above threshold is larger than in the results obtained from rate equations (Figure 48.3) and does here no longer scale with $1/\beta$. For $\beta = 0.1$, a soft kink appears in the input–output curve. For smaller values of β, the limited number of QD emitters cannot provide sufficient gain to reach lasing before saturation is reached. A cavity-Q factor of 32,000 and a cavity-enhanced spontaneous emission rate of $\tau_{\mathrm{sp}} = 50$ ps has been used.

directed into the laser mode. In contrast to the rate-equation result that we have discussed in the context of Figure 48.3, saturation is visible at strong pumping. This effect is important in QD nanolasers, since the small size of the resonator naturally limits the amount of gain material that can be brought into the cavity. A monotonous increase of the output intensity with pumping, as given by the rate equations (Equation 48.1), would be unphysical. As a consequence, for smaller values of β, the gain can become insufficient to reach lasing before saturation sets in. In the example in Figure 48.16, this is visible for $\beta = 0.01$, where the threshold is reached, but the emission remains largely thermal with $g^{(2)}(0) \approx 1.9$, and for $\beta = 0.001$, where the QD emission saturates long before the laser threshold at a low mean intracavity photon number of 0.1. Access to the photon autocorrelation function $g^{(2)}(0)$ provides an important tool in judging the emission properties far beyond what can be inferred from the emission intensity alone and is a significant advancement over the rate equations.

Deeper insight into cavity-QED effects in nanolasers can be obtained from the emission and correlation *dynamics*. The top panel in Figure 48.17 shows time resolved the emission pulse (solid curves) and autocorrelation function $g^{(2)}(t, \tau = 0)$ (symbols) following picosecond-short optical excitation of the nanolaser. Also here the high time resolution has been realized by using a streak-camera setup. Following a sufficiently strong excitation pulse, the statistical properties of the emission pulse change as function of time from thermal to coherent and back to thermal. Coherent emission is reached around the peak region of the emission pulse and forms a plateau when the excitation level is above the threshold region. The data agrees well with time-dependent results of the laser model (bottom panel), which are readily available from the time evolution of the coupled differential equations presented in Section 48.4. In contrast to conventional measurements that integrate over the emission pulse, here it is possible to track the degree of coherence during the pulse and to provide an understanding of the coherence it contains. In their combination, state-of-the art measurements and microscopic models can reveal new aspects of solid-state light sources, reflecting the emission dynamics and the dynamics of quantum-mechanical correlations down to the few-photon level. The explicit treatment of dynamical properties of nanolasers, such as carrier dynamics, turn-on delay, and modulation response requires to go beyond the simple approach in Equations 48.21 through 48.24 for the carrier scattering and use the more involved methods of Section 48.5 due to the sensitivity of these quantities [55].

48.7 Interemitter Coupling Effects and Superradiance in Nanolasers

When several emitters couple to a common light field, the interaction introduces correlations between them. The eigenstates of the interacting system are then no longer given by those of the individual systems, but by so-called Dicke states of the collective system. In 1954, Dicke has shown [76] that from certain collective states, emission can be enhanced or suppressed, which is referred to as superradiance and subradiance. In particular, for the Dicke state of the half inverted system of N two-level emitters, the recombination rate is proportional to $N(N-1)$ [77]. This quadratic behavior with emitter number has become one of the hallmarks of superradiance. Since its discovery by Dicke, superradiance has been extensively and continuously studied in a variety of systems. Most commonly, however, superradiance is associated with temporal modifications of the emission in spatially extended systems, such as clouds of atoms [78] or, when it comes to semiconductor systems, excitons in extended systems [79].

In an ensemble of QDs that are embedded in a microcavity, depending on the inhomogeneous broadening of the QDs and the cavity Q factor, a fraction of the QDs overlaps both spectrally and spatially with the mode. The common light field these resonant solid-state emitters are subjected to provides a mechanism to form inter-emitter correlations that can lead to sub- and superradiant effects in QD nanolasers. Laser models typically assume that emitters act individually and neglect such effects. In the following, we outline how to use the formalism introduced in Section 48.4 to formulate a laser theory that includes inter-emitter coupling in an approximative way and give a quantitative discussion of its effects. As it turns out, their

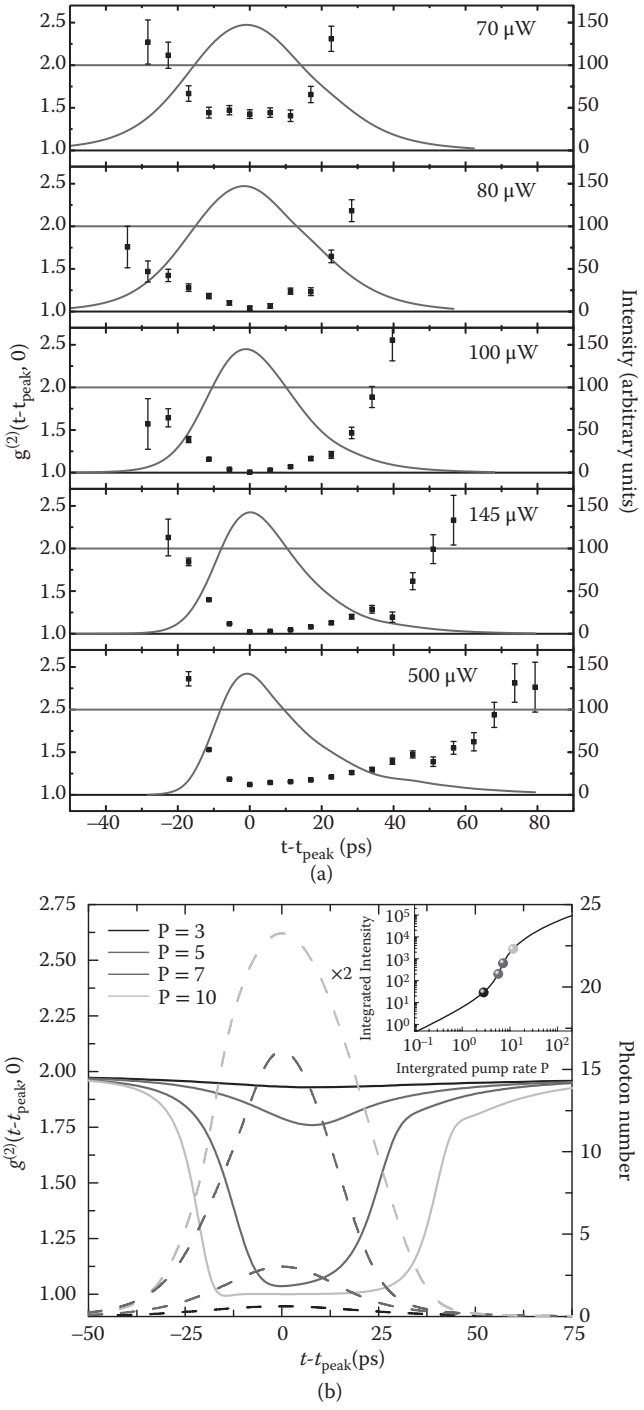


FIGURE 48.17 Time-resolved measurements (a) and theoretical results (b) for emission intensity and photon autocorrelation function following picos-second short optical excitation of a QD-micropillar laser. (Reprinted from Aßmann, M. et al., *Physical Review B* 81, 165314, 2010. With permission).

impact on both stationary and dynamical laser properties can be quite significant. A detailed description of the formalism is found in [80].

To this end, we introduce a formulation for operators that act on the electronic QD states in terms of the many-particle configurations $|i\rangle$, where i refers to configurations like ground state, exciton, trion, biexciton, various dark states, etc. These configuration states, some of which are shown in Figure 48.10, form the basis of the single-QD subspaces of the Hilbert space. Operators Q_{ij}^{α} are introduced that describe either transitions from configuration j to configuration i for $i \neq j$, or the probability that configuration i is realized for $i = j$ in QD α. This approach has advantages over descriptions that use single-particle operators c_ν and v_ν, like we did in Section 48.4. In doing so, many-electron configurations of one emitter are addressed by a single operator Q rather than many single-particle creation and annihilation operators. It further simplifies the systematic approximation scheme used to truncate the hierarchy of equations of motion with respect to correlations involving *different* QDs. Interemitter correlations are described by a hierarchy of expectation values $\langle Q_{ij}^{\alpha} Q_{kl}^{\beta} \rangle$, $\langle Q_{ij}^{\alpha} Q_{kl}^{\beta} Q_{mn}^{\gamma} \rangle$, … that involve transitions in QDs α and β, or QDs α, β and γ, etc. The cluster expansion approach facilitates a truncation at a certain order of interemitter coupling, e.g. on the level of pair-correlations between emitters given by $\langle Q_{ij}^{\alpha} Q_{kl}^{\beta} \rangle$ and illustrated in Figure 48.18. All quantum-mechanical expectation values containing electronic QD operators acting on more than two emitters are then neglected. On this level, which is the lowest order that contains interemitter coupling effects, a system of N identical QDs with four confined states each is described by about 300 coupled differential equations, also including photon correlations up to the fourth order providing access to the autocorrelation function $g^{(2)}(0)$. The generation of the equations of motion can be assisted by using computer algebra [80]. While the formalism may appear to involve a great deal of effort, it is highly efficient in the numerical evaluation, and computations for many emitters can typically be performed on a single workstation.

The impact of interemitter coupling on the steady-state properties of a continuously driven QD nanolaser can be seen in Figure 48.19. In the top panel, input–output curves are shown that have been obtained from two separate calculations, one including and one omitting QD–QD correlations. C_F in the bottom panel visualizes the difference in the output intensity I, i.e.,

$$C_F = \frac{I_{\text{rad. coupled QDs}}}{I_{\text{independent QDs}}} - 1. \tag{48.37}$$

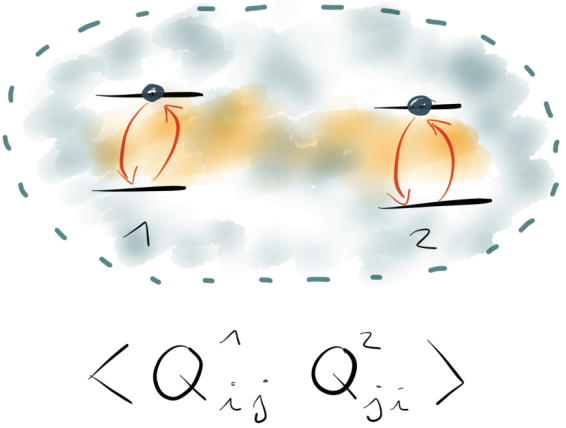


FIGURE 48.18 Illustration of photon-mediated inter-emitter dipole correlations between two emitters.

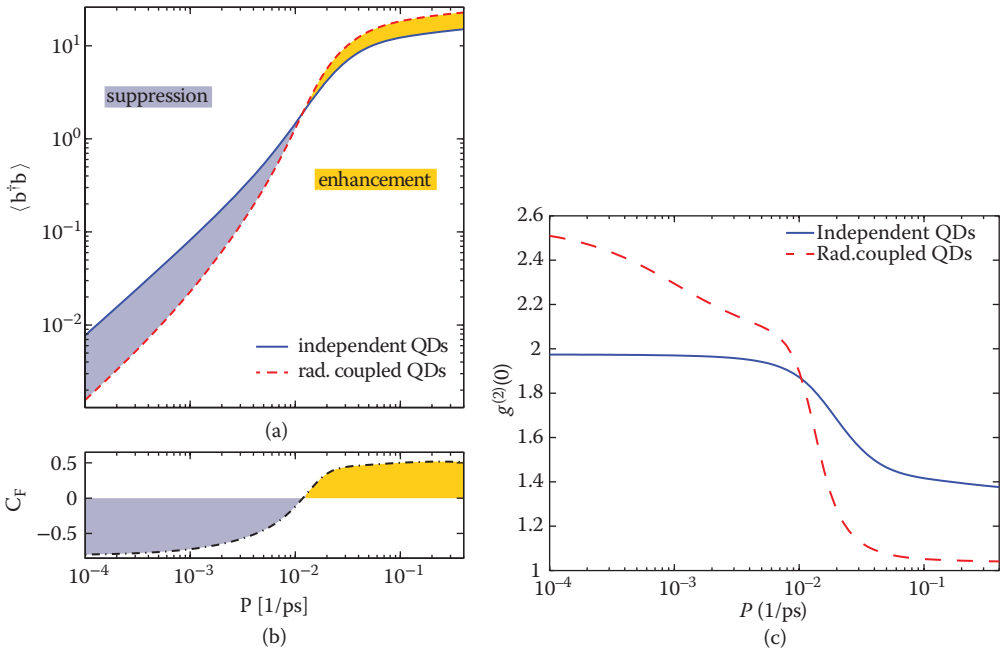


FIGURE 48.19 (a) Input–output curve for a QD nanolaser with typical specifications: 100 resonant emitters, cavity loss rate $\kappa = 0.05/\text{ps}$ (corresponding to $Q = 20,000$), carrier relaxation rate $\gamma_r = 0.05/\text{ps}$, and spontaneous losses into leaky modes $\gamma_{spont} = 0.01/\text{ps}$. (b) The cooperativity factor visualizes the difference caused by the radiative coupling. (Adapted from Leymann, H. A. M. et al., *Physical Review Applied*, 4, 044018, 2015. With permission.) (c) Corresponding autocorrelation functions. At low excitation, superthermal bunching is a signature of the subradiant emission regime. At high excitation, the threshold to lasing is not fully crossed if emitters act individually and radiative coupling is neglected. Dipole–dipole correlations between distant emitters create additional coherence in the superradiant regime.

The possibility to "switch" the coupling on an off is a particular advantage of the equation-of-motion method and allows to directly assess the impact of the radiative coupling. The following effects are observed:

1. In the low-excitation regime, the inter–emitter coupling creates a subradiant state with a reduced spontaneous emission rate and reduced photon output. Once the laser threshold is crossed, a superradiant state is formed which enhances the laser output *in addition* to stimulated emission provided by photons in the cavity. In sum, the effect on the input–output curve over the whole excitation regime is quite significant. In particular, in the given example, which has been calculated for a typical microcavity laser with 100 QD emitters as active medium, one would underestimate the true β factor by one order of magnitude when ignoring the radiative coupling
2. The input/output characteristics exhibit a different and untypical slope below threshold.
3. In the subradiant regime (low emission intensity), dipole correlations between pairs of QDs cause an increased probability to emit photons synchronously, leading to super-thermal bunching behavior in the statistical properties of the emission, as shown in the right panel of Figure 48.19.
4. Due to inter-emitter coupling, fewer QDs are required to reach the lasing threshold. In small systems, where achieving sufficient gain due to the limited number of emitters in the cavity has always been an issue, the presence of radiative coupling can explain why lasing is more easily reached than expected from conventional laser theories that do not include QD–QD correlations.

5. Coherent emission with $g^{(2)}(0) = 1$ is achieved at lower intracavity mean photon number in the presence of interemitter coupling. This may be interpreted as an increase in the "coherence per photon" due to the alignment of dipole correlations between emitters.

Interestingly, these theoretical predictions are hard to verify experimentally. A strong indicator for the presence of a dipole-correlated phase is the super-thermal photon bunching [81], which has recently been observed in time-resolved studies using a streak-camera setup in combination with microscopic theory [82]. In the same study, highly accelerated spontaneous emission and subradiant photon trapping have been demonstrated. In their sum, the observation of these criteria give a convincing account of the presence of strong inter-emitter coupling effects in a typical QD nanolaser.

48.8 Nonresonant Coupling and Lasing from Multiexciton States in Few-QD Systems

An inherent property of self-assembled QDs that are grown in the so-called *Stranski–Krastanov* mode [83] is inhomogeneous broadening. It refers to inhomogeneity in size, shape and material composition from one QD to the next, causing differences in the confinement potential and with it the electronic single-particle states. While the typical linewidth of QD transitions is tens to hundreds of μeV depending on temperature [2,84,85], the line of an ensemble of self-assembled QDs consists of the spectral lines of the individual emitters, which for many emitters then appears as a single line with a typical broadening of 20–50 meV. An illustrative demonstration of this behavior is given in Figure 48.20 [86]. Inhomogeneous broadening is generally seen as a weak point of self-assembled QDs, as it takes away some of the advantage of the narrow and well-defined transition energies that is characteristic for the single emitter. On the other hand, in cavity-QED it simplifies the task of creating spectral overlap between emitters and a cavity mode (cf. Figure 48.4) and compensates for spectral wandering, since the ensemble broadening is large compared to the cavity linewidth (typically 50–500 μeV). The influence of detuning of emitters in the ensemble and a cavity mode is greatly reduced by nonresonant coupling mechanisms discussed later. The influence of inhomogeneous broadening on QD laser properties has been studied in the framework of a QD laser theory in [27].

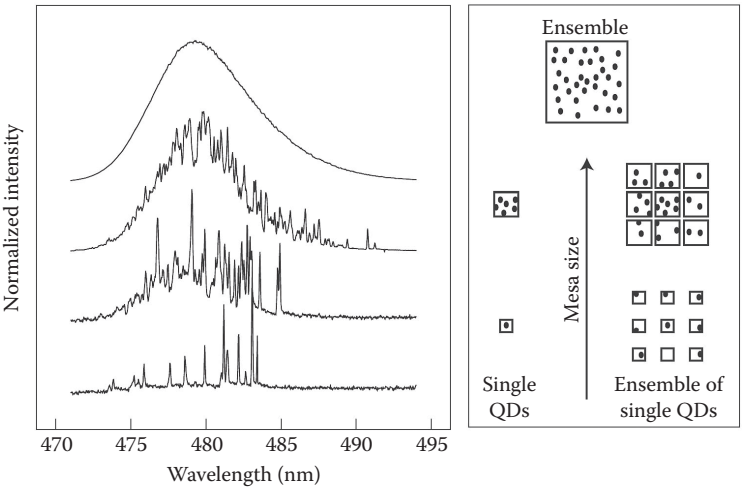


FIGURE 48.20 Experimental results illustrating inhomogeneous broadening in a quantum dot (QD) ensemble. With decreasing mesa size, the QD ensemble consists of fewer and fewer emitters (illustrated right). The broad inhomogenous emission spectrum for many emitters (top left) is revealed to consist of discrete lines as the number of emitters is reduced (bottom left). (Reprinted from Scheibner, M. et al., *Nature Physics*, 3, 106–110, 2007. With permission.)

In this section, we look into what happens if the resonator contains only a few emitters, and the gain spectrum consists of a multitude of separate sharp lines. Such a situation is depicted in the left panel of Figure 48.21 [87]. Then, the emission properties are determined by the interplay of various many-particle configurations of each QD emitter, and their contribution to the photon production depends on the relative spectral position to the cavity mode. While at first it seems contradictory that lasing can be achieved in such a system, strong emission enhancement can drive even a single emitter close to the regime of lasing. At the same time, nonresonant coupling mechanisms haven been identified to provide means for a detuned QD to emit photons into the cavity mode. Nonresonant cavity feeding plays a significant role in nanolasers with ensembles of emitters, but a quantification of its effect is nearly impossible to obtain in such a system. The few-emitter limit, however, offers the unique possibility to study non-resonant mode coupling in a highly controllable environment. In fact, the underlying physical process has long been elusive. Intensive research on this topic has identified three mechanisms that are responsible for the effect:

1. Accoustic phonons can bridge small energy gaps in the range of 1 meV between a detuned emitter resonance and the cavity mode [88–93]. The efficiency is determined by the imbalance between phonon-assisted cavity-photon emission, and the reverse process. Low temperature favors this asymmetry, because low phonon population makes the emission of a phonon more likely than its absorption. Therefore, phonon-mediated nonresonant coupling is more efficient for coupling blue-detuned emitter transitions to the mode [94].
2. At higher carrier densities, interactions with the quasi-continuous WL states provide means for Auger-like scattering processes that allow to bridge larger energy gaps of up to 10 meV [95–99]. While their efficiency is small in comparison to the ultrafast intraband relaxation [53,64,100–102], Auger-coupling is of central importance in the high-excitation regime of nanolasers.
3. QDs possess a rich electronic structure that gives rise to a multitude of multiexciton states, as illustrated in Figure 48.10. The Coulomb interaction energetically separates transitions from different multi-exciton states. This energetic separation, and the possibility of interband recombinations from higher confined states (e.g., the p shell) covers an even larger energetic window, as has been shown

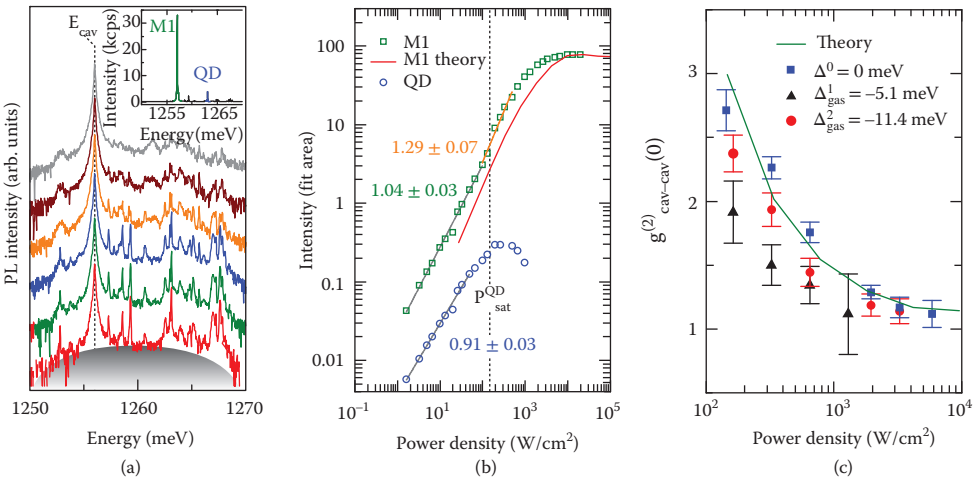


FIGURE 48.21 **(a)** Spectra of a few-quantum dot (QD) photonic-crystal cavity laser for excitation energies ranging from 0.14 to 5.9 kW/cm^2. **(b)** Input–output curves at the cavity-mode energy (squares) and the QD exciton (circles), the latter showing saturation while the cavity-mode emission continues to increase. **(c)** Second-order photon correlation function at the cavity mode for three different detunings of the cavity-mode energy. Irrespective of the detuning between QD transitions and mode, lasing is achieved at sufficiently high excitation. PL, photoluminescence. (Reprinted from Lichtmannecker, S. et al., *Scienti. Rep.*, 7, 2017).

in [99,103,104]. Thus, even if the QD exciton line, which dominates the spectrum at low excitation, is detuned from the cavity mode, a higher transition, e.g., from the biexciton, may be in perfect resonance with the mode and lead to efficient photon production at higher excitation.

There are several approaches to include the non-resonant coupling of detuned emitters in a laser theory, which superficially all have the same effect of producing additional photons in the resonator mode. In their origin they differ fundamentally, which is reflected in the saturation behavior and in the statistical properties of the emission. On the one hand, a "direct" cavity feeding is often used in the literature [105–108], where photons are directly generated in the cavity mode by an inverse cavity loss term. Such a feeding mechanism, acting together with the cavity loss, gives rise to a thermal photonic state with a temperature defined by the cavity-pump-to-loss ratio. This overlooks the fact that the background photons are generated through exciting an active medium, and hence are expected to have a coherent component [89]. On the other hand, detuned emitters can be explicitly included in the calculation, and their coupling to a detuned cavity mode via phonon or Auger coupling is facilitated by an effective system-bath coupling (Lindblad) term.This approach has the advantage that properties of the gain medium are preserved. The resulting cavity feeding is strongly non-linear with pump rate and exhibits saturation if the number of detuned emitters is limited. Furthermore, the overall emission can become coherent if the background contributions drive the laser above the threshold [89].

As an example, the left part of Figure 48.21 [109] shows the spectrum of a few-QD photonic-crystal nanocavity laser. Only few discrete QD lines are found in the vicinity of the mode (M1). The input–output characteristics is that of a nanolaser with a slight kink, hinting at a β-factor close to unity (middle panel), and coherent emission with $g^{(2)}(0) = 1$ being reached (right panel). Remarkably, the performance of the device is insensitive to tuning of the cavity-mode energy, which can be realized by gas deposition on the photonic crystal cavity. Evidence for this is found in the autocorrelation function in the right panel, which shows a clear transition to lasing with $g^{(2)}(0) = 1$ *independently* of the exact cavity position. In a system with few discrete emission channels this provides strong evidence for the importance of nonresonant coupling effects. Similar results have been reported in [110], to which the authors referred to as "self-tuning gain" effect. Moreover, in the spontaneous emission regime the coupling of various multiexcitonic emission channels from different emitters inside the cavity leads to super-thermal photon bunching with $g^{(2)}(0) > 2$. A density-matrix theory including quantum-mechanical correlations between these transitions, shown as green curve in the right panel, is in good agreement with the measured data. Neither rate equations, nor conventional laser models can unravel the intricate quantum-mechanical nature that governs the laser properties of few-emitter cavity-QED systems.

48.9 Outlook

With present-day nanolaser devices, we are closer than ever to reaching fundamental limits of light–matter interaction in semiconductor systems. Electronic excitations can be generated and used on the level of single excitons [111,112]. Brought into interaction with a single mode of microresonator, semiconductor nanolasers operate at the transition to the regime of quantum optics, where cavity-QED effects, such as photon blockade, photon-antibunching, and vacuum Rabi oscillations can coexist with lasing in different excitation regimes of the same device [106,113,114]. This offers fascinating prospects both from a fundamental and an applied point of view.

Acknowledgments

The authors would like to thank Joachim Piprek for the kind invitation to write this chapter. We also thank Andreas Beuthner for his help in creating the schematic figures. Christopher Gies and Frank Jahnke acknowledge financial support from the German Science Foundation (DFG).

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. WWC thanks the hospitality of the Technical University Berlin and travel support from the German Research Foundation via the collaborative research center 787.

References

1. Vahala, K. J. (2003). Optical microcavities. *Nature 424* (6950), 839–846.
2. Borri, P., W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg (2002). Relaxation and dephasing of multiexcitons in semiconductor quantum dots. *Physical Review Letters 89* (18), 187401.
3. Borri, P., W. Langbein, U. Woggon, V. Stavarache, D. Reuter, and A. D. Wieck (2005). Exciton dephasing via phonon interactions in InAs quantum dots: Dependence on quantum confinement. *Physical Review B 71* (11), 115328.
4. Florian, M., A. Steinhoff, C. Gies, and F. Jahnke (2016). Scattering-induced dephasing of many-particle transitions in semiconductor quantum dots. *Applied Physics B 122* (1), 1–7.
5. Blood, P. (2013). The laser diode: 50 years on. *IEEE Journal of Selected Topics in Quantum Electronics 19* (4), 1503201–1503201.
6. Ning, C. (2013). What is laser threshold? *IEEE Journal of Selected Topics in Quantum Electronics 19* (4), 1503604–1503604.
7. Rice, P. R. and H. J. Carmichael (1994). Photon statistics of a cavity-QED laser: A comment on the laser–phase-transition analogy. *Physical Review A 50* (5), 4318–4329.
8. Chow, W. W., M. Lorke, and F. Jahnke (2011). Will quantum dots replace quantum wells as the active medium of choice in future semiconductor lasers? *IEEE Journal of Selected Topics in Quantum Electronics 17*, 1349–1355.
9. Yokoyama, H. and S. D. Brorson (1989). Rate equation analysis of microcavity lasers. *Journal of Applied Physical 66*, 4801.
10. Musiał, A., C. Hopfmann, T. Heindel, C. Gies, M. Florian, H. A. M. Leymann, A. Foerster, C. Schneider, F. Jahnke, S. Höfling, M. Kamp, and S. Reitzenstein (2015). Correlations between axial and lateral emission of coupled quantum dot micropillar cavities. *Physical Review B 91* (20), 205310.
11. Munsch, M., N. S. Malik, E. Dupuy, A. Delga, J. Bleuse, J.-M. Gérard, J. Claudon, N. Gregersen, and J. Mørk (2013). Dielectric GaAs antenna ensuring an efficient broadband coupling between an InAs quantum dot and a Gaussian optical beam. *Physical Review Letters 110* (17), 177402.
12. Leymann, H. A. M., C. Hopfmann, F. Albert, A. Foerster, M. Khanbekyan, C. Schneider, S. Höfling, A. Forchel, M. Kamp, J. Wiersig, and S. Reitzenstein (2013). Intensity fluctuations in bimodal micropillar lasers enhanced by quantum-dot gain competition. *Physical Review A 87* (5), 053819.
13. Agrawal, G. P. and N. K. Dutta (1993). *Semiconductor Laser*. Norwell, MA: Kluwer.
14. Chuang, S. L. (1995). *Physics of Optoelectronic Devices*. New York, NY: Wiley.
15. Lohmeyer, H., K. Sebald, C. Kruse, R. Kröger, J. Gutowski, D. Hommel, J. Wiersig, N. Baer, and F. Jahnke (2006). Confined optical modes in monolithic II-VI pillar microcavities. *Applied Physics Letters 88* (5), 051101.
16. McCall, S. L., A. F. J. Levi, R. E. Slusher, S. J. Pearton, and R. A. Logan (1992). Whispering gallery mode microdisk lasers. *Applied Physics Letters 60* (3), 289–291.
17. Joannopoulos, J. D., P. R. Villeneuve, and S. Fan (1997). Photonic crystals: Putting a new twist on light. *Nature 386* (6621), 143–149.
18. Purcell, E. M. (1946). Spontaneous emission probability at radio frequencies. *Physical Review 69*, 681.

19. Schwab, M., H. Kurtze, T. Auer, T. Berstermann, M. Bayer, J. Wiersig, N. Baer, C. Gies, F. Jahnke, J. P. Reithmaier, A. Forchel, M. Benyoucef, and P. Michler (2006). Radiative emission dynamics of quantum dots in a single cavity micropillar. *Physical Review B 74* (4), 045323.
20. Aßmann, M., F. Veit, M. Bayer, M. v. d. Poel, and J. M. Hvam (2009). Higher-order photon bunching in a semiconductor microcavity. *Science 325* (5938), 297–300. PMID: 19608912.
21. Loudon, R. (2000). *The Quantum Theory of Light*. Oxford, UK: Oxford University Press.
22. Hanbury Brown, R. and R. Q. Twiss (1956). Correlation between photons in two coherent beams of light. *Nature 177*, 27–29.
23. Ulrich, S. M., C. Gies, J. Wiersig, S. Reitzenstein, C. Hofmann, A. Löffer, A. Forchel, F. Jahnke, and P. Michler (2007). Photon statistics of semiconductor microcavity lasers. *Physical Review Letters 98*, 043906.
24. Gardiner, C. and P. Zoller (2004). *Quantum Noise A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*. Heidelberg, Berlin, New York: Springer Verlag.
25. Gies, C., M. Florian, P. Gartner, and F. Jahnke (2012). Modelling single quantum dots in microcavities. In F. Jahnke (Ed.), *Quantum Optics with Semiconductor Nanostructures*. Philadelphia, PA: Woodhead Publishing Limited.
26. Ates, S., C. Gies, S. M. Ulrich, J. Wiersig, S. Reitzenstein, A. Löffler, A. Forchel, F. Jahnke, and P. Michler (2008). Influence of the spontaneous optical emission factor β on the first-order coherence of a semiconductor microcavity laser. *Physical Review B 78* (15), 155319.
27. Chow, W. W., F. Jahnke, and C. Gies (2014). Emission properties of nanolasers during the transition to lasing. *Light: Science & Applications 3* (8), e201.
28. Ritter, S., P. Gartner, C. Gies, and F. Jahnke (2010). Emission properties and photon statistics of a single quantum dot laser. *Optics Express 18* (10), 9909–9921.
29. Reitzenstein, S., C. Böckler, A. Bazhenov, A. Gorbunov, A. Löffer, M. Kamp, V. D. Kulakovskii, and A. Forchel (2008). Single quantum dot controlled lasing effects in high-q micropillar cavities. *Optics Express 16* (7), 4848–4857.
30. Gies, C., J. Wiersig, and F. Jahnke (2008). Output characteristics of pulsed and continuous-wave-excited quantum-dot microcavity lasers. *Physical Review Letters 101* (6), 067401.
31. Chow, W. and S. Koch (2005). Theory of semiconductor quantum-dot laser dynamics. *IEEE Journal of Quantum Electronics 41* (4), 495–505.
32. Kolarczik, M., N. Owschimikow, J. Korn, B. Lingnau, Y. Kaptan, D. Bimberg, E. Schöll, K. Lüdge, and U. Woggon (2013). Quantum coherence induces pulse shape modification in a semiconductor optical amplifier at room temperature. *Nature Communications 4*, 2953.
33. Gies, C., J. Wiersig, M. Lorke, and F. Jahnke (2007). Semiconductor model for quantum-dot-based microcavity lasers. *Physical Review A 75* (1), 013803.
34. Wiersig, J., C. Gies, F. Jahnke, M. Aßmann, T. Bestermann, M. Bayer, C. Kistner, S. Reitzenstein, S. H. A. Forchel, C. Kruse, J. Kalden, and D. Hommel (2009). Direct observation of correlations between individual photon emission events of a microcavity laser. *Nature 460*, 245.
35. Baer, N., C. Gies, J. Wiersig, and F. Jahnke (2006). Luminescence of a semiconductor quantum dot system. *European Physical Journal B 50*, 411.
36. Gregersen, N., T. Suhr, M. Lorke, and J. Mørk (2012). Quantum-dot nano-cavity lasers with Purcell-enhanced stimulated emission. *Applied Physics Letters 100* (13), 131107.
37. Lermer, M., N. Gregersen, M. Lorke, E. Schild, P. Gold, J. Mørk, C. Schneider, A. Forchel, S. Reitzenstein, S. Hing, and M. Kamp (2013). High beta lasing in micropillar cavities with adiabatic layer design. *Applied Physics Letters 102* (5), 052114.
38. Lorke, M., T. Suhr, N. Gregersen, and J. Mørk (2013). Theory of nanolaser devices: Rate equation analysis versus microscopic theory. *Physical Review B 87*, 205310.

39. Suhr, T., N. Gregersen, M. Lorke, and J. Mørk (2011). Modulation response of quantum dot nanolightemitting-diodes exploiting Purcell-enhanced spontaneous emission. *Applied Physics Letters 98* (21), 211109.
40. Chow, W. and S. Koch (1999). *Semiconductor-Laser Fundamentals* (1st ed.). Berlin: Springer-Verlag.
41. Coldren, L. and S. Corzine (1995). *Diode Lasers and Photonic Integrated Circuits*. New York, NY: Wiley.
42. Binder, R., H. Köhler, M. Bonitz, and N. Kwong (1997). Green's function description of momentum orientation relaxation of photoexcited electron plasmas in semiconductors. *Physical Review B 55*, 5110.
43. Bonitz, M., D. Kremp, D. Scott, R. Binder, W. Kraeft, and H. Köhler (1996). Numerical analysis of nonmarkovian effects in charge-carrier scattering: One-time kinetic equations. *Journal of Physics: Condensed Mattter 8*, 6057.
44. ElSayed, K., S. Schuster, H. Haug, G. Herzel, and K. Henneberger (1994). Subpicosecond plasmon response: Buildup of screening. *Physical Review B 49*, 7337.
45. Henneberger, K. (1988). Resonant laser excitation and electron-hole kinetics of a semiconductor: I. nonequilibrium green's function treatment and fundamental equations. *Physica A 150*, 419.
46. Jahnke, F. and S. W. Koch (1995). Many-body theory for semiconductor microcavity lasers. *Physical Review A 52*, 1712.
47. Bányai, L., P. Gartner, and H. Haug (1998). Self-consistent RPA retarded polaron Green function for quantum kinetics. *European Physical Journal B 1*, 209.
48. Gartner, P., L. Bányai, and H. Haug (1999). Two-time electron-LO-phonon quantum kinetics and the generalized Kadanoff-Baym approximation. *Physical Review B 60*, 14234.
49. Gartner, P., L. Bányai, and H. Haug (2002). Self-consistent RPA for the intermediate-coupling polaron. *Physical Review B 66*, 75205.
50. Hartmann, M. and W. Schäfer (1992). Real time approach to relaxation and dephasing processes in semiconductors. *Physical Status Solidi (B) 173*, 165.
51. Haug, H. (1992). Interband quantum kinetics with LO-phonon scattering in a laser-pulse-excited semiconductor I. Theory. *Physical Status Solidi (B) 173*, 139.
52. Jahnke, F. and S. W. Koch (1993). Theory of carrier heating through injection pumping and lasing in semiconductor microcavity lasers. *Optics Letters 18* (17), 1438–1440.
53. Lorke, M., T. R. Nielsen, and J. Mørk (2011). Switch-on dynamics of nanocavity laser devices. *Applied Physics Letters 99* (15), 151110.
54. Lingnau, B., K. Lüdge, E. Schöll, and W. W. Chow (2010). Many-body and nonequilibrium effects on relaxation oscillations in a quantum-dot microcavity laser. *Applied Physics Letters 97*, 111102.
55. Lorke, M., T. R. Nielsen, and J. Mørk (2010). Influence of carrier dynamics on the modulation bandwidth of quantum-dot based nanocavity devices. *Applied Physics Letters 97* (21), 211106.
56. Luedge, K., M. J. P. Bormann, E. Malic, P. Hoevel, M. Kuntz, D. Bimberg, A. Knorr, and E. Schoell (2008). Turn-on dynamics and modulation response in semiconductor quantum dot lasers. *Physical Review B 78*, 035316.
57. Luedge, K. and E. Scholl (2009). Quantum-dot lasers—Desynchronized nonlinear dynamics of electrons and holes. *IEEE Journal of Quantum Electronics 45* (11), 1396–1403.
58. Aßmann, M., F. Veit, M. Bayer, C. Gies, F. Jahnke, S. Reitzenstein, S. Höfling, L. Worschech, and A. Forchel (2010). Ultrafast tracking of second-order photon correlations in the emission of quantum-dot microresonator lasers. *Physical Review B 81* (16), 165314.
59. Lorke, M., W. Chow, T. R. Nielsen, J. Seebeck, P. Gartner, and F. Jahnke (2006). Anomaly in the excitation dependence of optical gain in semiconductor quantum dots. *Physical Review B 74*, 035334.
60. Lorke, M., T. Nielsen, J. Seebeck, P. Gartner, and F. Jahnke (2006). Influence of carrier-carrier and carrier-phonon correlations on optical absorption and gain in quantum-dot systems. *Physical Review B 73*, 85324.

61. Schäfer, W. and M. Wegener (2002). *Semiconductor Optics and Transport Phenomena* (1st ed.). Berlin: Springer-Verlag.
62. Mahan, G. (1990). *Many-Particle Physics*. New York, NY: Plenum Press.
63. Seebeck, J., T. Nielsen, P. Gartner, and F. Jahnke (2005). Polarons in semiconductor quantum dots and their role in the quantum kinetics of carrier relaxation. *Physical Review B 71*, 125327.
64. Nielsen, T. R., P. Gartner, and F. Jahnke (2004). Many-body theory of carrier capture and relaxation in semiconductor quantum-dot lasers. *Physical Review B 69* (23), 235314.
65. Benisty, H. (1995). Reduced electron-phonon relaxation rates in quantum-box systems: Theoretical analysis. *Physical Review B 51*, 13281.
66. Inoshita, T. and H. Sakaki (1996). Electron-phonon interaction and the so-called phonon bottleneck effect in semiconductor quantum dots. *Physica B 227*, 373.
67. Kurtze, H., J. Seebeck, P. Gartner, D. R. Yakovlev, D. Reuter, A. D. Wieck, M. Bayer, and F. Jahnke (2009). Carrier relaxation dynamics in self-assembled semiconductor quantum dots. *Physical Review B 80*, 235319.
68. Schuh, K., F. Jahnke, and M. Lorke (2011). Rapid adiabatic passage in quantum dots: Influence of scattering and dephasing. *Applied Physics Letters 99* (1), 011105.
69. Seebeck, J., T. Nielsen, P. Gartner, and F. Jahnke (2006). Quantum kinetic theory of phonon-assisted carrier transitions in nitride-based quantum-dot systems. *European Physical Journal B 49*, 167.
70. Seebeck, J., M. Lorke, P. Gartner, and F. Jahnke (2009). Carrier-carrier and carrier-phonon scattering in the low-density and low-temperature regime for resonantly pumped semiconductor quantum dots. *Physica Status Solidi (C) 6*, 488.
71. Uskov, A. V., C. Meuer, H. Schmeckebier, and D. Bimberg (2011). Auger capture induced carrier heating in quantum dot lasers and amplifiers. *Applied Physics Express 4* (2), 022202.
72. Goldmann, E., M. Lorke, T. Frauenheim, and F. Jahnke (2014). Negative differential gain in quantum dot systems: Interplay of structural properties and many-body effects. *Applied Physics Letters 104* (24), 242108.
73. Lorke, M., W. Chow, and F. Jahnke (2007). Reduction of optical gain in semiconductor quantum dots. *Proceedings of SPIE 6468*, 646818.
74. Lorke, M., F. Jahnke, and W. Chow (2007). Excitation dependence of gain and carrier induced refractive index changes in quantum dots. *Applied Physical Letters 90*, 51112.
75. Lorke, M., J. Seebeck, T. Nielsen, P. Gartner, and F. Jahnke (2006). Excitation dependence of the homogeneous linewidths in quantum dots. *Physical Status Solidi (C) 3*, 2393–2396.
76. Dicke, R. H. (1954). Coherence in spontaneous radiation processes. *Physical Review 93*, 99–110.
77. Mandel, L. and E. Wolf (1995). *Optical Coherence and Quantum Optics*. New York, NY: Cambridge University Press.
78. Bohnet, J. G., Z. Chen, J. M. Weiner, D. Meiser, M. J. Holland, and J. K. Thompson (2012). A steady-state superradiant laser with less than one intracavity photon. *Nature 484* (7392), 78–81.
79. Timothy Noe Ii, G., J.-H. Kim, J. Lee, Y. Wang, A. K. Wójcik, S. A. McGill, D. H. Reitze, A. A. Belyanin, and J. Kono (2012). Giant superfluorescent bursts from a semiconductor magneto-plasma. *Nature Physics 8* (3), 219–224.
80. Leymann, H. A. M., A. Foerster, F. Jahnke, J. Wiersig, and C. Gies (2015). Sub—And superradiance in nanolasers. *Physical Review Applied 4* (4), 044018.
81. Auffèves, A., D. Gerace, S. Portolan, A. Drezet, and M. F. Santos (2011). Few emitters in a cavity: From cooperative emission to individualization. *New Journal of Physics 13* (9), 093020.
82. Jahnke, F., C. Gies, M. Aßmann, M. Bayer, H.A.M. Leymann, A. Foerster, J. Wiersig, C. Schneider, M. Kamp, and S. Höfling (2016). Giant photon bunching, superradiant pulse emission and excitation trapping in quantum-dot nanolasers. *Nature Communications 7*, 11540.
83. Bimberg, D., M. Grundmann, and N. N. Ledentsov (1998). *Quantum Dot Heterostructures*. Hoboken, NJ: John Wiley & Sons.

84. Bayer, M. and A. Forchel (2002). Temperature dependence of the exciton homogeneous linewidth in In0.60Ga0.40As/GaAs self-assembled quantum dots. *Physical Review B 65* (4), 041308.
85. Moody, G., R. Singh, H. Li, I. A. Akimov, M. Bayer, D. Reuter, A. D. Wieck, and S. T. Cundiff (2013). Fifth-order nonlinear optical response of excitonic states in an InAs quantum dot ensemble measured with two-dimensional spectroscopy. *Physical Review B 87* (4), 045313.
86. Scheibner, M., T. Schmidt, L. Worschech, A. Forchel, G. Bacher, T. Passow, and D. Hommel (2007). Superradiance of quantum dots. *Nature Physics 3* (2), 106–110.
87. Lichtmannecker, S., M. Florian, T. Reichert, M. Blauth, M. Bichler, et al. (2017). A few-emitter solid-state multi-exciton laser. *Scientific Reports 7* (7420). doi: 10.1038/s41598-017-07097-9
88. Ates, S., S. M. Ulrich, A. Ulhaq, S. Reitzenstein, A. Löffler, S. Höfling, A. Forchel, and P. Michler (2009). Non-resonant dot—cavity coupling and its potential for resonant single-quantum-dot spectroscopy. *Nature Photonics 3* (12), 724–728.
89. Florian, M., P. Gartner, C. Gies, and F. Jahnke (2013). Phonon-mediated off-resonant coupling effects in semiconductor quantum-dot lasers. *New Journal of Physics 15* (3), 035019.
90. Hohenester, U., A. Laucht, M. Kaniber, N. Hauke, A. Neumann, A. Mohtashami, M. Seliger, M. Bichler, and J. J. Finley (2009). Phonon-assisted transitions from quantum dot excitons to cavity photons. *Physical Review B 80* (20), 201311.
91. Majumdar, A., E. D. Kim, Y. Gong, M. Bajcsy, and J. Vučković (2011). Phonon mediated off-resonant quantum dot–cavity coupling under resonant excitation of the quantum dot. *Physical Review B 84* (8), 085309.
92. Press, D., S. Götzinger, S. Reitzenstein, C. Hofmann, A. Löffler, M. Kamp, A. Forchel, and Y. Yamamoto (2007). Photon antibunching from a single quantum-dot-microcavity system in the strong coupling regime. *Physical Review Letters 98* (11), 117402.
93. Tarel, G. and V. Savona (2009). Emission spectrum of a quantum dot embedded in a nanocavity. *Physica Status Solidi (C) 6* (4), 902–905.
94. Roy, C. and S. Hughes (2011). Influence of electron–acoustic-phonon scattering on intensity power broadening in a coherently driven quantum-dot–cavity system. *Physical Review X 1* (2), 021009.
95. Chauvin, N., C. Zinoni, M. Francardi, A. Gerardino, L. Balet, B. Alloing, L. H. Li, and A. Fiore (2009). Controlling the charge environment of single quantum dots in a photonic-crystal cavity. *Physical Review B 80* (24), 241306.
96. Florian, M., P. Gartner, A. Steinhoff, C. Gies, and F. Jahnke (2014). Coulomb-assisted cavity feeding in nonresonant optical emission from a quantum dot. *Physical Review B 89* (16), 161302(R).
97. Karrai, K., R. J. Warburton, C. Schulhauser, A. Högele, B. Urbaszek, E. J. McGhee, A. O. Govorov, J. M. Garcia, B. D. Gerardot, and P. M. Petroff (2004). Hybridization of electronic states in quantum dots through photon emission. *Nature 427* (6970), 135–138.
98. Settnes, M., P. Kaer, A. Moelbjerg, and J. Mork (2013). Auger processes mediating the nonresonant optical emission from a semiconductor quantum dot embedded inside an optical cavity. *Physical Review Letters 111* (6), 067403.
99. Winger, M., T. Volz, G. Tarel, S. Portolan, A. Badolato, K. J. Hennessy, E. L. Hu, A. Beveratos, J. Finley, V. Savona, and A. Imamoğlu (2009). Explanation of photon correlations in the far-off-resonance optical emission from a quantum-dot–cavity system. *Physical Review Letters 103* (20), 207403.
100. Bockelmann, U. and T. Egeler (1992). Electron relaxation in quantum dots by means of auger processes. *Physical Review B 46*, 15574.
101. Efros, A. L., V. A. Kharchenko, and M. Rosen (1995). Breaking of the phonon bottleneck in nanometer quantum dots role of auger-like processes. *Solid State Communications 93*, 281.
102. Uskov, A. V., F. Adler, H. Schweizer, and M. H. Pilkuhn (1997). Auger carrier relaxation in self-assembled quantum dots by collisions with two-dimensional carriers. *Journal of Applied Physics 81*, 7895.

103. Dekel, E., D. Gershoni, E. Ehrenfreund, D. Spektor, J. M. Garcia, and P. M. Petroff (1998). Multiexciton spectroscopy of a single self-assembled quantum dot. *Physical Review Letters 80* (22), 4991–4994.
104. Laucht, A., M. Kaniber, A. Mohtashami, N. Hauke, M. Bichler, and J. J. Finley (2010). Temporal monitoring of nonresonant feeding of semiconductor nanocavity modes by quantum dot multiexciton transitions. *Physical Review B 81*, 241302.
105. del Valle, E., F. P. Laussy, and C. Tejedor (2009). Luminescence spectra of quantum dots in microcavities. II. Fermions. *Physical Review B 79*, 235326.
106. Nomura, M., N. Kumagai, S. Iwamoto, Y. Ota, and Y. Arakawa (2010). Laser oscillation in a strongly coupled single-quantum-dot–nanocavity system. *Nature Physics 6*, 279–283.
107. Valle, E. d. and F. P. Laussy (2011). Effective cavity pumping from weakly coupled quantum dots. *Superlattices and Microstructures 49* (3), 241–245.
108. Yao, P., P. K. Pathak, E. Illes, S. Hughes, S. Münch, S. Reitzenstein, P. Franeck, A. Löffer, T. Heindel, S. Höfling, L. Worschech, and A. Forchel (2010). Nonlinear photoluminescence spectra from a quantum-dot cavity system: Interplay of pump-induced stimulated emission and anharmonic cavity QED. *Physical Review B 81* (3), 033309.
109. Lichtmannecker, S., M. Florian, T. Reichert, M. Blauth, M. Bichler, F. Jahnke, J. J. Finley, C. Gies, and M. Kaniber (2016). A few-emitter solid-state multi-exciton laser. *ArXiv:1602.03998 [condmat]*.
110. Strauf, S., K. Hennessy, M. T. Rakher, Y.-S. Choi, A. Badolato, L. C. Andreani, E. L. Hu, P. M. Petroff, and D. Bouwmeester (2006). Self-tuned quantum dot gain in photonic crystal lasers. *Physical Review Letters 96*, 127404.
111. Mlynek, J. A., A. A. Abdumalikov, C. Eichler, and A. Wallraff (2014). Observation of Dicke superradiance for two artificial atoms in a cavity with high decay rate. *Nature Communications 5*, 5186.
112. Müller, M., S. Bounouar, K. D. Jöns, M. Glässl, and P. Michler (2014). On-demand generation of indistinguishable polarization-entangled photon pairs. *Nature Photonics 8* (3), 224–228.
113. Muñoz, C. S., E. del Valle, A. G. Tudela, K. Müller, S. Lichtmannecker, M. Kaniber, C. Tejedor, J. J. Finley, and F. P. Laussy (2014). Emitters of n-photon bundles. *Nature Photonics 8* (7), 550–555.
114. Gies, C. et al., Phys. Rev. A **96**, 023806 (2017).

49

Nonlinear Dynamics in Quantum Photonic Structures

49.1 Introduction 662

49.2 Maxwell-Pseudospin Method 667
Pseudospin Equation • Dissipative Dynamics and Master Pseudospin Equation • Vector Maxwell Equations and Coupling to Pseudospin Equations • Coherent Maxwell-Pseudospin Equations in 2D • Numerical Methodology

49.3 Simulation of Nonlinear Dynamics in Optical Waveguides and Semiconductor Microcavities 674
One-Dimensional Model of a Two-Level System • Passive Cavity Properties • Active Nonlinear Optical Waveguide/Cavity Properties • Resonant Nonlinearities and Gain Saturation • Coherent Propagation Effects, Self-Localization and Pattern Formation in 2D Planar Nonlinear Optical Waveguides and Semiconductor Microcavities

49.4 Optical Orientation in Quantum Wells and QDs 691
Circularly Polarized Optical-Pulse Interactions with Two-Level Quantum Systems: Selective Spin Initialization and Readout • Circularly Polarized Optical Pulse Interactions with Degenerate Four-Level Systems

49.5 Modeling Coherent Spin Generation, Manipulation, and Readout in Charged QDs 699
Single-Charged QD Embedded in a Waveguide • Coherent Spin Manipulation through Hot-Charged Exciton States in QDs

49.6 Quantum Stochastic Formalism for Modeling Cavity-Emitter Systems 707
Langevin Formalism for Modeling Quantum-Optical Effects • Spontaneous Emission of a QD Ensemble in a Semiconductor Microcavity and Onset of Lasing

49.7 Modeling Optical Rotation in Chiral Carbon Nanotubes with Master Maxwell-Pseudospin Equations 712
Natural Optical Activity • Faraday Rotation and Magneto-Chiral Effects

49.8 Conclusion 724

Gabriela Slavcheva
and
Mirella Koleva

49.1 Introduction

Quantum optics unites physical optics and the quantum field theory of light. The foundations of the latter were laid by Planck, who succeeded to explain theoretically the black-body radiation spectrum by postulating that the energy of a harmonic oscillator is quantized. In 1927, Dirac developed this theory by solving the long-standing problem of the wave-particle duality and unifying the wave and particle aspects of light by quantization of the electromagnetic radiation. Thus, according to Dirac's single-particle interpretation, the light emitted from a spontaneously emitting source with low photon density is described as single-photon emission events.

A fundamental consequence of light-field quantization is the appearance of vacuum (zero-point energy) fluctuations, or quantum noise, which has no classical analogue and stems from the Heisenberg uncertainty relations. However, some of the greatest successes of quantum theory—such as Planck's black-body spectrum and Einstein's theory of the photoelectric effect—do not actually require a full quantum-mechanical treatment and can instead be explained using semiclassical theory, whereby matter is quantized but light remains a classical wave. Only in the 1960s, motivated by the development of lasers, did researchers start to examine effects which truly required a quantum theory of light. For instance, light quantization is required to explain spontaneous emission, irreversible decay, and the origin of the laser spectral linewidth. The demand to develop further the quantum theory of light in order to explain these effects was met by the pioneering work of Glauber [1,2].

Vacuum quantum fluctuations are necessary to account for phenomena, such as the Lamb shift: a radiative correction to the energy spectrum of hydrogen, as well as for the attractive force between two uncharged conductive plates in a vacuum placed a few nanometres apart, known as the Casimir effect. Some examples of effects which the semiclassical approach fails to explain are quantum beat phenomena in a three-level atom, two-photon interferometry and production of "entangled" states of light, resonance fluorescence (quantum jumps) in the emission of atoms, and the so-called "nonclassical" states of light.

The classification of light states as classical and nonclassical stems from the statistical properties of radiation and, more specifically, from the second-order coherent properties of the light emitter. A crucial step in elucidating these properties was the correlation experiment of Hanbury Brown and Twiss [3], which was designed to measure the coincidence of individual photons at two detectors placed equidistantly from a half-silvered mirror splitting a beam of oncoming low-intensity light. Using the second-order correlation function, $g^{(2)}(\tau)$, quantifying the synchronicity with which photons arrive at the detectors, it was then possible to discern the manner in which different types of light sources emitted light. This method revealed that classical light sources—such as a thermal black-body radiation source or a lamp—exhibited Bose–Einstein statistics manifested by photon "bunching", i.e., an increased likelihood of the photons to arrive at the detectors at the same time, as though they had "bunched together" while traveling from the source to the detectors. On the other hand, a stable coherent light source (e.g., a laser) exhibits Poissonian statistics, which is manifested by a constant unity second-order correlation function, $g^{(2)}(\tau)$, meaning that photons are not intercorrelated and are equally likely to impinge on the detectors with a delay of any length with respect to each other. However, most interestingly of all, nonclassical states of light are described by sub-Poissonian statistics (a diminished likelihood of coincident arrival) called "antibunching": the photons are "diluted" as they travel and eventually emitted one by one. Antibunching is characteristic of resonant fluorescence from single atoms, molecules, or semicondictor nanocrystals and quantum dots (QDs), and is a signature of a single-photon emitter. For perfectly antibunched light, the second-order correlation function vanishes for coincident detections. This is irreconcilable with the classical wave optics and is considered a hallmark of pure quantum behavior.

Photon-number (Fock) states are states of light containing a specific number of photons. The single-photon state is a special case of a Fock state and is a squeezed-light state [4], a quantum state of light resulting from the Heisenberg relations with, e.g., reduced amplitude fluctuations at the expense of enhanced fluctuations of the phase. An ideal amplitude-squeezed source delivers a stream of photons at regular time intervals, which is the precise realization of a single-photon source [5].

Another quantum-mechanical property is entanglement. Quantum systems consisting of interacting subsystems become highly correlated and their individual constituents become "entangled". The global quantum state of an entangled system—for e.g., one consisting of two particles (e.g., a photon pair)—remains correlated even when the two particles are separated by large distances from each other and unable to interact or communicate in any known way. The most famous example of an entangled state is the Einstein–Podolsky–Rosen–Bohm [6] state, whereby two spin 1/2 particles reside in a state with zero total spin angular momentum: $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle)$. Notably, this state cannot be represented as a simple product of the pure states of the individual particles.

Many of these quantum effects have been observed only in carefully controlled atomic systems; however, with the advent of quantum photonics (the applied device-oriented field of quantum optics), recent advances in semiconductor technology have enabled the fabrication of solid-state devices that emit single photons, opening new avenues for nonclassical light state generation. For instance, the use of radiative cascades from QD excited states is a principal method for generation of quantum entanglement for applications of QDs as polarization-entangled sources. The exchange interaction and the fine structure splitting of the energy levels involved in radiative cascades from QD excited states are of fundamental importance, since they determine the polarization and entanglement of the emitted photons.

Quantum photonics has attracted great interest in recent years. This is due to its potential to revolutionize science and day-to-day life alike through enabling the implementation of faster algorithms, secure transmission of information, vast increase in data storage, and execution of more accurate measurements. One of the major challenges of modeling quantum photonic devices is working out a way of efficient simulation and control of realistic, "open" quantum systems (i.e., allowing for energy to flow in and out of the system) and devices for applications in quantum technologies. For optical quantum information processing to advance beyond demonstration experiments, the development of on-chip capabilities will be essential. Solid-state qubit architectures have emerged as a most promising physical implementation of quantum photonic circuits, taking advantage of the state-of-the-art semiconductor technologies. Reliable, integrated sources of nonclassical light, as well as realization of high-fidelity on-chip readout of solid-state qubits, are a critical requirement for next-generation quantum technologies [7].

Recently, significant progress has been made toward the achievement of high-efficiency solid-state single-photon sources [8–12]. Among these sources, QDs are particularly promising since they can be embedded in devices enabling electrical injection while maintaining efficient light emission. Furthermore, the dot doping—or the number and type of the charge carriers within the dot—can be controlled by applying electrical bias in a Schottky-type structure. A quantum bit (qubit) is a linear superposition of basis states $|0\rangle$ (ground level) and $|1\rangle$ (excited level) in a two-level system, given by: $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with α and β being complex numbers such that $\alpha^2 + \beta^2 = 1$. In a classical computer, the information is encoded in digital bit states 0 and 1. By contrast, in a quantum computer, the state of each bit is permitted to be any quantum-mechanical state of a qubit (manipulated using qubit rotations, also known as gate operations), leading to a massive parallelism and an exponential speed-up of the quantum computations.

The $|0\rangle$ and $|1\rangle$ states of a qubit could be encoded in the spin-up ($|\uparrow\rangle$) and spin-down ($|\downarrow\rangle$) states of a charge carrier. The spin-up and spin-down states of a single electron or hole confined in a charged semiconductor QD are a particularly attractive qubit candidate [13]. This arises from several advantages intrinsic to this quantum system. The first advantage is that the atomic-like electronic structure suppresses coupling of the spin to the solid-state quantum environment and thus impedes the spin decoherence processes. This leads to a very long spin coherence time (denoted T_2)—that is, the time within which the spin loses its phase information—and thus a long-lived quantum state, which is required for performing a large number of spin manipulations. Another important advantage is optical convertibility—i.e., the ability to switch through transfer of angular momentum between a "flying" qubit encoded in a photon and the stationary matter qubit of the spin confined in a QD. This enables the transport of qubits over large distances and facilitates the realization of scalable solid-state qubit architectures for optical quantum computing. Furthermore, this promising platform allows integrating the operations of optical initialization, manipulation,

and readout of a spin qubit on a chip. By employing a combination of optical orientation [14] and quantum coherent control [15] techniques, using ultrashort polarized light pulses one can address individual carrier spins in semiconductor QDs, initialize and manipulate them coherently through optically excited states (charged excitons), and detect the resulting spin qubit [16]. For quantum technologies applications, control of both QD position and emission energy is crucial. For instance, deterministic site control of QD location is indispensable for multiqubit operations and scalability. Therefore, new simulation tools are needed to determine the allowed tolerances in the dot positions and emission energies to ensure observation of the desired quantum-optical effects.

Quantum coherent control represents a universal approach for predictable manipulation of the properties of quantum systems, such as atoms and molecules. This technique has most recently been applied to QD solid-state systems. Coherent control in semiconductor nanostructures allows for coherent manipulation of the carrier wavefunctions on a time scale shorter than typical dephasing times. This is a prerequisite for successful implementation of ultrafast optical switching and quantum information processing. Quantum coherent control requires use of ultrashort pulses considerably shorter than the characteristic relaxation times in matter ($\tau \ll T_1, T_2$). This is equivalent to the photon-dipole coupling rate exceeding all dissipative rates in the system, a general condition for achieving the strong-coupling regime. The ultrashort pulses are usually characterized by high field amplitudes and, consequently, lead to nonlinear optical effects, such as coherent pulse propagation and self-induced transparency (SIT) [17]. The phenomenon of SIT is observable above a critical power threshold for a given pulse width: a high-intensity, ultrashort pulse propagating through a medium composed of an ensemble of resonant quantum two-level absorbers, whose relaxation times greatly exceed the pulse duration, travels at a reduced speed and unchanged shape with anomalously low energy loss. The absorbers are driven into the excited state by absorbing ultrashort pulse energy; by reradiating this energy into the pulse, they return to the ground state. Thus, the optical energy is carried through the medium not by the electromagnetic field, but by a coupled light-matter polariton wave. The polariton is a mixed light-matter quasiparticle resulting from the strong coupling of the optical wave to the medium's polarization. As a result, the pulse travels as a solitary wave, known as a SIT-soliton. This soliton is localized both in space and time, in contrast with the well-known nonlinear optics solitons which result from the interplay between nonlinearity and the medium's dispersion and/or diffraction. The condition of SIT is predicted by the remarkable pulse-area theorem (PAT) [18] which establishes a general criterion for stable, ultrashort pulse propagation in attenuating or amplifying media based on an integral quantity—namely, the pulse area. This phenomenon can be preserved to a great extent in solid-state systems—e.g., semiconductors—and has been experimentally demonstrated by picosecond pulse propagation in QD waveguides [19].

When the laser pulse's temporal width becomes comparable with the optical period (i.e., when the pulse envelope contains only a few optical cycles of the carrier wave), a transition to a qualitatively new "extreme nonlinear" regime of strong light-matter interactions is induced in which the electric field itself, rather than intensity envelope, drives the interaction. The majority of the current analytical approaches are based on the standard, slowly varying amplitude approximation (SVEA) and rotating-wave approximation (RWA). However, a number of works have demonstrated the limitations of these approximations, and new phenomena have been predicted on the basis of a non-perturbative approach (see, e.g., [20], and references therein). This requires development of new, non-perturbative theoretical and computational methods beyond the above approximations as shown in Section 49.2, which could account properly for the ultrafast carrier-wave dynamics.

The earlier high-intensity (i.e., high photon density) limit allows to achieve the strong-coupling regime in free space. However, in the opposite, low-intensity (i.e., few-photon) limit, the light-matter coupling of a single photon to a single atom in free space is weak, and despite attempts by several groups [21–23] to improve it, the experimental realization of high-fidelity excited-state preparation remains challenging. In order to improve the performance of single-photon sources, it is necessary to incorporate the emitters in cavities. The cavity helps to select a single mode out of the infinite number of radiation field modes coupled

to the emitter. As the emission is preferentially directed into a single mode, owing to the Purcell effect the spontaneous emission lifetime is reduced for emitters located at the antinode of the electric-field intensity in cavities with a small mode volume and high quality factor. Faster emission is desired because it makes the single-photon source more deterministic and reliable.

If the exciton-light coupling in a semiconductor microcavity is weak compared to all dissipative rates in the system, irreversible decay occurs—that is, energy cannot be alternately passed from light to matter. Depending on whether the quantum wells inside the semiconductor microcavity are situated near the node or antinode of the excitatory electric field, either an inhibited or enhanced spontaneous emission takes place. The cavity then operates in the weak coupling regime, whereby the dynamics of the light-matter interaction is dominated by incoherent processes and damping. Conversely, in the strong coupling limit, no irreversible decay takes place. Instead, a sequence of coherent reabsorption and reemission processes occurs and the energy oscillates between the exciton and photon modes, a phenomenon known as optical Rabi oscillations. The spectral signature of strong coupling is an anticrossing between the dot line (i.e., the spectral linewidth of the QD) and the cavity mode. The eigenmodes of the system are mixed entangled light-matter states, giving rise to quasiparticles known as microcavity exciton–polaritons which exhibit bosonic properties. Considerable scientific effort worldwide has been invested in exploration and control of a strongly coupled single QD in a cavity. The latter is considered a solid-state analogue of the atom-cavity system for which many quantum-optical effects have been demonstrated. Some examples include cavity field quantization [24] and photon blockade [25]. Embedding QDs in confined geometries—such as planar [26], micropillar [27], microdisk [28], microsphere [29], nanowire [30], or photonic crystal nanocavities [31]—provides a means for controlling the light-matter interactions. However, unlike isolated atoms, QDs are embedded in a solid-state environment, so confined charge carriers may experience additional interactions with phonons and other neighboring states and charges. These interactions may result in a departure from the behavior of an idealized two-level quantum system, but can also be exploited as additional control measures.

Cavity quantum electrodynamics (CQED), traditionally applied to atom-cavity systems, has gained new momentum after recent reports of quantum-optical experiments in solid-state systems demonstrating nonclassical light behavior, such as nonclassical photon statistics, nonlocality, and quantum entanglement [32,33]. These cutting-edge studies put to the test the very foundations of quantum mechanics and open up new avenues for on-demand generation and exploitation of various nonclassical states of light. So far, close to perfect single-photon interaction with a QD has been achieved in a cavity under the strong coupling regime [34,35].

A major goal of the simulation of quantum-photonic devices is to find an optimum design for combining photonic and solid-state devices on a chip. Integrated optical technologies for quantum computing based on linear optics [36] have been under active development in the recent decade. However, nonlinear optical interactions "on a chip" offer new, unexplored functionalities, and may play a key role in future technologies. In this respect, a major theoretical and computational challenge is the development of new, tractable models of the nonlinear quantum dynamics in realistic devices with complex geometries in multiple dimensions. This chapter reviews the evolution of the original methodology into our fully non-perturbative, mesoscopic, Maxwell's curl-pseudospin equation-based model of the optically induced spatiotemporal dynamics in quantum-photonic devices. This chapter contains further details about its numerical implementation and some concrete applications.

As we shall show in Section 49.2, our theoretical and computational approach differs substantially from the methods described in Chapter 46 of this book. In that chapter, a number of different approaches for calculation of the spontaneous emission have been employed. In the weak-coupling regime, a classical dipole model is developed whereby the dyadic Green's function is calculated and the spontaneous emission rate is found from Fermi's Golden Rule. Analytical expressions for Green's function are obtainable only for a limited number of device geometries, and more generally one has to resort to numerical methods (e.g., the finite-element method) to compute the electric field. This entails expanding the fields in

the eigenmodes in different layers of the specific geometry under consideration. Another approach for modeling cavity-emitter systems is CQED, which is based on the Jaynes–Cummings Hamiltonian, describing the interaction of a two-level system strongly coupled to a single mode of an electromagnetic field. Two approximations are made: (1) for a very low Q cavity, the optical field operators can be replaced with material excitation operators; the dissipation in the system (spontaneous emission rate and dephasing) is taken into account through Lindblad operators. Using the quantum regression theorem that is valid for Markovian processes, an analytical expression for the correlation function is obtained; (2) for a coupled cavity-emitter system, a more complicated master equation is written with an additional dissipation term describing the cavity photon loss. However, for losses weaker than the cavity-emitter coupling, the cavity can be adiabatically eliminated and this results in a master equation describing the emitter only but with modified spontaneous emission rates. By contrast, we use the full master equation for the system density matrix which is valid for all regimes: weak, intermediate, and strong coupling. In addition, in our case the longitudinal and transverse dissipation are taken into account not through Lindblad operators, but by considering the dynamics of population transfer and decoherence between the discrete levels. Another difference between our model presented in Section 49.2 and the model of Gregersen et al. is that in our case, the cavity-emitter (Rabi) coupling is not a model parameter, but is calculated from the initial pulse amplitude, duration, and dipole moment of the specific optical transition, as in the actual experiment. Furthermore, the cavity loss is not a phenomenological parameter either, as it is calculated numerically through the perfectly transparent boundary conditions of the specific cavity geometry. The master equation is solved self-consistently with the vector Maxwell equations for the pulse propagation.

Our approach, predominantly using quantum-optical description based on the Liouville equation of motion for the density matrix, is quite distinctive from the many-body theoretical methods applied in Chapter 48. For instance, the developed by the aforementioned authors quantum-optical semiconductor laser theory beyond the semiclassical approach is based on a correlation-expansion and truncation technique for the quantized light field interacting with carriers in the single-particle picture (electrons and holes) occupying the discrete QD levels. Dynamical equations for the carrier populations in the conduction and valence states of the QD, for the expectation value of the photon number operator of the laser mode and the photon-assisted polarization are derived using Heisenberg operator equations of motion for the creation and annihilation operators. These coupled equations contain carrier–photon and photon–photon correlations, providing access to the second-order correlation function, $g^{(2)}$. The interaction with the environment is accounted through quantum kinetic equations, beyond the Boltzmann equation approach, for the carrier dynamics due to carrier–phonon and carrier-Coulomb interactions. Many-body theory is applied to calculate the corresponding matrix elements. In the former case, effects such as polaron scattering and non-Markovian dynamics are shown for QD systems. This is a very thorough many-body quantized electromagnetic field approach; however, it does not take into account the specific macroscopic boundary conditions of the particular device geometry considered, as the light-matter interaction is described by a single laser mode coupled to the quantum system. An extension for multiple emitters coupled to a single common light mode is proposed, taking advantage of the many-particle picture description of the QD states, such as excitons, trions, multiexcitons, etc., to describe interemitter correlations and radiative coupling. This is somewhat similar to our approach in that it considers many-particle states (e.g., excitons); however, our approach is extended in that it also considers multiple cavity electromagnetic field modes. The cluster expansion approach in Chapter 48 by Gies et al. is in essence a generation of equations of motion for the photon correlations up to a given order of truncation of the expansion and thus providing access to $g^{(2)}$. The dissipation phenomena within this approach—such as the carrier–phonon and carrier–carrier Coulomb scattering—are calculated from a microscopic many-body model (compared to the Lindblad terms and our approach, where they are taken as phenomenological parameters), which is an advantage, as the theory is self-consistent.

In Section 49.2, we will give an overview of the strategy we have used to construct a model for describing coherent light-matter interactions in semiconductor nanostructures.

49.2 Maxwell-Pseudospin Method

49.2.1 Pseudospin Equation

In what follows, we will consider either resonant or near-resonant laser optical field interactions with atom-like discrete-level quantum systems, such as QDs. The optical field we consider is monochromatic and (near-)resonant with a given transition between a pair of atomic levels, thus effectively mimicking a two-level atom. As the two-level atom is conceptually equivalent to a half-spin particle in a magnetic field, the dynamics of the laser-atom interaction is governed by the same Bloch spin vector equation of motion [37] developed for describing nuclear magnetic resonance and the precession of a classical gyromagnet (or its quantum analogue, the spin vector, **S**) in a constant magnetic field, **B**:

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = f\mathbf{B} \times \mathbf{S}. \tag{49.1}$$

Here f guarantees the constant length of the spin vector, thus confining the spin dynamics to the surface of a Bloch sphere. (Note, however, that this equation is valid only for systems with equally spaced energy levels). The formal analogy between a spin and a nonspin (pseudospin) system was pointed out for the first time by Feynman and coworkers [38], showing that when coherent processes are involved in a two-level system, it is sufficient to consider a real three-vector, rather than the complex probability amplitudes in the Schrödinger equation or the complex density-matrix elements [37]. This real-vector representation provides an elegant and intuitive geometrical framework for comprehending the system dynamics in terms of rotations of a real state vector in the Hilbert space. Furthermore, in contrast to the wavefunction formalism, it describes both pure and mixed quantum states. Attempts to extend the formalism analogously for (N > 2)-level quantum systems fail to preserve the simple vector form of the basic equation of motion for the real state pseudospin (coherence, or Bloch) vector. The only known exception to this is the special case of excitation of a spin-J system in constant magnetic field (thus yielding a $(N = 2J + 1)$-level system), due to the equal spacing of the energy levels in this system [39]. Preserving this simple form of the vector equation has a number of advantages: (1) it accounts for the intrinsic symmetry of the underlying Hilbert space of the system and therefore is an exact description, independent of the strength, number, or time dependence of the external forces acting on the system; (2) similar to the two-level system, the dynamical evolution can be illustrated as a rotation in the real physical space of a real coherence vector.

The case of a three-level system is of particular interest, since it provides a useful framework for studying such phenomena as two-photon coherence, resonance Raman scattering, three-level echoes, three-level superradiance, coherent multistep photoionization/photodissociation, coherent population trapping, and electromagnetically induced transparency (EIT). A solution to this problem has been given by Elgin [40] by invoking the invariance of the state vector under rotations of the SU(3) transformation group. The dynamical evolution of a three-level system can be expressed in terms of an eight-dimensional real coherence vector, taking advantage of the group-theory and Gell-Mann's SU(3) generators, which had been developed for the quark triplet in the domain of high-energy physics [41].

A general solution for an N-level quantum system with arbitrary level spacing has been provided by Hioe and Eberly [42]. The starting point is the dynamical evolution of a quantum system described by the Liouville equation for the density operator in the Schrödinger picture

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = \left[\hat{H}, \hat{\rho}\right], \tag{49.2}$$

where $\hat{H}$ is the total system Hamiltonian:

$$\hat{H} = \hat{H}_0 + \hat{H}_{\mathrm{int}}(t), \tag{49.3}$$

with $\hat{H}_0$ being the unperturbed Hamiltonian of an N-level system: a diagonal matrix with eigenenergies of each level, $\hbar\omega_k$ $(k = 1, ..., N)$, along the main diagonal, and $\hat{H}_{int}$ is a time-dependent perturbation (not necessarily small).

Hioe and Eberly showed that the presence of unitary group generators in the time evolution of an N-level quantum system permits to describe the evolution in terms of the rotations of a real coherence vector. Expanding the system Hamiltonian and the density matrix operators in terms of λ-generators of the SU(N) Lie algebra, $\hat{\lambda}_k$, they derived the pseudospin equation of motion in the Heisenberg picture, depicting the time evolution of the real pseudospin vector, **S**, as a generalized rotation in $N^2 - 1$ Hilbert space:

$$\dot{S}_i = f_{ijk}\gamma_j S_k, \;\; i, j, k = 1, 2, ..., N^2 - 1, \tag{49.4}$$

where the dot stands for the time derivative, γ_j is the torque vector, and $f_{ijk} = \frac{1}{4}i\left[Tr\left(\hat{\lambda}_i\hat{\lambda}_k\hat{\lambda}_j\right) - Tr\left(\hat{\lambda}_i\hat{\lambda}_j\hat{\lambda}_k\right)\right]$, $i, j, k = 1, 2, ..., N^2 - 1$, is the fully antisymmetric tensor of the structure constants of the SU(N) group, whose asymmetry guarantees the constant length of the pseudospin vector, $|\mathbf{S}|$, in the (N^2-1)-dimensional Hilbert space. The coherence vector is expressed as the trace of the product of the density and the SU(N) group generators, whereas the torque vector is expressed as the trace of the product of the system Hamiltonian operators and the SU(N) group generators

$$S_j(t) = Tr\left(\hat{\rho}(t).\hat{\lambda}_j\right); \;\; \gamma_j(t) = \frac{1}{\hbar}Tr\left(\hat{H}(t).\hat{\lambda}_j\right). \tag{49.5}$$

The SU(N) group generators obey the group property

$$\left[\hat{\lambda}_j, \hat{\lambda}_k\right] = 2if_{jkl}\hat{\lambda}_l \tag{49.6}$$

and orthogonality relations, $Tr\left(\hat{\lambda}_j.\hat{\lambda}_k\right) = \delta_{jk}$, however their choice is not unique.

For the simplest case of a two-level system, $N = 2$, the λ-generators are simply the Pauli matrices $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$. This is obtained from Equation 49.4 when f_{ijk} is replaced by the Levi-Civita symbol, ε_{ijk}. For a dipole-coupling interaction $\hat{H}_{\text{int}} = e\mathbf{E}.\hat{\boldsymbol{r}}$, where $\mathbf{r}$ is the local displacement operator and e is the electron charge, and a linearly polarized dipole optical transition ($\mathbf{E} \,||\, \mathbf{r}$ with an optical selection rule for the z-axis projection of the total angular momentum, $\Delta J_z = 0$) the optical Bloch equations are easily recovered:

$$\dot{\mathbf{S}} = \boldsymbol{\gamma} \times \mathbf{S} \quad \begin{aligned} \dot{S}_1 &= \omega_0 S_2 \\ \dot{S}_2 &= -\omega_0 S_1 + 2\Omega_R S_3 \\ \dot{S}_3 &= -2\Omega_R S_2, \end{aligned} \tag{49.7}$$

where S_i $(i = 1, 2, 3)$ is the real pseudospin three-vector, and the torque vector is given by $\boldsymbol{\gamma} = \left(2\Omega_{\text{R}}, 0, \omega_0\right)$ from Equation 49.5. Here we have defined the Rabi frequency, $\Omega_{\text{R}} = \frac{\wp}{\hbar}E$, with $\wp = \langle i|e\hat{r}|j\rangle$, transition dipole matrix element of $i \to j$, and ω_0 is the atomic transition resonant frequency. In case of near-resonant excitation, ω_0 should be replaced by $\Delta\omega = \omega_0 - \omega$, where ω is the excitation frequency [43].

In the general case of an electric field vector noncollinear with the local displacement (dipole moment) vector, the optical transition is excited by circularly or elliptically polarized light obeying the selection rule $\Delta J_z = \pm 1$, where the sign "+/–" corresponds to right/left circularly polarized light. The dipole matrix elements are complex vectors yielding complex Rabi frequencies along the x- and y-direction in the plane determined by **E** and **r** (or equivalently, along the dipole moment vector): $\Omega_x = \frac{\wp}{\hbar}E_x$; $\Omega_y = \frac{\wp}{\hbar}E_y$. The

torque vector is given by $\boldsymbol{\gamma} = \left(\Omega_x, \Omega_y, \omega_0\right)$ and the resulting pseudospin equations read (see also [37])

$$\dot{\mathbf{S}} = \boldsymbol{\gamma} \times \mathbf{S} \quad \begin{aligned} \dot{S}_1 &= -\omega_0 S_2 - \Omega_y S_3 \\ \dot{S}_2 &= \omega_0 S_1 + \Omega_x S_3 \\ \dot{S}_3 &= \Omega_y S_1 - \Omega_x S_2 \,. \end{aligned} \tag{49.8}$$

The relationship between the real-state coherence vector components and the density-matrix components for the general N-level case is given by Equation 49.5, and for a two-level system reads

$$\begin{aligned} S_1 &= \hat{\rho}_{12} + \hat{\rho}_{21} = 2\mathrm{Re}\left(\hat{\rho}_{12}\right) \\ S_2 &= i\left(\hat{\rho}_{12} - \hat{\rho}_{21}\right) = -2\mathrm{Im}\left(\hat{\rho}_{12}\right) = 2\mathrm{Im}\left(\hat{\rho}_{21}\right) \\ S_3 &= \hat{\rho}_{22} - \hat{\rho}_{21} \,. \end{aligned} \tag{49.9}$$

The physical meaning of the first two components of the pseudospin vector are dispersive (in-phase) and absorptive (in-quadrature) parts of the dipole polarization, and the third component is the population difference between the two levels. Since the level occupation probability is conserved, i.e., $\rho_{11} + \rho_{22} = 1$, it can be easily shown that the length of the coherence vector is unity: $|S|^2 = S_1^2 + S_2^2 + S_3^2 = 1$, and the evolution of the pseudospin is "inscribed" by the pseudospin vector on the surface of the Bloch sphere.

Up until now, we have only considered coherent processes; however, in an open quantum system there is always energy dissipation. For a two-level system, it is easy to define phenomenological longitudinal relaxation time (also known as population relaxation time), T_1, and transverse relaxation time (also known as polarization-decay or dephasing time), T_2, leading to the well-known optical Bloch equations for a two-level system with damping for e.g., linearly polarized transitions:

$$\begin{aligned} \frac{\partial S_1}{\partial t} &= \omega_o S_2 - \frac{1}{T_2} S_1 \\ \frac{\partial S_2}{\partial t} &= -\omega_o S_1 + 2\Omega_x S_3 - \frac{1}{T_2} S_2 \\ \frac{\partial S_3}{\partial t} &= -2\Omega_x S_2 - \frac{1}{T_1}\left(S_3 - S_{30}\right), \end{aligned} \tag{49.10}$$

where $S_{30} = \pm 1$ is the initial population profile; "–" and "+" correspond, respectively, to all of the population in the ground and the excited state. Upon optical excitation, the former describes absorption, while the latter describes optical gain. However, it is not so easy to generalize the dissipation dynamics for a system with an arbitrary number of levels (N-level system), because each pair of levels within this system is characterized by their individual population transfer rates and dephasing rates. We show how to do this generalization next. The time evolution of the two-level quantum system is described by the generalized rotation of the real three-vector, **S**, in the three-dimensional Hilbert space; at all times the tip of **S** is just touching the surface of the Bloch sphere, shown in Figure 49.1.

49.2.2 Dissipative Dynamics and Master Pseudospin Equation

To include energy relaxation and decoherence processes in the system dynamics, we adopt an unconventional approach. We introduce longitudinal relaxation in the Liouville equation for the density matrix by considering first the diagonal population components of the Liouville equation, ρ_{ii}, and taking into account population transfer between all allowed dipole optical transitions in the system with corresponding rates.

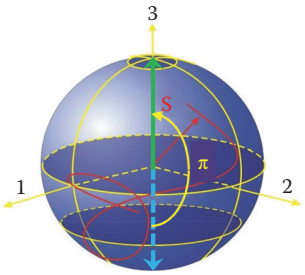


FIGURE 49.1 Geometrical representation of the time evolution of a two-level quantum system by generalized rotation of the real state pseudospin vector, **S**, in the 3D Hilbert space, drawing a trajectory on the $|\mathbf{S}| = const.$ sphere surface. A one-qubit rotation occurs when the vector is rotated through π from state 0 (spin down, dashed arrow) to state 1 (spin up, solid arrow along $\hat{3}$-axis), or vice versa.

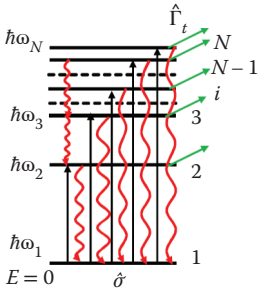


FIGURE 49.2 Schematic of an N-level quantum system with energy zero at the ground state. [Vertical solid arrows, coherent optical excitation; wavy lines, population relaxation $\hat{\sigma}$ (spontaneous emission); tilted arrows, dephasing $\hat{\Gamma}_t$ (decoherence).]

For each Liouville equation we can construct a longitudinal damping matrix of rank N: $\hat{\Gamma}_i$, $i = 1, ..., N$, whose elements are the population transfer rates between all pairs of levels. This allows us to introduce a diagonal matrix, $\hat{\sigma} = diag(Tr(\hat{\Gamma}_i.\hat{\rho}))$, which accounts for the longitudinal population relaxation. Similarly, we introduce an off-diagonal matrix, $\hat{\Gamma}_t$, of rank N with zero main diagonal components to account for the dephasing rates of all levels (Figure 49.2). Adding the relaxation and dephasing terms in the Liouville equation thus leads to the following master equation:

$$\frac{\partial\hat{\rho}}{\partial t} = \frac{1}{\hbar}\left[\hat{\rho}, \hat{H}\right] + \hat{\sigma} - \hat{\Gamma}_t.\hat{\rho}. \tag{49.11}$$

Using Equation 49.5, we can now rewrite the density-matrix equations in terms of the pseudospin vector, **S** [44]:

$$\frac{\partial S_j}{\partial t} = \begin{cases} f_{jkl}\Gamma_k S_l + \frac{1}{2}Tr\left(\hat{\sigma}.\hat{\lambda}_j\right) - \frac{1}{T_j}\left(S_j - S_{je}\right), & j = 1, 2, ...N\,(\mathrm{N}-1) \\ f_{jkl}\Gamma_k S_l + \frac{1}{2}Tr\left(\hat{\sigma}.\hat{\lambda}_j\right), & j = N\,(N-1)+1, ..., N^2-1, \end{cases} \tag{49.12}$$

where we have defined an equilibrium coherence vector, $\mathbf{S_E} = (S_{1e}, S_{2e}, ..., S_{(N^2-1)e})$, and $T_j, j = 1, 2...$, $N^2 - 1$ are phenomenologically introduced nonuniform decay times describing the relaxation of the components of the real state vector towards their equilibrium values, S_{je}. Because of the incoherent processes involved in maintaining the system at a definite level of excitation, the equilibrium system polarization is

zero. This means that the subset of $\mathbf{S_E}$ containing the polarization components of the equilibrium coherence vector is zero, i.e., $\mathbf{S_E^*} = (S_{1e}, S_{2e}, ..., S_{N(N-1)e}) = 0$. The second term in the second line of the Equation 49.12 containing the trace of the density matrix through $\hat{\sigma}$ can be rewritten in terms of nonuniform relaxation times, $T_j = T_{N(N-1)+1}, ..., T_{N^2-1}$, describing the relaxation of the population components of the real coherence vector toward their equilibrium values. Therefore, the total equilibrium coherence vector is of the form: $S_E = (0, 0, ..., 0, S_{N(N-1)+1}, ..., S_{N^2-1})$. We shall refer to Equation 49.12 as the master pseudospin equations.

49.2.3 Vector Maxwell Equations and Coupling to Pseudospin Equations

The optical wave propagation is described by the vector Maxwell equations:

$$\begin{aligned} \frac{\partial \mathbf{H}}{\partial t} &= -\frac{1}{\mu}\nabla \times \mathbf{E} \\ \frac{\partial \mathbf{E}}{\partial t} &= \frac{1}{\varepsilon}\nabla \times \mathbf{H} - \frac{1}{\varepsilon}\frac{\partial \mathbf{P}}{\partial t}, \end{aligned} \tag{49.13}$$

where $\mathbf{P}$ is the macroscopic medium polarization vector given by

$$\mathbf{P} = -eN_a Tr\,(\hat{\rho}.\hat{\boldsymbol{r}})\,, \tag{49.14}$$

with N_a signifying the density of the ensemble of resonant dipoles in the medium. Starting with a phenomenologically constructed dipole-coupling interaction Hamiltonian with coherent Rabi frequency terms describing optical transitions between pairs of levels, one can decompose the dipole moment operator, and represent it as a linear combination of λ-generators, whose coefficients are the corresponding Cartesian dipole moment vector components. Then, using Equation 49.14, we can find a relationship between the macroscopic polarization components, $P_{x,y,z}$, and the coherence vector components, S_j. To illustrate this procedure, suppose we have a transverse electromagnetic wave propagating along the z-axis exciting a given $i \rightarrow j\,(i < j)$ transition in an N-level system. Let the wave be circularly polarized in the plane perpendicular to z, with in-plane electric-field components E_x and E_y. Then, the (i, j) term of the interaction Hamiltonian will contain a complex coherent term $\hbar(\Omega_x \pm i\Omega_y)$, where the sign "+/−" depends on the helicity of the pulse, σ^+ or σ^-. (The sign convention is such that "+" and "−" correspond to, respectively, clockwise and counterclockwise rotation of the electric-field vector, when viewed from a direction opposite to that of the propagation of the pulse.) Note that, for the case of a linearly polarized excitation, the Rabi frequency will be real. The in-plane dipole moment vector can be decomposed into two Cartesian components: $e\mathbf{r} = e(\hat{r}_x\vec{e}_x + \hat{r}_y\vec{e}_y)$, where $\vec{e}_x$ and $\vec{e}_y$ are the unit vectors along the x- and y-axes. The coefficients $\hat{r}_x$ and $\hat{r}_y$ are $\hat{\lambda}$-generators; $\hat{r}_x$ contains unity at positions (i, j) and (j, i), whereas $\hat{r}_y$ contains the imaginary number, $-i$, at position (i, j) and i at position (j, i):

$$\mathbf{r} = r_0 \left\{ \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 & \cdot & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \end{pmatrix} \vec{e}_x + \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & -i & \cdot & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & i & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \end{pmatrix} \vec{e}_y \right\}, \tag{49.15}$$

where r_0 is the dipole length scale. The following general formula can be used to obtain the corresponding $\hat{\lambda}$-generators, depending on the number of discrete levels of the system, N:

$$\begin{aligned}\hat{r}_x &= \hat{\lambda}_{j-i+\sum_{d=1}^{i-1}(\mathrm{N}-d)} \\ \hat{r}_y &= \hat{\lambda}_{j-i+\sum_{d=1}^{i-1}(\mathrm{N}-d)+\frac{\mathrm{N}(\mathrm{N}-1)}{2}}.\end{aligned} \tag{49.16}$$

In fact, this linear combination reproduces the phenomenologically constructed dipole coupling interaction Hamiltonian, $\hat{H}_{int} = -e\mathbf{r}.\mathbf{E}$, when the dipole moment and the decomposed electric-field vector ($\vec{E} = E_x\vec{e}_x + E_y\vec{e}_y$) are substituted in

$$\hat{H}_{\text{int}} = \hbar \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \Omega_x - i\Omega_y & \cdot & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \Omega_x + i\Omega_y & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \end{pmatrix}. \tag{49.17}$$

The macroscopic polarization vector components, P_x and P_y, are then obtained from Equation 49.14, giving

$$\begin{aligned} P_x &= -\wp N_a Tr(\hat{\rho}.\hat{r}_x) \\ P_y &= -\wp N_a Tr(\hat{\rho}.\hat{r}_y). \end{aligned} \tag{49.18}$$

Now we can write down the vector Maxwell equations for a circularly polarized optical wave exciting the $i \rightarrow j$ transition of an N-level system, and thus inducing dipole polarizations P_x and P_y along the x- and y-directions, respectively:

$$\begin{aligned} \frac{\partial H_x(z,t)}{\partial t} &= \frac{1}{\mu}\frac{\partial E_y(z,t)}{\partial z} \\ \frac{\partial H_y(z,t)}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x(z,t)}{\partial z} \\ \frac{\partial E_x(z,t)}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_x(z,t)}{\partial t} \\ \frac{\partial E_y(z,t)}{\partial t} &= \frac{1}{\varepsilon}\frac{\partial H_x(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_y(z,t)}{\partial t}. \end{aligned} \tag{49.19}$$

For the special case of a transverse optical wave linearly polarized along x and propagating in the z-direction, the above system (Equation 49.19) is reduced to

$$\begin{aligned} \frac{\partial H_y(z,t)}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x(z,t)}{\partial z} \\ \frac{\partial E_x(z,t)}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_x(z,t)}{\partial t}. \end{aligned} \tag{49.20}$$

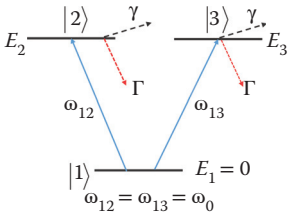


FIGURE 49.3 Energy-level scheme of a V-type three-level system. $\omega_{jk} = (E_j - E_k)/\hbar$, with energy zero chosen at the ground state $|1\rangle$. (Γ, population relaxation rate; γ, dephasing rate.)

We solve Maxwell's curl equations, (Equations 49.19 or 49.20), and the pseudospin equations (Equation 49.12) self-consistently, having established a link between them via the medium's polarization (Equation 49.18). We shall refer to this coupled system of first-order differential equations as the coherent Maxwell's curl-pseudospin model.

49.2.4 Coherent Maxwell-Pseudospin Equations in 2D

Description of nonlinear pulse propagation, interaction, and localization phenomena in multiple dimensions requires dipole coupling to at least a 2D resonant medium, whereby the electric field of a single-photon excitation is coupled to at least two distinct dipole optical transitions. We have previously shown that as a minimum requirement, it is sufficient to consider a degenerate three-level ensemble of dipoles in which two of the allowed dipole optical transitions are excited by each of the two components of the electric-field vector in the waveguide plane [45]. The pseudospin geometrical picture thus allows for adequate modeling of the interaction of an ultrashort laser pulse with a medium in two spatial dimensions. Therefore, we consider a V-type degenerate three-level system of resonant dipoles shown in Figure 49.3, for which a linearly polarized monochromatic electromagnetic wave induces polarizations (dipoles) in two orthogonal directions in the waveguide plane. Two cases arise: a transverse electric (TE) wave, for which $\boldsymbol{E} = (E_x, 0, 0)$ and $\mathbf{H} = (0, H_y, H_z)$, and transverse magnetic (TM) wave, for which $\mathbf{E} = (0, E_y, E_z)$ and $\mathbf{H} = (H_x, 0, 0)$. The respective vector Maxwell equations are

$$\begin{aligned}
\frac{\partial H_y}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x}{\partial z} \\
\frac{\partial H_z}{\partial t} &= \frac{1}{\mu}\frac{\partial E_x}{\partial y} \\
\frac{\partial E_x}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) - \frac{1}{\varepsilon}\frac{\partial P_x}{\partial t}
\end{aligned} \tag{49.21}$$

for the TE wave and

$$\begin{aligned}
\frac{\partial H_x}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_z}{\partial y} + \frac{1}{\mu}\frac{\partial E_y}{\partial z} \\
\frac{\partial E_y}{\partial t} &= \frac{1}{\varepsilon}\frac{\partial H_x}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_y}{\partial t} \\
\frac{\partial E_z}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_x}{\partial y} - \frac{1}{\varepsilon}\frac{\partial P_z}{\partial t}
\end{aligned} \tag{49.22}$$

for the TM wave. We can phenomenologically construct the system Hamiltonian corresponding to each of the cases above:

$$\hat{H}_{TE}(t) = \begin{pmatrix} 0 & \Omega_x & 0 \\ \Omega_x & \omega_0 & 0 \\ 0 & 0 & \omega_0 \end{pmatrix}; \hat{H}_{\mathrm{TM}}(t) = \begin{pmatrix} 0 & \Omega_y & \Omega_z \\ \Omega_y & \omega_0 & 0 \\ \Omega_z & 0 & \omega_0 \end{pmatrix}. \tag{49.23}$$

The master pseudospin equations (Equation 49.12) for a three-level system with $i, j, k = 1, 2, ..., 8$ are coupled to Maxwell's equations above through the macroscopic polarization, Equation 49.14. It thus becomes possible to derive the following relationships between the polarization components and the pseudospin vector components for the case of a TE wave:

$$P_x = -\wp N_a S_1, \tag{49.24}$$

and for a TM wave:

$$\begin{aligned} P_y &= -\wp N_a S_1 \\ P_z &= -\wp N_a S_3. \end{aligned} \tag{49.25}$$

49.2.5 Numerical Methodology

The Maxwell-pseudospin system is discretized on a Yee grid in 1D (modified Yee grid in 2D, see Figure 49.4) and solved directly in the time domain by the finite-difference time-domain (FDTD) method [46]. At each time step, a solution to all equations is obtained using a predictor-corrector iterative scheme, proven to be very efficient in solving simultaneously a large number of first-order differential equations [47]. The time evolution of a discrete multilevel system under an external perturbation can be viewed as a Goursat-type initial boundary value problem, which is well posed when the time history of the initial electric field is given along, e.g., the left (lower) boundary of the simulation domain in 1D (2D). We impose analytical absorbing boundary conditions at the simulation domain boundaries, based on the Engquist–Majda [48] one-way wave equations discretized by the Mur finite-difference scheme [49] using a two-term Taylor-series approximation [46]. To ensure numerical stability, the time step and the spatial discretization are chosen to satisfy the Courant stability criterion: $\frac{c\Delta t}{\Delta z} \leq 1$ in 1D and $\frac{c\Delta t}{\left(\frac{1}{(\Delta y)^2}+\frac{1}{(\Delta z)^2}\right)^{-1/2}} \leq 1$ in 2D [46]. The FDTD numerical technique has numerous advantages over other methods, such as the beam propagation method (BPM). For instance, modeling structures with rapid longitudinal variations (on the scale of the wavelength, λ) of, e.g., refractive index or energy-level population, is impossible to model with BPM, because it relies on the paraxial approximation which in this case is invalid. Moreover, FDTD can handle strong coupling between forward- and backward-propagating fields and can be used to describe general, nonharmonic electromagnetic fields. It can also be used to simulate ultrashort (sub-picosecond) pulses in situations where the pulse bandwidth is significant compared to material resonances.

49.3 Simulation of Nonlinear Dynamics in Optical Waveguides and Semiconductor Microcavities

In this section, we apply the vector Maxwell-pseudospin formalism to the problem of nonlinear optical pulse propagation in planar waveguides and microcavities with embedded quantum systems. We provide a solution in both 1D and 2D. As a first step, we perform validation studies against well-established theoretical and computational methods.

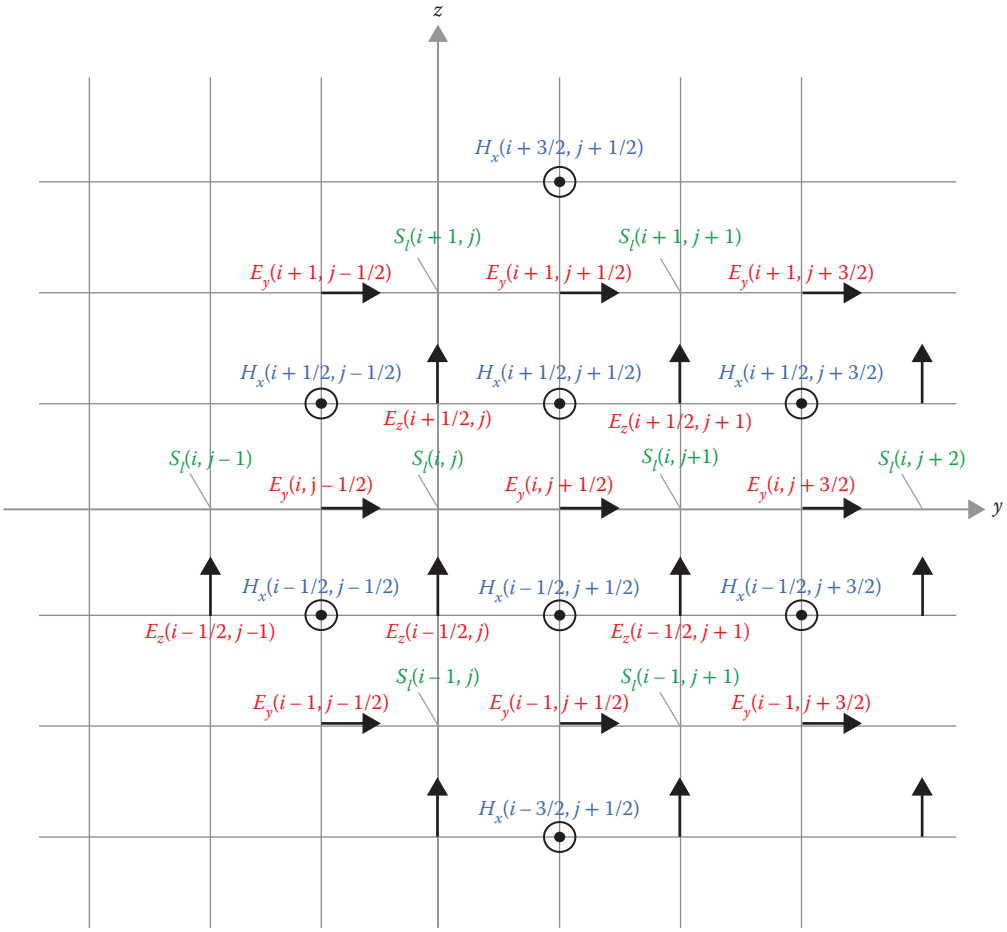


FIGURE 49.4 Modified Yee grid for discretization of the 2D Maxwell-pseudospin system for the purpose of carrying out numerical modeling. The grid is staggered, meaning that the electric and magnetic fields are calculated at every other node. This schematic is for a TM wave: the electric-field vector is in the y–z plane (arrows); the magnetic-field vector is pointing out of the page toward the reader (circle with dot). The electric- and magnetic-field components are spatially separated by $\Delta y/2$ and $\Delta z/2$ in the y–z plane, and temporally by $\Delta t/2$, with the electric-field components situated in the middle of the cells' edges, and magnetic-field components in the center of the cell [46]. The magnetic-field equation is solved at the spatial steps $(i+1/2)\Delta z$ and $(j+1/2)\Delta y$ over time step $(k+1/2)\Delta t$. The electric-field components, E_y and E_z, are solved at the spatial steps $((j+1/2)\Delta y, i\Delta z)$ and $(j\Delta y, (i+1/2)\Delta z)$ over time step $k\Delta t$. Quantum-system variables, S_l, are assigned to the empty nodes in the 2D Yee grid, while the electric-field component values at these empty mesh points are estimated by averaging over the electric fields located at the nearest neighbors in the plane.

49.3.1 One-Dimensional Model of a Two-Level System

Consider a two-level quantum system, such as fundamental heavy-hole transitions in a QD. The coherent Maxwell-pseudospin equations describing the interaction of an optical wave propagating along the z-axis and linearly polarized in a transverse direction along x are given by Equations 49.10, 49.20, and 49.24:

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu}\frac{\partial E_x}{\partial z}$$

$$\begin{aligned}
\frac{\partial E_x}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y}{\partial z} - \frac{\wp N_a}{\varepsilon T_2} S_1 + \frac{\wp N_a \omega_0}{\varepsilon} S_2 \\
\frac{\partial S_1}{\partial t} &= -\frac{1}{T_2} S_1 + \omega_0 S_2 \\
\frac{\partial S_2}{\partial t} &= -\omega_0 S_1 - \frac{1}{T_2} S_2 + 2\frac{\wp}{\hbar} E_x S_3 \\
\frac{\partial S_3}{\partial t} &= -2\frac{\wp}{\hbar} E_x S_2 - \frac{1}{T_1}\left(S_3 - S_{30}\right).
\end{aligned} \tag{49.26}$$

We apply an excitation pulse of duration T_p with either a hyperbolic secant or Gaussian envelope and central frequency, ω_0, in resonance with the two-level transition at the boundary $z = 0$ of the simulation domain. Its time dependence is thus

$$E_x(z=0,t) = \begin{cases} E_0 \operatorname{sech}\left(10\dfrac{t-\frac{T_p}{2}}{\frac{T_p}{2}}\right)\sin\left(\omega_0 t\right) \\ \\ E_0 \exp\left[\left(\dfrac{t-t_0}{t_d}\right)^2\right]\sin\left(\omega_0 t\right), \end{cases} \tag{49.27}$$

where E_0 is the initial pulse amplitude, t_0 is the time moment at which the Gaussian pulse is initially centered, and t_d is the characteristic $1/e$ Gaussian decay.

49.3.2 Passive Cavity Properties

To validate our model, we construct a simple, symmetric test Bragg microcavity containing 5 $\lambda/4$ GaAs/$Al_{0.1}Ga_{0.9}As$/AlAs layers, with λ being 1.4 μm (Figure 49.5f) and compute the passive (cold-)cavity modes and the electric-field standing-wave profile within the cavity. In order to compute the resonant cavity modes, we launch a short broad-band pulse of duration $\leq$ 100 fs from the left boundary of the structure and sample in time the electric-field evolution at the output on the right. This time trace is Fourier-transformed and normalized with respect to the Fourier transform of the input trace, thus yielding the transmission spectrum of the microcavity. The resonant frequency is extracted from the peak in the stop band of the spectrum (Figure 49.5d). Note that reducing the time step allows to resolve higher-order cavity modes. To obtain the electric-field standing-wave profile within the cavity, a continuous sine wave at the resonance frequency (obtained in the previous step) is pumped from the left boundary, and the time evolution of the intracavity field is monitored. The field amplitude increases in time and eventually reaches a steady value. The results shown in Figure 49.5f are compared with the cavity modes and field distributions computed by the transfer matrix method (TMM).

Using the above FDTD-based methodology, we will subsequently investigate passive cavity properties of more complex cavities. For example, we calculate the resonant cavity modes and standing-wave profile of a Coldren-type microcavity structure designed at $\lambda = 1.3$ μm and consisting of 10 GaAs/$Al_{0.9}Ga_{0.1}As$ pairs for the bottom distributed Bragg reflector (DBR) mirror, followed by an AlAs supply layer, a tuning $Al_{0.7}Ga_{0.3}As$ layer, a GaAs $\lambda/2$ cavity and 15 GaAs/$Al_{0.9}Ga_{0.1}As$ pairs for the top DBR mirror. The simulation results for the resonant cavity modes, namely the transmission spectrum as a function of λ and the electric-field standing-wave profile along the structure, are displayed in Figure 49.6.

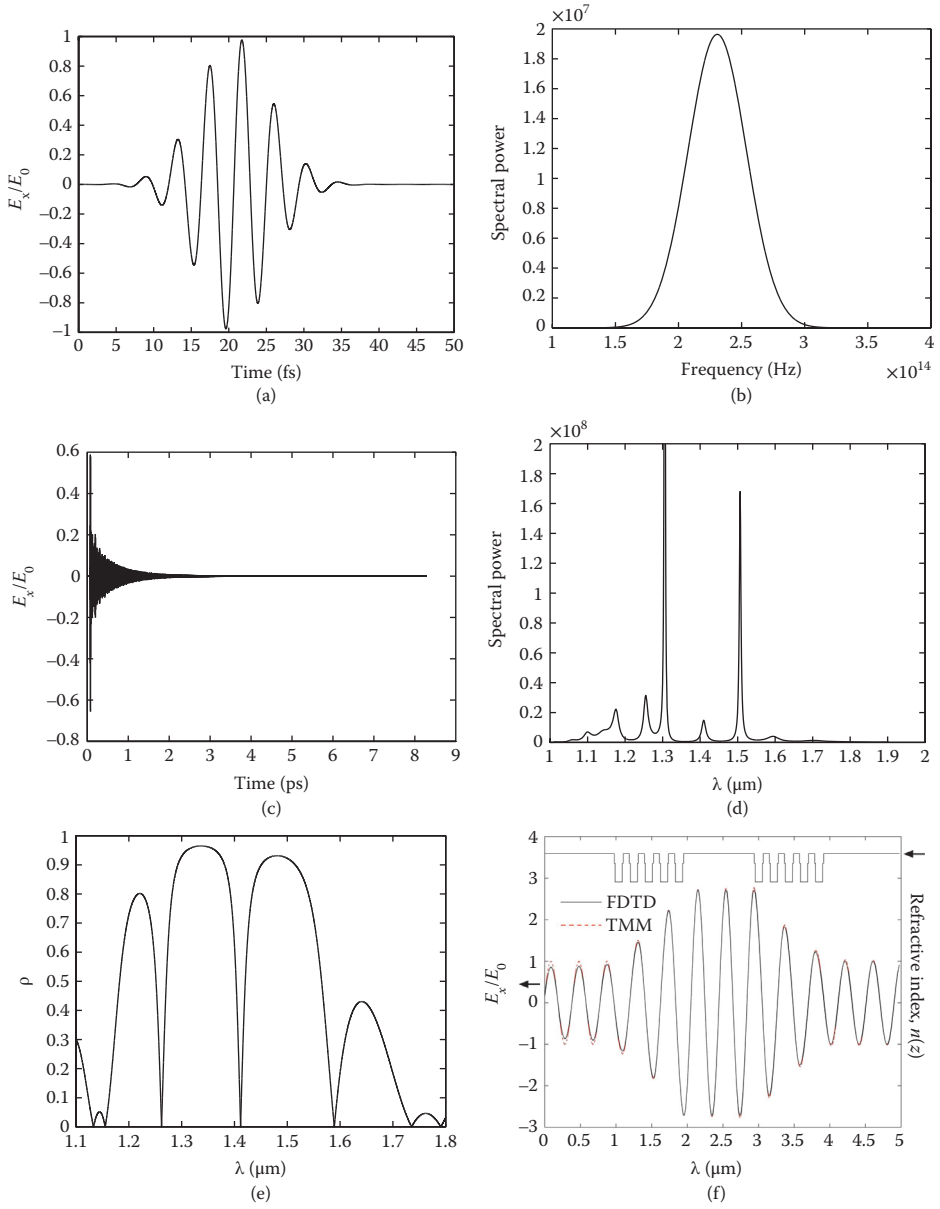


FIGURE 49.5 (a) Initial ultrashort pulse ($T_p \leq 100$ fs). (b) Fourier spectrum of the initial broad-band pulse. (c) Time trace of $E_x(t)$ at the output boundary of the structure. (d) Transmission spectrum as a function of wavelength, λ, computed as the discrete Fast Fourier transform (FFT) of the output time trace in (c), normalized with respect to the input FFT spectrum (b). The resonant cavity wavelength, $\lambda_0 \approx 1.4\,\mu$m, is extracted from the peak in the DBR cavity stop band. (e) Reflectivity spectrum as a function of wavelength computed by the TMM: the resonant cavity wavelength, $\lambda_0 \approx 1.4\,\mu$m is obtained from the dip in the reflectivity spectrum within the DBR cavity stop band. (f) Comparison of the standing-wave profile within the cavity, as obtained by the TMM (dashed curve) and FDTD (solid black curve) methods. The FDTD-inferred instantaneous field standing-wave profile is computed with continuous sine-wave excitation propagating from the left boundary at the cavity resonance frequency equivalent to $\lambda_{pump} = 1.4116\,\mu$m, obtained from (d) after a sufficiently large number of time steps 2×10^6 (each of duration $\Delta t = 3.33 \times 10^{-3}$ fs) with 31,300 spatial grid points (each spatial step of $\Delta z = 1$ nm), Courant number= 0.75. The scale on the right corresponds to the refractive-index profile of the DBR cavity (top).

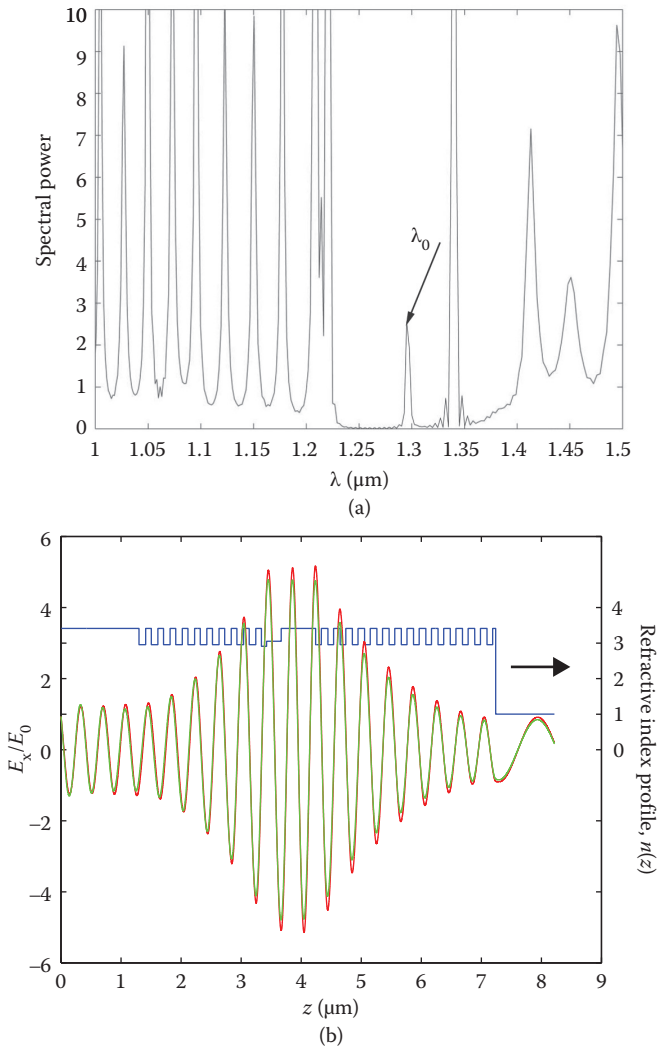


FIGURE 49.6 (a) FFT of the time trace of the signal detected at the output boundary of a Coldren-type microcavity structure, revealing the cavity's (nonnormalized) transmission spectrum. The resonant cavity mode is indicated with λ_0, and corresponds to approximately 1.3 μm. (b) Standing-wave profile after 6 million (dashed curve) and 12 million (solid curve) time steps superimposed on the refractive-index profile of the microcavity (scale on right-hand side). The time step in the simulation was $\Delta t = 3.33 \times 10^{-3}$ fs and the spatial step was $\Delta z = 1$ nm. The amplitude of the standing wave increases as the number of time steps is increased, and eventually reaches a steady maximum value (solid curve). The amplitude is augmented with time due to the build-up of the field caused by multiple reflections at the cavity mirrors.

49.3.3 Active Nonlinear Optical Waveguide/Cavity Properties

We have performed gain validation studies of our 1D semiclassical Maxwell-pseudospin model. This was done by establishing the ability of our model to recover the linear small-signal (unsaturated) gain and nonlinear (saturated) gain of a homogeneously broadened (HB) two-level system, and comparing the result against that of the density-matrix theory, which served as a benchmark (see, e.g., [50]). The test structure used for these validation studies was a 15 μm-long simulation domain consisting of a GaAs slab waveguide

with a gain region length of $L_g = 9\,\mu m$, refractive index $n_{GaAs} = 3.59$, and pumping wavelength of $1.5\,\mu m$ corresponding to the resonant transition frequency of a two-level system sandwiched between two free-space regions (Figure 49.7a). The GaAs slab was represented by an initially inverted ($S_{30} = 1$) two-level medium with resonant dipole density $N_a = 1 \times 10^{24}\,m^{-3}$, dipole transition matrix element, $\wp = 1 \times 10^{-29}\,C \cdot m$ and population relaxation and dephasing times, $T_1 = 100$ ps and $T_2 = 50$ fs, respectively. We then applied a switch-on pulse: a continuous, sinusoidal electric-field pulse, with pulse carrier frequency ω_0 in resonance with the two-level system, which is smoothly increased in amplitude from 0 up to unit amplitude ($E_0 = 1\ V \cdot m^{-1}$) within five periods:

$$E_x(t) = E_0 \sin(\omega_0 t) \begin{cases} 0, & \text{for } t < 0 \\ (1 - x^2)^4, & \text{for } 0 \le t \le 5T_p \\ 1, & \text{for } t > 5T_p, \end{cases} \tag{49.28}$$

with $x = (t - 5T_p)/5T_p$, and T_p being the period of one cycle of the sinusoid. The electric-field amplitude is clearly amplified upon the passage of the sinusoidal pulse across the gain region, thus demonstrating linear small-signal gain (Figure 49.7a). In the time domain, the application of an initial sinusoidal pulse with unit amplitude is amplified at the output end of the structure (Figure 49.7b). Let us compute the complex propagation factor between two points in the gain medium, z_1 and z_2, separated by one dielectric wavelength, $l = z_2 - z_1 = \lambda_0/n$:

$$e^{ik_c(z_2 - z_1)} = \frac{E_x(z_2, \omega)}{E_x(z_1, \omega)}, \tag{49.29}$$

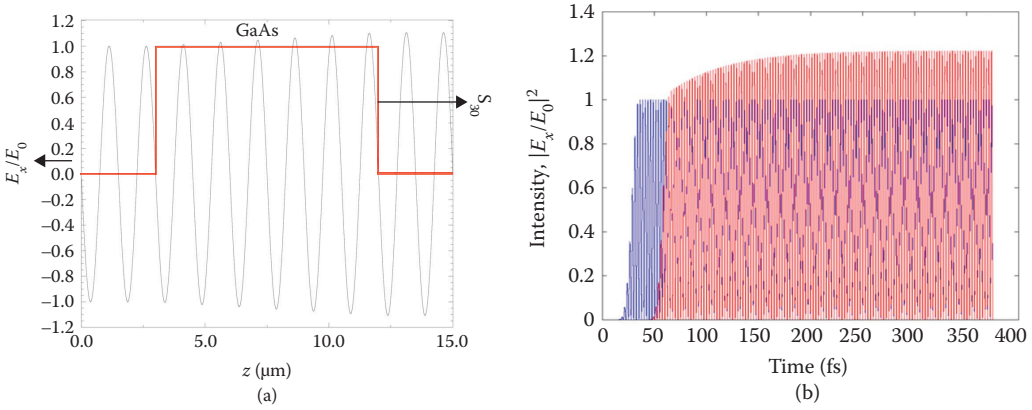


FIGURE 49.7 (a) Slight amplification of a sinusoidal electric-field switch-on pulse upon its passage through an initially inverted two-level GaAs slab medium of refractive index 3.59 after 1×10^6 time steps (each of duration 1×10^{-4} fs) with Courant stability parameter 0.75. The initial population profile across the structure is shown with a thick solid line (top-hat shape). The spatial discretization in this simulation was 1 Å. (b) Intensity evolution in time of the pulse at the input boundary of the simulated structure (fast-oscillating curve on the left) and the amplified time trace of the pulse at the output boundary (time-delayed curve on the right). The fast oscillations are at the resonant pulse carrier frequency, ω_0.

where $E_x(z_i, \omega)$ $i = 1, 2$, are the Fourier transforms of the time traces of the electric field sampled at the two points, and $k_c = \beta + i\gamma$ is the complex wavevector with β-phase shift and γ-gain/absorption coefficient, depending on its sign (+/−). Thus, substituting these back into the left-hand side of Equation 49.29, one can define the amplification and phase factors as $e^{-\gamma.l}$ and $e^{\beta.l}$, respectively. For the amplitude gain coefficient, the steady-state solutions of the density-matrix equations for a two-level system (see [50]) read:

$$\gamma^{ampl}(\omega) = -\frac{\omega_o}{c_o}\frac{N_a \wp^2}{2\varepsilon_o \hbar n_{\mathrm{GaAs}}} g(\omega), \tag{49.30}$$

where n_{GaAs} is the GaAs refractive index, c_0 is the speed of light, ε_0 is the dielectric permeability in vacuum, and the normalized Lorentzian lineshape function is given by: $g(\omega) = \frac{2T_2}{1+(\omega-\omega_o)^2 T_2^2}$, with normalization condition: $\int_{-\infty}^{\infty} g(\omega) d\omega = 1$. The "atomic" phase shift of the optical wave induced by the two-level system is given by

$$\Delta\varphi_{\mathrm{at}} = \Delta k.l = \frac{\omega}{c_o}\frac{N_a \wp^2 (\omega - \omega_o) T_2}{4\varepsilon_o \hbar n_{\mathrm{GaAs}}} g(\omega) = \beta l - n_{\mathrm{GaAs}}\frac{\omega}{c_o} l. \tag{49.31}$$

We can now compare the FDTD simulation results obtained with our Maxwell-pseudospin model (Equation 49.26) with the above analytic expressions (Equations 49.30 and 49.31); the results from the two methods are plotted in Figure 49.8. Our model is in remarkably good agreement with the density matrix-calculated gain/absorption coefficients (a) and phase shift (b) in a HB two-level system.

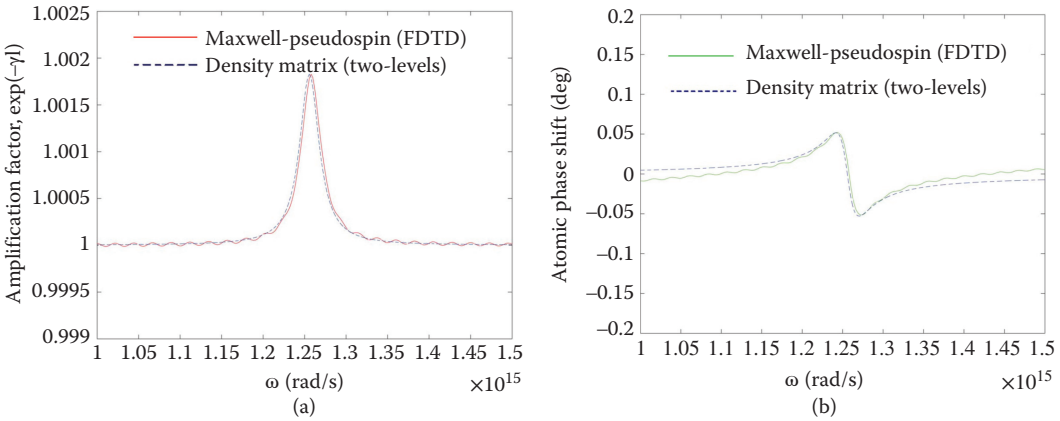


FIGURE 49.8 (a) Amplification spectrum of a switch-on pulse with carrier wave of angular frequency, ω, passing through a GaAs slab waveguide. The spectrum was calculated using two independent methods: by applying the analytical expressions of the density-matrix method to a HB two-level system (Equation 49.30) (dashed curve), and by finding the FDTD solution of the Maxwell-pseudospin set of equations (Equation 49.26) (solid curve). The ω corresponding to the amplification peak is resonant with the optical transition of the two-level quantum system making up the waveguide and is denoted ω_0. Note that the ripple effect observed in the solid curve is an artefact of the FDTD numerical method. (b) ω-dependence of the phase shift (in degrees) of the switch-on pulse emerging from the waveguide. The phase-shift spectrum was inferred independently from the density-matrix method (Equation 49.31) (dashed curve) and the Maxwell-pseudospin method (Equation 49.26) (solid curve). FDTD, finite-difference time-domain.

49.3.4 Resonant Nonlinearities and Gain Saturation

In Section 49.3.3, we employed the analytical steady-state density-matrix equations to perform a successful validation of our Maxwell-pseudospin model against the linear small-signal gain in a HB two-level system. In what follows, we shall show that our model also describes correctly the nonlinear gain dynamics of a HB two-level saturable gain medium. We keep the model structure, source excitation (i.e., input pulse) shape, and material parameters the same as in Section 49.3.3, so that the saturation condition at resonance is satisfied:

$$\frac{\wp^2 E_0^2 T_1 T_2}{\hbar^2} > 1. \tag{49.32}$$

We apply the same source excitation (Equation 49.28) but vary the amplitude, E_0, running separate simulations for different amplitudes, ranging from $E_0 = 5\times10^6\ \mathrm{V\cdot m^{-1}}$ to $E_0 = 1\times10^8\ \mathrm{V\cdot m^{-1}}$. The nonlinear gain dynamics is shown in Figure 49.9 [51]. The gain peaks and levels off after some time, due to the depletion of the population residing on the upper level. The higher the initial excitation, the faster the gain reaches the steady-state field amplitude value. We proceed to demonstrate quantitative agreement between density-matrix theory of saturation of a two-level system and our Maxwell-pseudospin model. We choose the initial electric-field amplitude $E_0 = 5\times10^7\ \mathrm{V\cdot m^{-1}}$, $T_1 = 10\,\mathrm{ps}$, and $T_2 = 10$ fs, again satisfying the resonant saturation condition (Equation 49.32). The saturation value of the population inversion was calculated in [50] to be

$$S_{3\mathrm{sat}} = \frac{1}{1+\left(\frac{E_0}{E_{s0}}\right)^2}, \quad E_{s0} = \frac{\hbar}{\wp\sqrt{T_1 T_2}}, \tag{49.33}$$

giving a value $S_{3\mathrm{sat}} = 0.3$ for the chosen parameter set. We can calculate the unsaturated intensity gain coefficient from the maximum intensity using $\gamma_{\mathrm{int}}^{\mathrm{unsat}} = \frac{ln(I\mathrm{max}/I_o)}{2L_g}$. By definition, the saturated intensity is the intensity of the signal passing through the laser medium for which the gain coefficient is reduced down to one-half of its original value [52]:

$$\gamma_{\mathrm{int}}^{\mathrm{sat}} = \frac{\gamma_{\mathrm{int}}^{\mathrm{unsat}}}{1+I/I_{\mathrm{sat}}}. \tag{49.34}$$

Given that the normalized saturated field intensity is $I_{\mathrm{sat}}/I_0 = exp(\frac{\gamma^{\mathrm{unsat}}}{2}L_g)$ [51], I_{sat}/I_0 is calculated to be 1.01. In order to make quantitative comparison with our model, we run the code with a large number of time steps (4.8×10^7, of 1×10^{-4} fs duration each) to ensure that the field and population dynamics have reached stationary values. In Figure 49.10a, we plot the final instantaneous electric-field distribution and population profile at subsequent time steps, showing the population relaxing to a steady-state value. In the time domain, the time trace at the waveguide output, (Figure 49.10b), and the population inside the gain region (Figure 49.10c), are plotted. It is easy to check that the steady-state values inferred from our model are in excellent agreement with the theoretically computed values. With this we conclude our corroboration of the ability of our model to reproduce correctly the gain saturation dynamics of a HB two-level quantum system.

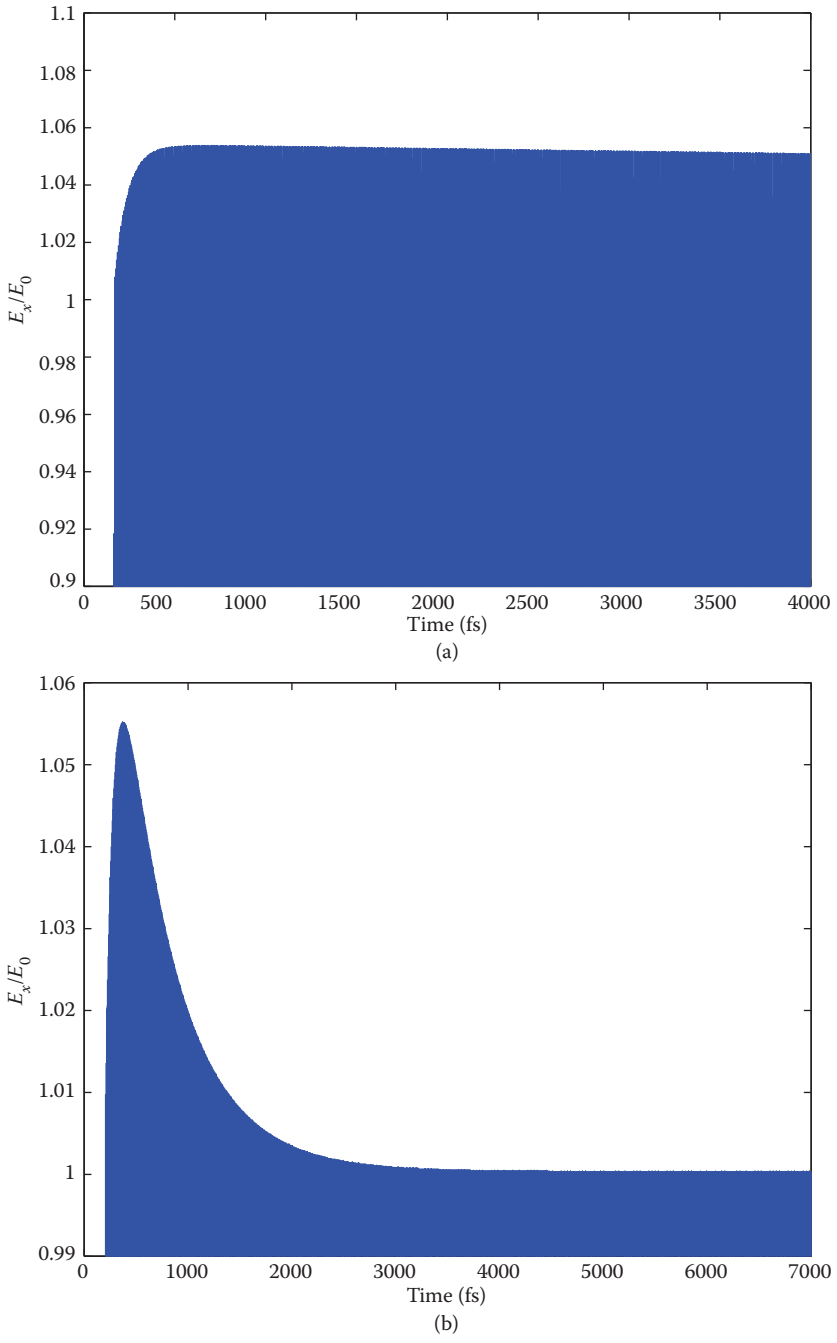


FIGURE 49.9 Gain saturation dynamics of a HB two-level quantum system excited by a source electric field with different amplitude: (a) $E_0 = 5 \times 10^6$ V · m^{-1}; (b) $E_0 = 5 \times 10^7$ V · m^{-1}. The time trace of the amplified electric field is sampled at the output of the simulation domain. In both cases, the amplitude of the oscillations (with frequency ω_0) stabilizes after some time. Note that after peaking, the gain in (a) continues to drop approximately linearly at a relatively slow rate over the entire time span shown on the graph, while the gain in (b) reaches a steady-state value after ~4 ps. (G. M. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

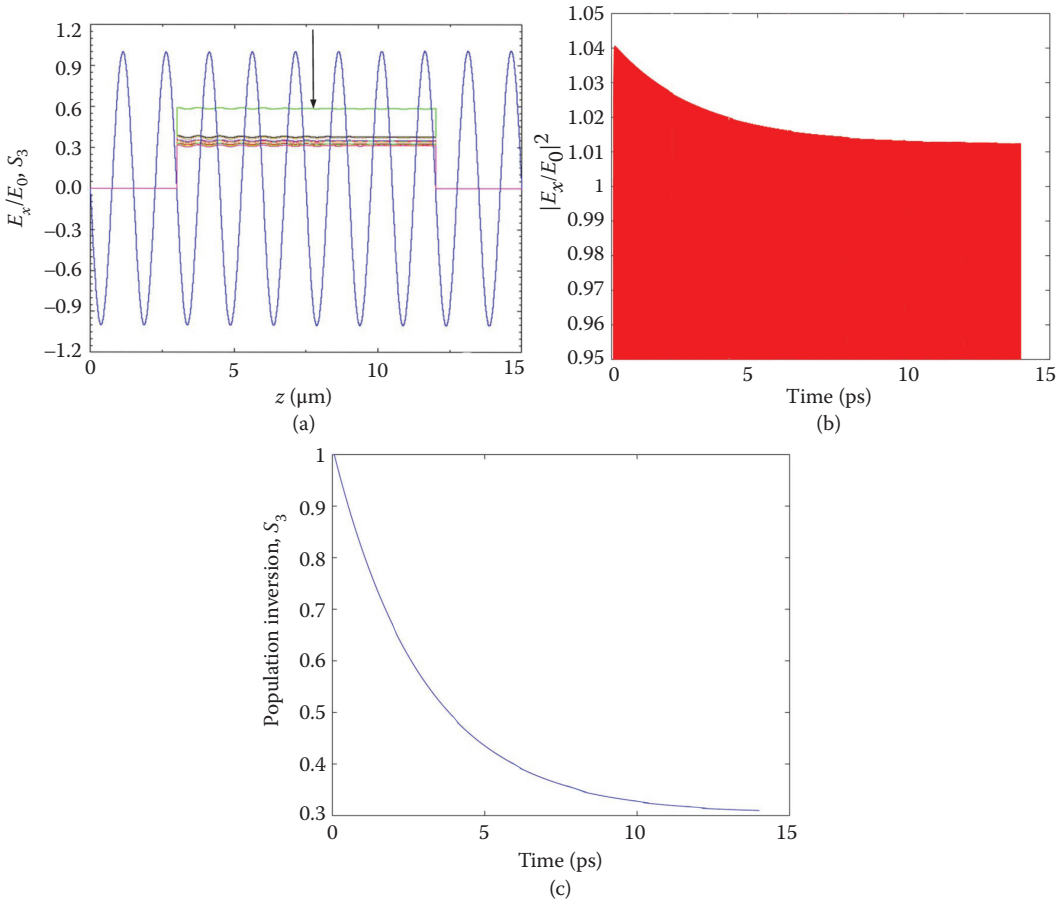


FIGURE 49.10 (a) Instantaneous electric-field amplitude profile across the waveguide (sinusoidal curve), obtained after 4.8×10^7 time steps (corresponding to a total time of ~4.8 ps), and the population inversion profile within the active region at several subsequent time steps (top-hat function with decreasing amplitude). The step in z is $\sim1.8\times10^{-10}$ m and the time step is $\sim1\times10^{-4}$ fs. The population is initially completely inverted: $S_{30} = 1$, but reduces over time (downward-pointing arrow), eventually converging to the theoretically obtained population steady-state value of 0.3. (b) Time evolution of the normalized field intensity showing convergence toward the theoretical value of 1.01. (c) Population relaxation over time, tending toward the theoretically predicted value of 0.3. (G. M. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

49.3.5 Coherent Propagation Effects, Self-Localization and Pattern Formation in 2D Planar Nonlinear Optical Waveguides and Semiconductor Microcavities

49.3.5.1 2D Nonlinear Optical Waveguides

We employ our 2D Maxwell-pseudospin model (see Section 49.2.4), based on self-consistent solution of the time evolution equations of a degenerate three-level quantum system, (Equation 49.4), for $j, k, l = 1, 2, ..., 8$ and Maxwell's curl equations for a TE-guided mode, Equation 49.21 (Equation 49.22 for a TM equivalent), propagating in a planar parallel-plate mirror optical waveguide, or planar semiconductor microcavity [45], of the type shown in Figure 49.11.

It is easy to show that by setting $S_1 = S_1, S_4 = -S_2, S_7 = S_3$, the TE case is equivalent to a two-level system case (Equation 49.7) and the equations for the remaining coherence vector components are decoupled

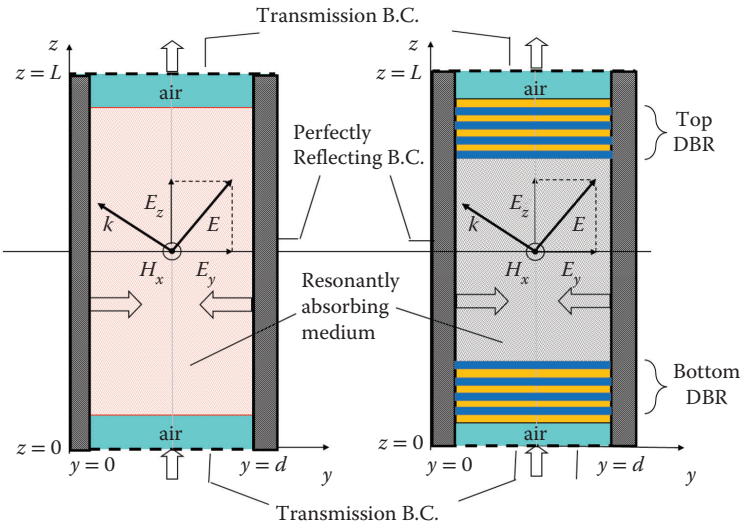


FIGURE 49.11 Planar parallel-mirror waveguide (left panel) and a semiconductor microcavity (right panel) filled with a resonantly absorbing three-level medium. Only TM-guided modes are shown. Perfectly transmitting boundary conditions are imposed on the upper/lower boundaries (upward-pointing arrows) and perfectly reflecting side boundaries are assumed (left/right arrows). The pulse initially propagates through a free-space region (air), then penetrates further, interacting with the degenerate three-level medium, and finally exits through another free-space region. B.C., boundary conditions; DBR, distributed Bragg reflector.

from this system, since there is no coupling to "level 3" in this case. In the FDTD numerical implementation, we use as a source field the TEM (i.e., TM_0) guided mode of a parallel-mirror slab waveguide. The electric field is a plane wave linearly polarized in the transverse x-direction perpendicular to the waveguide plane

$$E_y\left(z=0,y,t\right)=\begin{cases} E_0\mathrm{sech}\left(10\Gamma\right)\sin\left(\omega_0 t\right) \\ E_0\exp\left[\left(\frac{t-t_0}{t_\mathrm{d}}\right)^2\right]\sin\left(\omega_0 t\right) \end{cases} \qquad E_z\left(z=0,y,t\right)=0\,, \tag{49.35}$$

where $\Gamma=\frac{t-\frac{T_\mathrm{p}}{2}}{\frac{T_\mathrm{p}}{2}}$. Here we have allowed for two electric-field envelopes: hyperbolic secant and Gaussian. We should therefore expect that the PAT [18] still holds. The pulse area at a given time is given by

$$\theta\left(z,t\right)=\frac{\wp}{\hbar}\int_{-\infty}^{t}A\left(z,t'\right)dt'\,, \tag{49.36}$$

where the electric-field envelope is as follows:

$$A\left(z,t\right)=E_0\mathrm{sech}\left(\frac{t-\frac{z}{\mathrm{v}}}{\tau}\right)=E_0\mathrm{sech}\left(10\Gamma\right). \tag{49.37}$$

The area under the entire pulse is given by

$$\theta_{\text{pulse}}(z) = \frac{\wp}{\hbar} \int_{-\infty}^{\infty} A\left(z, t'\right) \mathrm{d}t' . \tag{49.38}$$

According to PAT, pulses with areas less than π should be completely absorbed within several absorption lengths; however, pulses with an area of even multiples of π should continue to propagate undistorted as a solitary wave—known as SIT soliton—through the resonantly absorbing medium. We set all resonant dipoles initially to the ground state, i.e., $S_{7e} = -1, S_{8e} = -\frac{1}{\sqrt{3}}$. We then test the predictions of PAT by calculating the pulse amplitude from the PAT, which is given by the following relation for a pulse with a hyperbolic secant-shaped envelope:

$$E_0 = \frac{2\pi\hbar f_0}{\wp \arctan\left(\sinh\left(u\right)\right)\Big|_{-10}^{+10}} \frac{\theta_{\text{pulse}}(z=0)}{2\pi}, \tag{49.39}$$

where $f_0 = 1/\tau_{\text{p}}$ and $\tau_{\text{p}} = T_{\text{p}}/20$, with $T_{\text{p}} = 100$ fs being the pulse duration chosen for this simulation. If we set the initial pulse area $\theta(z=0)$ to 2π, we get a pulse amplitude corresponding to a soliton pulse of $E_0 = 4.2186 \times 10^9\ \text{V} \cdot \text{m}^{-1}$. The simulated waveguide structure is 150 μm long and 50 μm wide. 3D plots of the soliton field modulus (a) and population relaxation term S_7 (b) distributions in the waveguide y-z plane and a slice through the middle of the waveguide's width are displayed in Figure 49.12.

The stable SIT soliton propagation is shown in Figure 49.13a and b: the initial 2π pulse propagates in the resonantly absorbing medium while maintaining its hyperbolic secant shape without any loss. The atoms are driven into the excited state by absorbing ultrashort pulse energy, and by reradiating this energy to the field they return to the ground state. The corresponding population inversion exhibits complete transition from the ground level to the excited state and back to the initial state within one Rabi period. By contrast, for a 0.9π pulse with pulse amplitude $E_0 = 1.18957 \times 10^9\ \text{V} \cdot \text{m}^{-1}$, the initially symmetric pulse is distorted and reshaped during its propagation (Figure 49.13c). The initial equilibrium between the energy absorbed by the leading pulse edge and the energy reemitted by the trailing edge is violated and the pulse becomes increasingly asymmetric. In addition, the system is only partially inverted (Figure 49.13d). The reemitted energy of the continuously growing trailing edge is transmitted through the lower ($z = 0$) boundary. After a sufficiently long time, this pulse should eventually be completely absorbed, in agreement with the PAT.

Let us consider now the more interesting case of a TM_1-guided mode of a parallel-mirror slab waveguide, which cannot be reduced to the two-level system case. We solve self-consistently Maxwell's equations for a TM wave, Equation 49.22, with the pseudospin equations, (Equation 49.4), for a three-level system. As a source field we use the TM_1 slab waveguide mode

$$\begin{aligned}
E_y\left(z=0, y, t\right) &= E_o c_1 \cos\left(\frac{\pi y}{w}\right) \operatorname{sech}\left(10\Gamma\right) \sin\left(\omega_o t\right) = \tilde{E}_y\left(y, t\right) \sin\left(\omega_o t\right) \\
E_z\left(z=0, y, t\right) &= -E_o c_2 \sin\left(\frac{\pi y}{w}\right) \operatorname{sech}\left(10\Gamma\right) \cos\left(\omega_o t\right) = \tilde{E}_z\left(y, t\right) \sin\left(\omega_o t\right) \\
c_1 &= \frac{\sqrt{\frac{\omega_o^2 n^2}{c_o^2} - \left(\frac{\pi}{w}\right)^2}}{\omega_o \varepsilon_o n^2}; \qquad c_2 = \frac{\pi/w}{\omega_o \varepsilon_o n^2},
\end{aligned} \tag{49.40}$$

where w is the waveguide width. In order to obtain a stable soliton solution, we generalize the PAT to multiple spatial dimensions, defining the pulse area as the area below the modulus of the field envelope,

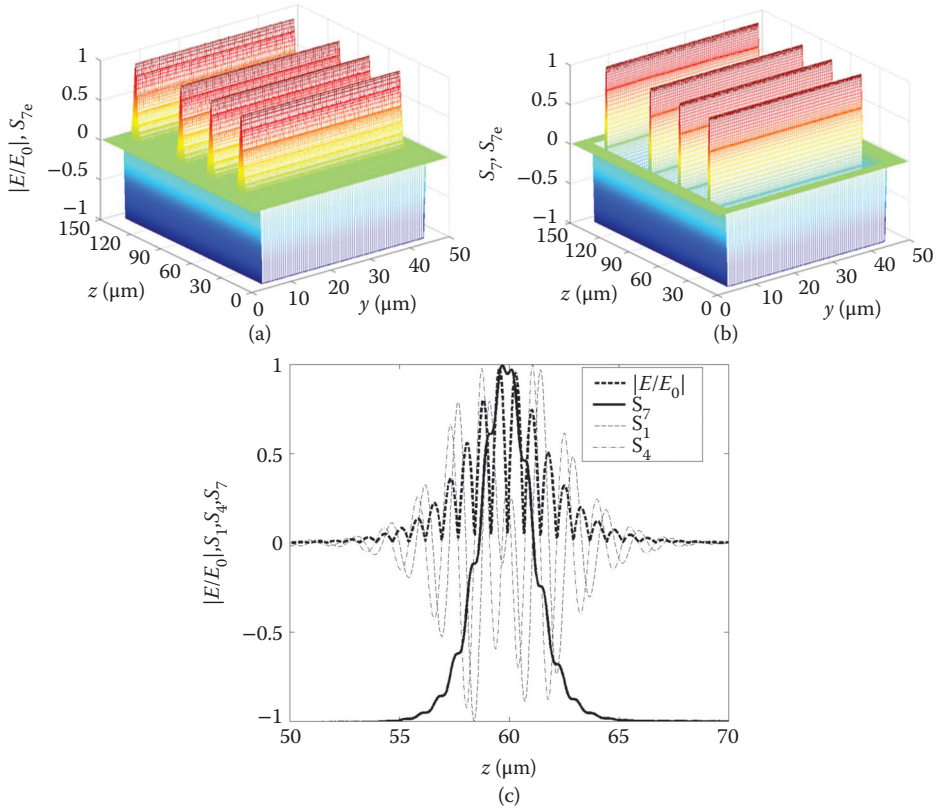


FIGURE 49.12 (a) Superimposed snapshots of the modulus of the electric field for a TEM guided mode propagating in a planar parallel-mirror waveguide at times $t = 150, 250, 350$, and 500 fs after the arrival of the pulse at the left boundary of the waveguide. The initial population profile, S_{7e}, is also shown as an indication of the 2D boundaries of the active medium; the resonant dipoles are initially in the ground state with $S_{7e} = -1$, absorbing the incoming field radiation. (b) Corresponding population inversion term S_7 distribution across the waveguide plane, showing Rabi flops. (c) Slice straight across the middle of the waveguide along the propagation direction, showing the modulus of the electric field, dispersive, and absorptive polarization components, S_1 and S_4, respectively, and the population, S_7, performing a full Rabi flop along the waveguide axis. $\Delta z = 30$ nm, $\Delta y = 500$ nm, $\Delta t = 9.989 \times 10^{-2}$ fs, the pulse wavelength is $\lambda = 1.5$ μm and the duration $T_p = 100$ fs. (Adapted with permission from G. Slavcheva et al., *Physical Review A*, 66, 063418, 2002. Copyright 2002 by the American Physical Society.)

which depends on the point in the waveguide plane we choose to consider:

$$\theta(y,z) = \frac{\wp}{\hbar}\int_{-\infty}^{\infty}\left|\tilde{E}(y,z,t')\right|dt' = \frac{\wp}{\hbar}\int_{-\infty}^{\infty}\sqrt{\tilde{E}_y^2+\tilde{E}_z^2}dt'. \tag{49.41}$$

Assuming that the initial pulse is injected into the active medium at a point $y = y_0$ at the boundary ($z = 0$), stable pulses in the absorbing medium are obtained when

$$\theta(y = y_0, z = 0) = \frac{\wp}{\hbar}\int_{-\infty}^{\infty}\sqrt{\tilde{E}_y^2(y_0,0,t') + \tilde{E}_z^2(y_0,0,t')}dt' = 2n\pi, \tag{49.42}$$

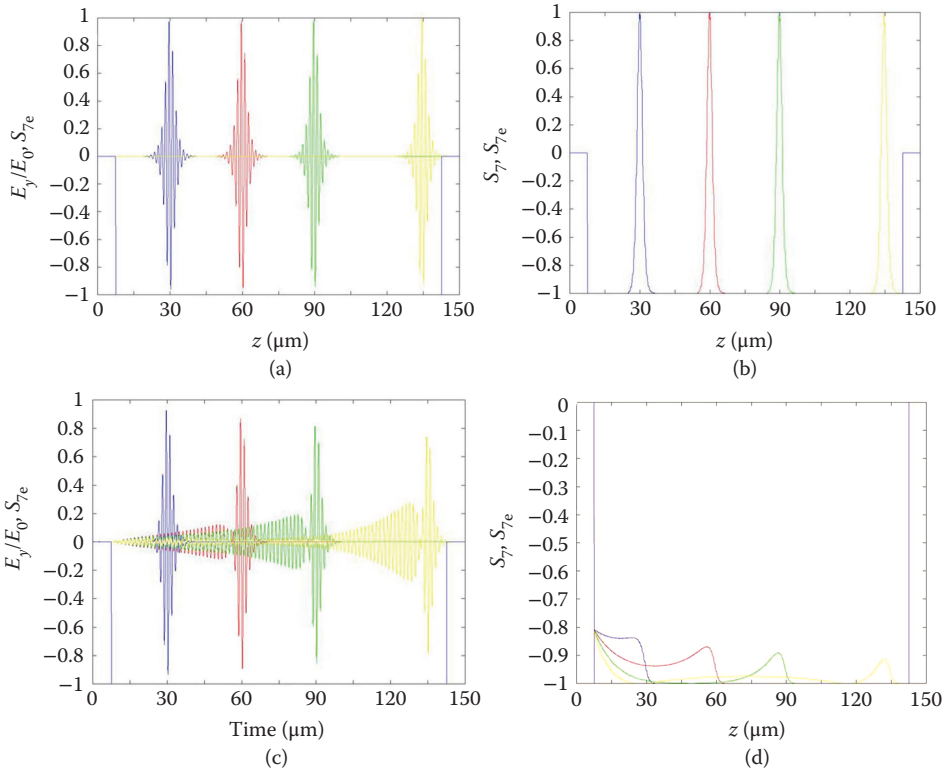


FIGURE 49.13 (a) Time evolution of a 2π initial pulse ($E_0 = 4.2186 \times 10^9$ V · m^{-1}) at $t = 150, 250, 350$, and 500 fs after impinging on the left boundary of a planar parallel-mirror waveguide. The soliton propagates undistorted through the resonant medium, the boundaries of which are shown by the initial population profile, $S_{7e} = -1$, i.e., all dipoles are initially in the ground state. (b) The corresponding population inversion term, S_7, exhibits full Rabi flops, whereby the absorbed energy is fully balanced by the emitted energy over one Rabi period. (c) Time evolution of a pulse with initial pulse area 0.9π ($< \pi$) ($E_0 = 1.18957 \times 10^9$ V · m^{-1}) at the same simulation times. In this case, the pulse becomes distorted due to imbalance of the absorbed and reemitted energies, and will eventually decay to zero in agreement with the PAT. (d) The resonant system is only partially inverted: there is no population swap between the ground and excited levels. (Adapted with permission from G. Slavcheva et al., *Physical Review A*, 66, 063418, 2002. Copyright 2002 by the American Physical Society.)

with n being an integer. A 2π pulse has an initial pulse amplitude $E_0 = 1.123 \times 10^7$ V · m^{-1}. The time evolution of the TM_1 soliton is shown in Figure 49.14.

49.3.5.2 2D Semiconductor Microcavities

Using the 2D model, we numerically investigate conditions of onset of SIT cavity soliton formation in a semiconductor cavity driven by a coherent ultrashort pulse launched from the lower boundary (see Figure 49.11, right panel) [53]. We have designed a number of 2D microcavity structures in order to test the influence of the cavity length on the localization phenomena and pattern formation. A typical 5λ-cavity (where the optical path length between the mirrors of the cavity is equal to an exact multiple of five wavelengths) designed at $\lambda = 1.3$ μm, filled with resonantly absorbing medium sandwiched between a bottom DBR (31.5 pairs of GaAs/AlAs) and a top DBR (24 pairs of $Al_{0.1}Ga_{0.9}As/Al_{0.8}Ga_{0.2}As$) is shown in Figure 49.15a and b. We apply a driving pulse excitation with duration 100 fs at the left boundary

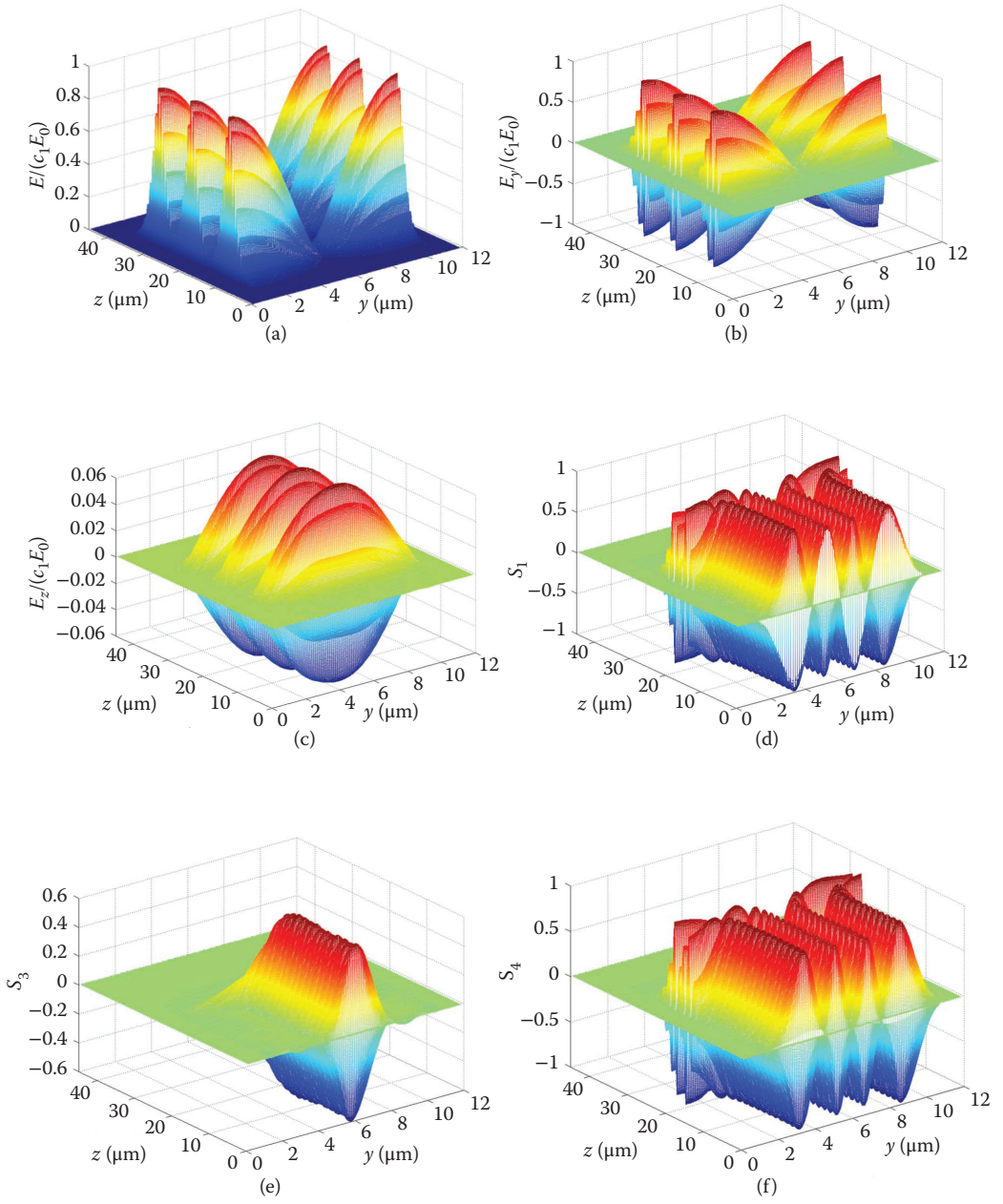


FIGURE 49.14 Time evolution of the (a) modulus, (b) E_y, and (c) E_z electric-field vector component of the TM_1 soliton at times $t = 90, 125$, and 155 fs after its arrival at the left boundary of a planar parallel-mirror waveguide. The 2D pulse travels undistorted in the resonant medium with refractive index, $n = 1$. The waveguide is 50 μm long and 9.9185 μm wide. Discretization grid: $\Delta z = 1.5$ nm; $\Delta y = 119.5$ nm; $\Delta t = 5 \times 10^{-3}$ fs. (d–g) Polarization components of significance (i.e., those corresponding to transitions from one level to another) in a degenerate three-level system. S_1 and S_4: in-phase dispersive components; S_3 and S_6: in-quadrature absorptive components. (*Continued*)

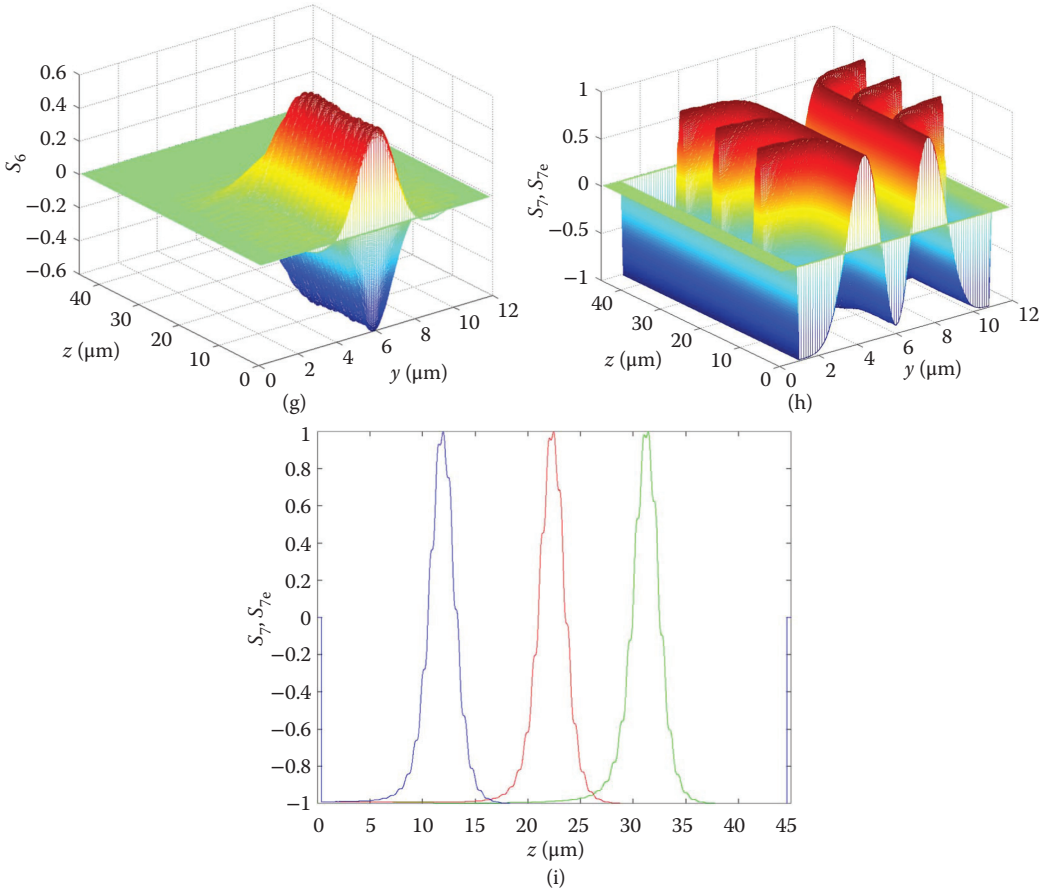


FIGURE 49.14 (*Continued*) (d–g) Polarization components of significance (i.e., those corresponding to transitions from one level to another) in a degenerate three-level system. S_1 and S_4: in-phase dispersive components; S_3 and S_6: in-quadrature absorptive components. (h) Population component S_7 showing population Rabi flops along the waveguide and in a transverse direction (two complete Rabi flops are visible). (i) lengthways cross section of (h) showing complete Rabi flops in the propagation direction, z.

($z = 0$) of the microcavity (Figure 49.15a), which is 12 μm long and $w = 9.92$ μm wide. The dipole number density is $N_a = 10^{24}\,\mathrm{m}^{-3}$ and the dipole matrix element, $\wp = 1 \times 10^{-29}$ C · m. The relaxation times are assumed uniform and equal to 100 ps, satisfying the SIT criterion $T_\mathrm{p} \ll T_i$, $i = 1, 2, \ldots, 8$. The initial pulse amplitude $E_0 = 4.8643 \times 10^9$ V · m^{-1} is calculated for a pulse area of 2π, assuming that the PAT still holds. As pointed out in [54–56], this is not a necessary condition for SIT in multilayered media. Even a relatively weak pulse would lead to SIT intracavity patterns after sufficiently long simulation times, due to field enhancement by constructive interference from multiple reflections in the DBR mirrors. The population inversion gradually builds up with time and begins to perform Rabi flopping. In Figure 49.15c through f, the electric-field modulus and the corresponding population inversion profile are plotted for the cavity region. The clearly visible instantaneous roll patterns for the electric field and population inversion, S_7, feature a quasi-stationary standing-wave (i.e., quiescent) SIT soliton. The cavity induces a quasi-stationary population "grating", and the number of the quasi-complete Rabi flops (10) of the population equals the number of half-wavelengths along the cavity length, $L_c = (10\lambda)/(2n)$. At the

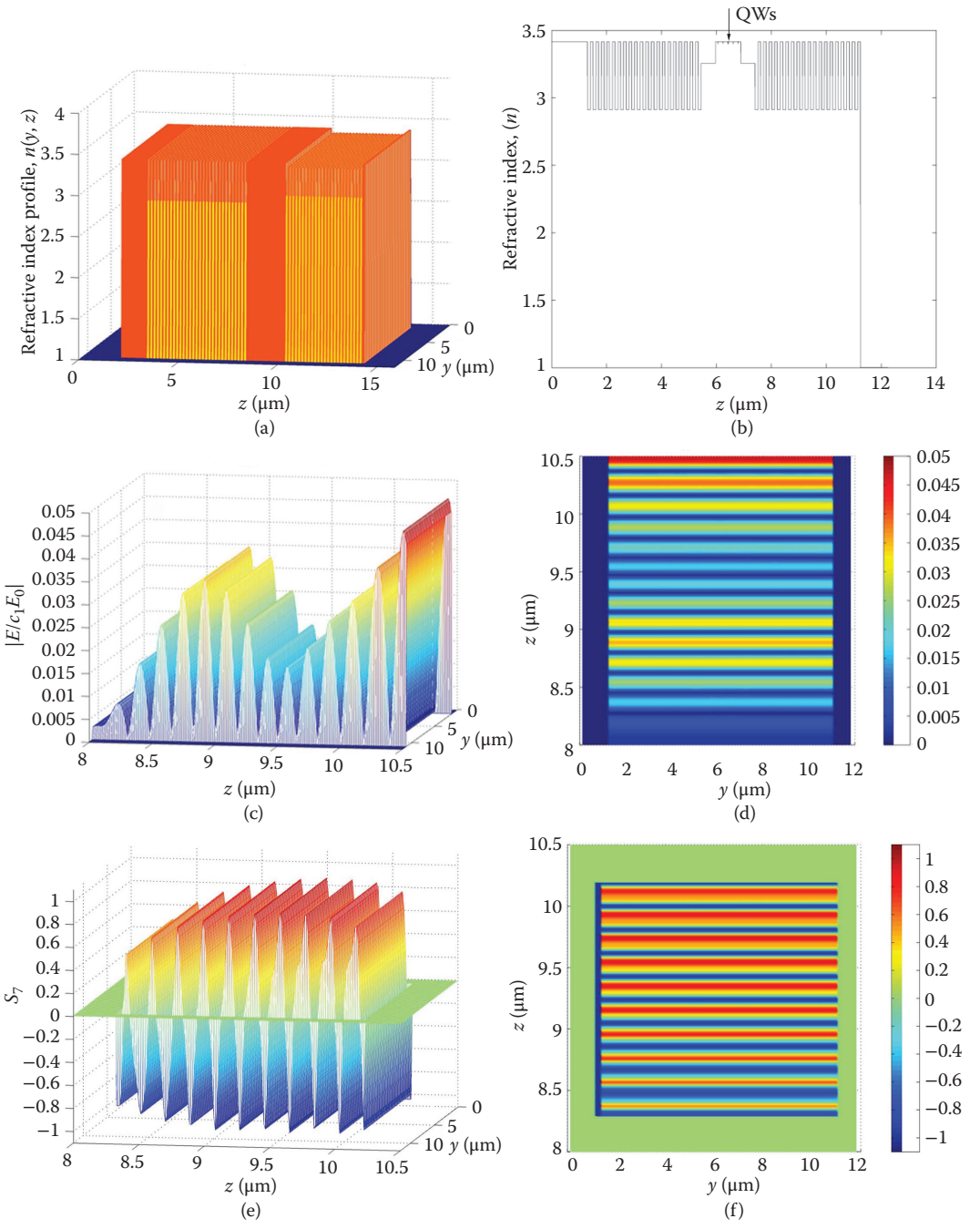


FIGURE 49.15 (a) 2D refractive-index profile across the semiconductor microcavity. (b) Longitudinal slice of (a) at a distance $y_0 = 6\,\mu m$ from the left interface, $y = 0$. (c) A snapshot at time $t = 3.6055 \times 10^{-13}$ s of the electric-field modulus distribution across the cavity. (d) Top view of (c) showing the roll patterns formed. (e) Snapshot of the corresponding population inversion profile across the cavity, showing 10 complete Rabi flops. (f) Top view showing the intracavity population "grating" (roll pattern). (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2003 IEEE.) (*Continued*)

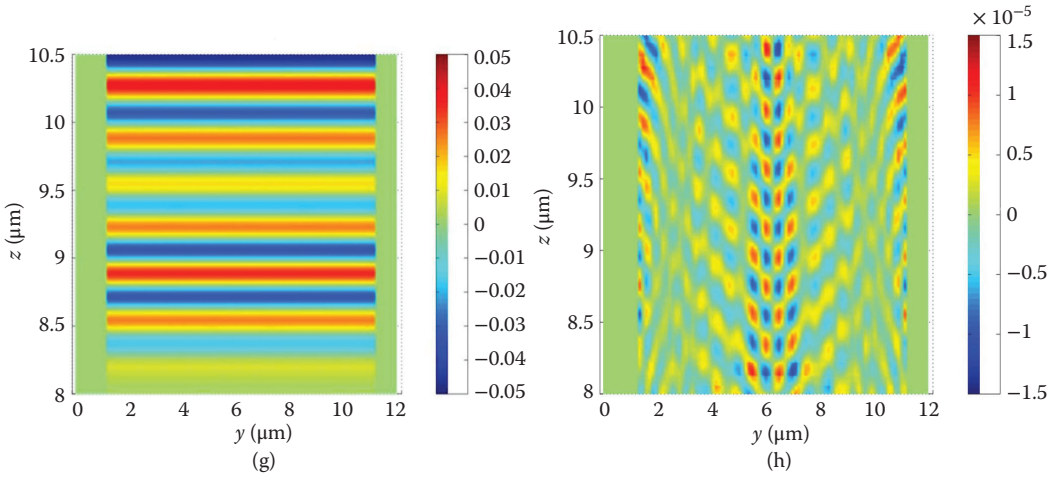


FIGURE 49.15 (*Continued*) (g) Top view of the E_y-component roll pattern. (h) Top view of the 2D pattern of the E_z electric-field component. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2003 IEEE.)

same time, a roll pattern for the E_y electric-field component, Figure 49.15g, and a more elaborate 2D one, Figure 49.15h, for the E_z field component are observed. Performing simulations with different driving pulse durations on a number of different microcavity designs, we demonstrate numerically the existence of a new type of 2D spatiotemporally localized standing-wave soliton ("light bullet"). The SIT cavity soliton results from the interplay of the resonant nonlinearity and the Bragg reflection interference and dispersion in the microcavity.

49.4 Optical Orientation in Quantum Wells and QDs

Circularly polarized optical excitations of semiconductor nanostructures excite particular interest due to their ability to transfer angular momentum and thus to drive the quantum system to a specific angular-momentum state. The conversion between the flying qubit encoded in photon polarization and the material qubit encoded, for example, in the carrier spin, is known as optical orientation, and this technique has been successfully applied to spin-polarized states in semiconductors [14]. This, in turn, opens up the possibility for faithful transmission of flying qubits between distant locations. If sufficiently short (and therefore high-intensity) optical pulses are used, so that they can interact with the system before it can be affected by its environment, the spin does not have time to decohere (i.e., lose its phase information) and could be coherently manipulated by such light pulses. In this high-intensity, coherent regime, optical Rabi oscillations take place; one-half of each oscillation consists of the population being driven from the ground to the excited state (or vice versa), hence performing a Rabi flop. A Rabi flop corresponds to a one-qubit rotation, whereby the spin is rotated through π; this entails a transition either from spin-down (state 0) to spin-up (state 1) or vice versa (Figure 49.1). The achievement of control and manipulation of the single charge-carrier spin dynamics and the collective many-body dynamics of multiple entangled spins is the key to the physical implementation of quantum computation.

49.4.1 Circularly Polarized Optical-Pulse Interactions with Two-Level Quantum Systems: Selective Spin Initialization and Readout

In bulk semiconductors and in semiconductor structures with reduced dimensionality, dipole optical transitions (including excitonic transitions) between energy levels are only allowable subject to satisfying the angular-momentum optical selection rules, $\Delta J_z = 0, \pm 1$, for linear and circular photon polarization respectively, where J_z is the total angular-momentum projection on the electromagnetic wave propagation direction coinciding with the QD growth axis. We shall consider optical transitions with $\Delta J_z = \pm 1$ that are excited by circularly (or elliptically) polarized light, and will therefore apply the pseudospin equations derived for circularly polarized light in Equation 49.8, coupled to Maxwell's vector equations for an electromagnetic wave propagating along the z-direction, circularly (elliptically) polarized in a plane perpendicular to z [57]:

$$\begin{aligned}
\frac{\partial H_x(z,t)}{\partial t} &= \frac{1}{\mu}\frac{\partial E_y(z,t)}{\partial z} \\
\frac{\partial H_y(z,t)}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x(z,t)}{\partial z} \\
\frac{\partial E_x(z,t)}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_x(z,t)}{\partial t} \\
\frac{\partial E_y(z,t)}{\partial t} &= \frac{1}{\varepsilon}\frac{\partial H_x(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_y(z,t)}{\partial t}.
\end{aligned} \tag{49.43}$$

We phenomenologically construct the Hamiltonian corresponding to initially absorbing energy-level schemes (i.e., $S_{30} = -1$), shown in Figure 49.16:

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}} = \hbar \begin{pmatrix} 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) \\ -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \omega_0 \end{pmatrix}. \tag{49.44}$$

The Hamiltonian for an initially amplifying two-level system (i.e., $S_{30} = 1$) is the transposed matrix. Here we have introduced a complex Rabi frequency, whose real and imaginary parts correspond to the $E_{x(y)}$ electric-field vector components, $\Omega_x = \frac{\wp}{\hbar}E_x$ and $\Omega_y = \frac{\wp}{\hbar}E_y$.

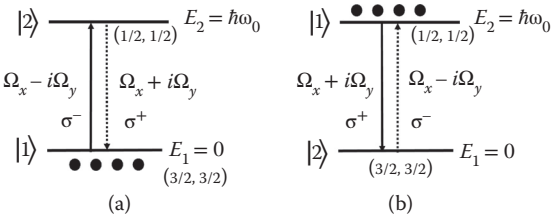


FIGURE 49.16 Energy-level diagram of (a) initially absorbing (i.e., entire charge-carrier population in the ground state) and (b) amplifying (i.e., entire charge-carrier population in the excited state) two-level system, with energy-level separation $\hbar\omega_0$ and allowed dipole optical transitions subject to the condition $\Delta J_z = \pm 1$. States are additionally labeled (J, J_z), where J is the total angular momentum and J_z is its projection along z. (a) A left circularly polarized optical field, σ^-, drives the population into the upper excited level, $|2\rangle$; the population is returned to the ground level by an optical pulse with opposite helicity, σ^+. (b) Having been initially prepared in the excited state, $|1\rangle$, the population in the system is driven by a right circularly polarized pulse, σ^+, to the ground state $|2\rangle$ and returned to the initial excited state by a left circularly polarized pulse, σ^-.

Employing the general methodology in Section 49.2.3, we can easily show that

$$\begin{aligned} P_x &= -\wp N_a S_1, \\ P_y &= -\wp N_a S_2 \,. \end{aligned} \tag{49.45}$$

We solve numerically by the FDTD method the system of equations, Equations 49.43 and 49.8, coupled through Equation 49.45. The source optical fields for left and right circularly polarized light are each given by two orthogonal, linearly polarized pulses, one time-delayed relative to the other by $\pi/2$, with carrier frequency ω_0 tuned in resonance with the two-level system:

$$\sigma^- \begin{cases} E_x(z=0,t) = \tilde{E}_x(t)\cos(\omega_o t) \\ E_y(z=0,t) = -\tilde{E}_y(t)\sin(\omega_o t) \end{cases} ; \; \sigma^+ \begin{cases} E_x(z=0,t) = \tilde{E}_x(t)\cos(\omega_o t) \\ E_y(z=0,t) = \tilde{E}_y(t)\sin(\omega_o t), \end{cases} \tag{49.46}$$

where the pulse shape is determined by the envelopes:

$$\tilde{E}_{x,y}(t) = \begin{cases} E_0 \mathrm{sech}(10\Gamma) \\ E_0 \exp\left[-\left(t-t_0\right)^2 / t_d^2\right], \end{cases} \tag{49.47}$$

which we choose to be either hyperbolic secant or Gaussian. We consider a 15 μm-long simulation structure filled with either a resonantly absorbing or amplifying ($S_{30} = -1$ or $S_{30} = +1$) two-level medium, assumed to be GaAs (refractive index, $n = 3.3827$). We use a pulse wavelength, $\lambda_0 = 1.5\,\mu\mathrm{m}$ and a pulse duration, $T_\mathrm{p} = 30\,\mathrm{fs}$; the dipole coupling $\wp = 1 \times 10^{-29}\,\mathrm{C \cdot m}$, and two-level system ensemble density $N_a = 1 \times 10^{24}\,\mathrm{m}^{-3}$ are taken from [47]. We restrict ourselves to the linear regime (initial electric-field amplitude, $E_0 = 1\,\mathrm{V \cdot m}^{-1}$). The material relaxation times are $T_1 = 100\,\mathrm{ps}$ and $T_2 = 70\,\mathrm{fs}$. We launch a left circularly polarized pulse, σ^-, from the left boundary of the simulation domain and sample the electric field at two locations, z_1 and z_2, along the active medium. The complex propagation factor (see Equation 49.29) over one dielectric wavelength $l = z_2 - z_1 = \lambda/n$ is calculated for the E_x and E_y components, and the results for the gain/absorption coefficient spectra are compared with the steady-state density-matrix theory spectra (see Equation 49.30). A snapshot of the electric-field components, E_x and E_y, at the input and output boundaries are shown in the first column of Figure 49.17. The corresponding Fourier spectra and gain/absorption coefficients are shown in the second and third columns, respectively, of Figure 49.17 for an initially absorbing (Figure 49.17a through f) and initially amplifying (gain) medium (Figure 49.17g through l). The optical wave interacts with the two-level system, driving the population into the upper state (or returning the population to the ground state) by absorbing (or emitting) pulse energy. Therefore, we clearly observe gain/absorption in the Fourier spectra. The gain/absorption coefficients coincide with those for linear polarization. In addition, we have shown that when pumping an absorbing system with a right circularly polarized pulse, the pulse fails to excite the system, and similarly, when pumping an initially inverted system (the whole population residing on the upper level) by a left circularly polarized pulse, the system remains unaffected. Therefore our model correctly describes the excitation using circularly polarized light of dipole optical transitions subject to the optical selection rules, $\Delta J = \pm 1$, in a two-level system. We have thus numerically demonstrated selective excitation of specific spin states by circularly polarized pulses with predefined helicity.

49.4.1.1 SIT and Polarized SIT Soliton Formation

The results in Section 49.4.1 were obtained in the linear (small-signal) regime. Let us now consider the nonlinear coherent regime of ultrashort pulse propagation through resonant media, whose atomic lifetimes greatly exceed the pulse duration (i.e., such that $T_1,\ T_2 \gg T_\mathrm{p}$). The simulation domain is 150 μm long and a two-level resonantly absorbing medium with refractive index $n = 1$ is embedded between two symmetric free-space regions, each with length 7.5 μm. We inject a resonant, left circularly polarized pulse,

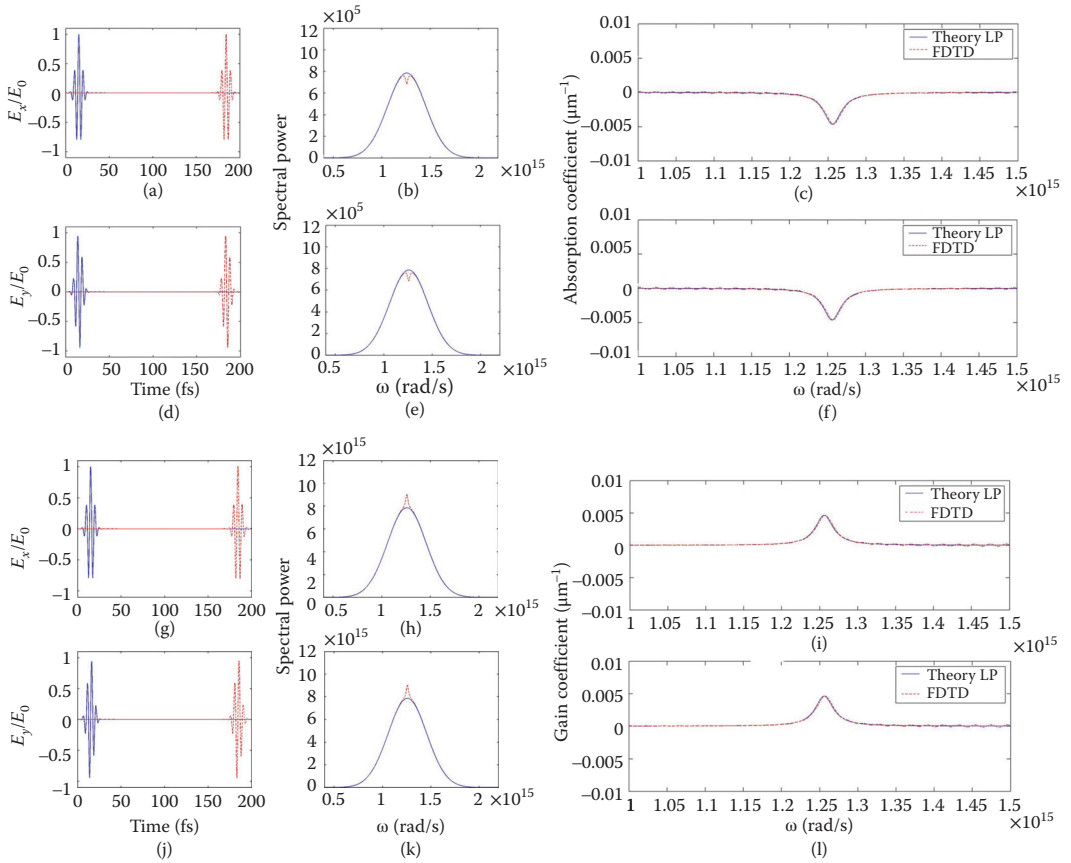


FIGURE 49.17 Time evolution at the input (solid line) and output (dashed line) structure boundaries of the electric-field components, E_x (a) and E_y (d), of a circularly polarized optical pulse with unit amplitude and pulse duration 30 fs for an initially absorbing two-level medium. (b), (e) Fourier spectra corresponding to the E_x and E_y field components, respectively, showing absorption of the pulse (dashed curve dip). (c), (f) Absorption coefficients for E_x and E_y, respectively; in this case, the dashed curve indicates the numerically calculated coefficients for linear polarization, while the solid curve signifies the analytically computed ones (Equation 49.30). (g–l) Respective plots for an initially amplifying (gain) medium; notations are the same as in (a–f). The gain is clearly visible in the Fourier spectra in (h) and (k) (see overshoot of dashed curve). (Adapted with permission from G. Slavcheva and O. Hess. *Physical Review A*, 72, 053804, 2005. Copyright 2005 by the American Physical Society.) (FDTD, finite-difference time-domain.)

σ^-, with hyperbolic secant envelope, since this is the well-known stable soliton solution of the Maxwell-Bloch system (see, e.g., [58]). We choose the electric-field component amplitudes so that the pulse area, according to the PAT, is 2π. This condition implies that $E_0 = 4.2186{\times}10^9$ V·m^{-1}. The time evolution of a 2π pulse is shown in Figure 49.18a. Both circularly polarized pulse components travel undistorted through the resonantly absorbing two-level medium as a solitary wave, and the population inversion is driven through complete Rabi flops by the leading and trailing pulse edges. An expanded view of a snapshot of a circularly polarized SIT soliton at the simulation time $t = 0.2$ ps is shown in Figure 49.18b. By virtue of the PAT, stable soliton solutions in gain (initially inverted, $S_{30} = 1$) medium are obtained for pulse areas which are odd multiples of π. In order to confirm this prediction, we launch a right circularly polarized pulse with amplitude $E_0 = 2.1093 \times 10^9$ V · m^{-1}, which corresponds to a pulse area of π. As expected, a π-pulse completely de-excites the population back to the ground state (Figure 49.18c); the pulse amplitude increases,

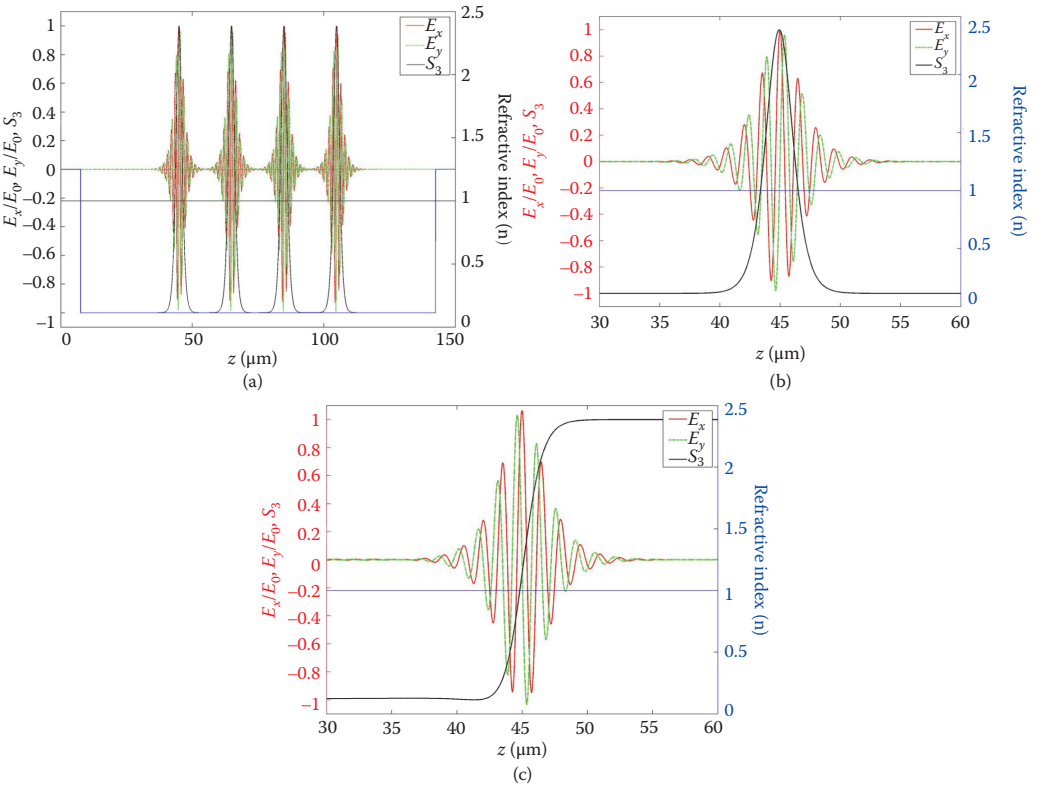


FIGURE 49.18 (a) Time evolution of a $2\pi, \sigma^-$ pulse propagating (from left to right) in a resonantly absorbing medium. The region where the initial charge carrier population profile is $S_{30} = -1$ demarcates the extent of the active medium on the diagram. The normalized electric-field components, E_x and E_y, and the instantaneous population profile, S_3, are plotted at the simulation times $t = 200, 267, 334$ and 400 fs from the moment the pulse is injected at the left boundary of the medium. The pulse travels without loss: it is a SIT soliton. (b) Expanded view of the σ^- SIT soliton in (a) at $t = 267$ fs showing the full Rabi flop of the population inversion. (c) A snapshot of a σ^+ SIT π pulse at $t = 267$ fs that completely inverts the population in the resonantly amplifying medium. (Adapted with permission from G. Slavcheva and O. Hess. *Physical Review A*, 72, 053804, 2005. Copyright 2005 by the American Physical Society.)

while the pulse duration decreases, thereby conserving the pulse area during the pulse propagation in the amplifying medium.

49.4.2 Circularly Polarized Optical Pulse Interactions with Degenerate Four-Level Systems

Let us consider now a more realistic discrete-level quantum system, such as excitonic transitions in bulk or low-dimensional semiconductor systems. For instance, we consider fundamental heavy-hole dipole optical transitions in QDs or quantum wells (QWs), whereby the band structure in the vicinity of the Γ-point ($\mathbf{k} = 0$) in the Brillouin zone (BZ) is isomorphic with a pair of two-level systems corresponding to σ^- and σ^+ heavy-hole excitonic transitions, as shown in Figure 49.19.

Consider a plane electromagnetic wave propagating along z direction, circularly polarized in a plane perpendicular to z, and tuned in resonance with the fundamental heavy-hole transition of the semiconductor system. The resonant nonlinearity is modeled by an ensemble of degenerate four-level systems of the type shown in Figure 49.19, with density N_a. The time-evolution of a four-level system under a dipole-coupling

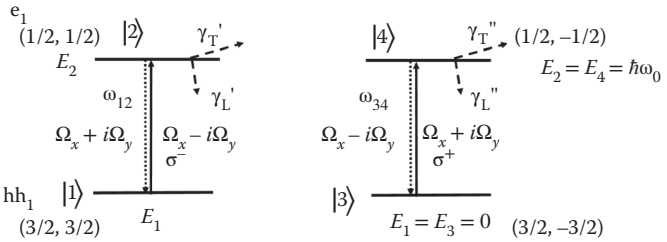


FIGURE 49.19 Energy-level diagram of σ^- and σ^+ heavy-hole transitions in a quantum well near the Γ-point ($\mathbf{k} = 0$) or between s-shell heavy-hole and electron levels in a QD. The energy separation between the upper and lower levels is $\hbar\omega_0$, and the allowed dipole transitions satisfy $\Delta J_z = \pm 1$. Ω_x and Ω_y are, respectively, the real and imaginary part of the Rabi frequencies associated with the coherent transitions. γ_L and γ_T are, respectively, the longitudinal relaxation and transverse dephasing rates of the excited states. e_1 is the bottommost electron energy level in the conduction band, and hh_1 is the topmost heavy-hole energy level in the valence band; together, these energy levels make up the fundamental bandgap.

external perturbation is described in terms of the SU(4) group generators by an equation of motion for a 15-dimensional real state pseudospin vector (see Equation 49.12 for $j, k, l = 1, 2, ..., 15$), where the dissipation in the system is accounted for by introducing nonuniform relaxation times, T_j. Without making any assumptions about the initial population redistribution between the levels, the system Hamiltonian is given by [59]

$$\hat{H} = \hbar \begin{pmatrix} 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & 0 & 0 \\ -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \omega_0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) \\ 0 & 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & \omega_0. \end{pmatrix}. \tag{49.48}$$

We investigate the nonlinear coherent propagation regime, where the SIT condition $T_\mathrm{p} \ll T_j$ is satisfied with $T_\mathrm{p} = 100\,\mathrm{fs}$ and $T_j = 100\,\mathrm{ps}$. A circularly polarized pulse, amplitude-modulated by a hyperbolic secant function and whose carrier-wave frequency is resonant with the optical transition (at $\lambda = 1.5\,\mu\mathrm{m}$), Equation 49.46 and 49.47, is applied at the left boundary $z = 0$ of a 150-μm-long simulation domain, with either resonantly absorbing or amplifying media embedded between two free-space regions of thickness 7.5 μm each. We choose the initial pulse to be a π-pulse, giving $E_0 = 1.093 \times 10^9\ \mathrm{V \cdot m^{-1}}$ according to the PAT. We wish to achieve selective excitation of dipole optical transitions with $\Delta J_z = -1$ or $\Delta J_z = +1$, with a view to manipulate a particular spin state simply by selecting the helicity of the impinging optical pulse. The two leftmost columns of Figure 49.20 show all four possible cases where the initial state population of each two-level system is either entirely in the ground or entirely in the excited state. The two rightmost columns of the same figure also show the outcome of pulse excitation of each respective system. The conclusion that can be drawn is that we can indeed coherently control the spin population of specific spin states in the four-level system by choosing the proper polarization of the injected pulse. We have numerically demonstrated SIT and SIT soliton propagation in a four-level system initially prepared in a state with uniformly distributed populations in either the ground or excited states of the two two-level systems [60]. The ultrashort pulse travels undistorted in the resonant, degenerate four-level medium, Figure 49.21a, and the population is locally driven by the electric field through full Rabi flops simultaneously in both two-level systems, Figure 49.21b (compare with the two-level system soliton, Figure 49.18b).

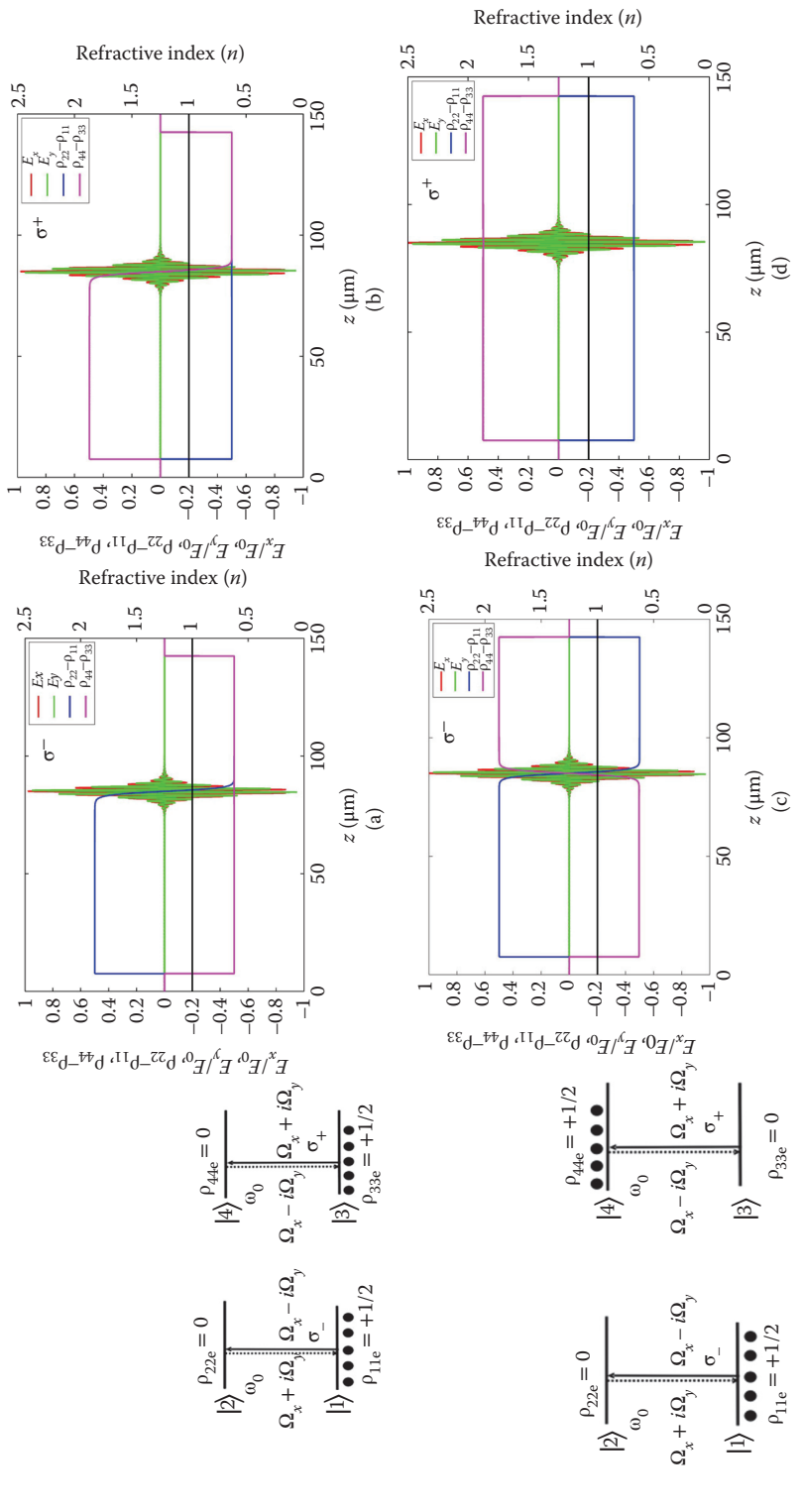


FIGURE 49.20 A snapshot of a circularly polarized pulse at time $t = 333$ fs after injection of the pulse at the medium's left boundary. E_x and E_y electric-field components and the population inversion $\rho_{22} - \rho_{11}$ in the first, and $\rho_{44} - \rho_{33}$ in the second two-level systems are plotted, tracing out the boundaries of the active medium with refractive index, $n = 1$. The leftmost two columns display the initial population distribution between the levels (equal split of the population between the two sets of two-level systems is assumed at the outset). (a) A σ^- π-pulse completely excites the first two-level system, but does not affect the second system. (b) A σ^+ π-pulse completely de-excites the initially inverted second two-level system, but does not affect the first one. (c) A σ^- π-pulse completely excites the first and de-excites the second system. (d) A σ^+ π-pulse does not affect either of the systems: the population in the first remains in the ground state and the population of the second remains in the excited state.

(*Continued*)

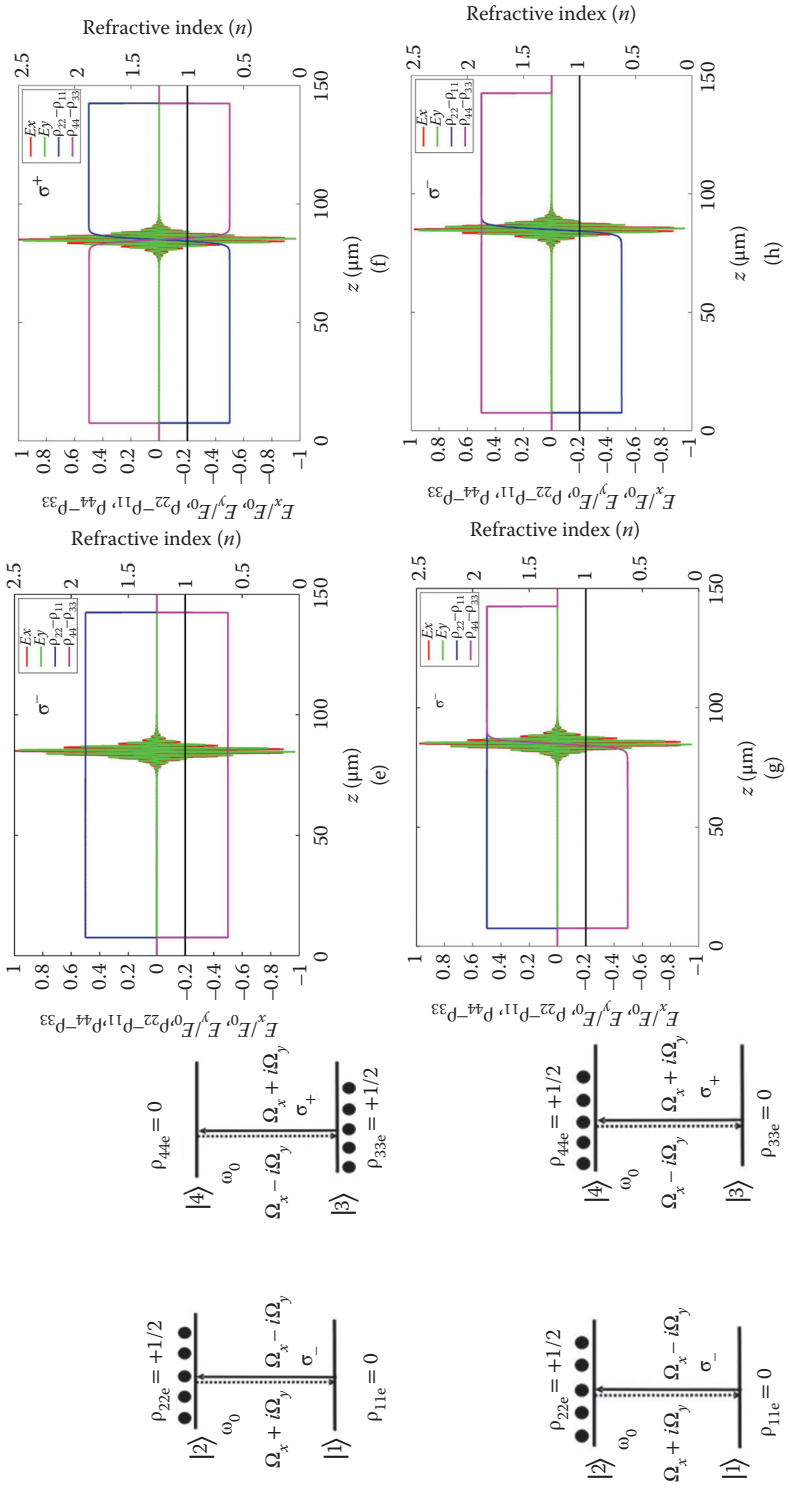


FIGURE 49.20 (*Continued*) (e) A σ^- π-pulse does not change the population in either of the systems. (f) A σ^+ π-pulse inverts simultaneously the population in both systems. (g) A σ^- π-pulse does not affect the first system, but de-excites the second. (h) A σ^+ π-pulse inverts the population in the first system, but does not affect the second system.

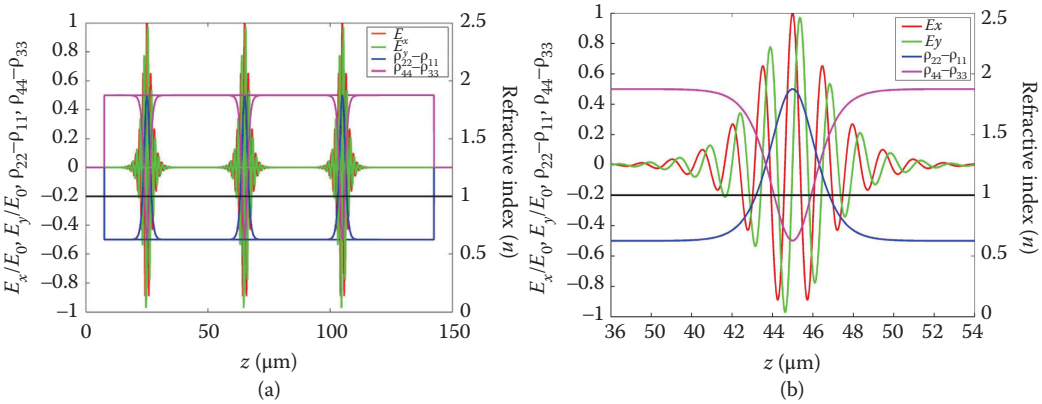


FIGURE 49.21 (a) Polarized SIT soliton propagation through a four-level medium resonant with the carrier wave of the injected pulse. The diagram shows a superimposition of the soliton at three different times: $t = 133, 266$ and 400 fs from the moment the pulse is injected at the left boundary of the medium. (b) A snapshot of the polarized soliton in the four-level medium, consisting of two pairs of two-level systems, at $t = 200$ fs. The medium is initially prepared in a state with the charge-carrier population residing in the lower-lying level, $|1\rangle$, of the first two-level system and in the upper level, $|4\rangle$, of the second two-level system (see Figure 49.19).

49.5 Modeling Coherent Spin Generation, Manipulation, and Readout in Charged QDs

In Section 49.4, we demonstrated coherent, selective spin excitation dependent on the helicity orientation of an impinging ultrashort optical pulse. In this section, we will show how spin states in low-dimensional semiconductor structures can be generated, manipulated, and read out using ultrashort circularly polarized pulses. Our goal will be to compute the time evolution of a discrete-level system and map the allowed dipole optical transitions that occur in a semiconductor nanostructure (e.g., a single QD or an ensemble of QDs) upon ultrashort, circularly polarized optical excitation. The pulse leaves behind a long-lived coherence, and we shall look for characteristic signatures in the dynamics that would allow us to determine the initial spin state and find useful regimes and schemes for high-fidelity optical manipulation of the spin confined in a QD. One promising approach is to address optically individual carrier spins in semiconductor QDs and manipulate them through optically excited states (charged excitons) by employing the techniques of coherent quantum control and optical orientation.

49.5.1 Single-Charged QD Embedded in a Waveguide

We will study n-doped QDs with an overall negative charge of a single electron. Injection of a single electron into a QD can be achieved either by modulation doping of the barrier region, adjusting the impurity doping within a delta-doped layer to transfer on average one electron per dot to the lowest energy states, or by electrical injection. A large class of self-assembled QDs posses quasi-cylindrical geometry about the quantization axis, z, and the electron and hole single-particle states in the conduction and valence band, respectively, can be approximated by harmonic oscillator potentials [61] (Figure 49.22). The resonant, circularly polarized optical excitation of a charged QD leads to the formation of a three-particle complex—a trion—consisting of two electrons occupying the same lowest conduction-band electron level in a spin singlet state, and a hole occupying the lowest valence-band hole level (Figure 49.22c and d). There is strong evidence from both experiment [62] and theory [63,64] that an intense resonant excitation of the trion transition suppresses the electron spin relaxation due to hyperfine interaction of the electron spins with

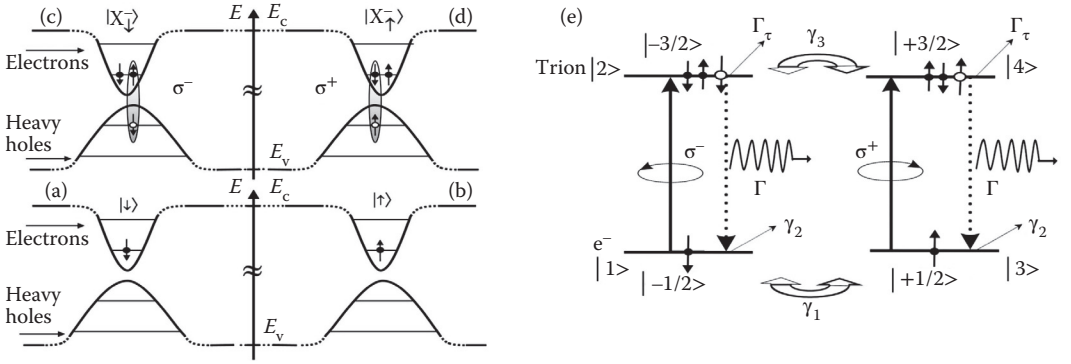


FIGURE 49.22 (a), (b) Schematic of the initial electron states (spin down and spin up) in a singly negatively charged QD and (c), (d) the ground singlet negatively charged exciton, X^-, which forms upon left circularly polarized, σ^-, and right circularly polarized, σ^+, optical excitation of the heavy-hole excitonic transitions. (e) Energy-level diagram of a negatively charged trion (a three-particle complex of two electrons and a hole), equivalent to (a–d); the levels are labeled with the J_z quantum number, and the allowed dipole transitions correspond to $\Delta J_z = \pm 1$ with an energy gap of $\hbar\omega_0$. [Solid upward arrows, coherent optical transitions excited by σ^- or by σ^+; dashed arrows, spontaneous emission rate, Γ; double-headed arrows, transitions due to electron- and hole-spin relaxation, with rates γ_1 and γ_3, respectively; γ_2 and Γ_τ, spin decoherence rates for electrons and holes, respectively.]

the lattice ions' nuclear spins. The latter limits the ability to measure accurately the electron spin orientation at low temperatures. It is obvious that the longer the spin lifetime, the better the chances for optical manipulation and readout of that spin state. We shall investigate both low- and high-intensity nonlinear excitation regimes.

In our treatment of a single quantum object, we shall assume that the ergodic hypothesis holds—i.e., the equivalence of time averages of an observable (in this case, the single-dot polarized, time-resolved photoluminescence TR [PL]) and the ensemble averages at one time over a large number of identical replicas of the dot system. Under such conditions, quantum mechanics allows the prediction of the single-system properties on the basis of macroscopic ensemble averages. Therefore, we shall assume that the time dependence of the optically induced coherent spin generation and subsequent relaxation in a single QD, averaged over a large number of successive measurements, is equivalent to the corresponding spin dynamics of an ensemble of identically prepared QDs. The ensemble of degenerate four-level systems describing the QD resonant nonlinearities is resonantly coupled to an optical wave propagating along the QD growth axis, z, and circularly polarized in a plane perpendicular to z. We use our general methodology to tackle this problem [44] and solve the 1D Maxwell's curl equations (Equation 49.43) coupled to the master pseudospin equations for a four-level system (Equation 49.12) through the induced macroscopic polarization given by

$$\begin{aligned} P_x &= -\wp N_a S_1 \\ P_y &= -\wp N_a S_7. \end{aligned} \tag{49.49}$$

The system under investigation is a GaAs/AlGaAs self-assembled, molecular beam epitaxy (MBE)-grown QD with a height of 5 nm and refractive index, $n_{\text{GaAs}} = 3.63$, sandwiched between two $Al_{0.3}Ga_{0.7}As$ barrier regions, each with a width of 50 nm and refractive index, $n_{\text{AlGaAs}} = 3.46$, at the trion fundamental transition resonance wavelength of 757 nm. The circularly polarized pulse, whose carrier wave is resonant with the trion transition, is described by a hyperbolic secant envelope (see Equation 49.46 and first line of Equation 49.47) and has a duration of 1.3 ps. The simulations are run repeatedly with a different initial

pulse amplitude, E_0, which is varied from a low value, $E_0 = 550\,\mathrm{V}\cdot\mathrm{m}^{-1}$, through $E_0 = 3\times10^6\,\mathrm{V}\cdot\mathrm{m}^{-1}$, representing a π-pulse completely inverting the spin population in the system, to $E_0 = 4\times10^7\,\mathrm{V}\cdot\mathrm{m}^{-1}$, corresponding to a pulse area of 12π and inducing six full Rabi flops. The fundamental trion transition dipole moment is estimated to be $\wp = 4.8\times10^{-28}\,\mathrm{C}\cdot\mathrm{m}$, and the 3D number density of the resonant four-level dipoles, $N_a = 2.5\times10^{24}\,\mathrm{m}^{-3}$, is calculated from the QD surface density to give on average one dot within the microscopic single-dot volume. The trion recombination (spontaneous emission) rate, $\Gamma = 400\,\mathrm{ps}^{-1}$, is taken from [62]; the electron spin-flip rate due to hyperfine interaction with the lattice ions' spins, $\gamma_1 = 0.5\,\mathrm{ns}^{-1}$, has been calculated in [65,66]; the hole spin-flip relaxation is widely assumed to be due to phonon-assisted processes, and an estimate of $\gamma_3 \sim 170\,\mathrm{ps}^{-1}$ [67,68] has been given. The trion-bound electron and hole spin phonon-assisted decoherences are taken as $\gamma_2 = 450\,\mathrm{ps}^{-1}$ and $\Gamma_\tau = 340\,\mathrm{ps}^{-1}$ [69]. We shall assume that the initial spin population resides either in state $|1\rangle$ with spin-down, or in state $|3\rangle$ with spin up. We initially excite the $|1\rangle \rightarrow |2\rangle$ transition by a σ^- pulse. The simulation results for the time evolution of the pulse's electric-field vector components and the populations of all four levels for three initial electric-field amplitudes, E_0, are summarized in Figure 49.23. Excitation of the $|3\rangle \rightarrow |4\rangle$ optical transition is, of course, identical to exciting the $|1\rangle \rightarrow |2\rangle$ transition. The simulated detection of the time-resolved signals shows that there is a sufficiently long time interval—on the order of 400 ps—within which it is possible to determine the spin of the initial states. Indeed, there are two reliable ways for initial spin-state identification: through examining the PL trace profile and through the presence or absence of a second echo pulse after a $\sigma^{+/-}$ π-pulse excitation. The observation of a nonmonotonic PL trace, rather than an exponential decay, is always associated with a spin-up state; in addition, the appearance of a second echo pulse post σ^- π-pulse excitation alludes to a spin up, whereas a repeated echo after a σ^+ π-pulse excitation is indicative of a spin down. The simulations show the onset of the high-intensity Rabi oscillations regime, which suppresses the spin-relaxation processes.

49.5.2 Coherent Spin Manipulation through Hot-Charged Exciton States in QDs

Optical manipulation of the spin of a QD-confined single electron or hole through the resonantly driven trion ground singlet transition (described in the previous section) is considered to be one of the most promising schemes for implementation of spin-based quantum computing, due to the extended spin lifetimes which are limited only by the hyperfine interaction. However, the requirement of a resonantly driven ground trion transition in an inhomogeneously broadened ensemble of charged QDs represents an obstacle for the scalability of architectures based on this quantum system. This problem can be overcome by optical excitation into the excited trion states, taking advantage of a recently discovered effect, herein referred to as "spin filtering". The effect is observed in p-doped QDs under nonresonant, circularly polarized optical excitation, and constitutes an enhanced, photo-induced circular dichroism in the excited trion-state emission, compared to the dichroism exhibited by the ground singlet trion state [70]. The degree of spin polarization is nearly doubled when a high-energy level in the QD is excited resonantly. This is due to an increased spin "injection" efficiency—i.e., the resonant excitation prevents relaxation processes from occurring and the desired spin is maintained. In addition, nonmonotonic dependence of the degree of spin polarization on the optical pulse power is observed, allowing to maximize it by optimizing the pulse characteristics. This spin-filtering effect is promising for realization of high-fidelity schemes for all-optical spin manipulation, since the increased polarization contrast for the two pulse helicities enables highly efficient selective excitation and readout of the spin-up and spin-down populations.

We consider quasi-resonant σ^- or σ^+ circularly polarized excitation of a p-doped QD ensemble into the p-shell in the presence of a resident s-shell hole, and subsequent cascade relaxation to the bright trion ground singlet state, whose decay is detected as PL. The optical excitation creates X^{+*} states, consisting of one electron–hole pair in the s-shell and a resident hole in the p-shell ($1e^1 1h^1 2h^1$), grouped in four degenerate doublets [71]. The exchange interactions can be revealed in the excited exciton spectra, and the energy-level diagram and the spin configurations of the X^{+*} trion are given in Figure 49.24.

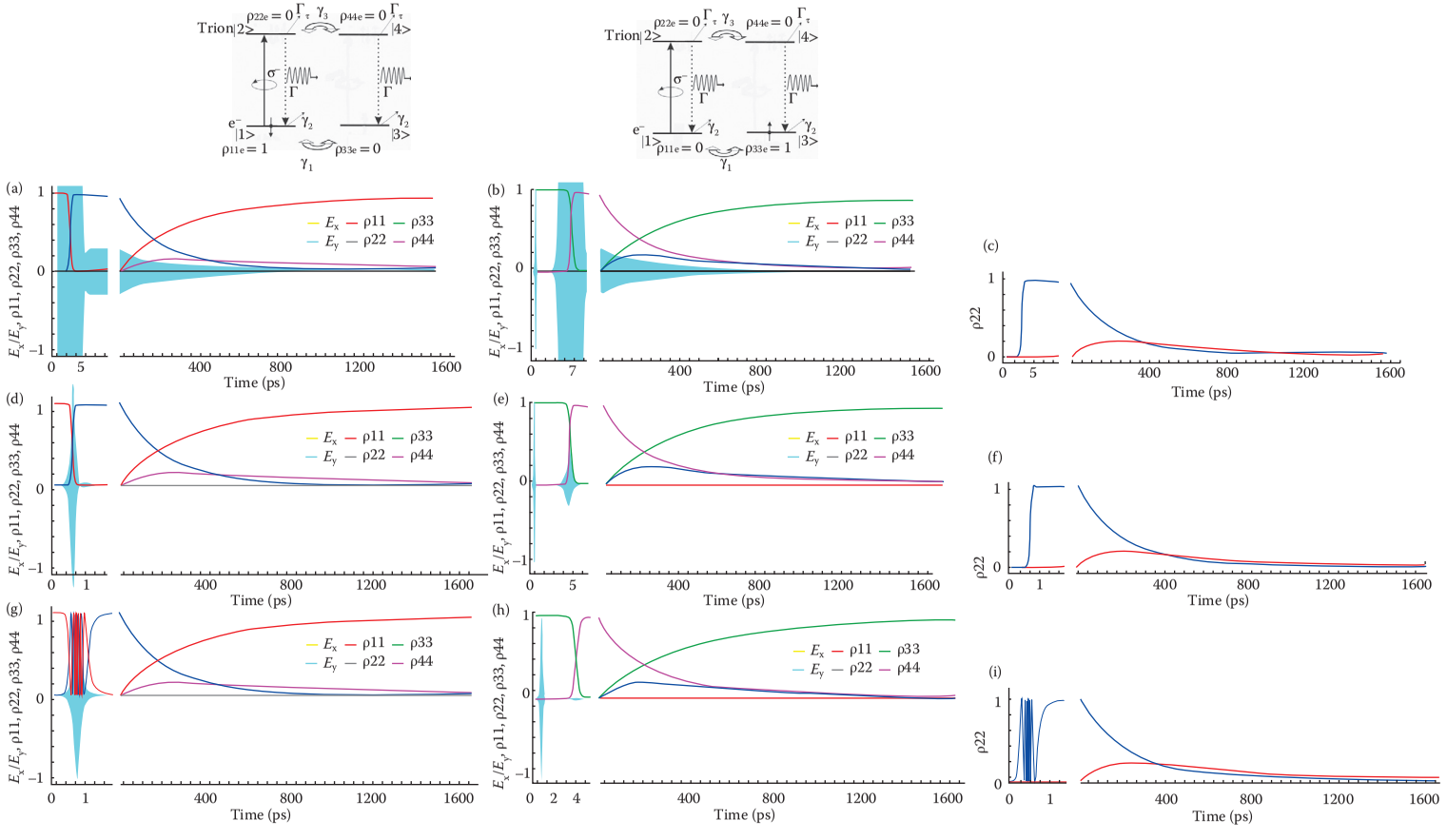


FIGURE 49.23 Top row: initial spin-down (left) and spin-up (right) state. The system is excited by a σ^- polarized pulse. (a–c) Excitation with $E_0 = 550\,\mathrm{V}\cdot\mathrm{m}^{-1}$. Time evolution of (a) initial spin-down state $|1\rangle$; (b) initial spin-up state in $|3\rangle$; ρ_{22} (blue curve) represents the trion $|-3/2\rangle$ state spin population proportional to σ^- polarized PL signal detected experimentally. (c, f, i) TRPL signal for $\Downarrow$ (blue curve) and $\Uparrow$ (red curve) exhibiting different dynamics in the short term: ρ_{22} populations from the graphs to their respective left are plotted; (d–f) Excitation with a π-pulse with $E_0 = 3\times10^6\,\mathrm{V}\cdot\mathrm{m}^{-1}$. Time-resolved dynamics for (d) spin-down initial state; (e) spin-up initial state. (g–i) Excitation with a 12π-pulse with $E_0 = 4\times10^7\,\mathrm{V}\cdot\mathrm{m}^{-1}$. Time-resolved dynamics for (g) spin-down and (h) spin-up initial states. Excitation of initial spin-down state results in six full spin population Rabi flops. (Adapted with permission from G. Slavcheva, *Physical Review B*, 77, 115347, 2008. Copyright 2008 by the American Physical Society.)

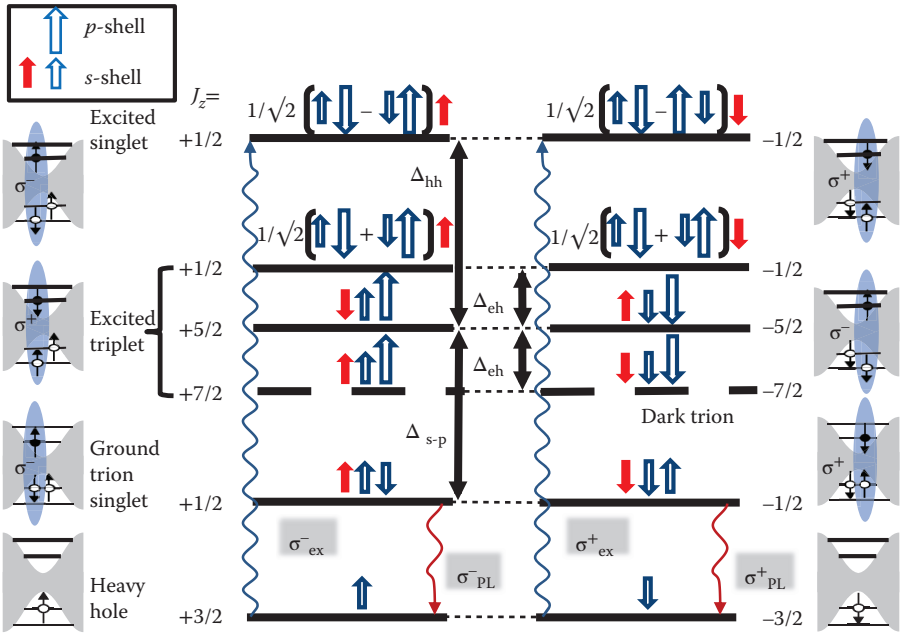


FIGURE 49.24 Energy levels and spin configurations of the hot X^{+*} trion states under σ^- (left six-level system) and σ^+ (right six-level system) optical excitation. The excited singlet is energetically split from the triplet by Δ_{hh}; the excited triplet states are separated from each other by energy Δ_{eh}. The levels are labeled with the total spin projection, J_z, of the electrons and holes. [Small solid arrows, s-shell electron-spin projection; small open arrows, s-shell hole-spin projection; large open arrows, p-shell hole-spin projection; upward wavy arrows, either σ^- or σ^+ excitation; downward wavy arrows, photoluminescence (PL); shaded ellipses, exciton coupling in the radiative states; Δ_{s-p}, energy separation between p-shell and s-shell trion states.]

Spin decoherence mechanisms that need to be considered include the electron–hole anisotropic exchange interaction (AEI), which remains relevant for excited trion states, and the hyperfine interaction between nuclei and either electrons [72] or holes through dipole–dipole interaction [73]. The two degenerate six-level systems for σ^- and σ^+ excitation are coupled through transverse spin decoherence mechanisms, as shown in Figure 49.25.

The system Hamiltonian of a circularly polarized pulse resonantly coupled to an ensemble of six-level resonant absorbers with density N_a is given by

$$\hat{H}^{\mp} = \hbar \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -\frac{1}{2}\left(\Omega_x \mp i\Omega_y\right) \\ 0 & \omega_0 - \Delta_{hh} - \Delta_{sp} & 0 & 0 & 0 & 0 \\ 0 & 0 & \omega_0 - \Delta_{hh} - \Delta_{eh} & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_0 - \Delta_{hh} & 0 & 0 \\ 0 & 0 & 0 & 0 & \omega_0 - \Delta_{hh} + \Delta_{eh} & 0 \\ -\frac{1}{2}\left(\Omega_x \pm i\Omega_y\right) & 0 & 0 & 0 & 0 & \omega_0 \end{pmatrix}, \tag{49.50}$$

where −/+ correspond to σ^-/σ^+ polarization, $\Omega_x = \wp\frac{E_x}{\hbar}$ and $\Omega_y = \wp\frac{E_y}{\hbar}$ are the time-dependent Rabi frequencies associated with the E_x and E_y electric-field components, and ω_0 and $\wp$ are the resonant transition frequency and the optical dipole matrix element of the ground to excited singlet state transition ($|1\rangle \rightarrow |6\rangle$).

In the presence of relaxation processes (Figure 49.25), under time-dependent external perturbation, e.g., a laser pulse, the time evolution of the six-level quantum system is governed by a master equation,

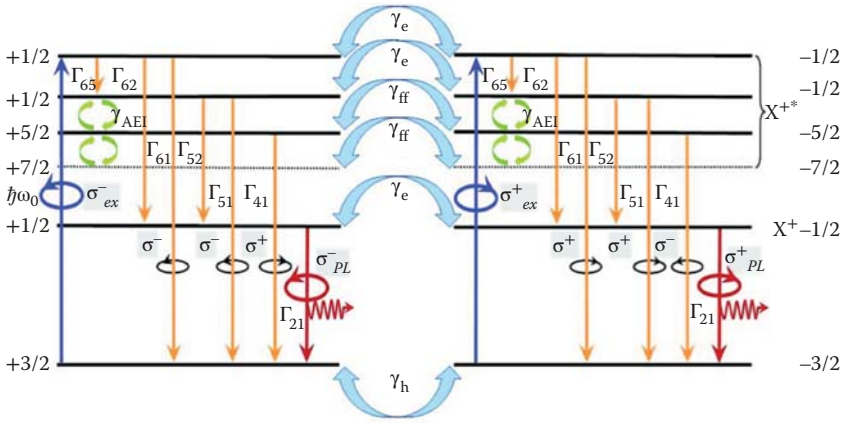


FIGURE 49.25 Discrete-level model of the X^{+*} states in a p-doped QD. The two degenerate six-level systems are coupled via spin decoherence mechanisms represented by double-headed arrows. [γ_h, hole-spin decoherence; γ_e, electron-spin decoherence; γ_{ff}, spin decoherence due to spin flip-flop processes, i.e., anisotropic electron-hole exchange interactions (AEI); ↺ or ↻, circularly polarized transitions; upward arrows, σ^-_{ex} and σ^+_{ex} resonant pumping; downward arrows, radiative or nonradiative transitions; downward arrows denoted Γ_{21}, detected polarized PL; vertical small curved arrows, spin flip-flop coupling due to AEI.]

Equation 49.12, for the $N^2 - 1 = 35$-dimensional real state pseudospin vector, S_j, coupled to Maxwell's curl equations, Equation 49.43 [74], through the macroscopic polarization induced by the electromagnetic pulse:

$$\begin{aligned} P_x &= -\wp N_a S_5 \\ P_y &= \mp\wp N_a S_{20} \end{aligned} \tag{49.51}$$

where −/+ again correspond to σ^- or σ^+ excitation, respectively.

The QD samples modeled are chosen to be identical to the ones used in TRPL experiments, namely MBE grown on a semi-insulating GaAs (001) substrate. The QD layers are sandwiched between two GaAs barriers and the dot areal density is approximately 2×10^{10} cm^{-2} with an average uncapped height of 4 nm, equivalent to a volume dot density, $N_a = 5\times10^{22}$ m^{-3}. The simulation domain consists of an InAs QD layer with nominal thickness given by the height of the typical QD (i.e., 4 nm) embedded between two GaAs barrier regions, each with thickness 50 nm (see Figure 49.26).

The circularly polarized Gaussian pulse at the left boundary of the simulation domain, $z = 0$, is modeled using two x and y linearly polarized waves with resonant carrier frequency, ω_0, phase-shifted by $\pi/2$ according to Equation 49.46 and bottom row of Equation 49.47. The system of equations is discretized in space and time, with spatial and temporal steps, $\Delta z = 1$ Å and $\Delta t = 3.33\times10^{-4}$ fs, respectively, and solved numerically directly in the time domain using the FDTD technique.

For comparison with the theory, PL experiments with short (i.e., 50 ps) σ^- or σ^+ optical excitation pulses with pumping wavelength $\lambda_{res} = 1065$ nm inducing a resonant transition to the excited dot states were carried out on QD ensembles nominally doped with one hole. The polarized PL was detected at the ground singlet X^+ transition, exhibiting a peak at $\lambda_{det} = 1148$ nm. In our simulations, an optical pulse of duration 50 ps is injected at $z = 0$. The pulse center frequency, ω_0, is tuned in resonance with the energy splitting between the ground heavy-hole level $|1\rangle$ and the excited singlet trion state $|6\rangle$: $\Delta E_{1\to6} = \Delta_{hh} + E_{X^+} + \Delta_{s-p}$, where Δ_{s-p} is the energy separation between the s- and p-shell trion states. The ground trion singlet energy, $E_{X^+} = 1.0815$ eV, is determined from the resonant PL spectra at λ_{det}. $\Delta_{s-p} \approx 73$ meV is inferred from the PL spectra [70], and $\Delta_{hh} = 12$ meV is taken in agreement with [71]. In order to calculate the Hamiltonian in Equation 49.50, $\Delta_{eh} \approx 0.5$ meV is used, as obtained in [71,75].

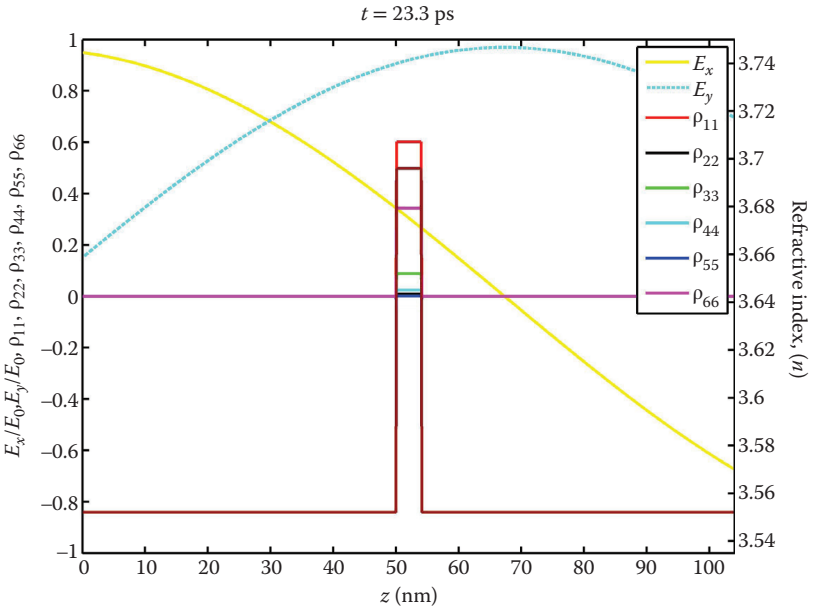


FIGURE 49.26 A snapshot of the spatial distribution of the circularly polarized electric-field components of an excitation pulse interacting with a six-level quantum system. The timing of the snapshot is $t = 23.3$ ps after the source pulse impinges on the left boundary ($z = 0$) of the medium incorporating the quantum system. The charge-carrier spin populations (diagonal density-matrix elements) of all six levels are shown; these indicate the fractional occupancy of each level compared to the total number of charge carriers spread across the different levels. The refractive-index profile of the simulation domain (bottommost curve, right-hand side scale) is produced by an InAs QD layer of thickness 4 nm embedded between two 50-nm-thick GaAs barrier regions.

The pulse area is chosen to be π, so that the pulse completely excites the ground-state spin population to the excited singlet level $|6\rangle$; this corresponds to choosing $E_0 = 2.69 \times 10^5\ \mathrm{V \cdot m^{-1}}$, assuming a dipole matrix element $\wp = 9.83 \times 10^{-29}\ \mathrm{C \cdot m}$. The dipole matrix element's value is comparable to the dipole moment of the X^+ ground singlet transition, and is calculated using the Fermi Golden Rule [50] from the energy separation between the levels and the spontaneous emission rate $\Gamma_{61} \sim 1.2\ \mathrm{ns^{-1}}$, which is close to the experimentally observed value.

The longitudinal relaxation times are taken as follows: the radiative spontaneous decay, $\tau_{21} \approx 1.27$ ns, is experimentally measured from a sample with one hole; spontaneous emission times, $\tau_{41} = 1.35$ ns and $\tau_{51} = \tau_{61} = 1.2$ ns are estimated from the energy-level separation using the Wigner–Weisskopf formula [50]. The nonradiative decay times are taken from theory [76], the transverse spin decoherence rates are obtained from experiment: $\tau_e = 500$ ps [72] and $\tau_h = 14$ ns [73], and a theoretical estimate for $\tau_{ff} = 125$ ps is made on the basis of the Heisenberg uncertainty relations. We should note that there are only two adjustable parameters in our model—namely the largely unknown nonradiative spin relaxation times, τ_{52} and τ_{65}.

The theoretically computed TRPL traces that are displayed in Figure 49.27 clearly exhibit photo-induced circular dichroism in the polarized TRPL emission; this is manifested as a different evolution of PL in time, depending on the pulse helicity used. In contrast to the widely used rate equations model which assumes mono-exponential decays, the dynamics of the quantum system described by the present model is nonlinear and, as a result, the simulated PL decay is not mono-exponential.

The experimentally detected TRPL trace shape strongly depends on detector characteristics: the faster the detector's response and the higher its sensitivity, the better the agreement between theory and

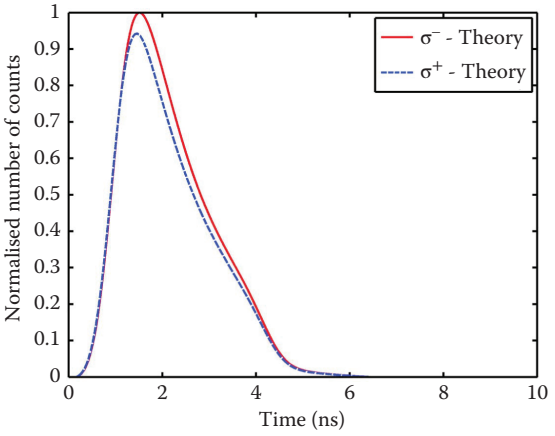


FIGURE 49.27 Theoretically computed polarized TRPL traces for σ^- and σ^+ excitation of a six-level QD. The presence of circular dichroism is clearly visible.

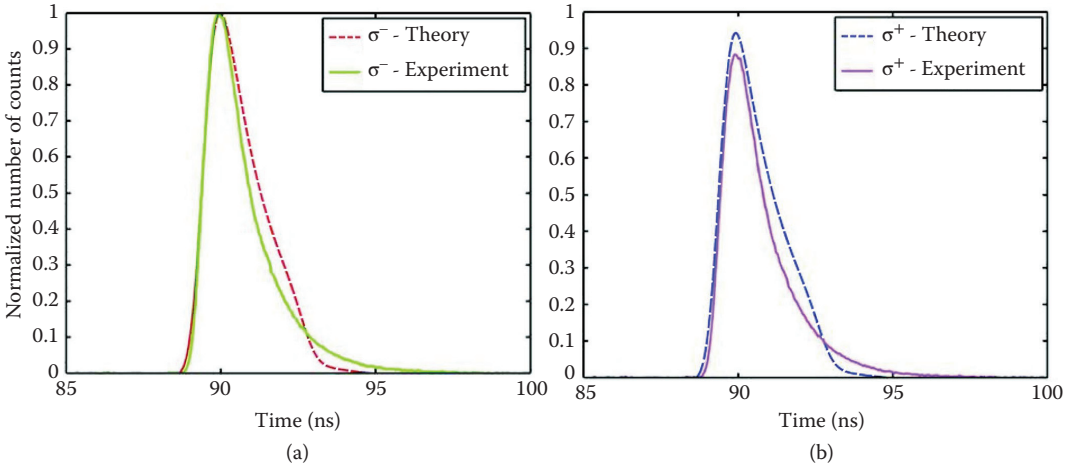


FIGURE 49.28 Comparison between the simulated TRPL trace (after convolution with the detector's response function) and experimentally detected TRPL trace for (a) σ^- excitation and (b) σ^+ excitation of a six-level QD.

experiment. For proper comparison with experiment, we convolve our theoretically calculated polarized traces with the detector's response function and plot them along with the experimental traces in Figure 49.28.

There is excellent agreement between the theoretical and the experimental results. The rising slope and initial decay after the peak in the experimental trace are almost perfectly reproduced by our simulations. The small discrepancy at longer times can be attributed to several factors. Firstly, it is not possible during the course of an experiment to determine precisely which hot trion state is being resonantly excited, so the assumed λ_{res} is only approximate. Secondly, Δ_{s-p} is inferred from the resonant PL spectra and is only a rough estimate, since the excited state does not correspond to a sharp spectral line, but rather a broadened line of finite width; in addition, the inhomogeneous broadening has not been taken into account. The lower peak height for σ^+ excitation can be attributed to the estimated value for the spin flip-flop coupling rate and requires further investigation.

Our theoretical model explains quantitatively the origin of the spin-filtering effect. The experimentally observed enhanced time-resolved circular dichroism in the excited trion state emission is shown to emerge

from the dynamical spin flip-flop coupling between the two degenerate sets of excited charged excitonic states. The theory is in very good agreement with experiment, thereby allowing to obtain an estimate for largely unknown intra- and intershell spin-relaxation timescales. The approach also allows to predict optimum pulse parameters, such as power and duration, for efficient control of spin dynamics.

49.6 Quantum Stochastic Formalism for Modeling Cavity-Emitter Systems

Future advances in information systems are expected to exploit the quantum nature of their constituent parts—e.g., in the development of quantum simulators and design of architectures for quantum cryptography. The implementation of these technologies necessitates a fuller understanding of the underlying quantum systems, which in turn will require several fundamental aspects of quantum theory to be tested.

One example of a quantum-optical phenomenon is quantum noise. In contrast to the more familiar classical noise, which can be regarded as spontaneous random fluctuations from a steady state, quantum noise arises from the Heisenberg uncertainty relations, which are fundamental to quantum theory [77]. The concept of quantum noise stems from the statistical interpretation of quantum mechanics and our inability to access any particular one out of the infinite number of degrees of freedom of the electromagnetic field propagating in free space. Quantum noise is closely related to spontaneous emission, irreversible decay and the origin of the spectral line width described by the theory of Weisskopf and Wigner [78], who are considered to have laid the foundations of the quantum noise theory.

At large photon densities, quantum fluctuations are necessary to explain laser linewidth and the threshold properties. By contrast, at low photon densities, light is generated by spontaneous emission described in terms of single-photon emission events within the single-particle Dirac [79] picture. Control over such noise effects will be crucial in the design of devices for quantum information processing.

49.6.1 Langevin Formalism for Modeling Quantum-Optical Effects

Considering a two-level system with a few degrees of freedom immersed in a radiation-field heat bath with an infinite number of degrees of freedom, we exploit the quantum-classical correspondence in the presence of noise and employ the Langevin formalism. The essence of the Langevin approach lies in the realization that the action that the many unknown "bath" (i.e., external) variables have on the system is to modify its deterministic equation of motion, and that this can be accommodated by the inclusion of apparently random terms, known as "Langevin forces."

49.6.2 Spontaneous Emission of a QD Ensemble in a Semiconductor Microcavity and Onset of Lasing

Consider the semiconductor microcavity structures shown in Figure 49.29, designed at $\lambda = 850\,\text{nm}$ and $\lambda = 1.29\,\mu\text{m}$. The cavity of the first structure (bottom DBR: 35.5 pairs $\text{AlAs}/\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$; top DBR: five pairs $\text{AlO}/\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$) is filled with a resonant two-level medium ($\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$), whereas within the second cavity, there are six ($\text{Ga}_{0.63}\text{In}_{0.37}\text{N}_{0.012}\text{As}$) quantum wells filled with the resonant medium. To describe the interaction of an optical wave with these systems, we use the semiclassical 1D Maxwell–Bloch equations in the real vector pseudospin picture, Equation 49.26, for an optical wave linearly polarized along x propagating in the z-direction. In order to model the spontaneous emission of two-level atoms, we add a random electric-field fluctuation, δE_x, at each time step:

$$\left(E_x(z,t) + \delta E_x(z,t)\right) = -\frac{1}{\varepsilon}\frac{\partial H_y}{\partial z} - \frac{N_a \wp}{\varepsilon T_2} S_1 + \frac{N_a \wp \omega_0}{\varepsilon} S_2 \,. \tag{49.52}$$

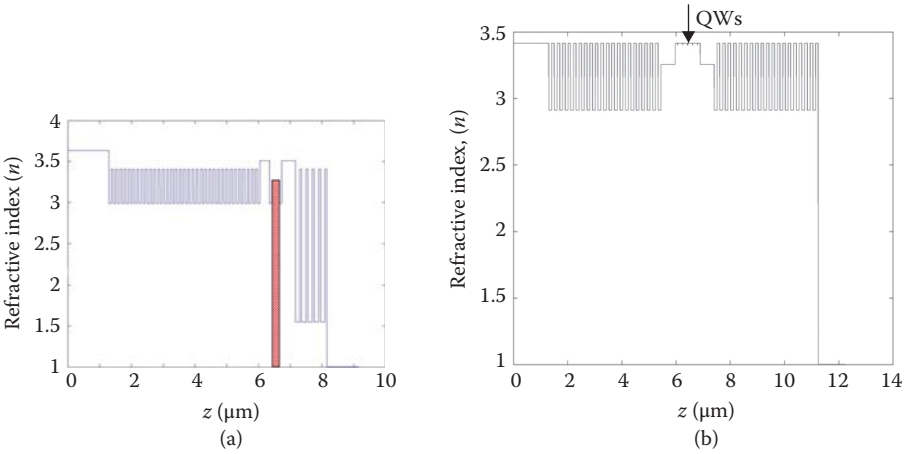


FIGURE 49.29 Refractive-index profile along a semiconductor microcavity (a) filled with gain medium (shaded area) with $n = 3.2736$ at $\lambda = 850$ nm; (b) with 6 GaInNAs-QWs, filled with gain medium of refractive index $n = 3.4$ at $\lambda = 1.29$ μm. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

We use a pseudorandom number generator based on the Box–Müller method for generating random deviates with a normal (Gaussian) distribution from uniformly distributed in the interval (0, 1) random numbers a and b. At each time step, j, white Gaussian noise with variance $\xi_E = \sigma^2 = 1 \times 10^{-3}\,\mathrm{V^2 m^{-2}}$ is implemented according to

$$E_x(z)_j = E_x(z)_j + \sqrt{-2\xi_E \ln(a)}\cos(2\pi b)\,. \tag{49.53}$$

White Gaussian noise has a mean zero and δ-correlated second moment of the random distribution:

$$\begin{aligned} &\langle \delta E_x(z,t)\rangle = 0 \\ &\left\langle \delta E_x(z,t)\,\delta E_x\left(z,t'\right)\right\rangle = \tilde{\xi}_E \delta\left(t-t'\right) = \xi_E R_{sp}\delta\left(\frac{t-t'}{\Delta t}\right), \end{aligned} \tag{49.54}$$

where an expression for the spontaneous emission rate has been derived:

$$R_{\mathrm{sp}} = \frac{\xi_E(\varepsilon_0\varepsilon)^2}{N_a^2 \wp^2 T_2^2}. \tag{49.55}$$

Consider the first microcavity: we assume that the resonant medium is initially inverted ($S_{30} = 1$), thus providing gain, and apply solely random white Gaussian noise at each point within the cavity according to Equation 49.53. The spatial distribution of the intracavity electric field as a function of time, and the time evolution of the electric field at the microcavity structure's output boundary are calculated. The parameter set used for this simulation is $\wp = 4.8 \times 10^{-28}\,\mathrm{C \cdot m}$, $n = 3.2736$, $N_a = 1 \times 10^{24}\,\mathrm{m^{-3}}$, initial field amplitude, $E_0 = 700\,\mathrm{V \cdot m^{-1}}$, and relaxation and dephasing times, $T_1 = 10$ ps and $T_2 = 70$ fs, respectively. The cavity field enhancement provided by the DBR mirrors leads to build-up of the electric field within the cavity borne out of the quantum noise. A snapshot of the intracavity electric-field build-up is shown in Figure 49.30, along with the corresponding Rabi flopping of the population dynamics. The amplitude that has arisen solely due to noise gradually builds up towards a standing-wave mode in the cavity.

After this initial amplitude build-up process, the lasing threshold is reached, and coherent oscillations appear at the output boundary. This is the onset of the lasing regime, shown for the unsaturated

$$\Omega_R^2 T_1 T_2 \approx 1.59 \tag{49.56}$$

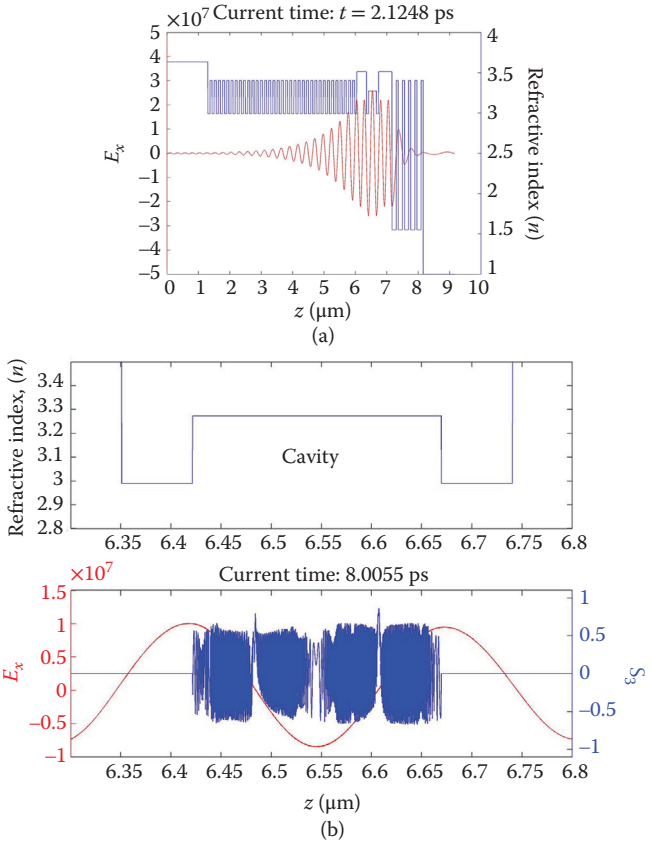


FIGURE 49.30 (a) Stationary electric-field standing-wave profile arisen from quantum noise in the cavity. (b) (Top) Refractive-index profile within the cavity and in its vicinity. (Bottom) A snapshot of the spatial distribution of the intracavity E_x electric field (left axis, slowly-varying sinusoidal curve) and the corresponding population inversion (right axis, fast-oscillating curve) exhibiting fast Rabi oscillations at $t = 8$ ps after the instant when the quantum noise arises. These oscillations result in the fast modulation of the electric-field envelope observed at the output boundary of the structure, which is shown in Figure 49.31a and b.

and saturated

$$\Omega_R^2 T_1 T_2 \approx 16241 \gg 1 \tag{49.57}$$

cases in Figure 49.31 (a,b), respectively. The charge-carrier population performs multiple Rabi flops, oscillating about a steady population inversion value close to zero. The net gain and absorption are therefore similar, resulting in a gain saturation in time.

Let us now consider the structure designed at $\lambda = 1.29\,\mu$m. The population in all six quantum wells is initially inverted into the upper state, $S_{30} = 1$, representing a nonlinear gain medium, while the barrier regions within the cavity are assumed to be in the ground state, $S_{30} = -1$. Random noise is generated within the cavity with a parameter set corresponding to the unsaturated case above (49.56). In Figure 49.32, a snapshot of the spatial distribution of the electric-field amplitude and the steady-state population inversion are shown. The population inversion relaxes to this value after a sufficiently large number of time steps. The time evolution of the electric field at the output boundary is shown in Figure 49.33, showing continuous build-up of coherent self-sustained oscillations and lasing. The lasing threshold is clearly discernible, and the electric-field envelope exhibits rapid damped relaxation oscillations with a decay rate,

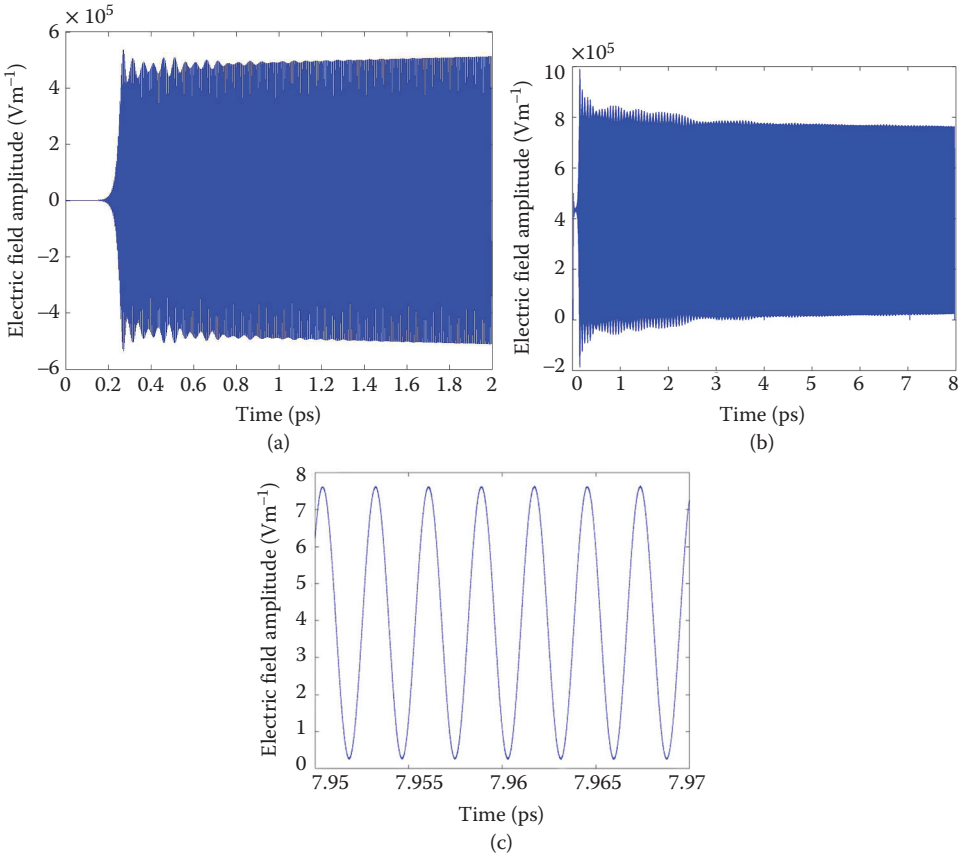


FIGURE 49.31 Time evolution of the electric field at the output boundary of a semiconductor microcavity, showing the noise-induced build-up of coherent oscillations and onset of lasing of the structure displayed in Figure 49.29a. Simulations were run for two different sets of relaxation and dephasing times (T_1 and T_2, respectively). (a) $T_1 = 10$ ps, $T_2 = 70$ fs; (b) $T_1 = 1$ ns, $T_2 = 10$ ps. (c) Expanded view of (b) in the steady-state region, showing single-mode oscillations and lasing at a frequency corresponding to $\lambda = 850$ nm. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

$\Gamma_R \approx 1.5 \times 10^{11}\ s^{-1}$. These rapidly varying oscillations are a new feature which is not predicted by the usual rotating-wave (RWA) and slowly-varying envelope (SVEA) approximations. The oscillations eventually reach the steady-state gain saturation value, $S_{3sat} = 0.0025$, at simulation times exceeding 16 ps.

We perform a Fourier transform of the output electric-field trace in Figure 49.33 and obtain the laser spectral linewidth, shown in Figure 49.34. The obtained linewidth exhibits superfine structure: the main mode is centered at a frequency ω_0 corresponding to $\lambda = 1.289\ \mu m$, and the wings exhibit satellite peaks at $\omega_0 + n\Omega_{rel}$. The frequency of the fast relaxation oscillations, $f_{rel} \approx 2.54$ THz, is determined from the figure.

We showed that an extension within the Langevin framework of our Maxwell-pseudospin model predicts the build-up of coherent, self-sustained oscillations and the onset of lasing, which is a sole consequence of noise being introduced into the cavity. This emphasizes the importance of quantum fluctuations for triggering the lasing regime of cavity operation. The simulations allow to investigate the lasing threshold behavior directly in the time domain and provide an estimate for the coherence time of the laser emission and the laser linewidth. We showed that several important parameters can be extracted from the simulations and that they can be optimized using numerical experiments on a variety of design geometries. Most importantly, the developed quantum stochastic approach is capable of modeling quantum effects, such as

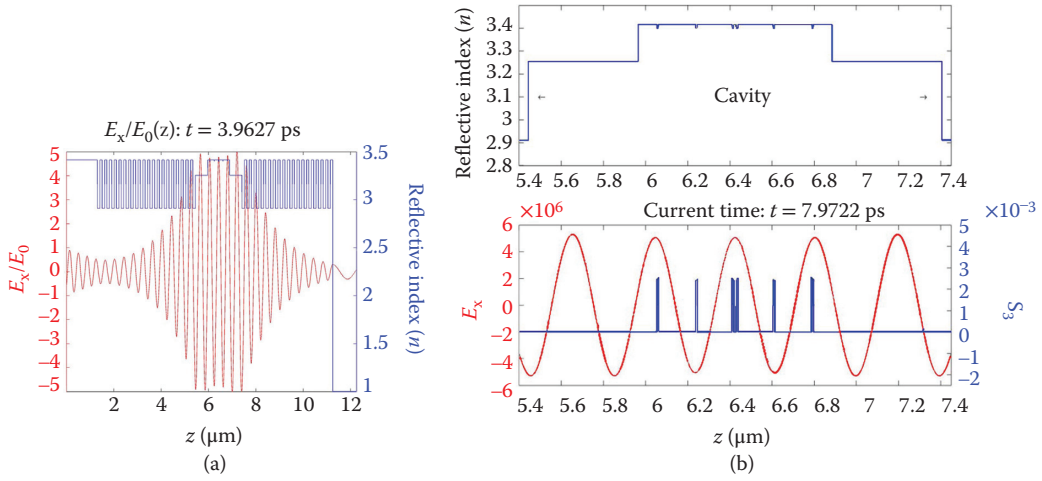


FIGURE 49.32 (a) Instantaneous profile of the electric-field standing wave borne out of quantum noise in the cavity. (b) (Top) Refractive-index profile implying the cavity boundaries and the location of the quantum wells. (Bottom) Spatial distribution of the stationary electric-field standing wave (left axis, sinusoidal curve) and the corresponding saturated population inversion, $S_{3sat} = 0.0025$, in the quantum wells within the cavity (right axis, spiky curve). (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

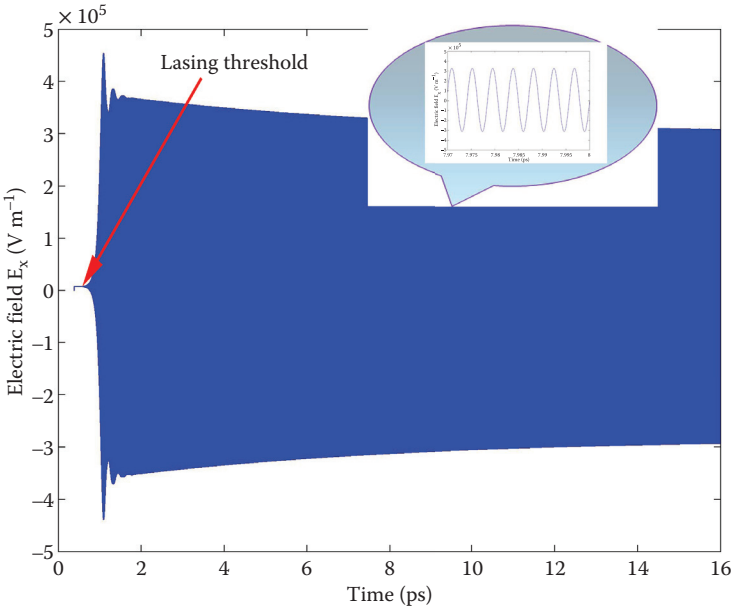


FIGURE 49.33 Build-up in time of coherent, self-sustained oscillations arisen from noise introduced into the cavity of the structure displayed in Figure 49.29b. The lasing threshold is apparent. The decay rate of the electric-field envelope is $\Gamma_R \approx 1.5 \times 10^{11}$ s^{-1}. Inset: expanded view of the fast carrier-frequency oscillations with $\lambda = 1.289$ μm; the steady-state region occurs for $t > 16$ ps. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

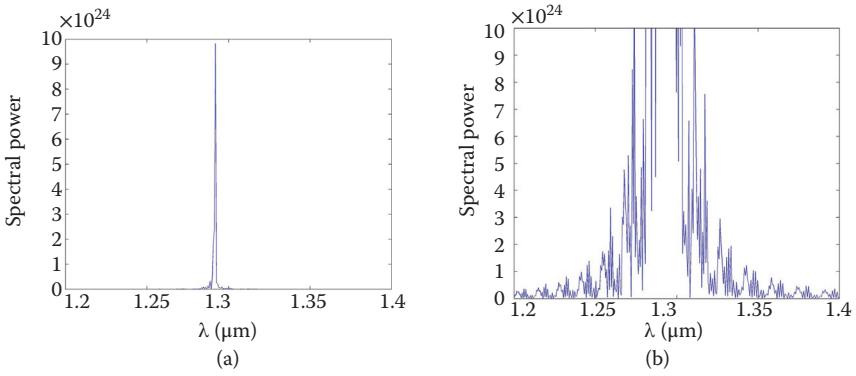


FIGURE 49.34 (a) Laser spectral line obtained by Fourier-transforming the output trace of the electric-field amplitude centered at the lasing frequency, corresponding to $\lambda = 1.289\,\mu m$. (b) Magnified view of the spectrum in (a) showing relaxation oscillation sidebands. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

spontaneous emission, and thus can be used to model the system dynamics in the few-photon limit as well as demonstrate and predict quantum-optical effects.

49.7 Modeling Optical Rotation in Chiral Carbon Nanotubes with Master Maxwell-Pseudospin Equations

In this section, we shall employ our vector Maxwell-pseudospin model [44] to investigate how chirality affects the ultrafast nonlinear optical and magneto-optical response of a single chiral CNT [80]. We shall show that a simple discrete-level model of the optically active states near the bandgap edge of a chiral nanotube is sufficient to explain the circular dichroism and birefringence, as well as to quantify the optical rotation per unit length (also known as specific rotatory power), either in the absence or in the presence of an external magnetic field permeating the tube. The model predicts a giant natural gyrotropy for the specific chirality considered in this case, which is comparable with or exceeding that of artificially fabricated helical photonic structures. This is remarkable because the latter are considered to achieve the largest optical rotation of all known materials: orders of magnitude larger than that of crystal birefringent materials or of liquid crystals. Finally, we will give a quantitative estimate of the nonlinear coherent magneto-chiral optical effect in an axial magnetic field.

Chirality is one of the main symmetries of the CNT geometry that determines the optical properties of single-walled CNTs (SWCNTs). SWCNTs are uniquely determined by the chiral vector, $C_h = na_1 + ma_2 \equiv (n, m)$, with $0 < |m| < n$, or equivalently by a pair of integer numbers (n, m) in the planar graphene hexagonal lattice unit-vector basis (Figure 49.35). Two helical forms (enantiomers), e.g., AL (5, 4) and AR (4, 5), can exist. Topologically, the CNT can be viewed as a graphene sheet rolled up along the chiral vector into a cylinder. Chirality serves as a primary classification criterion. Achiral nanotubes, whose mirror image is superimposable, are subdivided into two classes: zigzag ($m = 0$) and armchair ($m = n$) nanotubes. The rest of the nanotubes, whose mirror reflection is not superimposable, are subsumed into the most general class of chiral nanotubes.

There is a fundamental relationship between chirality and optical activity. The interaction of polarized light with chiral materials gives rise to the phenomenon of optical rotation (or optical activity), whereby the polarization plane is rotated continuously during the propagation of the light through the nanotube. In the presence of an external magnetic field, magnetically induced optical activity—also known as Faraday effect—takes place. Both effects manifest themselves as a rotation of the transmitted light; however, the origin of the two effects is fundamentally different. While the natural optical activity is a result of

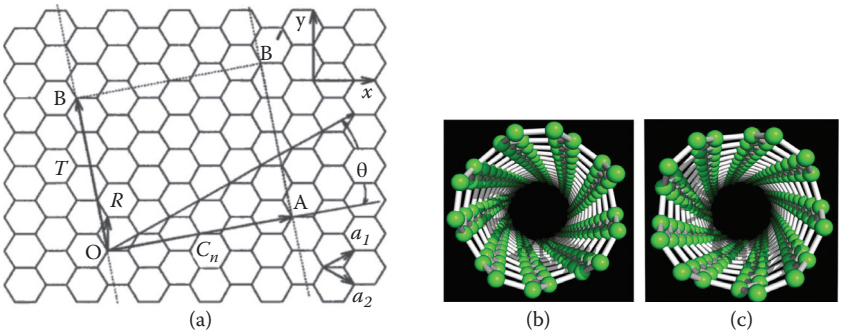


FIGURE 49.35 (a) Chiral vector, C_h (C_n on schematic) and chiral angle, $\theta = \angle\left(a_1, C_h\right)$, $0° < |\theta| < 30°$, where a_1 and a_2 are the elementary Bravais cell unit vectors. (b) Left-handed AL (5, 4) (with $m > n$; m, n – integers, see text) and (c) right-handed AR (4, 5) (with $m < n$) single-walled CNT molecular structure calculated by the tight-banding method [81]. Shown is a view along the tube axis looking in the negative z-direction. Chirality is determined by the rotation of the so-called armchair chains of carbon atoms either clockwise or counterclockwise when looking in the negative z-direction.

the nonlocal optical response of a medium lacking mirror symmetry, the magnetic optical activity results from time-reversal symmetry breaking by the magnetic field. The two phenomena are linked through the magneto-chiral optical effect which takes place when both symmetries are broken simultaneously [82].

The electronic band structure of an SWCNT is described by the quantization of the wavevector along the tube circumference perpendicular to the tube axis. This results in a discrete spectrum of allowed k-vector states forming subbands in the valence and conduction bands labeled by the quasi-angular momentum quantum number, μ [83]: $k_\perp = \frac{2\pi\mu}{L}$, $\mu = 0, \pm1, \pm2, ...$, where $L = |C_n|$ is the tube circumference (Figure 49.36a). We should note that due to symmetry in the graphene hexagonal unit cell, there are two equivalent points, K (with $\mu > 0$) and K′ (with $\mu < 0$) in the BZ (Figure 49.36c). The energy dispersion of a CNT can be visualized as intersections of the graphene conical dispersion at the quantized transverse wavevector, $k_\perp$ (Figure 49.36b).

The most widely studied excitation geometry in experiments on individual CNT is the one of linearly polarized excitation perpendicular to the tube axis, shown in Figure 49.37a. In this excitation geometry, the longitudinal E_z electric-field component is dominant, because the transverse E_x component is suppressed by depolarization effects, and linearly polarized intersubband optical transitions from the valence μ subband to the conduction μ subband with $\Delta m = 0$ can be excited (here m signifies the eigenvalue of the quasi-angular momentum operator projection, $\hat{J}_z$). We shall be interested, however, in the unconventional circularly polarized excitation geometry shown in Figure 49.37b, whereby a circularly x–y plane-polarized laser pulse propagates along the tube axis, exciting only one of the two allowed transitions, $\mu \to \mu \pm 1$, where "+/−" corresponds to right and left circularly polarized light, respectively (Figure 49.38b). The molecular structure and the allowed dipole optical intersubband transitions of a (20, 10) chiral CNT are shown in Figure 49.38.

We shall consider aggregates or bundles of aligned SWCNTs grown by chemical-vapor deposition in an electric field [84], or aligned in a polymer matrix [85]. We model the single carbon-nanotube bandgap-edge structure at the K point of the BZ (where the fundamental bandgap opens) by a large ensemble of identical four-level systems, corresponding to the allowed transitions of the two valence subbands closest to the Fermi level for the AL and AR nanotube enantiomers shown in Figure 49.39. We note that the energy-level diagrams for the left- and right-handed helical forms in Figure 49.39 are nonsuperimposable: absorption of σ^+ light excites the $\mu \to \mu + 1$ transition in AL-handed SWCNT, while absorption of light with the same helicity excites the $\mu \to \mu - 1$ transition in AR-handed SWCNT. The difference between the optical selection rules for left and right circularly polarized light gives rise to the optical activity [86].

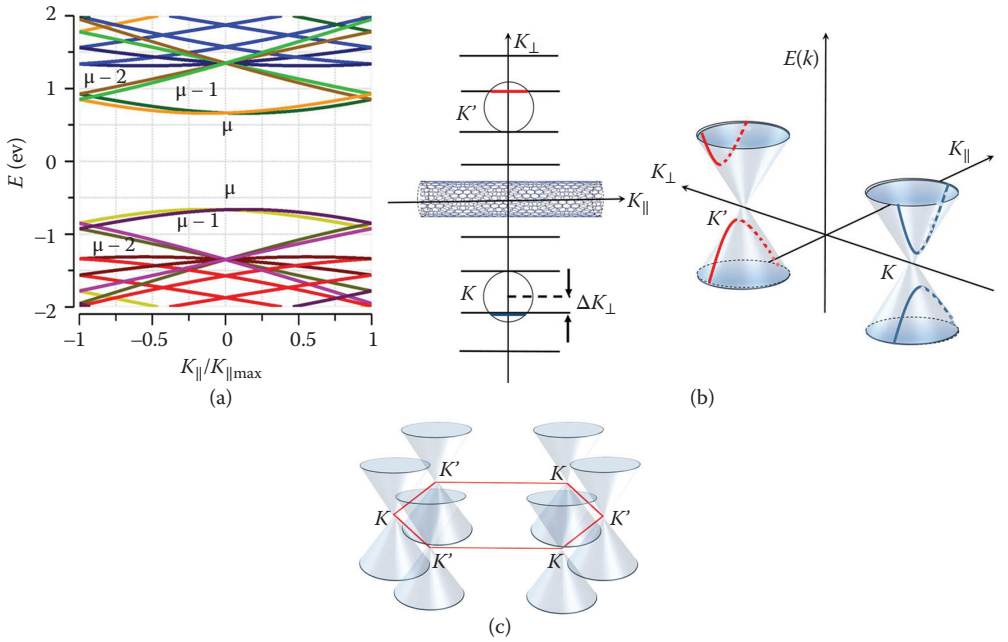


FIGURE 49.36 (a) Energy dispersion as a function of the longitudinal wavevector, $k_{||}$, normalized with respect to the wavevector at the boundary of the first BZ of a (5, 4) SWCNT. $k_{||}$ points along the tube axis. The tube for this chirality is semiconducting—i.e., a bandgap is opened at the center (Γ point) of the BZ. (b) Parabolic cross sections of the graphene linear dispersion, corresponding to points K and K′ in (a) representing the nanotube energy dispersion curves. (c) Unfolded 2D first BZ of graphene, showing the equivalent points K and K′ in the band structure.

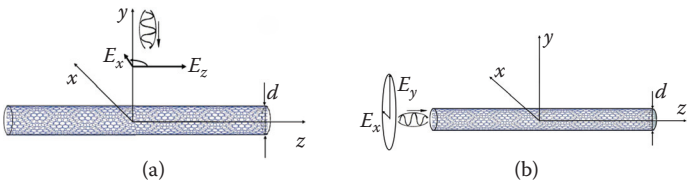


FIGURE 49.37 (a) Linearly polarized excitation perpendicular to the tube axis exciting intersubband optical transitions with $\Delta m = 0$. (b) Circularly polarized excitation along the tube axis exciting $\Delta m = \pm 1$ optical transitions.

49.7.1 Natural Optical Activity

In this section, we compute the natural optical activity rotation angle (rotatory power per unit length) in the absence of an external magnetic field and demonstrate the possibility to manipulate it by engineering the nanotube chirality. We consider the specific case of a chiral, left-handed AL (5, 4) SWCNT, which is chosen for illustration of the general method valid for an arbitrary chirality. The nanotube length is 500 nm and all edge effects will be ignored in our 1D model. We performed tight-binding calculations of the electronic band structure (Figure 49.36a) and determined the fundamental bandgap, $E_g = E_{\mu,\mu} = 1.321$ eV, corresponding to a pumping wavelength of $\lambda = 939$ nm. The tube diameter is 0.611 nm, and the chiral angle, 26.33°. The resonant transition energy for circularly polarized excitations is $E_{\mu,\mu\pm1} = 1.982$ eV, corresponding to a resonant wavelength, $\lambda_0 = 626.5$ nm.

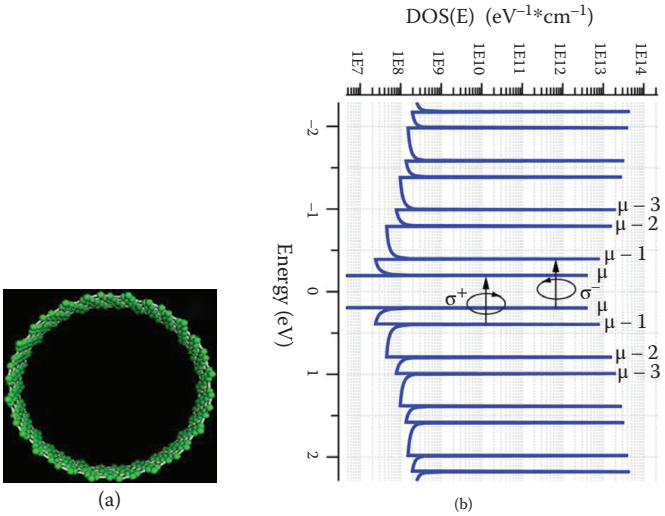


FIGURE 49.38 (a) Molecular structure of an AL (20, 10) left-handed chiral CNT. (b) 1D electron density of states (DOS) plotted as a function of energy at the K point of the BZ ($\mu > 0$). The allowed dipole optical transitions for circularly polarized light are denoted by arrows for left- (σ^-) and right-handed (σ^+) pulse helicity.

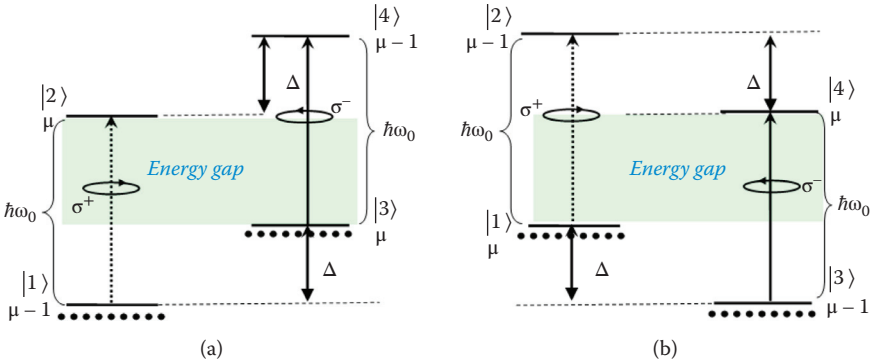


FIGURE 49.39 Energy-level structure at the K (or K') point of the lowest subbands labeled by the subband index μ for an (a) AL (le-handed) and (b) AR (right-handed) SWCNT. The fundamental energy gap is shaded. The dipole optical transitions excited by σ^- and σ^+ circularly polarized light are designated by arrows. Only one of the two transitions is allowed for circularly polarized light, denoted by solid arrow (by contrast, the forbidden transition is indicated by a dashed arrow). Valence-band states below the bandgap are populated. (ω_0, resonant transition frequency; Δ, energy separation between the lowest subband and the second lowest subband near the bandgap.)

We consider a circularly polarized ultrashort pulse of duration 60 fs matching experiments [87], and a pulse area of π, giving an initial electric-field amplitude $E_0 = 6.098 \times 10^8\ \mathrm{V \cdot m^{-1}}$. We inject the pulse from the left boundary of the simulation domain, consisting of the tube embedded between two free-space regions of thickness 50 nm each. The pulse is resonant with the energy $E_{\mu,\mu\pm1}$ of the allowed dipole optical transition and is coupled to an ensemble of identical, HB four-level systems (Figure 49.39a for a left-handed AL SWCNT) used to describe the optical transitions involving the lowest-energy subbands lying nearest to the bandgap edge (Figure 49.40a).

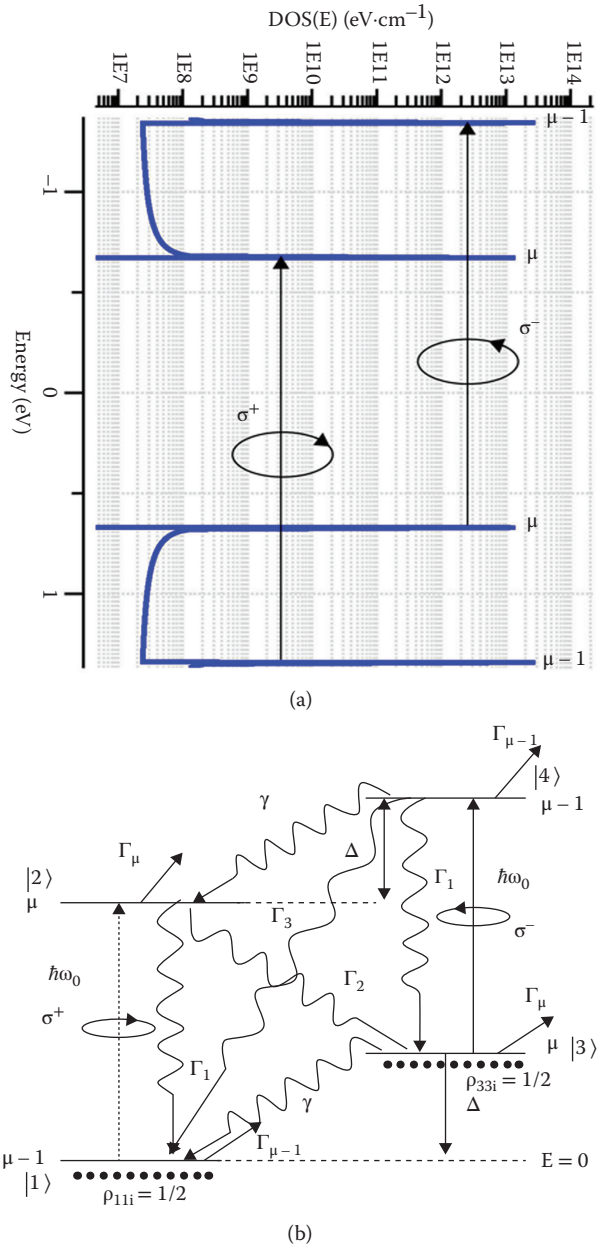


FIGURE 49.40 (a) Plot of the 1D density of states (DOS) showing the lowest-energy subbands near the bandgap edge involved in dipole optical transitions excited by circularly polarized light. Only one transition at a time can be excited by each helicity (σ^- or σ^+). (b) Energy-level diagram in the general case of a resonant optical excitation of $\mu - 1 \rightarrow \mu$ and $\mu \rightarrow \mu - 1$ interband transitions by a σ^+ and σ^- polarized pulse, respectively. The resonant transition energy is $\hbar\omega_0$. The initial population of the lowest valence states below the bandgap, ρ_{11i} and ρ_{33i}, is equally distributed between levels $|1\rangle$ and $|3\rangle$ (here, the subscript "i" denotes "initial"). (Δ, the energy separation between the first and the second lowest conduction (or valence) band states; wavy lines denote longitudinal relaxation processes between the levels, associated with population transfer; Γ_1, spontaneous emission (radiative decay) rate of $|2\rangle \rightarrow |1\rangle$ transitions, assumed to be equal to that of $|4\rangle \rightarrow |3\rangle$; Γ_2, spontaneous emission rate for the linearly polarized transition from the conduction μ subband to the valence μ subband; Γ_3, decay rate of the linearly polarized transition from the conduction $\mu - 1$ subband to the valence $\mu - 1$ subband; γ, intraband relaxation rate; Γ_μ and $\Gamma_{\mu-1}$, transverse relaxation (dephasing) rates.)

The corresponding energy-level diagram is shown in Figure 49.40b, with all coherent and spontaneous transitions indicated. Following our coherent vector Maxwell-pseudospin formalism, the system Hamiltonian for a four-level system that applies for either excitation by σ^- or σ^+ is given by

$$\hat{H} = \hbar \begin{pmatrix} 0 & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & 0 & 0 \\ -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & \omega_0 & 0 & 0 \\ 0 & 0 & \Delta & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) \\ 0 & 0 & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \Delta + \omega_0 \end{pmatrix}, \quad (49.58)$$

where $\Omega_x = \frac{\wp}{\hbar}E_x$ and $\Omega_y = \frac{\wp}{\hbar}E_y$ are the time-dependent Rabi frequencies associated with the E_x and E_y electric-field components, and $\wp$ is the optical dipole matrix element for $\mu \rightarrow \mu \pm 1$ transitions excited by circularly polarized light. This parameter is largely unknown. We have provided an estimate for it,

$$\wp \approx 3 \times 10^{-29}\ \mathrm{C \cdot m}, \quad (49.59)$$

based on an extension of the effective-mass method applied to chirality effects in carbon nanotubes (CNTs) [88], in agreement with the measured radiative lifetime of typical molecular transition in a single CNT [89], which is on the order of $\tau_{\mathrm{spont}} \approx 10$ ns.

We consider a four-level resonant medium consisting of a large number of identical aligned CNTs with an average density $N_a = 6.8 \times 10^{24}\ \mathrm{m}^{-3}$, calculated in such a way that the volume of the simulated nanotube with diameter 0.611 nm and length 500 nm contains just a single nanotube on average. A method called the effective medium approximation—used for describing the macroscopic properties of composite materials, such as the dielectric permittivity of a single CNT—has been successfully applied in FDTD modeling of the thermal radiative properties of vertical arrays of multi-walled CNTs [90]. We use two independent theoretical approaches to obtain an estimate for the effective dielectric constant of an isolated SWCNT: the aforementioned effective medium approximation [80] and tight-binding model calculations of the axial component of the imaginary part of the dielectric function tensor [91,92]. This gives a value for the refractive index, $n \approx 2.3$ along the nanotube axis. With dipole matrix element from (49.59), we use the expression for the spontaneous emission rate: $\tau_{\mathrm{spont}} = \frac{3\pi\varepsilon_0\hbar c^3}{n\omega_0^3\wp^2}$, taking into account the energy separation of each transition, to obtain an estimate for the relaxation rates: $\Gamma_1 = 2.907\ \mathrm{ns}^{-1}$, $\Gamma_2 = 9.812\ \mathrm{ns}^{-1}$, and $\Gamma_3 = 1.227\ \mathrm{ns}^{-1}$. We take the experimental value obtained in [87] for the intraband optical transitions, $\gamma = 130\ \mathrm{fs}^{-1}$. Since the dephasing rates for the states involved are largely unknown, we treat them as phenomenological parameters adopting the following values: $\Gamma_\mu = 800\ \mathrm{fs}^{-1}$ and $\Gamma_{\mu-1} = 1.6\ \mathrm{ps}^{-1}$; however, our simulations show that the dynamics is largely insensitive to the choice of these dephasing rates.

We consider separately the cases of an ultrafast resonant optical excitation of the $|1\rangle \rightarrow |2\rangle$ transition with σ^+ helicity, and of the $|3\rangle \rightarrow |4\rangle$ transition with σ^- helicity (Figure 49.40b). Note that in the former case, level $|4\rangle$ does not participate in the relaxation dynamics and the system is effectively a three-level Λ-system, rather than a four-level system. The initial population is assumed to be equally distributed between the lower-lying levels. The ultrashort circularly polarized source pulse is injected into the medium, and the temporal dynamics of the electric-field vector components and population of all four levels are sampled at four different locations along the nanotube z-axis. The spatially resolved temporal dynamics at a point $z = 300$ nm from the left boundary of the simulated structure for σ^+ and σ^- excitations is shown in Figure 49.41.

It would be interesting from an experimental point of view to test whether nanotube chirality could be determined from the ultrafast nonlinear response using ultrashort pulses with either helicity. We compare the Fourier spectra at the output boundary for each helicity of the optical pulse excitation in Figure 49.42. Both transmission spectra exhibit a sharp peak at the resonant wavelength, indicating

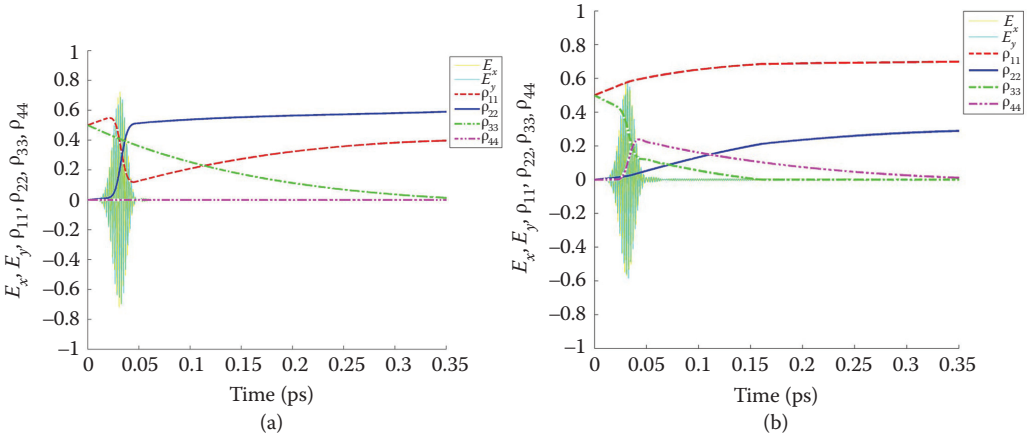


FIGURE 49.41 Time evolution of the electric-field components E_x and E_y, and the populations of all four energy levels of a SWCNT (a) for a circularly polarized σ^+ excitation pulse (note that level $|4\rangle$ is not involved in the dynamics) and (b) for a circularly polarized σ^- excitation pulse. The location is $z = 300$ nm. The left boundary of the SWCNT is at $z = 0$. The SWCNT is 500 nm long and is placed between two free-space regions, each 50 nm long.

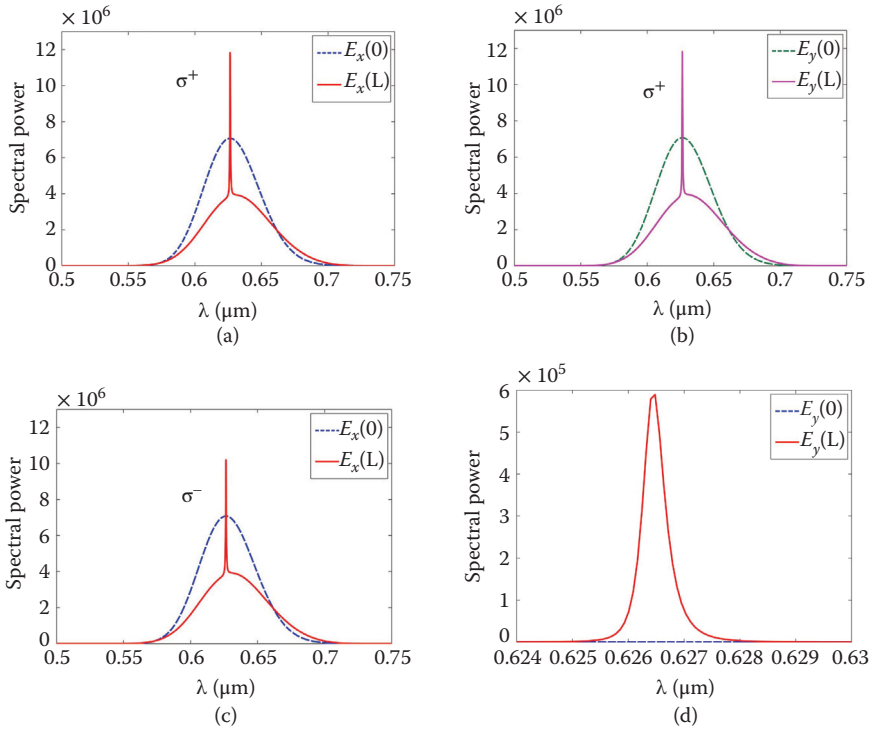


FIGURE 49.42 Transmission Fourier spectra plotted against the wavelength of the E_x and E_y electric-field components at the input, $E_{x,y}(0)$, and output, $E_{x,y}(L)$, of the structure for a (a), (b) σ^+ polarized pulse, and a (c), (d) σ^- polarized pulse. Note that the peak output amplitudes are approximately 20% higher in (a) and (b) compared to in (c) and (d). The simulated structure is as in Figure 49.41.

resonant amplification; however, the peak intensity corresponding to σ^+ excitation is nearly 20% higher. This difference can be exploited in an experiment aiming to determine unambiguously the chirality of a single CNT by first injecting a pulse of one helicity, then the other, and observing the consequences at the output. We show that the difference in the polarization-resolved transmission spectra of a linearly polarized pulse is much more pronounced and allows identification of the precise helical form of the SWCNT.

In order to demonstrate rotation of the electric-field polarization plane during the pulse propagation across the resonant four-level medium, we launch a linearly polarized pulse along x, expressed by:

$$X \begin{cases} E_x(z=0,t) = E_0 e^{-(t-t_0)^2/t_d^2} \cos(\omega_0 t) \\ E_y(z=0,t) = 0. \end{cases} \tag{49.60}$$

The system's temporal dynamics induced by the ultrashort linearly polarized pulse is shown in Figure 49.43a and b on an expanded scale, where the appearance of a second E_y component, and therefore also the appearance of optical rotation of the electric-field vector polarization, is clearly visible. The maximum amplitude of the E_y component continuously increases as the pulse propagates along the nanotube structure. We should therefore expect the maximum optical rotation angle to occur at the output boundary.

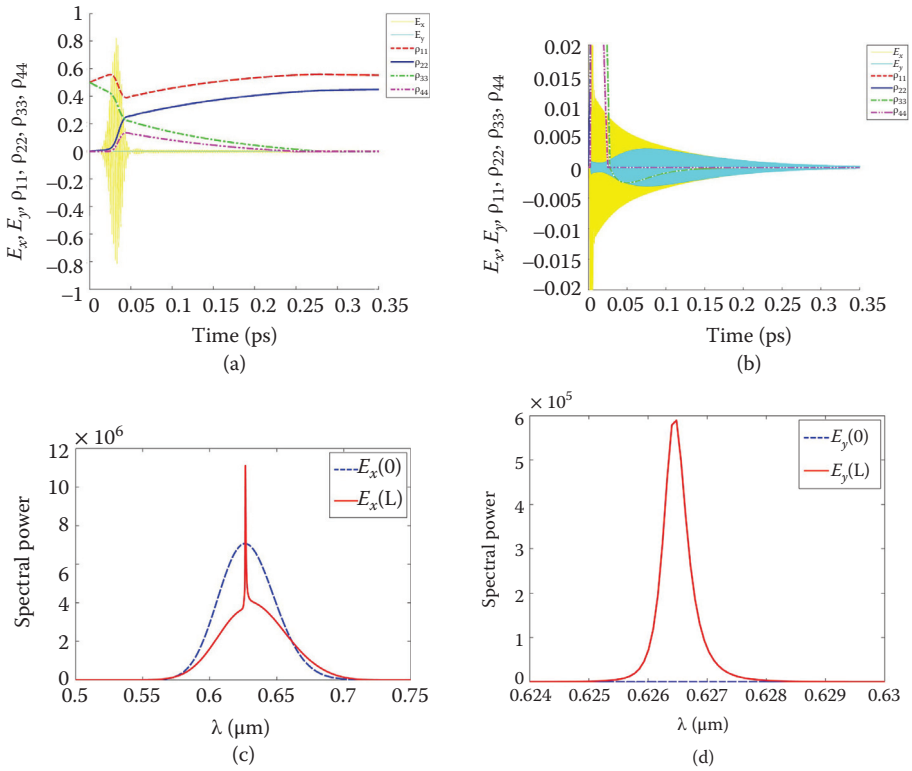


FIGURE 49.43 (a) Time evolution of a linearly polarized pulse with initial value $E_y = 0$, and the populations of all four levels at $z = 550$ nm from the left boundary. The simulated structure is as in Figure 49.41. (b) Magnified view of (a) showing the build-up of the E_y component with time, which is equivalent to a rotation of the electric-field polarization. The transmitted pulse is elliptically polarized and displays different power spectra for (c) the E_x component (sharp resonant peak superimposed on a broader line) and (d) the E_y component (single spectral line).

The polarization-resolved transmission spectra of the E_x and E_y components are quite distinct: while the former is similar to the spectrum obtained for circularly polarized optical excitation, the E_y spectrum exhibits a single spectral line. Conversely, for excitation with light of the opposite helicity (data not shown), the E_x and E_y profiles are interchanged. This difference in the polarization-resolved nonlinear optical response can therefore be used for unambiguous determination of the nanotube chirality.

In order to obtain a quantitative estimate of the natural optical activity in a single chiral CNT, we follow the general methodology developed in Section 49.3.3. We calculate the gain/absorption coefficient from the complex propagation factor, $e^{ik_c(z_2-z_1)}$, for the E_x and E_y electric-field components of a circularly polarized pulse over a distance of one dielectric wavelength, $l = z_2 - z_1 = \lambda_0/n$, where λ_0 is the resonant wavelength. The wavevector is $k_c = \beta + i\gamma$, where β is the phase shift per unit length induced in the optical pulse by the interaction with the resonant medium, and γ is the gain/absorption coefficient. The gain/absorption coefficient allows us to calculate the magnitude of the circular dichroism, whereas the phase shift represents a measure of the rotation angle. A comparison between the spatially resolved gain coefficients for σ^+ and σ^- pulse excitation is shown in Figure 49.44a; the analytically computed gain coefficient of a HB two-level system has also been plotted on the same graph, for reference. The maxima of the gain coefficients for both σ^+ and σ^- occur at the resonant wavelength. The circular dichroism, $\Delta A = G_R - G_L$, is therefore calculated as the average of the difference between the maximal gain coefficients for σ^+ and σ^- excitation (R and L subscripts, respectively) computed at several points along the nanotube. This yields a value for the circular dichroism of 0.083 μm^{-1}. By comparison, the absolute value of the circular dichroism of an artificial helicoidal bilayered structure varies in the range of 5–9 dB, which is equivalent to a linear amplitude gain/absorption coefficient of 0.58–1.04 μm^{-1}. The atomic phase shift ($\beta_{R(L),x(y)}l$) and the average rotation angle are plotted in Figure 49.44b. The specific rotatory power per unit length is calculated using $\rho = \pi(n_L - n_R)/\lambda_0$, where n_L and n_R are the refractive indices for left and right circularly polarized light. Computing this value gives ≈2962.24°/mm.

Although the circular dichroism is low compared to the artificial chiral photonic structures, the optical rotation, as computed from the phase shift, is enormous compared to other birefringent materials. The

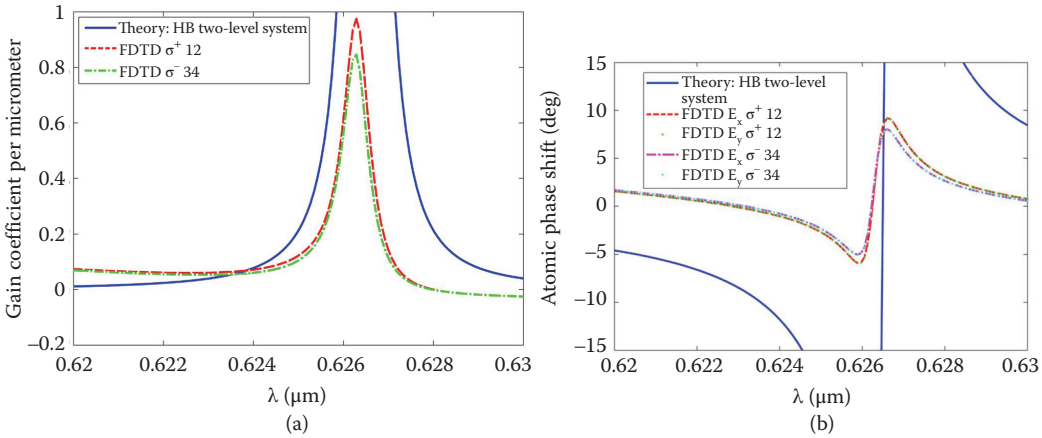


FIGURE 49.44 (a) Plot against wavelength of the spatially resolved gain coefficient per micron for a σ^+ (dashed curve) and σ^- (dash-dotted curve) circularly polarized ultrashort optical pulse, and the theoretical gain coefficient of a homogeneously broadened (HB) two-level system (solid curve) (Equation 49.30). The simulated structure is as in Figure 49.41. (b) Spectrum of the phase shift, averaged over the entire length of the structure, of the E_x (dashed curve) and E_y (dotted curve) electric-field components of σ^+ excitation, E_x (dash-dotted curve) and E_y (dotted curve) of σ^- optical excitation, and theoretical calculations of this phase shift (Equation 49.31), assuming a HB, resonant two-level system.

specific rotatory power of crystals varies widely from 2.24°/mm for $NaBrO_3$ to 522°/mm for $AgGaS_2$. Liquid substances exhibit much lower values of specific rotatory power—e.g., $\rho = -0.37°/\text{mm}$ for turpentine ($T = 10°$, $\lambda = 589.3\,\text{nm}$); $\rho = 1.18°/\text{mm}$ for corn syrup, etc. Cholesteric liquid crystals and sculptured thin films exhibit large rotatory power in the visible spectrum: ~1000°/mm and ~6000°/mm, respectively, which is comparable with the calculated rotatory power. We should note that the dependence of the optical rotation angle on chirality opens up the exciting possibility to engineer this angle through manipulating the chirality of the structure.

49.7.2 Faraday Rotation and Magneto-Chiral Effects

In this section, we shall develop a theoretical model of the resonant, coherent nonlinear optical activity when a static magnetic field, $B_{\|}$, threads the nanotube (Figure 49.45).

We shall be interested in the Faraday effect, or the rotation of the polarization of a plane-polarized electromagnetic wave propagating in a medium permeated by a static magnetic field, **B**, oriented along the direction of propagation. In the presence of an axial magnetic field, the electronic band structure of a single CNT—and the electronic states near the bandgap edge in particular—change significantly, owing to the combined action of two effects: the spin-B interaction, resulting in Zeeman splitting of the energy levels [93–95], and the appearance of the Aharonov–Bohm phase in the wavefunction [83,96,97]. The two symmetric subbands at the K (or K') point of the BZ are degenerate at $B_{\|} = 0$ (Figure 49.45b). An applied magnetic field along the nanotube axis lifts this degeneracy by shifting the energy levels. As a result, the bandgap of one of the subbands (K') becomes larger while the bandgap of the other subband (K) becomes smaller [94] (Figure 49.45c).

Without loss of generality, we shall consider the electronic states near the bandgap at the K point in the BZ. Therefore, the overall effect will be a bandgap reduction. We consider and calculate separately the contributions to the bandgap from the Zeeman splitting and the Aharonov–Bohm effect.

The effect of the Aharonov–Bohm flux on the bandgap is to induce oscillations in its energy. The energy bandgap thus varies between zero and a fixed value with a period of the flux quantum, $\Phi_0 = h/e$ [83,93],

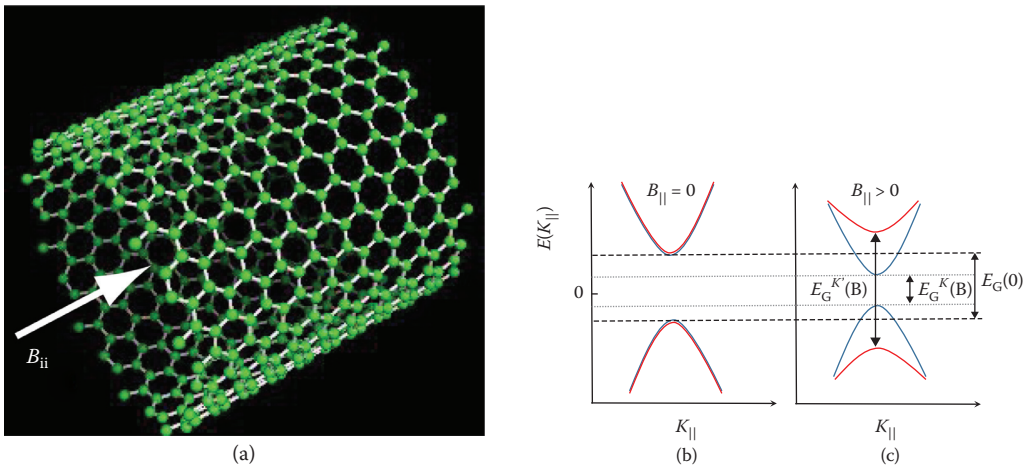


FIGURE 49.45 (a) Single (20, 10) chiral nanotube threaded by an axial magnetic field (a so-called Faraday configuration). (b) Magnetic energy bands in the absence of a magnetic field, $B_{||}$; the energy subbands at the point K (K') are degenerate. (c) When an axial magnetic field is applied, the degeneracy is lifted and the energy bandgap of the subband at K' becomes larger, while that of the other subband at K diminishes.

resulting in periodic oscillations of the magneto-optical absorption spectra given by

$$E_G(\Phi) = \begin{cases} 3E_G(0)\left|\frac{1}{3} - \frac{\Phi}{\Phi_0}\right|, & 0 \le \Phi/\Phi_0 \le 1/2 \\ 3E_G(0)\left|\frac{2}{3} - \frac{\Phi}{\Phi_0}\right|, & 1/2 \le \Phi/\Phi_0 \le 1, \end{cases} \tag{49.61}$$

where $E_G(0) = \hbar\omega_0$ is the energy bandgap at zero magnetic field and ω_0 is the resonant transition frequency.

For a magnetic field $B = 8$ T with a flux Φ threading a (5, 4) nanotube with a diameter 0.61145 nm, the ratio $\Phi/\Phi_0 = 0.00057$, and therefore the first of the equations (49.61) above holds. Due to the orbital Aharonov–Bohm effect, this leads to an energy-level shift, or bandgap renormalization of $E_{AB} = 3.37$ meV, corresponding to a resonant angular frequency, $\omega_{AB} = 5.12 \times 10^{12}$ rad/s. At a fixed value of the static magnetic field, the orbital Aharonov–Bohm effect leads to a uniform shift in the energy levels, so the resonant transition frequency ω_0 is replaced by $\omega_0 - \omega_{AB}$.

In an external magnetic field, the energy levels near the bandgap of Figure 49.40b split and the spin degeneracy is lifted. The resulting energy-level system can be split into two reduced systems of levels, each of which represents a mirror image of the other. The symmetry is broken only by the fact that the allowed optical transitions are different in each case (see Figure 49.46).

The Zeeman splitting, or the spin-B interaction energy, is given by

$$\mathrm{E}_z = \mu_B g_e \sigma B_{\|}\,, \tag{49.62}$$

where $\mu_B = \frac{e\hbar}{2m_e}$ is the Bohr magneton; the electron g-factor, g_e, is taken to be the same as that of pure graphite (≈ 2), $\sigma = \pm 1/2$ is the z-axis projection of the electron spin (spin-up/spin-down state), and m_e is the free electron mass.

We have performed simulations with magnetic fields in the range 2–100 T and found for the considered (5, 4) tube a nonmonotonous dependence of the optical rotation angle on the magnitude of **B**, with maximum rotation reached for magnetic fields in the range 8–10 T. We therefore selected $B_{\|} = 8$ T to demonstrate the maximum angle of Faraday rotation. Another reason for choosing this value is that it was the maximum magnetic field achievable by the static magnet available in the laboratory of our collaborators at the time when we performed these calculations. For $B_{\|} = 8$ T, the Equation (49.62) gives an energy shift of $E_z \approx 0.46$ meV, which corresponds to a Zeeman resonant frequency, $\omega_z = 7.026 \times 10^{11}$ rad/s.

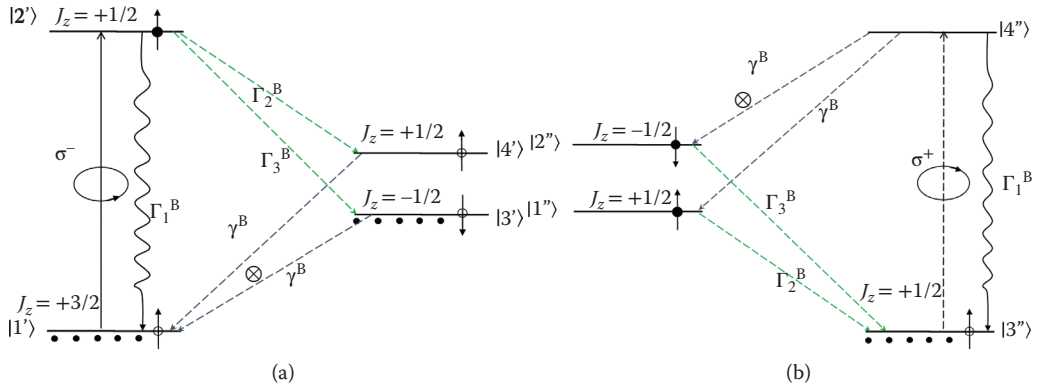


FIGURE 49.46 A four-level system at $B \neq 0$ splits into two nonsuperimposable reduced systems, involving levels energetically closest to the bandgap edge. (a) Left-handed reduced energy-level system excitable by a σ^- optical pulse. (b) Right-handed reduced energy-level system excitable by a σ^+ optical pulse. (Coherent and spontaneous transitions indicated as in Figure 49.40b. Forbidden dipole optical transitions are denoted by $\otimes$. In fact, there is a X-circle over the dashed lines in Figure 49.46, denoting forbidden transitions between levels).

The system Hamiltonians corresponding to the energy-level schemes displayed in Figure 49.46 are given by

$$\hat{H}' = \hbar \begin{pmatrix} \omega_z & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & 0 & 0 \\ -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \omega_0 - \omega_z & 0 & 0 \\ 0 & 0 & \Delta - \omega_z & 0 \\ 0 & 0 & 0 & \Delta + \omega_z \end{pmatrix} \tag{49.63}$$

and

$$\hat{H}'' = \hbar \begin{pmatrix} \omega_0 - \omega_z & 0 & 0 & 0 \\ 0 & \omega_0 + \omega_z & 0 & 0 \\ 0 & 0 & \Delta + \omega_z & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) \\ 0 & 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & \Delta + \omega_0 - \omega_z \end{pmatrix}, \tag{49.64}$$

respectively.

We employ our vector Maxwell-pseudospin master equations model for the four-level systems above. The simulation structure is the same as the one described for the $B = 0$ case. Note that due to the presence of a magnetic field, the energy levels $|1\rangle$, $|2\rangle$, $|3\rangle$ and $|4\rangle$ are split into two sets, which we choose to distinguish by adding a $'$ and $''$ to the notation; the first set (denoted by $'$) is accessed only using σ^- light, whereas the second (denoted by $''$) is accessible only via σ^+ light, due to the optical selection rules. We then resonantly excite the $|1'\rangle \rightarrow |2'\rangle$ and $|3''\rangle \rightarrow |4''\rangle$ transitions using, respectively, a left (σ^-) and right (σ^+) circularly polarized π-pulse of duration 60 fs. The pulse's central frequency is tuned in resonance with the transition frequency, $\omega_0 - \omega_{AB} - 2\omega_z$, and the pulse envelope is a Gaussian function. Owing to the combined effect of the Aharonov-Bohm and Zeeman energy-level shift, the dipole matrix element is modified. An estimate of the optical dipole matrix element in an axial magnetic field can be obtained from the theory developed in [88], taking into account the bandgap reduction at $B = 8$ T, giving $\wp = 3.6205 \times 10^{-29}$ C · m. We recalculate the relaxation times using the above mentioned magnetic-field dipole coupling, thus obtaining the following relaxation rates: $\Gamma_1 = 2.914\,\text{ns}^{-1}$, $\Gamma_2 = 9.79\,\text{ns}^{-1}$ and $\Gamma_3 = 9.77\,\text{ns}^{-1}$. The intraband relaxation rate, $\gamma = 130\,\text{fs}^{-1}$, and the dephasing rates, $\Gamma_\mu = 800\,\text{fs}^{-1}$ and $\Gamma_{\mu-1} = 1.6\,\text{ps}^{-1}$, are taken to be the same as for the zero-field case. Note that the initial charge-carrier populations are different for excitation by the σ^- and σ^+ pulses: while in the former case the initial population is assumed to be equally distributed between the valence-band levels $\rho_{1'1'i} = \rho_{3'3'i} = 1/2$ nearest the bandgap edge, in the latter case the whole population is in the single valence-band state $\rho_{4'4'i} = 1$.

Similar to the zero-field case, the time evolution of the electric-field components and populations of all four levels are sampled at several points along the nanotube axis, and the magnetic circular dichroism spectra of the gain/absorption coefficients and phase shifts are calculated and displayed in Figure 49.47. Note that the resonance is shifted toward longer wavelengths due to the bandgap reduction. We should point out that the magnetic circular dichroism spectra at $B = 8$ T are quite distinct from the zero magnetic field ones for natural optical activity (Figure 49.44). While the σ^- polarized pulse is amplified during its propagation along the nanotube, the σ^+ polarized pulse is absorbed, resulting in a much larger net circular dichroism. The average value of the magnetic circular dichroism is $0.706\,\mu\text{m}^{-1}$, an order of magnitude larger than the natural circular dichroism. The different behavior of the calculated gain and absorption spectra under σ^- compared to σ^+ polarized optical pulse excitation is a direct consequence of the different energy-level configuration describing the two cases. While the energy-level system for the σ^- excitation is a four-level system, the one corresponding to the σ^+ excitation is a three-level Λ-system, due to level $|2''\rangle$ being completely decoupled from the rest of the levels as a result of the dipole optical selection rules. The calculated spectra in the latter case are, in fact, a signature of EIT and coherent population trapping

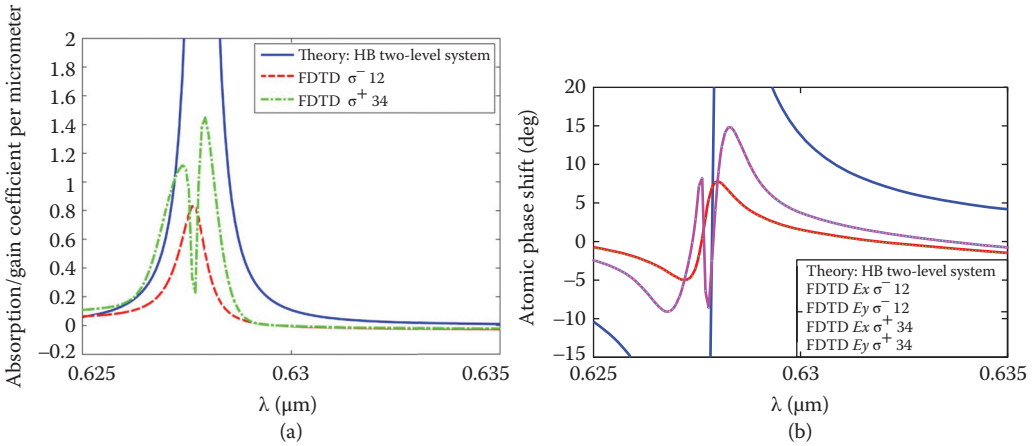


FIGURE 49.47 (a) Spectrum of the gain coefficient per micron, averaged spatially over the entire length of the structure, of the E_x and E_y electric-field vector components of σ^- (dashed curve) and σ^+ (dash-dotted curve) circularly polarized ultrashort optical excitation. The theoretical gain coefficient of a homogeneously broadened (HB), resonant two-level system (solid curve) at $B = 8\,\mathrm{T}$ is also shown. Note that, while the σ^- pulse is amplified during its propagation along the nanotube, the σ^+ pulse is absorbed, resulting in an absorption dip in the spectrum. (b) Spectrum of the phase shift, averaged spatially over the entire length of the structure, at $B = 8\,\mathrm{T}$ of the E_x (dashed curve) and E_y (dots) electric-field components under σ^- excitation, and of E_x (dash-dotted curve) and E_y (dots) under σ^+ circularly polarized ultrashort optical excitation. The phase spectra exhibit double-peaked refraction characteristic of EIT. The theoretical phase shift of a HB resonant two-level system (solid curve), is also given. (FDTD, finite-difference time-domain.)

effects in a three-level system [98], as will be explained subsequently. The absorption at resonance is close to zero, and the spectrum exhibits a dip (Figure 49.47a) similar to the absorption dip characteristic of EIT. The dip is caused by destructive interference between the excitation and emission fields, and leads to the trapping of the charge-carrier population in the ground state (hence, the term "coherent population trapping"). Whereas the phase-shift spectrum for a σ^- pulse excitation is of the type of a two-level atomic phase shift, the spectrum of the phase shift for σ^+ excitation is double peaked (Figure 49.47b). The latter is characteristic of the real part of the dielectric susceptibility in a three-level system exhibiting EIT. The predicted destructive interference in an external axial magnetic field after the passage of the ultrashort pulse is a direct consequence of the timescales of the processes involved in the relaxation dynamics and the phase shift which they introduce in the emission. The calculated average specific rotatory power in a magnetic field $B = 8\,\mathrm{T}$ is $-32580.4°/\mathrm{mm}$, corresponding to an average refractive-index anisotropy $(n_\mathrm{R} - n_\mathrm{L})$ of 6.497 (the purpose of the minus sign in the rotatory power is to denote counterclockwise rotation of the field when viewed from a direction opposite to that of its propagation). This is nearly an order of magnitude greater than the natural optical rotation. We should note that the calculated rotation is a combined effect of the chirality of the nanotube and the magnetic field–induced rotation, and thus can be considered as an estimate for the magneto-chiral effect in a single nanotube. Therefore, our model can be used to study the magnetic field and chirality dependence of the optical rotation angle.

49.8 Conclusion

We have developed a general methodology for description of ultrashort optical-pulse interaction with open, discrete multilevel quantum systems, based on a self-consistent solution of the vector Maxwell's equations and equations for the time evolution of the quantum system in the real-vector representation of the density-matrix geometrical picture. An advantage of this theoretical method is that the coupled vector Maxwell-pseudospin equations can be generated on the fly for a quantum system with an arbitrary number

of energy levels using the corresponding unique SU(N) Lie algebra generators. Unlike the widely accepted approach of using simplified two-level systems for description of the system dynamics, our approach allows to model multiple optical transitions and time-dependent gain effects that occur in a realistic quantum system. Our theoretical framework goes beyond the usual rate equations approach, which is unable to model and predict coherent quantum effects that could arise due to couplings and interplay between multiple optical transitions in the system. Most other approaches use the SVEA and RWA which consider only events occurring on a relatively slow timescale—such as the pulse envelope—but do not take account of the pulse's fast carrier wave. By contrast, the proposed approach allows to model rigorously ultrashort optical pulses down to a few optical cycles. Contrary to BPM, no approximations are made to the electromagnetic field equations and their solutions.

We have shown that the model can describe coherent propagation effects: self-induced and EIT, and self-localization effects, such as soliton and pattern formation in cavities. The method can be applied to a wide range of quantum-system geometries by imposing macroscopic boundary conditions and, hence, it can be used to simulate the performance of not only individual idealized devices, but whole interacting systems. We have shown that the model can successfully describe the complex dynamics of an open quantum system, and at the same time provides the capability to observe the feedback mechanisms—including non-Markovian memory effects—as they occur in space and time. In addition, within the Langevin model extension, we were able to model quantum-optical effects such as spontaneous emission. This approach can be used to simulate quantum correlations, nonclassical light generation and manipulation, and few-photon optical nonlinearities.

The model can describe interactions of matter with circularly/elliptically polarized optical pulses and was successfully applied to the problem of generation, manipulation, and readout of single spins and their dynamics. This success is critical for the development of quantum information technologies—currently a very hot research topic which could potentially revolutionize technological applications. Possible application areas include—but are not limited to—quantum computing and information technologies, such as spin-qubit manipulation in QDs, cold atoms, and NV centers in nanodiamond. CNT and arrays of aligned nanotubes have the potential to serve as functional elements in polarization-sensitive devices and polarization switches, as well as for metamaterial applications in the visible part of the electromagnetic spectrum. Predicted spatio-temporal coherent effects, such as cavity solitons, can be used as integrated low-loss optical switches and for all-optical information processing. Finally, the proposed methodology can be used as a design tool for future quantum-optical devices and for the advancement of quantum-photonic technologies.

References

1. R. J. Glauber. The quantum theory of optical coherence. *Physical Review*, 130:2529, 1963.
2. R. J. Glauber. *Quantum Optics and Electronics*. New York, NY: Gordon Breach, 1965.
3. R. Hanbury Brown and R. Q. Twiss. *Nature*, 177:27, 1956.
4. L. Mandel and S. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, 1995.
5. B. Lounis and M. Orrit. Single-photon sources. *Report on Progress in Physics*, 68:1129, 2005.
6. A. Einstein, B. Podolsky, and N. Rosen. Can quantum mechanical description of physical reality be considered complete? *Physical Review*, 47:777, 1935.
7. J. L. O'Brien, A. Furusawa, and J. Vuckovic. Photonic quantum technologies. *Nature Photonics*, 3:687, 2009.
8. T. M. Babinec, B. J. M. Hausmann, M. Khan, Y. Zhang, J. R. Maze, Ph. R. Hemmer, and M. LonÄar. A diamond nanowire single-photon source. *Nature Nanotechnology*, 5:195, 2010.
9. J. Claudon, J. Bleuse, N. S. Malik, M. Bazin, P. Jaffrennou, N. Gregersen, C. Sauvan, P. Lalanne, and J.-M. Gerard. A highly efficient single-photon source based on a quantum dot in a photonic nanowire. *Nature Photonics*, 4:174, 2010.

10. D. Dalacu, K. Mnaymneh, V. Sazonova, P. J. Poole, G. C. Aers, J. Lapointe, R. Cheriton, A. J. SpringThorpe, and R. Williams. Deterministic emitter-cavity coupling using a single-site controlled quantum dot. *Physical Review B*, 82:033301, 2010.
11. K. G. Lee, X. W. Chen, H. Eghlidi, P. Kukura, R. Lettow, A. Renn, V. Sandoghdar, and S. Götzinger. A planar dielectric antenna for directional single-photon emission and near-unity collection efficiency. *Nature Photonics*, 5:166, 2011.
12. S. Strauf, N. G. Stoltz, M. T. Rakher, L. A. Coldren, P. M. Petroff, and D. Bouwmeester. High-frequency single-photon source with polarization control. *Nature Photonics*, 1:704, 2007.
13. D. Loss and D. P. diVincenzo. Quantum computation with quantum dots. *Physical Review A*, 57:120, 1998.
14. B. Meier and B. P. Zakharchenya editors. *Optical Orientation*. North-Holland, Amsterdam: Elsevier, 1984.
15. W. S. Warren, H. Rabitz, and M. Dahleh. Coherent control of quantum dynamics: The dream is alive. *Science*, 259:1581, 1993.
16. K. De Greve, D. Press, P. L. McMahon, and Y. Yamamoto. Ultrafast optical control of individual quantum dot spin qubits. *Reports on Progress in Physics*, 76:092501, 2013.
17. M. Sargent III, M. O Scully, and W. E. Lamb Jr. *Laser Physics*. Reading, MA: Addison-Wesley Publishing Company Advanced Book Program, 1974.
18. S. L. McCall and E. L. Hahn. Self-induced transparency by pulsed coherent light. *Physical Review Letters*, 18:908, 1967.
19. S. Schneider, P. Borri, W. Langbein, U. Woggon, J. Förstner, A. Knorr, R. L. Sellin, D. Ouyang, and D. Bimberg. Self-induced transparency in InGaAs quantum-dot waveguides. *Applied Physics Letters*, 83:3668, 2003.
20. S. Hughes. Breakdown of the area theorem: Carrier-wave Rabi flopping of femtosecond optical pulses. *Physical Review Letters*, 81:3363, 1998.
21. D. Pinotsi and A. Imamoglu. Single photon absorption by a single quantum emitter. *Physical Review Letters*, 100:093603, 2008.
22. M. Sondermann, R. Maiwald, H. Konermann, N. Lindlein, U. Peschel, and G. Leuchs. Design of a mode converter for efficient light-atom coupling in free space. *Applied Physics B*, 89:489, 2007.
23. S. J. van Enk. Atoms, dipole waves, and strongly focused light beams. *Physical Review A*, 69:043813, 2004.
24. M. Brune, F. Schmidt-Kaler, A. Maali, J. Dreyer, E. Hagley, J. M. Raimond, and S. Haroche. Quantum Rabi oscillation: A direct test of field quantization in a cavity. *Physical Review Letters*, 76:1800, 1996.
25. K. M. Birnbaum, A. Boca, R. Miller, A. D. Boozer, T. E. Nortup, and H. J. Kimble. Photon blockade in an optical cavity with one trapped atom. *Nature*, 436, 87:87, 2005.
26. R. Young, R. M. Stevenson, P. Atkinson, K. Cooper, D. A. Ritchie, A. J. Shields. *Advances in Solid State Physics*, Vol. 46:55. Berlin: Springer Verlag, 2008.
27. O. Gazzano, S. Michaelis de Vasconcellos, C. Arnold, A. Nowak, E. Galopin, I. Sagnes, L. Lanco, A. Lemaître, and P. Senellart. Bright solid-state sources of indistinguishable single photons. *Nature communications*, 4:1425, 2012.
28. K. Srinivasan and O. Painter. Linear and nonlinear optical spectroscopy of a strongly coupled micro disk-quantum dot system. *Nature*, 450:862, 2007.
29. N. Le Thomas, U. Woggon, O. Schöps, M. V. Artemyev, M. Kazes, and U. Banin. Cavity QED with semiconductor nanocrystals. *Nano Letters*, 6:557, 2006.
30. M. Heiss, Y. Fontana, A. Gustafsson, G. Wust, C. Magen, D. D. O'Regan, J. W. Luo, B. Ketterer, S. Conesa-Boj, A. V. Kuhlmann, J. Houel, E. Russo-Averchi, J. R. Morante, M. Cantoni, N. Mazari, J. Arbiol, A. Zunger, R. J. Warburton, and A. Fontcuberta i Morral. Self-assembled quantum dots in a nanowire system for quantum photonics. *Nature Materials*, 12:439, 2013.

31. H. Kim, R. Bose, T. C. Shen, G. S. Solomon, and E. Waks. A quantum logic gate between a solid-state quantum bit and a photon. *Nature Photonics*, 7:373, 2013.
32. A. Faraon, I. Fushman, D. Englund, N. Stoltz, P. Petroff, and J. Vučković. Coherent generation of non-classical light on a chip via photon-induced tunnelling and blockade. *Nature Physics*, 4:859, 2008.
33. J. Kasprzak, S. Reitzenstein, E. A. Muljarov, C. Kistner, C. Schneider, M. Strauss, S. Hing, A. Forchel, and W. Langbein. Up on the Jaynes-Cummings ladder of a quantum-dot/microcavity system. *Nature Materials*, 9:304, 2010.
34. P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, and A. Imamoglu. A quantum dot single-photon turnstile device. *Science*, 290:2282, 2000.
35. C. Santori, D. Fattal, J. Vuckovic, G. S. Solomon, and Y. Yamamoto. Indistinguishable photons from a single-photon device. *Nature*, 419:594, 2002.
36. E. Knill, R. Laamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409:46, 2001.
37. L. Allen and J. H. Eberly. *Optical Resonance and Two-Level Atoms*. New York, NY: Wiley, 1975.
38. R. P. Feynman, F. L. Vernon, and R. W. Hellwarth. Geometrical representation of the Schrödinger equation for solving laser problems. *Journal of Applied Physics*, 28:49, 1957.
39. R. J. Cook and B. W. Shore. Coherent dynamics of N-level atoms and molecules. III. An analytically soluble periodic case. *Physical Review A*, 20:539, 1979.
40. J. N. Elgin. Semiclassical formalism for the treatment of three-level systems. *Physics Letters*, 80A:140, 1980.
41. M. Gell-Mann and Y. Neeman. *The Eightfold Way*. New York, NY: Benjamin, 1964.
42. F. T. Hioe and J. H. Eberly. N-level coherence vectors and higher conservation laws in quantum optics and quantum mechanics. *Physical Review Letters*, 47:838, 1981.
43. S. L. McCall and E. L. Hahn. Self-induced transparency. *Physical Review*, 183:457, 1969.
44. G. Slavcheva. Model for the coherent optical manipulation of a single spin state in a charged quantum dot. *Physical Review B*, 77:115347, 2008.
45. G. Slavcheva, J. M. Arnold, I. Wallace, and R. W. Ziolkowski. Coupled Maxwell-pseudospin equations for investigation of self-induced transparency effects in a degenerate three-level quantum system in two dimensions: Finite-difference time-domain study. *Physical Review A*, 66:063418, 2002.
46. A. Taove. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Norwood, MA: Artech House, 2000.
47. R. W. Ziolkowski, J. M. Arnold, and D. M. Gogny. Ultrafast pulse interactions with two-level atoms. *Physical Review*, 52:3082, 1995.
48. B. Engquist and A. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Mathematics of Computation*, 31:629, 1977.
49. G. Mur. Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations. *IEEE Transactions on Electromagnetic Compatibility*, 23:377, 1981.
50. A. Yariv. *Quantum Electronics*. New York, NY: Wiley, 1989.
51. G. M. Slavcheva, J. M. Arnold, and R.W. Ziolkowski. FDTD simulation of the nonlinear gain dynamics in active optical waveguides and semiconductor microcavities. *IEEE Journal of Selected Topics in Quantum Electronics*, 10(5):1052, 2004.
52. A. E. Siegman. *Lasers*. Mill Valley, CA: University of Science Books, 1986.
53. G. Slavcheva, J. M. Arnold, and R. W. Ziolkowski. Ultrashort pulse lossless propagation through a degenerate three-level medium in nonlinear optical waveguides and semiconductor microcavities. *IEEE Journal of Selected Topics in Quantum Electronics*, 9(3):929, 2003.
54. M. Blaaboer, G. Kurizki, and B. A. Malomed. Spatiotemporally localized solitons in resonantly absorbing Bragg reflectors. *Physical Review E*, 62:R57, 2000.

55. A. Kozhekin and G. Kurizki. Self-induced transparency in Bragg reflectors: Gap solitons near absorption resonances. *Physical Review Letters*, 74:5020, 1995.
56. A. Kozhekin, G. Kurizki, and B. A. Malomed. Standing and moving gap solitons in resonantly absorbing Bragg gratings. *Physical Review Letters*, 81:3647, 1998.
57. G. Slavcheva and O. Hess. Dynamical model of coherent circularly polarized optical pulse interactions with two-level quantum systems. *Physical Review A*, 72:053804, 2005.
58. A. I. Maimistov, A. M. Besharov, O. Elyutin, and Yu. M. Sklyarov. Present state of self-induced transparency theory. *Physics Reports*, 191:1–108, 1990.
59. G. Slavcheva and O. Hess. Spin-dependent dynamics of ultrafast polarised optical pulse propagation in coherent semiconductor quantum systems. *Physica Status Solidi C*, 3:2414, 2006.
60. G. Slavcheva and O. Hess. All-optical coherent control of spin dynamics in semiconductor quantum dots. *Optical and Quantum Electronics*, 31(12–14):973, 2007.
61. A. Wojs, P. Hawrylak, S. Fafard, and L. Jacak. Electronic structure and magneto-optics of self-assembled quantum dots. *Physical Review B*, 54:5604, 1996.
62. A. Greilich, R. Oulton, E. A. Zhukov, I. A. Yugova, D. R. Yakovlev, M. Bayer, A. Shabaev, A. L. Efros, I. A. Merkulov, V. Stavarache, D. Reuter, and A. Wieck. Optical control of spin coherence in singly charged (In,Ga)As/GaAs quantum dots. *Physical Review Letters*, 96:227401, 2006.
63. P. Chen, C. Piermarocchi, L. J. Sham, D. Gammon, and D. G. Steel. Theory of quantum optical control of a single spin in a quantum dot. *Physical Review B*, 69:075320, 2004.
64. A. Shabaev, A. L. Efros, D. Gammon, and I. A. Merkulov. Optical readout and initialization of an electron spin in a single quantum dot. *Physical Review B*, 68:201305(R), 2003.
65. A. V. Khaetskii, D. Loss, and L. Glazman. Electron spin decoherence in quantum dots due to interaction with nuclei. *Physical Review Letters*, 88:186802, 2002.
66. I. A. Merkulov, A. L. Efros, and M. Rosen. Electron spin relaxation by nuclei in semiconductor quantum dots. *Physical Review B*, 65:205309, 2002.
67. D. V. Bulaev and V. Loss. Spin relaxation and decoherence of holes in quantum dots. *Physical Review Letters*, 95:076805, 2005.
68. T. Flissikowski, I. A. Akimov, A. Hundt, and F. Henneberger. Single-hole spin relaxation in a quantum dot. *Physical Review B*, 68:161309(R), 2003.
69. S. Economou, R.-B. Liu, L. J. Sham, and D. G. Steel. Unified theory of consequences of spontaneous emission in a λsystem. *Physical Review B*, 71:195327, 2005.
70. M. Taylor, E. Harbord, P. Spencer, E. Clarke, G. Slavcheva, and R. Murray. Optical spin-filtering effect in charged InAs/GaAs quantum dots. *Applied Physics Letters*, 97:171907, 2010.
71. K. Kavokin. Fine structure of the quantum-dot trion. *Physica Status Solidi: A*, 195:592, 2003.
72. P. F. Braun, X. Marie, L. Lombez, B. Urbaszek, T. Amand, P. Renucci, V. K. Kalevich, K. V. Kavokin, O. Krebs, P. Voisin, and Y. Masumoto. Direct observation of the electron spin relaxation induced by nuclei in quantum dots. *Physical Review Letters*, 94:116601, 2005.
73. B. Eble, C. Testelin, P. Desfonds, F. Bernardot, A. Balocchi, T. Amand, A. Miard, A. Lemaître, X. Marie, and M. Chamarro. Hole-nuclear spin interaction in quantum dots. *Physical Review Letters*, 102:146601, 2009.
74. G. Slavcheva. Model of coherent optical spin manipulation through hot trion states in p-doped InAs/GaAs quantum dots. *ArXiv:1301.7018v1*, 2013.
75. T. Warming, E. Siebert, A. Schliwa, E. Stock, R. Zimmermann, and D. D. Bimberg. Hole-hole and electron-hole exchange interactions in single InAs/GaAs quantum dots. *Physical Review B*, 79:125316, 2009.
76. G. A. Narvaez, G. Bester, and A. Zunger. Carrier relaxation mechanisms in self-assembled (In, Ga) As/GaAs quantum dots: Efficient P→S auger relaxation of electrons. *Physical Review B*, 74:075403, 2006.
77. C. W. Gardiner and P. Zoller. *Quantum Noise*, 3rd ed. Berlin: Springer, 2004.

78. V. Weisskopf and E. Wigner. Berechnung der natürlichen linienbreite auf grund der diracschen lichttheorie. *Zeitschrift für Physik*, 63:54, 1930.
79. P. A. M. Dirac. The quantum theory of the emission and absorption of radiation. *Proceedings of the Royal Society London A*, 114:243, 1927.
80. G. Slavcheva and Ph. Roussignol. Nonlinear coherent magneto-optical response of a single chiral carbon nanotube. *New Journal of Physics*, 12:103004, 2010.
81. L. Yang, M. P. Anantram, J. Han, and J. P. Lu. Band-gap change of carbon nanotubes: Effect of small uniaxial and torsional strain. *Physical Review B*, 60:13874, 1999.
82. G. L. A. Rikken and E. Raupach. Observation of magneto-chiral dichroism. *Nature*, 390:493, 390.
83. A. Ajiki and T. Ando. Electronic states of carbon nanotubes. *Journal of the Physical Society of Japan*, 62:1255, 1993.
84. Y. Zhang, A. Chang, J. Cao, Q. Wang, W. Kim, Y. Li, N. Morris, E. Yenilmez, J. Kong, and H. Daia. Electric-field-directed growth of aligned single-walled carbon nanotubes. *Applied Physics Letters*, 79:3155, 2001.
85. L. Jin, C. Bower, and O. Zhou. Alignment of carbon nanotubes in a polymer matrix by mechanical stretching. *Applied Physics Letters*, 73:1197, 1998.
86. Ge. G. Samsonidze, A. Grüneis, R. Saito, A. Jorio, A. G. Souza Filho, G. Dresselhaus, and M. S. Dresselhaus. Interband optical transitions in left- and right-handed single-wall carbon nanotubes. *Physical Review B*, 69:205402, 2004.
87. J.-S. Lauret, C. Voisin, G. Cassabois, C. Delalande, Ph. Roussignol, O. Jost, and L. Capes. Ultrafast carrier dynamics in single-wall carbon nanotubes. *Physical Review Letters*, 90:057404, 2003.
88. E. L. Ivchenko and B. Spivak. Chirality effects in carbon nanotubes. *Physical Review B*, 66:155404, 2002.
89. F. Wang, G. Dukovic, L. E. Brus, and T. F. Heinz. Time-resolved fluorescence of carbon nanotubes and its implication for radiative lifetimes. *Physical Review Letters*, 92:177401, 2000.
90. H. Bao, X. Ruan, and T. S. Fisher. Optical properties of ordered vertical arrays of multi-walled carbon nanotubes from FDTD simulations. *Optics Express*, 18:6347, 2010.
91. V. N. Popov and L. Henrard. Comparative study of the optical properties of single-walled carbon nanotubes within orthogonal and nonorthogonal tight-binding models. *Physical Review B*, 70:115407, 2004.
92. V. N. Popov. Curvature effects on the structural, electronic and optical properties of isolated single-walled carbon nanotubes within a symmetry-adapted non-orthogonal tight-binding model. *New Journal of Physics*, 6:17, 2004.
93. J. Jiang, J. Ding, and D. Y. Xing. Zeeman effect on the electronic spectral properties of carbon nanotubes in an axial magnetic field. *Physical Review B*, 62:13209, 2000.
94. E. D. Minot, Y. Yaish, V. Sazonova, and P. L. McEuen. Determination of electron orbital magnetic moments in carbon nanotubes. *Nature*, 428:536, 2004.
95. F. L. Shuy, C. P. Chang, R. B. Chen, C. W. Chiu, and M. F. Lin. Magnetoelectronic and optical properties of carbon nanotubes. *Physical Review B*, 67:045405, 2003.
96. H. Ajiki. Magnetic-field effects on the optical spectra of a carbon nanotube. *Physical Review B*, 65:233409, 2002.
97. R. Saito, G. Dresselhaus, and M. S. Dresselhaus. Magnetic energy bands of carbon nanotubes. *Physical Review B*, 50:14698, 1994.
98. K. J. Boller, A. Imamoglu, and S. E. Harris. Observation of electromagnetically induced transparency. *Physical Review Letters*, 66:2593, 1991.

X

Mathematical Methods

50 Drift-Diffusion Models *Patricio Farrell, Nella Rotundo, Duy Hai Doan, Markus Kantner, Jürgen Fuhrmann, and Thomas Koprucki* 733
Introduction • How to Get Started: Numerical Solution of the Stationary 1D Semiconductor Equations • Van Roosbroeck System • Discretizing the van Roosbroeck System • Nonlinear Solvers • Mesh Generation for the Finite Volume Method • Time Discretization • Contact Terminal Currents • An Alternative: The Finite Element Method • Extensions and Outlook

51 Monte Carlo Device Simulations *Katerina Raleva, Abdul R. Shaik, Raghuraj Hathwar, Akash Laturia, Suleman S. Qazi, Robin Daugherty, Dragica Vasileska, and Stephen M. Goodnick* ... 773
Introduction • Bulk Monte Carlo Method Description • Particle-Based Device Simulation • Representative Simulation Results • Outlook

52 Photonics *Frank Schmidt* 807
Landscape of Numerical Methods • Maxwell's Equations in Nondispersive, Chargeless Media • FDTD Method • Fourier Modal Method • Finite Element Method • DG Method

50

Drift-Diffusion Models

50.1 Introduction 733
50.2 How to Get Started: Numerical Solution of the Stationary 1D Semiconductor Equations 734
1D van Roosbroeck Model • Thermodynamic Equilibrium and Local Charge Neutrality • Discretizing the Poisson Problem • Solving the Discrete Nonlinear Poisson Problem • Discretizing the Continuity Equations Using the Scharfetter–Gummel Scheme • Solving the Full Discretized van Roosbroeck System • Why Not Use Central Finite Differences for the Fluxes?
50.3 Van Roosbroeck System 745
The van Roosbroeck System of Equations • Initial and Boundary Conditions • Recombination Processes • Thermodynamic Equilibrium • Non-Boltzmann Statistics and Generalized Einstein Relation • Free Energy and Dissipation Rate • Existence and Uniqueness Results • Maximum Principle
50.4 Discretizing the van Roosbroeck System 753
Finite Volume Method in Higher Dimensions • Scharfetter–Gummel Fluxes and Their Non-Boltzmann Generalizations • Flux Expressions Consistent with the Thermodynamic Equilibrium • Free Energy and Dissipation Rate • Existence, Uniqueness, and Convergence • Maximum Principle
50.5 Nonlinear Solvers 759
Newton's Method • Gummel's Method
50.6 Mesh Generation for the Finite Volume Method 761
Boundary Conforming Delaunay Meshes on Polyhedral Domains • Tensor Product Approaches • Open Problems
50.7 Time Discretization 763
50.8 Contact Terminal Currents 763
50.9 An Alternative: The Finite Element Method 765
50.10 Extensions and Outlook 766
Additional Physical Models • Coupling the van Roosbroeck System to Other Models • Methods for Doping Optimization • Alternative Modeling Approaches for Carrier Transport

Patricio Farrell
Nella Rotundo
Duy Hai Doan
Markus Kantner
Jürgen Fuhrmann
and
Thomas Koprucki

50.1 Introduction

The semiconductor technology is undeniably one of the most important branches in modern industry. Apart from its obvious significance to our daily lives, this technology is an excellent example of a broad collaboration among various disciplines. The development of novel technologies and devices has not only

been driven by engineers, but also by physicists, mathematicians, and numerical analysts. On the one hand, rigorous mathematical models allow sophisticated predictions, which might be difficult to observe experimentally. On the other hand, numerical simulations have the potential to optimize device designs without the costly and time-consuming development of prototypes.

In 1950, van Roosbroeck introduced the fundamental semiconductor device equations [1] as a system of three nonlinearly coupled partial differential equations (PDEs). They describe the semiclassical transport of free electrons and holes in a self-consistent electric field using a drift-diffusion approximation. Since then, the so-called *van Roosbroeck system* (frequently also called *drift-diffusion system*) became the standard model to describe the current flow in semiconductor devices at a macroscopic scale. Typical devices modeled by these equations range from diodes and transistors to LEDs, solar cells, and lasers [2,3]. In recent years, the emergence of quantum nanostructures [4–7] and organic semiconductors [8–10] has invigorated the research activities in this area.

In this chapter, we focus on solving the van Roosbroeck system numerically. From a mathematical point of view, the challenge lies in the strong nonlinearities, the drift dominated nature of the underlying physics, the formation of internal and boundary layers as well as the need to accurately mirror qualitative physical properties, such as nonnegativity of carrier densities and consistency with thermodynamical principles. The finite difference scheme invented by Scharfetter and Gummel [11] was the first to deal appropriately with most of these difficulties for charge carrier densities described by Boltzmann statistics. Later on, Fichtner, Rose, and Bank [12] extented it to higher dimensions as a finite volume scheme, providing geometric flexibility. Selberherr combined a lot of these results along with his own contributions in his textbook [13] which has become one of the standard reference works in the field.

However, often the Boltzmann approximation is insufficient and one has to impose Fermi–Dirac statistics or other constitutive laws for the charge carrier densities. For this reason, we present recent attempts to generalize the Scharfetter–Gummel approach while still reflecting the physics correctly.

In order to enable readers not familiar with semiconductor simulations to quickly grasp the main points in the discretization process of the van Roosbroeck system, we start by studying a one-dimensional (1D) problem in Section 50.2. Even though we have to postpone some of the solid-state physics to the third section, we can already introduce many key concepts of the standard finite difference ansatz proposed by Scharfetter and Gummel, which we will interpret as a 1D finite volume scheme.

In Section 50.3, we introduce the full van Roosbroeck system in higher spatial dimensions along with its physical concepts. In Section 50.4, we look at its numerical solution. We present the two-point flux finite volume method for the stationary case, which allows to use the Scharfetter–Gummel scheme in higher dimensions and discuss its generalization to more general carrier statistics. The remaining sections are devoted to special numerical aspects: nonlinear solvers, mesh generation, time stepping, the correct computation of terminal currents, and the finite element method. In the final section, we discuss generalizations of the approach and its embedding into more complex physical models.

50.2 How to Get Started: Numerical Solution of the Stationary 1D Semiconductor Equations

This section serves as a quick introduction to convey the main idea behind the van Roosbroeck system and its discretization. The van Roosbroeck system consists of three equations modeling the charge carrier flow as well as the electrostatic potential distribution in a semiconductor device. However, since the physics behind these equations is rather cumbersome and the focus of this chapter are numerical methods, we omit some of these physical details for now and start with a relatively simple 1D problem. Thus, we can highlight the key elements (and difficulties!) in the discretization of the van Roosbroeck system. This approach may appeal to newcomers to the field of optoelectronics. Experienced readers may jump directly to Section 50.3, where we will dive into the physical meaning of this model.

Here, we wish to model the stationary state of a 1D semiconductor device consisting of a homogeneous material on an interval $\Omega = [0, L]$ with Ohmic contacts at each end. This configuration already covers the simulation of p-n and p-i-n junctions.

50.2.1 1D van Roosbroeck Model

The stationary 1D van Roosbroeck system consists of three nonlinear ordinary differential equations for the unknown electrostatic potential $\psi(x)$, the quasi-Fermi potential for electrons $\varphi_n(x)$ and the quasi-Fermi potential for holes $\varphi_p(x)$. It links Poisson's equation for the electric field to the continuity equations for the carrier densities as follows:

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\varepsilon_s \frac{\mathrm{d}}{\mathrm{d}x}\psi\right) = q\Big(C + p(\psi, \varphi_p) - n(\psi, \varphi_n)\Big), \tag{50.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = \quad qR(\psi, \varphi_n, \varphi_p), \tag{50.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} j_p = -qR(\psi, \varphi_n, \varphi_p). \tag{50.3}$$

The fluxes or current densities are given by

$$j_n = -q\mu_n n(\psi, \varphi_n)\frac{\mathrm{d}}{\mathrm{d}x}\varphi_n \qquad \text{and} \qquad j_p = -q\mu_p p(\psi, \varphi_p)\frac{\mathrm{d}}{\mathrm{d}x}\varphi_p. \tag{50.4}$$

The electron density n and the hole density p are related to the electric potential ψ as well as the quasi-Fermi potentials of electrons and holes via

$$n(\psi, \varphi_n) = N_c \exp\left(\frac{q(\psi - \varphi_n) - E_c}{k_\mathrm{B} T}\right) \quad \text{and} \quad p(\psi, \varphi_p) = N_v \exp\left(\frac{q(\varphi_p - \psi) + E_v}{k_\mathrm{B} T}\right). \tag{50.5}$$

We stress here that in (50.5) we have used the so-called Boltzmann approximation. In general, the exponentials will be replaced by some monotonically increasing *statistical distribution function* $\mathcal{F}$. This function will be discussed in greater detail in Section 50.3.

Next, we explain the constants and functions in the van Roosbroeck system. The elementary charge q and the Boltzmann constant k_B are universal constants. The (absolute) dielectric permittivity $\varepsilon_\mathrm{s} = \varepsilon_0 \varepsilon_\mathrm{r}$ is given as the product of the vacuum dielectric permittivity ε_0 and the relative permittivity of the semiconductor ε_r in static (low frequency) limit. The carrier mobilities μ_n and μ_p, the conduction and valence band densities of states N_c and N_v as well as the conduction and valence band-edge energies E_c and E_v are assumed to be constant even though in general they can vary with the material (e.g., due to abrupt or graded heterojunctions). The temperature T is also assumed to be constant; in general it can be space and even time dependent. The doping profile $C = C(x)$ describes material properties and the function $R(\psi, \varphi_n, \varphi_p)$ models the recombination and generation of electron–hole pairs. The electric field $\mathbf{E}$ is determined by the derivative (in higher dimensions the gradient) of the electrostatic potential

$$\mathbf{E} = -\frac{\mathrm{d}}{\mathrm{d}x}\psi\, \mathbf{e}_x,$$

where $\mathbf{e}_x$ denotes the unit normal along the x axis. A more detailed discussion on the physics behind these different quantities will be given in Section 50.3.

In the literature, authors frequently use a different set of unknowns for the van Roosbroeck system by replacing the quasi-Fermi potentials with the carrier densities n and p. However, we prefer to use the quasi-Fermi potentials for several reasons: they allow to easily model abrupt or graded heterostructures

and appear naturally in the thermodynamic description of the van Roosbroeck system since the negative gradients of the quasi-Fermi potentials are the driving force of the currents [14]. Even mathematically, they are more beautiful as they make it possible to write the whole van Roosbroeck system in a gradient form [15] consistent with the principles of nonequilbrium thermodynamics [16]. Also carrier densities as opposed to quasi-Fermi potentials vary drastically in magnitude, which may lead to numerical difficulties.

The system of Equations 50.1 through 50.3 needs to be supplied with boundary conditions at $x = 0$ and $x = L$. For applied external voltages U_1 and U_2, we require Dirichlet boundary conditions at the *Ohmic contacts*, i.e.,

$$\psi(0) = \psi_0(0) + U_1, \qquad \psi(L) = \psi_0(L) + U_2, \tag{50.6}$$

$$\varphi_n(0) = U_1, \qquad \varphi_n(L) = U_2, \tag{50.7}$$

$$\varphi_p(0) = U_1, \qquad \varphi_p(L) = U_2. \tag{50.8}$$

We discuss the meaning of the potential ψ_0 in Section 50.2.2.

50.2.2 Thermodynamic Equilibrium and Local Charge Neutrality

The goal is to enforce that the van Roosbroeck system of (50.1) through (50.3) is consistent with the *thermodynamic equilibrium*, which is a physical state defined by vanishing currents, namely

$$j_n = j_p = 0 \qquad \text{implying} \qquad \varphi_0 := \varphi_n = \varphi_p = \text{const.} \tag{50.9}$$

Without loss of generality, we can set $\varphi_0 = 0$. We also discuss the thermodynamic equilibrium in Section 50.3.4. As a consequence, the three differential equations (50.1) through (50.3) reduce to the 1D nonlinear Poisson equation

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\varepsilon_s \frac{\mathrm{d}}{\mathrm{d}x}\psi\right) = q\left(N_v \exp\left(\frac{E_v - q\psi}{k_\mathrm{B}T}\right) - N_c \exp\left(\frac{q\psi - E_c}{k_\mathrm{B}T}\right) + C\right). \tag{50.10}$$

We supply it with Dirichlet boundary conditions (50.6) for zero voltages $U_1 = U_2 = 0$. The solution of Equation 50.10 with these boundary conditions is the *built-in potential* denoted with ψ_eq. In general, we cannot expect to find an analytic expression for ψ_eq.

Another important physical concept is *local charge neutrality*, characterized by a vanishing left-hand side in the Poisson equation. Combined with the equilibrium condition (50.9), this leads to

$$0 = q\left(N_v \exp\left(\frac{E_v - q\psi_0}{k_\mathrm{B}T}\right) - N_c \exp\left(\frac{q\psi_0 - E_c}{k_\mathrm{B}T}\right) + C\right), \tag{50.11}$$

which can be solved for ψ_0 by solving a quadratic equation and omitting the unphysical solution, yielding

$$\psi_0(x) = \frac{E_\mathrm{c} + E_\mathrm{v}}{2q} - \frac{1}{2}U_\mathrm{T}\log\left(\frac{N_c}{N_v}\right) + U_\mathrm{T}\,\mathrm{arcsinh}\left(\frac{C}{2N_\mathrm{intr}}\right), \tag{50.12}$$

where the intrinsic carrier density N_intr and the *thermal voltage* U_T are given by

$$N_\mathrm{intr}^2 = N_c N_v \exp\left(-\frac{E_c - E_v}{k_\mathrm{B}T}\right) \qquad \text{and} \qquad U_T = \frac{k_\mathrm{B}T}{q}. \tag{50.13}$$

The potential ψ_0 serves two purposes. In Section 50.2.4, it will be used to obtain an initial guess for the nonlinear iteration to solve a discretized version of Equation 50.10. On the other hand, modeling (ideal) Ohmic contacts requires local charge neutrality. Hence, the boundary values in (50.6) are obtained by evaluating the function ψ_0 at the contacts $x = 0$ and $x = L$. The voltage difference between the boundary values $\psi_0(0)$ and $\psi_0(L)$ is called *built-in voltage*.

50.2.3 Discretizing the Poisson Problem

Our goal is to find a discrete solution to Equation 50.1. For this, we will use the *finite volume method* which naturally preserves many important physical properties. To see this, the reader may have a quick look in the table of contents and compare the next two sections devoted to the van Roosbroeck system and its numerical solution. It will become apparent that for many aspects the continuous model and its discretized counterpart are closely related.

We discretize the interval using a (possibly) nonuniform mesh of the form

$$0 =: x_1 < \cdots < x_N := L.$$

We associate each node x_k with a *control volume* ω_k defined by the interior subintervals

$$\omega_k = \left[x_{k-1,k}, x_{k,k+1}\right] \quad \text{for} \quad k = 2, \ldots, N-1$$

and the boundary subintervals

$$\omega_1 = \left[x_1, x_{1,2}\right] \quad \text{and} \quad \omega_N = \left[x_{N-1,N}, x_N\right].$$

The boundaries of the kth control volume are given by

$$x_{k-1,k} := \frac{x_{k-1} + x_k}{2} \qquad \text{and} \qquad x_{k,k+1} := \frac{x_k + x_{k+1}}{2}$$

for $k = 2, \ldots, N-1$. We point out that we have generated a disjoint partition of the interval, $\Omega = \bigcup_{k=1}^{N} \omega_k$. This discretization along with its notation is illustrated in Figure 50.1.

We introduce a finite volume discretization for Poisson's equation. Using the fundamental theorem of calculus, we integrate Equation 50.1 over the interior control volumes ω_k and obtain Gauss's law of electrodynamics:

$$-\varepsilon_s \left(\left.\frac{\mathrm{d}\psi}{\mathrm{d}x}\right|_{x_{k,k+1}} - \left.\frac{\mathrm{d}\psi}{\mathrm{d}x}\right|_{x_{k-1,k}} \right) = \int_{\omega_k} q\Big(C + p(\psi, \varphi_p) - n(\psi, \varphi_n)\Big)\,\mathrm{d}x. \tag{50.14}$$

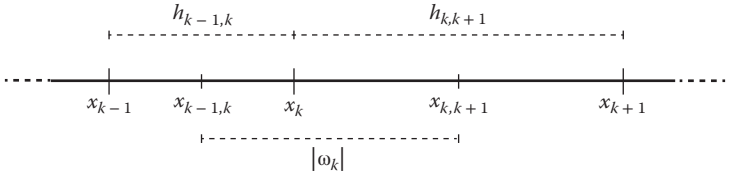


FIGURE 50.1 1D discretization around the kth node along with the corresponding notation.

This equation constitutes a balance law that can be interpreted as follows: the total charge in the kth control volume is given by the difference of the electric displacement $D = -\varepsilon_s \frac{\mathrm{d}}{\mathrm{d}x}\psi$ at its interfaces.

So far, we have derived an integral form of Poisson's equation for every interior control volume. We need approximations of the electric displacements at the boundary of each control volume. For this we employ central finite differences:

$$\left.\frac{\mathrm{d}\psi}{\mathrm{d}x}\right|_{x_{k,k+1}} \approx \frac{\psi_{k+1} - \psi_k}{h_{k,k+1}}, \tag{50.15}$$

with the distance between two neighboring nodes given by

$$h_{k,k+1} := x_{k+1} - x_k.$$

Moreover, the integral over the net charge density is approximated by the rectangle method, evaluating the integrand at node x_k times the size of the control volume $|\omega_k| = x_{k,k+1} - x_{k-1,k}$. That is, the balance equation (50.14) is approximated by

$$-\varepsilon_s \left(\frac{\psi_{k+1} - \psi_k}{h_{k,k+1}} - \frac{\psi_k - \psi_{k-1}}{h_{k-1,k}} \right) = q\Big(C_k + p(\psi_k, \varphi_{p;k}) - n(\psi_k, \varphi_{n;k}) \Big) |\omega_k|$$

or, slightly rearranged,

$$\begin{aligned} 0 = -\varepsilon_s \left(\frac{1}{h_{k,k+1}}\psi_{k+1} - \left[\frac{1}{h_{k,k+1}} + \frac{1}{h_{k-1,k}} \right] \psi_k + \frac{1}{h_{k-1,k}}\psi_{k-1} \right) \\ - q\Big(C_k + p(\psi_k, \varphi_{p;k}) - n(\psi_k, \varphi_{n;k}) \Big) |\omega_k|, \end{aligned} \tag{50.16}$$

where $k = 2, \ldots, N-1$. The subindices denote evaluation at a given node. So, for example, ψ_k means $\psi(x_k)$ as well as $\varphi_{n;k}$ indicates $\varphi_n(x_k)$. We use this nodal notation frequently within this chapter. To incorporate the Dirichlet boundary conditions (50.6), we simply impose two more equations for the endpoints, namely

$$\psi_1 = \psi_0(0) + U_1 \qquad \text{and} \qquad \psi_N = \psi_0(L) + U_2. \tag{50.17}$$

50.2.4 Solving the Discrete Nonlinear Poisson Problem

Using the finite volume discretization scheme we have just described, we solve the nonlinear Poisson equation (50.16) with Dirichlet boundary conditions (50.17) in thermodynamic equilibrium ($U_1 = U_2 = 0$) to obtain the built-in potential ψ_{eq}. Recalling that there are $N-2$ interior and two boundary nodes, we end up with a system of N variables which can be summarized by the nonlinear discrete system

$$\mathbf{0} = \mathbf{F}_1(\boldsymbol{\psi}_{\mathrm{eq}}),$$

where $\boldsymbol{\psi}_{\mathrm{eq}} = (\psi_{\mathrm{eq}}(x_k))_{k=1}^N$ is the vector of the nodal values. Solving this system is not straightforward due to its inherent nonlinearity. One approach to obtain a solution is Newton's method, which we discuss in more detail in Section 50.5. Newton's method converges quadratically—if one has a starting guess sufficiently close to the solution [17].

50.2.4.1 Solution Procedure: Nonlinear Poisson Problem

1. The easiest way to obtain a starting guess for the solution of Equation 50.16 with Dirichlet boundary conditions (50.17) is to neglect the left-hand side. That is, we use the local charge neutrality condition (50.11) for each node, which we can solve explicitly for ψ_0 via (50.12). With this function, we determine the components (nodal values) of our starting guess for the nonlinear Poisson problem $\boldsymbol{\psi_0}$ via

$$\boldsymbol{\psi_0} = (\psi_0(x_k))_{k=1}^N.$$

2. Solving the discrete nonlinear Poisson problem via Newton's method with the starting guess $\boldsymbol{\psi_0}$ yields the built-in potential vector solution $\boldsymbol{\psi}_{\text{eq}}$.

This solution procedure for the nonlinear Poisson equation for a p-i-n structure is illustrated in Figure 50.2. In the following, we address how to discretize the continuity equations. A task which turns out to be nontrivial.

50.2.5 Discretizing the Continuity Equations Using the Scharfetter–Gummel Scheme

Now, we turn our attention to the continuity equations (50.2) and (50.3). We proceed just as for Poisson's equation: we integrate over the control volumes $\omega_k = [x_{k-1,k}, x_{k,k+1}]$, apply the fundamental theorem of calculus for the left-hand sides and approximate the integral of the recombination term as before. For the interior control volumes we obtain

$$\begin{aligned} 0 &= j_{n;k,k+1} - j_{n;k-1,k} - qR(\psi_k, \varphi_{n;k}, \varphi_{p;k})|\omega_k|, \\ 0 &= j_{p;k,k+1} - j_{p;k-1,k} + qR(\psi_k, \varphi_{n;k}, \varphi_{p;k})|\omega_k|, \end{aligned} \tag{50.18}$$

where $k = 2, \ldots, N-1$. Similar to before, these equations constitute a balance law in integral form which we can interpret as follows: the carrier densities within the control volume ω_k change either due to in- and outflow or by recombination.

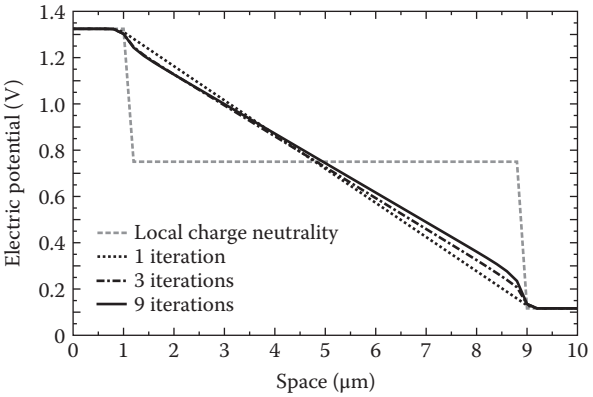


FIGURE 50.2 Convergence of the electric potential toward the thermodynamic equilibrium solution (built-in potential) for a GaAs p-i-n diode on an equidistant mesh with 51 nodes. The locally electroneutral initialization quickly converges to the built-in potential by means of a Newton iteration. The device has total length of 10 μm (1 μm n-doped with $N_D = 10^{16}$ cm^{-3}, 8 μm intrinsic, 1 μm p-doped with $N_A = 10^{17}$ cm^{-3}. Standard parameters for GaAs at $T = 300$ K: $E_v = 0$ eV, $E_c = 1.424$ eV, $N_v = 9.0 \times 10^{18}$ cm^{-3}, $N_c = 4.7 \times 10^{17}$ cm^{-3}, and $\varepsilon_r = 12.9$.

The fluxes (50.18) need to be expressed in terms of our set of unknowns, namely the nodal values of the electrostatic and the quasi-Fermi potentials. Deriving a suitable approximation for the current expressions $j_{n;k,k+1}$ and $j_{p;k,k+1}$ describing the carrier exchange between neighboring control volumes ω_k and ω_{k+1} is crucial for numerical simulations. Since standard discretization techniques such as central finite differences may lead to unphysical behavior (see Section 50.2.7), the derivation of such an approximation requires particular care.

For the Boltzmann approximation (50.5), a numerically stable scheme was invented in 1969 by Scharfetter and Gummel [11]. In fact, it has been discovered independently several times [18–20]. Its derivation is based on integrating the flux along the interval $[x_k, x_\ell]$ between two neighboring discretization nodes x_k and x_ℓ under the assumption of constant current density and constant electric field. That is, either $\ell = k-1$ or $\ell = k+1$. We illustrate the approach for the electron current density j_n.

In order to obtain a flux approximation, we will make two important assumptions, namely

1. The electric field along each edge is constant. This implies that the derivative of the electric potential $\frac{\mathrm{d}\psi}{\mathrm{d}x}$ along the edge $[x_k, x_\ell]$ is approximated by $\frac{\psi_\ell - \psi_k}{h_{k,\ell}}$, where again $\psi_\ell = \psi(x_\ell)$ and $\psi_k = \psi(x_k)$. This is consistent with the central difference approximation for the electric field in the Poisson equation, see (50.15).
2. The current density j_n is assumed to be constant along the edge.

The second assumption means that the current fluxes through the boundary from one control volume to another are conserved, which implies that the derivative of the flux along the interval $[x_k, x_\ell]$ vanishes. This means that we assume that there is no recombination along the edge between neighboring nodes. Hence, we obtain for the electron flux (the hole flux follows analogously)

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = \frac{\mathrm{d}}{\mathrm{d}x}\left(-q\mu_n N_c \exp\left(\frac{q(\psi(x) - \varphi_n(x)) - E_c}{k_\mathrm{B} T}\right)\frac{\mathrm{d}}{\mathrm{d}x}\varphi_n(x)\right) = 0. \tag{50.19}$$

We supply the differential equation with the boundary conditions

$$\varphi_n(x_k) = \varphi_{n;k} \qquad \text{and} \qquad \varphi_n(x_\ell) = \varphi_{n;\ell}. \tag{50.20}$$

The values $\varphi_{n;k}$ and $\varphi_{n;\ell}$ have to be determined. But for now we assume they are given. Equations 50.19 and 50.20 constitute a two-point boundary value problem. Upon integrating once and multiplication by the integrating factor

$$M(x) = \exp\left(-\frac{\psi_\ell - \psi_k}{h_{k,\ell} U_\mathrm{T}} x\right),$$

we obtain for the integration constant $j_{n;k,\ell}$ the equation

$$j_{n;k,\ell} M(x) = q\mu_n N_c U_\mathrm{T} \frac{\mathrm{d}}{\mathrm{d}x}\left(M(x)\exp\left(\frac{q(\psi(x) - \varphi_n(x)) - E_c}{k_\mathrm{B} \mathrm{T}}\right)\right).$$

Integrating once more from x_k to x_ℓ and using the boundary conditons (50.20) yields after some algebraic manipulations the Scharfetter–Gummel expression for the current flux along the edge $[x_k, x_\ell]$ between

neighboring control volumes

$$j_{n;k,\ell} = j_{n;k,\ell}(\psi_k, \psi_\ell, \varphi_{n;k}, \varphi_{n;\ell}) = -\frac{q\mu_n U_T}{h_{k,\ell}} \left[B\left(-\frac{\psi_\ell - \psi_k}{U_T}\right) N_c \exp\left(\frac{q\left(\psi_k - \varphi_{n;k}\right) - E_c}{k_B T}\right) - B\left(\frac{\psi_\ell - \psi_k}{U_T}\right) N_c \exp\left(\frac{q\left(\psi_\ell - \varphi_{n;\ell}\right) - E_c}{k_B T}\right)\right], \tag{50.21}$$

where we have introduced the *Bernoulli function*

$$B(x) = \frac{x}{\exp(x) - 1}.$$

This function smoothly interpolates between zero and $-x$, see Figure 50.3. Remembering the relationships (50.5), the flux can be brought into the simpler form

$$j_{n;k,\ell} = -\frac{q\mu_n U_T}{h_{k,\ell}} \left[B\left(-\frac{\psi_\ell - \psi_k}{U_T}\right) n_k - B\left(\frac{\psi_\ell - \psi_k}{U_T}\right) n_\ell \right]. \tag{50.22}$$

An important property of this scheme is its consistency with the thermodynamic equilibrium. That is, similarly to (50.9), we have

$$j_{n;k,\ell} = 0 \quad \text{for} \quad \varphi_{n;k} = \varphi_{n;\ell} = \text{const.} \tag{50.23}$$

For vanishing electric fields corresponding to $\delta\psi_{k,\ell} := \psi_\ell - \psi_k = 0$, the flux $j_{n;k,\ell}$ reduces to a purely diffusive flux

$$j_{n;k,\ell} = -q\left(-D_n \frac{n_\ell - n_k}{h_{k,\ell}}\right).$$

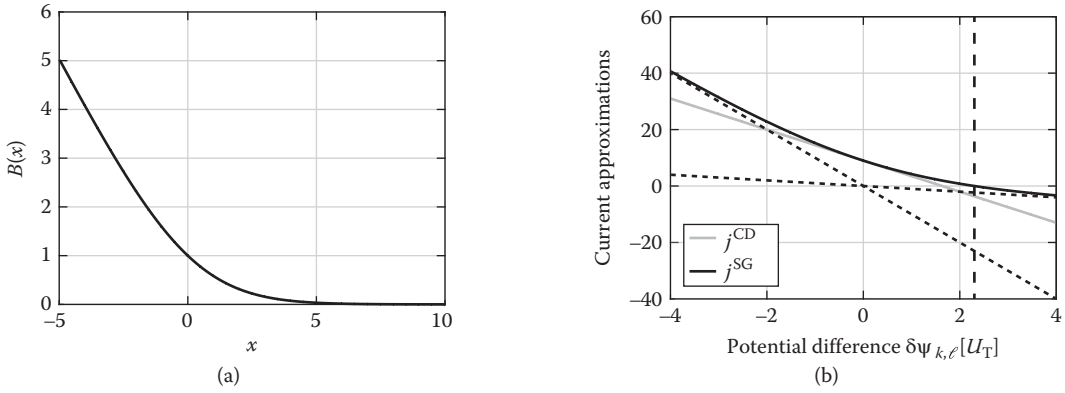


FIGURE 50.3 On the left the Bernoulli function $B(x)$ is shown. The picture on the right shows different current approximations for $q\mu_n U_T/h_{k,\ell} = 1$, $n_\ell = 10$, and $n_k = 1$ in terms of the potential difference $\delta\psi_{k,\ell}$. The dashed black line denotes the equilibrium potential difference. The dotted lines show the drift currents, i.e., the asymptotics of the Scharfetter–Gummel flux (50.24) for large positive and for large negative arguments. The central-difference flux is denoted with j^{CD} and the Scharfetter–Gummel flux with j^{SG}.

The diffusion constant D_n is defined via the so-called *Einstein relation* $D_n = \mu_n U_T$, which we discuss in Section 50.3.5 in more detail. For large electric fields, we obtain asymptotically a drift flux

$$j_{n;k,\ell} = \begin{cases} -q\mu_n n_k \frac{\psi_\ell - \psi_k}{h_{k,\ell}}, & \text{for } \delta\psi_{k,\ell} \gg +U_T, \\ -q\mu_n n_\ell \frac{\psi_\ell - \psi_k}{h_{k,\ell}}, & \text{for } \delta\psi_{k,\ell} \ll -U_T. \end{cases} \tag{50.24}$$

These two extreme cases show that the Scharfetter–Gummel scheme naturally encompasses drift and diffusive currents depending on the strength of the local electric field, see Figure 50.3.

Using analogous considerations, we can derive a current expression for the holes

$$j_{p;k,\ell} = \frac{q\mu_p U_T}{h_{k,\ell}} \left[B\left(\frac{\psi_\ell - \psi_k}{U_T} \right) p_k - B\left(-\frac{\psi_\ell - \psi_k}{U_T} \right) p_\ell \right].$$

Since we have an approximation for the fluxes at the control volume boundary, we can derive from the discrete system (50.18) nonlinear equations for the interior nodes in terms of our set of unknowns $(\psi, \varphi_n, \varphi_p)$.

We stress once more that the previous discussion is only valid for the Boltzmann approximation made in (50.5). A discussion on more general distribution functions follows in Sections 50.3.5 and 50.4.2.

50.2.6 Solving the Full Discretized van Roosbroeck System

So far we have described how to derive a set of discrete and nonlinear equations from the 1D van Roosbroeck system ((50.1)–(50.3)), namely

$$\begin{aligned} 0 &= -\varepsilon_s \left(\frac{1}{h_{k,k+1}} \psi_{k+1} - \left[\frac{1}{h_{k,k+1}} + \frac{1}{h_{k-1,k}} \right] \psi_k + \frac{1}{h_{k-1,k}} \psi_{k-1} \right) \\ &\qquad - q\Big(C_k + p(\psi_k, \varphi_{p;k}) - n(\psi_k, \varphi_{n;k}) \Big) |\omega_k|, \\ 0 &= j_{n;k,k+1}(\psi_k, \psi_{k+1}, \varphi_{n;k}, \varphi_{n;k+1}) - j_{n;k-1,k}(\psi_{k-1}, \psi_k, \varphi_{n;k-1}, \varphi_{n;k}) - qR(\psi_k, \varphi_{n;k}, \varphi_{p;k}) |\omega_k|, \\ 0 &= j_{p;k,k+1}(\psi_k, \psi_{k+1}, \varphi_{p;k}, \varphi_{p;k+1}) - j_{p;k-1,k}(\psi_{k-1}, \psi_k, \varphi_{p;k-1}, \varphi_{p;k}) + qR(\psi_k, \varphi_{n;k}, \varphi_{p;k}) |\omega_k|, \end{aligned} \tag{50.25}$$

with $k = 2, \ldots, N-1$. For the boundary nodes x_1 and x_N, the equations are given by the boundary conditions ((50.6)–(50.8)):

$$\begin{aligned} \psi_1 &= \psi(0) = \psi_0(0) + U_1, & \varphi_{n,1} &= \varphi_n(0) = U_1, & \varphi_{p,1} &= \varphi_p(0) = U_1, \\ \psi_N &= \psi(L) = \psi_0(L) + U_2, & \varphi_{n,N} &= \varphi_n(L) = U_2, & \varphi_{p,N} &= \varphi_p(L) = U_2. \end{aligned}$$

Just as for the Poisson problem we assume there are N control volumes in total. We end up with a system of $3N$ variables which we can summarize by the nonlinear discrete system

$$\mathbf{0} = \mathbf{F}(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) := \begin{pmatrix} \mathbf{F}_1(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_2(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_3(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \end{pmatrix},$$

where

$$\boldsymbol{\psi} = (\psi(x_k))_{k=1}^N, \qquad \boldsymbol{\varphi}_n = (\varphi_n(x_k))_{k=1}^N, \qquad \boldsymbol{\varphi}_p = (\varphi_p(x_k))_{k=1}^N.$$

Again, we employ Newton's method to solve this system. We aim to guarantee an appropriate starting guess by starting from thermodynamic equilibrium where the three equations reduce to one. The solution of the full van Roosbroeck system with an applied bias is then obtained via an iteration technique that we state here.

50.2.6.1 Solution Procedure: Full van Roosbroeck System

1. We assume thermodynamic equilibrium and set the constant quasi-Fermi potentials to $\varphi_0 = 0$. With the solution procedure introduced in Section 50.2.4, we obtain a solution vector $\boldsymbol{\psi}_{\text{eq}}$ for the nonlinear Poisson problem with Dirichlet boundary conditions.
2. We denote with $\boldsymbol{\varphi}_n^{\mathbf{0}}$ and $\boldsymbol{\varphi}_p^{\mathbf{0}}$ the constant vectors, consisting of the equilibrium value for the quasi-Fermi potentials, namely

$$\boldsymbol{\varphi}_n^{\mathbf{0}} = (\varphi_0)_{i=1}^N \qquad \text{and} \qquad \boldsymbol{\varphi}_p^{\mathbf{0}} = (\varphi_0)_{i=1}^N.$$

   We choose zero boundary voltages $U_1 = U_2 = 0$ and remember that we have obtained $\psi_0(0)$ and $\psi_0(L)$ via the charge neutrality condition. Thus we have determined the boundary conditions ((50.6)–(50.8)). Using the equilibrium vector $(\boldsymbol{\psi}_{\text{eq}}, \boldsymbol{\varphi}_n^{\mathbf{0}}, \boldsymbol{\varphi}_p^{\mathbf{0}})$ as a new starting guess, we can now solve the full discrete system (50.25) with these boundary conditions via Newton's method, yielding a solution $(\boldsymbol{\psi}^{\mathbf{1}}, \boldsymbol{\varphi}_n^{\mathbf{1}}, \boldsymbol{\varphi}_p^{\mathbf{1}})$. That is, we compute the equilibrium state of the full van Roosbroeck system using the equilibrium of the Poisson equation as a starting guess. We point out that solution and starting guess should agree up to machine precision by construction.
3. We slightly increase U_1 and U_2 to generate a "small" bias $\delta U = U_2 - U_1$. What small here means is not very precise and needs to be determined via trial and error. Now we solve the full discrete system (50.25) via Newton's method with the starting guess $(\boldsymbol{\psi}^{\mathbf{1}}, \boldsymbol{\varphi}_n^{\mathbf{1}}, \boldsymbol{\varphi}_p^{\mathbf{1}})$ to obtain a new solution $(\boldsymbol{\psi}^{\mathbf{2}}, \boldsymbol{\varphi}_n^{\mathbf{2}}, \boldsymbol{\varphi}_p^{\mathbf{2}})$.
4. Now we iterate the last step: gradually increase the bias and use the old solution as starting guess for the new one.

Figures 50.4 and 50.5 show the numerical results for the p-i-n structure, already introduced in Section 50.2.4. They depict the band edge energies, the quasi-Fermi energy levels $E_{F_n} = -q\varphi_n$ and $E_{F_p} = -q\varphi_p$ as well as the carrier densities for different bias values. Additionally, the current voltage characteristics for different n-dopings are shown.

50.2.7 Why Not Use Central Finite Differences for the Fluxes?

Thus far, we have discussed the Scharfetter–Gummel scheme to approximate the flux at the control volume interface. One might ask why not use a simpler flux approximation, for example, central finite differences. This question we address now.

For simplicity we assume a linear electrostatic potential. In this case

$$\frac{\mathrm{d}}{\mathrm{d}x}\psi = \psi(1) - \psi(0) =: \delta\psi.$$

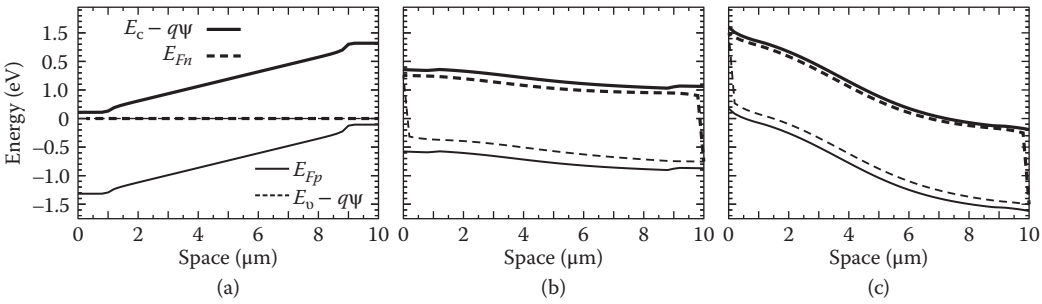


FIGURE 50.4 Computed band edge energy and quasi-Fermi levels for a GaAs p-i-n diode at different values of the applied voltage. (a) Thermodynamic equilibrium (off-state, 0V), (b) $U = 1.5$ V (flat band), (c) $U = 3.0$ V. Device has total length of 10 μm (1 μm n-doped with $N_D = 10^{16}$ cm^{-3}, 8 μm intrinsic, 1 μm p-doped with $N_A = 10^{17}$ cm^{-3}. The parameters used are standard parameters for GaAs as given in the caption of Figure 50.2 and $\mu_n = 8500\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$, $\mu_p = 400\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$, $\tau_n = \tau_p = 1$ ns, $C_n = C_p = 1 \times 10^{-29}\,\text{cm}^6\,\text{s}^{-1}$, and $r_{\text{spont}} = 1 \times 10^{-10}\,\text{cm}^3\,\text{s}^{-1}$.

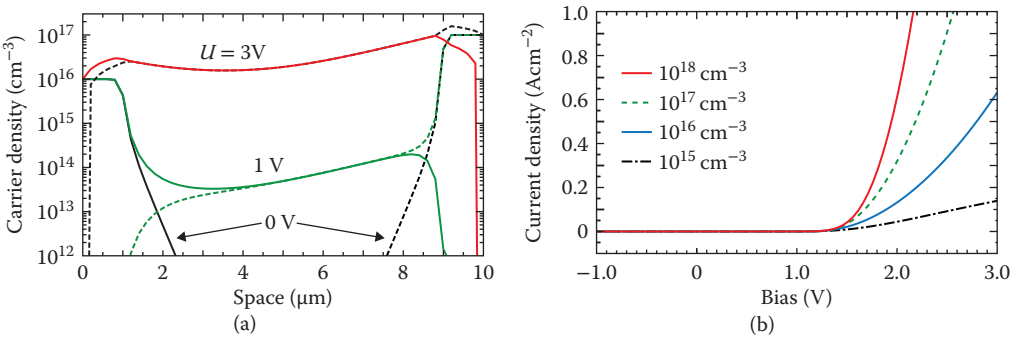


FIGURE 50.5 (a) Carrier densities in the p-i-n-diode at different applied voltages. Electron densities are shown as solid lines, holes are illustrated by dashed lines. One observes how the intrinsic region gets populated by free carriers while increasing the applied bias. (b) Current–voltage curves for the same device with altered values of the donor density N_D. Same parameters as in Figure 50.4.

We focus on the continuity equation for electrons on the domain $\Omega = [0, 1]$ which results in a linear differential equation of second order,

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = qD_n \frac{\mathrm{d}^2}{\mathrm{d}x^2} n - q\mu_n \delta\psi \frac{\mathrm{d}}{\mathrm{d}x} n = 0. \tag{50.26}$$

In this equation, we have introduced the drift-diffusion form of the (electron) flux

$$j_n = q\left(D_n \frac{\mathrm{d}}{\mathrm{d}x} n - n\mu_n \frac{\mathrm{d}}{\mathrm{d}x}\psi\right), \tag{50.27}$$

where we have used again the Einstein relation $D_n = \mu_n U_T$. We omit the physical derivation of the flux here. It is only important to know that this way of characterizing the flux is equivalent to Equation 50.4.

We have already seen how to solve this type of equation numerically by solving *local* two-point boundary value problems for the flux, which yields the Scharfetter–Gummel scheme (50.22). Of course, there are many different ways to discretize the flux (50.27) at a control volume interface $x_{k,\ell}$, needed in (50.18). For example, looking at Equation 50.27 one might consider replacing the derivative of the electron density $\frac{\mathrm{d}}{\mathrm{d}x} n$ with a central finite difference and the density n with an average. That is, at the control volume interface

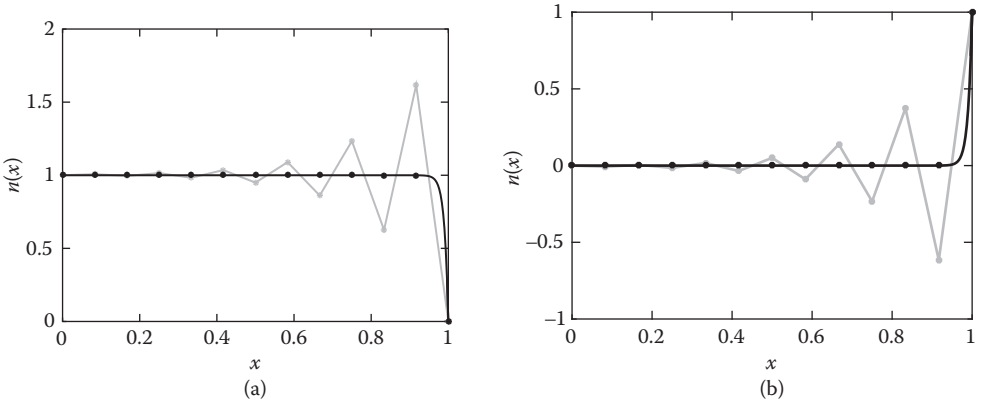


FIGURE 50.6 The exact solution (black) of Equation 50.26 and its central difference approximation (gray) on a uniform mesh with mesh width $h = 1/12$. The boundary conditions (50.28) lead to a numerical solution which violates the maximum principle (left). On the other hand, the boundary conditions (50.29) produce a numerical solution which violates the positivity of the density (right). Here $qD_n = 0.01$ and $q\mu_n\delta\psi = 1$.

we make the approximation

$$j^{\mathrm{CD}}_{n;k,\ell} = qD_n \frac{n_\ell - n_k}{h_{k,\ell}} - q\mu_n \delta\psi \, \frac{n_\ell + n_k}{2} = -\frac{q\mu_n U_{\mathrm{T}}}{h_{k,\ell}} \left(n_k - n_\ell + \frac{\delta\psi_{k,\ell}}{U_{\mathrm{T}}} \, \frac{n_\ell + n_k}{2} \right).$$

This, however, is a very bad idea since the scheme will become unstable, resulting under certain conditions in large oscillations. To show this, we supply Equation 50.26 with two different types of boundary conditions:

$$n(0) = 1 \qquad \text{and} \qquad n(1) = 0 \tag{50.28}$$

as well as

$$n(0) = 0 \qquad \text{and} \qquad n(1) = 1. \tag{50.29}$$

Then, the discretization using central finite differences shows large oscillations for $\mu_n\delta\psi/D_n = 100$, which can be seen in Figure 50.6. For the first set of boundary conditions, we can see that the *maximum principle* is violated, i.e., the solution exceeds both boundary values [21,22]. For the second set of boundary conditions, the numerically computed density becomes negative. Both observations constitute huge violations of basic physical laws and are unacceptable in applications.

The problem here is that the diffusion constant D_n is small compared to the drift constant $\mu_n\delta\psi$. We point out, that in this special 1D case of vanishing recombination the Scharfetter–Gummel scheme, on the other hand, yields *exact* nodal values due to its construction. The Scharfetter–Gummel and the central difference flux are compared in Figure 50.3.

50.3 Van Roosbroeck System

In this section, we make several generalizations to the van Roosbroeck system ((50.1)–(50.3)). In particular, we make the model time dependent and allow higher spatial dimensions. Before stating the equations, we discuss the physical concepts we want to encode in the model.

As in the 1D case, the van Roosbroeck system consists of three equations: the Poisson equation and two continuity equations for the carrier densities. The Poisson equation describes the electric field $\mathbf{E} = -\nabla\psi$ which is generated by a scalar electric potential $\psi(\mathbf{x}, t)$ in the presence of a free charge carrier density. In a (doped) bipolar semiconductor device, this charge density has three ingredients: the density of free (negatively charged) electrons $n(\mathbf{x}, t)$ occupying the conduction band, the density of (positively charged) holes $p(\mathbf{x}, t)$ occupying the valence band and the density of ionized built-in dopants $C(\mathbf{x}) = N_{\mathrm{D}}^{+}(\mathbf{x}) - N_{\mathrm{A}}^{-}(\mathbf{x})$, where N_{D}^{+} denotes the density of singly ionized donor atoms and N_{A}^{-} is the density of singly ionized acceptor atoms.

The continuity equations (in differential form), on the other hand, model the flow of the charge carrier densities due to diffusion and drift governed by the self-consistent electric field, which is generated by the net charge density. Furthermore, the recombination and generation of electron–hole pairs influences the electron and hole densities.

50.3.1 The van Roosbroeck System of Equations

For a bounded spatial domain $\Omega \subset \mathbb{R}^d$ where $d \in \{1, 2, 3\}$, the van Roosbroeck system consists of three coupled nonlinear PDEs of the form

$$-\nabla \cdot \left(\varepsilon_s \nabla\psi\right) = q\left(p - n + C\right), \tag{50.30}$$

$$-q\partial_t n + \nabla \cdot \mathbf{j}_n = qR, \tag{50.31}$$

$$q\partial_t p + \nabla \cdot \mathbf{j}_p = -qR, \tag{50.32}$$

for $\mathbf{x} \in \Omega$ and $t \in [0, T]$. The current densities in Equations 50.31 and 50.32 are given by the usual expressions

$$\mathbf{j}_n = -q\mu_n n \nabla\varphi_n \qquad \text{and} \qquad \mathbf{j}_p = -q\mu_p p \nabla\varphi_p. \tag{50.33}$$

That is, the negative gradients of the quasi-Fermi potentials are the driving forces of the currents [14]. These relationships correspond directly to Equation 50.4.

Within the framework of effective mass approximation [2], the densities of free carriers in a solid are given by

$$n = N_c \mathcal{F}\left(\frac{q(\psi - \varphi_n) - E_{\mathrm{c}}}{k_{\mathrm{B}}T}\right) \qquad \text{and} \qquad p = N_v \mathcal{F}\left(\frac{E_v - q(\psi - \varphi_p)}{k_{\mathrm{B}}T}\right). \tag{50.34}$$

Equations 50.34 indicate that the electric potential effectively leads to a bending of the energy band edge levels and thus a nonlinear, self-consistent coupling to the carrier densities is achieved. We assume a globally constant temperature for both carrier species and the crystal lattice.

The function $\mathcal{F}$ describes the occupation of energy states in the semiconductor. Since it plays an important role in the following discussion, we will define and discuss it later in Section 50.3.5. We only point out here that Equations 50.34 is analogous to Equations 50.5 when choosing for $\mathcal{F}$ the exponential function, the so-called *Boltzmann approximation*.

50.3.2 Initial and Boundary Conditions

The system ((50.30)–(50.32)) must be supplemented with initial and boundary conditions. The initial conditions at time $t = 0$ are given by the initial distributions ψ^I, φ_n^I and φ_p^I, i.e.,

$$\psi(\mathbf{x}, 0) = \psi^I(\mathbf{x}), \qquad \varphi_n(\mathbf{x}, 0) = \varphi_n^I(\mathbf{x}), \qquad \varphi_p(\mathbf{x}, 0) = \varphi_p^I(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega.$$

Regarding the boundary conditions, we briefly discuss the most important case where the boundary of the domain Ω can be decomposed into Ohmic contacts, a gate contact and *artificial* interfaces, i.e.,

$$\partial\Omega = \left(\bigcup_{\alpha=1}^{N_O} \Gamma_{O,\alpha} \right) \cup \Gamma_G \cup \Gamma_A.$$

Semiconductor–metal interfaces, such as Ohmic contacts, are modeled by Dirichlet boundary conditions. For any Ohmic contact $\Gamma_{O,\alpha}$ with $\alpha = 1, \ldots, N_O$, we set for all $\mathbf{x} \in \Gamma_{O,\alpha}$ and $t \in [0, T]$

$$\psi(\mathbf{x}, t) = \psi_0(\mathbf{x}) + U_\alpha(t), \tag{50.35}$$

$$\varphi_n(\mathbf{x}, t) = U_\alpha(t), \tag{50.36}$$

$$\varphi_p(\mathbf{x}, t) = U_\alpha(t), \tag{50.37}$$

where U_α denotes the corresponding externally applied contact voltage. The value ψ_0 at the boundary is defined by local charge neutrality similar to (50.11), where the exponential is now replaced by $\mathcal{F}$:

$$0 = N_v \mathcal{F}\left(\frac{E_v - q\psi_0(\mathbf{x})}{k_B T} \right) - N_c \mathcal{F}\left(\frac{q\psi_0(\mathbf{x}) - E_c}{k_B T} \right) + C(\mathbf{x}).$$

We just remark that, in general, this equation yields no analytical solution, and therefore its solution needs to be obtained by a nonlinear solver. The boundary conditions for the more advanced nonlinear semiconductor–metal interfaces (Schottky contacts) can be found in [13,23].

Gate contacts are modeled by Robin boundary conditions for the electrostatic potential and homogeneous Neumann boundary conditions for the quasi-Fermi potentials

$$\begin{aligned} \varepsilon_s \nabla\psi(\mathbf{x}, t) \cdot \boldsymbol{\nu} + \tfrac{\varepsilon_{ox}}{d_{ox}}(\psi(\mathbf{x}, t) - U_G(t)) = 0, \\ \mathbf{j}_n(\mathbf{x}, t) \cdot \boldsymbol{\nu} = \mathbf{j}_p(\mathbf{x}, t) \cdot \boldsymbol{\nu} = 0, \end{aligned} \qquad \text{for all } \mathbf{x} \in \Gamma_G \text{ and } t \in [0, T], \tag{50.38}$$

where ε_{ox} and d_{ox} are the absolute dielectric permittivity and the thickness of the oxide, respectively. The function $U_G(t)$ is the applied voltage at the outside of the insulating gate oxide at Γ_G. Here $\boldsymbol{\nu}$ denotes the outer normal vector on the interface.

On the remaining (artificial) interfaces, one typically imposes homogeneous Neumann boundary conditions (natural boundary conditions), namely

$$\nabla\psi(\mathbf{x}, t) \cdot \boldsymbol{\nu} = \mathbf{j}_n(\mathbf{x}, t) \cdot \boldsymbol{\nu} = \mathbf{j}_p(\mathbf{x}, t) \cdot \boldsymbol{\nu} = 0 \qquad \text{for all } \mathbf{x} \in \Gamma_A \text{ and } t \in [0, T]. \tag{50.39}$$

50.3.3 Recombination Processes

The net recombination rate R on the right-hand side of Equations 50.31 and 50.32 describes the radiative and nonradiative generation or recombination of carriers due to thermal excitation and various scattering

effects. We assume that the recombination rate $R(n,p)$ is given by the sum of the most common processes, namely the Shockley–Read–Hall recombination R_{SRH}, the spontaneous radiative recombination R_{rad} and the Auger recombination R_{Auger}. All of these rates are of the form

$$R(n,p) = r(n,p)np\left(1 - \exp\left(\frac{q\varphi_n - q\varphi_p}{k_{\mathrm{B}}T}\right)\right), \tag{50.40}$$

where $r(n,p)$ is a model-dependent generation-recombination rate [2,13,24]. In Figure 50.7, one finds the definitions of these rates together with a schematic illustration of the corresponding processes. For Boltzmann statistics, Equation 50.40 is equivalent to the widely used $R(n,p) = r(n,p)(np - N_{\mathrm{intr}}^2)$, where N_{intr} is the intrinsic carrier density defined in (50.13).

50.3.4 Thermodynamic Equilibrium

The thermodynamic equilibrium is characterized by vanishing current fluxes $\mathbf{j}_n(\mathbf{x},t) = \mathbf{j}_p(\mathbf{x},t) = \mathbf{0}$. As a consequence, $R(n,p) = 0$ and the quasi-Fermi potentials assume constant values which due to (50.40) are equal:

$$\varphi_0 := \varphi_n = \varphi_p = \mathrm{const}.$$

Without loss of generality, we set $\varphi_0 = 0$. Therefore, the van Roosbroeck system of (50.30)–(50.32) reduces to the nonlinear Poisson equation

$$-\nabla\cdot\varepsilon_s\nabla\psi = q\left(N_v\mathcal{F}\left(\frac{E_v - q\psi}{k_{\mathrm{B}}T}\right) - N_c\mathcal{F}\left(\frac{q\psi - E_c}{k_{\mathrm{B}}T}\right) + C\right). \tag{50.41}$$

We supply it with Dirichlet boundary conditions (50.35) for zero voltages $U_\alpha = 0$ for $\alpha = 1,\ldots,N_O$ and homogeneus Neumann boundary conditions (50.39) elsewhere. The solution of (50.41) with these boundary conditions defines the *built-in potential* denoted with ψ_{eq}.

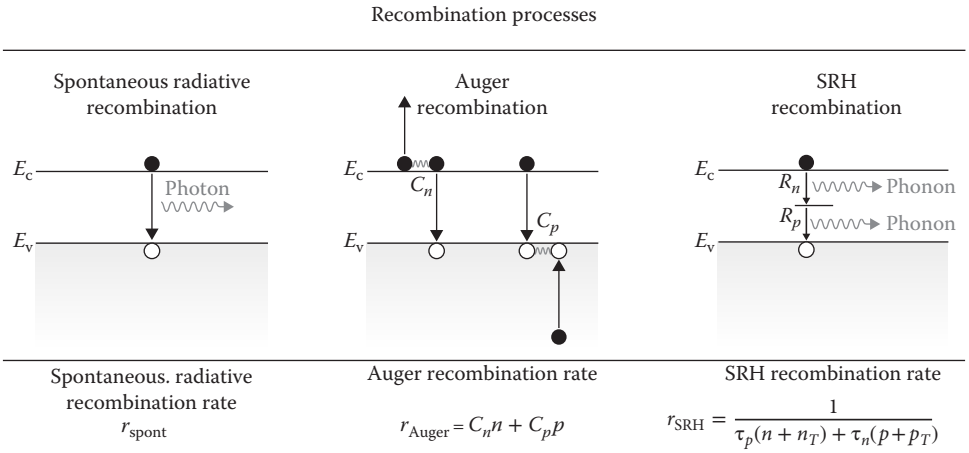


FIGURE 50.7 This figure illustrates three different recombination processes. The carrier life times τ_n, τ_p, the reference carrier densities n_T, p_T, and the coefficients $r_{\mathrm{spont}}, C_n, C_p$ are material-dependent parameters.

50.3.5 Non-Boltzmann Statistics and Generalized Einstein Relation

Now, we address the statistical distribution function $\mathcal{F}$ which appears in Equation 50.34. It describes the relationship between the density of the free carriers as well as the electrostatic potential and the quasi-Fermi potentials [2].

Assuming that the electrons in the conduction band are in quasi equilibrium, i.e., they are described by the quasi-Fermi level E_{F_n}, the electron density can be introduced as a convolution integral of the density of states DOS(E) with the Fermi–Dirac distribution function

$$n = \int_{-\infty}^{\infty} \mathrm{DOS}(E) \frac{1}{\exp\left(\frac{E-E_{F_n}}{k_\mathrm{B}T}\right)+1} \mathrm{d}E. \tag{50.42}$$

It is possible to express this convolution as a product of the effective density of states N_c and a nondimensionalized statistical distribution function $\mathcal{F}$, i.e.,

$$n(\eta) = N_c \mathcal{F}(\eta). \tag{50.43}$$

We consider organic and inorganic semiconductors. For inorganic, three-dimensional (3D) bulk semiconductors with parabolic bands the statistical distribution function is given by the Fermi–Dirac integral of order 1/2 [2], which can be approximated by the Blakemore [25] or the Boltzmann distribution in the low density limit. For organic semiconductors, it is given by the Gauss–Fermi integral [10], which reduces to a Boltzmann type distribution function (up to some normalizing factor) in the low-density limit or a Blakemore distribution function for vanishing variance σ. The latter corresponds to a δ-shaped density of states [10,26], describing a single electronic transport level. See Figure 50.8 for definitions and a relationship between these functions and Figure 50.9 for a graphical representation.

For inorganic, 3D bulk semiconductors assuming parabolic bands with effective mass m_e^* the effective density of states is given by

$$N_c = 2\left(\frac{m_e^* k_\mathrm{B} T}{2\pi\hbar^2}\right)^{3/2},$$

where T is the temperature and $\hbar$ the Planck constant. For organic semiconductors, the effective density of states N_c is given by the density of transport states N_t.

Introducing a general relationship between carrier density as well as quasi-Fermi and electrostatic potential via Equation 50.34 has quite a few implications. The nonlinear Poisson equation (50.41) becomes more complicated. In Section 50.2.4, we have studied how to design useful starting guesses for the Boltzmann approximation. For general distribution functions $\mathcal{F}$, it is no longer possible to obtain an explicit expression for ψ_0, like we have achieved in Equation 50.12. One rather needs to solve a nonlinear equation, for example, via Newton's methods at a given point $\mathbf{x}$. For highly doped regions, the Boltzmann approximation strongly overestimates the carrier densities if $\eta > 0$, see Figure 50.9. This implies that the more realistic Fermi–Dirac distribution leads to a higher built-in voltage. Moreover, from a numerical point of view, the Scharfetter–Gummel method needs to be adjusted. We discuss this in Section 50.4.2.

Finally, the distribution function $\mathcal{F}$ influences the ratio between diffusion and drift. To see this, we write the currents in *drift-diffusion* form. Using the relations for the carrier densities (50.34), we obtain

$$\mathbf{j}_n = -q\mu_n n\nabla\psi + qD_n\nabla n, \qquad \mathbf{j}_p = -q\mu_p p\nabla\psi - qD_p\nabla p.$$

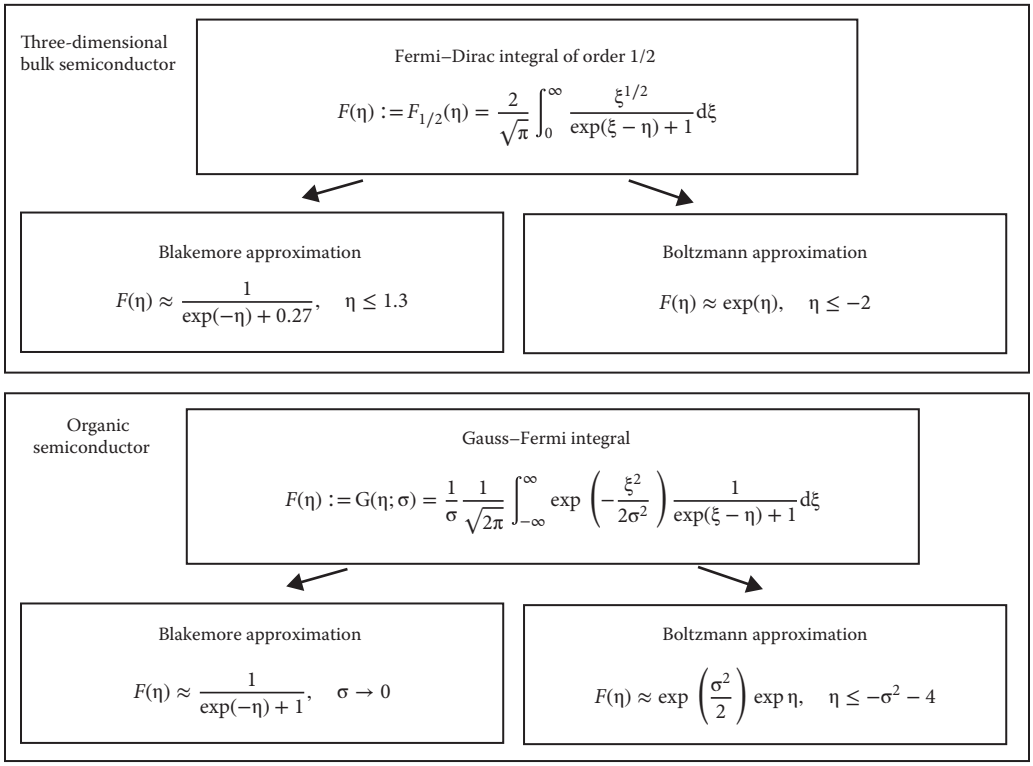


FIGURE 50.8 This figure shows two important choices of the statistical distribution function: for three-dimensional bulk semiconductors and for organic semiconductors [10] described by a Gaussian density of states with variance σ. Both can be approximated using the Blakemore or Boltzmann distribution function.

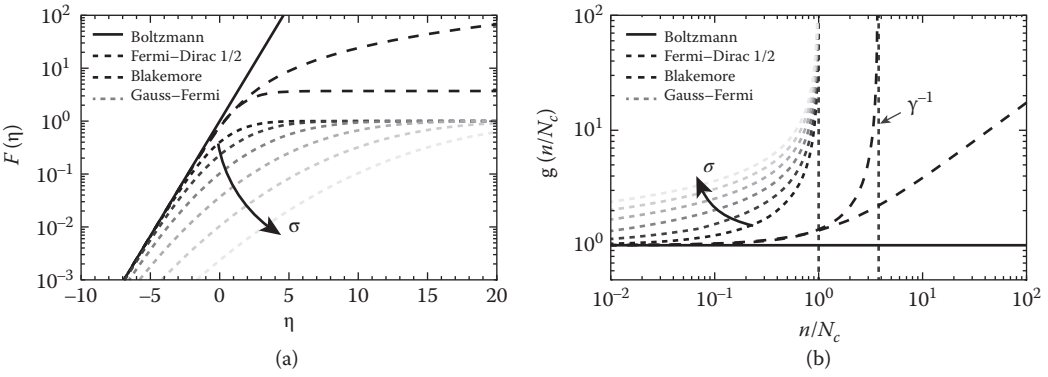


FIGURE 50.9 (a) Some frequently considered distribution functions for the density of free carriers. For the Blakemore and the Gauss–Fermi functions one obtains a saturation of the carrier density at high values of η. (b) Illustration of the diffusion enhancement factors for the distribution functions shown in (a). In the case of the Blakemore and the Gauss–Fermi functions the saturation values of the carrier density directly lead to a divergent behavior of the nonlinear diffusion factor.

The diffusion coefficients D_n and D_p are connected to the carrier mobilities by a *generalized* Einstein relation

$$\frac{D_n}{\mu_n} = \frac{k_B T}{q} g\left(\frac{n}{N_c}\right), \qquad \frac{D_p}{\mu_p} = \frac{k_B T}{q} g\left(\frac{p}{N_v}\right) \tag{50.44}$$

with a density-dependent nonlinear factor

$$g(\xi) = \xi \left(\mathcal{F}^{-1}\right)'(\xi),$$

leading in general to a nonlinear diffusion coefficient. For the Boltzmann approximation, we immediately see that $g(\xi) = 1$, which gives the classical Einstein relation $D_n = k_B T \mu_n / q$. For non-exponential distribution function and $\xi \geq 1$, however, we note $g(\xi) \geq 1$, see [27]. For this reason, we call this factor *diffusion enhancement* as proposed in [9]. The dependence of the diffusion enhancement g on the density is depicted in Figure 50.9 for various distribution functions.

50.3.6 Free Energy and Dissipation Rate

We consider consistency with fundamental principles of thermodynamics to be a quality measure for models describing natural processes. So it is of significant interest to study the consistency for both continuous and discretized models. Due to our simplifying assumption of constant temperature, we cannot have energy conservation in the proper sense. The second law of thermodynamics for nonequilibrium processes [16,28] requires nonnegativity of the local entropy production which, multiplied by the temperature T, gives the *dissipation rate*.

For the sake of readability, we discuss these concepts for Boltzmann statistics [29] here. See [30] and [31] for more general statistics. For a triple (ψ, n, p) and a thermodynamic equilibrium solution $(\psi_{eq}, n_{eq}, p_{eq}) := (\psi_{eq}, n(\psi_{eq}), p(\psi_{eq}))$ of the full van Roosbroeck system the *free energy* is defined as

$$\begin{aligned}\mathbb{F}(\psi, n, p) = &\int_\Omega \left(n \log \frac{n}{n_{eq}} - n + p \log \frac{p}{p_{eq}} - p + n_{eq} + p_{eq} \right) d\mathbf{x} \\ &+ \int_\Omega \frac{\varepsilon_s}{2} \left| \nabla \left(\psi - \psi_{eq}\right) \right|^2 d\mathbf{x} + \frac{1}{2} \int_{\Gamma_G} \frac{\varepsilon_{ox}}{d_{ox}} \left(\psi - \psi_{eq}\right)^2 ds.\end{aligned} \tag{50.45}$$

For a transient solution $(\psi(t), n(t), p(t))$ of ((50.30)–(50.32)), the function $\mathbb{L}(t) = \mathbb{F}(\psi(t), n(t), p(t))$ decays exponentially as t tends to infinity and one has

$$\mathbb{L}(t) = \mathbb{L}(0) - \int_0^t \mathbb{D}(\tau) d\tau,$$

where $\mathbb{D}(t)$ is the nonnegative dissipation rate [29]

$$\mathbb{D}(t) = \int_\Omega \left(n\mu_n |\nabla \varphi_n|^2 + p\mu_p |\nabla \varphi_p|^2 + r(n,p)(np-1)\log(np) \right) d\mathbf{x} \geq 0 \tag{50.46}$$

depending on t via the time evolution of n and p. This result confirms the consistency with the second law of thermodynamics. Incidentally, the function $\mathbb{L}$ is a Lyapunov function, allowing in certain situations to prove the global stability of the thermodynamical equilibrium. Furthermore, it can be used as a tool to prove uniqueness of solutions to the system ((50.30)–(50.32)). This leads us naturally to the following subsection.

50.3.7 Existence and Uniqueness Results

When modeling complex physical phenomena, it is often necessary to simplify the underlying physical model. Therefore, a sound mathematical investigation is necessary to assess the implications of such simplifications. Take, for example, the electrostatic potential of a device. Physically, we know that such a potential exists. Hence, a sound model should guarantee its existence—also from a mathematical point of view. If the existence cannot be shown, then the model does not represent the physical world accurately. Apart from existence, uniqueness of the solution is often desirable as well as continuous dependence on the initial data.

The mathematical technique used to prove existence and uniqueness depends on the device geometry and the model. The first existence result on the van Roosbroeck system was shown by Mock [32]. Since then several results have been obtained by Gajewski and Gröger, we refer the interested reader to [33,34]. In [34], the key tool to show the existence and uniqueness of the time-dependent system is based on finding a Lyapunov function. Moreover, Gajewski and Gröger presented the first result which considered Fermi–Dirac statistics instead of Boltzmann statistics. There are other important existence results. We would like to mention explicitly the results studied by Markowich see, for example, [35], Jüngel [36], and Jerome [37].

We would like to show the reader that these results are not of purely analytical interest but can also be used to design numerical methods. Gummel's method [38], for example (see Section 50.5), is based on the same fixed point iteration technique used to obtain the existence result which we discuss next. The connection to numerics can be found in [37]. We would like to give the reader an idea regarding the proof of a standard result without going into the tricky details. We follow ideas from Markowich's textbook [39].

We consider the stationary van Roosbroeck system with Boltzmann statistics. Using a scaling that can be found in [39, Section 2.4] and making a change of variables, we can rewrite the steady-state van Roosbroeck system in the following way:

$$-\lambda^2 \Delta \psi = \delta^2 e^{-\psi} v - \delta^2 e^{\psi} u + C, \tag{50.47}$$

$$\nabla \cdot (\mu_n e^{\psi} \nabla u) = \tilde{R}, \tag{50.48}$$

$$\nabla \cdot (\mu_p e^{-\psi} \nabla v) = \tilde{R}. \tag{50.49}$$

The densities n and p are related to the so-called *Slotboom variables* u and v via $n = \delta^2 e^{\psi} u$ and $p = \delta^2 e^{-\psi} v$. In Equations 50.47 through 50.49, the parameter λ denotes the normalized characteristic Debye length of the device, δ^2 indicates the scaled intrinsic carrier density and $\tilde{R}$ is the scaled recombination term. It is of the form $\tilde{R} = c(\psi, u, v)(uv - 1)$ with $c(\psi, u, v) > 0$. The chosen scaling and change of variables make it possible to rewrite the boundary conditions. We consider only Ohmic contacts and homogeneous Neumann boundaries.

In order to prove the existence of solutions for the system ((50.47)–(50.49)) equipped with suitable boundary conditions, we need to make several technical assumptions. We refer the interested reader to [39] where these assumptions are well explained as well as physically and mathematically justified. Here, we would rather like to focus on the key idea of the proof which is exploited numerically, omitting mathematical detail. The existence proof is based on an iteration scheme which considers the Poisson equation (50.47) decoupled from the continuity equations (50.48) and (50.49). The proof consists of the following steps:

1. We fix some $u_0, v_0 > 0$, and consider the semilinear elliptic problem

$$-\lambda^2 \Delta \psi = \delta^2 e^{-\psi} v_0 - \delta^2 e^{\psi} u_0 + C, \text{ in } \Omega, \qquad \text{(plus BC for } \psi \text{ on } \partial\Omega). \tag{50.50}$$

   Using standard analytical results for semilinear elliptic equations (namely the Leray–Schauder fixed point theorem and the maximum principle [40]) we prove that there exists a unique solution of the problem (50.50) that we denote with ψ_1.

2. We insert the solution ψ_1 of Equation 50.50 into the decoupled linear elliptic equations

$$\nabla \cdot (\mu_n e^{\psi_1} \nabla u) = c(\psi_1, u_0, v_0)(uv_0 - 1), \text{ in } \Omega, \qquad \text{(plus BC for } u \text{ on } \partial\Omega), \tag{50.51}$$

$$\nabla \cdot (\mu_p e^{-\psi_1} \nabla v) = c(\psi_1, u_0, v_0)(u_0 v - 1), \text{ in } \Omega, \qquad \text{(plus BC for } v \text{ on } \partial\Omega). \tag{50.52}$$

   We point out that in the right-hand side of Equation 50.51 the rate c depends only on ψ_1 determined from the previous step and on the fixed u_0, v_0. The other term, $uv_0 - 1$, depends linearly on the unknown u. Analogous considerations are valid for Equation 50.52. Physically, this means that in the factor describing relaxation to equilibrium, $uv - 1$, we freeze the hole density, more precisely the Slotboom variable v_0, in the continuity equation for the electrons and vice versa. Thanks to the nonnegativity of $c(\psi_1, u, v)$ there exist unique solutions u_1 and v_1 to Equations 50.51 and 50.52. This is ensured by standard results for linear elliptic equations (see, e.g., [40]).
3. Based on the first two steps, we can now define a map H which maps (u_0, v_0) onto (u_1, v_1). This map is known as *Gummel map*. It is possible to prove that this map has a fixed point (u^*, v^*) which determines a (weak) solution (ψ^*, u^*, v^*) of the coupled system ((50.47)–(50.49)). The electrostatic potential ψ^* that solves Poisson's equation (50.47) can be determined using the first step by substituting (u_0, v_0) with (u^*, v^*).

Finally, we discuss the uniqueness of solutions to the van Roosbroeck system. It is well known that for the steady-state system, the uniqueness of the solution cannot be shown without additional assumptions, for example, on the applied voltage. This is no surprise, as in fact, some semiconductor devices (e.g., thyristors [41]) are designed to have multiple steady states. However, for a device in thermal equilibrium (see Section 50.3.4), we have $u = v = 1$. In this case, there is a unique function which satisfies the Poisson equation (50.47). Hence, the solution $(\psi_e, 1, 1)$ is the unique equilibrium solution of the system ((50.47)–(50.49)), see [32]. For sufficiently small bias voltages, the uniqueness of the solution is shown in [39] under some specific assumptions on the recombination rates. All recombination mechanisms appearing in Figure 50.3.3 fulfill these assumptions. However, impact ionization rates, for example, are excluded.

50.3.8 Maximum Principle

An important mathematical tool, often used for proofs similar to the ones in the previous section, is the maximum principle. Intuitively, the maximum principle states that if the domain is bounded and the right-hand side of the elliptic equation is positive, then the maximum of the solution is attained on the boundary of the domain [40].

50.4 Discretizing the van Roosbroeck System

In this section, we introduce a method for discretizing the van Roosbroeck system which is close to the physicist's approach to derive PDEs based on a subdivision of the computational domain into *representative elementary volumes* or *control volumes*. The *two-point flux* finite volume method described here can be seen as a straightforward generalization, preserving the properties of the 1D Scharfetter–Gummel scheme in higher dimensions. Figure 50.10 shows a 2D simulation. The 2D variant of this approach was introduced as *box method* in [12]. Historically, it goes back to [42]. The 3D variant of this method was probably first investigated in [29,43].

The method has two main ingredients: a geometry-based approach to obtain a system describing communicating control volumes and a consistent description of the fluxes between two adjacent control volumes. These are discussed in the following two subsections. We finish this section by describing various properties of the finite volume scheme.

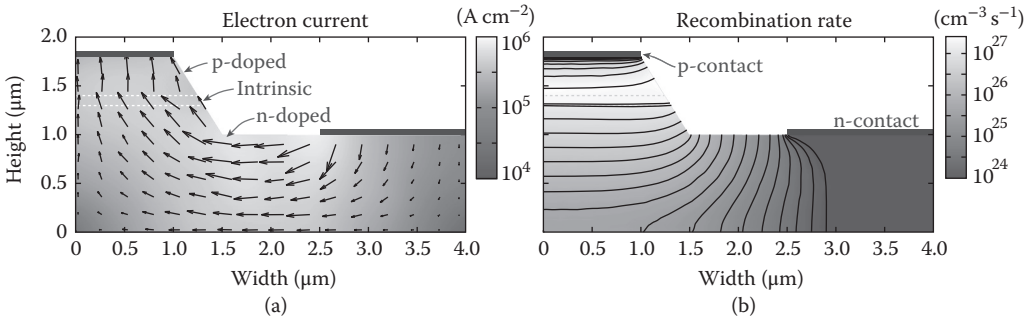


FIGURE 50.10 Exemplary results of a 2D simulation. The example is inspired by a ridge-waveguide laser with Ohmic contacts on the top of a mesa and at the side. For simplicity, we restrict to a homogeneous material which is here GaAs with the parameters as given in Figure 50.2. Within the mesa, an intrinsic domain is enclosed by a p-doped top-layer ($N_A = 2 \times 10^{18}$ cm^{-3}) and the n-doped substrate ($N_D = 2 \times 10^{18}$ cm^{-3}). The boundaries of the doped domains are shown in the pictures as dashed lines. (a) At an applied voltage of 1.5 V (flat band conditions) a significant electron current flow can be observed. The absolute value of the current density is shown by the gray scale, the arrows indicate the direction of particle flow (in the case of electrons the current density vector points in the opposite direction). (b) Plot of the total recombination rate at the same bias. The maximal recombination rate is observed in the vicinity of the intrinsic domain.

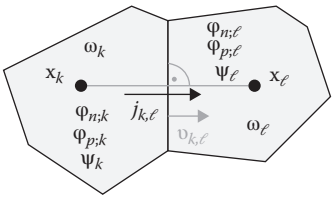


FIGURE 50.11 Two adjacent control volumes ω_k and ω_ℓ with corresponding data.

50.4.1 Finite Volume Method in Higher Dimensions

We partition the domain into N nonintersecting, convex polyhedral control volumes ω_k such that $\Omega = \bigcup_{k=1}^{N} \omega_k$. We associate with each control volume a node $\mathbf{x}_k \in \omega_k$. If the control volume intersects with the boundary of our domain, we demand that the node lies on the boundary, that means $\mathbf{x}_k \in \partial\Omega \cap \omega_k$. We assume that the partition is *admissible* [44], i.e., the *edge* $\overline{\mathbf{x}_k \mathbf{x}_\ell}$ with length $h_{k,\ell}$ is orthogonal to $\partial\omega_k \cap \partial\omega_\ell$. Thus, the normal vectors to $\partial\omega_k$ can be calculated by $\boldsymbol{\nu}_{k,\ell} = (\mathbf{x}_\ell - \mathbf{x}_k)/\|\mathbf{x}_\ell - \mathbf{x}_k\|$. See Figure 50.11 for details.

The set of all nodes and control volumes is called the *mesh* of the domain Ω. For a discussion of constructive ways to obtain such mesh partitions, see Section 50.6. In order to keep the equations simple, we introduce two abbreviations,

$$\eta_n\left(\psi, \varphi_n\right) = \frac{q\left(\psi - \varphi_n\right) - E_c}{k_B T} \quad \text{and} \quad \eta_p\left(\psi, \varphi_p\right) = \frac{E_v - q\left(\psi - \varphi_n\right)}{k_B T},$$

and assume that we have only a single gate contact with gate voltage U_G. We integrate the van Roosbroeck system ((50.30)–(50.32)) over ω_k and apply Gauss's divergence theorem, resulting in the integral equations

$$-\int_{\partial\omega_k} \varepsilon_s \nabla\psi \cdot \boldsymbol{\nu} \mathrm{d}s + \int_{\partial\omega_k \cap \Gamma_G} \frac{\varepsilon_{ox}}{d_{ox}}(\psi - U_G)\mathrm{d}s = q\int_{\omega_k} \left(C - N_c \mathcal{F}\left(\eta_n\left(\psi, \varphi_n\right)\right) + N_v \mathcal{F}\left(\eta_p\left(\psi, \varphi_p\right)\right)\right) \mathrm{d}\mathbf{x},$$

$$-q\frac{\partial}{\partial t}\int_{\omega_k} N_c\mathcal{F}\left(\eta_n\left(\psi,\varphi_n\right)\right)\mathrm{d}\mathbf{x}+\int_{\partial\omega_k}\mathbf{j}_n\cdot\boldsymbol{\nu}\mathrm{d}s=q\int_{\omega_k}R d\mathbf{x},$$

$$q\frac{\partial}{\partial t}\int_{\omega_k} N_v\mathcal{F}\left(\eta_p\left(\psi,\varphi_p\right)\right)\mathrm{d}\mathbf{x}+\int_{\partial\omega_k}\mathbf{j}_p\cdot\boldsymbol{\nu}\mathrm{d}s=-q\int_{\omega_k}R\mathrm{d}\mathbf{x},$$

for $k=1,\ldots,N$. Here, $\boldsymbol{\nu}$ is outward-pointing unit normal to the control volume ω_k; see Figure 50.11. These equations represent an integral form of the van Roosbroeck system on every control volume. In particular, the first equation is Gauss's law of electrodynamics. The other two equations are balance laws for the carrier densities. The densities in each control volume change only due to in- and outflow through the boundary or recombination.

Next, the surface integrals are split into the sum of integrals over the planar interfaces between the control volume ω_k and its neighbors. Employing one point quadrature rules for the surface and volume integrals, we deduce the finite volume scheme

$$\begin{aligned}\sum_{\omega_\ell\in\mathcal{N}(\omega_k)}|\partial\omega_k\cap\partial\omega_\ell|D_{k,\ell}&=q|\omega_k|\left(C_k-N_c\mathcal{F}\left(\eta_n\left(\psi_k,\varphi_{n;k}\right)\right)+N_v\mathcal{F}\left(\eta_p\left(\psi_k,\varphi_{p;k}\right)\right)\right)\\&\quad+|\partial\omega_k\cap\Gamma_G|\frac{\varepsilon_{\mathrm{ox}}}{d_{\mathrm{ox}}}(U_G-\psi_k),\end{aligned}\tag{50.53}$$

$$-q|\omega_k|N_c\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}\left(\eta_n\left(\psi_k,\varphi_{n;k}\right)\right)+\sum_{\omega_\ell\in\mathcal{N}(\omega_k)}|\partial\omega_k\cap\partial\omega_\ell|j_{n;k,\ell}=q|\omega_k|R_k,\tag{50.54}$$

$$q|\omega_k|N_v\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}\left(\eta_p\left(\psi_k,\varphi_{p;k}\right)\right)+\sum_{\omega_\ell\in\mathcal{N}(\omega_k)}|\partial\omega_k\cap\partial\omega_\ell|j_{p;k,\ell}=-q|\omega_k|R_k.\tag{50.55}$$

In the above mentioned formulas, $\mathcal{N}\left(\omega_k\right)$ denotes the set of all control volumes neigboring ω_k. In 2D, the measure $|\partial\omega_k\cap\partial\omega_\ell|$ corresponds to the length of the boundary line segment and in 3D to the area of the intersection of the boundary surfaces. The measure $|\omega_k|$ is in 2D given by the area and in 3D by the volume of the control volume ω_k. The unknowns ψ_k, $\varphi_{n;k}$, and $\varphi_{p;k}$ are approximations of the electric potential as well as the quasi-Fermi potentials for electrons and holes evaluated at node $\mathbf{x}_k$ like already introduced for the 1D case in Section 50.2. Accordingly, R_k is defined as

$$R_k=R\left(N_c\mathcal{F}\left(\eta_n\left(\psi_k,\varphi_{n;k}\right)\right),N_v\mathcal{F}\left(\eta_p\left(\psi_k,\varphi_{p;k}\right)\right)\right).$$

The doping is defined by the integral average

$$C_k=\frac{1}{|\omega_k|}\int_{\omega_k}C(\mathbf{x})\mathrm{d}\mathbf{x},$$

which can be estimated by its nodal value $C(\mathbf{x}_k)$. The numerical fluxes $D_{k,\ell}$, $j_{n;k,\ell}$ and $j_{p;k,\ell}$ approximate $-\varepsilon\nabla\psi\cdot\boldsymbol{\nu}_{k\ell}$, $\mathbf{j}_n\cdot\boldsymbol{\nu}_{k\ell}$ and $\mathbf{j}_p\cdot\boldsymbol{\nu}_{k\ell}$, respectively, on the interfaces between two adjacent control volumes ω_k and ω_ℓ, see Figure 50.11. As in the 1D case, we assume that these fluxes can be expressed as nonlinear functions depending on the values $\psi_k,\varphi_{n;k},\varphi_{p;k}$ and $\psi_\ell,\varphi_{n;\ell},\varphi_{p;\ell}$.

The electric displacement flux is approximated by

$$D_{k,\ell}=-\varepsilon_s\frac{\psi_\ell-\psi_k}{h_{k,\ell}},$$

where

$$h_{k,\ell} = \|\mathbf{x}_\ell - \mathbf{x}_k\|$$

is the edge length, compare with Equation 50.15.

For the Boltzmann approximation, the numerical charge carrier fluxes can be approximated via the Scharfetter–Gummel expression (50.21). In theory, any classical technique such as central differences or upwinding could be used to discretize the numerical fluxes. However, the former suffers from instability issues (as seen in Section 50.2.7) and neither scheme is consistent with the thermodynamic equilibrium. We will explain what this means in Section 50.4.3. In the next section, we discuss thermodynamically consistent schemes for non-Boltzmann statistics.

The finite volume scheme ((50.53)–(50.55)) yields a nonlinear system of $3N$ equations depending on $3N$ variables. We can directly substitute Dirichlet boundary values in the equations. In practice, however, this way of handling Dirichlet values is very technical to implement. Exploiting floating point arithmetic, the Dirichlet penalty method [45] provides a reasonable alternative. Physically, it replaces the Dirichlet boundary conditions by gate boundary conditions with very high oxide permittivity.

50.4.2 Scharfetter–Gummel Fluxes and Their Non-Boltzmann Generalizations

In 1D, we have already discussed the highly effective flux discretization scheme proposed by Scharfetter and Gummel for Boltzmann statistics [11]. They derive the numerical flux by locally solving a linear two-point boundary value problem. The 1D idea immediately carries over to higher dimensions if we insert the Scharfetter–Gummel (50.21) into the discrete system ((50.53)–(50.55)). However, if the Boltzmann approximation is not valid anymore, we can no longer derive the flux like Scharfetter and Gummel suggested.

This motivated the work in [46] where the Scharfetter–Gummel idea was generalized to a large class of nonlinear convection-diffusion problems, allowing to define consistent numerical fluxes from nonlinear two-point boundary value problems. Unfortunately, these *generalized Scharfetter–Gummel schemes* cannot be expressed by closed formulas. Sometimes, however, the local fluxes can be obtained iteratively [27]. It is also possible to approximate the two-point boundary value problems by simpler ones (e.g., by freezing some coefficients). This leads to *modified Scharfetter–Gummel schemes*. We address some of these schemes now.

For the sake of readability, we provide the formulas for electrons only. The formulas for holes follow similarly.

50.4.2.1 Generalized Scharfetter–Gummel Schemes

By the construction of our mesh, it suffices to study the 1D flux j_n along the edge $\overline{\mathbf{x}_k\mathbf{x}_\ell}$. For general distribution functions $\mathcal{F}$, we want to solve the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = \frac{\mathrm{d}}{\mathrm{d}x}\Big(- q\mu_n N_c \mathcal{F}\left(\eta_n\left(\psi, \varphi_n\right)\right) \nabla \varphi_n \Big) = 0$$

on the interval $[0, h_{k,\ell}]$ with boundary conditions

$$\varphi_n(0) = \varphi_{n;k} \quad \text{and} \quad \varphi_n\left(h_{k,\ell}\right) = \varphi_{n;\ell},$$

where $\varphi_{n;k}$ and $\varphi_{n,\ell}$ are the values of the quasi-Fermi potentials at nodes $\mathbf{x}_k$ and $\mathbf{x}_\ell$, respectively, see Figure 50.11. We assume now that the flux is constant between both nodes and denote it with $j_{n;k,\ell}$. Integrating

twice leads to an integral equation [27] for the unknown current, namely

$$\int_{\eta_{n;k}}^{\eta_{n;\ell}} \left(\frac{j_{n;k,l}/j_0}{\mathcal{F}(\eta)} + \frac{\psi_\ell - \psi_k}{U_T} \right)^{-1} d\eta = 1, \tag{50.56}$$

where $j_0 = q\mu_n N_c \dfrac{U_T}{h_{k,\ell}}$ and the limits are given by $\eta_{n;k} = \eta_n\left(\psi_k, \varphi_{n;k}\right)$ and $\eta_{n;\ell} = \eta_n\left(\psi_\ell, \varphi_{n;\ell}\right)$. For strictly monotonously increasing $\mathcal{F}(\eta)$ this equation has always a unique solution [47].

For the Boltzmann approximation $\mathcal{F}(\eta) = \exp(\eta)$, this integral equation can be solved analytically and yields the classical Scharfetter–Gummel expression for the flux (50.21). In [27], it was shown that for the Blakemore distribution function $\mathcal{F}(\eta) = \frac{1}{\exp(-\eta)+\gamma}$, the integral equation yields a fixed point equation

$$\hat{j}_{n;k,\ell} = B\left(\gamma\hat{j}_{n;k,\ell} + \frac{\psi_\ell - \psi_k}{U_T}\right) e^{\eta_{n;\ell}} - B\left(-\left[\gamma\hat{j}_{n;k,\ell} + \frac{\psi_\ell - \psi_k}{U_T}\right]\right) e^{\eta_{n;k}} \tag{50.57}$$

for the nondimensionalized edge current $\hat{j}_{n;k,\ell} = j_{n;k,l}/j_0$. The right-hand side is a Scharfetter–Gummel expression where the argument of the Bernoulli function is shifted by $\gamma\hat{j}_{n;k,\ell}$. Hence, this reduces to the classical Scharfetter–Gummel scheme for $\gamma = 0$.

Since the Bernoulli function is strictly decreasing, this fixed point equation possesses a unique solution $\hat{j}_{n;k,\ell}$. If we want to use the flux given by (50.57) in the discrete system ((50.53)–(50.55)) , we need to solve for the fluxes twice (once for electrons and once for holes) on each discretization edge $\overline{\mathbf{x}_k\mathbf{x}_\ell}$. A few Newton steps are sufficient to solve this equation iteratively.

Even though (50.57) is restricted to the Blakemore approximation, it provides a useful scheme in the context of organic semiconductors. There it arises naturally as a model for materials with δ-shaped density of states [10,26], describing a single electronic transport level, see Figures 50.8 and 50.9. This has been described in Section 50.3.5.

Unfortunately, for a general statistical distribution function, no corresponding equation has been derived so far. Therefore, in [47] it was proposed to use piecewise approximations of $\mathcal{F}$ by Blakemore type or rational approximations of Padé type in order to obtain piecewise integrable expression from the local boundary value problem.

50.4.2.2 Modified Scharfetter–Gummel Schemes

Bessemoulin-Chatard [48] derived a finite volume scheme for convection-diffusion problems by averaging the nonlinear diffusion term appropriately. This idea was generalized to more general distribution functions in [49], introducing a logarithmic average of the nonlinear diffusion enhancement

$$g_{n;k,\ell} = \frac{\eta_{n;k} - \eta_{n;\ell}}{\log \mathcal{F}\left(\eta_{n;k}\right) - \log \mathcal{F}\left(\eta_{n;\ell}\right)}$$

along the discretization edge. Using the generalized Einstein relation (50.44), one immediately observes that the diffusion enhancement g can be seen as a modification factor of the thermal voltage U_T. Replacing U_T in the Scharfetter–Gummel expression (50.21) by $U_T^* = U_T g_{n;k,\ell}$, we deduce the following modified Scharfetter–Gummel scheme

$$j_{n;k,\ell} = -\frac{q\mu_n U_T}{h_{k,\ell}} g_{n;k,\ell} \left(N_c \mathcal{F}\left(\eta_{n;k}\right) B\left(-\frac{\psi_\ell - \psi_k}{U_T g_{n;k,\ell}}\right) - N_c \mathcal{F}\left(\eta_{n;\ell}\right) B\left(\frac{\psi_\ell - \psi_k}{U_T g_{n;k,\ell}}\right)\right), \tag{50.58}$$

approximating the current along the edge.

Previously, we have replaced the thermal voltage by a suitable average along the edge. Now, we want to approximate $\mathcal{F}(\eta)$ along the edge by an exponential (Boltzmann approximation) and adjusting the effective density of states N_c accordingly, i.e.,

$$N_c \mathcal{F}(\eta) \approx N^*_{c;k,\ell} \exp(\eta).$$

This choice makes it possible to use the original Scharfetter–Gummel flux (50.21) by replacing N_c with $N^*_{c;k,\ell}$. One choice for the modified density of states is

$$N^*_{c;k,\ell}(\eta^*) = N_c \frac{\mathcal{F}(\eta^*)}{\exp(\eta^*)},$$

where $\eta^* \in [\eta_k, \eta_\ell]$, assuming $\eta_k \leq \eta_\ell$. In practice, we might consider taking the geometric average between $N_c(\eta_k)$ and $N_c(\eta_\ell)$, which leads to another modified Scharfetter–Gummel scheme:

$$j_{n;k,\ell} = -\frac{q\mu_n U_T}{h_{k,\ell}} N_c \sqrt{\frac{\mathcal{F}\left(\eta_{n;k}\right)\mathcal{F}\left(\eta_{n;\ell}\right)}{\exp(\eta_{n;k})\exp(\eta_{n;\ell})}} \left(B\left(-\frac{\psi_\ell - \psi_k}{U_T}\right) e^{\eta_{n;k}} - B\left(\frac{\psi_\ell - \psi_k}{U_T}\right) e^{\eta_{n;\ell}} \right). \quad (50.59)$$

The idea behind this scheme was introduced in [50] for the numerical solution of the generalized Nernst–Planck system which is similar to the van Roosbroeck system ((50.30)–(50.33)). A variant of this scheme for Fermi–Dirac statistics is described in [14,51] and numerically implemented in the semiconductor device simulation package WIAS-TeSCA [52].

50.4.3 Flux Expressions Consistent with the Thermodynamic Equilibrium

As introduced in Section 50.3.4, the currents of holes and electrons vanish in the thermodynamic equilibrium when no bias is applied. If a numerical scheme for zero bias boundary conditions results into vanishing numerical fluxes, we call it *consistent with the thermodynamical equilibrium*. In practice, one can examine this consistency by checking whether $j_{n;k,\ell} = j_{p;k,\ell} = 0$ if $\varphi_{n;k} = \varphi_{n;\ell}$ and $\varphi_{p;k} = \varphi_{p;\ell}$. Violating this property causes unphysical dissipation (spurious Joule heating) in the steady state attained for zero bias boundary conditions, which by definition is supposed to be the thermodynamic equilibrium [48].

All schemes introduced in Section 50.4.2 are consistent with the thermodynamical equilibrium. Indeed, let us assume that the quasi-Fermi potentials φ_n and φ_p between two adjacent control volumes ω_k and ω_ℓ are equal. The consistency of the classical Scharfetter–Gummel (50.21) and the scheme using a modified density of states (50.59) are obvious since the Bernoulli function satisfies $B(-x) = \exp(x)\, B(x)$.

Due to the definition from the solution of the two-point boundary value problem, the generalized Scharfetter–Gummel scheme defined in Equation 50.56 and its specialization for the Blakemore approximation (50.57) are consistent with the thermodynamic equilibrium. And finally, the logarithmic average of the diffusion enhancement is the only possible average which guarantees consistency with thermodynamics in the scheme (50.58). For details see [27,47,49].

50.4.4 Free Energy and Dissipation Rate

For Boltzmann statistics, it is rather straightforward to define discrete analoga of the free energy (50.45) and the dissipation rate (50.46). The positivity of the discrete dissipation rate was shown in [29], and the exponential decay of the free energy to its equilibrium value was proven in [53]. An overview of the entropy

method for finite volume schemes has been given in [54]. First results on more general statistics functions in this respect have been obtained in [55].

These pioneering works strongly indicate that the chosen discretization approach results in discrete models, which are consistent with the structural assumptions of nonequilibrium thermodynamics. A full account of these issues in the context of general statistics functions remains an open research topic.

50.4.5 Existence, Uniqueness, and Convergence

There are very few existence proofs for the solutions of the discretized system ((50.53)–(50.55)). For the Boltzmann approximation, Gärtner [56] proved that the discretized steady-state system has a solution, which becomes unique if a small bias is applied. A similar result for Fermi statistics has been obtained in [47].

A convergence theory for the finite volume scheme for the full discrete system ((50.53)–(50.55)) and general flux functions is still missing. However, practical experience and a number of results make its convergence plausible.

For example, in one space dimension, second-order convergence in the discrete maximum norm for the Scharfetter–Gummel scheme has been shown in [19]. Under the assumption that second derivatives of the continuous solution exist, in [57] for moderately sized drift terms and two-dimensional (2D), square grids, first-order convergence for the simple upwind scheme (see e.g., [21]), and second-order convergence for the exponential fitting scheme in the L_2-norm has been shown. Reinterpretations of the finite volume Scharfetter–Gummel scheme as a nonstandard finite element method allowed to obtain convergence estimates for Scharfetter–Gummel schemes on Delaunay grids (see Section 50.6) [58,59]. For a general approach to the convergence theory of finite volume schemes, see [44]. In [46], weak convergence (no order estimate) for a generalization of the Scharfetter–Gummel scheme to nonlinear convection-diffusion problems has been shown. A convergence proof for a variant of the van Roosbroeck system discretized with the simple upwind scheme was given in [60].

50.4.6 Maximum Principle

When applied to the drift-diffusion formulation, compared to various variants of the stabilized finite element method, the two-point flux finite volume scheme is outstanding in the sense that it guarantees positivity of densities and absense of unphysical oscillations [61,62]. It allows to carry over the discussions of Section 50.2.7 to the higher dimensional case.

50.5 Nonlinear Solvers

In Section 50.2, we have already briefly mentioned that we need to solve a nonlinear discrete system of the form

$$\mathbf{0} = \mathbf{F}(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) := \begin{pmatrix} \mathbf{F}_1(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_2(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_3(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \end{pmatrix}.$$

Here, we discuss two ways to do this: Newton's method and *Gummel's iteration method* [38], which is based on a blockwise decoupling of the discrete system.

50.5.1 Newton's Method

Assuming that a good starting guess has been provided, for example, by following the ideas in Section 50.2.6, Newton's method constructs a new $(k+1)$th iterate from the kth one by solving the linear system

$$\begin{pmatrix} \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\psi}} & \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\varphi}_n} & \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\varphi}_p} \\ \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\psi}} & \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\varphi}_n} & \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\varphi}_p} \\ \frac{\partial \mathbf{F}_3}{\partial \boldsymbol{\psi}} & \frac{\partial \mathbf{F}_3}{\partial \boldsymbol{\varphi}_n} & \frac{\partial \mathbf{F}_3}{\partial \boldsymbol{\varphi}_p} \end{pmatrix}^k \begin{pmatrix} \boldsymbol{\delta\psi}^k \\ \boldsymbol{\delta\varphi}_n^k \\ \boldsymbol{\delta\varphi}_p^k \end{pmatrix} = - \begin{pmatrix} \mathbf{F}_1(\boldsymbol{\psi}^k, \boldsymbol{\varphi}_n^k, \boldsymbol{\varphi}_p^k) \\ \mathbf{F}_2(\boldsymbol{\psi}^k, \boldsymbol{\varphi}_n^k, \boldsymbol{\varphi}_p^k) \\ \mathbf{F}_3(\boldsymbol{\psi}^k, \boldsymbol{\varphi}_n^k, \boldsymbol{\varphi}_p^k) \end{pmatrix} \tag{50.60}$$

for the update vector

$$\begin{pmatrix} \boldsymbol{\delta\psi}^k \\ \boldsymbol{\delta\varphi}_n^k \\ \boldsymbol{\delta\varphi}_p^k \end{pmatrix} := \begin{pmatrix} \boldsymbol{\psi}^{k+1} - \boldsymbol{\psi}^k \\ \boldsymbol{\varphi}_n^{k+1} - \boldsymbol{\varphi}_n^k \\ \boldsymbol{\varphi}_p^{k+1} - \boldsymbol{\varphi}_p^k \end{pmatrix}.$$

Then the new iterate is obtained by adding the previous vector to the update. The advantage of Newton's method is that it converges quadratically if the starting guess is sufficiently close to the solution [17]. This allows to obtain highly accurate discrete solutions at low additional cost. The major drawback is that it might converge very slowly or even fail to converge if the starting guess is too far from the actual solution. Damping—multiplying the update with a factor less than 1—is known to increase the convergence region. Another remedy is parameter embedding where one slowly changes a parameter, always using the old solution as a new starting guess. We have already employed this embedding technique in Section 50.2.6 when gradually increasing the bias voltage in the solution procedure.

An important advantage of finite element and finite volume discretizations is the fact that they create only next-neighbor couplings in the discretized systems. The resulting linearized systems are therefore sparse, i.e., the maximum number of nonzero elements in a matrix row is bounded by a constant independent of the number of discretization cells, making it possible to use highly economic storage schemes. For 2D applications, one can usually solve the resulting sparse linear system (50.60) via sparse direct solvers, such as PARDISO [63–66] or UMFPACK [67,68]. Direct solvers calculate a representation of the system matrix as a product of easily solvable lower and upper triangular matrices, which unfortunately are not anymore sparse, increasing memory consumption and computational time especially in large 3D applications. This fill-in phenomenon is avoided by preconditioned Krylov subspace methods [69], which compared to the direct solvers need significant effort in tuning and adaptation to the problem at hand.

50.5.2 Gummel's Method

Gummel [13,38] suggested decoupling the three equations in the van Roosbroeck system. He devised an iterative method at the continuous level, which for Boltzmann statistics and drift-diffusion form leads to alternating between solving linear differential equations for the electric potential as well as the charge carrier densities. Suppose, one already knows the iterate (ψ^k, n^k, p^k). In order to obtain the electric potential at the new level, ψ^{k+1}, one solves

$$0 = -\frac{1}{q}\nabla \cdot \left(\varepsilon_s \nabla \psi^{k+1}\right) + \left(n^k + p^k\right)\left(\frac{\psi^{k+1} - \psi^k}{U_\mathrm{T}}\right) + n^k - p^k - C.$$

This formulation is motivated by linearizing a fixed point problem that one obtains from the Poisson equation [13]. Once one has solved for the electrostatic potential at the new level, one can successively

solve the continuity equations from

$$0 = \nabla \cdot \mathbf{j}_n(\psi^{k+1}, n^{k+1}) - qR(\psi^{k+1}, n^{k+1}, p^k)$$
$$0 = \nabla \cdot \mathbf{j}_p(\psi^{k+1}, p^{k+1}) + qR(\psi^{k+1}, n^{k+1}, p^{k+1}).$$

Gummel's method is known to have a larger convergence region and a slower asymptotic convergence rate compared to Newtons's method. Depending on the software environment chosen, it also may be easier to implement. Note that standard existence and uniqueness proofs for the van Roosbroeck system rely on a similar decoupling strategy, see Section 50.3.7.

50.6 Mesh Generation for the Finite Volume Method

While appearing to be just a technical footnote compared to the questions of modeling, analysis and discretization, mesh generation in reality is a hard problem which deserves special attention. We focus on methods to construct admissible subdivisions of the computational domain Ω as used in Section 50.4.

50.6.1 Boundary Conforming Delaunay Meshes on Polyhedral Domains

Assume a partition (triangulation in 2D, tetrahedralization in 3D) of the polyhedral domain $\Omega = \bigcup_{k=1}^{N_\Sigma} \Sigma_k$ into nonoverlapping simplices Σ_k as it is commonly used for finite element methods [70]. We require that this simplicial partition of Ω has the *boundary conforming Delaunay property* [71]. For a triangulation of a 2D domain this property is equivalent to:

(1) For any two triangles with a common edge, the sum of their respective angles opposite to that edge is less or equal to 180°.
(2) For any triangle sharing an edge with $\partial\Omega$, its angle opposite to that edge is less or equal to 90°.

For a given vertex $x_i \in X$, where X is the set of vertices, the *Voronoi cell* around X is the set

$$V_i = \left\{ \mathbf{x} \in \mathbb{R}^d \,:\, \|\mathbf{x} - \mathbf{x}_i\| < \|\mathbf{x} - \mathbf{x}_j\| \text{ for all } \mathbf{x}_j \in X \text{ with } \mathbf{x}_j \neq \mathbf{x}_i \right\}.$$

The *Voronoi diagram*—the set of Voronoi cells for all vertices in X—is dual to the *Delaunay triangulation* of the point set X in the sense that for each edge $\overline{\mathbf{x}_i\mathbf{x}_j}$ in the Delaunay triangulation, $\partial V_i \cap \partial V_j \neq \emptyset$. If the simplicial partition is boundary conforming Delaunay, the restricted Voronoi cells $\omega_i = \Omega \cap V_i$ are well defined and can be obtained by joining the circumcenters of the simplices adjacent to the vertex $\mathbf{x}_i$. These restricted Voronoi cells provide an admissible control volume partition as required in Section 50.4. We note that in order to implement the finite volume method as described in Section 50.4, there is no need for an explicit construction of the control volumina ω_i as geometrical objects. Given the simplical partition, it is sufficient to base the calculations on the simplicial contributions $s_{ij}^k = |\partial\omega_i \cap \partial\omega_j \cap \Sigma_k|$ and $|\omega_i \cap \Sigma_K|$, and to use a simplex-based assembly loop as common for finite elements [29]. Figure 50.12 shows (boundary conforming) Delaunay triangulations.

There are several efficient algorithms to construct Delaunay triangulations for a given point set X [73,74]. These are good starting points of devising meshing algorithms in general. Many different problems complicate the generation of a mesh. For example, the boundary conforming Delaunay property is rather difficult to achieve, in particular it requires the careful insertion of additional points on the boundary. In 3D, slivers (very flat tetrahedra) must be avoided. And finally, it may be very complicated to fulfill additional requirements like constraints on the minimum angle or the local element size. Though there are still unsolved problems, the triangle ([72], 2D, free for noncommercial use) and TetGen ([75], 3D, open source) mesh generators help to create boundary conforming Delaunay meshes based on algorithms which are proven to

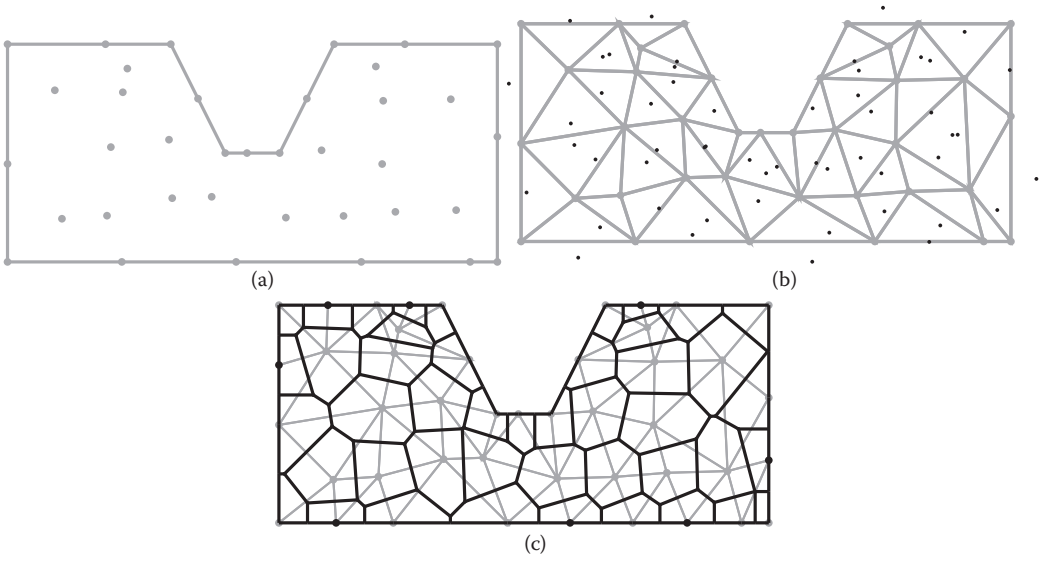


FIGURE 50.12 (a) Piecewise linear description of computational domain with given point cloud. (b) Delaunay triangulation of domain and triangle circumcenters. As some boundary triangles have angles larger than 90° angles opposite to the boundary, their circumcenters (small dots) are outside of the domain. (c) Boundary conforming Delaunay triangulation with automatically inserted additional points at the boundary (large dots), and Voronoi and restricted Voronoi cells (black). Created using triangle [72].

deliver meshes with the desired properties in finite time for a broad class of geometries given by a piecewise linear description of the boundary.

Other mesh generation approaches, in particular the advancing front [76] and the octree method [77], are similarly widespread. However, while popular in finite element community, their design makes it much harder to use them as a starting point for generating boundary conforming Delaunay meshes.

50.6.2 Tensor Product Approaches

In simplified geometrical situations, rectangular and cuboid mesh structures are a straightforward way to generate admissible finite volume partitions of the computational domain [12]. Extrusion of a 2D boundary conforming Delaunay base mesh into a 3D mesh consisting of prisms (and optional subsequent subdivision of these prisms into tetrahedra) provides another method to create admissible finite volume partitions [78].

50.6.3 Open Problems

Charge densities and potentials in semiconductors exhibit interior and boundary layers, i.e., thin regions charaterized by rapid gradients in one direction and slow gradients in the orthogonal direction. This may happen in space charge regions in the vicinity of gate contacts and at p-n junctions. These layers and other possible "hotspots" due to recombination, for example, call for local mesh adaptation. Element size control via the mesh generators using *a priori* knowledge is often used in this situation [72,75]. The development of reliable a posteriori mesh refinement criteria is an open problem in the context of the strongly coupled van Roosbroeck system. A physically motivated heuristic approach based on the equidistribution of the dissipation rate was suggested in [29] and investigated in [79].

Mesh generation tools mostly fail to create boundary conforming Delaunay meshes which use thin, anisotropic simplices to resolve interior and boundary layers with an optimal number of points. A partial

remedy in some situations can be provided by node offsetting [80], a technique still under development. In certain cases, tensor product approaches [78] give reasonable results.

50.7 Time Discretization

In order to discretize the time derivative in the van Roosbroeck system ((50.30)–(50.32)), we define a series of discrete time values $0 = t^0, t^1, \ldots, t^M = T$ with step lengths $\tau^m = t^{m+1} - t^m$. We describe the temporal semi-discretization of the (spatially) continuous problem. For practical purposes it needs to be combined with the space discretization approach described in Section 50.4. The implict Euler time discretization scheme assumes given values $\psi^m, \varphi_n^m, \varphi_p^m, n^m, p^m$ of the variables at time t^m. Values at the subsequent time t^{m+1} are calculated by solving the nonlinear system of equations

$$-\nabla \cdot (\varepsilon_s \nabla \psi^{m+1}) = q(C - n^{m+1} + p^{m+1}), \tag{50.61}$$

$$-q\frac{n^{m+1} - n^m}{\tau^m} - q\nabla \cdot \mu_n n^{m+1} \nabla \varphi_n^{m+1} = qR(n^{m+1}, p^{m+1}), \tag{50.62}$$

$$q\frac{p^{m+1} - p^m}{\tau^m} - q\nabla \cdot \mu_p p^{m+1} \nabla \varphi_p^{m+1} = -qR(n^{m+1}, p^{m+1}), \tag{50.63}$$

where

$$n^{m+1} = N_c \mathcal{F}(\eta_n(\psi^{m+1}, \varphi_n^{m+1})),$$
$$p^{m+1} = N_v \mathcal{F}(\eta_p(\psi^{m+1}, \varphi_p^{m+1})).$$

As a consequence, the nonlinear system ((50.61)–(50.63)), which is structurally similar to Equations 50.53 through 50.55, has to be solved for each time step by iterative methods described in Section 50.5. An initial guess for these methods can be obtained from the old time step t^m or a linear extrapolation involving data from several previous time steps [81].

It is advisable to choose the size of the time steps τ^m according to multiple criteria, including the convergence radius of Newton's method, the free energy, and local changes of the potentials.

At first glance, the computational cost of having to solve a nonlinear system for each time step appears to be rather high. However, this scheme is stable independent of the time step size and allows to carry over all advantages of the finite volume scheme from the stationary to the transient case. The transient discrete solution converges to the stationary one, in particular, for the corresponding boundary condition to the thermodynamic equilibrium solution. The free energy of the system decays during this approach to the equilibrium. In several cases, it has been proven that this decay is exponential [55,82]. By preserving the physics, the implicit Euler scheme fits well to the thermodynamically motivated discretization schemes. The use of other schemes implies giving up some of these physical properties. If the resulting deviations from thermodynamic properties can be kept under control, higher order schemes (e.g., backward differentiation formula (BDF) methods [83]) might help to reduce computational costs due to a higher temporal convergence order.

Due to the perceived complexity and high computational costs of solving coupled nonlinear PDE systems, linear implicit methods, mixing linearization, and time discretizations are alternatives. Most prominent is the scheme devised in [84].

50.8 Contact Terminal Currents

In simulations, one is usually interested in IV-curves, i.e., the dependency of terminal currents on applied voltages. Therefore, calculating terminal currents accurately is crucial to a successful postprocessing of the

simulated field data. The *total current density* is given by

$$\mathbf{j} = \mathbf{j}_d + \mathbf{j}_n + \mathbf{j}_p,$$

where $\mathbf{j}_d = -\varepsilon_s \nabla(\partial_t \psi) = \partial_t \mathbf{D}$ is the *displacement current density*. Taking the time derivative of the Poisson equation (50.30) yields

$$\begin{aligned} -\nabla \cdot \varepsilon_s \nabla(\partial_t \psi) = q(\partial_t n - \partial_t p) &= -\nabla \cdot \mathbf{j}_n - qR - \nabla \cdot \mathbf{j}_p + qR \\ \nabla \cdot \partial_t \mathbf{D} &= -\nabla \cdot \mathbf{j}_n - \nabla \cdot \mathbf{j}_p, \end{aligned}$$

resulting in $\nabla \cdot \mathbf{j} = 0$, which physically implies charge conservation.

Given a set of contacts $\Gamma_\alpha \subset \partial\Omega$ for $\alpha = 1, \ldots, N_C$, the terminal current through contact α is defined as

$$I_\alpha = \int_{\Gamma_\alpha} \mathbf{j} \cdot \boldsymbol{\nu} \, \mathrm{d}s.$$

The correct calculation of this integral presupposes a regularity of the solution and its derivative which is not supported by analytical theory. Therefore, in [51,52] a test function–based technique was proposed, which uses the weak formulation of the problem. For a physical motivation of this approach and an interpretation as a generalized Shockley–Ramo theorem we refer to [85]. For any α, let the function $T_\alpha : \Omega \to \mathbb{R}$ solve the boundary value problem

$$-\nabla^2 T_\alpha = 0 \qquad \text{in } \Omega, \tag{50.64}$$

$$\nabla T_\alpha \cdot \boldsymbol{\nu} = 0 \qquad \text{on } \partial\Omega \setminus \bigcup_{\beta=1}^{N_C} \Gamma_\beta, \tag{50.65}$$

$$T_\alpha = \delta_{\alpha\beta} \qquad \text{on } \Gamma_\beta \text{ (for } \beta = 1, \ldots, N_C), \tag{50.66}$$

where $\delta_{\alpha\beta}$ denotes the Kronecker delta. As a consequence,

$$\begin{aligned} I_\alpha = \int_{\Gamma_\alpha} \mathbf{j} \cdot \boldsymbol{\nu} \, \mathrm{d}s &= \int_{\partial\Omega} T_\alpha \mathbf{j} \cdot \boldsymbol{\nu} \, \mathrm{d}s = \int_\Omega \nabla \cdot (T_\alpha \mathbf{j}) \, \mathrm{d}\mathbf{x} \\ &= \int_\Omega \nabla T_\alpha \cdot \mathbf{j} \, \mathrm{d}\mathbf{x} + \int_\Omega T_\alpha \nabla \cdot \mathbf{j} \, \mathrm{d}\mathbf{x} \\ &= \int_\Omega \nabla T_\alpha \cdot (\mathbf{j}_n + \mathbf{j}_p) \, \mathrm{d}\mathbf{x} + \partial_t \int_\Omega \nabla T_\alpha \cdot \mathbf{D} \, \mathrm{d}\mathbf{x}. \end{aligned}$$

Given the semi-discretization in time from Section 50.7, one arrives at

$$I_\alpha^{m+1} = \int_\Omega \nabla T_\alpha \cdot (\mathbf{j}_n^{m+1} + \mathbf{j}_p^{m+1}) \, \mathrm{d}\mathbf{x} + \frac{1}{\tau^m} \int_\Omega \nabla T_\alpha \cdot \left(\mathbf{D}^{m+1} - \mathbf{D}^m \right) \, \mathrm{d}\mathbf{x}.$$

Like in Section 50.4, we approximate the system (50.64)–(50.66) by a finite volume discretization. Using partial integration and the notation introduced in Section 50.4 we come up with an expression for the total

current through the α-th contact of the form

$$I_\alpha^{m+1} \approx \sum_{\substack{\omega_\ell \in \mathcal{N}(\omega_k) \\ k<\ell}} |\partial\omega_k \cap \partial\omega_\ell| \left(j_{n;k,\ell}^{m+1} + j_{p;k,\ell}^{m+1} + \frac{D_{k,\ell}^{m+1} - D_{k,\ell}^{m}}{\tau^m} \right) (T_{\alpha;k} - T_{\alpha;\ell}). \tag{50.67}$$

50.9 An Alternative: The Finite Element Method

We have put special emphasis on the finite volume method. However, it is worth pointing out that there are other approaches. The finite difference method that has been used to solve the van Roosbroeck system [13] on tensor product meshes is equivalent to the previously introduced finite volume method.

The most popular ansatz to deal with unstructured meshes is the finite element method [70]. It starts with a *weak formulation* of the steady-state van Roosbroeck problem ((50.30)–(50.32)). Each equation of the system is multiplied with a so-called test function and integrated in space. Integrating by parts, each equation of the steady-state system can be restated as the following problem:

$$\text{find } u \in U \text{ such that } a(u, v) = f(v) \text{ for all } v \in V,$$

where U is a function space containing the ansatz functions u and V is a function space containing the test functions v. The map $a : U \times V \to \mathbb{R}$ has the structure

$$a(u, v) = \int_\Omega \alpha(u) \nabla u \cdot \nabla v \, \mathrm{d}\mathbf{x} - \int_{\partial\Omega} \nabla u \cdot \mathbf{n} \, v \, ds \tag{50.68}$$

and the functional $f : V \to \mathbb{R}$ is continuous. The idea of a finite element method is to approximate this continuous problem with a discrete version. We approximate the infinite dimensional spaces U and V with finite dimensional subspaces U_h and V_h spanned by basis functions with localized support. Then we obtain a finite dimensional version of Equation 50.68, namely

$$\text{find } u_h \in U_h \text{ such that } a_h(u_h, v_h) = f(v_h) \text{ for all } v_h \in V_h.$$

As it is sufficient to fulfill this condition for finitely many localized basis functions in V_h, we obtain a sparse (containing many zeros) nonlinear system of equations for the basis coefficients of u_h, assuming we are able to calculate the integrals. One usually approximates these integrals by quadrature rules.

Auf der Maur [86] suggested to write the stationary van Roosbroeck system in this framework, using the quasi-Fermi potential formulation of the flux. In this case one has

$$a^\psi(\psi, v) = \int_\Omega \varepsilon_s \nabla\psi \cdot \nabla v \, \mathrm{d}\mathbf{x} + \int_{\Gamma_G} \frac{\varepsilon_{\mathrm{ox}}}{d_{\mathrm{ox}}} (\psi - U_{\mathrm{G}}) v \, \mathrm{d}s, \qquad f^\psi(v) = q \int_\Omega (p - n + C) v \, \mathrm{d}\mathbf{x}, \tag{50.69}$$

$$a^{\varphi_n}(\varphi_n, v) = \int_\Omega \mu_n n \nabla\varphi_n \cdot \nabla v \, \mathrm{d}\mathbf{x}, \qquad f^{\varphi_n}(v) = \int_\Omega R v \, \mathrm{d}\mathbf{x}, \tag{50.70}$$

$$a^{\varphi_p}(\varphi_p, v) = \int_\Omega \mu_p p \nabla\varphi_p \cdot \nabla v \, \mathrm{d}\mathbf{x}, \qquad f^{\varphi_p}(v) = -\int_\Omega R v \, \mathrm{d}\mathbf{x}. \tag{50.71}$$

At the Dirichlet parts of the boundary, namely at the Ohmic contacts $\Gamma_{\mathrm{O},\alpha}$ introduced in Section 50.3, we assume that test functions v vanish and the ansatz functions denoted by ψ, φ_n, and φ_p fulfill the Dirchlet boundary conditions ((50.35)–(50.37)). Note, that the gate boundary conditions (50.38) have been incorparated into the boundary part of the quadratic form (50.69).

We point out that for Poisson's equation $\alpha^{\psi}(\psi) = \varepsilon_s$ and for the continuity equations we have $\alpha^{\varphi_n}(\varphi_n) = \mu_n n$ and $\alpha^{\varphi_p}(\varphi_p) = \mu_p p$, respectively. Due to the formulation of the problem with quasi-Fermi potentials as unknowns, this means that $\alpha^{\varphi_n}(\varphi_n)$ and $\alpha^{\varphi_p}(\varphi_p)$ introduce nonlinearities in a^{φ_n} and a^{φ_p}. The integrals on the left-hand side of Equations 50.69 through 50.71 have no analytical expressions and are approximated using quadrature rules [70].

Essential for this finite element scheme is the underlying potential-based formulation of the flux in terms of φ_n and φ_p. If one attempts to apply the finite element method using the density-based, drift-diffusion formulation of the flux, one encounters a lot of technical difficulties and one has to stabilize the finite element method [61,87]. A convergence proof of the finite element method for a similar system in a somewhat different context (and without recombination) is available in [88].

50.10 Extensions and Outlook

In this chapter, we discussed the most important numerical solution techniques of the van Roosbroeck system, describing the transport of electrons and holes in semiconductors in a self-consistent electric field. This approach provides the core functionalities for a solver, which can be enhanced by models describing additional physical phenomena and more complex device structures including, for example, heterostructures. We mention a number of additional extensions and point to the corresponding literature.

50.10.1 Additional Physical Models

Many devices require a more accurate description of the carrier mobilities. For example, models of ionized impurity scattering, high-field drift velocity saturation and similar effects [13,89] introduce dependencies of the mobilities on carrier densities and the electric field strength, resulting in even more nonlinear couplings in the system.

The carrier transport in organic semiconductors is governed by hopping transport between the energy states of neighboring molecules. The van Roosbroeck system can be used to describe this effectively [8,9]. This may require the use of statistical distribution functions that reflect the distribution of the energy transport levels such as Blakemore for δ-shaped densities of states or the Gauss–Fermi integral [10]. In this context, the correct treatmeant of the diffusion enhancement is vital. One also has to account for the nonlinear mobilities [8,90] related to the energetic disorder characteristic for organic semiconductors.

Spin-polarized drift-diffusion models have been proposed in [91] for the description of spintronic devices. They generalize the van Roosbroeck system, introducing spin-resolved densities for electron and holes and additional mechanisms describing the spin relaxation. They have been recently studied also from a numerical point of view in [92–94].

50.10.2 Coupling the van Roosbroeck System to Other Models

When modeling lasers one has to couple the van Roosbroeck system to equations for the optical modes. This introduces additional recombination processes, describing stimulated emission. Additionally, one has to consider balance equations for the photon number. A comprehensive model (also including heating effects) along with a numerical solution strategy is given in [14].

In semiconductor devices with embedded nanostructures (such as quantum dots and quantum wells), one has to couple the drift-diffusion equations to equations describing the dynamics of the carriers localized in the nanostructures [4,5,7]. A multiscale approach for coupling atomistic with continuum drift-diffusion models is presented in [95].

If heating effects become important, the van Roosbroeck system needs to be extended by an energy transport model for the heat flow in the device. For a thermodynamically consistent extension of the van Roosbroeck system to account for this effect we refer to [14,31].

50.10.3 Methods for Doping Optimization

An important task when designing semiconductor devices is finding a suitable doping profile. This is an analytical and numerical challenge. Electric properties of the device can be improved by optimizing the doping profiles using suitable objective functionals [96–98]. Recently, the approach introduced in [96] has been extented for the optimization of doping profiles in lasers [99].

50.10.4 Alternative Modeling Approaches for Carrier Transport

The van Roosbroeck system uses implicitly the assumption that the carrier ensemble is locally in quasi equilibrium. When this assumption is not met, for example, for hot electrons, one has to consider alternative approaches. The most common ones are hydrodynamic models [100] and approximations of the Boltzmann transport equation which are derived from a spherical harmonics expansion [101]. The book [102] gives a mathematically oriented overview of these topics.

Acknowledgments

The authors are grateful for innumerable discussions with K. Gärtner and H. Gajewski, who helped to shape their still incomplete knowledge on the topic. This work has been supported by ERC-2010-AdG no. 267802 "Analysis of Multiscale Systems Driven by Functionals" (N.R.), by the Deutsche Forschungsgemeinschaft DFG within CRC 787 "Semiconductor Nanophotonics" (T.K., N.R., M.K) and partially funded in the framework of the project "Macroscopic Modeling of Transport and Reaction Processes in Magnesium-Air-Batteries" (grant 03EK3027D) under the research initiative "Energy storage" of the German Federal government.

References

1. W. van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell Syst. Tech. J.*, 29(4):560–607, 1950.
2. S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*, 3rd edition. Wiley, 2006. ISBN: 978-0-471-14323-9.
3. J. Piprek. *Optoelectronic Devices: Advanced Simulation and Analysis*. Berlin: Springer, 2005.
4. M. Grupen and K. Hess. Simulation of carrier transport and nonlinearities in quantum-well laser diodes. *IEEE J. Quant. Electron.*, 34(1):120–140, 1998.
5. S. Steiger, R. G. Veprek, and B. Witzigmann. Unified simulation of transport and luminescence in optoelectronic nanostructure. *J. Comput. Electron.*, 7(4):509–520, 2008.
6. T. Koprucki, H.-C. Kaiser, and J. Fuhrmann. Electronic states in semiconductor nanostructures and upscaling to semi-classical models. In A. Mielke, editor, *Analysis, Modeling and Simulation of Multiscale Problems*, pp. 365–394. Heidelberg, Berlin: Springer, 2006.
7. T. Koprucki, A. Wilms, A. Knorr, and U. Bandelow. Modeling of quantum dot lasers with microscopic treatment of Coulomb effects. *Opt. Quant. Electron.*, 42:777–783, 2011.
8. R. Coehoorn, W. F. Pasveer, P. A. Bobbert, and M. A. J. Michels. Charge-carrier concentration dependence of the hopping mobility in organic materials with Gaussian disorder. *Phys. Rev. B.*, 72(15):155206, 2005.
9. S. L. M. van Mensfoort and R. Coehoorn. Effect of Gaussian disorder on the voltage dependence of the current density in sandwich-type devices based on organic semiconductors. *Phys. Rev. B.*, 78(8):085207, 2008.
10. G. Paasch and S. Scheinert. Charge carrier density of organics with Gaussian density of states: Analytical approximation for the Gauss-Fermi integral. *J. Appl. Phys.*, 107(10):104501, 2010.

11. D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon Read diode oscillator. *IEEE Trans. Electron. Dev.*, 16(1):64–77, 1969.
12. W. Fichtner, D. J. Rose, and R. E. Bank. Semiconductor device simulation. *SIAM J. Sci. Stat. Comput.*, 4(3):391–415, 1983.
13. S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Berlin: Springer, 1984.
14. U. Bandelow, H. Gajewski, and R. Hünlich. Fabry–Perot lasers: Thermodynamics-based modeling. In J. Piprek, editor, *Optoelectronic Devices*. Berlin: Springer, 2005.
15. A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift diffusion systems. *Nonlinearity*, 24(4):1329, 2011.
16. S. R. de Groot and P. Mazur. *Non-Equilibrium Thermodynamics*. North Holland: Dover, 1962.
17. P. Deuflhard *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, volume 35. Berlin: Springer, 2011.
18. D. N. Allen and R. V. Southwell. Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder. *Quart. J. Mech. Appl. Math.*, 8:129–145, 1955.
19. A. M. Il'in. A difference scheme for a differential equation with a small parameter multiplying the second derivative. *Mat. Zametki.*, 6:237–248, 1969.
20. J. S. Chang and G. Cooper. A practical difference scheme for Fokker-Planck equations. *J. Comput. Phys.*, 6(1):1–16, 1970.
21. H.-G. Roos, M. Stynes, and L. Tobiska. *Robust Numerical Methods for Singularly Perturbed Differential Equations*, volume 24. Berlin: Springer, 2008.
22. K. W. Morton. *Numerical Solution of Convection-Diffusion Problems: Applied Mathematics*. London: Taylor & Francis, 1996.
23. D. Schröder. *Modelling of Interface Carrier Transport for Device Simulation*. Wien: Springer, 1994.
24. S. Y. Karpov. *Visible Light-Emitting Diodes*, Chapter 14. In J. Piprek, editor, *Nitride Semiconductor Devices: Principles and Simulation*, pp. 303–325. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2007.
25. J. S. Blakemore. The parameters of partially degenerate semiconductors. *Proc. Phys. Soc. London A.*, 65:460–461, 1952.
26. M. Gruber, E. Zojer, F. Schürrer, and K. Zojer. Impact of materials versus geometric parameters on the contact resistance in organic thin-film transistors. *Adv. Funct. Mater.*, 23(23):2941–2952, 2013.
27. T. Koprucki and K. Gärtner. Discretization scheme for drift-diffusion equations with strong diffusion enhancement. *Opt. Quant. Electron.*, 45(7):791–796, 2013.
28. R. Haase. *Thermodynamics of Irreversible Processes*. Berlin: Addison-Wesley, 1968.
29. H. Gajewski and K. Gärtner. On the discretization of van Roosbroeck's equations with magnetic field. *Z. Angew. Math. Mech.*, 76(5):247–264, 1996.
30. H. Gajewski and K. Gröger. Semiconductor equations for variable mobilities based on Boltzmann statistics or Fermi-Dirac statistics. *Math. Nachr.*, 140(1):7–36, 1989.
31. G. Albinus, H. Gajewski, and R. Hünlich. Thermodynamic design of energy models of semiconductor devices. *Nonlinearity*, 15(2):367, 2002.
32. M. S. Mock. On equations describing steady-state carrier distributions in a semiconductor device. *Comm. Pure Appl. Math.*, 25(25):781–792, 1972.
33. H. Gajewski and K. Gröger. On the basic equations for carrier transport in semiconductors. *J. Math. Anal. Appl.*, 113:12–35, 1986.
34. H. Gajewski. On existence, uniqueness and asymptotic behavior of solutions of the basic equations for carrier transport in semiconductors. *Z. Angew. Math. Mech.*, 65(2):101–108, 1985.
35. P. A. Markowich. A nonlinear eigenvalue problem modelling the avalanche effect in semiconductor diodes. *SIAM J. Math. Anal.*, 6:1268–1283, 1985.
36. A. Jüngel. On the existence and uniqueness of transient solutions of a degenerate nonlinear drift-diffusion model for semiconductors. *Math. Models Methods Appl. Sci.*, 04:677, 1994.

37. J. W. Jerome. *Analysis of Charge Transport. A Mathematical Study of Semiconductor Devices*. Berlin Heidelberg: Springer, 1996.
38. H. K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Electron Dev.*, 11(10):455–465, 1964.
39. P. A. Markowich. *The Stationary Semiconductor Device Equations*. Berlin: Springer-Verlag, 1986.
40. D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Berlin: Springer, 2001. Reprint of the 1998 edition.
41. I. Rubinstein. *Electro-Diffusion of Ions*, volume 11. Philadelphia, PA: SIAM, 1990.
42. R. H. Macneal. An asymmetrical finite difference network. *Quart. Math. Appl.*, 11:295–310, 1953.
43. P. Fleischmann and S. Selberherr. Three-dimensional Delaunay mesh generation using a modified advancing front approach. In *Proceedings of IMR97*, pp. 267–278, 1997.
44. R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of Numerical Analysis*, P.G. Ciarlet and J.L. Lions (Eds.), 7:713–1018, 2000.
45. J. W. Barrett and C. M. Elliott. Finite element approximation of the Dirichlet problem using the boundary penalty method. *Numer. Math.*, 49(4):343–366, 1986.
46. R. Eymard, J. Fuhrmann, and K. Gärtner. A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local Dirichlet problems. *Numer. Math.*, 102(3):463–495, 2006.
47. K. Gärtner. Existence of bounded discrete steady state solutions of the van Roosbroeck system with monotone Fermi-Dirac statistic functions. *J. Comput. Electron.*, 14(3):773–787, 2015.
48. M. Bessemoulin-Chatard. A finite volume scheme for convection–diffusion equations with nonlinear diffusion derived from the Scharfetter–Gummel scheme. *Numer. Math.*, 121(4):637–670, 2012.
49. T. Koprucki, N. Rotundo, P. Farrell, D. H. Doan, and J. Fuhrmann. On thermodynamic consistency of a Scharfetter–Gummel scheme based on a modified thermal voltage for drift-diffusion equations with diffusion enhancement. *Opt. Quant. Electron.*, 47(6):1327–1332, 2015.
50. J. Fuhrmann. Comparison and numerical treatment of generalised Nernst-Planck models. *Comput. Phys. Commun.*, 196:166–178, 2015.
51. H. Gajewski. *Analysis und Numerik von Ladungstransport in Halbleitern*. WIAS Report No. 6. Berlin: WIAS, 1993. ISSN 0942-9077.
52. H. Gajewski, M. Liero, R. Nürnberg, and H. Stephan. *WIAS-TeSCA User Manual 1.2*. WIAS Technical Report No. 14. Berlin: WIAS, 2016. ISSN: WIAS 1618-7776.
53. A. Glitzky and K. Gärtner. Energy estimates for continuous and discretized electro-reaction–diffusion systems. *Nonlinear Anal.*, 70(2):788–805, 2009.
54. C. Chainais-Hillairet. Entropy method and asymptotic behaviours of finite volume schemes. In J. Fuhrmann, M. Ohlberger, and C. Rohde, editors, *Finite Volumes for Complex Applications VII – Methods and Theoretical Aspects*, pp. 17–35. Berlin: Springer, 2014.
55. M. Bessemoulin-Chatard and C. Chainais-Hillairet. Exponential decay of a finite volume scheme to the thermal equilibrium for drift–diffusion systems. *J. Numer. Math.*, 2016. (Online first, doi: 10.1515/jnma-2016-0007)
    M. Bessemoulin-Chatard and C. Chainais-Hillairet. Exponential decay of a finite volume scheme to the thermal equilibrium for drift–diffusion systems. *arXiv preprintarXiv:1601.00813*, 2016.
56. K. Gärtner. Existence of bounded discrete steady-state solutions of the van Roosbroeck system on boundary conforming delaunay grids. *SIAM J. Sci. Comput.*, 31(2):1347–1362, 2009.
57. R. D. Lazarov, I. D. Mishev, and P. S. Vassilevski. Finite volume methods for convection diffusion problems. *SIAM J. Numer. Anal.*, 33(1):31–55, 1996.
58. J. J. H. Miller and S. Wang. An analysis of the Scharfetter-Gummel box method for the stationary semiconductor device equations. *RAIRO-Math. Num.*, 28(2):123–140, 1994.
59. J. Xu and L. Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Math. Comput. Am. Math. Soc.*, 68(228):1429–1446, 1999.

60. C. Chainais-Hillairet, J.-G. Liu, and Y.-J. Peng. Finite volume scheme for multidimensional drift-diffusion equations and convergence analysis. *ESAIM: Math. Model. Numer. Anal. Math. Num.*, 37(2):319–338, 2003.
61. R. E. Bank, W. M. Coughran Jr, and L. C. Cowsar. The finite volume Scharfetter-Gummel method for steady convection diffusion equations. *Comput. Vis. Sci.*, 1(3):123–136, 1998.
62. M. Augustin, A. Caiazzo, A. Fiebach, J. Fuhrmann, V. John, A. Linke, and R. Umla. An assessment of discretizations for convection-dominated convection-diffusion equations. *Comp. Meth. Appl. Mech. Eng.*, 200:3395–3409, 2011.
63. A. Kuzmin, M. Luisier, and O. Schenk. Fast methods for computing selected elements of the greens function in massively parallel nanoelectronic device simulations. In F. Wolf, B. Mohr, and D. Mey, editors, *Euro-Par 2013 Parallel Processing*, volume 8097, *Lecture Notes in Computer Science*, pp. 533–544. Berlin: Springer, 2013.
64. O. Schenk, M. Bollhöfer, and R. A. Römer. On large-scale diagonalization techniques for the Anderson model of localization. *SIAM Rev.*, 50(1):91–112, 2008.
65. O. Schenk, A. Wächter, and M. Hagemann. Matching-based preprocessing algorithms to the solution of saddle-point problems in large-scale nonconvex interior-point optimization. *Comput. Optim. Appl.*, 36(2–3):321–341, 2007.
66. O. Schenk. PARDISO version 5.0.0. http://www.pardiso-project.org. Accessed on February 22, 2016.
67. T. A. Davis. *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms 2)*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2006.
68. T. Davis. SuiteSparse version 4.5.1. http://faculty.cse.tamu.edu/davis/suitesparse.html. Accessed on February 22, 2016.
69. Y. Saad. *Iterative Methods for Sparse Linear Systems*. Philadelphia, PA: SIAM, 2003.
70. P. Ciarlet. *The Finite Element Method for Elliptic Problems*. Philadelphia, PA: SIAM, 2002.
71. H. Si, K. Gärtner, and J. Fuhrmann. Boundary conforming Delaunay mesh generation. *Comput. Math. Math. Phys.*, 50(1):38–53, 2010.
72. J. Shewchuk. Triangle: A two-dimensional quality mesh generator and Delaunay triangulator. http://www.cs.cmu.edu/~quake/triangle.html. Accessed on December 1, 2015.
73. H. Edelsbrunner. *Algorithms in Combinatorial Geometry*. Heidelberg: Springer-Verlag, 1987.
74. S.-W. Cheng, T. K. Dey, and J. R. Shewchuk. *Delaunay Mesh Generation*. Boca Raton, FL: Chapman & Hall/CRC, 1st edition, 2012.
75. H. Si. TetGen version 1.5. http://tetgen.org/. Accessed on 2015-12-01.
76. J. Schöberl. Netgen an advancing front 2d/3d-mesh generator based on abstract rules. *Comput. Vis. Sci.*, 1(1):41–52, 1997.
77. M. S. Shephard and M. K. Georges. Automatic three-dimensional mesh generation by the finite octree technique. *Int. J. Numer. Meth. Eng.*, 32(4):709–749, 1991.
78. K. Gärtner and R. Richter. DEPFET sensor design using an experimental 3d device simulator. *Nucl. Instrum. Meth. A.*, 568(1):12–17, 2006.
79. B. Schmithüsen, K. Gärtner, and W. Fichtner. Grid adaptation for device simulation according to the dissipation rate. In *Simulation of Semiconductor Processes and Devices 1998*, pp. 197–200. Berlin: Springer, 1998.
80. J. Krause, N. Strecker, and W. Fichtner. Boundary-sensitive mesh generation using an offsetting technique. *Int. J. Numer. Meth. Eng.*, 49(1–2):51–59, 2000.
81. A. Fiebach, A. Glitzky, and A. Linke. Uniform global bounds for solutions of an implicit voronoi finite volume method for reaction–diffusion problems. *Numer. Math.*, 128(1):31–72, 2014.
82. A. Glitzky. Exponential decay of the free energy for discretized electro-reaction–diffusion systems. *Nonlinearity*, 21(9):1989, 2008.
83. E. Hairer, S. P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations: Nonstiff Problems. v. 2: Stiff and Differential-Algebraic Problems*. Berlin: Springer Verlag, 2010.

84. M. S. Mock. A time-dependent numerical model of the insulated-gate field-effect transistor. *Solid-State Electron.*, 24(10):959–966, 1981.
85. P. D. Yoder, K. Gärtner, U. Krumbein, and W. Fichtner. Optimized terminal current calculation for Monte Carlo device simulation. *IEEE Trans. Comput-Aided Design Integr. Circuits Syst.*, 16(10):1082–1087, 1997.
86. M. Auf Der Maur. *A Multiscale Simulation Environment for Electronic and Optoelectronic Devices*. PhD thesis, University of Tor Vergata, 2008.
87. V. John and P. Knobloch. On spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: Part I–A review. *Comput. Methods Appl. Mech. Eng.*, 196(17):2197–2215, 2007.
88. A. Prohl and M. Schmuck. Convergent discretizations for the Nernst-Planck-Poisson system. *Numer. Math.*, 111(4):591–630, 2009.
89. V. Palankovski and R. Quay. *Analysis and Simulation of Heterostructure Devices*. Computational Microelectronics. Vienna: Springer Science & Business Media, 2004.
90. H. Bässler. Charge transport in disordered organic photoconductors a Monte Carlo simulation study. *Phys. Stat. Sol. (B).*, 175(1):15–56, 1993.
91. J. Fabian, S. C. Erwin, and I. Zutic. Bipolar spintronics: Fundamentals and applications. *IBM J. Res. Dev.*, 50(1):121–139, 2006.
92. A. Glitzky and K. Gärtner. Existence of bounded steady state solutions to spin-polarized drift-diffusion systems. *SIAM J. Math. Anal.*, 41(6):2489–2513, 2010.
93. A. Jüngel, C. Negulescu, and P. Shpartko. Bounded weak solutions to a matrix drift diffusion model for spin-coherent electron transport in semiconductors. *Math. Models Methods Appl. Sci.*, 25(5):929–958, 2015.
94. C. Chainais-Hillairet, A. Jüngel, and P. Shpartko. A finite-volume scheme for a spinorial matrix drift-diffusion model for semiconductors. *Numer. Methods for Partial Differ. Equ.*, 32(3):819–846, 2016.
95. M. Auf der Maur, A. Pecchia, G. Penazzi, F. Sacconi, and A. Carlo. Coupling atomistic and continuous media models for electronic device simulation. *J. Comput. Electron.*, 12(4):553–562, 2013.
96. M. Hinze and R. Pinnau. An optimal control approach to semiconductor design. *Math. Models Methods Appl. Sci.*, 12(1):89–107, 2002.
97. M. Hinze and R. Pinnau. Second-order approach to optimal semiconductor design. *J. Optim. Theory. Appl.*, 133(2):179–199, 2007.
98. M. Burger and R. Pinnau. Fast optimal design of semiconductor devices. *SIAM J. Appl. Math.*, 64(1):108–126, 2003.
99. D. Peschka, N. Rotundo, and M. Thomas. Towards doping optimization of semiconductor lasers. To appear in. *J. Comput. Theor. Trans.*, 45(5):410–423, 2016.
100. T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. IEEE*, 91(2):251–274, 2003.
101. S.-M. Hong and C. Jungemann. A fully coupled scheme for a Boltzmann-Poisson equation solver based on a spherical harmonics expansion. *J. Comput. Electron.*, 8(3):225–241, 2009.
102. A. Jüngel. *Transport Equations for Semiconductors*, volume 773. Berlin: Springer, 2009.

51

Monte Carlo Device Simulations

Katerina Raleva
Abdul R. Shaik
Raghuraj Hathwar
Akash Laturia
Suleman S. Qazi
Robin Daugherty
Dragica Vasileska
and
Stephen M. Goodnick

51.1 Introduction 773
Transistor Scaling • More Moore Technology • Importance of Modeling and Simulation
51.2 Bulk Monte Carlo Method Description 779
51.3 Particle-Based Device Simulation 791
Calculation of the Current • Ohmic Contacts • Time-Step • PM Coupling
51.4 Representative Simulation Results 796
Bulk Monte Carlo Simulations of Different Materials • Modeling Thermal Effects in SOI Devices Using Particle-Based Device Simulator
51.5 Outlook 803

51.1 Introduction

The semiconductor device–based electronics industry has been the largest industry in the world with global sales of over a trillion dollars since 1998. The revolution in the semiconductor industry, a large portion of the electronics industry, began in 1947 with the fabrication of bipolar devices on slabs of polycrystalline germanium (Ge). Single-crystalline materials were later proposed and introduced that made possible the fabrication of grown junction transistors. Migration to silicon (Si)–based devices was initially hindered by the stability of the Si/SiO_2 materials system, necessitating a new generation of crystal pullers with improved environmental controls to prevent SiO_2 formation. Later, the stability and low interface-state density of the Si/SiO_2 materials system provided passivation of junctions and eventually, the migration from bipolar devices to field-effect devices in 1960. By 1968, both complementary metal–oxide–semiconductor (CMOS) devices and poly-Si gate technology that allowed self-alignment of the gate to the source/drain of the device had been developed. These innovations permitted a significant reduction in power dissipation and a reduction of the overlap capacitance, improving frequency performance and resulting in the essential components of the modern CMOS device.

Since the invention of the bipolar transistor in 1947, the number and variety of semiconductor devices have increased tremendously as advanced technology, new materials, and broadened comprehension have been applied to the creation and innovation of new devices. Dr. William Shockley suggested in the fifties the use of semiconductors of different bandgap for the fabrication of heterostructure devices. Professor Herbert Kroemer's contributions to heterostructures—from heterostructures bipolar transistors to lasers—culminated in a Nobel Prize in Physics in 2000 and have paved the way for novel heterostructure devices including those in silicon. The unique properties of semiconductor materials have enabled the development of a wide variety of ingenious devices that have literally changed our world. To date, there are about

TABLE 51.1 Major Semiconductor Device Discoveries

- 1874: Metal-semiconductor contact
- 1907: Light-emitting diode
- 1947: Bipolar junction transistor (BJT)
- 1954: Solar cell
- 1957: Heterojunction bipolar transistor (HBT)
- 1958: Tunnel diode
- 1959: Integrated circuits
- 1960: Field-effect transistors (FETs)
- 1962: Semiconductor lasers
- 1966: Metal-semiconductor FET (MESFET)
- 1967: Nonvolatile semiconductor memory
- 1974: Resonant tunneling diode (RTD)
- 1980: Modulation (MOD) FET
- 1994: Room-temperature single-electron memory cell (SEMC)
- 2001: 15 nm MOSFET
- 2006: 3 nm FinFET
- 2012: Single atom transistor in Si technology

60 major devices, with over 100 device variations related to them. A list of most of the basic semiconductor devices discovered and used over the past century with the date of their introduction is shown in Table 51.1.

The metal–oxide–semiconductor field-effect transistor (MOSFET) and related integrated circuits (ICs) now constitute about 90% of the semiconductor device market. Combining Si with the elegance of the field-effect transistor (FET) structure has allowed simultaneously making devices smaller, faster, and cheaper—the mantra that has driven the modern semiconductor microelectronics industry. Indeed, the single factor driving continuous device improvement has been the semiconductor industry's relentless effort to reduce the cost per function on a chip. The way this is done is to put more devices on a chip while either reducing manufacturing costs or holding them constant. This leads to three methods of reducing the cost per function. The first is *transistor scaling*, which involves reducing the transistor size in accordance with some goal, i.e., keeping the electric field constant from one generation to the next. With smaller transistors, more can fit into a given area than in previous generations. The second method is *circuit cleverness*, which is associated with the physical layout of the transistors with respect to each other. If the transistors can be packed into a tighter space, then more devices can fit into a given area than before. The third method is to make the die larger. More devices can be fabricated on a *larger die*. All the while, the semiconductor industry is constantly looking for technological breakthroughs to decrease the manufacturing cost. All of this effort serves to reduce the cost per function on a chip.

51.1.1 Transistor Scaling

Device engineers are most concerned with the method of scaling introduced in Section 51.1. The semiconductor industry has been so successful in providing continued system performance improvement year after year that the Semiconductor Industry Association (SIA) has been publishing roadmaps for semiconductor technology since 1992. These roadmaps represent a consensus outlook of industry trends, taking history as a guide. SIA roadmaps [1] incorporate participation from the global semiconductor industry, including the United States, Europe, Japan, Korea, and Taiwan. They basically affirm the desire of the industry to continue with *Moore's law* [2], which is often stated as the doubling of transistor performance and quadrupling of the number of devices on a chip every 3 years. The phenomenal progress signified by Moore's law has been achieved through scaling of the MOSFET from larger to smaller physical dimensions [3].

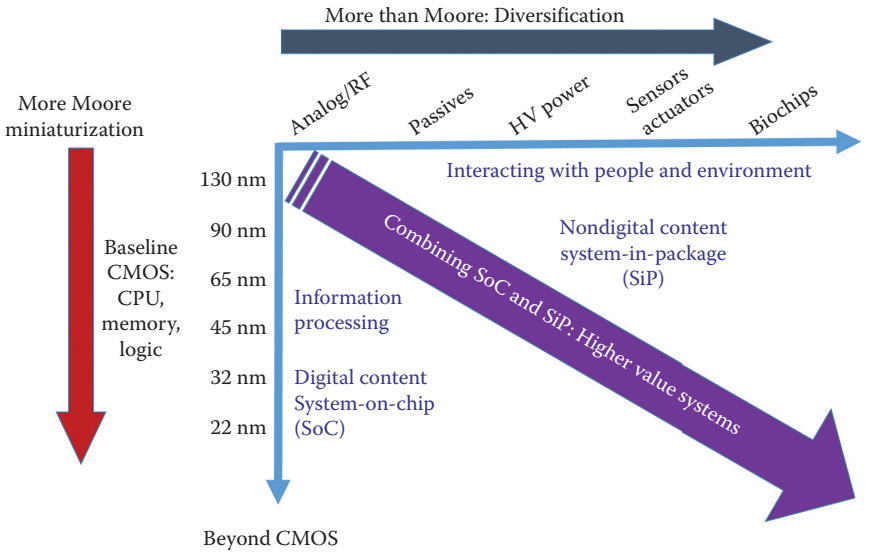


FIGURE 51.1 Schematic description of More Moore, More than Moore, and beyond complementary metal–oxide–semiconductor (CMOS). (RF, radio frequency; HV, high-voltage. [Reprinted from ITRS Roadmap for Semiconductors. Available at: www.itrs.org. With permission.])

The *More Moore* domain is internationally defined as an attempt to further develop advanced CMOS technologies and reduce the associated cost per function by introducing strain and silicon-on-insulator (SOI) technology (see Figure 51.1). Almost 70% of the total semiconductor components market is directly impacted by advanced CMOS miniaturization achieved in the More Moore domain. This 70% comprises three component groups of similar size, namely microprocessors, mass memories, and digital logic. As an example, in recent years, technology has advanced to fabricate ICs at 14-nm gate length commercially [4]. Fab industry giants such as Intel, TSMC, Samsung, and Global Foundries have plans to fabricate ICs at 10 nm by 2017. Samsung has already fabricated and tested 128 Mb SRAM in 10 nm [5] and is hoping to commercialize the process by the end of 2016. This aggressive scaling of technology is possible because of the advent of FinFETs and fully-depleted silicon-on-insulator (FD-SOI) device technology [6].

More than Moore (MtM) refers to a set of technologies that enable nondigital micro-/ nanoelectronic functions. MtM radio frequency (RF), high-voltage power, solid-state lighting (SSL), medical ultrasound, biochips and microfluidics, energy scavenging, electronic imaging, sensors, and actuators on CMOS platforms are some of the applications. They are based on, or derived from, Si technology but do not necessarily scale with Moore's Law. MtM devices typically provide conversion of nondigital as well as nonelectronic information, such as mechanical, thermal, acoustic, chemical, optical, and biomedical functions to digital data, and vice versa.

51.1.2 More Moore Technology

For digital circuits, a figure of merit for MOSFETs for unloaded circuits is CV/I, where C is the gate capacitance, $V$the voltage swing, and I the current drive of the MOSFET. For loaded circuits, the current drive of the MOSFET is of paramount importance. Keeping in mind both the CV/I metric and the benefits of a large current drive, one notes that device performance may be improved [7] by (1) inducing a larger charge density for a given gate voltage drive; (2) enhancing the carrier transport by improving the mobility, saturation velocity, or ballistic transport; (3) ensuring device scalability to achieve a shorter channel length; and (4) reducing parasitic capacitances and resistances. For capitalizing these opportunities, the

proposed technology options generally fall into two categories: new materials and new device structures. In many cases, the introduction of a new material requires the use of a new device structure, or vice versa. To fabricate devices beyond current scaling limits, IC companies have simultaneously pushed the planar, bulk Si CMOS design while incorporating alternative gate stack materials (high-k dielectric [8] and metal gates), band engineering methods (using strained Si [9] or SiGe [10]), and alternative transistor structures. The concept of a band-engineered transistor is to enhance the mobility of electrons and/or holes in the channel by modifying the band structure of Si in the channel in a way such that the physical structure of the transistor remains substantially unchanged. This enhanced mobility increases the transistor transconductance (g_m) and on-drive current (I_{on}). A SiGe layer or a strained Si on a relaxed SiGe layer is used as an enhanced-mobility channel layer. It has already been demonstrated experimentally that at $T = 300$ K (room temperature), effective hole enhancement of about 50% can be achieved using the SiGe technology [10]. Intel has adopted strained Si technology for the first time for its 90-nm process [11]. This resulted in nearly 20% performance improvement per technology generation, with only a few additional process steps. The challenge in identifying suitable high-k dielectrics and metal gates for both conventional p-channel MOS (PMOS) and n-channel MOS (NMOS) transistors has led to early adoption of alternative transistor designs. These include primarily partially-depleted (PD) SOI and FD-SOI devices. Today, there is extensive research in double-gate (DG) structures and FinFET transistors [12], which have better electrostatic integrity and, theoretically, have better transport properties than single-gated FETs. A FinFET is a form of a DG transistor having surface conduction channels on two opposite vertical surfaces and having current flow in the horizontal direction. The channel length is given by the horizontal separation between source and drain and is usually determined by a lithographic step combined with a side-wall spacer etch process. Many innovative structures, involving structural challenges such as fabrication on nanometer-scale fins and nanometer-scale planarization over an entire wafer are currently being implemented. Some new and revolutionary technology such as carbon nanotubes or molecular transistors might be on the horizon, but it is not clear, in view of the predicted future capabilities of CMOS, whether they will be competitive. In conclusion, the semiconductor industry is approaching the end of an era of scaling gains by rote shrinkage of device dimensions and entering a postscaling era, a new phase of CMOS evolution in which innovation is demanded simply to compete. The trends in benefits to density, performance, and power will be continued through such innovations. Rather than coming to a close, a new era of CMOS technology is just beginning.

51.1.3 Importance of Modeling and Simulation

As semiconductor devices are scaled into nanoscale regimes, first, velocity saturation starts to limit the carrier mobility due to pronounced intervalley scattering; when the device dimensions are scaled below 200 nm, velocity overshoot (velocity of the carriers in the channel is larger than the steady-state saturation velocity) starts to dominate the device behavior leading to larger on-state currents. Alongside with the developments in the semiconductor nanotechnology, in recent years, there has been significant progress in physically based modeling of semiconductor devices. There are two issues that make simulation important. Product cycles are getting shorter with each generation, and the demand for production wafers shadows development efforts in the factory. Consider the product cycle issue first. In order for companies to maintain their competitive edge, products have to be taken from design to production in less than 18 months. As a result, the development phase of the cycle is getting shorter. Contrast this requirement with the fact that it takes 2–3 months to run a wafer lot through a factory, depending on its complexity. The specifications for experiments run through the factory must be near the final solution. While simulations may not be completely predictive, they provide a good initial guess. This can ultimately reduce the number of iterations during the device development phase. The second issue that reinforces the need for simulation is the production pressures that factories face. To meet customer demand, development factories are making way for production space. It is also expensive to run experiments through a production facility. Such resources could have otherwise been used to produce sellable products. Again, device simulation can be

used to decrease the number of experiments run through a factory. Device simulation can be used as a tool to guide manufacturing down the right path, thereby decreasing the development time and costs. Besides offering the possibility to test hypothetical devices that have not (or could not) yet been manufactured, device simulation offers unique insight into device behavior by allowing the observation of phenomena that cannot be measured on real devices. It is related to, but usually separate from, process simulation, which deals with various physical processes such as material growth, oxidation, impurity diffusion, etching, and metal deposition inherent in device fabrication leading to ICs. Device simulation is distinct from another important aspect of computer-aided design (CAD), device modeling, which deals with compact behavioral models for devices and subcircuits relevant for circuit simulation in commercial packages such as simulation program with integrated circuit emphasis (SPICE). Device simulation can provide the parameters that are used to generate the compact behavioral models, and when coupled with process simulation and circuit simulation, it provides a hierarchical approach to technology computer-aided design (TCAD).

The main components of semiconductor device simulation at any level of approximation are illustrated in Figure 51.2 [13]. There are two main kernels, which must be solved self-consistently with one another, the *transport equations* governing charge flow, and the *fields* driving charge flow. Both are coupled strongly to one another, and hence must be solved simultaneously. The fields arise from external sources, as well as the charge and current densities which act as sources for the time-varying electric and magnetic fields obtained from the solution of Maxwell's equations. Under appropriate conditions, only the quasi-static electric fields arising from the solution of Poisson's equation are necessary. The fields, in turn, are driving forces for charge transport as illustrated in Figure 51.3 for the various levels of approximation within a hierarchical structure ranging from compact modeling at the top to an exact quantum-mechanical description at the bottom. For devices for which gradual channel approximation cannot be used due to the two-dimensional (2D) nature of the electrostatic potential and the electric fields driving the carriers from source to drain, drift-diffusion models have been exploited. These models are valid, in general, for large devices in which the fields are not that high so that there is no degradation of the mobility due to the electric field. The validity of the drift-diffusion models can be extended to take into account the velocity saturation effect with the introduction of field-dependent mobility and diffusion coefficients. When velocity overshoot becomes important (for length scales less than 200 nm—see Figure 51.4), the drift-diffusion model is no longer valid and the hydrodynamic model must be used.

The hydrodynamic model has been the workhorse for technology development and several high-end commercial device simulators have appeared including Silvaco, Synopsys, Crosslight, etc. The advantages of the hydrodynamic model are that it allows quick simulation runs, but the problem is that the amount of the velocity overshoot depends upon the choice of the energy relaxation time. The smaller the device,

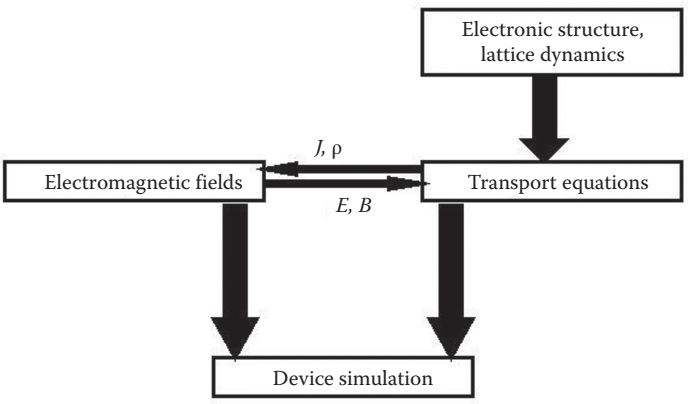


FIGURE 51.2 A schematic description of the building blocks of device simulator.

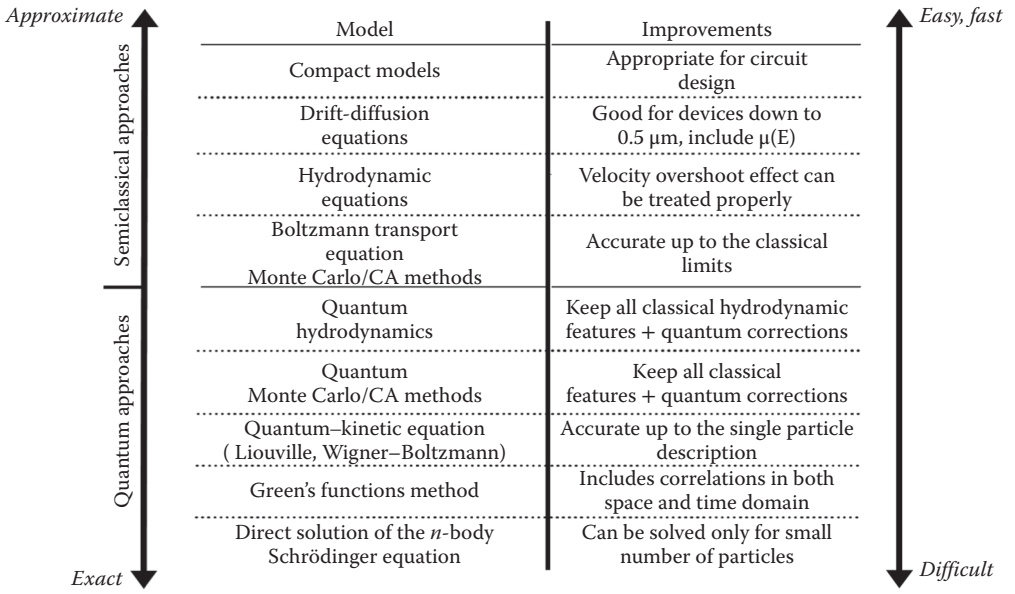


FIGURE 51.3 Illustration of the hierarchy of transport models.

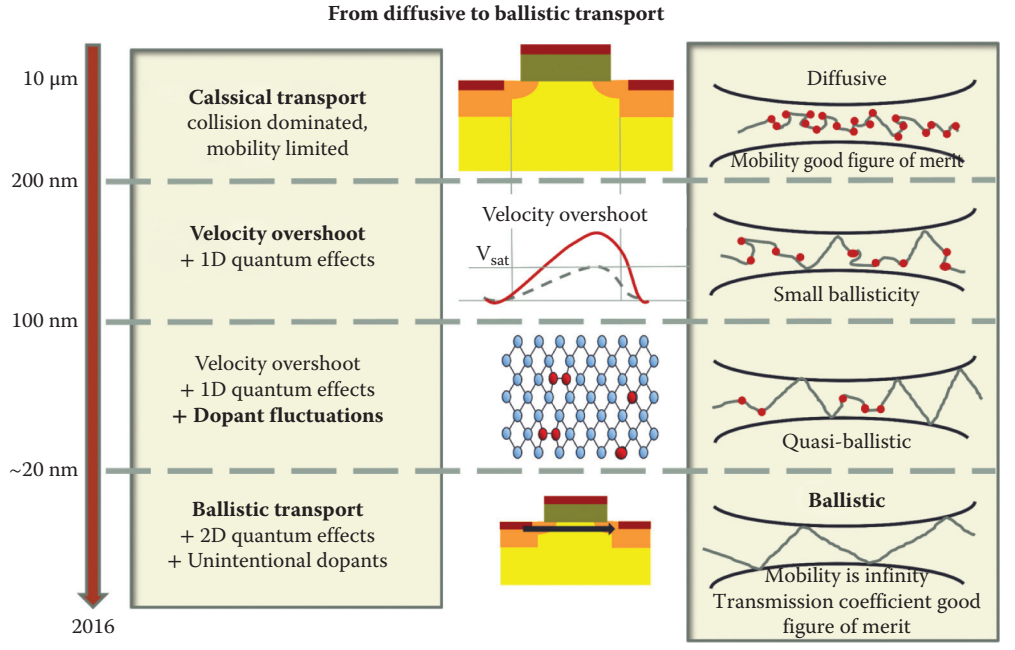


FIGURE 51.4 From diffusive to ballistic transport when scaling semiconductor devices into nanometer scale.

the larger the deviation when using the same set of energy relaxation times. A standard way in calculating the energy relaxation times is to use bulk Monte Carlo simulations. However, the energy relaxation times are material-, device geometry-, and doping-dependent parameters, so their determination ahead of time is not possible.

To avoid the problem of the proper choice of the energy relaxation times in hydrodynamic models, a direct solution of the Boltzmann transport equation (BTE)

$$\frac{\partial f}{\partial t} + \frac{1}{\hbar}\nabla_k E(k) \cdot \nabla_r f + \left(-\frac{e\mathbf{E}}{\hbar}\right) \cdot \nabla_k f = \left(\frac{\partial f}{\partial t}\right)_{\mathbf{coll}} \tag{51.1}$$

using the Monte Carlo method is the best method of choice for nanoscale devices in which position and momentum can still be treated as independent variables. In Equation 51.1, $f(r, k, t)$ is the semiclassical distribution function, $E(k)$ the dispersion relation for the electrons, and $\mathbf{E}$ the electric field. When, for example, collisional broadening of the states is important, quantum transport approaches must be pursued.

This chapter focuses on Monte Carlo methods for the solution of the BTE and is organized as follows. Monte Carlo solution of the BTE for homogeneous systems is given in Section 51.2. Particle-based device simulator is described in Section 51.3. Representative simulation results for bulk material and for semiconductor devices are given in Section 51.4. Conclusions and future directions of research are presented in Section 51.5.

51.2 Bulk Monte Carlo Method Description

The ensemble Monte Carlo (EMC) techniques have been used for well over 40 years as a numerical method to simulate nonequilibrium transport in semiconductor materials and devices and have been the subject of numerous books and reviews [14–16]. In the applications to transport problem, a random walk is generated using the random number–generating algorithms common to modern computers to simulate the stochastic motion of particles subject to collision processes. This process of random walk generation is part of a very general technique used to evaluate integral equations and is connected to the general random sampling technique used in the evaluation of multidimensional integrals [17].

The basic technique as applied to transport problems is to simulate the free particle motion (referred to as the *free flight*) terminated by instantaneous random *scattering events*. The Monte Carlo algorithm consists of generating random free-flight times for each particle, choosing the type of scattering occurring at the end of the free flight, changing the final energy and momentum of the particle after scattering, and then repeating the procedure for the next free flight. Sampling the particle motion at various times throughout the simulation allows for the statistical estimation of physically interesting quantities such as the single particle distribution function, the average drift velocity in the presence of an applied electric field, the average energy of the particles, etc. By simulating an *ensemble* of particles, representative of the physical system of interest, the nonstationary time-dependent evolution of the electron and hole distributions under the influence of a time-dependent driving force may be simulated. This particle-based picture, in which the particle motion is decomposed into free flights terminated by instantaneous collisions, is basically the same approximate picture underlying the derivation of the semiclassical BTE. In fact, it may be shown that the one-particle distribution function obtained from the random walk Monte Carlo technique satisfies the BTE for a homogeneous system in the long-time limit [18]. This semiclassical picture breaks down when quantum-mechanical effects become pronounced, and one cannot unambiguously describe the instantaneous position and momentum of a particle. In the following, we describe first the standard Monte Carlo algorithm used to simulate charge transport in semiconductors. We then discuss how this basic model for charge transport within the BTE is self-consistently solved with the appropriate field equations to perform particle-based device simulation (Section 51.3).

In the bulk Monte Carlo method [19], particle motion is assumed to consist of free flights terminated by instantaneous scattering events, which change the momentum and energy of the particle after scattering. So, the first task is to generate free flights of random time duration for each particle. To simulate this process, the probability density, $P(t)$, is required, in which $P(t)\mathrm{d}t$ is the joint probability that a particle will arrive at time t without scattering after a previous collision occurring at time $t = 0$, and then suffer a collision in a

time interval dt around time t. The probability of scattering in the time interval dt around t may be written as $\Gamma[\mathbf{k}(t)]\mathrm{d}t$, where $\Gamma[\mathbf{k}(t)]$ is the scattering rate of an electron or hole of wavevector $\mathbf{k}$. The scattering rate, $\Gamma[\mathbf{k}(t)]$, represents the sum of the contributions from each individual scattering mechanism, which are usually calculated quantum mechanically using perturbation theory [20]. The implicit dependence of $\Gamma[\mathbf{k}(t)]$ on time reflects the change in $\mathbf{k}$ due to acceleration by internal and external fields. For electrons subject to time-independent electric and magnetic fields, the time evolution of $\mathbf{k}$ between collisions is represented as

$$\mathbf{k}(t) = \mathbf{k}(0) - \frac{e(\mathbf{E} + \mathbf{v} \times \mathbf{B})t}{\hbar}, \tag{51.2}$$

where $\mathbf{E}$ is the electric field, $\mathbf{v}$ the electron velocity, and $\mathbf{B}$ the magnetic flux density. In terms of the scattering rate, $\Gamma[\mathbf{k}(t)]$, the probability that a particle has not suffered a collision after a time t is given by $\exp\left(-\int_0^t \Gamma\left[k(t')\right] \mathrm{d}t'\right)$. Thus, the probability of scattering in the time interval dt after a free flight of time $t$may be written as the joint probability:

$$P(t)\mathrm{d}t = \Gamma[k(t)] \exp\left[-\int_0^t \Gamma\left[k\left(t'\right)\right] \mathrm{d}t'\right] \mathrm{d}t. \tag{51.3}$$

Random flight times may be generated according to the probability density $P(t)$ above using, for example, the pseudo-random number generator implicit on most modern computers, which generate uniformly distributed random numbers in the range [0, 1]. Using a direct method, random flight times sampled from $P(t)$ may be generated according to

$$r = \int_0^{t_r} P(t)\mathrm{d}t, \tag{51.4}$$

where r is a uniformly distributed random number and t_r is the desired free-flight time. Integrating Equation 51.4 with $P(t)$ given by Equation 51.3 above yields

$$r = 1 - \exp\left[-\int_0^{t_r} \Gamma\left[k(t')\right] dt'\right]. \tag{51.5}$$

Since $1 - r$ is statistically the same as r, Equation 51.5 may be simplified to

$$-\ln r = \int_0^{t_r} \Gamma\left[\mathbf{k}\left(t'\right)\right] \mathrm{d}t'. \tag{51.6}$$

Equation 51.6 is the fundamental equation used to generate the random free-flight time after each scattering event, resulting in a random walk process related to the underlying particle distribution function. If there is no external driving field leading to a change of $\mathbf{k}$ between scattering events (e.g., in ultrafast photoexcitation experiments with no applied bias), the time dependence vanishes, and the integral is trivially evaluated. In the general case, where this simplification is not possible, it is expedient to introduce the so-called self-scattering method [21] in which one introduces a fictitious scattering mechanism whose

scattering rate always adjusts itself in such a way that the total (self-scattering plus real scattering) rate is energy-independent:

$$\Gamma = \Gamma\left[\mathbf{k}(t')\right] + \Gamma_{\text{self}}\left[\mathbf{k}(t')\right], \tag{51.7}$$

where $\Gamma_{\text{self}}[\mathbf{k}(t')]$ is the self-scattering rate. The self-scattering mechanism itself is defined such that the final state before and after scattering is identical. Hence, it has no effect on the free-flight trajectory of a particle when selected as the terminating scattering mechanism, yet results in the simplification of Equation 51.6 such that the free flight is given by

$$t_r = -\frac{1}{\Gamma}\ln r. \tag{51.8}$$

The constant total rate (including self-scattering) Γ must be chosen at the start of the simulation interval (there may be multiple such intervals throughout an entire simulation) so that it is larger than the maximum scattering encountered during the same time interval. In the simplest case, a single value is chosen at the beginning of the entire simulation (constant gamma method), checking to ensure that the real rate never exceeds this value during the simulation. Other schemes may be chosen that are more computationally efficient, and which modify the choice of Γ at fixed time increments [22].

The algorithm described above determines the random free-flight times during which the particle dynamics is treated semiclassically. For the scattering process itself, we need the type of scattering (i.e., impurity, acoustic phonon, photon emission, etc.) which terminates the free flight, and the final energy and momentum of the particle(s) after scattering. The type of scattering that terminates the free flight is chosen using a uniform random number between 0 and Γ, and using this pointer to select among the relative total scattering rates of all processes including self-scattering at the final energy and momentum of the particle:

$$\Gamma = \Gamma_{\text{self}}[n,\mathbf{k}] + \Gamma_1[n,\mathbf{k}] + \Gamma_2[n,\mathbf{k}] + \cdots \Gamma_N[n,\mathbf{k}], \tag{51.9}$$

with n the band index of the particle (or sub-band in the case of reduced-dimensionality systems), and $\mathbf{k}$ the wavevector at the end of the free flight. This process is illustrated schematically in Figure 51.5.

Once the type of scattering terminating the free flight is selected, the final energy and momentum (as well as band or subband) of the particle due to this type of scattering must be selected. For elastic scattering processes such as ionized impurity scattering, the energy before and after scattering is the same. For the interaction between electrons and the vibrational modes of the lattice described as quasi-particles known as phonons, electrons exchange finite amounts of energy with the lattice in terms of emission and absorption of phonons. For determining the final momentum after scattering, the scattering rate, $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$ of the jth scattering mechanism is needed, where n and m are the initial and final band indices, and $\mathbf{k}$ and $\mathbf{k}'$ are the particle wavevectors before and after scattering. Defining a spherical coordinate system

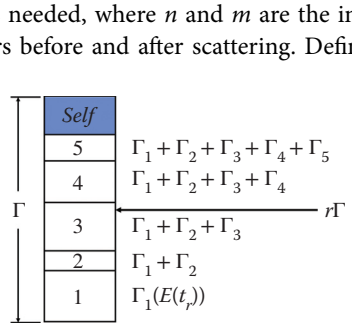


FIGURE 51.5 Selection of the type of scattering terminating a free flight in the Monte Carlo algorithm.

around the initial wavevector $\mathbf{k}$, the final wavevector $\mathbf{k}'$ is specified by $|\mathbf{k}'|$ (which depends on conservation of energy) as well as the azimuthal and polar angles, ϕ and θ around $\mathbf{k}$. Typically, the scattering rate, $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$, only depends on the angle θ between $\mathbf{k}$ and $\mathbf{k}'$. Therefore, ϕ may be chosen using a uniform random number between 0 and 2π (i.e., $2\pi r$), while θ is chosen according to the angular dependence for scattering arising from $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$. If the probability for scattering into a certain angle $P(\theta)d\theta$ is integrable, then random angles satisfying this probability density may be generated from a uniform distribution between 0 and 1 through a prescribed recipe. Otherwise, a rejection technique may be used to select random angles according to $P(\theta)$. Common scattering mechanisms that contribute to transport are summarized in Figure 51.6. The corresponding scattering rate expressions for general nonparabolic bands are summarized in Table 51.2 [23–25].

A general Monte Carlo code is then developed as follows: First, a subroutine is typically called that contains all material and scattering rates parameters for the scattering mechanisms included in the theoretical model. After the material and run parameters are read, in the first step of the Monte Carlo simulation procedure, it is necessary to construct scattering tables for the Γ, L, and X valleys (for GaAs as a prototypical example) that initialize a series of events that are summarized in Figure 51.7. Then, for each energy, the cumulative scattering rates for each valley are stored in separate look-up tables and renormalized according to the maximum scattering rate (including self-scattering) that occurs over the range of energies stored.

Having constructed the scattering table and after renormalizing the table, examples of which are given for GaAs in Figures 51.8 and 51.9 for the Γ, L, and X valleys, the next step is to initialize the carrier's wavevector and energy and the initial free-flight time. This is accomplished by calling the initialization subroutine. Energy and wavevector histograms of the initial carrier energy and the components of the wavevector along the x-, y-, and z-axes are shown in Figure 51.10. For good statistics, the number of particles simulated is typically taken to be 10,000, and one can see the statistical fluctuation of these average quantities associated with the finite number of particles. Notice that the variance is inversely proportional to the square-root of the number of particles. Also note that the initial y-component for the wavevector is symmetric around the y-axis, which means that the average wavevector along the y-axis is zero, which should be expected since the electric field along the y-component is zero at $t = 0$. Identical distributions have been obtained for the x- and for the z-components of the wavevector. Also note that the energy distribution has the Maxwell–Boltzmann form, as it should be expected. One can also estimate from this graph that the average energy of the carriers is on the order of $(3/2)k_B T$ (k_B is the Boltzmann constant and T the temperature).

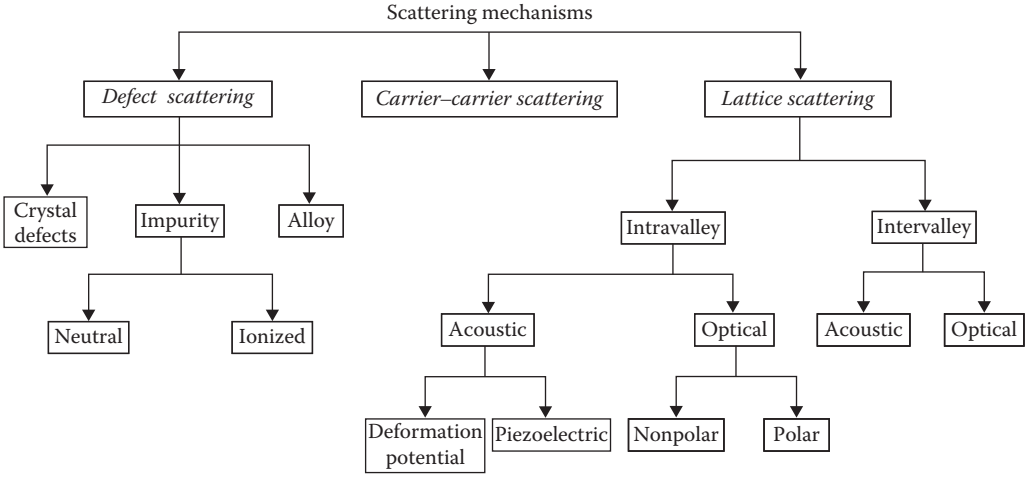


FIGURE 51.6 Scattering mechanisms in a typical semiconductor.

TABLE 51.2 Scattering Rate Expressions Γ(k) = W(k) = W(E) for Electrons for General Nonparabolic Energy Bands

1. *Ionized impurity scattering:* Scattering from ionized Coulomb charges in the Brooks–Herring approach.

$$W(E) = \left(\frac{\sqrt{2}e^4 N_I m_d^{\frac{3}{2}}}{\pi\varepsilon_s^2\gamma^4}\right) \times \left(\sqrt{E(1+\alpha E)} \times (1+2\alpha E)\right) \times \left(\frac{1}{q_D^2\left(q_D^2 + \frac{8m_d E(1+\alpha E)}{\hbar^2}\right)}\right) \quad q_D = \sqrt{\frac{e^2(n+p)}{\varepsilon_s K_B T_L}},$$

where n and p are the electron and hole concentrations, respectively, k_B is the Boltzmann constant, m_d is the density of states mass, T_L is the lattice temperature, ε_s is the permittivity of the material, E is the energy, and α is the nonparabolicity factor.

2. *Alloy disorder scattering* (e.g., $Al_xGa_{1-x}As$): Scattering due to the presence of a random alloy.

$$W(E) = \left(\frac{x(1-x)a^3}{\pi}\right) \times \left(\frac{D_{alloy}^2 d}{\hbar^4}\right) \times m_d\sqrt{2m_d E(1+\alpha E)} \times (1+2\alpha E),$$

where x is the mole fraction, d the lattice disorder ($0 \le d \le 1$), a the lattice constant, and D_{alloy} the alloy disorder scattering potential.

3. *Dislocation scattering* (e.g., GaN): Scattering due to crystal defects/dislocations (Coulomb-type interaction).

$$W(E) = \left(\frac{N_{dis} m_d e^4}{4\hbar^3\varepsilon^2 c^2}\right) \times \left(\frac{\lambda^4}{\left(1+\frac{8\lambda^2 m_d E(1+\alpha E)}{\hbar^2}\right)^{\frac{3}{2}}}\right) \times \left(1 + \frac{4\lambda^2 m_d E(1+\alpha E)}{\hbar^2}\right) \times (1+2\alpha E) \quad \lambda = \sqrt{\frac{\varepsilon K_B T_L}{e^2 n'}},$$

where n' is the effective screening concentration and N_{dis} the line dislocation density.

4. *Acoustic deformation potential scattering:* Describes fluctuations in the band structure due to deformation of the unit cell because of the strain due to the different amounts of displacement of the atoms (in the same direction).

$$W(E) = \left(\frac{2\pi D_{ac}^2 K_B T_L}{\hbar C_l}\right) \times \left(\frac{(2m_d)^{\frac{3}{2}}\sqrt{E(1+\alpha E)}}{4\pi^2\hbar^3}\right) \times (1+2\alpha E),$$

where D_{ac} is the acoustic deformation potential constant and C_l is the sound velocity.

5. *Piezoelectric scattering:* Dipole-type interaction in zinc-blende materials in which there is charge transfer between the two atoms in the unit cell. The optical modes of vibration (vibration of the atoms in opposite direction) directly affect the magnitude of the dipole moment, which in turn leads to scattering potential.

$$W(E) = \left(\frac{m_d^{\frac{1}{2}} K_B T_L}{4\sqrt{2}\pi\rho V_s^2\hbar^2}\right) \times \left(\frac{1+2\alpha E}{\sqrt{E(1+\alpha E)}}\right) \times \left(\frac{e e_{pz}}{\varepsilon_\infty}\right)^2 \times \ln\left(1 + \frac{8m_d E(1+\alpha E)}{\hbar^2 q_D^2}\right) \quad q_D = \sqrt{\frac{e^2(n+p)}{\varepsilon_s K_B T_L}},$$

where ρ is the crystal density, v_s sound velocity, e_{pz} the piezoelectric constant, and ε_∞ the high-frequency dielectric constant.

6. *Polar optical phonon (POP) scattering:* In polar materials (where there is transfer of charge from one atom to the other, which in turn, leads to a dipole moment), optical modes of vibration cause direct modification of the dipole moment, which then leads to polar optical phonon scattering which is a Coulomb-type interaction.

$$W(E) = \left(\frac{\sqrt{m_d}e^2 W_{LO}}{4\sqrt{2}\pi\hbar\varepsilon_p}\right) \times \left(N_0 + \frac{1}{2} \mp \frac{1}{2}\right) \times \left(\frac{1+2\alpha E_k'}{\gamma_k}\right) \times F\left(E_k, E_k'\right)$$

$$N_0 = \frac{1}{W_{\frac{\hbar_{LO}}{e^{K_B T_L}-1}}} \qquad \varepsilon_p = \frac{1}{\frac{1}{\varepsilon_{high}} - \frac{1}{\varepsilon_{low}}} \qquad F(E_k,\ E_k') = \ln\left(\mathrm{mod}\left(\frac{\sqrt{\gamma_k}+\sqrt{\gamma_{k'}}}{\sqrt{\gamma_k}-\sqrt{\gamma_{k'}}}\right)\right)$$

$$\gamma_K = E_k(1+\alpha)E_k \quad E_k = E_k \pm \hbar W_{LO},$$

where W_{LO} is the optical phonon energy and the indices high/low stand for high frequency/static dielectric constants.

7. *Intervalley phonon scattering:* Scattering between equivalent or nonequivalent valleys.

$$W(E) = \left(\frac{\pi D_{ij}^2 Z_j}{\rho W_{ij}}\right) \times \left(n(W_{ij})\frac{1}{2} \mp \frac{1}{2}\right) \times \left(\frac{(2m_d)^{\frac{3}{2}}\sqrt{E_f(1+\alpha E_f)}}{4\pi^2\hbar^3}\right) \times (1+2\gamma E_f) \qquad E_f = F \pm \hbar w_{ij} - \Delta E_{ij},$$

where D_{ij} is the coupling constant, Z_j the number of final valleys, ρ the crystal density, and w_{ij} the energy of the phonon involved in the process.

Sources: M. Lundstrom, Fundamentals of Carrier Transport, Cambridge University Press, London, 2000; B. K. Ridley, *Quantum Processes in Semiconductors, Oxford University Press, 1982; D. Vasileska nanoHUB.org page: nanohub.org/groups/dragica_vasileska/semiclassical* For more details on the derivation of these formulas, please refer to Refs. [26–28].

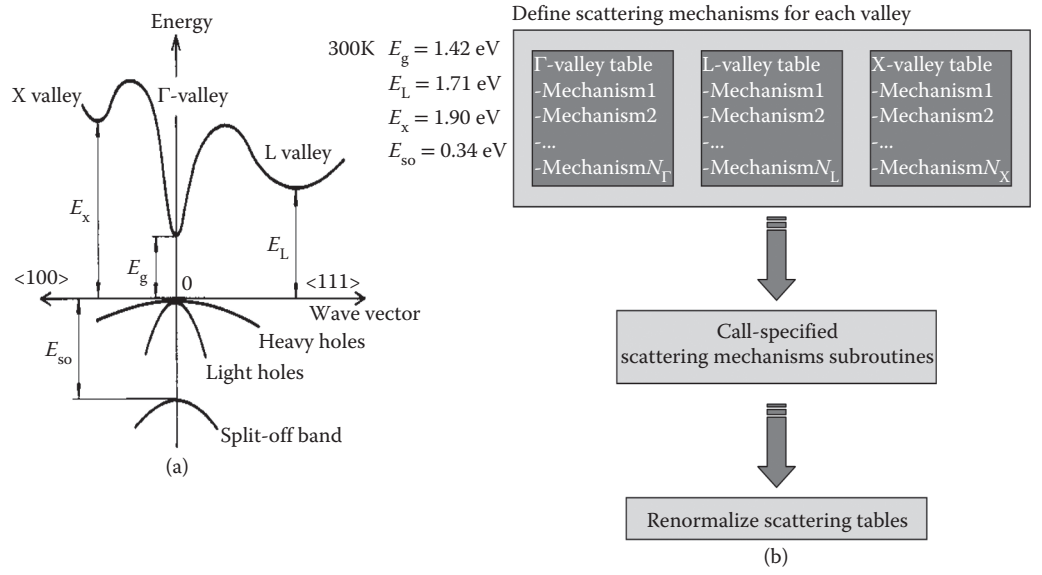


FIGURE 51.7 (a) Schematics of a bandstructure for GaAs. (b) Procedure for the creation of the scattering tables for electron transport.

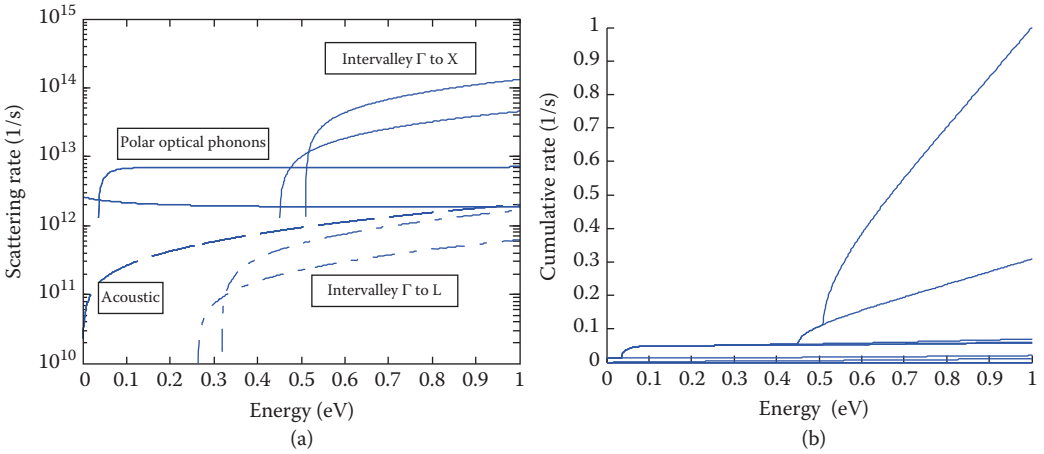


FIGURE 51.8 (a) Scattering rates for the Γ valley. For simplicity, we have omitted Coulomb scattering in these calculations. In (a) the dashed line corresponds to the acoustic phonon scattering rate, solid lines correspond to polar optical phonon scattering (absorption and emission), and the dashed-dotted line corresponds to intervalley scattering from Γ to L valleys. Since the L valley is along the [111] direction, there are eight equivalent directions, and since these valleys are shared, there are a total of four equivalent L valleys. The dotted line corresponds to scattering from Γ to X valleys. The X valleys are in the [100] direction and since there are six equivalent [100] directions and the valleys are shared between Brillouin zones, there are three equivalent X valleys. (b) Normalized cumulative scattering table for the Γ valley. Everything above the top line up to Γ = 1 is self-scattering so it is advisable when checking the scattering mechanisms to first check whether the scattering mechanism chosen is self-scattering or not. This is in particular important for energies below 0.5 eV for this particular scattering table when the Γ to X intervalley scattering (absorption and emission) takes over.

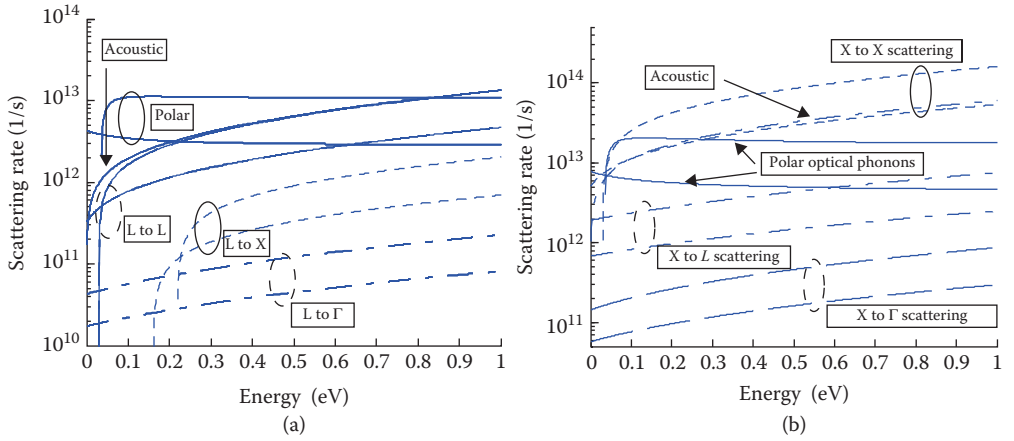


FIGURE 51.9 Scattering rates for the L (a) and X (b) valleys used to create the corresponding normalized scattering tables (not shown here).

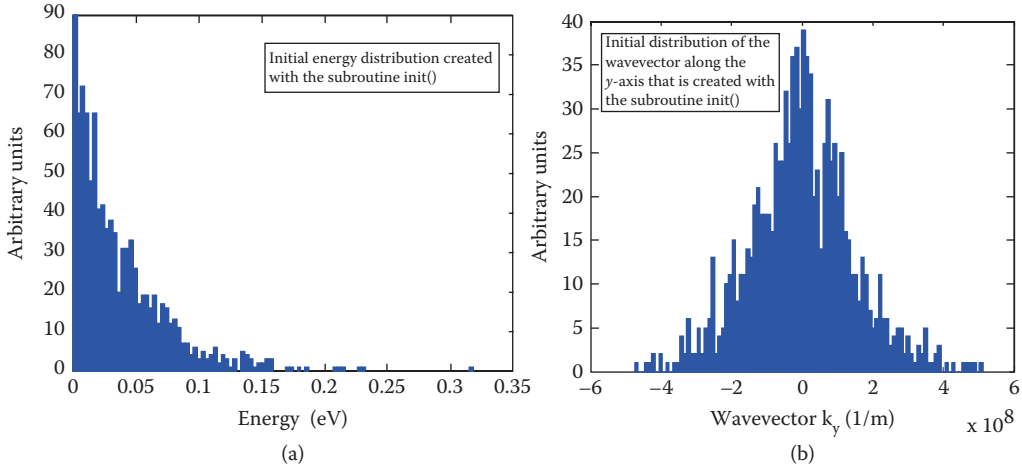


FIGURE 51.10 Initial carrier distribution for an ensemble of 10,000 particles. (a) Energy distribution. (b) Distribution of wavevector k_y.

When the initialization process is finished, the main free-flight-scatter procedure takes place until the completion of the simulation time. There are two components in this process: first, the carriers accelerate freely due to the electric field, accomplished by calling the **drift()** subroutine (which implements Equation 51.2), and then their free-flights are interrupted by random scattering events that are managed by the **scatter_carrier()** subroutine. The flowchart for performing the free-flight-scatter process within one time-step Δt is shown diagrammatically in Figure 51.11.

In the **scatter_carrier()** subroutine from Figure 51.11, first the scattering mechanism terminating the free flight is chosen, to which certain attributes are associated such as the change in energy after scattering. For inelastic scattering processes, we have the change in energy due to emission or absorption of phonons, for example. Also, the nature of the scattering process is identified as isotropic or anisotropic. Note that when performing acoustic phonon and intervalley scattering for GaAs, both of which are isotropic scattering processes, no coordinate system transformation is needed to determine the final wavevector after scattering. Because polar optical phonon and Coulomb scattering mechanisms are anisotropic, it is necessary to do a rotation of the coordinate system, scatter the carrier in the rotated system, and then perform

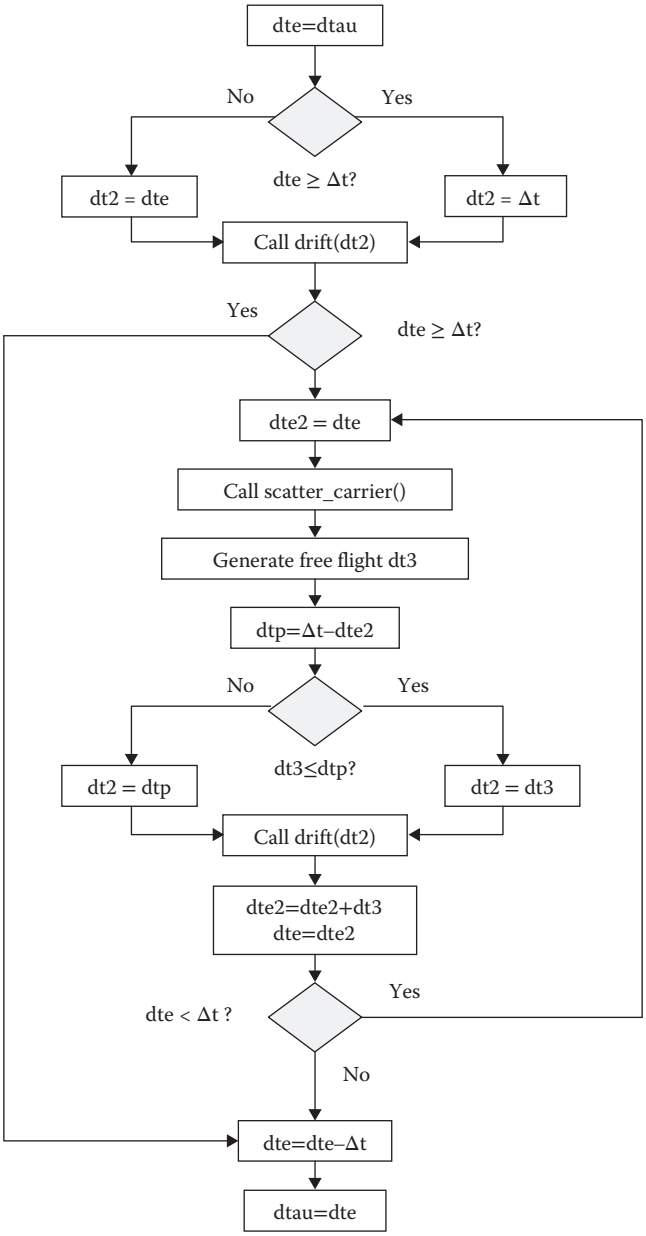


FIGURE 51.11 Free-flight-scatter procedure within one time-step. Time variable dtau is the remaining free-flight time from previous time-step Δt.

inverse coordinate transformation. This procedure is needed because it is much easier to determine final carrier momentum in the rotated coordinate system in which the initial wavevector k is aligned with the z-axis. For this case, one can calculate that the final polar angle of scattering with polar optical phonons for parabolic bands in the rotated coordinate system as

$$\cos\theta = \frac{(1+\xi)-(1+2\xi)^r}{\xi}, \quad \xi = \frac{2\sqrt{E_k\left(E_k \pm \hbar\omega_0\right)}}{\left(\sqrt{E_k}-\sqrt{E_k \pm \hbar\omega_0}\right)^2}, \tag{51.10}$$

where E_k is the carrier energy, $\hbar\omega_0$ the polar optical phonon energy, and r a random number uniformly distributed between 0 and 1. The final angle for scattering with ionized impurities (Coulomb scattering) and for parabolic bands is

$$\cos\theta = 1 - \frac{2r}{1 + 4k^2 L_D^2 (1 - r)}, \tag{51.11}$$

where k is the carrier wavevector and L_D is the Debye screening length. The azimuthal angle for both scattering processes is simply calculated using $\varphi = 2\pi r$. The importance of properly calculating the polar angle θ after scattering to describe small angles of deflection in the case of Coulomb or polar optical phonon scattering is illustrated in Figure 51.13. The histogram of the polar angle (from 0 to $\pi = 3.141592654$) after scattering for electron–polar optical phonon scattering is presented in Figure 51.12, from where we can clearly see the preference for small angle deflections that are characteristic for any Coulomb-type interaction (polar optical phonon is in fact electron–dipole interaction). Graphical representation of the determination of the final angle after scattering for both isotropic and anisotropic scattering processes is shown in Figure 51.13.

The direct technique described above can be applied when the integrals describing $\cos\theta$ can be analytically calculated. For most cases of interest, the integral cannot be easily inverted. In these cases, a rejection technique may be employed. The procedure for the rejection technique goes as follows:

- Choose a maximum value C, such that $C > f(x)$ for all x in the interval (a, b).
- Choose pairs of random numbers, one between a and b $\left(x_1 = a + r_1(b - a)\right)$ and another one$f_1 = r_1' C$ between 0 and C, where r_1 and r_1' are random numbers uniformly distributed between 0 and 1.
- If $f_1 \leq f(x_1)$, then the number x_1 is accepted as a suitable value, otherwise it is rejected.

The three steps described above are schematically shown in Figure 51.14. For $x = x_1$, $r_1 C$ is larger than $f(x_1)$ and in this case, if this represents the final polar angle for scattering, this angle is rejected and a new sequence of two random numbers is generated to determine x_2 and $r_2 C$. In this second case, $f(x_2) > r_2 C$ and the polar angle $\theta = x_2$ is selected (for polar angle selection $a = 0$ and $b = \pi$).

After the simulation is completed, typical results to check are the velocity–time, the energy–time, and the valley occupation versus time characteristics, such as those shown in Figure 51.15, where the velocity–time characteristics for applied electric fields ranging from 0.5 to 7 kV/cm, with an electric field increment

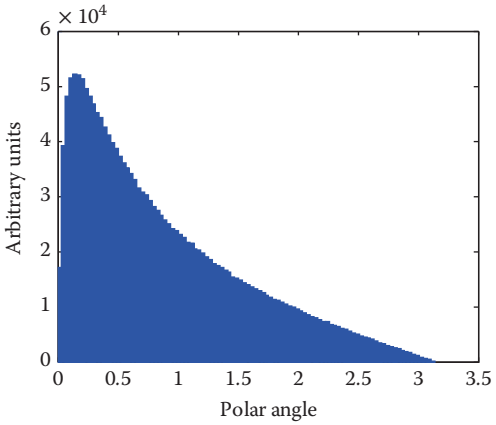
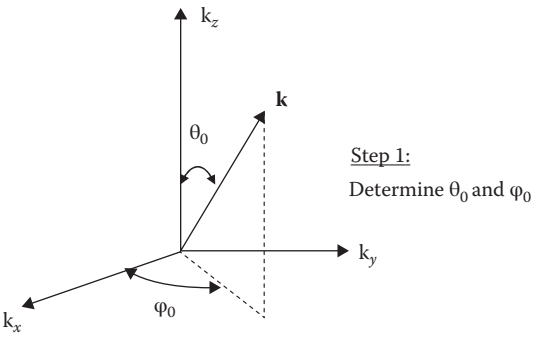


FIGURE 51.12 Histogram of the polar angle for electron–polar optical phonon scattering.

1. Isotropic scattering processes

$$\cos\theta = 1 - 2r,\ \varphi = 2\pi r$$

2. Anisotropic scattering processes (Coulomb, POP)

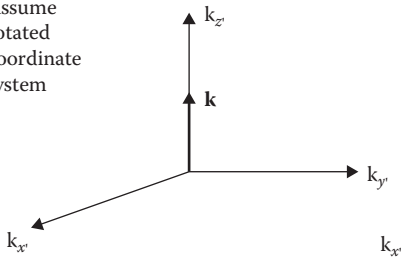


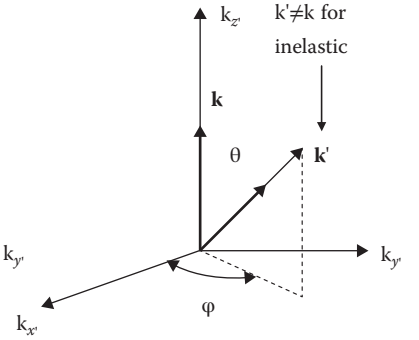
Step 1:

Determine θ_0 and φ_0

Step 2:

Assume rotated coordinate system





Step 3:

perform scattering

$$\cos\theta = \frac{(1+\xi)-(1+2\xi)^r}{\xi},\quad \xi = \frac{2\sqrt{E_k(E_k \pm \hbar\omega_0)}}{\left(\sqrt{E_k}-\sqrt{E_k \pm \hbar\omega_0}\right)^2}\ \text{POP}$$

$$\cos\theta = 1 - \frac{2r}{1 + 4k^2L^2_D(1-r)}\ \text{Coulomb}$$

$\varphi = 2\pi r$ for both

Step 4:

$k_{xp} = k'\sin\theta\cos\varphi$, $k_{yp} = k'\sin\theta^*\sin\varphi$, $k_{zp} = k'\cos\theta$

Return back to the original coordinate system:

$k_x = k_{xp}\cos\varphi_0\cos\theta_0 - k_{yp}\sin\varphi_0 + k_{zp}\cos\varphi_0\sin\theta_0$

$k_y = k_{xp}\sin\varphi_0\cos\theta_0 + k_{yp}\cos\varphi_0 + k_{zp}\sin\varphi_0\sin\theta_0$

$k_z = -k_{xp}\sin\theta_0 + k_{zp}\cos\theta_0$

FIGURE 51.13 Description of final angle selection for isotropic and anisotropic scattering processes using the direct technique.

of 0.5 kV/cm, are shown. These clearly demonstrate that after a transient phase, the system reaches a stationary steady state, after which time we can start taking averages for calculating steady-state quantities.

From the results shown in Figure 51.15, one can see that steady state is achieved for larger time intervals when the electric field value is increased and the carriers are still sitting in the Γ valley. Afterward, time needed to get to steady state decreases. This trend is related to the valley repopulation and movement of the

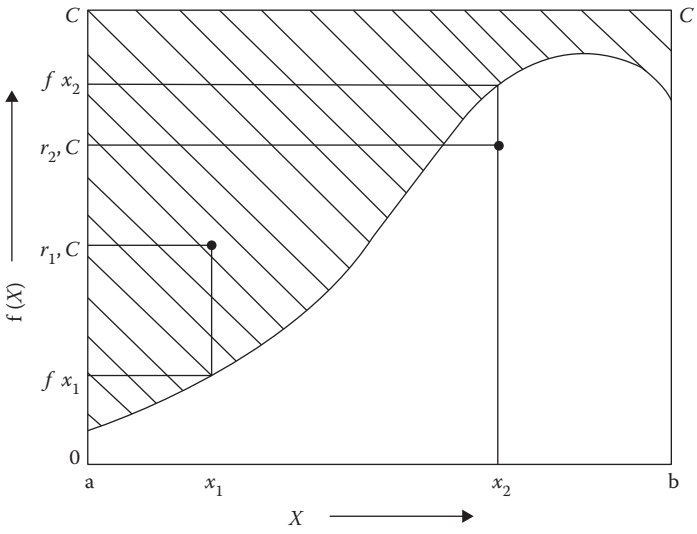


FIGURE 51.14 Schematic description of the rejection technique.

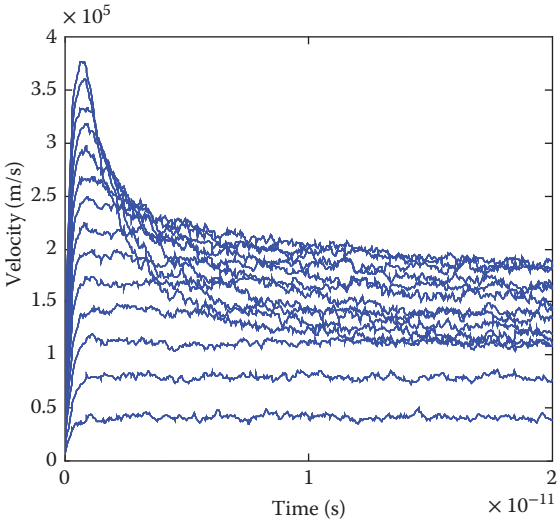


FIGURE 51.15 Time evolution of the drift velocity for electric field strengths ranging between 0.5 and 7 kV/cm, in 0.5 kV/cm increments.

carriers from Γ, into the X and finally into the L valley. The steady-state velocity-field and valley population versus electric field characteristics are shown in Figures 51.16 and 51.17, respectively. One can clearly see on the velocity–field characteristics that a low field mobility of about 8000 cm^2/V-s is correctly reproduced for GaAs without the use of any adjustable parameters.

At this point, it is advisable to check the energy and wavevector histograms (Figure 51.18) to ensure that the energy range chosen in the scattering tables is correct or not for the particular maximum electric field strength being considered, which gives the worst case scenario. Since, as already noted, we apply the electric field in the *y*-direction, for comparative purposes, we plot the histograms of the *x*-component of the wavevector, *y*-component of the wavevector, and the histogram of the final carrier energy distribution

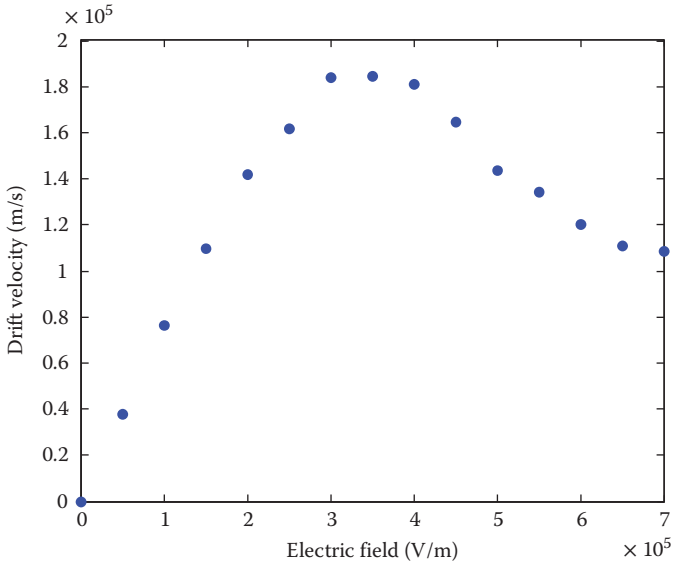


FIGURE 51.16 Steady-state drift velocity versus electric field.

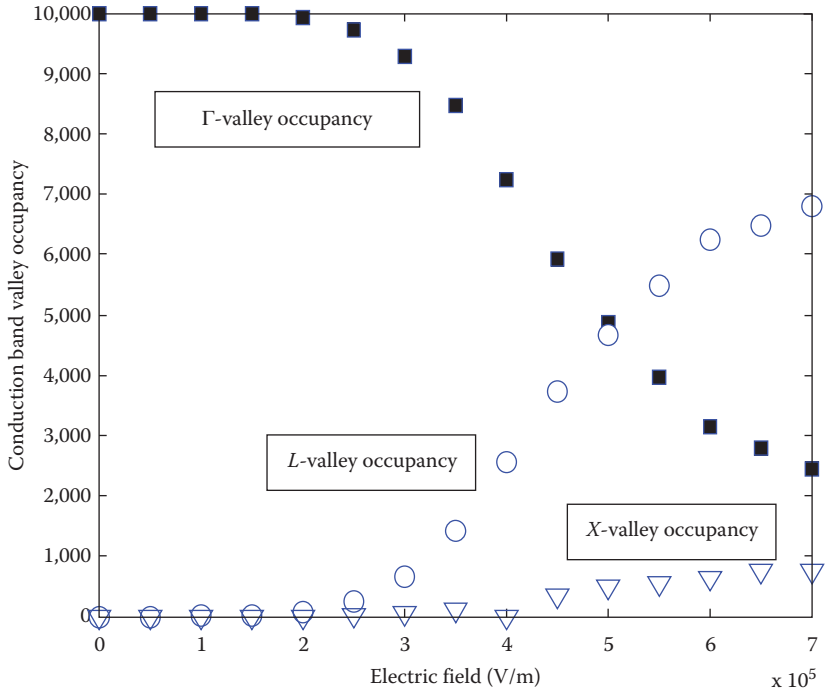


FIGURE 51.17 Different valley occupancy versus electric field.

for which a drifted Maxwellian form is evident. Since there is no field applied in the x-direction, we see that the average wavevector in the x-direction is 0. Due to the application of the field in the y-direction, there is a finite positive shift in the y-component of the velocity, which is yet another signature for the displaced Maxwellian form of the energy distribution in the bottom histogram.

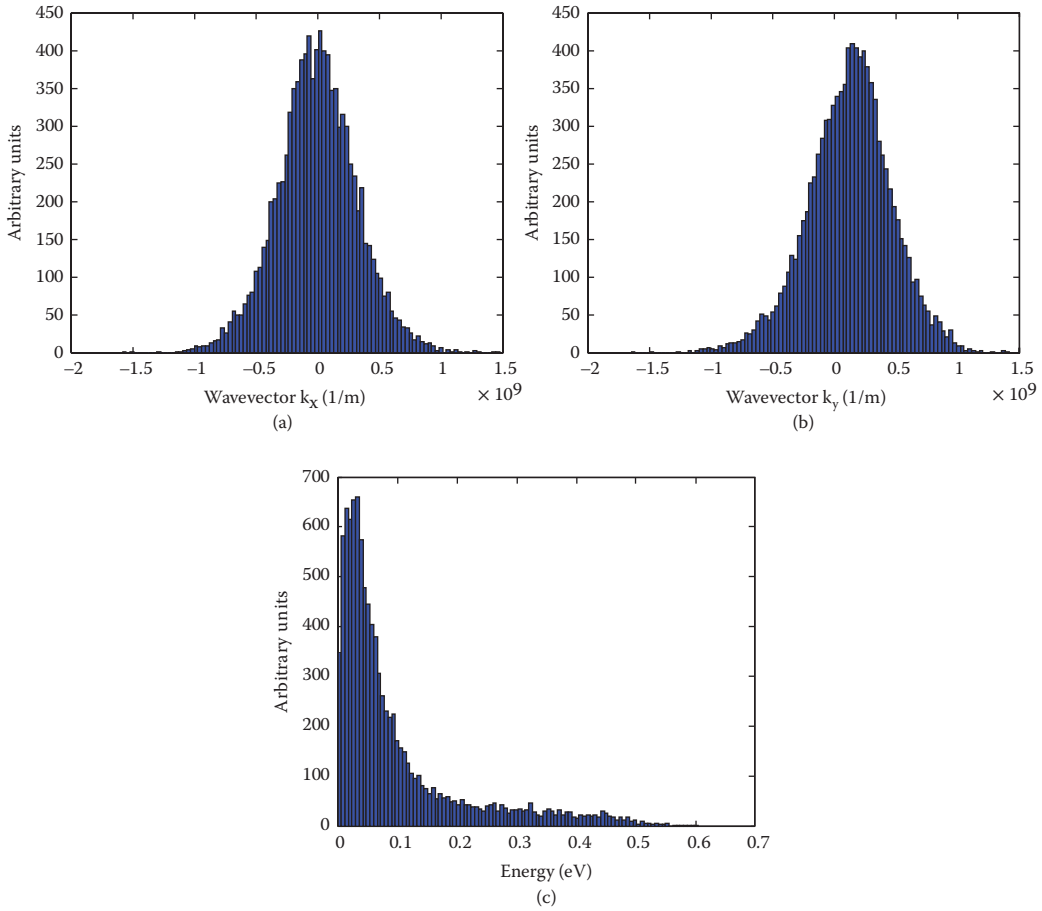


FIGURE 51.18 (a) Histogram of the x-component of the wavevector. (b) Histogram of the y-component of the wavevector. (c) Histogram of the carrier energy. Applied electric field is 7 kV/cm.

51.3 Particle-Based Device Simulation

In Section 51.2, we introduced the numerical solution of the BTE using Monte Carlo method. Within a device, both the transport kernel and the field solver are coupled to each other. In Poisson's equation, the field associated with the potential is the driving force accelerating particles in the Monte Carlo phase, for example, while the distribution of mobile (both electrons and holes) and fixed charges (e.g., donors and acceptors) provides the source of the electric field. Below, we give a brief description of the Monte Carlo particle-based device simulator development.

Within the particle-based EMC method, with its time-marching algorithm, Poisson's equation may be decoupled from the BTE over a suitably small time step (typically less than the inverse plasma frequency corresponding to the highest carrier density in the device). Over this time interval, carriers accelerate according to the frozen field profile from the previous time-step solution of Poisson's equation, and then Poisson's equation is solved at the end of the time interval with the frozen configuration of charges arising from the Monte Carlo phase [6,26]. Note that Poisson's equation is solved on a mesh, whereas the solution of charge motion using EMC occurs over a continuous range of coordinate space in terms of the particle position. Therefore, a particle-mesh (PM) coupling is needed for both the charge assignment and the force

interpolation. The PM coupling is broken into four steps: (1) assign particle charge to the mesh, (2) solve the Poisson equation on the mesh, (3) calculate the mesh-defined forces, and (4) interpolate to find forces on the particle.

The motion in real space of particles under the influence of electric fields is somewhat more complicated due to the band structure. The velocity of a particle in real space is related to the E-$\mathbf{k}$ dispersion relation defining the bandstructure as

$$\begin{aligned} \mathbf{v}(t) &= \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{1}{\hbar}\nabla_k E(\mathbf{k}(t)) \\ \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} &= \frac{q\mathbf{E}(\mathbf{r})}{\hbar} \end{aligned}, \tag{51.12}$$

where the rate of change of the crystal momentum is related to the local electric field acting on the particle through the acceleration theorem expressed by the second equation. In turn, the change in crystal momentum, $\mathbf{k}(t)$, is related to the velocity through the gradient of E with respect to $\mathbf{k}$. If one has to use the full bandstructure of the semiconductor, then integration of these equations to find $\mathbf{r}(t)$ is only possible numerically, using, for example, a Runge–Kutta algorithm [27].

To simulate the steady-state behavior of a device, the system must be initialized in some initial condition, with the desired potentials applied to the contacts, and then the simulation proceeds in a time stepping manner until steady state is reached. This process may take several picoseconds of simulation time, and consequently several thousand time-steps based on the usual time increments required for stability. Clearly, the closer the initial state of the system is to the steady-state solution, the quicker the convergence. If one is, for example, simulating the first bias point for a transistor simulation, and has no *a priori* knowledge of the solution, a common starting point for the initial guess is to start out with charge neutrality, i.e., to assign particles randomly according to the doping profile in the device and based on the super-particle charge assignment of the particles, so that initially the system is charge neutral on the average. (For 2D device simulation, one should keep in mind that each particle actually represents a rod of charge into the third dimension.) Subsequent simulations at the same device at different bias conditions can use the steady-state solution at the previous bias point as a good initial guess. After assigning charges randomly in the device structure, charge is then assigned to each mesh point using the nearest-grid-point (NGP), cloud-in-cell (CIC), or nearest-element-cell (NEC) PM coupling methods, and Poisson's equation is solved. The forces are then interpolated on the grid, and particles are accelerated over the next time-step. A flowchart of a typical Monte Carlo device simulation is shown in Figure 51.19.

As the simulation evolves, charge will flow in and out of the contacts, and depletion regions internal to the device will form until steady state is reached. The charge passing through the contacts at each time step can be tabulated, and a plot of the cumulative charge as a function of time gives the steady-state current. Figure 51.20 shows the particle distribution in three dimension of a metal-semiconductor field-effect transistor (MESFET), where the dots indicate the individual simulated particles for two different gate biases. Here, the heavily doped MESFET region (shown by the inner box) is surrounded by semi-insulating GaAs forming the rest of the simulation domain. Figure 51.20a corresponds to no net gate bias (i.e., the gate is positively biased to overcome the built-in potential of the Schottky contact), while Figure 51.20b corresponds to a net negative bias applied to the gate, such that the channel is close to pinch-off. One can see the evident depletion of carriers under the gate under the latter conditions.

51.3.1 Calculation of the Current

The device output current can be determined using two different yet consistent methods. First, by keeping track of the charges entering and exiting each terminal/contact, the net number of charges over a period of the simulation can be used to calculate the terminal current. The net charge crossing a terminal boundary

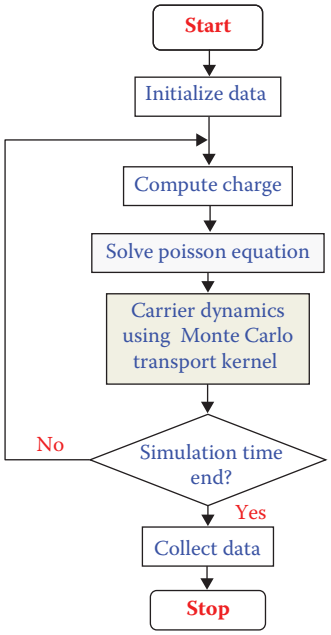


FIGURE 51.19 Flowchart of a typical particle-based device simulation.

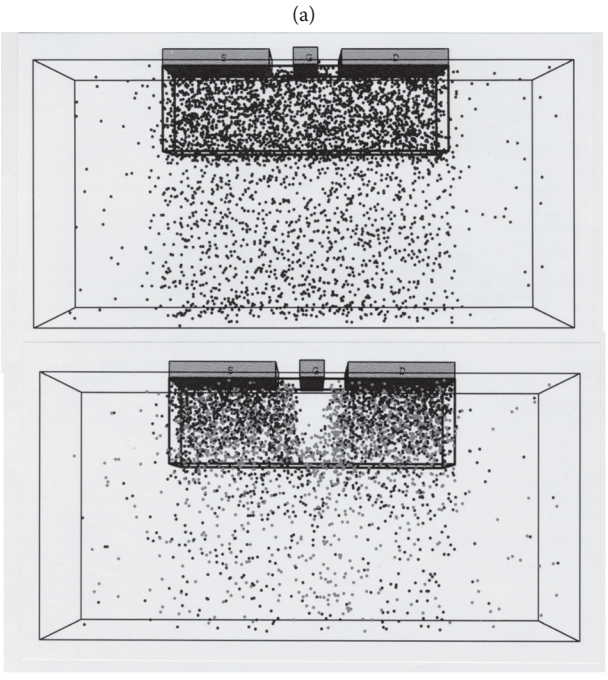


FIGURE 51.20 Example of the particle distribution in an MESFET structure simulated in three dimension using an EMC approach. (a) The device with zero gate voltage applied and (b) the device with a negative gate voltage applied, close to pinch-off.

is determined by

$$Q(t) = e(n_{\text{abs}}(t) - n_{\text{injec}}(t)) + \varepsilon \int E_y(x,t)\mathrm{d}y, \tag{51.13}$$

where n_{abs} is the number of particles that are absorbed by the contact (exit), n_{injec} the number of particles that have been injected at the contact, and E_y the vertical field at the contact. The second term in Equation 51.13 on the right-hand side is used to account for the displacement current due to the changing field at the contact. Equation 51.13 assumes the contact is at the top of the device and that the fields in the x- and z- directions are negligible. The charge e in Equation 51.13 should be multiplied by the particle charge if it is not unity. The slope of $Q(t)$versus time gives a measure of the terminal current. In steady state, the current can be found by

$$I = \frac{\mathrm{d}Q(t)}{\mathrm{d}t} = \frac{e(n_{\text{net}})}{\Delta t}, \tag{51.14}$$

where $n_{\text{net}} = n_{\text{abs}} - n_{\text{injec}}$ is the net number of particles exiting the contact over a fixed period of time Δt.

In a second method, the sum of the electron velocities in a portion of the channel region of the device is used to calculate the current. The electron current density through a cross-section of the device is given by

$$J = env_{\text{d}}, \tag{51.15}$$

where v_{d} is the average electron drift velocity and n the carrier concentration. If there are a total of N particles in a differential volume, $\mathrm{d}V = \mathrm{d}L \cdot \mathrm{d}A$,the current found by integrating Equation 51.15 over the cross-sectional area, $\mathrm{d}A$, is

$$I = \frac{eNv_{\text{d}}}{\mathrm{d}L}, \text{ or } I = \frac{e}{\mathrm{d}L}\sum_{i=1}^{N} v_x(i), \tag{51.16}$$

where $v_x(i)$ is the velocity along the channel of the ith electron. The device is divided into several sections along the x-axis, and the number of electrons and their corresponding velocity is added for each section after each free flight. The total x-velocity in each section is then averaged over several time-steps to determine the current for that section. Total device current can be determined from the average of several sections, which gives a much smoother result compared to counting terminal charges. By breaking the device into sections, individual currents can be compared to verify that there is conservation of particles (constant current) throughout the device. In addition, sections near the source and drain regions may have a high y-component in their velocity and should be excluded from the current calculations. Finally, by using several sections in the channel, the average energy and velocity of electrons along the channel can be observed to ensure the proper physical characteristics. The two methods for the calculation of the current are performed, as illustrated in Figure 51.17, on a 50-nm channel length MOSFET device.

Extrapolating the slope of the curve shown in Figure 51.21a, that represents the cumulative electron charge that enters/exits the source/drain contact, leads to source/drain current of 0.5205/0.5193 mA/μm. When compared with the results shown in Figure 51.21b, it is evident that both current measurement techniques discussed in this section give current values with relative error less than 1%.

51.3.2 Ohmic Contacts

Another issue that has to be addressed in particle-based simulations is the real space boundary conditions for the particle part of the simulation. Reflecting boundary conditions are usually imposed at the artificial

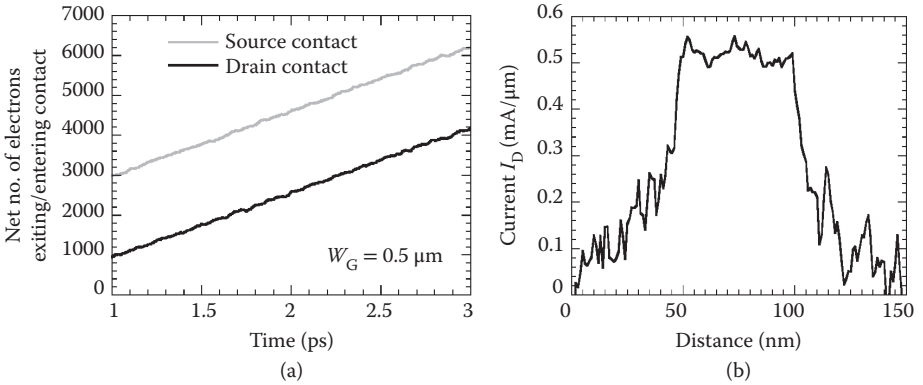


FIGURE 51.21 (a) Net charge entering/exiting the source/drain contact. (b) Average current along the channel. The gate length of the device being modeled equals 50 nm. $V_G = 1.4$ V and $V_D = 1$ V in these simulations. Width of the gate is $W = 0.5$ μm.

boundaries. As far as the ohmic contacts are concerned, they require more careful consideration because electrons crossing the source and drain contact regions contribute to the corresponding terminal current. Commonly employed models for the contacts include the following [28]:

- Electrons are injected at the opposite contact with the same energy and wavevector **k**. If the source and drain contacts are in the same plane, as in the case of MOSFET simulations, the sign of **k**, normal to the contact will change. This is an unphysical model, however [29].
- Electrons are injected at the opposite contact with a wavevector randomly selected based upon a thermal distribution. This is also an unphysical model.
- Contact regions are considered to be in thermal equilibrium. The total number of electrons in a small region near the contact are kept constant, with the number of electrons equal to the number of dopant ions in the region. This is a very good model most commonly employed in actual device simulations.
- Another method uses "reservoirs" of electrons adjacent to the contacts. Electrons naturally diffuse into the contacts from the reservoirs, which are not treated as part of the device during the solution of Poisson's equation. This approach gives results similar to the velocity weighted Maxwellian but at the expense of increased computational time due to the extra electrons simulated. It is an excellent model employed in few most sophisticated particle-based simulators.

There are also several possibilities for the choice of the distribution function—Maxwellian, displaced Maxwellian, and velocity-weighted Maxwellian [10].

51.3.3 Time-Step

As in the case of solving the drift-diffusion, hydrodynamic, or full Maxwell's equations, for a stable Monte Carlo device simulation, one has to choose the appropriate time step, Δt, and the spatial mesh size (Δx, Δy, and/or Δz). The time-step and the mesh size may correlate to each other in connection with the numerical stability. For example, as discussed in the context of solving drift-diffusion simulations, the time-step Δt must be related to the plasma frequency:

$$\omega_p = \frac{e^2 n}{\varepsilon_s m^*}, \tag{51.17}$$

where n is the carrier density. From the viewpoint of the stability criterion, Δt must be much smaller than the inverse plasma frequency. The highest carrier density specified in the device model is used to estimate

Δt. If the material is a multivalley semiconductor, the smallest effective mass to be experienced by the carriers must be used in Equation 51.17 as well. In the case of GaAs, with the doping of 5×10^{17} cm^{-3}, $\omega_p \cong 5 \times 10^{13}$; hence, Δt must be smaller than 0.02 ps.

The mesh size for the spatial resolution of the potential is dictated by the charge variations. Hence, one has to choose the mesh size to be smaller than the smallest wavelength of the charge variations. The smallest distance is approximately equal to the Debye length, given as

$$\lambda_D = \sqrt{\frac{\varepsilon_s k_B T}{e^2 n}}. \tag{51.18}$$

The highest carrier density specified in the model should be used to estimate λ_D from the stability criterion. The mesh size must be chosen to be smaller than the value given by Equation 51.18. In the case of GaAs, with the doping density of 5×10^{17} cm^{-3}, $\lambda_D \cong 6$ nm.

Based on the discussion above, the time step (Δt) and the mesh size (Δx, Δy, and/or Δz) can be specified separately. However, the Δt chosen must be checked again by calculating the distance l_{max}, defined as

$$l_{max} = \mathbf{v}_{max} \times \Delta t, \tag{51.19}$$

where $\mathbf{v}_{max}$ is the maximum carrier velocity that can be approximated by the maximum group velocity of the electrons in the semiconductor (on the order of 10^8 cm/s). Therefore, the distance l_{max} is regarded as the maximum distance the carriers can propagate during Δt. The time step chosen must be small enough so that l_{max} is smaller than the spatial mesh size chosen using Equation 51.19. This is because large Δt chosen may cause substantial change in the charge distribution, while the field distribution in the simulation is only updated every Δt.

51.3.4 PM Coupling

As mentioned earlier, the position of charge as described by the EMC algorithm is continuous, whereas Poisson's equation is solved on a mesh, hence the charge associated with the individual particles must be mapped onto the field mesh in some manner. The charge assignment and force interpolation schemes usually employed in self-consistent Monte Carlo device simulations are the NGP, CIC, and NEC schemes [30]. In the NGP scheme, the particle position is mapped into the charge density at the closest grid point to a given particle. This has the advantage of simplicity, but leads to a noisy charge distribution, which may exacerbate numerical instability. Alternately, within the CIC scheme, a finite volume is associated with each particle spanning several cells in the mesh, and a fractional portion of the charge per particle is assigned to grid points according to the relative volume of the "cloud" occupying the cell corresponding to the grid point. This method has the advantage of smoothing the charge distribution due to the discrete charges of the particle-based method, but may result in an artificial "self-force" acting on the particle, particularly if an inhomogeneous mesh is used. For the case of inhomogeneous mesh or spatially varying dielectric permittivity, the NEC scheme is the best choice. Within the NEC scheme, the charge is moved at the center of the cell and then CIC-like charge assignment to the node points is performed. The PM coupling sequence is presented in Figure 51.22.

51.4 Representative Simulation Results

51.4.1 Bulk Monte Carlo Simulations of Different Materials

In most semiconductors, in order to properly simulate high field transport, it is necessary to consider more than one conduction band valley. To calculate the drift velocity along any direction, the effective mass along

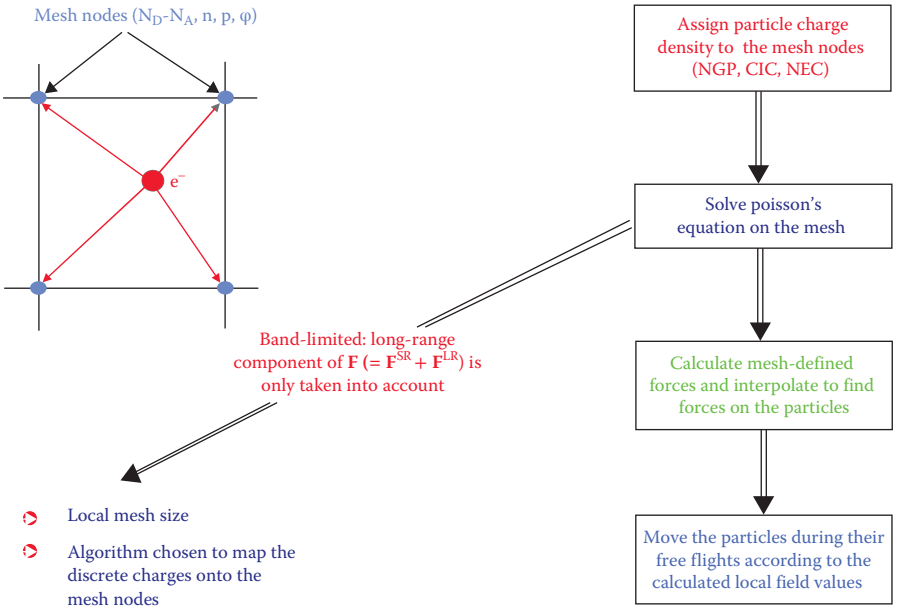


FIGURE 51.22 Particle-mesh coupling sequence. (SR, short range; LR, long range.)

that particular direction is required. In GaAs, for example, a subvalley of the L valley lies along the [111] direction. We know the effective masses along the transverse and longitudinal directions of the subvalley but we do not know the effective mass along the [100] direction. This makes it hard to calculate the drift velocity along the [100] direction.

Assume the Monte Carlo simulation is run on the x, y, z- coordinate system where the x-direction is [100], y-direction is [010], and z-direction is [001]. Let the three perpendicular directions that describe the subvalley be $[a_1, b_1, c_1]$, $[a_2, b_2, c_2]$, and $[a_3, b_3, c_3]$. The electrons in the Monte Carlo simulation will be drifted according to the x, y, z coordinate system so there will be k_x the wavevector along [100], k_y the wavevector along [010], and k_z the wavevector along [001]. The drift velocity of the electron is first calculated along the three directions that describe the subvalley and along which we know the effective masses.

$$v_{\mathrm{d}}, [[a]_1 b_1 c_1] = \frac{\hbar k_{[[a]_1 b_1 c_1]}}{m_1 (1 + 2\alpha E)} v_d, [[a]_1 b_1 c_1] = \frac{\hbar k_{[[a]_1 b_1 c_1]}}{m_1 (1 + 2\alpha E)} \tag{51.20}$$

Here m_1 is the effective mass of the electron along $[a_1, b_1, c_1]$. The final expression of drift velocity along that direction is then calculated by calculating $k_{[[a]_1 b_1 c_1]} k_{[[a]_1 b_1 c_1]}$ using a simple transformation of coordinates to give [31]

$$k_{[[a]_1 b_1 c_1]} = \frac{k_x a_1}{\sqrt{(a_1^2 + b_1^2 + c_1^2)}} + \frac{k_y b_1}{\sqrt{(a_1^2 + b_1^2 + c_1^2)}} + \frac{k_z a_1}{\sqrt{(a_1^2 + b_1^2 + c_1^2)}}. \tag{51.21}$$

Using similar methods we get $v_{\mathrm{d}}, \left[[a]_2 b_2 c_2\right]$ and $v_{\mathrm{d}}, \left[[a]_3 b_3 c_3\right]$. The coordinates system is then transformed once again back to the x,y,z coordinate system to get the drift velocities along x,y, and z.

$$
\begin{aligned}
v_x &= \frac{a_1 v_{\mathrm{d},[[a]_1 b_1 c_1]}}{\sqrt{(a_1^2 + b_1^2 + c_1^2)}} + \frac{a_2 v_{\mathrm{d},[[a]_2 b_2 c_2]}}{\sqrt{(a_2^2 + b_2^2 + c_2^2)}} + \frac{a_3 v_{\mathrm{d},[[a]_3 b_3 c_3]}}{\sqrt{(a_3^2 + b_3^2 + c_3^2)}} \\
v_y &= \frac{b_1 v_{\mathrm{d},[[a]_1 b_1 c_1]}}{\sqrt{(a_1^2 + b_1^2 + c_1^2)}} + \frac{b_2 v_{\mathrm{d},[[a]_2 b_2 c_2]}}{\sqrt{(a_2^2 + b_2^2 + c_2^2)}} + \frac{b_3 v_{\mathrm{d},[[a]_3 b_3 c_3]}}{\sqrt{(a_3^2 + b_3^2 + c_3^2)}} \\
v_x &= \frac{c_1 v_{\mathrm{d},[[a]_1 b_1 c_1]}}{\sqrt{(a_1^2 + b_1^2 + c_1^2)}} + \frac{c_2 v_{\mathrm{d},[[a]_2 b_2 c_2]}}{\sqrt{(a_2^2 + b_2^2 + c_2^2)}} + \frac{c_3 v_{\mathrm{d},[[a]_3 b_3 c_3]}}{\sqrt{(a_3^2 + b_3^2 + c_3^2)}}
\end{aligned}
\tag{51.22}
$$

Also, as the three directions are mutually perpendicular we have

$$
\begin{aligned}
a_1 a_2 + b_1 b_2 + c_1 c_2 &= 0 \\
a_1 a_3 + b_1 b_3 + c_1 c_3 &= 0 \\
a_3 a_2 + b_3 b_2 + c_3 c_2 &= 0.
\end{aligned}
\tag{51.23}
$$

For N electrons in the simulation, an average drift velocity is then calculated as

$$
\langle v_x \rangle = \frac{1}{N} \sum_{i=1}^{N} v_{x,i}, \quad \left\langle v_y \right\rangle = \frac{1}{N} \sum_{i=1}^{N} v_{y,i}, \quad \langle v_z \rangle = \frac{1}{N} \sum_{i=1}^{N} v_{z,i}, \tag{51.24}
$$

where each drift velocity depends on the subvalley the electron is currently in. The mobility is calculated as the slope of the velocity versus field curve for low fields. The range of "low fields" varies from material to material, but in all cases, it is the range of fields for which the velocity linearly varies with the field.

The results presented in Figure 51.23 shows the difference in velocity of electrons for different materials. For GaN, the peak velocity of 3×10^7 cm/s occurs at around 150 kV/cm so that in the electric field range shown in Figure 51.19 it is clear that for electric fields below 30 kV/cm, GaN carriers in the bulk GaN materials are still in the low-field regime.

As the drift velocities saturate at much smaller electric fields in Si and Ge, they are plotted on a separate graph (see Figure 51.24) [32].

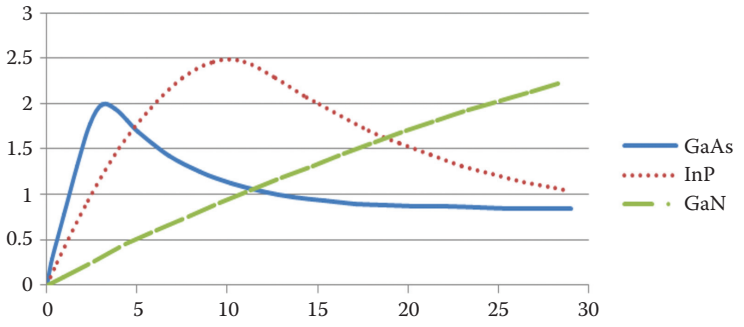


FIGURE 51.23 Drift velocity (10^7 cm/s) versus electric field (kV/cm).

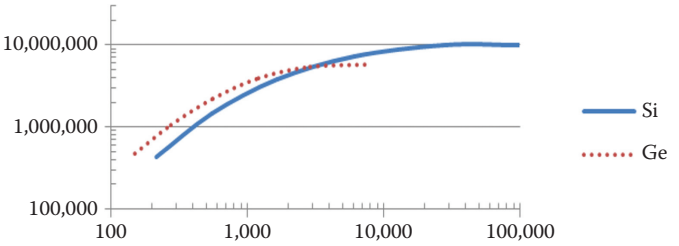


FIGURE 51.24 Electron drift velocity (cm/s) versus electric field (V/cm) in Si and Ge along the principal crystallographic directions.

51.4.2 Modeling Thermal Effects in SOI Devices Using Particle-Based Device Simulator

There are many applications of the Monte Carlo approach to device simulations that can be found in the literature. Here, we show the example of Monte Carlo device simulation to self-heating which requires coupling of the Monte Carlo device solver self-consistently with a solver for the phonon bath. Self-heating (despite the fact that supply voltages are reduced when reducing the size of the channel of nanoscale transistors) is still a problem, since there are sufficiently large electric fields in the smallest devices that can accelerate the carriers to average energies of 0.5–0.8 eV. These hot carriers then interact with the lattice and transfer their energy to the phonon bath (both acoustic and optical), thus heating the lattice and increasing the lattice vibrations (scattering). This leads to a negative feedback on the carrier drift velocity or mobility (if transport is scattering dominated). Thus, we may conclude that when transistors are in the ON-state, there is a strong interaction of the electrons with the lattice on a very short time scale (in particular with the optical phonons). The optical phonons in a longer time scale decay into acoustic phonons through a harmonic multiple phonon processes, thus removing the heat from the hot spot. Therefore, to properly treat the operation of nanoscale devices it is necessary to, at least, consider the electrons within the BTE picture and, in the lowest approximation, via energy balance equations including the acoustic and the optical phonon baths separately. This is what the Arizona State University (ASU) group has implemented couple of years ago [33]. Our ultimate goal is to have coupled the developed phonon BTE solver [34] with the electron BTE solver in a self-consistent manner.

51.4.2.1 Theoretical Model and Computation Details

A schematic description of our 2D/3D electrothermal simulator is shown in Figure 51.25. It is important to note that the EMC transport kernel for the electrons is in the loop for the solution of the energy balance equations for the acoustic and optical phonon temperatures. The criterion for convergence of the whole self-consistent loop (measured in terms of Gummel cycles) is convergence in the current to the third significant digit.

Between iterations, there are variables that are being exchanged between the electron and phonon solvers. Namely, after running the EMC for the electrons in the device for about 5–10 ps for 2D device analysis and 10 ps for the 3D nanowire simulations, the electron transport solver passes to the phonon energy balance solver the information about the spatial variation of the electron density, spatial variation of the average drift velocity, and spatial variation of the average electron temperature, where an assumption is made that the drift energy is much smaller than the thermal energy. These variables are then used in the solution of the energy balance equations for the acoustic and optical phonons to get updated values for the spatial variation of the optical phonon and acoustic phonon temperatures, which are then used as inputs, in the choice of the scattering table for the electrons (proper scattering table is chosen with the acoustic and optical phonon temperatures at the NGP). Then, when the proper table is selected, based on the energy of the electron and a selection of a random number, the scattering mechanism is selected.

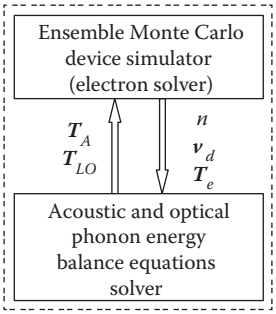


FIGURE 51.25 Exchange of variables between the two kernels. The electron density (n), electron drift velocity (v_d), and electron temperature (T_e), obtained from the electron solver, are input variables for the phonon kernel. The output variables of the phonon solver are acoustic and optical phonon temperature profiles (T_A and T_{LO}, respectively). They enter in the beginning of the free-flight-scattering phase through phonon temperature-dependent scattering tables.

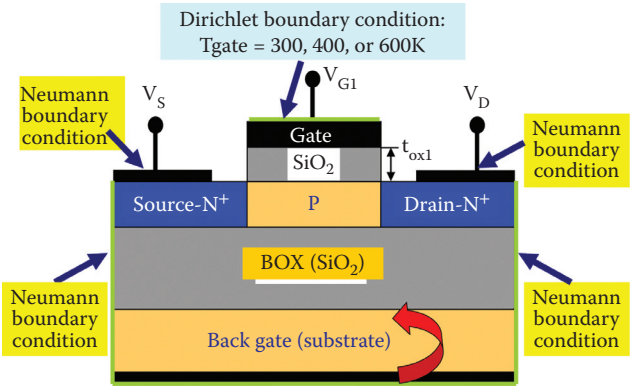


FIGURE 51.26 Schematic of the simulated device structure which illustrates the boundary conditions used in some of the simulations presented. Note also that the boundary conditions at the bottom of the substrate can be mapped as a boundary condition at the bottom of the BOX.

The question of the proper boundary conditions for the electronic part of the problem is rather clear and has been discussed in many papers and texts in the literature [35]. The problem in properly specifying the phonon boundary conditions is the selection of acoustic and optical phonon temperatures either at the artificial boundaries or at the contacts. Typically, acoustic phonon temperature is equated with the lattice temperature. To better understand the choice of boundary conditions we have considered in various works [15], we want to point out that lattice temperature is analogous to electrostatic potential and heat flux is analogous to current. As it is well known, when considering electrical behavior of the device, at least one node within the structure has to have Dirichlet boundary conditions specified (specify the potential for Poisson). Analogously, for the lattice temperature, we need at least one node that is a thermal contact and whose temperature is set to 300 K. As illustrated in Figure 51.26 for the case of FD-SOI devices, the bottom of the BOX is assumed to be at thermal equilibrium. Also, the gate contact is assumed to be at thermal equilibrium, but not necessarily at 300 K. For example, in one set of simulations, we vary the temperature on the gate to be 300, 400, and 600 K. Now comes the question: What happens to the source and drain contacts and the side artificial boundaries? Specifying Dirichlet boundary conditions at the ohmic contacts is not accurate from a standpoint that there is current flowing through the contacts, and since the contacts have finite resistance, there will be Joule heating (so the problem becomes unconstrained). The best solution to this, as we have done in several studies, is to extend the metal contact to become part of the simulation

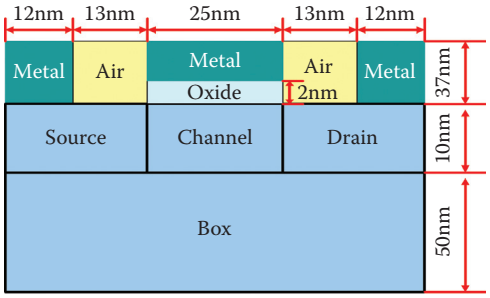


FIGURE 51.27 Device structure with extended domain. The geometrical dimensions are for the simulated 25-nm channel length FD-SOI nMOSFET.

domain and to apply at the very ends of the contact isothermal boundary conditions (see Figure 51.27). With the top and bottom specified, we have vertical transport of heat through the structure.

The next question is, is there lateral transfer of heat across the artificial boundaries? Here we can consider two cases: case when the neighboring device is ON, in which case Neumann boundary conditions are appropriate (derivative of the temperature), and case when the neighboring device is OFF, in which case Dirichlet boundary conditions are appropriate. In the case when Neumann boundary conditions are applied across the artificial boundaries, the heat transport remains vertical, but for the case when Dirichlet boundary conditions are applied across the boundary, the heat transport has both vertical and horizontal component. The case of Neumann boundary conditions can happen in analog circuits such as current mirror in which both transistors are simultaneously ON, whereas the case of Dirichlet boundary conditions occurs in digital circuits in which most of the time the transistors are OFF.

Yet another issue that deserves attention in getting physically correct results is the proper choice of the thermal conductivity for thin Si slabs and for nanowires. Asheghi and coworkers [36] and Li Shi and coworkers [37] have demonstrated via experimental measurements that thermal conductivities of thin Si films and Si nanowires, respectively, strongly depend of the geometry as for smaller geometries phonon boundary scattering can reduce the thermal conductivity of the Si film or nanowire by a factor of 10 or more from its bulk value. Moreover, thermal conductivity is temperature-dependent quantity. We made extensive efforts, using the theory of Sondhaimer for conductivity of metals [38], to arrive at an empirical formula that simultaneously described the thickness and temperature dependence of the thermal conductivity [39]. Our empirical expression perfectly matched the experimental data of Asheghi and coworkers [16]. In the model, it was assumed that phonon boundary scattering was perfectly diffusive.

51.4.2.2 Simulation Results

To validate our premise that self-heating effects become smaller as we scale devices into the nanometer regime (due to more pronounced velocity overshoot effects and thinner BOX layer that leads to smaller thermal resistance), in Table 51.3 we present the parameters of the devices being simulated. Simulation results for the lattice temperature profile in the Si film for devices with channel lengths from 25 to 180 nm are shown in Figure 51.28. We clearly see that the hot spot is in the channel for 180-nm channel length device and moves into the drain contact for 25-nm channel length device. The current degradation for isothermal temperatures on the gate of 300, 400, and 600 K are shown in Figure 51.29. These are the worst case scenario results as we use Neumann boundary conditions at the side boundaries.

In obtaining the results presented in Figures 51.28 and 51.29, constant thermal conductivity value of 13 W/m-K was used in the simulations [40]. The use of constant thermal conductivity model with value of 13 W/m-K leads to underestimation (25 nm channel length) or overestimation (180 nm channel length) of the average maximum temperature of the lattice in the channel region of the device.

TABLE 51.3 Parameters for Various Simulated Device Technology Nodes (Constant Field Scaling)

L(nm)	t_{ox} (nm)	t_{si} (nm)	t_{box} (nm)	N_{ch} (cm^{-3})	$V_{GS} = V_{DS}$ (V)	I_D (mA/ μm)
25	2	10	50	1×10^{18}	1.2	1.82
45	2	18	60	1×10^{18}	1.2	1.41
60	2	24	80	1×10^{18}	1.2	1.14
80	2	32	100	1×10^{17}	1.5	1.78
90	2	36	120	1×10^{17}	1.5	1.67
100	2	40	140	1×10^{17}	1.5	1.57
120	3	48	160	1×10^{17}	1.8	1.37
140	3	56	180	1×10^{17}	1.8	1.23
180	3	72	200	1×10^{17}	1.8	1.03

Note: *Parameters of the simulated structure given in Figure 51.26 are L, gate length; t_{ox}, gate oxide thickness; t_{Si}, active Si-layer thickness; t_{box}, BOX thickness; N_{ch}, channel doping concentration; I_D, isothermal current value (300 K).*

51.4.2.3 Summary of Results

In this section, we have presented modeling of self-heating effects in FD-SOI devices. Larger velocity overshoot and smaller BOX thickness lead to smaller degradation due to self-heating effects in 25-nm channel length FD-SOI devices. The amount of self-heating significantly depends upon the magnitude of the thermal conductivity used. Bulk values are inadequate and proper thickness and temperature dependence of

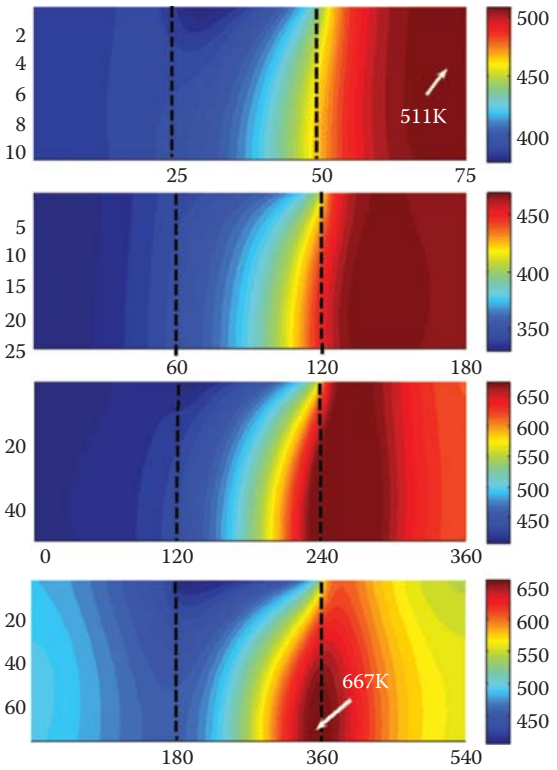


FIGURE 51.28 Lattice temperature profile in the active Si-layer for the FD-SOI devices from Table 51.2 ranging from 25 nm (top) to 180 nm (bottom) channel length. Shown here is only the thin Si film of the SOI device, namely the active area of the device.

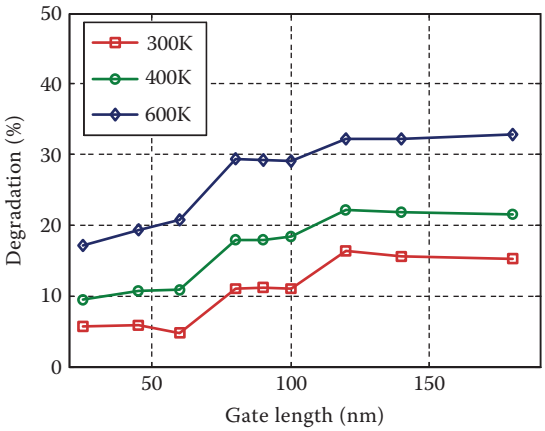


FIGURE 51.29 Current degradation (abs(I_{th}−I_o)/I_o, where I_{th} is the on current when the self-heating model is incorporated and I_o is current in the isothermal case) versus technology generation for the FD-SOI devices from Table 51.2 ranging from 25- to 180-nm channel length. Different curves correspond to different temperatures on the gate electrode (bottom set of results: T_{gate} = 300 K, middle set of results: T_{gate} = 400 K and top set of results: T_{gate} = 600 K).

the thermal conductivity must be taken into consideration. The choice of the BOX material makes significant impact on the amount of observed self-heating. Both diamond and Si are good materials to be used as BOX. Diamond, when compared to aluminum nitride (AlN) is a better heat spreader.

51.5 Outlook

The drift-diffusion solvers are applicable in situations in which the bias conditions and the device geometry are such that electric fields are relatively low and velocity saturation model is applicable [41]. Situations in which drift-diffusion models are applicable are the Si-based power MOSFET devices, bipolar junction transistors, light-emitting diodes (that are used more and more in solid-state lightning), and crystalline solar cells, just to name a few. In some of these devices, such as power transistors and light-emitting diodes (LEDs) used for solid-state lighting, it is of paramount importance to incorporate self-heating models within the drift-diffusion framework [42]. The accuracy of simple heating models in conjunction with drift-diffusion models is to some degree questionable so that it is in many circumstances justifiable to use particle-based device simulators with more exact self-heating models [43].

On the other hand, the hydrodynamic models do not suffer from the limitations of the drift-diffusion approaches and the incorporation of the additional energy balance equation allows one to include velocity overshoot in the model. Velocity overshoot and nonstationary transport are key features of conventional MOSFET devices with gate lengths of 200 nm and below [44]. However, the magnitude of the velocity overshoot observed via simulations depends strongly upon the choice of the energy relaxation time, mostly in sub-100-nm channel length devices. This, in turn, affects the magnitude of the drain current. The reason for such drastic differences in the results when different energy relaxation times are used is due to the fact that the energy relaxation time is a material-dependent as well as device geometry-dependent parameter. So, to calculate better estimates for the velocity overshoot, higher moments of the BTE are needed [45]. These, in turn involve parameters that are more and more ambiguous on the expense of increased computational cost and when the computational cost of hydrodynamic models exceeds one of particle-based device simulators, there is no point in using moment methods. In these circumstances, the direct solution of the BTE via the Monte Carlo method becomes a method of choice. Thus, we might conclude that it is advisable to use particle-based device simulators when nanoscale devices are concerned [33].

But how far down in the scaling can we go? Particle-based device simulators capture on an equal footing both ballistic and diffusive transport, so if the ballisticity factor in the device increases [46], there

is no problem that ballistic transport is effectively captured with particle-based device simulators [47]. Quantum-mechanical size quantization effects can also be captured by solving in slices the corresponding one-dimensional (1D) or 2D Schrödinger equation if one is concerned with conventional or FD-SOI MOSFETs, or nanowire transistors, respectively. What cannot be captured with particle-based device simulators is if there are local strains and stresses in the ultra-nanoscale devices, but that can also be cured via coupling of Monte Carlo device simulators with atomistic models [48,49] for bandstructure calculation.

In summary, Monte Carlo device simulators are a powerful tool for modeling devices ranging from the nanoscale regime to the microscale regime. What cannot be modeled with particle-based device simulators are resonant tunneling diodes in which quantum interference effects dominate the device behavior [50]. Efforts have been made along this direction as well [51], but the inclusion of the quantum-mechanical phase alongside with well-defined particle trajectory still remains open field of research.

References

1. ITRS International Technology Roadmap for Semiconductors. Available at: http://www.itrs.org.
2. G. Moore. Progress in digital integrated electronics. *IEDM Tech. Digest*, 11–13, 1975.
3. R. Dennard, F. H. Gaensslen, H. N. Yu, L. Rideout, E. Bassous, and A. R. LeBlanc. Design of Ion-Implanted MOSFET's with very small physical dimensions. *IEEE J. Solid State Circ.*, Vol. 9 (5), 256–268, 1974.
4. E. Fayneh, M. Yuffe, E. Knoll, M. Zelikson, M. Abozaed, Y. Talker, Z. Shmuely, and S. A. Rahme. 14 nm 6th-generation core processor SoC with low power consumption and improved performance. *IEEE International Solid-State Circuits Conference*, San Francisco, pp. 72–74, 2016.
5. T. Song, W. Rim, S. Park, Y. Kim, J. Jung, G. Yang, S. Baek, J. Choi, B. Kwon, Y. Lee, S. Kim, G. Kim, H. S. Won, J. H. Ku, S. S. Paak, E. Jung, S. S. Park, and K. Kim. A 10 nm FinFET 128 Mb SRAM with assist adjustment system for power, performance, and area optimization. *IEEE International Solid-State Circuits Conference*, San Francisco, pp. 306–309, 2016.
6. H.-Y. Chen, F.-C. Chen, Y.-L. Chan, K.-N. Yang, F.-L. Yang, and C. Hu. Method of fabricating a necked FINFET device. US 7122412 B2, 2006.
7. L. Geppert. A quantum leap for photonics. *IEEE Spectr.*, 28–33, April 9, 2004.
8. W. Zhu, J. P. Han, and T. P. Ma. Mobility measurement and degradation mechanisms of MOSFETs made with ultrathin high-k dielectrics. *IEEE Trans. Electron Dev.*, Vol. 51, 98–105, 2004.
9. J. Welser, J. L. Hoyt, and J. F. Gibbons. Temperature and scaling behavior of strained-Si N-MOSFET's. *IEDM Tech. Dig.,* 1000–1002, 1992.
10. R. Oberhuber, G. Zandler, and P. Vogl. Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFET's. *Phys. Rev. B.*, Vol. 58, 9941–9948, 1998.
11. R. Chau, J. Kavalieros, B. Doyle, A. Murthy, N. Paulsen, D. Lionberger, D. Barlage, R. Arghavani, B. Roberts, and M. Doczy. A 50 nm depleted-substrate CMOS transistor (DST). *IEDM Techn. Dig.*, 29.1.1–29.1.4, 2001.
12. Y.-K. Choi, K. Asano, N. Lindert, V. Subramanian, T.-J King., J. Bokor, and C. Hu. Ultrathin-body SOI MOSFET for deep-sub-tenth micron era. *IEEE Electron Dev. Lett.*, Vol. 21 (5), 254–255, 2000.
13. D. Vasileska and S. M. Goodnick. Computational electronics. *Mater. Sci. Eng. R Rev. J.*, Vol. R38, 181–236, 2002.
14. C. Jacoboni and L. Reggiani. The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Rev. Mod. Phys.*, Vol. 55, 645–705, 1983.
15. C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation*. Vienna, VA: Springer-Verlag, 1989.
16. K. Hess. *Monte Carlo Device Simulation: Full Band and Beyond*. Boston, MA: Kluwer Academic Publishing, 1991.
17. M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods*. New York, NY: Wiley, 1986.

18. D. K. Ferry. *Semiconductors*. New York, NY: Macmillan, 1991.
19. D. Vasileska, S. M. Goodnick, and G. Klimeck. *Computational Electronics: Semiclassical and Quantum Transport Modeling*. New York, NY: Taylor & Francis, 2010.
20. D. K. Ferry. *Quantum Mechanics: An Introduction for Device Physicists and Electrical Engineers*. New York, NY: Taylor & Francis, 2001.
21. H. D. Rees. Calculation of distribution functions by exploiting the stability of the steady state. *J. Phys. Chem. Solids*, Vol. 30, 643, 1969.
22. R. M. Yorston. Free-flight time generation in the Monte Carlo simulation of carrier transport in semiconductors. *J. Comp. Phys*., Vol. 64, 177, 1986.
23. M. Lundstrom. *Fundamentals of Carrier Transport*. London: Cambridge University Press, 2000.
24. B. K. Ridley. *Quantum Processes in Semiconductors*. Oxford: Oxford University Press, 1982.
25. D. Vasileska. https://nanohub.org/groups/dragica_vasileska/semiclassical.
26. K. Tomizawa. *Numerical Simulation of Submicron Semiconductor Devices*. Boston, MA: Artech House, 1993.
27. M. V. Fischetti and S. E. Laux. Band structure, deformation potentials and carrier mobility in strained Si, Ge and SiGe alloys. *J. Appl. Phys.*, Vol. 80, 2234–2252, 1996.
28. T. Gonzalez and D. Pardo. Physical models of ohmic contact for Monte Carlo device simulation. *Solid State Electron*., Vol. 39, 555–562, 1996.
29. P. A. Blakey, S. S. Cherensky, and P. Sumer. *Physics of Submicron Structures*. New York, NY: Plenum Press, 1984.
30. S. E. Laux. On particle-mesh coupling in Monte Carlo device simulation. *IEEE Trans. Comp. Aided Des. Int. Circ. Sys*., Vol. 15, 1266, 1996.
31. R. Hathwar. Generalized Monte Carlo tool for investigating low-field and high field properties of materials using non-parabolic band structure model. MS Thesis, Arizona State University, June 2011.
32. S. S. Qazi. Electrical and thermal transport in alternative device technologies. MS Thesis, Arizona State University, November 2013.
33. K. Raleva, D. Vasileska, S. M. Goodnick, and M. Nedjalkov. Modeling thermal effects in nanodevices. *IEEE Trans. Electron Dev.*, Vol. 55 (6), 1306, 2008.
34. A. R. Shaik. Multi-scale study of heat transfer using Monte Carlo technique for phonon transport. MS Thesis, Arizona State University, April 2016.
35. D. Vasileska, K. Raleva, and S. M. Goodnick. Self-heating effects in nanoscale FD SOI devices: The role of the substrate, boundary conditions at various interfaces, and the dielectric material type for the box. *IEEE Trans. Electron Dev.*, Vol. 56, 3064, 2009.
36. W. Liu and M. Asheghi. Phonon-boundary scattering in ultrathin single-crystal silicon layers. *Appl. Phys. Lett.*, Vol. 84, 3819, 2004.
37. D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar. Thermal conductivity of individual silicon nanowires. *Appl. Phys. Lett*., Vol. 83, 2934, 2003.
38. E. H. Sondheimer. The mean free path of electrons in metals. *Adv. Phys*., Vol. 1, 1, 1952, reprinted in *Adv. Phys.*, 50, 499, 2001.
39. D. Vasileska, K. Raleva, and S. M. Goodnick. Electrothermal studies of FD SOI devices that utilize a new theoretical model for the temperature and thickness dependence of the thermal conductivity. *IEEE Transactions on Electron Devices*, Vol. 57, 726–728, 2010.
40. K. Raleva and D. Vasileska. The importance of thermal conductivity modeling for simulations of self-heating effects in FD SOI devices. *J. Comput. Elec*., Vol. 12 (4), 601–610, 2013.
41. D. Vasileska and S. M. Goodnick, Editors. *Nano-Electronic Devices: Semiclassical and Quantum Transport Modeling.* Vienna, VA: Springer Verlag, June 2011.
42. *www.silvaco.com*.
43. D. Vasileska. Modeling thermal effects in nano-devices. *Microelectron. Eng.*, Vol. 109 (9), 163–167 2013.

44. T. Grasser, T. W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. IEEE*, Vol. 91 (2), 251–274, 2003.
45. T. Grasser, H. Kosina, C. Heitzinger, and S. Selberherr. Characterization of the hot electron distribution function using six moments. *J. Appl. Phys,* Vol. 91, 3869, 2002.
46. M. Lundstrom. Elementary scattering theory of the Si MOSFET. *IEEE Electron Dev. Lett.,* Vol. 18, 361–363, 1997.
47. M. V. Fischetti and S. E. Laux. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects. *Physical Review B*, Vol. 38(14), 9721, 1988.
48. T. Frauenheim, G. Seifert, M. Elstner, T. Niehaus, C. Köhler, M. Amkreutz et al. Atomistic simulations of complex materials: Ground-state and excited-state properties. *J. Phys. Condens. Matter*, Vol. 14 (11), 3015, 2002.
49. G. Klimeck, S. S. Ahmed, H. Bae, N. Kharche, S. Clark, B. Haley, S. Lee et al. Atomistic simulation of realistically sized nanodevices using NEMO 3-D—Part 1: models and benchmarks. *IEEE Trans. Electron Dev.,* Vol. 54 (9), 2079–2089, 2007.
50. R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic. Single and multiband modeling of quantum electron transport through layered semiconductor devices. *J. Appl. Phys.*, Vol. 81 (12), 7845–7869, 1997.
51. L. Shifren and D. K. Ferry. Wigner function quantum Monte Carlo. *Physica B: Condens. Matter*, Vol. 314 (1), 72–75, 2002.

52

Photonics

52.1 Landscape of Numerical Methods 808
Time-Harmonic versus Time-Dependent Modeling • Exact versus Approximate Models • Order of Approximation and Flexibility of Meshes • Computing Aspects: CPU-Time versus Memory • Special Adaptation to Photonics

52.2 Maxwell's Equations in Nondispersive, Chargeless Media 809
Scattering, Mode Propagation, and Resonance Problems • Simplification: 2D Geometries and 2D Fields • Transformation Rules • Perfectly Matched Layers

52.3 FDTD Method 817
Historical Notes • Concept: Discrete Differential Operators on Staggered Grids • Spatial Discretization of Maxwell's Equations • Time Evolution • Dispersion, Stability, and Courant–Friedrichs–Lewy Condition • Discrete Div-Conditions • Convergence

52.4 Fourier Modal Method 823
Historical Notes • Concept: Segmentwise Waveguide Treatment • Transfer-Matrix Approach • Scattering-Matrix Approach • Layered Media Stack as Model Case • FMM in 2D • Convergence

52.5 Finite Element Method 833
Historical Notes • Concept: Patchwise Tangential Continuous, Polynomial Approximation on Unstructured Grids • 1D: Weak Form and Variational Formulation • 1D: Linear Finite Elements, Local Matrices, and Assembly Process • 2D and 3D: Weak Form and Variational Formulation • 2D and 3D: Linear Finite Elements, Local Matrices, and Assembly Process • Convergence

52.6 DG Method 844
Historical Notes • Concept: Combine Adaptive Spatial FEM Discretization with the Time Treatment in Finite Volume Methods • Scalar Advection Equation as Model Equation: The Finite Volume Method in 1D • Maxwell's Equations in Conservation Form • Concept of DG-Algorithms for 3D Maxwell's Equations • Convergence

Frank Schmidt

This chapter deals with numerical simulation in photonics. It provides an overview of the most widely used methods, explains their concepts, and gives some details, as long as it is necessary for a qualitative understanding. It is far from being complete but tries to describe the landscape of simulation tools. Parts of this review are closely related to the book *Numerical Methods in Photonics* [1], which itself is a comprehensive representation of the field covered by many very detailed monographs. The methods we describe

in the following are the *finite-difference time-domain FDTD method*, closely related to the *finite integration technique* (FIT), the *Fourier modal method* (FMM), the *finite element method* (FEM), and the *discontinuous galerkin* (DG) method. We omit other very useful methods often specialized to certain applications, in particular:

- Beam propagation method (BPM), for simulation of paraxial beam propagation [2]
- Integral methods such as boundary integral methods, for simulation problems in grating design [3]
- Plane wave expansion (PWE) for simulations of band structures [4]

and refer the reader to the given monographs.

52.1 Landscape of Numerical Methods

There are many aspects to consider if one has to choose a numerical method to solve a particular problem in photonics. We now discuss a few.

52.1.1 Time-Harmonic versus Time-Dependent Modeling

In the most general case, one wants to solve time-dependent Maxwell's equations with general, i.e., nonlocal, nonlinear, dispersive media. Both FDTD and DG methods in time provide such general solutions. On the other hand, especially in the case of high accuracy requirements, strict conditions on the maximal stepsizes in time may cause a high amount of central processing unit (CPU)-time consumption. Therefore, in case of linear and nondispersive problems, it is often useful to consider time-harmonic problems, and, if needed, to combine time-harmonic solutions via Fourier transform with different frequencies to obtain a time-dependent behavior.

52.1.2 Exact versus Approximate Models

It is common to approximate Maxwell's equations by structurally simpler equations. Especially, the simplification in terms of polarizations have a long tradition, where the full vectorial equations are replaced by semivectorial or scalar approximations. The scalar Helmholtz equation, e.g., is well investigated and a number of efficient numerical approaches exists, hence at first glance it seems reasonable to try Helmholtz solvers instead of Maxwell solvers. But the full vectorial Maxwell solvers available today do not have significant disadvantages with respect to speed or memory consumption, thus it is obsolete in many cases to use such semivectorial or scalar approximations. The case is different for special situations such as paraxial field propagations along distinguished propagation directions. As long as reflections can be neglected, in the case of waveguide tapers for example, so-called BPM are useful. They map a time-harmonic three-dimensional (3D) scattering problem approximately to a time-like evolution problem with 2D cross sections, which results in a substantial reduction of the computational effort.

52.1.3 Order of Approximation and Flexibility of Meshes

All methods described in the following offer, at least in principle, high-order approximations in space and time, thus resulting in high-order convergence. Nevertheless, this is not always used in practice. FDTD methods, e. g., are typically linked to a second-order discretization in time and space on staggered grids. Therefore, once geometrical flexibility is needed, one would typically, instead of enhancing FDTD, directly apply DG. Implementation of DG is more involved than of FDTD, but offers the flexibility of FEM meshes by design and yields the time evolution with costs comparable to FDTD.

52.1.4 Computing Aspects: CPU-Time versus Memory

The time-dependent solvers FDTD and DG can be treated by explicit time integrators. Thus, the evolution procedure results in much less memory consumption than that is typical for FMM or FEM, where either full or large sparse matrices have to be inverted. Hence, on memory-restricted systems, FDTD or DG are advantageous. The price to pay is longer computing times.

52.1.5 Special Adaptation to Photonics

FDTD, DG, and FEM are general methods for solving electromagnetic problems; they are not specially tailored to photonics. Since mode propagation is often the central physical effect, numerical methods based on the interaction of waveguide or cavity modes are often very effective. The dominant class of methods here are the FMMs. Subdividing the structure into waveguide sections and relating waveguide modes to each other result in algorithms which often offer surprisingly fast convergence in the range of lower, but in practice sufficient accuracy.

52.2 Maxwell's Equations in Nondispersive, Chargeless Media

In the most general case, the numerical methods of photonics solve time-dependent Maxwell's equations with general media, i.e., with nonlocal, nonlinear, dispersive media. For a concise and clear presentation of the different numerical methods, we use only a subset of the general equations. Our discussions are based on

$$\begin{aligned} \nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \cdot \mathbf{D} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{52.1}$$

completed by the corresponding constitutive equations

$$\mathbf{D} = \varepsilon \mathbf{E} \qquad \mathbf{B} = \mu \mathbf{H} \qquad \mathbf{J} = \sigma \mathbf{E}. \tag{52.2}$$

These equations describe the fields in linear, nondispersive, and chargeless media, where the magnetic field $\mathbf{H}$, the electric field $\mathbf{E}$, the electric displacement $\mathbf{D}$, and the magnetic flux density $\mathbf{B}$ are vectorial quantities defined for the entire space and positive times, $\mathbf{H}, \mathbf{E}, \mathbf{D}, \mathbf{B} : \mathbb{R}^3 \times \mathbb{R}^+ \longrightarrow \mathbb{R}^3$, and the permittivity ε, the permeability μ, and the conductivity σ are scalar, piecewise continuous functions in space, $\varepsilon, \mu, \sigma : \mathbb{R}^3 \longrightarrow \mathbb{R}$.

The field equations combined with the constitutive equations yield the system, on which our representation of the FDTD method and the DG method are based:

$$\begin{pmatrix} -\sigma & \nabla\times \\ \nabla\times & \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \frac{\partial}{\partial t} \left[\begin{pmatrix} \epsilon & \\ & -\mu \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} \right]. \tag{52.3}$$

The other methods, FEM, FMM, etc., are based on the time-harmonic form of Equation 52.3. The concept is to use a Fourier transform of the fields in Equation 52.3 for a single frequency ω. Assuming the single frequency dependence to be $\exp(-i\omega t)$, the Fourier transformed fields become complex but live in space

only, $\mathbf{H}, \mathbf{E} : \mathbb{R}^3 \longrightarrow \mathbb{C}^3$. We do not introduce a special notation for the time-harmonic fields, since the meaning becomes clear from the context. In the frequency domain, Equation 52.3 reads

$$\begin{pmatrix} & \nabla\times \\ \nabla\times & \end{pmatrix}\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = -i\omega\left[\begin{pmatrix} \epsilon - \frac{\sigma}{i\omega} & \\ & -\mu \end{pmatrix}\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}\right]. \tag{52.4}$$

52.2.1 Scattering, Mode Propagation, and Resonance Problems

Maxwell's equations take different forms depending on the different modeling situations. In the case of time-dependent problems, these are the different types of linear and nonlinear scattering problems. In the case of time-harmonic problems, where mostly but not exclusively, linear problems are considered, we have scattering, mode propagation, and resonance problems. Time-dependent solutions to Equation 52.3 including also nonlocal, dispersive, or nonlinear materials are the most general classical solutions to Maxwell's equations—the corresponding computational effort is usually relatively large. Solutions to the time-harmonic system (52.4) are often much cheaper.

52.2.1.1 Scattering on Unbounded Domain

Figure 52.1 shows the situation of a time-harmonic scattering problem, the situation for time-dependent scattering is very similar. There is a source field $\mathbf{E}_{\text{source}}$ traveling through the unbounded space, being itself a solution to Maxwell's equations (Equation 52.4). If no scatterer would be present, this would be the only field we have. But if it hits an object, scattered fields $\mathbf{E}_{\text{scattered}}$ (or in short $\mathbf{E}_{\text{s}}$) are generated. We need to compute the field inside a computational domain Ω, which we also call the interior domain, in contrast to the exterior domain surrounding Ω. Further, we denote the boundary of the computational domain with $\partial\Omega$.

Let the source field $\mathbf{E}_{\text{source}}(\mathbf{r})$, with $\mathbf{r}$ in the exterior domain, $\mathbf{r} \in \mathbb{R}^3 \setminus \bar{\Omega}$ be given. Then, the general scattering problem in 3D is defined by the following conditions:

1. Interior problem. The field on the computational domain obeys Equation 52.4.
2. Exterior problem. The fields in the exterior domain are a superposition of the source and the scattered fields $\mathbf{E}_{\text{source}}$ and $\mathbf{E}_{\text{s}}$. The source field and the scattered field must also obey Maxwell's equations.
3. Continuity of the tangential data and their normal derivative along the boundary of the computational domain.
4. In the case of time-harmonic fields, the Silver–Müller radiation condition for the scattered field

$$\lim_{r\to\infty} r\left(\left(\nabla \times \mathbf{E}_{\text{s}}\right) \times \mathbf{r}^0 - i\left|\mathbf{k}\right| \mathbf{E}_{\text{s}}\right) = 0 \tag{52.5}$$

holds true uniformly in all directions $\mathbf{r}$. Here, $\mathbf{r}^0$ denotes the radial vector of unit length.

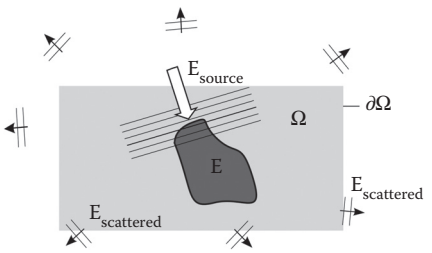


FIGURE 52.1 Scattering problem.

Typically, the radiation condition is not directly applied. But the methods that treat the exterior part of the problem ensure that the scattered fields satisfy the condition. The most prominent and successful method here is the perfectly matched layer (PML) method, see the Section 52.2.4.

52.2.1.2 Resonance Problems

In resonance problems, we consider Equation 52.4 as an eigenvalue problem with the unknown eigenvalue ω and the unknown resonance mode, the eigenvector, $(\mathbf{E}, \mathbf{H})^T$. The special challenge lies in the treatment of the exterior domain. Either the boundary conditions on the boundary of the computational domain are explicitly known, e. g., as a magnetic or electric wall, or methods from scattering problems are applied.

52.2.1.3 Mode Propagation Problems

Again, we consider Equation 52.4 as an eigenvalue problem. This time, we specialize the problem further to z-invariant problems: $\varepsilon = \varepsilon(x, y)$. We take a fixed frequency ω and look for solutions of the type $\mathbf{E}(x, y, z) = \mathbf{E}(x, y)e^{ik_z z}$ with a (unknown) phase velocity k_z. Two common forms of the eigenvalue problem for k_z are the following: Since any z-dependence is given by factors $\exp(ik_z z)$, one can introduce a specialized nabla-operator $\nabla_{k_z} = \mathbf{e}_x\partial_x + \mathbf{e}_y\partial_y + ik_z\mathbf{e}_z$ and Equation 52.4 reads

$$\begin{pmatrix} & \nabla_{k_z}\times \\ \nabla_{k_z}\times & \end{pmatrix}\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix} = -i\omega\left[\begin{pmatrix}\epsilon - \frac{\sigma}{i\omega} & \\ & -\mu\end{pmatrix}\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix}\right].$$

Here, the eigenvector consists of both $\mathbf{E}$ and $\mathbf{H}$, the eigenvalue k_z appears in linear form. Alternately, eliminating, e. g., the magnetic field, and using the decompositions $\mathbf{E} = \mathbf{E}_\perp + \mathbf{e}_z E_z$ and $\nabla = \nabla_\perp + i\mathbf{e}_z k_z$, one derives from Equation 52.4 the modal eigenvalue problem

$$\begin{bmatrix} \frac{1}{\varepsilon}\nabla_\perp\times\frac{1}{\mu}\nabla_\perp\times & +\frac{k_z^2}{\varepsilon\mu} & \frac{ik_z}{\varepsilon\mu}\nabla_\perp \\ i\frac{k_z}{\varepsilon}\nabla_\perp\frac{1}{\mu}\cdot & & -\frac{1}{\varepsilon}\nabla_\perp\cdot\frac{1}{\mu}\nabla_\perp \end{bmatrix}\begin{bmatrix}\mathbf{E}_\perp\\ E_z\end{bmatrix} = \omega^2\begin{bmatrix}\mathbf{E}_\perp\\ E_z\end{bmatrix} \tag{52.6}$$

The eigenvector consists just of the $\mathbf{E}$-field, but this is a quadratic eigenvalue problem in k_z. The same procedure for the magnetic field yields

$$\begin{bmatrix} \frac{1}{\mu}\nabla_\perp\times\frac{1}{\varepsilon}\nabla_\perp\times & +\frac{k_z^2}{\varepsilon\mu} & \frac{ik_z}{\varepsilon\mu}\nabla_\perp \\ i\frac{k_z}{\mu}\nabla_\perp\frac{1}{\varepsilon}\cdot & & -\frac{1}{\mu}\nabla_\perp\cdot\frac{1}{\varepsilon}\nabla_\perp \end{bmatrix}\begin{bmatrix}\mathbf{H}_\perp\\ H_z\end{bmatrix} = \omega^2\begin{bmatrix}\mathbf{H}_\perp\\ H_z\end{bmatrix}. \tag{52.7}$$

52.2.2 Simplification: 2D Geometries and 2D Fields

In a 2D geometry, the permittivity ε depends only on two coordinates, say the x- and y-coordinates: $\varepsilon = \varepsilon(x, y)$, just as in the mode propagation case in Section 52.2.1. In the mode propagation case, all fields depend on a common phase factor $\exp(ik_z z)$. If we set $k_z = 0$, i. e., we consider only fields with transversal position dependence $\mathbf{E} = \mathbf{E}(x, y)$, we obtain from (Equations 52.6 and 52.7) the decoupled 2D equations

$$\begin{aligned}\nabla_\perp\times\nabla_\perp\times\mathbf{E}_\perp(x, y) &= \varepsilon(x, y)\mu\omega^2\mathbf{E}_\perp(x, y)\\ -\nabla_\perp\cdot\nabla_\perp E_z(x, y) &= \varepsilon(x, y)\mu\omega^2 E_z(x, y).\end{aligned} \tag{52.8}$$

If we additionally assume uniformity in y-direction, we find the eigenvalue problem for waveguides with a 1D cross section

$$\partial_x^2 E_z + \varepsilon(x)\mu\omega^2 E_z = k_y^2 E_z. \tag{52.9}$$

The same applies to the magnetic field:

$$\nabla_\perp \times \frac{1}{\varepsilon(x,y)} \nabla_\perp \times \mathbf{H}_\perp = \mu\omega^2 \mathbf{H}_\perp(x,y)$$

$$\varepsilon(x)\partial_x \frac{1}{\varepsilon(x)} \partial_x H_z + \varepsilon(x)\mu\omega^2 H_z = k_y^2 H_z. \tag{52.10}$$

These equations are simpler than the original ones and they build the basis of many (approximate) solvers. They are simpler in the operator structure and they are linear in the eigenvalue k_y^2.

Note that, conventionally, the propagation direction of waveguide modes along the axis of waveguides is chosen to be the z-axis. To meet this convention, one has to interchange the y- and z-indices in Equations 52.9) and 52.10.

As a consequence of decoupling, we can split the entire field globally into a transversal electric (TE or s-polarization) and a transversal magnetic (TM or p-polarization) field:

$$\begin{pmatrix} \mathbf{E}(x,y) \\ \mathbf{H}(x,y) \end{pmatrix} = \underbrace{\begin{pmatrix} E_z(x,y)\mathbf{e}_z \\ H_x(x,y)\mathbf{e}_x + H_y(x,y)\mathbf{e}_y \end{pmatrix}}_{\text{s-polarization (TE)}} + \underbrace{\begin{pmatrix} E_x(x,y)\mathbf{e}_x + E_y(x,y)\mathbf{e}_y \\ H_z(x,y)\mathbf{e}_z \end{pmatrix}}_{\text{p-polarization (TM)}}. \tag{52.11}$$

52.2.3 Transformation Rules

Transformation techniques often yield considerable simplifications, e.g.,

- Exploiting symmetries by adapting coordinate systems (cylindrical, spherical coordinate systems, etc.)
- Using precomputations on reference elements as in FEM, cf. Section 52.5
- Handling of PML, cf. Section 52.2.4, by a complex coordinate stretching

The transformation rules describe the change of the equations when we change a local coordinate system. The PML, seen as a complex scaling of coordinates, can be treated as change of a coordinate system. The computations in FEM simplify if they are done in the reference coordinate systems and then mapped to the physical systems.

Geometric quantities and field quantities transform differently. For an easy *algebraic* notation, it is useful to write geometric quantities in row vectors and field quantities in column vectors of their coefficients. Expressions like $\mathrm{d}\mathbf{l} \cdot \mathbf{E}$ with $\mathrm{d}\mathbf{l} = (\mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$ (a row vector of path components) and $\mathbf{E} = (E_x, E_y, E_z)^T$ (a column vector of field components) have then the right algebraic form and can be programed directly as they appear in the formula.

52.2.3.1 Mapping of Geometric Quantities

Let an arbitrary domain mapping $Q : \widehat{K} \to K$ be given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = Q\left(\hat{x}, \hat{y}, \hat{z}\right), \tag{52.12}$$

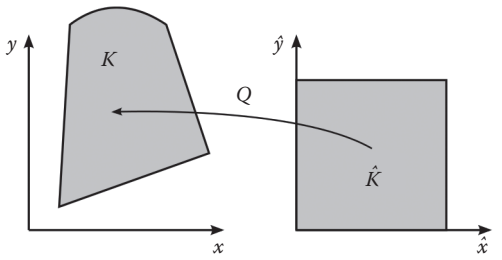


FIGURE 52.2 Example of a domain mapping. A square is mapped to a curvilinear quadrilateral. Quantities belonging to the reference domain are marked with a hat.

see Figure 52.2. This way we map *points* from the reference domain to *points* in the original domain, where the points are defined by the coefficients $\hat{x}, \hat{y}, \hat{z}$ in the reference system and by the coefficients x, y, z in the original system.

We study how infinitesimally small path, face, and volume elements $\mathrm{d}\mathbf{l}$, $\mathrm{d}\mathbf{A}$, $\mathrm{d}V$, transform. In 3D, we describe $\mathrm{d}\mathbf{l}$ and $\mathrm{d}\mathbf{A}$ by three-coefficient row vectors. A linearization of the mapping at point $\left(\hat{x}, \hat{y}, \hat{z}\right)$ yields

$$\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix} = J\left(\hat{x}, \hat{y}, \hat{z}\right) \begin{pmatrix} \mathrm{d}\hat{x} \\ \mathrm{d}\hat{y} \\ \mathrm{d}\hat{z} \end{pmatrix}, \qquad J\left(\hat{x}, \hat{y}, \hat{z}\right) := \begin{pmatrix} \partial_{\hat{x}} Q_1 & \partial_{\hat{y}} Q_1 & \partial_{\hat{z}} Q_1 \\ \partial_{\hat{x}} Q_2 & \partial_{\hat{y}} Q_2 & \partial_{\hat{z}} Q_2 \\ \partial_{\hat{x}} Q_3 & \partial_{\hat{y}} Q_3 & \partial_{\hat{z}} Q_3 \end{pmatrix}. \tag{52.13}$$

Defining the path element as the row vector $\mathrm{d}\mathbf{l} = \left(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z\right)$, we obtain the path element transformation

$$\mathrm{d}\mathbf{l} = \mathrm{d}\hat{\mathbf{l}} J^T. \tag{52.14}$$

A similar computation for a face element $\mathrm{d}\mathbf{A} = \left(\mathrm{d}A_x, \mathrm{d}A_y, \mathrm{d}A_z\right)$ spanned by two path elements gives, with $|J|$ the determinant of J, the transformation rule

$$\mathrm{d}\mathbf{A} = \mathrm{d}\hat{\mathbf{A}} J^{-1} |J| \tag{52.15}$$

and for a volume element $\mathrm{d}V$ spanned by three path elements we obtain

$$\mathrm{d}V = \mathrm{d}\hat{V} |J|. \tag{52.16}$$

52.2.3.2 Transformation of Grad, Curl, Div

We use the nabla notation describing a column vector

$$\mathbf{grad}\, v(x, y, z) = \nabla_{xyz} v(x, y, z) = \begin{pmatrix} \partial_x v(x, y, z) \\ \partial_y v(x, y, z) \\ \partial_z v(x, y, z) \end{pmatrix}.$$

Let an arbitrary function scalar $v(x, y, z)$ be given which lives on K and allows for the computation of the gradient. We compute the scalar quantity $\mathrm{d}\mathbf{l} \cdot \nabla_{xyz} v$. This is a physical quantity with the meaning of a voltage (e.g., $\mathbf{E} \cdot \mathrm{d}\mathbf{s}$) or a current ($\mathbf{H} \cdot \mathrm{d}\mathbf{s}$). This quantity must not change if we change the coordinate system, hence, we require

$$\mathrm{d}\mathbf{l} \cdot \nabla_{xyz} v(x, y, z) = \mathrm{d}\hat{\mathbf{l}} \cdot \nabla_{\hat{x}\hat{y}\hat{z}} \hat{v}\left(\hat{x}, \hat{y}, \hat{z}\right),$$

TABLE 52.1 Transformation of Geometric and Field Quantities

	Geometric Quantity	Field Quantity	Nabla Operation
Segment	$\mathrm{d}\mathbf{l} = \mathrm{d}\hat{\mathbf{l}} J^T$	$\mathbf{E} = J^{-T}\hat{\mathbf{E}}$ $\mathbf{H} = J^{-T}\hat{\mathbf{H}}$	$\nabla_{xyz} v = J^{-T}\nabla_{\hat{x}\hat{y}\hat{z}}\hat{v}$
Face	$\mathrm{d}\mathbf{A} = \mathrm{d}\hat{\mathbf{A}}\,\lvert J\rvert\, J^{-1}$	$\mathbf{D} = \frac{J}{\lvert J\rvert}\hat{\mathbf{D}}$ $\mathbf{B} = \frac{J}{\lvert J\rvert}\hat{\mathbf{B}}$	$\nabla_{xyz} \times \mathbf{v} = \frac{J}{\lvert J\rvert}\nabla_{\hat{x}\hat{y}\hat{z}} \times \hat{\mathbf{v}}$
Volume	$\mathrm{d}V = \mathrm{d}\hat{V}\,\lvert J\rvert$	$\rho = \frac{1}{\lvert J\rvert}\hat{\rho}$	$\nabla_{xyz} \cdot \mathbf{v} = \frac{1}{\lvert J\rvert}\nabla_{\hat{x}\hat{y}\hat{z}} \cdot \hat{\mathbf{v}}$

where $\hat{v}\left(\hat{x},\hat{y},\hat{z}\right)$ in the reference domain follows from $v\left(\hat{x},\hat{y},\hat{z}\right)$ in the original domain by $v(x,y,z) = v\left(Q\left(\hat{x},\hat{y},\hat{z}\right)\right) = (v \circ Q)\left(\hat{x},\hat{y},\hat{z}\right)$, hence $\hat{v} = v \circ Q$, based on the coordinate transform Equation 52.12. It follows from Equation 52.14 that

$$\mathrm{d}\hat{\mathbf{l}} J^T \cdot \nabla_{xyz} v(x,y,z) = \mathrm{d}\hat{\mathbf{l}} \cdot \nabla_{\hat{x}\hat{y}\hat{z}}\hat{v}\left(\hat{x},\hat{y},\hat{z}\right).$$

Since this should hold true for all functions v, we find the mapping of the gradient

$$\nabla_{xyz} = J^{-T}\nabla_{\hat{x}\hat{y}\hat{z}}. \tag{52.17}$$

The coefficient of the nabla operator transform like the coefficients of the field strength. The same way one computes the transformations rules for **curl** and **div**, as collected in Table 52.1.

52.2.3.3 Transformation of μ and ε

Rewriting Maxwell's **curl** equations from the physical to the reference domain we get

$$\frac{J}{|J|}\nabla_{\hat{x}\hat{y}\hat{z}} \times \hat{\mathbf{H}} = i\omega\varepsilon J^{-T}\hat{\mathbf{E}}$$

$$\frac{J}{|J|}\nabla_{\hat{x}\hat{y}\hat{z}} \times \hat{\mathbf{E}} = -i\omega\mu J^{-T}\hat{\mathbf{H}},$$

which shows that the entire transformation can be traced back to a mapping of the material quantities

$$\hat{\varepsilon} : = |J|\, J^{-1}\varepsilon J^{-T} \tag{52.18}$$

$$\hat{\mu} : = |J|\, J^{-1}\mu J^{-T}. \tag{52.19}$$

52.2.4 Perfectly Matched Layers

A standard problem in the simulation of scattering problems is the treatment of scattered fields outside the domain of interest Ω_{int}, cf. Figure 52.3. The predominantly used method is the PML technique introduced by Berenger [5]. The PML treatment of Berenger relies on a split field formulation of Maxwell's equations in the PML region. This has been shown in [6] to be only weakly stable. In contrast, the uniaxial PML (UPML) [7] does not require such a splitting of Maxwell's fields and maintains the strong well-posedness of the original Maxwell system. Both formulations are equivalent to an elegant and general approach: the stretched-coordinate PML [8,9]. We describe briefly only the latter one. It has among others two remarkable properties

1. It can be used in a similar way in higher space dimensions.
2. It can be plugged into existing codes provided that the codes allow for sufficient general materials and field representations.

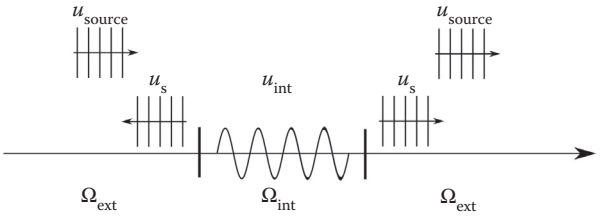


FIGURE 52.3 Sketch of a 1D time-harmonic scattering problem. The source fields are given, the interior field u_{int} and the scattered field u_S has to be computed.

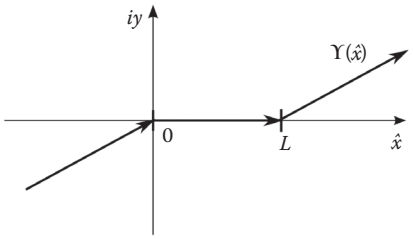


FIGURE 52.4 Complex PML path. Within the interior domain $(0, L)$ the x-coordinate is real but becomes complex outside.

52.2.4.1 PML in 1D

Let us consider the right exterior domain $x > L$, where L marks the right boundary point of the situation depicted in Figure 52.3. The general solution of the scattered field in the scalar 1D version of Equation 52.8 is

$$u_s = c_- e^{-ikx} + c_+ e^{ikx}, \quad x > L.$$

We have to discriminate between the incoming version of the scattered field $c_- \exp(-ikx)$, which we need to drop and the outgoing scattered field $c_+ \exp(ikx)$. Here, x is the real space coordinate. Now, we allow x to be a path in the complex plane, not just on the real axis. We parameterize path x with the real parameter $\hat{x}$ to derive a path in the complex plane; now $\hat{x}$ takes the role of the reference coordinate as in Section 52.2.3:

$$\begin{aligned} x(\hat{x}) &= L + (1 + i\sigma)\hat{x} \\ \sigma &> 0, \text{ real, fixed} \\ \hat{x} &> 0, \text{ real reference coordinate;} \end{aligned} \tag{52.20}$$

compare Figure 52.4, which yields the following mapping:

$$\begin{aligned} c_- \exp(-ikx) &\mapsto c_- \exp(-ikL)\exp(-ik\hat{x})\exp(\sigma k\hat{x}) \\ c_+ \exp(ikx) &\mapsto c_+ \exp(ikL)\exp(ik\hat{x})\exp(-\sigma k\hat{x}). \end{aligned}$$

The first expression, the incoming field, blows up for $\hat{x} \to \infty$, due to the factor $\exp(\sigma k\hat{x})$, whereas the second one decreases exponentially with $\exp(-\sigma k\hat{x})$. So for $\hat{x}$ large enough, we have a clear discrimination:

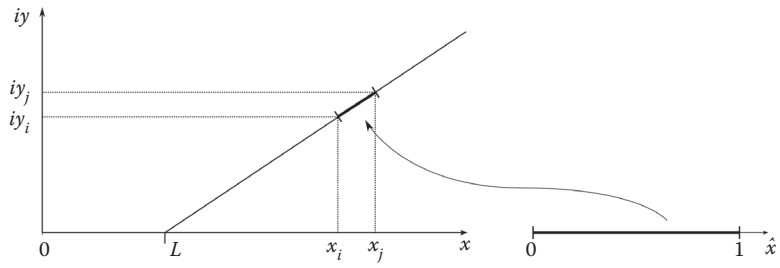


FIGURE 52.5 Transformation from the unit interval to the complex-valued PML interval.

nearly the entire field value comes from the left traveling scattered field. Hence, a proper formulation of Sommerfeld's radiation condition in terms of the complexified path $x(\hat{x})$ is

$$u\left(x\left(\hat{x}\right)\right) \longrightarrow 0 \ \text{ as } \hat{x} \longrightarrow \infty.$$

Let us stay in the right exterior domain and introduce the complex distance variable x:

$$x(\hat{x}) = \hat{x} + i\sigma\hat{x}.$$

We consider the complexification of the spatial variable within the Helmholtz equation:

$$\partial_x^2 u + k^2 u = 0 \qquad \Rightarrow \qquad \frac{1}{(1+i\sigma)^2}\frac{\partial^2 u}{\partial \hat{x}^2} + k^2 u = 0.$$

This step is decisive. If we are allowed to generalize the real derivative to a complex derivative, i. e., if the derivative is independent of the direction in the complex plane, we can use the linear path transformation, see Figure 52.5, (or others) to obtain the last equation.

This transformation is what we in fact exploit within the algorithmic realization. We define the complexification of the real x-coordinate to a complex coordinate x by means of the real reference coordinate $\hat{x}$

$$x = \begin{cases} \hat{x}\,(1+i\sigma)\,, & \hat{x} < 0 \\ \hat{x}, & 0 \le \hat{x} < L \\ L + \left(\hat{x} - L\right)(1+i\sigma)\,, & \hat{x} \ge L. \end{cases} \tag{52.21}$$

52.2.4.2 PML in 2D and 3D

The (tensor product type) generalization of Equation 52.20 to 2D is obviously the following (σ_x, σ_y are real, positive constants):

$$x = \begin{cases} \hat{x}\left(1+i\sigma_x\right), & \hat{x} < 0 \\ \hat{x}, & 0 \le \hat{x} < L\,, \\ L + \left(\hat{x} - L\right)\left(1+i\sigma_x\right), & \hat{x} \ge L \end{cases} \quad y = \begin{cases} \hat{y}\left(1+i\sigma_y\right), & \hat{y} < 0 \\ \hat{y}, & 0 \le \hat{y} < L \\ L + \left(\hat{y} - L\right)\left(1+i\sigma_y\right), & \hat{y} \ge L. \end{cases} \tag{52.22}$$

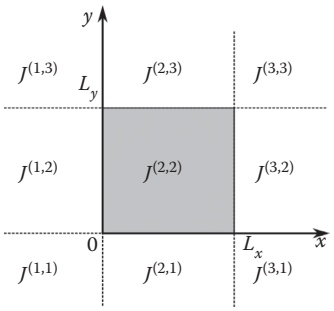


FIGURE 52.6 PML-regions in 2D lexicographically enumerated.

The Jacobian of the mapping $(\hat{x}, \hat{y}) \mapsto (x, y)$ depends on the position, see Figure 52.6, and evaluates to, e.g.,

$$J^{(2,2)} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \qquad J^{(3,2)} = \begin{pmatrix} 1 + i\sigma_x & \\ & 1 \end{pmatrix} \qquad J^{(3,3)} = \begin{pmatrix} 1 + i\sigma_x & \\ & 1 + i\sigma_y \end{pmatrix}.$$

The realization of the PML now consists in using the transformed permittivity (Equation 52.18) and permeability (Equation 52.19), where the Jacobians are computed from the complex stretching of the space variables (Equation 52.22). We proceed now as follows. We consider a subdomain of the exterior, say $\Omega^{(2,1)} \subset \mathbb{R}^2$ from Figure 52.6 and the complex solution $u^{(2,1)} : \Omega^{(2,1)} \to \mathbb{C}$ living there. Then, we use the mapping from the real to the complex coordinates to define a complex-valued spatial domain: $\Omega^{(2,1)} \to \Omega^{c,(2,1)} \subset \mathbb{C}^2$. On the complexified domain $\Omega^{c,(2,1)}$, we have a complex extension of $u^{(2,1)}$ which we denote by $u^{c,(2,1)}$. The solution $u^{c,(2,1)}$ is different to the one on the real space $\Omega^{(2,1)}$, it is the complex continuation of the original function. The key point is that our original function $u^{(2,1)}(x, y)$ has a complex extension by its nature, and we point to its values by introducing complex instead of only real spatial coordinates.

52.3 FDTD Method

The FDTD method is the most widely used method for the simulation of electromagnetic field propagation. It combines algorithmic simplicity with a deep universality. Nonlinear, nonlocal, and dispersive material responses can be treated. The reason for this generality lies in the fact that the structure of the FDTD scheme is very close to the structure of Maxwell's equations and has several algorithmic features:

- It is simple to implement and fast.
- High-order discretizations are feasible.
- It is explicit in time.
- There is a comprehensive theory.

Besides these pros, there are also severe drawbacks:

- The simple local approximation based on Cartesian staggered grids restricts geometric flexibility unless the grid is prohibitively fine (staircasing problem).
- The inherent second-order accuracy in time and space of the original method limits long-time computations on large domains.

52.3.1 Historical Notes

The initial and most important step was done by Yee in 1966 [10]. He used a so-called staggered mesh discretization in space and time, see the following subsections, and replaced the differential operators **grad**, **curl**, and **div** by finite-difference approximations to obtain a discrete approximation of second-order accuracy. Since then, a huge number of contributions improved the method and extended the fields of applications. An interesting view on the derivation of the FDTD equations has been given by Weiland in 1977 [11] starting from an integral formulation of Maxwell's equations. This approach is called FIT and yields the same set of equations. We use this approach to motivate the finite difference equations.

A number of efforts have been aimed at addressing the above-mentioned problems. The staircasing problem has been treated in many papers, e.g., [12]. High-order methods have been proposed, e.g., in [13, 14].The *generalized finite difference method* introduced by Bossavit [15] in 2001 takes ideas from differential geometry and topology and stimulates new interpretations of classical finite element schemes in terms of finite differences. Clemens and Weiland [16] follow a similar spirit in the framework of FIT to extend classical schemes.

A comprehensive overview of FDTD methods and further improvements along with detailed applications is given in [17] which is the standard reference book on FDTD methods. A thoroughly different approach to time-dependent problems is given by the class of DG methods which by its inherent construction offer high-order approximations in space and time as well as flexible meshes, cf. Section 52.6.

52.3.2 Concept: Discrete Differential Operators on Staggered Grids

Figure 52.7 explains the basic construction principle. We consider two entangled loops. One, say the light gray loop, is associated with the electric field, the other one with the magnetic field (left part of Figure 52.7). These loops will be used later for line integrals. The straight edges of the loops themselves are the edges of a cube, see the middle part of Figure 52.7. Then, by adding cubes related to the electric and magnetic fields, the complete spatial Yee mesh is generated. It can be considered as composed from two independent Cartesian meshes where the "electric" mesh and the "magnetic" mesh are shifted compared to each other by half of an elementary edge length of a cube.

52.3.2.1 Degrees of Freedom

A central question in any numerical scheme is the proper definition of the degrees of freedom (DoF). Typically, the stability of a method essentially relies on this. In FDTD methods, it is common to project the fields to the edges and to take the values at the midpoints of the edges as DoF. For example, we project $\mathbf{E}$ to the ith edge $\mathbf{e}_i$, $\mathbf{E} \cdot \mathbf{e}_i$, take the value at the midpoint, and denote it as E_i. We motivate the DoFs in the following by approximations of integral expressions (e.g., by the midpoint rule) as it is done in the FIT. This offers additional insight into the structure of the equations. The same results are obtained if the differential operators are approximated by finite difference stencils.

52.3.2.2 Discrete Gradient

For a direct discretization of Maxwell's equations, it is not necessary to consider the discrete approximation of the gradient but it completes the insight into the structure of the subsequent discretization process. Consider an edge $\mathbf{e}$ of length h connecting two neighboring vertices v_1, v_2 of the Yee mesh. The fundamental theorem of calculus

$$\int_{v_1}^{v_2} \mathbf{grad}\, \phi \, \mathrm{d}\mathbf{e} = \phi(v_2) - \phi(v_1),$$

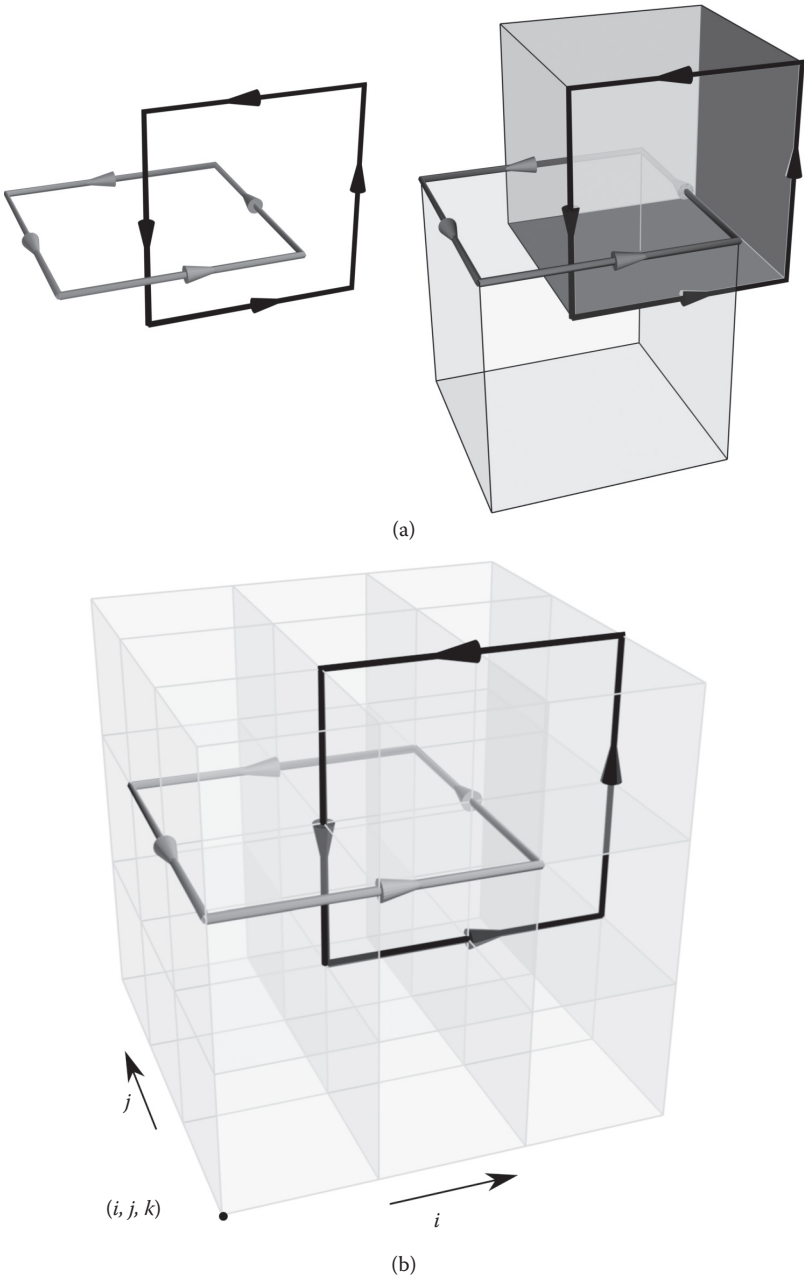


FIGURE 52.7 Schematic interaction of electric and magnetic fields and its realization with staggered grids.

motivates the definition of the discrete gradient operator $\mathbf{grad}_{\mathrm{h}}$:

$$\mathbf{grad}_{\mathrm{h}}\,\phi\big|_{\mathbf{e}} \;:=\; \frac{\mathbf{e}^0}{h}\left(\phi(v_2)-\phi(v_1)\right) \;\text{ forall edges e,}$$

where $\mathbf{e}^0$ is the normalization of $\mathbf{e}$ to unit length. The $\mathbf{grad}_{\mathrm{h}}$ -operator maps quantities defined on *vertices* to quantities defined on *edges*.

52.3.2.3 Discrete Curl

The definition of the discrete **curl** -operator, $\mathbf{curl}_\mathrm{h}$, relies on the Stokes theorem

$$\int_{\mathbf{f}} \mathbf{curl}\,\mathbf{E}\;\mathrm{d}\mathbf{f} = \int_{\partial\mathbf{f}} \mathbf{E}\;\mathrm{d}\mathbf{e}$$

with $\mathbf{f}$ a plane, rectangular face of a cube and $\partial\mathbf{f}$ its boundary consisting of four edges $\mathbf{e}_i$. The discrete version reads

$$\mathbf{curl}_\mathrm{h}\,\mathbf{E}\big|_{\mathbf{f}} \;:=\; \frac{\mathbf{f}^0}{h_1 h_2}\sum_{i=1}^{4} \iota_{e_i,f}\,\big|\mathbf{e}_i\big|\,E_i \;\text{ forall faces } \mathbf{f}.$$

The iota-symbol $\iota_{e_i,f}$ denotes the incidence number of edge $\mathbf{e}_i$ with respect to the orientation of the face $\mathbf{f}$. It holds $\iota_{e_i,f} = 1$, if edge $\mathbf{e}_i$ shares the same direction as that induced by the face vector $\mathbf{f}$, otherwise $\iota_{e_i,f} = -1$. The $\mathbf{curl}_\mathrm{h}$ -operator maps quantities defined on *edges* to quantities defined on *faces*. For the light gray loop of Figure 52.7, e.g., starting with the edge in front of the figure and assuming the edge lengths of the rectangle are all the same $h_1 = h_2 = h$, this is

$$\mathbf{curl}_\mathrm{h}\,\mathbf{E}\big|_{\mathbf{f}} \;:=\; \frac{\mathbf{f}^0}{h}\left(E_1 + E_2 - E_3 - E_4\right).$$

52.3.2.4 Discrete Divergence

The definition of the discrete **div** -operator, $\mathbf{div}_\mathrm{h}$, relies on the Gauss theorem

$$\int_{V} \mathbf{div}\,\mathbf{D}\;dV = \int_{\partial V} \mathbf{D}\;\mathrm{d}\mathbf{f}.$$

Here, V is the volume of the cube, ∂V its boundary consisting of 6 faces, and $\mathbf{f}$ is the outward pointing surface vector. The discrete version reads

$$\mathbf{div}_\mathrm{h}\,\mathbf{D}\big|_c = \frac{1}{h_1 h_2 h_3}\sum_{i=1}^{6} \iota_{e_f,c}\,\big|\mathbf{f}_i\big|\,D_i \;\text{ for all cells } c.$$

The iota-symbol $\iota_{f_i,c}$ denotes the incidence number of face $\mathbf{f}_i$ with respect to the orientation of cell c, defined positive by the outward orientation of all surface vectors. Hence, it holds $\iota_{f_i,c} = 1$, if face $\mathbf{f}_i$ points outward c, otherwise $\iota_{f_i,c} = -1$. The $\mathbf{div}_\mathrm{h}$ -operator maps quantities defined on *faces* to quantities defined on *cells*.

52.3.2.5 Exact Sequences

As a direct consequence of the fact that the Yee mesh is orientable, it follows

$$\begin{aligned} \mathbf{curl}_\mathrm{h}\;\cdot\mathbf{grad}_\mathrm{h} &= 0 \\ \mathbf{div}_\mathrm{h}\;\cdot\mathbf{curl}_\mathrm{h} &= 0. \end{aligned} \tag{52.23}$$

Hence, the discrete operators share the same properties as the continuous ones. For the Yee mesh, the property follows easily from a direct computation.

52.3.3 Spatial Discretization of Maxwell's Equations

Let us denote by $\mathbf{E}_\mathrm{h}$ the column vector of all DoFs associated with the primal mesh and by $\mathbf{H}_\mathrm{h}$ the column vector of all DoFs associated with the dual mesh. Then, the spatial discrete version of Equation 52.3, setting $\sigma = 0$ for simplicity, reads

$$\varepsilon \partial_t \mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}} = \mathbf{curl}_\mathrm{h}\, \mathbf{H}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{d}} \qquad \text{for all faces } \mathbf{f}_\mathrm{d} \text{ of the dual mesh,} \tag{52.24}$$

$$\mu \partial_t \mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}} = -\,\mathbf{curl}_\mathrm{h}\, \mathbf{E}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{p}} \qquad \text{for all inner faces } \mathbf{f}_\mathrm{p} \text{ of the primal mesh.} \tag{52.25}$$

Here, $\mathbf{e}_\mathrm{p}$ denotes the edge of the primal mesh that penetrates face $\mathbf{f}_\mathrm{d}$ of the dual mesh and $\mathbf{e}_\mathrm{d}$ denotes the edge of the dual mesh going through $\mathbf{f}_\mathrm{p}$. For implementations, it is convenient to consider both the primal and the dual meshes as submeshes of a common Cartesian mesh.

52.3.4 Time Evolution

The time evolution of the semidiscretized equations (Equations 52.24 and 52.25) can be computed by a number of different time integrators. Frequently used is the so-called leapfrog scheme. It fits perfectly to the spatial discretization and offers the advantage of being an explicit method.

52.3.4.1 Leapfrog Scheme in Time

We introduce a uniform mesh in time

$$0 = t_0 < t_{\frac{1}{2}} < t_1 < \ldots < t_i < t_{i+\frac{1}{2}} < t_{i+1} \ldots < t^N = T,$$

with $t_{i+1} - t_i = \Delta t$ for all time steps and indicate the discrete time by a superscript at the quantities to use, e. g., the discrete electric field vector at time $t_{n+\frac{1}{2}}$ is $\mathbf{E}_\mathrm{h}^{n+\frac{1}{2}}$. Using a central finite difference approximation of the time derivative in Equation 52.24, we obtain its fully discretized version

$$\mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n+\frac{1}{2}} = \mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n-\frac{1}{2}} + \frac{\Delta t}{\varepsilon}\, \mathbf{curl}_\mathrm{h}\, \mathbf{H}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{d}}^{n} \qquad \text{for all faces } \mathbf{f}_\mathrm{d} \text{ of the dual mesh.} \tag{52.26}$$

The time derivative is centered around n, hence it is of second-order accuracy. Using the updated electric field of Equation 52.26, Equation 52.25 is updated accordingly:

$$\mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}}^{n+1} = \mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}}^{n} - \frac{\Delta t}{\mu}\, \mathbf{curl}_\mathrm{h}\, \mathbf{E}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{p}}^{n+\frac{1}{2}} \qquad \text{for all inner faces } \mathbf{f}_\mathrm{p} \text{ of the primal mesh.} \tag{52.27}$$

Equations 52.26 and 52.27 are typical FDTD update equations.

52.3.5 Dispersion, Stability, and Courant–Friedrichs–Lewy Condition

Let us consider a monochromatic plane wave of frequency ω in a homogeneous medium, with speed of light $c = 1/\sqrt{\varepsilon\mu}$, ε and μ independent of position, and a wavevector $\mathbf{k} = (k_x, k_y, k_z)$. Then the plane wave

$$\mathbf{E} e^{-i\omega t + i\mathbf{k}\mathbf{r}}$$

with a constant vector **E** is a solution of Maxwell's equations without sources, if the dispersion relation

$$\omega^2 = \mathbf{k}^2 c^2 \tag{52.28}$$

and orthogonality condition

$$\mathbf{kE} = 0$$

hold true. The first one follows from the **curl**-**curl** equations, the second one from the divergence condition. For the Yee mesh, there is a discrete counterpart. The discrete plane wave satisfies Equations 52.26 and 52.27 takes the natural form

$$\mathbf{E}_\mathrm{h} e^{-i\omega t_j + i\mathbf{k}\mathbf{r}_\mathrm{h}}.$$

The corresponding discrete dispersion relation, however, contains the stepsizes in time, Δt, and space, h_x, h_y, h_z,

$$\frac{1}{\Delta t^2}\sin^2\frac{\omega\Delta t}{2} = c^2\left(\left(\frac{1}{h_x^2}\sin^2\frac{k_x h_x}{2}\right) + \left(\frac{1}{h_y^2}\sin^2\frac{k_y h_y}{2}\right) + \left(\frac{1}{h_z^2}\sin^2\frac{k_z h_z}{2}\right)\right). \tag{52.29}$$

For a derivation see, e. g.,[17] or [1]. Clearly, this discrete dispersion relation approaches Equation 52.28 as $\Delta t, h_x, h_y, h_z \to 0$ as it must be. But the sin-function of the left-hand side of Equation 52.29 has a further consequence. Let k_x, k_y, k_z be real, hence the plane wave is spatially bounded everywhere as in the case of the the continuous plane wave. Multiplying with Δt^2 from the left-hand side of Equation 52.29 might result in a real number larger than 1, if Δt is large enough. Then ω onthe left-hand side becomes imaginary, and since the expression involves $\sin^2$, both positive and negative imaginary frequencies occur. The positive one will cause an exponential increase of the discrete plane wave in time, hence the solution becomes unstable. Consequently, for stability we must require

$$c^2\Delta t^2\left(\frac{1}{h_x^2} + \frac{1}{h_x^2} + \frac{1}{h_z^2}\right) \le 1$$

or, for uniform meshes, $h = h_x = h_y = h_z$,

$$\Delta t \le \frac{h}{c\sqrt{3}}.$$

These inequalities are the Courant–Friedrichs–Lewy (CFL) conditions for the wave equation.

52.3.6 Discrete Div-Conditions

It is a remarkable property that the Yee scheme preserves the **div**-conditions of Equation 52.1 for homogeneous media in discrete form [18]. Computing the discrete divergence of the updated Equation 52.26

$$\mathbf{div}_\mathrm{h}\,\varepsilon\mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n+\frac{1}{2}} = \mathbf{div}_\mathrm{h}\,\varepsilon\mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n-\frac{1}{2}} + \Delta t\,\mathbf{div}_\mathrm{h}\ \mathbf{curl}_\mathrm{h}\,\mathbf{H}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{d}}^{n} \quad \text{for all cells of the dual mesh.}$$

The last term vanishes due to property (Equation 52.23), hence we have the convergence

$$\mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n+\frac{1}{2}} = \mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n-\frac{1}{2}} = \ldots = \mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{-\frac{1}{2}} \quad \text{for all cells of the dual mesh.}$$

The divergence attributed to a cell is preserved over all time steps. For the magnetic field, we obtain the corresponding result the same way:

$$\mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}}^{n+1} = \mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}}^{n} = \ldots = \mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}}^{0} \quad \text{for all cells of the primal mesh.}$$

If the initial data at time step $n = -1/2$ and $n = 0$ are divergence-free,

$$\mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}^{-\frac{1}{2}} = 0 \qquad \mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}^{0} = 0,$$

this property is preserved for all following time steps.

52.3.7 Convergence

Let, as before, $\mathbf{E}_\mathrm{h}$ be the vector of field values attributed to the edges of the primal mesh with values computed at the midpoint of edges. Let us denote the vector of the exact field values, projected to the edges, by $\mathbf{E}$; hence $\mathbf{E}_\mathrm{h}$ approximates $\mathbf{E}$ pointwise. We define the following, mesh-dependent discrete norm

$$\left\|\mathbf{E} - \mathbf{E}_h\right\|^2 \; : \; = h^3 \left(\mathbf{E}_\mathrm{h} - \mathbf{E}\right)^2,$$

and the analog for $\mathbf{H}$ with values attributed to the centroids of the cell faces.The following holds true. Let $\mathbf{E}$ and $\mathbf{H}$ be three times continuously differentiable in Ω. Then, there exists a constant C independent of h such that

$$\left\|\mathbf{E} - \mathbf{E_h}\right\| + \left\|\mathbf{H} - \mathbf{H_h}\right\| \leq Ch^2.$$

FDTD converges with second order. This classical result extends also to nonuniform meshes as shown by Monk and Süli [19].

52.4 Fourier Modal Method

This summary of FMMs follows Chapter 6 of [1]; for more details see that textbook and the references given therein.

52.4.1 Historical Notes

The FMM is a family of methods that belong to the class of spectral methods for the frequency domain and has been well established for more than three decades now. The term spectral method means that the field approximation is realized with functions which have a global support. In contrast, in FDTD methods, we consider only the pointwise-defined functions, and in FEM methods, we have basis functions with a very small, local support.

The origins of the FMM dates back to the 1960s and were devoted to grating analyses [20,21]. In 1981, Moharam and Gaylord established in their seminal paper [22] a computational framework they called rigorous coupled wave analysis (RCWA). Other synonyms or variations of the method include the PWE, the eigenmode expansion (EME), and eigenmode expansion technique (EET). A typical observation in RCWA methods has been that the convergence of TM fields is notably worse than that of TE fields. In 1996,

Lalanne and Granet and coworkers [23,24] found a solution to overcome this TM convergence problem and in the same year Li published his factorization rules [25] formalizing the results. Currently, different proposals are being investigated to establish a fast convergent 3D FMM, e. g., [26].

52.4.2 Concept: Segmentwise Waveguide Treatment

The family of FMMs are characterized by the following building blocks:

1. The geometry is subdivided into a sequence of waveguide-type sections, see Figure 52.8b and c.
2. The field within each waveguide section is represented by up and down propagating modes as well as evanescent modes.
3. The transversal fields of the waveguide-type eigenmodes are computed in Fourier representation.
4. The interaction of the individual sections is realized by posing proper continuity conditions at the interfaces and a scattering-matrix formalism.

52.4.3 Transfer-Matrix Approach

We consider the fields in two subsequent layers i and $i+1$, cf. Figure 52.9. At the moment, we do not focus on the special structure of the electromagnetic field except that we can decompose it into upward ($\mathbf{E}^+$) and downward ($\mathbf{E}^-$) propagating parts with respect to the z-direction. Let $\mathbf{E}_i^{\pm}(x,z)$ be arbitrary normalized solutions of Maxwell's equations split into propagation directions and $a_i^{\pm} \in \mathbb{C}$ their weights. It holds that

$$\begin{array}{lcccc} \text{Field in layer } i & : & a_i^+\mathbf{E}_i^+(x,z) & + & a_i^-\mathbf{E}_i^-(x,z) \\ \Downarrow \text{ implies} & & & & \\ \text{Field in layer } i+1 & : & a_{i+1}^+\mathbf{E}_{i+1}^+(x,z) & + & a_{i+1}^-\mathbf{E}_{i+1}^-(x,z). \end{array} \tag{52.30}$$

Given the complete field in one layer, Maxwell's equations including continuity conditions enforce the complete field in the other layer. This can be expressed by a transfer-matrix relation of the coefficients

$$\mathbf{A}_{i+1}\begin{pmatrix} a_{i+1}^+ \\ a_{i+1}^- \end{pmatrix} = \mathbf{A}_i \begin{pmatrix} a_i^+ \\ a_i^- \end{pmatrix}, \tag{52.31}$$

where $\mathbf{A}_i$, $\mathbf{A}_{i+1}$ are complex 2×2 matrices. We introduced this symmetric form since the count sequence of layers (top-down or bottom-up) is arbitrary. The typical application of the 1D transfer-matrix method is, given n layers of finite thickness with indices 0 and $n+1$ referring to substrate and superstrate, the

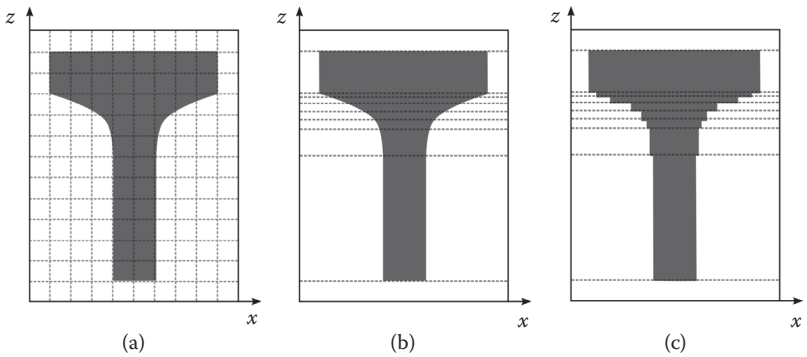


FIGURE 52.8 (a) A schematic geometry and a Cartesian mesh as used in FDTD methods. (b) Slicing of the geometry into segments along z. (c) Waveguide-like approximation of the geometry used in Fourier modal methods.

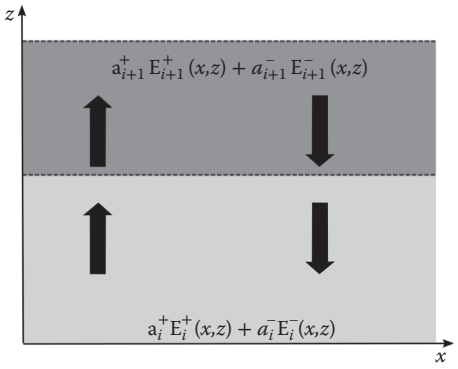


FIGURE 52.9 Superposition of up and down propagating fields in two subsequent layers.

update procedure

$$\begin{pmatrix} a_{n+1}^{+} \\ a_{n+1}^{-} \end{pmatrix} = \mathbf{T}_{n,n+1} \cdot \ldots \cdot \mathbf{T}_{i,i+1} \cdot \ldots \cdot \mathbf{T}_{0,1} \begin{pmatrix} a_0^{+} \\ a_0^{-} \end{pmatrix}.$$

This procedure may become unstable due to the exponentials involved in $T_{i,i+1}$.

52.4.4 Scattering-Matrix Approach

The scattering-matrix approach uses merely the same representation as the transfer-matrix approach, except that the computation of the coefficients $a_i^{\pm}$ is reordered (Figure 52.10). The scattering-matrix notation corresponding to Equation 52.31 is

$$\begin{pmatrix} a_i^{-} \\ a_{i+1}^{+} \end{pmatrix} = \mathbf{S}_{i,i+1} \begin{pmatrix} a_i^{+} \\ a_{i+1}^{-} \end{pmatrix} \tag{52.32}$$

with the obvious computation of the coefficients of $\mathbf{S}_{i,i+1}$ from $\mathbf{A}_i$ and $\mathbf{A}_{i+1}$. The notion scattering follows from the fact that we can consider the coefficients $\left(a_i^{+}, a_{i+1}^{-}\right)$ as fields *incident* to the layer separating boundary and the coefficients $\left(a_i^{-}, a_{i+1}^{+}\right)$ as fields *scattered* away from the boundary, respectively. The scattering matrix for an n-layer system is derived as follows. We rewrite Equation 52.31 as

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_i\left(d_i\right) & -\mathbf{A}_{i+1}(0) \end{pmatrix} \begin{pmatrix} a_i^{+} \\ a_i^{-} \\ a_{i+1}^{+} \\ a_{i+1}^{-} \end{pmatrix}.$$

Here, the top coefficient a_i^{+} of the column vector and the bottom coefficient a_{i+1}^{-} are the incident coefficients. We repeat this for the next layer, increasing the index i by 1

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_i\left(d_i\right) & -\mathbf{A}_{i+1}(0) & & \\ & \mathbf{A}_{i+1}\left(d_{i+1}\right) & -\mathbf{A}_{i+2}(0) & \end{pmatrix} \begin{pmatrix} a_i^{+} \\ a_i^{-} \\ a_{i+1}^{+} \\ a_{i+1}^{-} \\ a_{i+2}^{+} \\ a_{i+2}^{-} \end{pmatrix}.$$

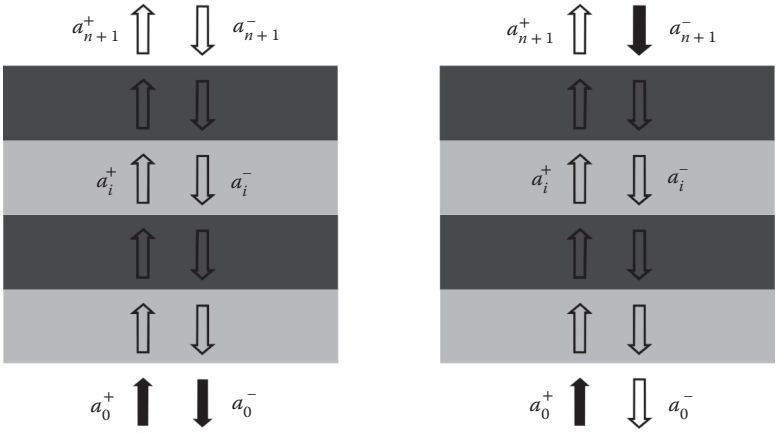


FIGURE 52.10 (a) Transfer-matrix approach. The data a_0^+, a_0^- are given (filled arrows), the others are computed. (b) Scattering-matrix approach. The data a_0^+, a_{n+1}^- are given, the others are computed.

Again, the top and bottom coefficients of the column vector are the incident coefficients of the layer system. Moving these coefficients to the left-hand side, deleting the first and last column of the system matrix, and inverting it yields the scattering-matrix formulation with scattering matrix **S**:

$$\mathbf{S}\begin{pmatrix} a_0^+ \\ 0 \\ \vdots \\ 0 \\ a_{n+1}^- \end{pmatrix} = \begin{pmatrix} a_0^- \\ a_1^+ \\ a_1^- \\ \vdots \\ \\ a_{n+1}^+ \end{pmatrix}. \tag{52.33}$$

52.4.5 Layered Media Stack as Model Case

A layered media stack is a 1D problem, if we choose the z-axis normal to the faces of the layers. To discuss the structure of FMM algorithms, we construct the standard transfer-matrix algorithm in scattering-matrix formulation from the viewpoint of the FMM. To find the field representation in layer i, we have to solve Maxwell's equations in layer i, i. e., either the equation for TE or TM. Let us consider the TE case. The modal ansatz

$$\text{layer } i: \quad E_y(x,z) = E(x)e^{ik_z z} = E_0 e^{ik_x x + ik_z z}$$

satisfies Equation 52.9 if $n_i^2 k_0^2 - k_x^2 = k_{z,i}^2$. Hence, given k_x, we have

$$\text{layer } i: \quad E_{y,i}(x,z) = a_i^+ e^{ik_x x + ik_{z,i} z} + a_i^- e^{ik_x x - ik_{z,i} z}, \tag{52.34}$$

which is the first line in Equation 52.30. Equally, we have for layer $i+1$

$$\text{layer } i+1: \quad E_{y,i+1}(x,z) = a_{i+1}^+ e^{ik_x x + ik_{z,i+1} z} + a_{i+1}^- e^{ik_x x - ik_{z,i+1} z}.$$

We compute matrices $\mathbf{A}_i$ and $\mathbf{A}_{i+1}$. Let the boundary between the two layers be placed at $z=0$. The tangential electric and magnetic fields must be continuous. It follows for the electric field

$$a_i^+ e^{ik_x x} + a_i^- e^{ik_x x} = a_{i+1}^+ e^{ik_x x} + a_{i+1}^- e^{ik_x x} \quad \text{for all } x \in \mathbb{R},$$

hence

$$a_i^+ + a_i^- = a_{i+1}^+ + a_{i+1}^-. \tag{52.35}$$

Given the electric field in each layer, we compute the tangential component of the magnetic field

$$\begin{aligned} \text{layer } i: \quad \omega\mu H_{z,i}(x,z) &= a_i^+ k_{z,i} e^{ik_x x + ik_{z,i} z} - a_i^- k_{z,i} e^{ik_x x - ik_{z,i} z} \\ \text{layer } i+1: \quad \omega\mu H_{z,i+1}(x,z) &= a_{i+1}^+ k_{z,i+1} e^{ik_x x + ik_{z,i+1} z} - a_{i+1}^- k_{z,i+1} e^{ik_x x - ik_{z,i+1} z} \end{aligned}$$

and obtain the continuity condition

$$k_{z,i} a_i^+ - k_{z,i} a_i^- = k_{z,i+1} a_{i+1}^+ - k_{z,i+1} a_{i+1}^-. \tag{52.36}$$

Consequently, taking Equations 52.35 and 52.36,

$$\begin{pmatrix} 1 & 1 \\ k_{z,i+1} & -k_{z,i+1} \end{pmatrix} \begin{pmatrix} a_{i+1}^+ \\ a_{i+1}^- \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ k_{z,i} & -k_{z,i} \end{pmatrix} \begin{pmatrix} a_i^+ \\ a_i^- \end{pmatrix}. \tag{52.37}$$

Repeating this calculation with a finite layer thickness d_i between the two boundaries, we obtain

$$A_i(d_i) = \begin{pmatrix} 1 & 1 \\ k_{z,i} & -k_{z,i} \end{pmatrix} \begin{pmatrix} e^{ik_{z,i} d_i} & 0 \\ 0 & -e^{ik_{z,i} d_i} \end{pmatrix}. \tag{52.38}$$

Collecting the steps, we obtain the classical transfer-matrix algorithm in scattering formulation Algorithm 52.1. In 1D it is not necessary to save the eigenvectors since they are the same for all layers. But in 2D and 3D, the eigenvectors are different in each layer, so we introduced the step here for full parallelism with Algorithm 52.1.

Algorithm 52.1: Transfer-Matrix Algorithm in Scattering Formulation

---

Require: a_0^+ and a_{n+1}^- to define source fields.
  for all layers **do**
    Solve each layer's eigenvalue problem (Equation 52.9).
    Save: eigenvectors $\exp(ik_x x)$ and eigenvalues $k_{z,i}$
    for later field representation (Equation 52.34).
    Compute the matrices A_i by (Equation 52.38).
  end for
  Assemble the system (Equation 52.33).
  Solve.
  return coefficients $a_0^-, \ldots, a_{n+1}^+$.

52.4.6 FMM in 2D

We generalize the foregoing, where ε has been constant in each layer. Now, let ε vary with x, each layer may have a different dependence $\varepsilon_i(x)$. The eigenvalue problems are Equations 52.9 and 52.10 with layerwise z-independent permittivities. In RCWA and FMM methods, these eigenvalue equations are solved based on a Fourier representation of the fields and the permittivities. This implies essentially two properties:

- Advantage: There is no lateral mesh needed. The nonlocal (or spectral) basis functions of fields and permittivities are simply constructed which results in a straightforward implementation.
- Disadvantage: The convergence of RCWA and FMM depend on the convergence properties of the Fourier representation which are not advantageous for nonsmooth functions.

52.4.6.1 Fourier Series Representation of Functions and Products of Functions

We consider a function $f(x)$ which is periodic with L: $f(x + L) = f(x)$. Hence, $f(x)$ has a Fourier series representation

$$f(x) = \sum_{m=-\infty}^{\infty} f_m e^{im\frac{2\pi}{L}x}. \tag{52.39}$$

We denote its Fourier transform by $\widehat{f}$

$$\widehat{f}(k) = \frac{1}{L}\int_0^L f(x)e^{-ikx}\,dx \text{ and compute } f_m = \widehat{f}\left(\frac{2\pi}{L}m\right), \quad m \in \mathbb{Z}.$$

We consider a second function $g(x)$, also periodic with interval L, and study the product $h(x) = f(x)g(x)$, which is also L-periodic and consequently has a Fourier representation

$$h(x) = \sum_{m=-\infty}^{\infty} h_m e^{im\frac{2\pi}{L}x}.$$

In terms of the Fourier coefficients of f and g, this is

$$\begin{aligned} h(x) &= \left(\sum_{m=-\infty}^{\infty} f_m e^{im\frac{2\pi}{L}x}\right)\left(\sum_{n=-\infty}^{\infty} g_n e^{in\frac{2\pi}{L}x}\right) = \sum_{m=-\infty}^{\infty}\left(\sum_{n=-\infty}^{\infty} f_{m-n}g_n\right)e^{im\frac{2\pi}{L}x} \\ h_m &= \sum_{n=-\infty}^{\infty} f_{m-n}g_n. \end{aligned} \tag{52.40}$$

We write the last equation in matrix-vector form

$$\mathbf{h} = \mathbf{Fg}$$

$$\mathbf{F} = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & f_0 & f_{-1} & f_{-2} & f_{-3} & \cdots \\ \cdots & f_1 & f_0 & f_{-1} & f_{-2} & \cdots \\ \cdots & f_2 & f_1 & f_0 & f_{-1} & \cdots \\ \cdots & f_3 & f_2 & f_1 & f_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

where the convolution matrix **F** is called the Toeplitz matrix.

52.4.6.2 Fourier Series Convergence in Dependence of the Smoothness of the Functions

In electromagnetics, we have to deal with piecewise smooth functions which have different global smoothness properties. Roughly, we have three classes:

Discontinuous: Functions that are piecewise smooth, but may have jumps or singularities. Example: Normal field components in a 3D setting with material discontinuities.

Continuous: Functions that are continuous but not differentiable at positions, where different materials meet. Example: H_y in TM fields as solution of, e.g., Equation 52.10.

Smooth: Functions that are continuous *and* the first derivative exists everywhere, even at positions, where different materials meet. Example: E_y in s-polarization, e.g., Equation 52.9.

52.4.6.3 Truncation

For an algorithmic application, the infinite series has to be truncated. The truncated Fourier series (Equation 52.39) then reads

$$f^M(x) = \sum_{m=-M}^{M} f_m e^{im\frac{2\pi}{L}x}, \quad f_m = \frac{1}{L}\int_0^L f(x)e^{-ik_m x}\,dx, \quad k_m = \frac{2\pi}{L}m \tag{52.41}$$

and the symmetric truncation of (Equation 52.40) yields

$$h^M(x) = \sum_{m=-M}^{M}\left(\sum_{n=-M}^{M} f_{m-n}g_n\right)e^{im\frac{2\pi}{L}x}, \quad h_m^M = \sum_{n=-M}^{M} f_{m-n}g_n.$$

Not only does the number of terms of the convolution sum change, but the individual coefficients h_n^M change too. The truncation changes the problem from an infinite to a finite one. The finite dimensional problem will differ from the exact, infinite dimensional one, as always in numerics, and the quality of this difference depends on the nature of the applied series truncation. The truncated matrix-vector form reads

$$\mathbf{h}^M = \mathbf{F}^M\mathbf{g}^M \qquad \text{“direct rule”} \tag{52.42}$$

$$\mathbf{F}^M = \left(f_{m-n}\right)_{-M\le m,n\le M} \tag{52.43}$$

with F^M the finite dimensional Toeplitz matrix and $f_{m-n} = 0$ for $|m-n| > M$. As shorthand for $\mathbf{F}^M$ generated by coefficients f_m we also write $[\![f]\!] := \mathbf{F}^M$.

52.4.6.4 Li's Factorization Rules

It has turned out that the originally observed slow convergence rate of TM problems is related to the smoothness properties of the products of field and permittivity and a special, polarization-dependent treatment can be used to remove the difficulty [23–25]. To this end, we introduce the discontinuity-types shown in Figure 52.11:

1. Type: I The product $f(x)g(x)$ results in $h(x)$ which exhibits also jump discontinuities.
2. Type: II The product $f(x)g(x)$ yields a continuous function $h(x)$. The discontinuities are removed in $h(x)$.

For a formulation of the FMM algorithm, taking into account the accelerated TM technique, we need additionally, the finite Fourier series representation of the inverse function $f^\dagger$:

$$f^\dagger(x) := \frac{1}{f(x)}$$

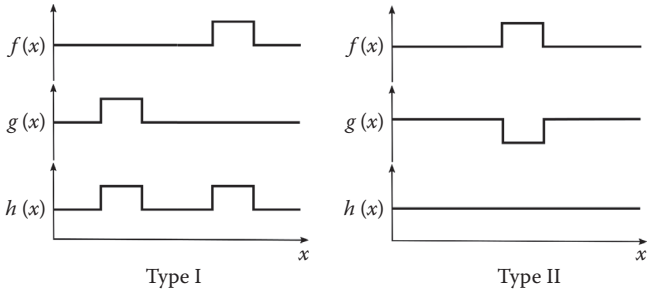


FIGURE 52.11 Two different types of jump discontinuities. Type I: The product $f(x)g(x)$ results in $h(x)$ which exhibits also jump discontinuities. Type II: The product $f(x)g(x)$ yields a continuous function $h(x)$.

together with its matrix-vector representation as counterpart to Equation 52.42

$$\mathbf{h}^M = \left(\mathbf{F}^{M\dagger}\right)^{-1} \mathbf{g}^M \qquad \text{“inverse rule.”} \tag{52.44}$$

The factorization rules are as follows:

1. Type: I If the discontinuities are not removable (Type I), apply the direct rule (Equation 52.42) for fastest convergence.
2. Type: II If the discontinuities are removable (Type II), apply the inverse rule (Equation 52.44) for fastest convergence.

52.4.6.5 Eigenvalue Problems in Fourier Space and Application of the Factorization Rules

The eigenvalue problems Equations 52.9 and 52.10 with ε independent of z hold true in each layer. We consider the TE equation Equation 52.9. Its continuous Fourier transform reads

$$\partial_x^2 E_y + \varepsilon(x)\mu\omega^2 E_y = k_z^2 E_y \quad \overset{\text{Fourier transform}}{\Longrightarrow} \quad (ik_x)^2 \widehat{E}_y + \mu\omega^2 \widehat{\varepsilon} * \widehat{E}_y = k_z^2 \widehat{E}_y. \tag{52.45}$$

Restricting $E_y(x)$ to functions periodic over L (discretization in k-space), truncating the number of Fourier basis functions, and using the Toeplitz matrix to compute the convolution we obtain the eigenvalue equation in Fourier space

$$(i\mathbf{K}_x)^2 \widehat{\mathbf{E}}_y + \mu\omega^2 \left[\!\left[\widehat{\varepsilon}\right]\!\right] \widehat{\mathbf{E}}_y = k_z^2 \widehat{\mathbf{E}}_y. \tag{52.46}$$

Equation 52.46 yields the generalization of the plane waves (TE) used in Equation 52.34 for the derivation of the transfer-matrix algorithm. The same procedure applies to the TM case, where strict consideration of the factorization rules is important. We compute

$$\varepsilon(x)\partial_x \frac{1}{\varepsilon(x)} \partial_x H_y + \varepsilon(x)\mu\omega^2 H_y = k_z^2 H_y.$$

$$\overset{\text{Fourier transform}}{\Longrightarrow} \quad \widehat{\varepsilon} * (ik_x)\widehat{\left(\frac{1}{\varepsilon}\right)} * (ik_x)\widehat{H}_y + \mu\omega^2 \widehat{\varepsilon} * \widehat{H}_y = k_z^2 \widehat{H}_y. \tag{52.47}$$

For proper discretization based on the factorization rules, we rewrite Equation 52.47, factoring out $\widehat{\varepsilon}$,

$$\widehat{\varepsilon} * \left((ik_x)\widehat{\left(\frac{1}{\varepsilon}\right)} * (ik_x) + \mu\omega^2 \right) \widehat{H}_y = k_z^2 \widehat{H}_y. \tag{52.48}$$

First, $\partial_x H_y$ and $1/\varepsilon$ are discontinuous, their product is continuous, hence factorization rule II applies. Second, the large parenthesis on the left-hand side of Equation 52.48 must be discontinuous since ε is discontinuous. The product, however, is continuous, since the right-hand side is continuous. Hence, the factorization rule type II applies again:

$$\left[\!\left[\widehat{\left(\frac{1}{\varepsilon}\right)}\right]\!\right]^{-1} \left(i\mathbf{K}_x \left[\!\left[\widehat{(\varepsilon)}\right]\!\right]^{-1} (i\mathbf{K}_x) + \mu\omega^2 \right) \widehat{\mathbf{H}}_y = k_z^2 \widehat{\mathbf{H}}_y. \tag{52.49}$$

52.4.6.6 Continuity Condition

We derive the continuity condition Equation 52.31 for the 2D case. In two subsequent waveguide sections, the multimode generalization of Equation 52.30 holds:

$$\begin{array}{ccccccc} \text{field in section } i & : & \sum_m a^+_{i,m}\mathbf{E}^+_{i,m}(x,z) & + & \sum_m a^-_{i,m}\mathbf{E}^-_{i,m}(x,z) \\ \Downarrow \text{ implies} & & & & & & \\ \text{field in section } i+1 & : & \sum_m a^+_{i+1,m}\mathbf{E}^+_{i+1,m}(x,z) & + & \sum_m a^-_{i+1,m}\mathbf{E}^-_{i+1,m}(x,z). \end{array} \tag{52.50}$$

The generalization from plane waves (Equation 52.34) to waveguide modes reads

$$\text{section } i: \quad E_{y,i}(x,z) = \sum_m a^+_{i,m}\mathbf{E}_{i,m}(x)e^{ik_{z,i,m}z} + \sum_m a^-_{i,m}\mathbf{E}_{i,m}(x)e^{-ik_{z,i,m}z}. \tag{52.51}$$

Each mth normalized mode in section i has a Fouriermode representation $\mathbf{E}_{i,m} = \sum_l b_{i,m,l} \exp ik_{x,l}x$. Here, the weights $b_{i,m,l}$ (ith section, mth waveguide mode, lth Fouriermode of this waveguide mode) are, in case of TE fields, the coefficients of the eigenvectors (Equation 52.46), i.e., $b_{i,m,l} = \left(\widehat{\mathbf{E}}_y\right)_l$, if $\widehat{E}_y$ is the mth waveguide mode in the ith section. Hence,

$$E_{y,i}(x,z) = \sum_m a^+_{i,m} \sum_l b_{i,m,l} e^{ik_{x,i,m,l}x} e^{ik_{z,i,m}z} + \sum_m a^-_{i,m} \sum_l b_{i,m,l} e^{ik_{x,i,m,l}x} e^{-ik_{z,i,m}z}.$$

Setting the interface position between sections i and $i+1$ to $z = 0$ and requiring continuity of E_y yields

$$\sum_{m=-M}^{M} b_{i,m,l}\left(a^+_{i,m} + a^-_{i,m}\right) = \sum_{m=-M}^{M} b_{i+1,m,l}\left(a^+_{i+1,m} + a^-_{i+1,m}\right), \qquad -M \le l \le M.$$

In the same way, one computes the 2D counterpart to Equation 52.36

$$\sum_{m=-M}^{M} b_{i,m,l}\left(k_{z,i,m}a^+_{i,m} - k_{z,i,m}a^-_{i,m}\right) = \sum_{m=-M}^{M} b_{i+1,m,l}\left(k_{z,i+1,m}a^+_{i+1,m} - k_{z,i+1,m}a^-_{i+1,m}\right)$$
$$-M \le l \le M$$

Introducing the matrix

$$\mathbf{B}_i = \left(b_{i,l,m}\right)^T = \left(\widehat{\mathbf{E}}_{y,-M}, \quad \cdots \quad , \widehat{\mathbf{E}}_{y,m}, \quad \cdots \quad , \widehat{\mathbf{E}}_{y,M} \right)^T$$

allows writing the two continuity conditions in compact block matrix form:

$$\begin{pmatrix} \mathbf{B}_i & \\ & \mathbf{B}_i \end{pmatrix} \begin{pmatrix} \mathbf{I}_\mathrm{d} & \mathbf{I}_\mathrm{d} \\ \operatorname{diag} \mathbf{k}_{z,i} & -\operatorname{diag} \mathbf{k}_{z,i} \end{pmatrix} \begin{pmatrix} \mathbf{a}_i^+ \\ \mathbf{a}_i^- \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{B}_{i+1} & \\ & \mathbf{B}_{i+1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_\mathrm{d} & \mathbf{I}_\mathrm{d} \\ \operatorname{diag} \mathbf{k}_{z,i+1} & -\operatorname{diag} \mathbf{k}_{z,i+1} \end{pmatrix} \begin{pmatrix} \mathbf{a}_{i+1}^+ \\ \mathbf{a}_{i+1}^- \end{pmatrix}$$

In this form, the similarity to the continuity condition Equation 52.37 of the transfer-matrix algorithm becomes obvious. For compact algorithmic notation, we define

$$\mathbf{A}_i := \begin{pmatrix} \mathbf{B}_i & \\ & \mathbf{B}_i \end{pmatrix} \begin{pmatrix} \mathbf{I}_\mathrm{d} & \mathbf{I}_\mathrm{d} \\ \operatorname{diag} \mathbf{k}_{z,i} & -\operatorname{diag} \mathbf{k}_{z,i} \end{pmatrix} \tag{52.52}$$

and have exactly as in the 1D case

$$\mathbf{A}_i \begin{pmatrix} \mathbf{a}_i^+ \\ \mathbf{a}_i^- \end{pmatrix} = \mathbf{A}_{i+1} \begin{pmatrix} \mathbf{a}_{i+1}^+ \\ \mathbf{a}_{i+1}^- \end{pmatrix}.$$

Collecting the steps, we obtain the FMM algorithm in 2D, Algorithm 52.2.

Algorithm 52.2: Fourier Modal Method in 2D

Require: $\mathbf{a}_0^+$ and $\mathbf{a}_{n+1}^-$ to define source fields.
  for all sections i **do**
    Solve each sections's eigenvalue problem, either (Equation 52.46), TE, or (Equation 52.49), TM.
    Save: eigenvectors $\widehat{\mathbf{E}}_{y,i,m}$, eigenvalues $k_{z,i,m}$
    for later field representation (Equation 52.51).
    Compute the matrices A_i by (Equation 52.52).
  end for
  Assemble the system (Equation 52.33).
  Solve.
  return vectors $\mathbf{a}_0^-, \ldots, \mathbf{a}_{n+1}^+$.

52.4.7 Convergence

The two significant approximation errors in FMM are the geometrical approximation by staircasing and the trunction in the number of Fourier modes. These problems have been addressed in numerous papers (see also the introduction). Especially, the representation of permittivities, which are not continuous in lateral direction, by global Fourier modes poses a convergence problem. Since in this case, the corresponding

Fourier coefficients decay only as $1/n$, where n is the number of the Fouriermode, the spatial approximation is worse than the $1/n^2$ we know, e. g., from the FDTD method. On the other hand, physics very often behaves in a way that only a few modes are sufficient to describe reasonably wave propagation and scattering. In these cases, FMM has been proven to be a very efficient method.

52.5 Finite Element Method

Several excellent books treating the finite element analysis of Maxwell's equations are available; the books of Peter Monk [27] and Leszek Demkowicz [28] are very recommendable. The summary of FEM here follows Chapter 8 of [1].

52.5.1 Historical Notes

The deep understanding of the relation between the scalar Helmholtz case, which enforces the use of classical, nodal based Lagrange shape functions, and the vectorial Maxwell case relies greatly on the work of Whitney and Nédélec [29,30], Bosssavit [31], and Webb [32]. Many important contributions of other authors have added to the field, which makes the FEM today a very flexible, effective, and well-studied method for electrodynamical simulations. An early survey of finite element studies with focus on nano-optical applications has been given in [33].

52.5.2 Concept: Patchwise Tangential Continuous, Polynomial Approximation on Unstructured Grids

The FEM is based on a *variational formulation* of Maxwell's equations that involves integral expressions on the computational domain. The FEM does not use any approximation to Maxwell's equations itself, except the discretization of the geometry, but approximates the solution space in which one tries to find a reasonable approximation to the exact solution. This is in contrast to the finite difference method which approximates Maxwell's equations directly via an approximation of the differential operators by finite difference stencils. In some respects, the FEM is comparable to FMMs: one takes a number of trial functions (the Fourier modes in case of FMM) and asks in which way a superposition of them gives a reasonable approximation to the exact solutions.

The solution space of FEM, however, is obtained by subdividing the computational domain into small patches and by providing a number of polynomials on each patch for the approximation of the solution. The patches together with the local polynomials defined on them are called *finite elements*. The most common examples of finite elements are triangles and rectangles in 2D and tetrahedrons and cubes in 3D together with constant, linear, quadratic, and cubic polynomials. These locally defined polynomial spaces have to be pieced together to ensure tangential continuity of the electric and magnetic field across the boundaries of neighboring patches. Once these local approximations with proper continuity conditions have been defined, these are inserted into the variational equation. This results in a linear system whose solution is a piecewise polynomial approximation of the exact solution. Hence, the two basic ingredients in each FEM for Maxwell's equations are

1. A variational formulation of Maxwell's equations.
2. A suitable construction of finite elements based on polynomials defined on local geometric patches to transform the variational formulation into a discrete, algebraic problem.

This concept has a number of remarkable properties:

- Complex geometrical shapes can be treated without geometrical approximations, e.g., rounding can be well and easily approximated, cf. Figure 52.12.

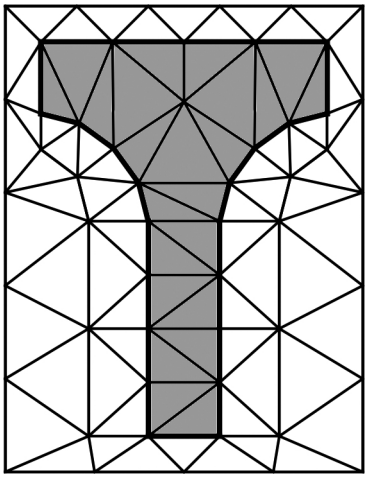


FIGURE 52.12 Sketch of a 2D triangulation for FEM, cf. also Figure 52.8 with staircase approximation. The triangles are shaped and placed such that the geometry is well covered by relatively few triangles.

- The finite element mesh can easily be adapted to the behavior of the solution, e.g., to singularities at corners.
- High-order approximations are available and ensure fast convergence.

52.5.3 1D: Weak Form and Variational Formulation

For the introduction of the finite element concept, we start with the most simple approximation of the time-harmonic scattering problem—the scalar Helmholtz equation in 1D, defined on the whole real axis, see Figure 52.3. The exterior domain should have a spatially invariant k. We let our computational domain Ω be the interval $(0, L)$. The source field is a plane wave coming from the left. A possible problem formulation reads: Find $u(x)$, $x \in \Omega = (0, L)$ such that

$$\partial_x^2 u(x) + k^2 u(x) = 0, \qquad x \in \mathbb{R} \tag{52.53}$$

$$u_{\text{source}}(x) = e^{ikx}, \qquad x \in \mathbb{R} \setminus \Omega. \tag{52.54}$$

Let us choose the interval $(0, L)$. We multiply Equation 52.53 with the complex conjugate of a so-called test function $v(x)$ and integrate over the domain Ω to arrive at

$$\int_0^L v^*(x) \left(\partial_x^2 u(x) + k^2 u(x)\right) dx = 0 \tag{52.55}$$

where nothing has been specified so far about $v(x)$, except that we must be allowed to compute the integral. This formulation cannot be equivalent to Equation 52.53, since very special choices of v would equate the integral to zero even for functions, which do not satisfy Equation 52.53. Hence, in order to have the chance to get something equivalent to Equation 52.53, Equation 52.55 must hold true for many functions $v(x)$. We call the class of functions from which we take $v(x)$, the *test space*. In contrast, the space of functions containing the desired solution $u(x)$ is called the *trial* or *ansatz space*. In general, both spaces can be different. It is subject to the convergence theory of finite elements to define precisely the test and trial

spaces and to conclude convergence rates. For our purpose here, it is sufficient to define these spaces in a way that makes the construction of the numerical algorithms reasonable. In the first attempt, we require $\int_0^L v^* v \, dx < \infty$, which is the class of square-integrable functions, the $L^2(0, L)$-functions. We will make the choice of the space more precise below, but only in a framework that introduces usual FEM notions and gives a first look at the topic. A comprehensive treatment can be found, e. g., in [27]. The functions $u(x)$ should be twice differentiable since the second derivative is contained in Equation 52.53. Hence, we seek them in the class $C^2(0, L)$. So in a first attempt to construct a variational equation, we formulate find a function $u(x) \in C^2(0, L)$ such that Equation 52.55 holds true for all $v \in L^2(0, L)$.

52.5.3.1 Weak Form

We perform an integration by parts on the left-hand side of Equation 52.55:

$$\int_0^L v^*(x) \left(\partial_x^2 u(x) + k^2 u(x)\right) dx = \int_0^L \left(-\partial_x v^*(x) \partial_x u(x) + k^2 v^*(x) u(x)\right) dx + v^*(x) \partial_x u(x) \Big|_0^L . \tag{52.56}$$

It is convenient to abbreviate the negative of the integral without boundary terms by

$$a_{\text{int}}(v, u) := \int_0^L \left(\partial_x v^*(x) \partial_x u(x) - k^2 v^*(x) u(x)\right) dx \tag{52.57}$$

to rewrite (Equation 52.56) to

$$\int_{\Omega_{\text{int}}} v^*(x) \left(\partial_x^2 u(x) + k^2 u(x)\right) dx =: -a_{\text{int}}(v, u) + \sum_{x \in \partial\Omega_{\text{int}}} v^*(x) \partial_n u(x). \tag{52.58}$$

The term Equation 52.57 is called a *sesquilinear form* because it is antilinear in its first argument (the v, due to complex conjugation) and linear in its second argument (the u). Hence, $a_{\text{int}}(v, u)$ maps two complex functions via integration to a complex number. We may repeat this for the scattered part of the exterior field

$$a_{\text{int}}(v, u) - \sum_{x \in \partial\Omega_{\text{int}}} v^*(x) \partial_n u(x) = 0 \quad \text{in } \Omega_{\text{int}} \tag{52.59}$$

$$a_{\text{ext}}(v, u_s) - \sum_{x \in \partial\Omega_{\text{int}}} v^*(x) \partial_n u_s(x) = 0 \quad \text{in } \mathbb{R} \setminus \Omega_{\text{int}}, \tag{52.60}$$

where we ensure (see discussion below) that the boundary term at infinity vanishes and only the boundary $\partial\Omega_{\text{int}}$ contributes to the equation. Using on this boundary, $u = u_{\text{source}} + u_s$ and introducing an auxiliary function g with

$$\begin{aligned} u_s &= \underbrace{u_s + g}_{w} - g \\ g(\partial\Omega_{\text{int}}) &= u_{\text{source}}(\partial\Omega_{\text{int}}) \\ g(\infty) &= 0 \end{aligned} \tag{52.61}$$

we obtain

$$a_{\text{int}}(v,u) - \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n\left(u_{\text{source}} + u_{\text{s}}\right) = 0 \quad \text{in } \Omega_{\text{int}} \tag{52.62}$$

$$a_{\text{ext}}(v,w-g) - \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u_{\text{s}}(x) = 0 \quad \text{in } \mathbb{R}\setminus\Omega_{\text{int}}. \tag{52.63}$$

Adding these two equations, taking into account the different normal directions, we end up with

$$a_{\text{int}}(v,u) + a_{\text{ext}}(v,w) = \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u_{\text{source}} + a_{\text{ext}}(v,g). \tag{52.64}$$

Next, using the function u, which lives only on the interior domain, and the function w, which lives only on the exterior domain, we define a joint, continuous function, again called u. This gives rise to the variational formulation of the scattering problem.

52.5.3.2 Variational Formulation

Let the source field data u_{source} and $\partial_n u_{\text{source}}$ at the boundary $\partial\Omega$ of the computational domain Ω be given. The variational form of the scattering problem reads

$$\text{Find } u\in V: \quad a(v,u) = \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u_{\text{source}} + a_{\text{ext}}(v,g) \quad \text{for all } v\in V, \tag{52.65}$$

with $a(\cdot,\cdot) = a_{\text{int}}(\cdot,\cdot)+a_{\text{ext}}(\cdot,\cdot)$, with a_{int} and a_{ext} given by Equation 52.59, Equation 52.60 and g defined by Equation 52.61. Note that in Equation 52.60 we neglected the contribution from the outer boundary $\partial\Omega$ext. The reason is that the expression is established for the scattered field u_{s} and we perform the integration over the complex extended exterior domain as described in Section 52.2.4. The complex extension leads to an exponential decay of u_{s}, hence the contribution of $\partial\Omega$ext vanishes.

52.5.4 1D: Linear Finite Elements, Local Matrices, and Assembly Process

The discretization of a given variational form to get the FEM equations is a straight forward procedure: just restrict the infinite dimensional test and trial space V to a finite dimensional space V_h:

$$\begin{aligned} \text{Find } u\in V: \quad a(v,u) &= b(v) \qquad \text{for all } v\in V. \\ &\downarrow \\ \text{Find } u_h\in V_h: \quad a(v_h,u_h) &= b(v_h) \qquad \text{for all } v_h\in V_h. \end{aligned} \tag{52.66}$$

The space V_h is not uniquely predetermined. The different types of Galerkin-type methods differ in the construction of the elements of V_h. Let V_h be spanned by N linearly independent functions $v_j(x)$, $j = 1,\ldots,N$. We represent the discrete approximation of the exact solution u by $u_h(x) = \sum_{j=1}^{N} u_j v_j(x)$, insert this into Equation 52.66:

$$a\left(v_h, \sum_{j=1}^{N} u_j v_j(x)\right) = b(v_h) \tag{52.67}$$

and get the linear system

$$i = 1, \ldots, N: \qquad \sum_{j=1}^{N} a_{ij} u_j = b_i \qquad \text{with } a_{ij} = a\left(v_i, v_j\right) \quad \text{and } b_i = b\left(v_i\right). \tag{52.68}$$

52.5.4.1 Linear Finite Elements

We want to construct the matrix $\left(a_{ij}\right)$ as a sparse matrix, i.e., a matrix with predominantly zero entries, which is a very useful numerical property. To this end we construct test functions v_j that exist only on small patches. The most common finite elements are the *piecewise linear "hat"*-functions shown in Figure 52.13a:

$$v_i(x) = \begin{cases} \frac{x - x_{i-1}}{h_{i-1}} & \text{for} \quad x \in \left[x_{i-1}, x_i\right] \\ -\frac{x - x_{i+1}}{h_i} & \text{for} \quad x \in \left[x_i, x_{i+1}\right] \\ 0 & \text{elsewhere,} \end{cases} \tag{52.69}$$

where we introduced the lengths of the subdomains with $h_{i-1} = x_i - x_{i-1}$ and $h_i = x_{i+1} - x_i$.

52.5.4.2 Assembly Process: Patchwise Computation of $A = \left(a_{ij}\right)$

We start from Equation 52.67 with the summation moved to front of $a(\cdot, \cdot)$:

$$i = 1, \ldots, N \qquad \sum_{j=1}^{N} a\left(v_i(x), v_j(x)\right) u_j = b(v_i).$$

Next we denote $a(\cdot, \cdot)$ restricted to patch K by $a(\cdot, \cdot)|_K = a_k(\cdot, \cdot)$ and rewrite the above in terms of a sum over patches:

$$i = 1, \ldots, N \qquad \sum_{j=1}^{N} \sum_{k} a_k\left(v_i(x), v_j(x)\right) u_j = b(v_i).$$

Interchanging the order of summation we have

$$i = 1, \ldots, N \qquad \sum_{k} \sum_{j=1}^{N} a_k\left(v_i(x), v_j(x)\right) u_j = b(v_i).$$

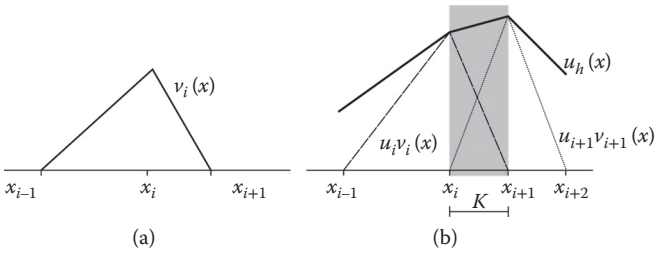


FIGURE 52.13 (a) Piecewise linear test function $v_i(x)$. (b) Piecewise linear function u_h composed from trial functions $v_i(x)$ and $v_{i+1}(x)$. Only trial functions v_i and v_{i+1} have support on the interval K, but no other trial function.

In principle, this is a summation over k different matrices $A_k = a_k\left(v_i(x), v_j(x)\right)$, each of dimension $N \times N$ multiplied with a vector $u = (u)_i$ of dimension N, hence

$$\left(\sum_k A_k\right) u = b(v_i).$$

The matrices A_k have full dimension. However, since they are computed only over a single patch K they have entries different from zero only if both v_i and v_j have support over K, hence most entries but a few are zero. Suppose we have a set of (few) trial functions $v_{i_1}, v_{i_2}, \ldots, v_{i_n}$ with index set $\{i_1, i_2, \ldots, i_n\}$ which contribute to A_k assigned to patch K. We count these indices as follows:

$$\begin{aligned} \text{local DOFs} &\rightarrow \text{global DOFs} \\ (1, 2, \ldots, n) &\rightarrow \left(i_1, i_2, \ldots, i_n\right) \end{aligned}$$

and store the connectivity data, using MATLAB ® notation, in an element connectivity vector

$$c = [\mathrm{i}_1, \mathrm{i}_2, \ldots, \mathrm{i}_n]. \tag{52.70}$$

We have such a mapping for each patch connecting the *local* DOFs to the *global* ones. With the help of the local DOFs we define dense, so-called *local* matrices $A_k^{\mathrm{loc}} \in \mathbb{C}^{n\times n}$ of small dimension n and corresponding *local* vectors $b_k^{\mathrm{loc}} \in \mathbb{C}^n$:

$$\begin{aligned} A_k^{\mathrm{loc}}(l, m) &:= A_k\left(i_l, i_m\right) \quad \text{for all indices with non-vanishing values of } A_k, \text{ and} \\ b_k^{\mathrm{loc}}(l) &:= b_k(i_l). \end{aligned}$$

The process of distributing the entries of the local matrices to the global matrices is called index scattering. The MATLAB ® statement

$$A(c, c) = A^{\mathrm{loc}}$$

scatters the entries of the local matrix A^{loc} to their global counterparts.

52.5.4.3 Computation of Local Matrices

For a convenient computation of the local vectors and matrices, it is useful to decompose the form $a(v, u)$, e.g., Equation 52.57, into the so-called *stiffness* and *mass* parts $s(v, u)$, $m(v, u)$, respectively:

$$a(v, u) = s(v, u) - m(v, u) \quad \text{with}$$

$$s(v, u) := \int_0^L \partial_x v^* \partial_x u \quad \text{and} \quad m(v, u) := \int_0^L k^2 v^* u.$$

The terms stiffness and mass come from mechanical systems and have no meaning in electrodynamics. But, as we will see, both parts have very different mathematical properties.

Let us compute the stiffness matrix based on the hat-functions. We consider the patch K from Figure 52.13 which gives us the local⟶global assignment of indices

$$(1, 2) \longrightarrow (i, i + 1).$$

Using Equation 52.69, we get

$$s^{\text{loc}} = \begin{pmatrix} s(v_i, v_i)\big|_K & s(v_i, v_{i+1})\big|_K \\ s(v_{i+1}, v_i)\big|_K & s(v_{i+1}, v_{i+1})\big|_K \end{pmatrix}.$$

We compute directly

$$s^{\text{loc}} = \frac{1}{h_i}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and the in same way} \quad m^{\text{loc}} = h_i \begin{pmatrix} \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} \end{pmatrix}.$$

The structure of a finite element algorithm is given in Algorithm 52.3.

52.5.5 2D and 3D: Weak Form and Variational Formulation

The Maxwell scattering problem with PML is derived in complete analogy to Section 52.5.3. Let the source field data $\mathbf{E}_{\text{source}}$ and $\nabla \times \mathbf{E}_{\text{source}}$ at the boundary $\partial\Omega$ of the computational domain Ω be given. The variational form of the scattering problem reads:

$$\text{Find } \mathbf{E} \in V: \quad a(\mathbf{v}, \mathbf{E}) = \int_{\partial\Omega} d\mathbf{s} \cdot \left(\mathbf{v}^* \times \left[\left(\mu^c\right)^{-1} \nabla \times \mathbf{E}_{\text{source}}\right]\right) + a_{\text{ext}}(\mathbf{v}, \mathbf{g}) \quad \text{for all } \mathbf{v} \in V, \tag{52.71}$$

with

$$\begin{aligned} a(\cdot,\cdot) &= a_{\text{int}}(\cdot,\cdot) + a_{\text{ext}}(\cdot,\cdot) \\ a_{\text{int}}(\mathbf{v}, \mathbf{E}) &= \int_{\Omega} dx\, dy\, \left(\nabla \times \mathbf{v}^*\right) \cdot \mu^{-1} (\nabla \times \mathbf{E}) - \omega^2 \int_{\Omega} dx\, dy\, \mathbf{v}^* \varepsilon \mathbf{E} \\ a_{\text{ext}}(\mathbf{v}, \mathbf{g}) &= \int_{\Omega_{\text{ext}}} dx\, dy\, \left(\nabla \times \mathbf{v}^*\right) \cdot \left(\mu^c\right)^{-1} \left(\nabla \times \mathbf{g}\right) - \omega^2 \int_{\Omega_{\text{ext}}} dx\, dy\, \mathbf{v}^* \varepsilon^c \mathbf{E} \end{aligned}$$

and $\mathbf{g}$ is an arbitrary tangentially continuous function with

$$\begin{aligned} \mathbf{g}(\partial\Omega) &= \mathbf{E}_{\text{source}}(\partial\Omega) \\ \mathbf{g}(\infty) &= 0. \end{aligned}$$

The transformed permeabilities and permittivities are

$$\left(\mu^c\right)^{-1} = \frac{1}{|J|} J^T \mu^{-1} J, \qquad \varepsilon^c = |J|\, J^{-1} \varepsilon J^{-T},$$

with J the Jacobian of the complex extension. Here, we let Ω_{ext} be the original, real exterior domain and the PML-complexification is hidden in the transformed materials μ^c and ε^c, see Equations 52.18 and 52.19.

52.5.6 2D and 3D: Linear Finite Elements, Local Matrices, and Assembly Process

The basic scheme of deriving the numerical algorithm is exactly the same as in 1D, see Section 52.5.4, and also Algorithm 52.3:

1. Construct a mesh that describes the geometry plus surrounding domain. The meshes in 2D are triangular or quadrilateral meshes, in 3D these are meshes based on tetrahedrons, bricks, prisms, and pyramids. Typically, special mesh generators are employed. The construction of high-quality meshes is often a nontrivial task and a research field in its own right.
2. On each patch of the mesh, say on a triangle, a polynomial space is defined which is the basis of the construction of so-called shape functions. The local space of shape functions must be capable of approximating the solutions on the patch. Given a polynomial space, different families of shape functions can be constructed which all span the same space. A local matrix and a local right-hand side vector is attributed to each patch. The computation of these involves mapping from reference elements.
3. The shape functions of neighboring patches are glued together to build the discrete test and trial functions v_i. These functions must be tangentially continuous across the patch boundaries to be a global basis function of the solution space for Maxwell's equations. One DoF is attributed to each v_i.
4. Having obtained all test and trial functions v_i, the assembly process is carried out as in 1D, see Equations 52.67 and 52.68. The scattering of local to global matrices depends on the neighborhood relation of patches.

Algorithm 52.3: Basic Finite Element Algorithm

construct a tessellation $\bar{\Omega} = \cup_i \bar{K}_i$
for all patches K_i **do**
  map $\widehat{K} \to K_i$
  compute local matrices $\mathbf{A}_i^{\text{loc}}$
  compute local vectors $\mathbf{b}_i^{\text{loc}}$
  scatter the local matrices and vectors to global ones
  update global matrices $\mathbf{A}$ and right-hand side $\mathbf{b}$
end for
solve $\mathbf{A}\mathbf{u}_h = \mathbf{b}$
return $\mathbf{u}_h$

52.5.6.1 Finite Elements

Figure 52.14 recalls the construction of the global test and trial functions from the local shape functions. The process is the same in higher dimensions, but there are many more possibilities for how to do it.

We illustrate these possibilities by means of some examples in 2D. First, we consider a mesh built from quadrilaterals, see Figure 52.15a. For a technical simplification of finite element algorithms, each quadrilateral is considered to be a map from a unit square, see Figure 52.16. The test and trial functions are constructed over the unit square and then mapped to their original positions. The lowest order trial functions are linear functions, in the case of a unit square, we use simply tensor products of 1D linear functions, see Section 52.5.4.

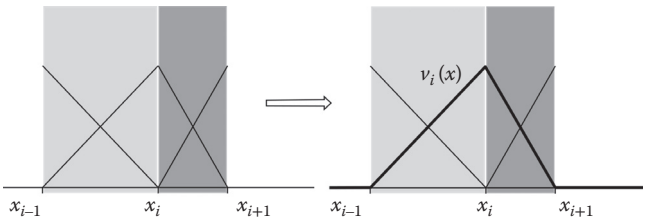


FIGURE 52.14 Construction of a global trial function v_i. Two neighboring intervals with two linear shape functions, each have the common boundary point x_i. One of the shape functions of each interval is taken and glued to the other and continued to a global, continuous trial function with local support.

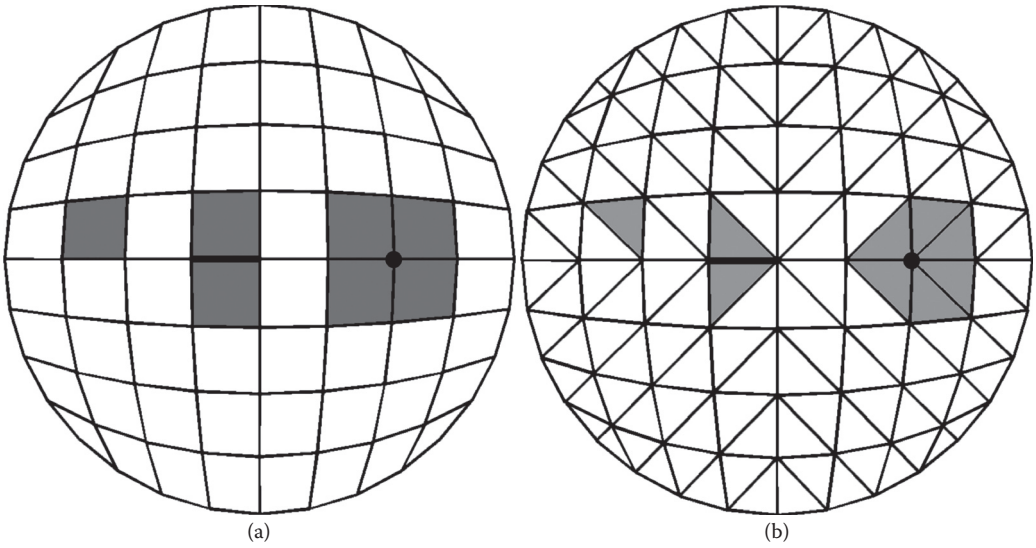


FIGURE 52.15 Structured mesh based on quadrilaterals (a) and a triangular mesh with the same vertices (b). The gray shaded areas show the geometric possibilities of constructing trial functions based on the interior of a single patch, neighboring patches with a common edge and neighboring patches with a common vertex.

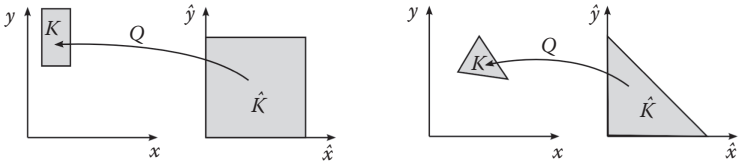


FIGURE 52.16 Mappings from reference elements to desired shapes.

52.5.6.2 Scalar Helmholtz Equation

Scalar trial functions v_i are used to approximate the solution of Helmholtz equations (Equation 52.8). In the case of a quadrilateral mesh with a unit square as reference element, the tensor product of the 1D shape functions gives four bilinear shape functions:

$$\left\{N_i : i = 1, \ldots, 4\right\} = \left\{\left(1-\hat{x}\right)\left(1-\hat{y}\right), \hat{x}\left(1-\hat{y}\right), \hat{x}\hat{y}, \left(1-\hat{x}\right)\hat{y}\right\}.$$

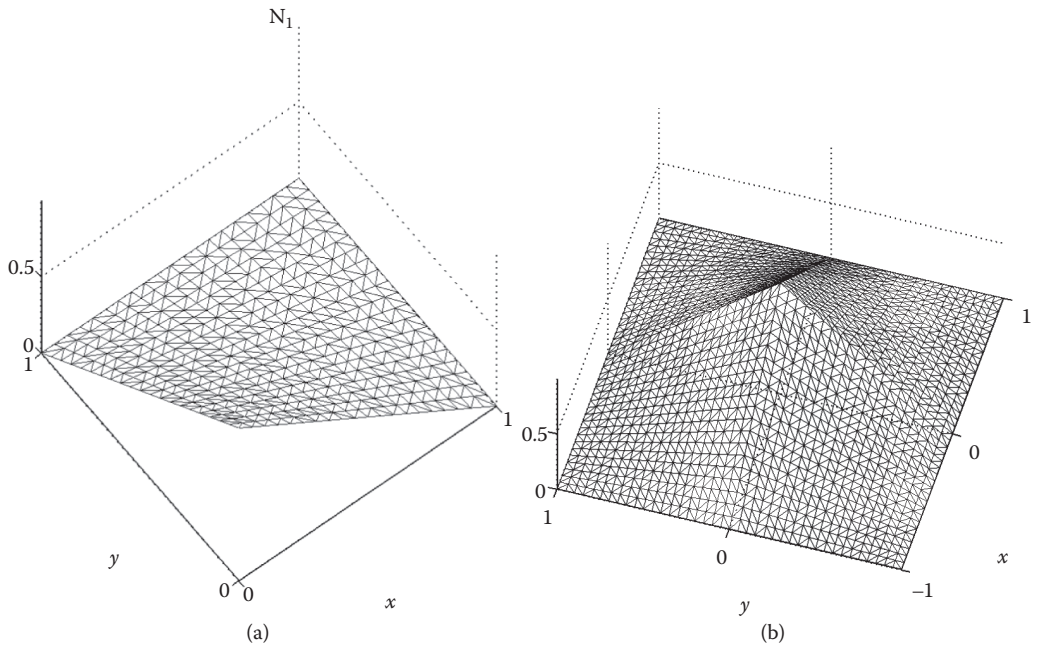


FIGURE 52.17 Bilinear shape function N_1 on a unit square and the composed trialfunction.

Figure 52.17a shows the shape function N_1, which is associated to the vertex $(0, 0)$ of the reference coordinate system $\widehat{K}$. In the spirit of the 1D construction (Figure 52.14), neighboring shape functions will be connected to continuous, global trial functions v_i as shown in Figure 52.17b. The common vertex is the defining feature of the resulting global trial function. The DoF attributed to this function is usually taken as its nodal value at the vertex position.

52.5.6.3 Maxwell's Equations

In the case of Maxwell's equations, the fields are generally no longer continuous. They are tangentially continuous but may have jumps or even singularities in normal field components at material discontinuities. Hence, nodal values, at least along material discontinuities, do not represent the physical situation properly. Therefore, the lowest order vectorial shape functions for Maxwell'equations are so-called edge functions. They are constructed in a way that they allow for an easy extension to tangential continuous, global trial functions. The four lowest order edge functions attributed to the unit square are shown in Figure 52.18.

52.5.6.4 Computation of Local Matrices and Assembly Process

This is very similar to the procedure discussed for the 1D Helmholtz problem Section 52.5.4. According to the Maxwell scattering formulation (Equation 52.71), we have to evaluate the variational formulation (Equation 52.71) on a finite dimensional test and trial space V_h. The computation of the local matrices and vectors goes the same way as in Section 52.5.4 with

$$a_{i,j} = a\left(\mathbf{v}_i, \mathbf{v}_j\right) = \underbrace{\int_{\Omega_{\mathrm{cd}}} dx\, dy\, \left(\nabla \times \mathbf{v}_i^*\right) \cdot \mu^{-1} \left(\nabla \times \mathbf{v_j}\right)}_{\text{stiffness part } s_{ij}} - \underbrace{\omega^2 \int_{\Omega_{\mathrm{cd}}} dx\, dy\, \mathbf{v}_i^* \varepsilon \mathbf{v}_j}_{\text{mass part } m_{ij}}$$

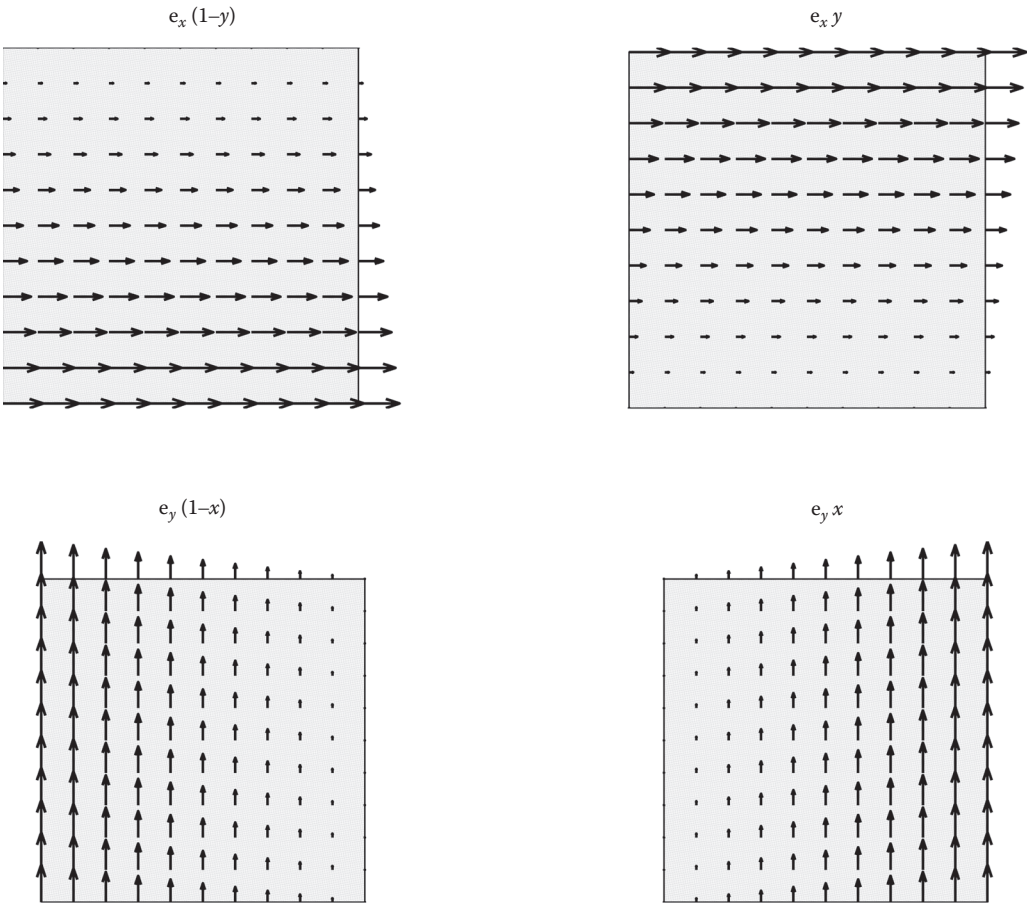


FIGURE 52.18 Vectorial, first-order edge elements of the unit square.

and

$$(b)_i = \int_{\partial\Omega} d\mathbf{f} \cdot \left(\mathbf{v}_i^* \times \left[\mu^{-1}\nabla \times \mathbf{E}_{\text{source}}\right]\right) + a_{\text{ext}}\left(\mathbf{v}_i, \mathbf{g}\right). \tag{52.72}$$

52.5.7 Convergence

The convergence analysis for edge element discretizations of time-harmonic scattering problems yields very useful insights into the behavior of numerical codes. The analysis itself is a challenging task since even in the lossless case the variational equations are neither elliptic nor definite. As a typical example we refer to Monk [34], who investigated a lossless scattering problem with sources on a bounded domain with electric boundary conditions. The error estimation uses the **curl** -norm

$$\left\|\mathbf{E} - \mathbf{E}_h\right\|_{\mathbf{curl}}^2 := \left\|\mathbf{E} - \mathbf{E}_h\right\|^2 + \left\|\nabla \times (\mathbf{E} - \mathbf{E}_h)\right\|^2,$$

where $\|\cdot\|$ is the standard L^2-norm, h denotes again a typical meshsize and as subscript it indicates discrete approximation. Depending on the properties of the computational domain and the smoothness of the exact

solution there are constants C and δ independent of h and a constant $h_0 > 0$ independent of $\mathbf{E}$ and $\mathbf{E}_h$ such that for all $0 < h < h_0$

$$\left\|\mathbf{E} - \mathbf{E}_h\right\|_{\mathbf{curl}} \leq \frac{1}{1 - Ch^{1/2+\delta}} \inf_{\mathbf{v}_h \in V_h} \left\|\mathbf{E} - \mathbf{v}_h\right\|_{\mathbf{curl}}.$$

This result shows that for h less than a threshold value h_0, the FEM scheme converges to the exact solution, but in the worst case without a polynomial rate. However, if the solution is smooth quantified as $\mathbf{E} \in H^s(\mathbf{curl}\,; \Omega)$ and s is a smoothness parameter, then, we obtain in fact the desired polynomial convergence if we use high-order finite elements

$$\left\|\mathbf{E} - \mathbf{E}_h\right\|_{\mathbf{curl}} \leq Ch^s.$$

52.6 DG Method

The FDTD method has become the method of choice in physics and engineering when time-dependent Maxwell's equations have to be solved. The reasons are its simplicity, robustness and sufficient accuracy for most practical problems. On the other hand, it has severe limitations. These are mainly the difficultie incorporating complicated geometry due to its link to staggered grids unless the grid is prohibitively fine, the inherent second-order accuracy in time and space. Both topics have been addressed successfully in the frame work of FDTM methods, see the section about FDTD. However, DG methods present a successful alternative with the two main features geometrical flexibility and high-order approximations both in time and space inherently anchored in the method from the beginning.

52.6.1 Historical Notes

As any other method described in this chapter, the DG method has a long history with many contributions from many sides. Reed and Hill [35] applied a first form of the DG method to the neutron transport equation in 1973.

Since then, it has undergone a fast development and finds numerous applications to e.g., the Euler equations of gas dynamics, the shallow water equations, the equations of magneto-hydrodynamics, the compressible Navier–Stokes equations, and Maxwell's equations only to mention a few; see e.g., [36] for an overview and analysis of many of these applications.

The combination of the DG method for spatial discretization and of an explicit Runge–Kutta scheme for time-integration has been introduced by Cockburn and Shu in several papers, [37] maybe being one of the latest. [38] gives an introduction to the DG method and contains a more complete list of references on the (RKDG) method; it can be downloaded online. We also mention [39–41], which is a review paper about RKDG methods, as references on the DG method. [42] gives a review of the DG method applied to nanophotonics.

52.6.2 Concept: Combine Adaptive Spatial FEM Discretization with the Time Treatment in Finite Volume Methods

The FEM treats successfully complicated geometries in a natural way. There is no basic obstacle in combining local high-order approximations on unstructured meshes composed from different elements like tetrahedrons, prisms, bricks, and pyramids. The difficulties here lie on the side of implementations and is linked to the complexity of the problem: nontrivial geometries have to be described, decomposed into geometric patches, a detailed and efficient bookkeeping has to be realized for fast traversing. A direct application of FEM to time-dependent problems are possible but has a severe drawback: due to the occurrence of the mass matrix in conjunction with the time-derivative, an inversion of the mass matrix is needed. This

can be done in several ways; nevertheless, it causes a problem not present in FDTD methods. The reason for the necessity to invert a mass matrix lies in the construction of finite elements. Typically, neighboring elements couple to each other in a way that the resulting system matrices are sparse but yield a global coupling of DoF which does not break down to structures which are cheaply invertible.

Finite volume methods, on the other hand, offer explicit time-integration schemes like those in FDTD methods. The key idea is to localize most of all spatial discretization into cells which are treated completely independent from each other. The cells are connected to each other by an explicit kind of continuity statement. This connection has to satisfy natural requirements like continuity or convergence properties but is not derived from first principles. There are many choices which can be adapted to given situations. The most intriguing feature is that by construction, each cell "sees" its own time evolution which results in an explicit scheme. The key idea of DG methods now is to combine this kind of time-treatment with a high-order spatial discretization localized to individual cell which are connected only by explicitly forced interface conditions on the cell boundaries.

52.6.3 Scalar Advection Equation as Model Equation: The Finite Volume Method in 1D

Consider the scalar advection equation

$$\begin{aligned} \partial_t u + \partial_x f(u) &= 0 \\ f(u) &= au, \quad a \text{ is a real constant,} \\ u(x,0) &= u_0(x), \quad x \in \mathbb{R}, \end{aligned} \tag{52.73}$$

where $\partial_x f(u)$ is the 1D model for divergence in higher dimensions $\nabla \cdot \mathbf{F}(\mathbf{u})$. The solution to Equation 52.73 is $u(x,t) = u_0(x-at)$. If $a > 0$, this is a right traveling wave, if $a < 0$, this is a left traveling wave. In order to analyze the stability of the FVM described below, we consider energy conservation. We multiply the wave equation (Equation 52.73) with u and integrate over a subdomain $\Omega_i = (x_i, x_{i+1})$:

$$\begin{aligned} \frac{d}{dt}\frac{1}{2}u^2 + \frac{d}{dx}\frac{1}{2}au^2 &= 0 \\ \frac{d}{dt}\int_{x_i}^{x_{i+1}} \frac{1}{2}u^2\,dx + \int_{x_i}^{x_{i+1}} \frac{d}{dx}\frac{1}{2}au^2\,dx &= 0 \\ \frac{d}{dt}\|u\|^2_{\Omega_i} &= -a\left(u^2\left(x_{i+1}\right) - u^2\left(x_i\right)\right). \end{aligned} \tag{52.74}$$

52.6.3.1 Direction of Information Flow

From Equation 52.74 we deduce the direction of information flow and stability under perturbation. Let $a > 0$, the wave is right propagating. At a fixed time, we consider the function $u(x)$ on the interval (x_i, x_{i+1}). Since (Equation 52.73) is linear, we can subtract a constant such that we obtain a new $u(x)$ with $u(x_i) = 0$. When we consider this value as a given zero boundary condition at the left side, we read from (Equation 52.74) $\frac{d}{dt}\|u\|^2_{\Omega_i} \leq 0$ independently from the right value. This right value might be the result of a computation, hence it might be contaminated by an error. Independent of contamination, there is no increase in the norm of u, hence a perturbation does not lead to instability. If we had chosen the right boundary to be the place for the exact boundary condition, the opposite would be the case. There would always be an increase in $\|u\|^2_{\Omega_i}$ in time.

52.6.3.2 Link to Finite Volume Methods

The whole computational domain Ω is decomposed into small control volumes $\bar{\Omega} = \cup\bar{\Omega}_i$, $\Omega_i \cap \Omega_j = 0$ for $i \neq j$. We consider one subdomain Ω_i and use the conservation form of Equation 52.73 applying Gauss's theorem:

$$\partial_t \int_{\Omega_i} u\, dV + \int_{\partial\Omega_i} f dS = 0, \tag{52.75}$$

where in higher dimensions the surface integral takes the form $\int_S \mathbf{f} \cdot \mathbf{n}\, dS$ and $\mathbf{n}$ is the outwardly directed unit normal vector. Next, let the surface $\partial\Omega_i$ of cell Ω_i be subdivided into surfaces S_k which connect cell i to neighboring cells. Let $|\Omega_i|$ be the volume of Ω_i. We introduce the cell average u_i as an unknown target quantity

$$\partial_t u_i + \frac{1}{|\Omega_i|} \sum_k \int_{S_k} f dS = 0 \tag{52.76}$$

$$u_i = \frac{1}{|\Omega_i|} \int_{\Omega_i} u\, dV. \tag{52.77}$$

On the 1D mesh this results in the numerical scheme:

$$\partial_t u_i + \frac{1}{|\Omega_i|} \left(f_{i+1/2} - f_{i-1/2}\right) = 0; \tag{52.78}$$

see Figure 52.19. As in the FDTD method, the solution is based on a combination of quantities defined in primal meshes (the function u_h at integer positions) and on dual meshes (the discrete flux f_h at half counting positions). The main question is how do we define the values of $f = au$ at positions $x_{i-1/2}$ and $x_{i+1/2}$?

The answer depends, as discussed above, on the direction of wave propagation. It is natural to select the points in the solution at current time that are "upwind" of the solution at the position i for increasing time to maintain the causal connection. Depending then on the direction in which the solution is translated, we have different numerical schemes.

52.6.3.3 Finite Volume Method with First-Order Upwind Scheme

If $a > 0$, there is propagation in the positive x-direction. It holds $u_{i-1/2} \approx u_{i-1}$ and $u_{i+1/2} \approx u_i$, hence $f_{i+1/2} = au_i, f_{i-1/2} = au_{i-1}$, and the scheme (Equation 52.78) would be

$$\partial_t u_i + \frac{a}{\Delta x}\left(u_i - u_{i-1}\right) = 0.$$

In contrast, if $a < 0$, there is propagation in the negative x-direction. It holds $u_{i-1/2} \approx u_i$ and $u_{i+1/2} \approx u_{i+1}$, hence $f_{i+1/2} = au_{i+1}, f_{i-1/2} = au_i$, and the scheme (Equation 52.78) would be

$$\partial_t u_i + \frac{a}{\Delta x}\left(u_{i+1} - u_i\right) = 0.$$

Let us fix position $x_{i+1/2}$. The approximation of the flux just used is called the numerical flux, f^{num},

$$f^{\text{num}}_{i+1/2}\left(u_i, u_{i+1}\right) = \begin{cases} au_i & \text{if } a > 0 \\ au_{i+1} & \text{if } a < 0. \end{cases} \tag{52.79}$$

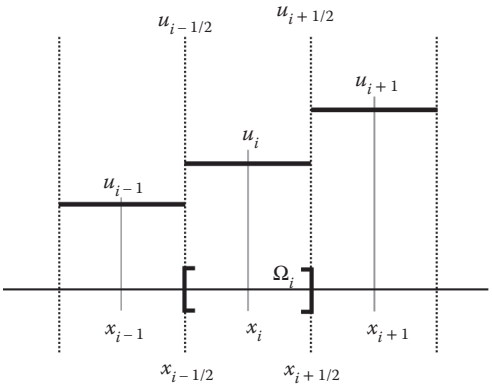


FIGURE 52.19 Vertex-centered finite volume method in 1D.

This can be compactly denoted as

$$f^{\text{num}}_{i+1/2}\left(u_i, u_{i+1}\right) = a \underbrace{\frac{1}{2}\left(u_i + u_{i+1}\right)}_{:=\{u\}_{i+1/2}} + \frac{|a|}{2} \underbrace{\left(u_i - u_{i+1}\right)}_{:=[\![u]\!]_{i+1/2}}$$

$$= a\{\!\{u\}\!\}_{i+1/2} + \frac{|a|}{2}\,[\![u]\!]_{i+1/2}\,.$$

We describe the typical finite volume discretization and give explicitly the 1D scheme:

1. Subdivide the domain into finite control volumes Ω_i. In one space dimension this is $\Omega_i = \left(x_{i-1/2}, x_{i+1/2}\right)$.
2. The conservation law is computed independently on each individual control volume Ω_i (52.76)

$$\partial_t u_i = -\frac{1}{|\Omega_i|}\sum_k \int_{S_k} f dS$$

with u_i the approximation to the average (Equation 52.77)

$$u_i = \frac{1}{|\Omega_i|}\int_{\Omega_i} u\, dV.$$

The resulting upwind scheme is, with flux (Equation 52.79),

$$\partial_t u_i = -\frac{1}{|\Omega_i|}\left(f_{i+1/2} - f_{i-1/2}\right)$$

$$= -\frac{1}{|\Omega_i|}\left(f^{\text{num}}_{i+1/2}\left(u_i, u_{i+1}\right) - f^{\text{num}}_{i+1/2}\left(u_{i-1}, u_i\right)\right). \tag{52.80}$$

3. Solve the semidiscrete system (Equation 52.80) with a suitable time-integrator. There are many possibilities, the most often used scheme in conjunction with time-dependent Maxwell's equations are low-storage Runge–Kutta methods. The resulting scheme after discretization in time must be a so-called monotonic scheme. Monotonic numerical schemes are ones which, given an initial distribution which is monotonic, produce a monotonic distribution after advection. A consequence of this property is that monotonic schemes neither create new extrema in the solution nor amplify existing extrema.

52.6.3.4 DG Method in 1D

The goal is to replace the low-order (piecewise constant) solution by a high-order approximation. We consider cell Ω_i. Exactly as in the FEM, we multiply (Equation 52.73) with a test function $v \in V$:

$$\int_{\Omega_i} v\partial_t u \, dx + \int_{\Omega_i} v\partial_x f(u) \, dx = 0.$$

An integration by parts gives

$$\int_{\Omega_i} v\partial_t u \, dx - \int_{\Omega_i} \partial_x v f(u) \, dx = - \int_{\partial\Omega_i} v f(u) \, ds.$$

Again as in FEM, we obtain the numerical method in *weak form* if we restrict the general space V to a finite dimensional space (of possibly high order) $V_h \subset V$:

$$\int_{\Omega_i} v_h \partial_t u_h \, dx - \int_{\Omega_i} \partial_x v_h f(u_h) \, dx = - \int_{\partial\Omega_i} v_h f^{\text{num}}(u_h) \, ds \quad \text{for all } v_h \in V_h. \tag{52.81}$$

The corresponding strong version results, if we undo the partial integration:

$$\int_{\Omega_i} v_h \partial_t u_h \, dx + \int_{\Omega_i} v_h \, \partial_x f(u_h) \, dx = \int_{\partial\Omega_i} v_h \left(f\left(u_h\right) - f^{\text{num}}(u_h) \right) \, ds \quad \text{for all } v_h \in V_h. \tag{52.82}$$

52.6.4 Maxwell's Equations in Conservation Form

We generalize the above step by step and present the framework of DG for Maxwell's equation's in 3D. The matrix-valued flux function $\mathbf{F} := \left(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\right)$ is composed of three column vectors of length n, $\mathbf{F}_i : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^n$, with $i = 1, 2, 3$, and $\mathbf{F}_i = \left(F_{1i}, \ldots, F_{ni}\right)^T$. The divergence $\nabla \cdot \mathbf{F}$ of this matrix-valued field $\mathbf{F} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^{n\times 3}$ is defined as

$$\nabla \cdot \mathbf{F} := \begin{pmatrix} \partial_{x_1} F_{11} & +\partial_{x_2} F_{12} + & \partial_{x_3} F_{13} \\ & \vdots & \\ \partial_{x_1} F_{n1} & +\partial_{x_2} F_{n2} + & \partial_{x_3} F_{n3} \end{pmatrix}.$$

The generalization of the 1D advection equation (Equation 52.73) reads

$$\begin{aligned} \mathbf{Q}\partial_t u + \nabla \cdot \mathbf{F}(\mathbf{u}) &= 0 \\ u(x, 0) &= u_0(x), \quad x \in \mathbb{R}^3. \end{aligned} \tag{52.83}$$

Next, we bring Maxwell's equations into this conservation form. From

$$\begin{pmatrix} \varepsilon & \\ & \mu \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} & \nabla\times \\ -\nabla\times & \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}$$

we read

$$\mathbf{Q} = \begin{pmatrix} \varepsilon & \\ & \mu \end{pmatrix} \qquad \mathbf{F}_i = \begin{pmatrix} -\mathbf{e}_i \times \mathbf{H} \\ \mathbf{e}_i \times \mathbf{E} \end{pmatrix}$$

with $\mathbf{e}_i$ being one of the three Cartesian unit vectors. Hence, there is in fact a conservation form (Equation 52.83) of Maxwell's equations.

52.6.5 Concept of DG-Algorithms for 3D Maxwell's Equations

The 3D algorithm follows the 1D concept. The only missing link is the upwind flux for the situation of Maxwell's equations. Following the same strategy as in 1D, Hesthaven and Warburg [43] derive the following expression:

$$\mathbf{n} \cdot \left(\mathbf{F} - \mathbf{F}^{\text{num}}\right) = \begin{pmatrix} \bar{Z}^{-1}\mathbf{n} \times \left(-Z^{+}\,[\![\mathbf{H}]\!] + \mathbf{n} \times [\![\mathbf{E}]\!]\right) \\ \bar{Y}^{-1}\mathbf{n} \times \left(-Y^{+}\,[\![\mathbf{E}]\!] + \mathbf{n} \times [\![\mathbf{H}]\!]\right) \end{pmatrix}. \tag{52.84}$$

Here Z and Y are the wave impedance and admittance, respectively,

$$Z^{\pm} = \sqrt{\frac{\mu^{\pm}}{\varepsilon^{\pm}}}, \quad Y^{\pm} = \left(Z^{\pm}\right)^{-1}, \quad \bar{Z} = \frac{Z^{+} + Z^{-}}{2}, \quad \bar{Y} = \frac{Y^{+} + Y^{-}}{2}.$$

The superscript "−" denotes the limit from the interior to the cell boundary, and the superscript "+" the limit from the neighboring cell to the boundary.

Algorithm 52.4: Basic Discontinuous Galerkin Algorithm

1. Discretization in space
construct a tessellation $\bar{\Omega} = \cup_i \bar{\Omega}_i$
for all patches Ω_i **do**
  map $\widehat{K} \to K_i$
  compute local matrices on local volumes
  assemble them (independently) to a global matrix
  compute local matrices on interior interfaces
  connect them to the global matrix
  compute local matrices on boundary interfaces
  connect them to build up the right-hand side
end for
2. Time evolution
use a time integrator (e. g., Runge–Kutta, leap-frog) to solve the semidiscrete, block-diagonal system in time
return vector of field values and fluxes in time

52.6.6 Convergence

We consider Maxwell's equations with homogeneous, isotropic materials and a convex domain $\Omega \subset \mathbb{R}^3$ and arbitrary smooth initial data. This is a simplified model case which has no singularities. In this case optimal L^2-convergence for a fixed time interval is obtained [44]:

$$\left\|\mathbf{u} - \mathbf{u}_h\right\| \leq Ch^{p+1}.$$

This is the same as we would have, under the same conditions, for finite elements.

Symbol Definition

$\mathbf{r}^0$ Vector of unit length.

Ω Computational domain.

$\partial\Omega$ Boundary of Ω.

$\mathbf{E}_\perp$ $\mathbf{e}_x E_x + \mathbf{e}_y E_y$

$\nabla_\perp$ $\mathbf{e}_x \partial_x + \mathbf{e}_y \partial_y$

∇_{k_z} $\mathbf{e}_x \partial_x + \mathbf{e}_y \partial_y + ik_z \mathbf{e}_z$

$\iota_{e,f}$ incidence number edge-face

c speed of light

$\mathbf{k}$ wavevector

$\mathbf{r}$ position

$[\![f]\!]$ Töplitz matrix generated by f

$[\![\cdot]\!]$ jump across a face

$\{\!\{\cdot\}\!\}$ average across a face

References

1. Andrei V Lavrinenko, Jesper Lagsgaard, Niels Gregersen, Frank Schmidt, and Thomas Søndergaard. *Numerical Methods in Photonics*. Boca Raton, FL: CRC Press, 2014.
2. Junji Yamauchi. *Propagating Beam Analysis of Optical Waveguides*. Electronic and Electrical Engineering Research Studies. Optoelectronics and Microwaves Series. Exeter, UK: Research Studies Press, 2003.
3. Jean-Claude Nédélec. *Acoustic and Electromagnetic Equations: Integral Representations for Harmonic Problems*, Vol. 144. Springer Science & Business Media, 2013.
4. Richard A Norton and Robert Scheichl. Planewave expansion methods for photonic crystal fibres. *Applied Numerical Mathematics*, 63:88–104, 2013.
5. Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994.
6. Saul Abarbanel and David Gottlieb. On the construction and analysis of absorbing layers in CEM. *Applied Numerical Mathematics*, 27(4):331–340, 1998.
7. Stephen D Gedney. An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices. *IEEE Transactions on Antennas and Propagation*, 44(12):1630–1639, 1996.
8. Weng Cho Chew and William H Weedon. A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates. *Microwave and Optical Technology Letters*, 7(13):599–604, 1994.
9. Fernando L Teixeira and Weng C Chew. General closed-form PML constitutive tensors to match arbitrary bianisotropic and dispersive linear media. *IEEE Microwave and Guided Wave Letters,* 8(6):223–225, 1998.

10. Kane S Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas Propagation*, 14(3):302–307, 1966.
11. Thomas Weiland. A discretization model for the solution of Maxwell's equations for six component fields. *Archiv Elektronik und Uebertragungstechnik*, 31:116–120, 1977.
12. Adi Ditkowski, K Dridi, and Jan S Hesthaven. Convergent Cartesian grid methods for Maxwell's equations in complex geometries. *Journal of Computational Physics*, 170(1):39–80, 2001.
13. Jan S Hesthaven. High-order accurate methods in time-domain computational electromagnetics: A review. *Advances in Imaging and Electron Physics*, 127:59–125, 2003.
14. Allen Taflove. *Advances in Computational Electrodynamics, the Finite-Difference time Domain*. Boston: Artech House, 1998.
15. Alain Bossavit. 'Generalized finite differences' in computational electromagnetics. *Progress in Electromagnetics Research*, 32:45–64, 2001.
16. Markus Clemens and Thomas Weiland. Discrete electromagnetism with the finite integration technique. *Progress in Electromagnetics Research*, 32:65–87, 2001.
17. Allen Taflove and Susan C Hagness. *Computational Electrodynamics: The Finite Difference Time-Domain Method*, 3rd edition. Boston, MA: Artech House Publishers, 2005.
18. Anders Bondeson, Thomas Rylander, and Par Ingelstrm. *Computational Electromagnetics*, Vol. 51. New York, NY: Springer Verlag, 2005.
19. Peter Monk and Endre Süli. A convergence analysis of Yee's scheme on nonuniform grids. *SIAM Journal on Numerical Analysis*, 31(2):393–412, 1994.
20. Theodor Tamir, HC Wang, and AA Oliner. Wave propagation in sinusoidally stratified dielectric media. *IEEE Transactions on Microwave Theory and Techniques*, 12(3):323–335, 1964.
21. C Yeh, KF Casey, and ZA Kaprielian. Transverse magnetic wave propagation in sinusoidally stratified dielectric media. *IEEE Transactions on Microwave Theory and Techniques*, 13(3):297–302, 1965.
22. MG Moharam and TK Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *Journal of the Optical Society of America*, 71(7):811–818, 1981.
23. Philippe Lalanne and G Michael Morris. Highly improved convergence of the coupled wave method for tm polarization. *JOSA A*, 13(4):779–784, 1996.
24. G Granet and B Guizal. Efficient implementation of the coupled-wave method for metallic lamellar gratings in tm polarization. *Journal of the Optical Society of America A*, 13(5):1019–1023, 1996.
25. Lifeng Li. Use of Fourier series in the analysis of discontinuous periodic structures. *Journal of the Optical Society of America A*, 13(9):1870–1876, 1996.
26. Jens Küchenmeister. Three-dimensional adaptive coordinate transformations for the Fourier modal method. *Optics Express*, 22(2):1342–1349, 2014.
27. Peter Monk. *Finite element Methods for Maxwell's Equations*. Oxford University Press, 2003.
28. Leszek Demkowicz. *Computing with hp-Adaptive Finite Elements: One and Two Dimensional Elliptic and Maxwell Problems*, Vol. 1. CRC Press, 2006.
29. Jean-Claude Nédélec. Mixed finite elements in R^3. *Numerische Mathematik*, 35(3):315–341, 1980.
30. Jean-Claude Nédélec. A new family of mixed finite elements in R^3. *Numerische Mathematik*, 50:57–81, 1986.
31. Alain Bossavit. Whitney forms: A class of finite elements for three-dimensional computations in electromagnetism. *IEE Proceedings A (Physical Science, Measurement and Instrumentation, Management and Education, Reviews)*, 135(8):493–500, 1988.
32. Jon P Webb. Hierarchal vector basis functions of arbitrary order for triangular and tetrahedral finite elements. *IEEE Transactions on Antennas and Propagation*, 47(8):1244–1253, 1999.
33. Jan Pomplun, Sven Burger, Lin Zschiedrich, and Frank Schmidt. Adaptive finite element method for simulation of optical nano structures. *Physica Status Solidi (B)*, 244(10):3419–3434, 2007.
34. Peter Monk. A simple proof of convergence for an edge element discretization of Maxwell's equations. In *Computational Electromagnetics*, pp. 127–141. Springer, 2003.

35. WH Reed and TR Hill. Triangular mesh methods for the neutron transport equation. Report LA-UR-73-479. Los Alamos, NM: Los Alamos Scientific Laboratory, 1973.
36. Bernardo Cockburn, George E Karniadakis, and Chi-Wang Shu. *The Development of Discontinuous Galerkin Methods*. Lecture notes in computational science and engineering. Berlin, Heidelberg: Springer-Verlag, 2000.
37. Bernardo Cockburn and Chi-Wang Shu. The Runge–Kutta discontinuous Galerkin method for conservation laws V: Multidimensional systems. *Journal of Computational Physics*, 141(2):199–224, 1998.
38. Bernardo Cockburn. An introduction to the discontinuous Galerkin method for convection-dominated problems. In *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, pp. 151–268. Berlin Heidelberg: Springer, 1998.
39. Bernardo Cockburn, George E Karniadakis, and Chi-Wang Shu. *Discontinuous Galerkin Methods: Theory, Computation and Applications*. Berlin Heidelberg: Springer, 2000.
40. Bernardo Cockburn and Chi-Wang Shu. Runge–Kutta discontinuous Galerkin methods for convection-dominated problems. *Journal of Scientific Computing*, 16(3):173–261, 2001.
41. Jan S Hesthaven and Tim Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer Science & Business Media, 2007.
42. Kurt Busch, Michael Koenig, and Jens Niegemann. Discontinuous Galerkin methods in nanophotonics. *Laser & Photonics Reviews*, 5(6):773–809, 2011.
43. Jan S Hesthaven and Timothy Warburton. Nodal high-order methods on unstructured grids: I. Time-domain solution of Maxwell's equations. *Journal of Computational Physics*, 181(1):186–221, 2002.
44. Marcus J Grote, Anna Schneebeli, and Dominik Schtzau. Interior penalty discontinuous Galerkin method for Maxwell's equations: Optimal l2-norm error estimates. *IMA journal of Numerical Analysis*, 28(3):440–468, 2008.

Index

A

Absorber relaxation time, 212–213
Absorber-to-gain cross section ratio, 214
Absorption dip, 724
Absorption spectrum, 220, 319, 364–367, 504
ACML, *see* Anticolliding mode-locking
Active ML, 184, 185, 187
Active nonlinear optical waveguide/cavity properties, 678–680
Additive pulse ML, 184
AEI, *see* Anisotropic exchange interaction
AM ML, *see* Amplitude-modulation ML
Amorphous silicon (a-Si), 498–499
Amplification, 87, 187, 215, 680
Amplified spontaneous emission (ASE), 84, 92, 623
Amplitude-modulation ML (AM ML), 183, 226
Andronov–Hopf bifurcation, 197–198
Angle of incidence (AOI), 353, 389, 392, 421, 478
Anisotropic exchange interaction (AEI), 703
Antibunching, 662
Anticolliding mode-locking (ACML), 216
Antireflection coating (ARC), 393, 446, 479, 480 485, 487
AOI, *see* Angle of incidence
APDs, *see* Avalanche photodiodes
ARC, *see* Anti reflection coating
Artificial interfaces, 747
ASE, *see* Amplified spontaneous emission
Auger recombination, 6
  process, 453, 542, 545, 550, 552
  rate, 317
Avalanche photodiodes (APDs), 295–297, 633

B

BA, *see* Broad-area
Backward differentiation formula (BDF) methods, 763
Backward propagation, 65, 73
Bandgap edge, 715–716, 723
Band structure, 249–250, 338, 363, 424–429, 529, 695, 776
Beam parameter product (BPP), 59, 60, 82, 83
Beam propagation factor, 82
Beam propagation method (BPM), 41–42, 65, 310, 314–315, 674, 808
Beam propagation ratio, 59–61
Beam quality, 59, 61, 66, 76, 82–83
Beam spoilers, 64
  GG tapered laser with, 68–72
  GG tapered laser without, 73–75
Beer–Lambert law, 482, 485, 504, 505, 510
Bernoulli function, 741, 757
Bessel functions, 271
BHJ, *see* Bulk heterojunction
Bidirectional symmetric filter, 135
Bipolar junction transistor (BJT), 300
Blakemore approximation, 757, 758
Bloch spin vector equation of motion, 667
Bloch theorem, 561, 566, 567
Boltzmann approximation, 52, 735, 740, 746
  numerical charge carrier fluxes, 756
Boltzmann transport equation (BTE), 244, 451, 779
Box–Müller method, 708
BPM, *see* Beam propagation method
BPP, *see* Beam parameter product
Bragg grating, 16, 32, 116, 117, 119, 156, 159, 209
Bragg stop-band width, 117–118
Bragg waveguide gratings, 37
Bragg waveguide SLM laser, 141–144
Brillouin zone (BZ), 695
Broad-area (BA)
  devices, 59
  lasers, 66, 85–86
BTE, *see* Boltzmann transport equation
Built-in potential, 736, 748
Built-in voltage, 737, 749
Bulk heterojunction (BHJ), 449, 502
Butler–Volmer equation, 527
BZ, *see* Brillouin zone (BZ)

C

Cadmium telluride (CdTe) cells, 497, 499–501
Carrier capture, 34–35, 210, 451, 454

Carrier density, 10, 71, 128, 163–165, 744
Carrier dynamics, in quantum-dot systems
carrier–carrier Coulomb interaction, 643–646
carrier–phonon interaction, 642–644
carrier population functions, 641
Fermi–Dirac distribution functions, 639, 640
scattering process, 640
Carrier leakage, 8, 9, 19, 20
Carrier rate equations, 167–169
Carrier reservoir (CR), 167–169
Carrier-selective contacts, 451–452
Cartesian coordinate system, 265, 273, 566
Cascading scheme, 236, 240
Casimir effect, 662
Catastrophic optical mirror damage (COMD) threshold, 85
Cavity field quantization, 665
Cavity modes, 41–42, 96, 144, 623, 809
in Lang–Kobayashi (LK) model, 178
resonant, 676
Cavity quantum electrodynamics (CQED), 594–596, 665–666
CdTe cells, *see* Cadmium telluride cells
CFL conditions, *see* Courant–Friedrichs–Lewy conditions
Charge carrier escape, 449–451
Chiral carbon nanotubes (CNTs), optical rotation in
chirality and optical activity, 712
density of states, 716
dipole optical intersubband transitions, 713, 715
energy dispersion, 713–714
Faraday rotation and magneto-chiral effects, 721–724
single-walled CNTs, 712–713, 717
Chirality, 712–714, 717
CIC scheme, *see* Cloud-in-cell scheme
CIGS cells, *see* Copper indium gallium diselenide cells
Circularly polarized electric-field components, 704–705
Circularly polarized optical-pulse interactions
degenerate four-level systems, 695–699
two-level quantum systems, 692–695
Circularly polarized ultrashort pulse, 715, 720, 724
Cleaved-coupled-cavity (C^3) laser, 136–137
Cloud-in-cell (CIC) scheme, 792, 796
CMOS devices, *see* Complementary metal-oxide-semiconductor devices
Coherent control, semiconductor nanostructures, 664
Coherent Maxwell-pseudospin equations, 673–674
Coherent population
effects, 225–226
trapping, 724
Coherent pulse propagation, 664
Coherent ray tracing, 93
laser diodes simulation tool coupling with, 94–95
Coherent spin manipulation, 701–707
Coherent vector Maxwell-pseudospin formalism, 717
Colliding pulse mode-locking (CPM), 216
COMD threshold, *see* Catastrophic optical mirror damage threshold
Complementary metal-oxide-semiconductor (CMOS) devices, 773, 775, 776
Complex refractive index, 320, 421–422
Compound cavity, 205, 216
Compressively strained multiple quantum wells (CS-MQW), 127
Computed band edge energy, 744
Concentrating photovoltaics (CPV), 410
Confinement factor, 84
Contact terminal currents, 763–765
Continuous wave (CW)
excitation, 638
operation, 8, 16
simulation tool, 82
Continuous-wave-excited lasers, 638–639
Control volumes, 753–754
Conventional lasers, 629
Conversion efficiency, 23
Copper indium gallium diselenide (CIGS) cells, 497, 500, 501
Coupled-cavity SLM lasers
band-pass filtering model, 133–136
cleaved-coupled-cavity laser, 136–137
discrete mode laser, 139
etched slotted laser, 137–139
general optical filtering model, 133–136
governing equations, 139–140
performance comparisons among, 140–141
Coupled laser systems, 170–172
Coupled Schrödinger–Poisson solver, 240, 458
Coupling losses, 20, 25, 69, 193, 203
Courant–Friedrichs–Lewy (CFL) conditions, 821–822
CPM, *see* Colliding pulse mode-locking
CPV, *see* Concentrating photovoltaics
CQED, *see* Cavity quantum electrodynamics
CR, *see* Carrier reservoir
Cross-talk heating effects, 163
Crystalline silicon solar cells
models for, 403–406
numerical simulation of, 402–403
CS-MQW, *see* Compressively strained multiple quantum wells
CW, *see* Continuous wave

D

Dark current, 338–341
DBR, *see* Distributed Bragg reflector
DBR-tapered laser, *see* Distributed Bragg reflector-tapered laser
DDE model, *see* Delay-differential equation model
Decimation technique, 220, 223–225

Defect voltage, 18, 19, 22
Delaunay triangulation, 761–762
Delay-differential equation (DDE) model, 154, 193–206
Density-dependent nonlinear factor, 751
Density matrix model, 239
Density of states (DOS), 222, 242, 342, 403, 442, 513, 563, 749–750
Dephasing time, 226, 239, 664, 669, 679, 708, 710
Detector QE modeling, 347–348
  edge coupling, for quantum well infrared photodetectors, 351–353
  InAs/GaInSb IR photodiode, 350–351
  modeling validation, 348–349
  photonic crystal slab quantum well infrared photodetectors, 355–356
  planar solar cell, 349–350
  plasmonic-enhanced quantum well infrared photodetectors, 353–355
  quantum grid infrared photodetector, 353
Device-level electrothermal simulation, 247–249
  band structure, 249–250
  heat generation rate, 251–252
  J-F curve, 250
  modal gain, 250–251
  results, 252–254
  threshold, 250–251
DFB laser, *see* Distributed feedback laser
DG method, *see* Discontinuous galerkin method
DG transistor, *see* Double-gate transistor
Dielectric permittivity, 209, 218, 422, 717, 735
Diffusion enhancement, 751
Diffusion time, of carriers, 288
Diffusion voltage, 395
Dipole length scale, 672
Dipole matrix element, 250, 668, 689, 703, 705, 723
Direct-diode laser systems, 81, 82, 86, 87, 106
Direct gap semiconductors, 285
Dirichlet boundary conditions, 800–801
Discontinuous galerkin (DG) method, 808, 809
  algorithm, 849
  convergence analysis, 849–850
  finite volume methods, 844–847
  in one dimensional, 848
  for three dimensional Maxwell's Equations, 848–849
Discrete mode laser, 139
Discrete nonlinear Poisson problem, 738–739
Discretizing van Roosbroeck system
  control volumes, 753–754
  existence, uniqueness, and convergence, 759
  finite volume method in higher dimensions, 754–756
  flux expressions with thermodynamic equilibrium, 758
  free energy and dissipation rate, 758–759
  maximum principle, 759
  Scharfetter–Gummel fluxes and non-Boltzmann generalizations, 756–758
  2D simulation, 753–754
Dispersion, 38, 821–822
Dispersive element, 189, 223
Displacement current density, 54, 764
Dissipation rate, 751
Dissipative dynamics, 669–671
Distributed Bragg reflector (DBR), 60, 261, 563, 571, 595–596
  laser, 12, 16, 118–120, 226
  mirror, 676
Distributed Bragg reflector (DBR)-tapered laser, 87, 89, 91
  analysis and discussion, 102–105
  parasitic reflections with off-axis alignment, 102
  parasitic reflections with on-axis alignment, 97–102
  unintentional reflections on, 95–97
Distributed feedback (DFB) laser, 12, 16, 111, 120–125
Distributed time-domain, 206–210
Doping, 250, 296, 346, 755
  optimization, 767
Double-gate (DG) transistor, 776
Double-trench resonant tunneling SLM laser, 144–146
Drift-diffusion calculation, 423–424
  interface recombination velocities, 429–430
  intrinsic carrier concentration, 424–429
  tunnel diodes, 430–431
Drift-diffusion models, 309, 530, 543–544
  contact terminal currents, 763–765
  definition of, 511
  density of states, 514–516
  finite element method, 765–766
  finite volume method, 761–763
  hopping process, 517, 518
  nonlinear solvers, 759–761
  1D van Roosbroeck Model, 735–736
  parabolic approximation, 514
  semiconductor device equations, 512–513
  thermodynamic formulation, 517
  time discretization, 763
  van Roosbroeck system, *see* van Roosbroeck system
Drift velocity, 289
Drude–Lorentz model, 610–611
DSC, *see* Dye-sensitized solar cell
Dyadic Green's function, 591, 592, 665
Dye-sensitized solar cell (DSC), 500–502, 527–528

E

EA, *see* Evolutionary algorithm
EAM, *see* Electroabsorption modulator
EBL, *see* Electron-beam lithography
Edge coupling, for quantum well infrared photodetectors, 351–353

Edge-emitting lasers, 36, 133, 193, 202, 261–263, 563–564, 631
Edge-emitting semiconductor laser diodes, 110, 125, 153
Effective interface roughness, 244–245
Efficiency analysis, 8–9
Einstein relation, 403, 423, 513, 514, 516, 742
Einstein summation convention, 567
Electrical energy, 395–396
Electrical phenomena, 262
Electric displacement flux, 755–756
Electricfield envelope, 684
Electroabsorption modulator (EAM), 187, 364
Electroluminescent (EL) refrigeration
  active layer thickness, 551–552
  boundary conditions, 546–547
  cooling cycling for, 542–543
  device structure and equations, 543–545
  doping types and concentrations on, 552–553
  feasibility of, 541–542
  GaAs-based, 542, 557–558
  leakage current and carrier blocking layer, 549–551
  light extraction, 553–554
  photon recycling, 554–557
  recombination and generation processes, 545–546
  steady-state analysis, 547–549
Electromagnetic theory, 477, 478
Electron-beam lithography (EBL), 118
Electron emission, of longitudinal optical phonon, 236
Electron-hole pairs, 3, 7
  photogeneration of, 445–449
Electronic bandstructure solver, 240–241
Electronic nanostructures, in solar cell, 481
Electronic transport, 238
  in mid-infrared quantum cascade lasers, 237
  nonequilibrium phonons, ensemble Monte Carlo with, 240–243
  quantum techniques, 239–240
  semiclassical techniques, 238–239
  in terahertz quantum cascade lasers, 237
Electron leakage current, 33–34
Electron spin-flip rate, 701
Electro-optic effect modulators, 371–372
  Mach–Zehnder modulator, 372–374
  traveling-wave design, 374
Electrorefractive modulator, 372, 374–375
EL refrigeration, *see* Electroluminescent refrigeration
EMC, *see* Ensemble Monte Carlo
Energy conservation, 53–55
Energy-level diagram, two-level quantum systems, 692
Energy yield, 409
Ensemble Monte Carlo (EMC), 239, 779, 799
  with nonequilibrium phonons, 240
Entanglement, 663, 665
EQE, *see* External quantum efficiency
Etched slotted laser, 137–139
Evolutionary algorithm (EA)
  description of, 616–617
  flowchart, 617
  zero group velocity points, 617–618
Exciton dissociation, 449
External cavity simulation tool, 93
External differential efficiency, 26
External quantum efficiency (EQE), 288, 398, 433, *See also* Slope efficiency

F

Fabry–Pérot (FP) laser, 16, 28, 111–113
  cavity, 5, 6
  edge-emitting, 203
Faraday rotation, 721–724
Fast-axis collimation (FAC) lens, 96
Fast Fourier transform (FFT), 315
FCA, *see* Free-carrier absorption
FCCC model, *see* Full cavity coherent coupling model
FCIC model, *see* Full cavity incoherent coupling
FD-SOI device technology, *see* Fully-depleted silicon-on-insulator device technology
FDTD, *see* Finite-difference time-domain (FDTD)
FEM, *see* Finite element method
Fermi–Dirac distribution function, 504, 513, 546, 639, 640, 748, 749
Fermi's golden rule, 242, 446, 504, 592
Fermi voltage, 18, 29, 34
FET, *see* Field-effect transistor
FF, *see* Fill factor
FFT, *see* Fast Fourier transform
Field-effect transistor (FET), 774
Filamentation
  longitudinal-lateral approximate projected equations, 42–43
  nonlinearities due to virtual transitions, 44–45
  nonlinearity induced by lateral spatial hole burning, 45–47
Fill factor (FF), 398
Final field equation, 39–41
FinFET transistor, 776
Finite-difference method, *see* Shooting method
Finite-difference time-domain (FDTD)
  algorithmic features and drawbacks, 817
  convergence, 823
  discrete differential operators, 818–820
  discrete div-conditions, 822–823
  dispersion and stability, 821–822
  historical notes, 818
  leapfrog scheme in time, 821
  method, 313–314, 674, 676, 680, 808, 809
  nanoplasmonic laser, 618–620
  numerical implementation, 683–684
  single CNT, 717

Finite element method (FEM), 765–766
convergence analysis, 843–844
historical notes, 833
in one dimensional, 834–839
patchwise tangential continuous and polynomial spaces, 833–834
in two and three dimensional, 839–843
Finite volume method (FVM), 737
mesh generation
boundary conforming Delaunay property, 761–762
open problems, 762–763
tensor product approaches, 762
in one dimensional, 845–847
time treatment in, 844–845
First-principles model, 127
FK effect, *see* Franz–Keldysh effect
Flared unstable cavity lasers, *see* Tapered lasers
Flicker noise, 292
Fluorine-doped tin oxide (FTO), 501
FMM, *see* Fourier modal method
FO, *see* Fourier optics
Fourier modal method (FMM), 809
convergence, 832–833
historical notes, 823–824
layered media stack, 826–827
scattering-matrix approach, 825–826
segmentwise waveguide treatment, 824
transfer-matrix approach, 824–826
in two dimensional, 828–832
Fourier optics (FO), 93
Fourier transform, 36, 240, 315, 348, 676, 680, 710, 712, 809
Fox–Li approach, 42
FP laser, *see* Fabry–Pérot laser
Franz–Keldysh (FK) effect, 363, 365
Free-carrier absorption (FCA), 6, 8, 33, 84, 365
Free energy, 517, 751, 758–759, 763
FTO, *see* Fluorine-doped tin oxide
Full cavity coherent coupling (FCCC) model, 65
Full cavity incoherent coupling (FCIC) model, 65
Full discretized van Roosbroeck system, 742–743
Full width at half maximum (FWHM), 133, 250
Fully-depleted silicon-on-insulator (FD-SOI) device technology, 800, 802
Fully microscopic quantum-kinetic models, 459–461
FVM, *see* Finite volume method
FWHM, *see* Full width at half maximum

G

Gain compression, 4, 169–170
Gain dispersion, 187, 207–208
and group velocity dispersion parameters, 214
model of, 158–161
self-phase modulation in, 215
Gain-guided (GG) tapered laser
with beam spoilers, 68–72
feature, 61–62
without beam spoilers, 73–75
Gain-to-absorber saturation energy ratio, 191
Galerkin-type methods, 836
GaN/AlGaN PIN photodiode, 327–331
Gate contact, 747, 762, 800
Gate operations, 663
Gauss–Fermi integral, 766
Gaussian beam, 61, 68, 77
Gaussian density of states (GDOS), 515
Gaussian function, 68, 723
Gaussian noise, 708
Gauss's law, 266, 367, 820, 846
GDOS, *see* Gaussian density of states
Generalized finite difference method, 818
Generation-recombination (GR) noise, 292
GFs, *see* Green's functions
GG tapered laser, *see* Gain-guided tapered laser
Global detailed balance theories, for ideal systems, 454–457
Gm, *see* Modal gain
Governing equations, 17–19
Green's functions (GFs), 643
dyadic, 591, 592, 665
nonequilibrium, 238–240
Green's identity, 53
GR noise, *see* Generation-recombination noise
Group velocity dispersion (GVD), 187, 214
Gummel map, 753
Gummel's method, nonlinear solvers, 760–761

H

Half-cavity (HC) model, 65
Hanbury, Brown, and Twiss (HBT)
experiment, 585–586
measurement, 633, 634
Harmonic oscillator, 662
potentials, 699–700
Hartree–Fock factorization, 637–638
Haus theory, 190, 196–197
HBT, *see* Hanbury, Brown, and Twiss
HCGs, *see* High-contrast gratings
HC model, *see* Half-cavity model
Heat sources, spatial distributions of, 51–52
HED, *see* Horizontal electric dipole
Heterojunctions, incorporation of, 303
HgCdTe-Based p-n junction photodiode, 323–327
High-contrast gratings (HCGs), 575–576
grating mirrors, cavities with, 578–580
manufacturing sensitivity of, 580–582
subwavelength reflectivity, 576–579

High-power laser diodes, 81
beam quality, role of, 82–83
external cavity simulation tool, 93
lateral cavity design, 85–86
power scaling, 82–83
simulation tool coupling with coherent ray tracing tools, 94–95
Speclase, 92–93
tapered laser diode, 86–92
vertical cavity design, 83–85
High-power lasers
high-power broad-area semiconductor laser, 16
nonlinear modes and filamentation
longitudinal-lateral approximate projected equations, 42–43
nonlinearities due to virtual transitions, 44–45
nonlinearity induced by lateral spatial hole burning, 45–47
optical field
balance of radiative energy, 39–41
beam propagation method, 41–42
cavity modes, 41–42
dispersion, 38
final field equation, 39–41
nonlinear susceptibility, 38–39
roundtrip operator, 41–42
spontaneous emission, 39
three-dimensional equations, 35–37
optical power
basic relations, 24–26
LSH and impact of series resistance and internal loss, 28–31
nonthermal power saturation effects, 31–35
Rigrod model, 26–28
treat-power-as-a-parameter method, 28
phenomenological model
governing equations, 17–19
parameters determination, 19–22
simulated versus. measured CW characteristics, 23–24
thermodynamic-based energy-transport model
basic equations, 48–51
energy conservation, 53–55
spatial distributions of heat sources, 51–52
Hole spin-flip relaxation, 701
Homogeneously broadened (HB) two-level quantum system
analytically computed gain coefficient, 720
density-matrix method, 680
gain saturation dynamics of, 681–682
linear small-signal gain in, 681
theoretical gain coefficient of, 724
Hong–Ou–Mandel (HOM) experiment, 586, 587
Horizontal electric dipole (HED), 600
Horn ridge waveguide (HRW) DFB laser, 146–147
Hot-charged exciton states, 701, 703–707
Huygen's integral, 42
Hybrid ML, 184, 187, 194
Hybrid models, 457–459

I

IAM, *see* Incident angle modifier
ICRs, *see* Intracavity reflectors
Ideal fill factor, 407
Ideal *I-V* model, 396–398
ideal single-junction solar cells, 399
quantum efficiency, 398–399
IG laser, *see* Index-guided laser
IIR, *see* Infinite impulse response
InAs/GaInSb IR photodiode, 350–351
Incident angle modifier (IAM), 392–393
Incident photocurrent efficiency (IPCE), 511
Index-guided (IG) laser, 60–62
narrow tapered section, 75–76
Indirect gap semiconductors, 285–286
Indium tin oxide (ITO), 499–501
Infinite impulse response (IIR), 208
Infrared photodiodes
HgCdTe-Based p-n junction photodiode, 323–327
$InP/In_{0.53}Ga_{0.47}As/InP$ PIN photodiode, 322–323
Injection barrier, 236
Injector level, 236, 239
Inorganic semiconductors, 529, 748, 749
$InP/In_{0.53}Ga_{0.47}As/InP$ PIN photodiode, 322–323, 325
Instantaneous optical modes, 174–176
Integrated circuits (ICs), 774
Interface recombination velocities, 429–430
Internal loss, 28–31
Internal quantum efficiency (IQE), 398, 434, *see also* Quantum efficiency
Intersymbol interference (ISI), 111
Intracavity reflectors (ICRs), 216
Intrinsic carrier density/concentration, 424–429, 748
Inverse Fermi integrals, 29, 49
Ionization thresholds, 295
IPCE, *see* Incident photocurrent efficiency
IQE, *see* Internal quantum efficiency
ISI, *see* Intersymbol interference
ITO, *see* Indium tin oxide

J

J-F curve, 250, 252
Joule heat, 50, 52, 84, 105, 800
Junction capacitance effect, 289, 293, 296
J-V curve calculation, 418, 419, 431–432

K

Kerr lens ML, 184
Kinetic Monte Carlo (kMC) method, 530–531
  algorithm, 520–525
  1D disordered system by, 525–527
  infrequent-event system, 519
  transport models, hierarchy of, 518
k·p method, 240–241
Kramers–Kronig relationship, 127, 363–364

L

Lambert's law, 285
Langevin formalism, 707
LA phonons, *see* Longitudinal acoustic phonons
Large optical cavity (LOC) tapered laser, 82, 84
  ELoD2 structure, 85, 95, 96
Large-signal lumped iterative model, 203
Large-signal time-domain approach, 193–206
Laser diode
  efficiency analysis, 8–9
  high-power, *see* High-power laser diodes
  laser cavity designs, 12
  optical gain and optical loss, 3–6
  photonic crystal, *see* Photonic crystal laser diodes
  rate equation analysis, 9–12
  semiconductor laser diodes, 3
  temperature effects, 7–8
  threshold current and slope efficiency, 6–7
Laser geometry, 215–217
Lasing filaments, 42
Lasing frequency, 110
Lateral modes, intensity distributions of, 275–276
LCOE, *see* Levelized cost of electricity
LDOS, *see* Local density of states
Leakage currents, 33–34
LED, *see* Light-emitting diode
Levelized cost of electricity (LCOE), 410
Light-emitting diode (LED), 541–542
Light intensity behavior, of solar cells, 406–408
Light states, 662
$LiNbO_3$, 372
Linear electro-optic effect, *see* Pockel's effect
Linear finite elements, 837, 840–841
Liouville equation, 669–670
Local charge neutrality, 736–737
Local density of states (LDOS), 462–463, 488, 592–593
LOC tapered laser, *see* Large optical cavity tapered laser
Longitudinal acoustic (LA) phonons, 242, 642
Longitudinal-lateral approximate projected equations, 42–43
Longitudinal optical (LO) phonon, 236, 642
Longitudinal spatial horn burning (LSH/LSHB) effect, 28–31, 121
Long-wavelength/mid-wavelength (LW/MW) two-color HgCdTe, 331–334
LO phonon, *see* Longitudinal optical phonon
Lorentzian approximation, 159–160
Lorentzian form, 218
LSH/LSHB effect, *see* Longitudinal spatial horn burning effect
Luminescent coupling, 434–437
Lumped series resistance, 400, 407
LW/MW two-color HgCdTe, *see* Long-wavelength/mid-wavelength two-color HgCdTe
Lyapunov function, 751, 752

M

Mach–Zehnder modulator, 372–374
Macroscopic polarization vector, 671, 672
Magneto-chiral effects, 721–724
$MAPbX_3$, *see* Methylammonium lead halide perovskites
Masetti model, 425
Master pseudospin equation, 669–671
MA system, *see* Mode approximation system
Material absorption modeling
  material parameters, 343–344
  numerical examples, 344–347
Material gain dispersion, 158–161
Maxwell–Bloch approach, 610–613
Maxwell–Bloch equations, 166, 635, 707
Maxwell-pseudospin method
  coherent equations in 2D, 673–674
  coupling to pseudospin equations, 671–673
  dissipative dynamics and master pseudospin equation, 669–671
  numerical methodology, 674
  pseudospin equation, 667–669
  vector Maxwell equations, 671–673
Maxwell's equations, 264
  curl equations, 314
  edge functions, 842, 843
  homogeneous, 35
  perfectly matched layers, 814–817
  resonance and mode propagation problem, 811
  scattering problem, 810–811
  spatial discretization of, 821
  time-dependent, 808, 809
  transformation rules, 812–814
  two dimensional geometries and fields, 811–812
MBE, *see* Molecular beam epitaxy
MCPM, *see* Multiple colliding pulse mode-locking
MD, *see* Modal discrimination
MESFET, *see* Metal-semiconductor field-effect transistor
Mesh discretization, 419–420
Metal-organic chemical vapor deposition (MOCVD), 235

Metal-organic vapor phase epitaxy (MOVPE), 15
Metal-oxide-semiconductor field-effect transistor (MOSFET), 774, 803
Metal-semiconductor field-effect transistor (MESFET), 792, 793
Metal-semiconductor-metal photodetectors (MSM photodetectors), 298–299
Methylammonium lead halide perovskites ($MAPbX_3$), 503
Microcavity exciton–polaritons, 665
Microcavity pillar geometry, 595–596
Microcrystalline silicon (μc-Si), 498–499
Micropillar lasers, 629
Mid-infrared quantum cascade lasers (mid-IR QCLs), 235
  electron transport in, 237
mid-IR QCLs, *see* Mid-infrared quantum cascade lasers
Minority carrier lifetime, 315–317
MJSCs, *see* Multijunction solar cells
ML semiconductor lasers, *see* Mode-locked semiconductor lasers
MMM, *see* Mode-matching method
MOCVD, *see* Metal-organic chemical vapor deposition
Modal discrimination (MD), 84, 100
Modal gain (Gm), 250–251
Mode approximation (MA) system, 178–179
Mode intensity, 176, 269
Model-dependent generation-recombination rate, 748
Modeling cavity-emitter systems, 666, 707–712
Modeling coherent spin, in QDs
  hot-charged exciton states, 701, 703–707
  single-charged particles in waveguide, 699–701
Mode-locked (ML) semiconductor lasers
  absorber relaxation time, 212–213
  coherent population effects, 225–226
  delay-differential equations model, 193–206
  features, 185–187
  frequency and time-frequency treatment, 192
  gain dispersion
    and absorber compression coefficients, 215
    and group velocity dispersion parameters, 214
  general principles of, 183–185
  large-signal time-domain approach, 193–206
  laser geometry and saturable absorber location, 215–217
  microscopic input in, 217–222
  operating point, 210–212
  pumping current, 212
  self-consistent pulse profile, 187–192
  self-phase modulation, 215
  in single-section lasers, 226–227
  small-signal time-domain models, 187–192
  traveling-wave models, 206–210, 223–225
Mode-matching method (MMM), 115
Mode partition noise (MPN), 105, 111
Molecular beam epitaxy (MBE), 15, 235
  grown QD, 700
Monolithic diode lasers, 214
Monte Carlo algorithm, 523–525
Monte Carlo device simulations
  of different materials, 796–799
  electron-polar optical phonon scattering, 787
  electron transport, scattering tables for, 784
  energy and wavevector histograms, 791
  ensemble Monte Carlo techniques, 779
  final angle selection, 787–788
  flow chart description of, 786
  initial carrier distribution, 785
  modeling and simulation, 776–779
  particle-based device simulation, *see* Particle-based device simulation
  pseudo-random number generator, 780
  rejection technique and drift velocity, 789
  scattering rate expressions, 783
  self-scattering mechanism, 781–782
  steady-state drift velocity versus. electric field, 790
Moore's law, 774–775
More than Moore (MtM) devices, 775–776
MOSFET, *see* Metal-oxide-semiconductor field-effect transistor
MOVPE, *see* Metal-organic vapor phase epitaxy
MPN, *see* Mode partition noise
MQW, *see* Multiple quantum well
MSLs, *see* Multisection semiconductor lasers
MSM photodetectors, *see* Metal-semiconductor-metal photodetectors
MtM devices, *see* More than Moore devices
Muller's algorithm, 121
Multijunction solar cells (MJSCs)
  complex refractive index, 421–422
  drift-diffusion calculation, *see* Drift-diffusion calculation
  J-V curve calculation, 431–432
  luminescent coupling, 434–437
  mesh discretization, 419–420
  overview, 415–417
  QE calculation, 433–434
  quantum yield, 422–423
  simulation workflow, 418–419
  transfer matrix method, 420–421
Multiple colliding pulse mode-locking (MCPM), 216
Multiple quantum well (MQW), 235, 338, 366
Multisection semiconductor lasers (MSLs), 154, 163, 167–168
  and coupled laser systems, 170–172
  nontrivial MSL device, 172–174

N

Nanoplasmonic laser
  block diagram of, 610
  Drude–Lorentz model, 610–611
  evolutionary algorithm, 616–618
  finite-difference time-domain method, 618–620
  lasing dynamics, 623
  loss compensation, 620–622
  multilevel Maxwell–Bloch theory, 611–613
  stopped-light lasing, 623–625
  transfer matrix method, 613–616
Nanostructure solar cells (NSSC)
  charge separation, 449–452
  electron-hole pairs, photogeneration of, 445–449
  fully microscopic quantum-kinetic models, 459–461
  global detailed balance theories, for ideal systems, 454–457
  hybrid models, 457–459
  in photovoltaics, 441–444
  recombination, 452–454
  single quantum well, 461–465
Nanowire solar cells
  current density, 476
  efficiency evaluation, 479–481
  electronic model, 482–483
  material parameters, 493
  modal analysis, 485–487
  nonequilibrium, 477–479
  optical models, 481–482
  with p-n junction, 475–476
  Purcell effect, 488
  radial versus. axial junction arrangement, 489–491
  Shockley–Queisser efficiency, 487
  spectral absorptivity, 484–485
  surface recombination, 491–492
  thermal equilibrium, 477
Narrow broad-area (NBA) laser diodes, 86
Narrow index-guided tapered laser, 75–76
National Renewable Energy Laboratory (NREL), 498
Natural optical activity, *see* Optical activity
NBA laser diodes, *see* Narrow broad-area laser diodes
Nearest-element-cell (NEC) scheme, 792, 796
Nearest-grid-point (NGP) scheme, 792, 796
NEC scheme, *see* Nearest-element-cell scheme
NEGF, *see* Nonequilibrium Green's functions
NEP, *see* Noise equivalent power
Neumann boundary conditions, 528, 747, 748, 801
New's model, 195–197
Newton's method, nonlinear solvers, 760, 763
NGP scheme, *see* Nearest-grid-point scheme
N-level quantum system, 667, 670
Noise equivalent power (NEP), 289–290
Non-concentrated solar cells, 388–389
  irradiation, 389–390
  temperature, 390–391
Nonequilibrium Green's functions (NEGF), 238–240, 483
Nonequilibrium phonons
  electronic bandstructure solver, 240–241
  ensemble Monte Carlo with, 240
  transport kernel with, 241–243
Non-Gaussian beams, 61
Nonideal *I-V* model
  parallel resistance, 399–400
  recombination, 400–402
Nonlinear gain saturation, 165–167
Nonlinearity
  due to virtual transitions, 44–45
  gain compression, 31–32
  induced by lateral spatial hole burning, 45–47
  susceptibility, 38–39
Nonlinear modes
  due to virtual transitions, 44–45
  induced by lateral spatial hole burning, 45–47
  longitudinal-lateral approximate projected equations, 42–43
Nonlinear Poisson problem, 738–739
Nonlinear solvers
  Gummel's method, 760–761
  Newton's method, 760
Nonthermal power saturation effects, 31–35
Normalized saturated field intensity, 681
NREL, *see* National Renewable Energy Laboratory
NSSC, *see* Nanostructure solar cells

O

OBPF, *see* Optical band-pass filter
ODE, *see* Ordinary differential equation
Ohmic contacts, 28, 289, 350, 395, 418, 432, 735–737, 747, 752, 754, 794–795
One-dimensional (1D) laser simulator, 63
One-dimensional scalar approach, 266–270
Optical activity
  bandgap edge, 715–716
  chirality, 714
  cholesteric liquid crystals and sculptured thin films, 721
  circularly polarized ultrashort pulse, 715
  coherent vector Maxwell-pseudospin formalism, 717
  electric-field components, time evolution, 717–718
  Fourier spectra, 717–718
  linearly polarized pulse, 719
  polarization-resolved transmission spectra, 720
  rotation angle, 714
  wavevector, 720

Optical band-pass filter (OBPF), 134
Optical beam absorption method, 310–311
Optical feedback, 4, 12, 82, 98, 102, 153, 172, 205, 261
Optical field
  balance of radiative energy, 39–41
  beam propagation method, 41–42
  cavity modes, 41–42
  dispersion, 38
  final field equation, 39–41
  nonlinear susceptibility, 38–39
  roundtrip operator, 41–42
  spontaneous emission, 39
  three-dimensional equations, 35–37
Optical gain, 3–6
Optical generation models, 309–310
  beam propagation method, 314–315
  FDTD method, 313–314
  optical beam absorption method, 310–311
  raytracing, 311
  transfer matrix method, 312–313
Optical loss, 3–6
Optical microconcentrators, of solar cell, 480–481
Optical modulators, 363–364
  electro-optic effect, 371–374
  plasma-effect, 374
  QCSE-based, *see* QCSE-based modulators
  resonant, 377–378
  Si electroabsorption versus electrorefractive effect, 374–376
Optical phenomena, 262
Optical power
  basic relations, 24–26
  LSH and impact of series resistance and internal loss, 28–31
  nonthermal power saturation effects, 31–35
  Rigrod model, 26–28
  treat-power-as-a-parameter method, 28
Optical Rabi oscillations, 665, 691
Optical source, brightness of, 59
Optical waveguides and semiconductor microcavities
  active nonlinear optical waveguide/cavity properties, 678–680
  passive cavity properties, 676–678
  resonant nonlinearities and gain saturation, 681–683
  2D nonlinear optical waveguides, 683–687
  2D semiconductor microcavities, 687–691
  two-level quantum system, 675–676
OpticStudio, 93–95
OPV cells, *see* Organic photovoltaics cells
Ordinary differential equation (ODE), 154
Organic photovoltaics (OPV) cells, 500, 502, 528–529
Organic semiconductors, 748, 766
Oxide aperture diameter, impact of, 277–278

P

Parallel-mirror slab waveguide, 685–686
PARDISO, direct solvers, 760
Partial differential equations (PDEs), 154
Partially gain-coupled antiphase DFB, 122, 123, 133
Partially gain-coupled in-phase DFB, 122, 123
Partially loss-coupled antiphase-grating DFB, 122, 124
Partially loss-coupled in-phase DFB, 122, 124
Particle-based device simulation
  current calculation, 792, 794, 795
  flowchart of, 793
  ohmic contacts, 794–795
  particle-mesh coupling, 791–792, 796, 797
  thermal effects in silicon-on-insulator devices, 799–803
  time-step and mesh size, 795–796
Particle current flow, 48
Particle-mesh (PM) coupling, 791–792, 796, 797
Passive cavity properties, 676–678
Passive ML (PML), 184
PAT, *see* Pulse-area theorem (PAT)
PBCs, *see* Periodic boundary conditions
PBG, *see* Photonic bandgap
PC-EELs, *see* Photonic-crystal edge-emitting lasers
PCS QWIP, *see* Photonic crystal slab quantum well infrared photodetectors
PDA, *see* Planar dielectric antenna
PDEs, *see* Partial differential equations
PDF, *see* Probability density function
p-doped QDs, 701, 704
PDs, *see* Photodetectors
PECVD, *see* Plasma-enhanced chemical vapor deposition
Peltier heat, 50, 52, 54
Perfectly matched layer (PML), 37, 814–817
Periodic boundary conditions (PBCs), 240, 484, 565
Permittivity spectrum, 370
Perovskites-based solar cells, 500, 503, 529–530
PhC waveguide, *see* Photonic crystal waveguide
Phenomenological model
  governing equations, 17–19
  parameters determination, 19–22
  simulated versus measured CW characteristics, 23–24
Photoconductive detectors, 290–292
Photoconductive gain, 291, 341
Photocurrent, 341–342, 394
Photodetectors (PDs), 283, *see also* Quantum well infrared photodetectors
Photodiode
  p-i-n, 293–294
  p-n, *see* p-n junction photodiodes
  visible, 327

Photogeneration of electron-hole pairs
absorption coefficient, 446–449
photon flux, 445–446
Photoluminescent (PL) refrigeration, 541
Photon, 4
blockade, 665
density, 10
flux, 445–446
Photon recycling, 433–437, 546–548, 553–558
Photonic bandgap (PBG), 563, 564
Photonic-crystal edge-emitting lasers (PC-EELs), 563–565
Photonic crystal laser diodes
periodic structures, 561–563
photonic bandgap, 563, 564
photonic-crystal edge-emitting lasers, 563–565
vertical-cavity surface-emitting lasers, *see* Vertical-cavity surface-emitting lasers
Photonic crystal slab quantum well infrared photodetectors (PCS QWIP), 355–356
Photonic crystal (PhC) waveguide, 598–599
Photonic nanowire geometry, 597–598
Photonics, numerical methods of, 808–809
Photon-number (Fock) states, 662
Photon recycling, 434–437, 546, 554–557
Phototransistors, 290, 300
Photovoltaics, nanostructure solar cells in, 441–444
Photovoltaic (PV) technologies, 497
Physical phenomena, 262–263
Picosecond pulse propagation, 664
p-i-n-diode, 744
p-i-n photodiode, 293–294
Planar dielectric antenna (PDA), 600
Planar GaAs solar cell, 349–350
Planar semiconductor microcavity, 683–684
Planck constant, 17, 156, 310, 434, 445, 504, 545, 749
Plane-wave admittance method (PWAM), 565–570
Plane-wave reflection transfer (PWRT), 569–570
Plasma-effect modulators, 363, 374
Plasma-enhanced chemical vapor deposition (PECVD), 499
Plasmonic-enhanced quantum well infrared photodetectors, 353–355
PL refrigeration, *see* Photoluminescent refrigeration
PM coupling, *see* Particle-mesh coupling
PML, *see* Passive ML; Perfectly matched layer
P-N junction photodiodes, 293
absorption coefficient model, 318–320
basic equations, 309
beam propagation method, 314–315
FDTD method, 313–314
infrared photodiodes, 322–327
minority carrier lifetime, 315–317
optical beam absorption method, 310–311
overwiew, 307–308
physical mechanism, 320–322
raytracing, 311
transfer matrix method, 312–313
trap concentration versus energy, 317–318
two-color heterojunction photodiodes, 331–334
ultraviolet photodiodes, 327–331
visible photodiodes, 327
Pockel's effect, 363, 371
Poisson-drift-diffusion model, 367–368
Poisson equation, 48, 53
Poissonian statistics, 662
Polarization-decay, 669
Polarization density, 35, 38
Population relaxation time, 669
Power-current (P-I) characteristics, 19, 20, 23, 30, 66–67
Power saturation, 23
laser parameters for, 30
Power scaling, 82–83
Probability density function (PDF), 521, 522
Pseudospin equation, 667–669
Pseudospin vector, 669
Pulse-area theorem (PAT), 664
Pulse center frequency, 704
Pumping current, 212
Purcell effect, 590, 594, 595, 602
nanowire solar cells, 488
Purely gain-coupled DFB, 122, 124, 125, 128, 133
Purely refractive index-coupled uniform-grating DFB laser, 122, 123, 128–131
PV technologies, *see* Photovoltaic technologies
PWAM, *see* Plane-wave admittance method
PWRT, *see* Plane-wave reflection transfer

Q

QCLs, *see* Quantum cascade lasers
QCSE, *see* Quantum-confined Stark effect
QCSE-based modulators
absorption spectrum, 367
Franz–Keldysh effect, 365
numerical results, 370–371
permittivity spectrum, 370
Poisson-drift-diffusion model, 367–368
quantum-confined stark effect, 365–367
Schrödinger equation, for QW states, 368–369
structure, 364–365
QDIP, *see* Quantum dot infrared photodetector
QD nanolasers, *see* Quantum-dot nanolasers
QE, *see* Quantum efficiency
QED, *see* Quantum electrodynamics
QGIP, *see* Quantum grid infrared photodetector
Q-switching instability, 198, 203, 206, 211
Quantum cascade lasers (QCLs)
lasing in, 236–237
models, 237–238

Quantum cascade lasers (QCLs) (*Continued*)
overview, 235–236
recent developments in, 237
thermal transport in, 238
Quantum coherent control, 664
Quantum-confined Stark effect (QCSE), 220–221
Quantum defect heat, 52
Quantum dot infrared photodetector (QDIP), 300–301
Quantum dot (QD) lasers, carrier rate equations, 167–169
Quantum-dot (QD) nanolasers
carrier dynamics, *see* Carrier dynamics, in quantum-dot systems
inhomogeneous broadening in, 651
input-output curves and autocorrelation function, 646–647
interemitter coupling effects and superradiance in, 647–651
laser rate equations and β factor, 628–631
lasing regime and threshold, 635
nonresonant cavity feeding, 652–653
photonic and electronic density of states, 632
Purcell enhancement, 631
semiconductor laser theory, *see* Quantum-optical semiconductor laser theory
spontaneous emission, 627–628
streak-camera measurements, 633
Quantum dots (QDs)
antibunching, 662
embedding, 665
modeling coherent spin in, 699–707
spontaneous emission
Box–Müller method, 708
intracavity electric-field build-up, 708–709
output electric-field trace, 710–712
refractive-index profile, 707–708
semiconductor microcavity, electric field, 708–710
slowly-varying envelope approximations, 710
steady-state population inversion, 709, 711
Quantum efficiency (QE), 288, 322, 398–399
calculation of, 433–434
Quantum electrodynamics (QED), 628
Quantum grid infrared photodetector (QGIP), 353
Quantum nanostructures, 286–287
incorporation of, 303
Quantum-optical semiconductor laser theory
continuous-wave-excited lasers, 638–639
correlation-expansion technique, 635
dynamical laser equations, 637–638
semiclassical and fully quantum-mechanical approach, 635–637
Quantum photonic structures
chiral carbon nanotubes, optical rotation in, 712–724
Maxwell-pseudospin method, 667–674
modeling cavity-emitter systems, 707–712
modeling coherent spin in QDs, 699–707
optical waveguides and semiconductor microcavities, 674–691
quantum wells and QDs, 691–699
Quantum regression theorem, 602, 634, 666
Quantum stochastic formalism, modeling cavity-emitter systems
Langevin formalism, 707
QD spontaneous emission, 707–712
quantum fluctuations, 707
quantum noise, 707
Quantum techniques, 239–240
Quantum well infrared photodetectors (QWIPs), 300–301
absorption coefficient, 342
dark current, 338–341
detector QE modeling, *see* Detector QE modeling
material parameters, 343–344
numerical examples, 344–347
overwiew, 337–338
photoconductive gain, 341
photocurrent, 341–342
resonator, *see* Resonator-quantum well infrared photodetector
Quantum wells (QWs), 286–287
energy levels, 66
Quantum wells (QWs) and QDs
circularly polarized optical-pulse interactions
degenerate four-level systems, 695–699
two-level quantum systems, 692–695
optical orientation, 691
Rabi flop, 691
Quantum yield, 422–423
Quarter-wavelength phase-shifted (QWPS) DFB, 122, 123, 125, 128–133
Quasi-3D (three-dimensional) model, 62–65
Quasi-Fermi level separation, 417, 478
Quasi-Fermi potentials, 735–736, 740
Quasi-monochromatic electromagnetic field, 109–110
QWIPs, *see* Quantum well infrared photodetectors
QWPS DFB, *see* Quarter-wavelength phase-shifted DFB
QWs, *see* Quantum wells

R

Radiative energy, balance of, 39–41
Radiative recombination, 452–453, 545
Radio frequency (RF) linewidth, 187
Raman effect, 44
Random-phase approximation (RPA), 643
Rate equation analysis, 9–12
Raytracing method, 310, 311
RCE structure, *see* Resonant-cavity-enhanced structure
RCWA, *see* Rigorous coupled wave analysis

Recombination phenomena, 263
Refractive index, 71, 169–170
  nonlinearities, 184
Refractive-index profile, 707–708
Representative elementary volumes, 753
Resonant-cavity-enhanced (RCE) structure, 302–303
Resonant modulators, 377–378
Resonant nonlinearities, 681–683
Resonator-quantum well infrared photodetector (R-QWIP)
  design, 356–358
  realization, 358–361
Responsivity, 288
RF linewidth, *see* Radio frequency linewidth
Ridge waveguide (RW)
  lasers, 59, 60, 68–75
  and tapered laser diode, 86–92
Rigorous coupled wave analysis (RCWA), 576, 823, 828
Rigrod model, 26–28
Robin boundary conditions, 747
Rotating-wave approximation (RWA), 664
Roundtrip operator, 41–42
RPA, *see* Random-phase approximation
R-QWIP, *see* Resonator-quantum well infrared photodetector
RW, *see* Ridge waveguide
RWA, *see* Rotating-wave approximation (RWA)

S

SA, *see* Saturable absorber
SAC lens, *see* Slow-axis collimation lens
SAGCM structure, *see* Separate absorption, grading, charge sheet, and multiplication structure
SAGM, *see* Separate absorption grading and multiplication
SAM, *see* Separate absorption and multiplication
SA Speclase, *see* Stand-alone Speclase
Saturable absorber (SA)
  implementation, 187
  location, 215–217
  mechanism, 225–226
  reverse bias, 186
  self-starting mechanism, 184
SB, *see* Schottky barrier
Scalar Helmholtz equation, 841–842
Scattering-matrix approach, 825–826
Scharfetter–Gummel fluxes, 756–758
Schottky barrier (SB), 298
Schrödinger equation, for QW states, 368–369
SCOWL diodes, *see* Slab-coupled optical waveguide laser diodes
SCP, *see* Self-consistent profile
Self-consistent profile (SCP), 187–192
Self-heating, 8, 799, 801–803
  due to free-carrier absorption/optical pumping, 101
  due to Joule heating, 100
Self-induced transparency (SIT), 664
  and polarized SIT soliton formation, 693–695
  soliton, 664, 685, 687
    polarized, formation, 693–695
Self-phase modulation, 184, 187, 215, 218
Semiclassical techniques, 238–239
Semiconductor diode lasers, 184
Semiconductor Industry Association (SIA), 774
Semiconductor laser
  brightness, 59
  devices, 155
  diodes, 3
Semiconductor luminescent refrigerator, 541, 542
Semiconductor optical amplifier (SOA), 60
Semiconductors, 284–285
Semiconductor saturable absorber mirror (SESAM) chip, 202
Separate absorption and multiplication (SAM), 296
Separate absorption grading and multiplication (SAGM), 296
Separate absorption, grading, charge sheet, and multiplication (SAGCM) structure, 296
Series resistance, 28–31
SESAM chip, *see* Semiconductor saturable absorber mirror chip
SHB, *see* Spatial hole burning
Shockley-Queisser model
  of nanowire array, 476, 481, 487
Shockley–Read–Hall (SRH), 401
  process, 128
  recombination, 6, 545
Shooting method, 369
Short-scale spatial hole burning, 209
SIA, *see* Semiconductor Industry Association
Side-mode-suppression-ratio (SMSR), 119
Si electroabsorption versus electrorefractive effect, 374–376
Silicon-based cells, 498–500
Silicon-on-insulator (SOI)
  technology, 775, 799–803
  waveguide, 120
Single-charged QD embedded in waveguide, 699–701
Single longitudinal mode (SLM) laser, 111–113
  Bragg waveguide SLM laser, 141–144
  characteristics, examples, 128–133
  classification, 113
  coupled-cavity lasers, *see* Coupled-cavity SLM lasers
  distributed Bragg reflector laser, 118–120
  distributed feedback lasers, 120–125
  double-trench resonant tunneling SLM laser, 144–146
  governing equations
    for design, modeling, and simulation, 125–128
    and solution technique, 139–140

Single-charged QD embedded in waveguide (*Continued*)
grating analysis, 113–118
horn ridge waveguide DFB laser, 146–147
problems in, 141
Single-mode photon density, 10
Single-photon emission events, 662, 587, 662
Single-photon source (SPS)
cavity quantum electrodynamics, 594–596
classical dipole, 590–592
coupling strength, 603–604
definition of, 585
efficiency, 587–588
emitter-cavity master equation, 602–603
indistinguishability, 586–587, 601
photon extraction strategies, 590
planar dielectric antenna, 600
pure single-photon emission, 585–586
quantum emitter master equation, 601–602
simulation techniques, 593–594
spectral wandering, 603
spontaneous emission, 592–593
spontaneous parametric down-conversion process, 588–590
waveguide quantum electrodynamics approach, 596–599
Single quantum well (SQW), 461
characteristics, 464–465
dynamics, 463–464
states, 462–463
Si P-Njunction photodiode, 327, 328
SIT, *see* Self-induced transparency
SL, *see* Superlattice
Slab-coupled optical waveguide laser (SCOWL) diodes, 85
SLM laser, *see* Single longitudinal mode laser
Slope efficiency, 6–9, 237
Slotboom variables, 752
Slow-axis collimation (SAC) lens, 96
Slowly-varying envelope (SVEA) approximations, 664, 710
Small-signal time-domain models, 187–192
Small wavelength chirp, 119
SMSR, *see* Side-mode-suppression-ratio
SOA, *see* Semiconductor optical amplifier
SOI, *see* Silicon-on-insulator
Solar cells optics, photocurrent of, 394
Solar spectrum, 387–388
Solid-state optical refrigeration, 541
Space-time folding technique, 223–225
Spatial hole burning (SHB), 62, 71, 163–165
nonlinearity induced by, 45–47
Spatially inhomogeneous carrier density, 163–165
SPDC process, *see* Spontaneous parametric down-conversion process
Specific rotatory power, 712
Speclase, 92–95
Spectrally selective laser cavity, 216
Spin-B interaction energy, 722
Spin decoherence mechanisms, 703
Spin-filtering effect, 701, 706–707
Spin-polarized drift diffusion models, 766
Spontaneous emission, 39, 128
Spontaneous parametric down-conversion (SPDC) process, 588–590
SPP modes, *see* Surface plasmon polariton modes
SPS, *see* Single-photon source
SQW, *see* Single quantum well
SRH, *see* Shockley–Read–Hall
Stand-alone (SA) Speclase, 93
Standing wave-induced gratings, 209
Stark effects, 44
Stationary 1D semiconductor equations
discrete nonlinear Poisson problem, 738–739
discretize Poisson problem, 737–738
1D van Roosbroeck system, 735–736
full discretized van Roosbroeck system, 742–743
linear electrostatic potential, 743–745
Scharfetter–Gummel scheme, 739–742
thermodynamic equilibrium and local charge neutrality, 736–737
Stationary ML equation, 189
Statistical distribution function, 735
Steady state, 176–178
above threshold, 25–26
around threshold, 25
conditions, 10–11
Steady-state population inversion, 709, 711
Stranski–Krastanov mode, 651
Superlattice (SL), 244
Surface plasmon polariton (SPP) modes, 609, 616
SVEA, *see* Slowly-varying envelope approximations
Synchronous ML, 184

T

Tapered laser diode
DBR-tapered laser, *see* Distributed Bragg reflector-tapered laser
lateral beam quality of, 86–92
Tapered lasers
fast axis, 60
GG tapered laser, *see* Gain-guided tapered laser
simulation examples
comparison between devices, 76–78
device geometries and simulation parameters, 67–68
GG tapered laser with beam spoilers, 68–72
GG tapered laser without beam spoilers, 73–75
narrow index-guided tapered laser, 75–76

simulation model
calibration and comparison with experiments, 65–67
physical effects, 63
quasi-3D model, 62–65
symmetry, 65
slow axis, 60
Taper losses, 69, 75, 77
TAT, *see* Thermally assisted tunneling
TCAD, *see* Technology computer-aided design
TCO layer, *see* Transparent conductive oxide layer
Technology computer-aided design (TCAD), 777
TE cooling, *see* Thermoelectric cooling
TE field, *see* Transverse electric field
TEM field, *see* Transverse electromagnetic field
Temperature behavior, of solar cells, 408
Temperature effects, 7–8
Temperature range, 20, 245
Terahertz quantum cascade lasers (THz QCLs), 237, 239
Thermal black-body radiation source, 662
Thermal detuning, 161–163
Thermal lensing, 42, 47, 62
Thermally assisted tunneling (TAT), 338, 339, 450
Thermal phenomena, 263
Thermal transport, 243–244, 246–247
device-level electrothermal simulation, 247–249
in quantum cascade lasers, 238
quantum cascade lasers active core, 245–246
superlattice, 244
twofold influence, of effective interface roughness, 244–245
Thermodynamic-based energy-transport model
basic equations, 48–51
energy conservation, 53–55
spatial distributions of heat sources, 51–52
Thermodynamic equilibrium
flux expressions with, 758
and local charge neutrality, 736–737
Thermodynamic model, 309
Thermoelectric (TE) cooling, 542, 543
Thin-film technologies
cadmium telluride-based cells, 499–501
coherent and incoherent calculation, 509
copper indium gallium selenide-based cells, 500, 501
definition of, 497
drift-diffusion model, *see* Drift-diffusion model
dye-sensitized solar cell, 500–502, 527–528
incident photocurrent efficiency, 511
in-plane and orthogonal components, 507
kinetic Monte Carlo method, *see* Kinetic Monte Carlo method
morphology and grain effects, 530–531
optical absorption, 498, 503–504
organic heterojunction solar cell, 510
organic photovoltaics cells, 502, 528–529
perovskites-based cells, 503, 529–530
propagation constant, 506
silicon-based cells, 498–499
transfer matrix method, 505, 508
Thomson heat, 50, 52
Three-dimensional equations, 35–37
Three-dimensional scalar approach, 273–277
Threshold current, 6–7
Threshold phonons, 250–251
THz QCLs, *see* Terahertz quantum cascade lasers
Time-bandwidth product, 184–185
Time-dependent effects, 9
Time-dependent Rabi frequencies, 703, 717
Time discretization, 763
Time-domain lumped models, 187
Time-harmonic versus time-dependent modeling, 808
Titanium dioxide (TiO_2), 501–503, 527, 528
TM field, *see* Transverse magnetic field
TMM, *see* Transfer matrix method
TPP method, *see* Treat-power-as-a-parameter method
Transfer matrix method (TMM), 312–313, 420–421, 482, 486, 676–677, 824–827
nanoplasmonic laser, 613–616
Transistor scaling, 774–775
Transmission spectrum, 676, 678
Transparent conductive oxide (TCO) layer, 499, 500
Transport kernel, 240, 791, 799
with nonequilibrium phonons, 241–243
Transverse electric (TE) field, 36, 614, 673–674
Transverse electromagnetic (TEM) field, 505
Transverse magnetic (TM) field, 36, 614, 673–674
Traveling-wave model (TWM), 154–155, 200, 374
considerations, 206–210
material gain dispersion, 158–161
modifications of, 167–170
nonlinear gain saturation, 165–167
numerical efficiency of, 223–225
numerical simulations of, 174–179
in solitary laser, 155–158
spatially inhomogeneous carrier density, 163–165
thermal detuning, 161–163
Treat-power-as-a-parameter (TPP) method, 28, 29
Truncated matrix-vector, 829
Tunnel diodes, 430–431
TWM, *see* Traveling-wave model
Two-color heterojunction photodiodes, 331–334
2D Nonlinear optical waveguides
electric-field envelope, 684
FDTD numerical implementation, 683–684
parallel-mirror slab waveguide, 685–686
planar semiconductor microcavity, 683–684
SIT soliton, 685, 687
waveguide width, 685
2D Semiconductor microcavities, 687–691
Two-dimensional scalar approach, 270–273

Twofold influence of effective interface roughness, 244–245
Two-level quantum system, 675–676
Two-photon absorption, 32–33, 44
Two-photon interferometry, 662

U

Ultrafast resonant optical excitation, 717
Ultraviolet photodiodes, 327–331
UMFPACK, direct solvers, 760
Uniaxial perfectly matched layers (UPML), 814
Urbach tails, 514

V

Vacuum quantum fluctuations, 662
Valence band (VB) mixing model, 66
van Roosbroeck system, *see also* Discretizing van Roosbroeck system
  coupling to other model, 766
  equations, 746
  existence and uniqueness results, 752–753
  free energy and dissipation rate, 751
  initial and boundary conditions, 747
  maximum principle, 753
  non-Boltzmann statistics and generalized Einstein relation, 749–751
  recombination processes, 747–748
  thermodynamic equilibrium, 748
Varshni formula, 7
VB mixing model, *see* Valence band mixing model
VCSELs, *see* Vertical-cavity surface-emitting lasers
VECSEL, *see* Vertical external cavity surface-emitting laser
Vector maxwell equations, 671–673
VED, *see* Vertical electric dipole
Vertical-cavity surface-emitting lasers (VCSELs), 12, 563, 564, 627
  fundamental equations, 264–265
  high-contrast gratings, *see* High-contrast gratings
  for light confinement, 570–574
  light confinement in, 263–264
  one-dimensional scalar approach, 266–270
  overwiew, 261–262
  oxide aperture diameter, impact of, 277–278
  physical phenomena, 262–263
  plane-wave admittance method, 565–569
  plane-wave reflection transfer, 569–570
  structure, 265–266
  three-dimensional scalar approach, 273–277
  two-dimensional scalar approach, 270–273
Vertical electric dipole (VED), 600
Vertical external cavity surface-emitting laser (VECSEL), 202
Virtual transitions, nonlinearity due to, 44–45
Visible photodiodes, 327
Voltage-current characteristic, 18, 22, 66
Voronoi diagram, 761

W

WA-FDBPM, *see* Wide-angle finite-difference beam propagation method
Waveguide photodetector (WPD), 302
Waveguide QED approach, *see* Waveguide quantum electrodynamics approach
Waveguide quantum electrodynamics (waveguide QED) approach, 596–599
Waveguide width, 685
Wavelength selectivity, 287, 302
Wavevector, 563–564, 720, 781–782, 789–791
Wetting layer (WL) distribution, 645
Wide-angle finite-difference beam propagation method (WA-FDBPM), 62
Wigner–Weisskopf formula, 705
WL distribution, *see* Wetting layer distribution
WPD, *see* Waveguide photodetector
Wronski–Staebler effect, 499

Y

Yee grid for discretization, 674, 675

Z

Zeeman splitting, 721, 722
Zero group velocity (ZGV) points, 617–618, 621, 623
Zero-reflection, 117