Progress In Electromagnetics Research, PIER 11, 301–344, 1995

TWO-DIMENSIONAL NUMERICAL SIMULATION OF SEMICONDUCTOR LASERS

Z.-M. Li

1. Introduction

Semiconductor laser is an important optoelectronic device as the light source for optical systems today. Many papers have been written on two-dimensional (2D) numerical simulations of semiconductor lasers in the last decade or so (see, for example, [1–3]. However,the work on laser simulation is still primitive in terms of volume and depth compared with its counterparts in simulation of conventional electronic devices. With the popularity of quantum well and strained quantum well lasers,the complexity in their design and analysis has increased significantly and it becomes important to develop or apply accurate numerical models to the design of modern semicondutor lasers.

A semiconductor laser is a diode and, therefore, also an electronic device. In this sense all the previous experience [4,5] generated in numerical simulation of VLSI electronic devices can be borrowed directly to laser simulation. However several important issues must be resolved before the conventional techniques can be applied. These are discussed as follows (see also the schematic diagram in Fig. 1 showing the relation between various processes and governing equations):

- 1) The optical field or optical mode must be solved. A unique feature for laser simulation is that the optical field is coupled to the electronic behavior of the device and, therefore,should be solved in a selfconsistent manner together with the electronic simulation.
- 2) An additional rate equation must be solved. The rate equation is the link between the optical power and electronic behavior.
- 3) An accurate gain function should be computed for the material of interest. The need for an accurate gain model is more obvious in the case of strained quantum well laser since a key incentive of using a strained quantum well is to improve the gain of the material.

This chapter details the physical models and numerical techniques needed to develop a 2D quantum well laser model. By 2D model it is meant a description of a 2D cross section in the transverse and lateral directions of a laser, or a small segment of a 2D cross-section in the longitudinal direction, e.g., one period of a DFB laser (see Figure 2). The model of longitudinal modal behavior is out of the scope of this chapter.

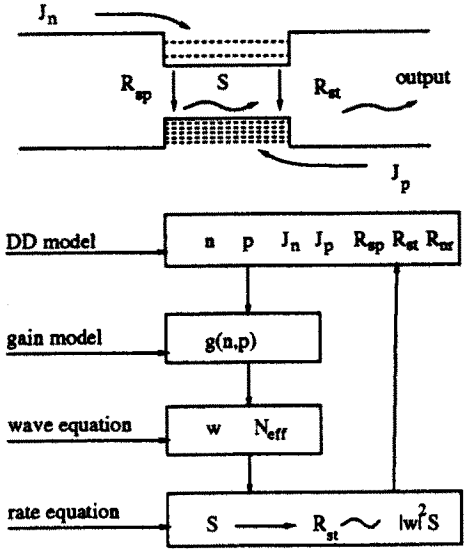
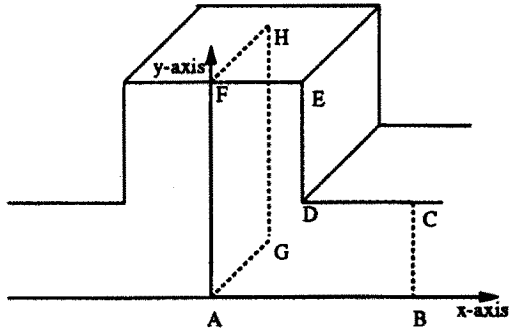


Figure 1. Physical processes and related equations of a quantum well laser.



Choice of 2D Section for Simulation:

1) ABCDEF (xy-plane).

2) AGHF (zy-plane).

Figure 2. Schematics of different 2D cross sections in a laser.

The organization of this chapter is as follows. Section 2 presents the basic equations needed to describe the semiconductor laser behavior. Section 3 details the discretization schemes used to solved the equations. Section 4 discusses the numerical techniques used to solve the discretized equations. Section 5 gives details in the modeling of quantum well optical gain in a form suitable for 2D simulations. An example of 2D simulation is briefly described in Section 6.

2. Basic Equations

2.1 Basic Differential Equations.

The basic equations used to describe the semiconductor laser behavior are the Poisson's equation:

$$-\nabla.\left(\frac{\epsilon_0\epsilon_{dc}}{q}\nabla V\right) = -n+p+N_D(1-f_D)-N_A f_A+\sum_j N_{tj}(\delta_j-f_{tj}), \quad (1)$$

and the current continuity equations for electrons and holes:

$$\nabla \cdot J_n - \sum_j R_n^{tj} - R_{sp} - R_{st} - R_{au} = \frac{\partial n}{\partial t} + N_D\frac{\partial f_D}{\partial t}, \quad (2)$$

$$\nabla \cdot J_p + \sum_j R_p^{tj} + R_{sp} + R_{st} + R_{au} = \frac{-\partial p}{\partial t} + N_A\frac{\partial f_A}{\partial t}. \quad (3)$$

where the symbols are defined as follows.

- V: electrostatic potential;
- n: electron concentration;
- p: hole concentrations;
- R_n^{tj} : recombination rate due to electron trapping;
- R_p^{tj} : recombination rate due to hole trapping;
- R_{st} : recombination rate due to stimulated emission;

- R_{au} : Auger recombination rate;
- f_A : occupancy of shallow acceptor;
- f_D : occupancy of shallow donor;
- f_{tj} : occupancy of deep level trap (of the j'th level);
- N_A : shallow acceptor concentration;
- N_D : shallow donor concentration;
- N_{tj} : concentration of deep level trap (of the j'th level);
- J_n : electron flux density;
- J_p : hole flux density;
- t : time;
- ϵ_0 : dielectric constant of vacuum;
- ϵ_{dc} : relative dielectric constant of a semiconductor at zero frequency (DC);
- q : electron charge;

These equations govern the electrical behavior (e.g., I-V characteristics) of a semiconductor laser. They are also called drift-diffusion equations and are widely used in the electronic device simulation [4,5].

In addition two more equations must be solved to describe the optical behavior (i.e., the optical field distribution and photon number in the laser cavity):

$$\nabla^2 W + k_0^2(\epsilon - \beta^2)W = 0, \tag{4}$$

$$\frac{c}{\beta_1}[g_m - \alpha_{int} - \alpha_m]S + c_m \int R_{sp} dv = \frac{\partial S}{\partial t} \tag{5}$$

where W is the complex wave amplitude and S is the total photon number. ϵ is the position and concentration dependent dielectric constant at the optical frequencies, or the square of the local refractive index. β is the effective refractive index which must be found from the solution of the wave Eq. (4). c_m is the fraction of spontaneous emission coupled into the lasing mode. g_m is the modal gain and α_{int} is the internal loss. Both of these are weighted average of the local material gain and loss over the 2D segment of interest. Therefore Eq. (5) is an integral equation. Equation (4) is the complex scalar wave equation and

can be derived from the Maxwell's equations for the transverse/lateral cross section of a laser [6].

Sometimes it is desirable to simulate a small segment of a laser in the longitudinal direction, e.g., to study the gain distribution within a period of a gain coupled DFB laser. In this case Eq. (4) is only valid in the transverse direction (see Figure 2) and the distribution of optical field in the longitudinal direction should be determined with a suitable model such as transfer matrix method or coupled mode theory [6].

Another set of basic equations further defines the electron and hole current in terms of the carrier concentration and the quasi-Fermi levels [7]:

$$J_n = n\mu_n \nabla E_{fn}/q, \tag{6}$$

$$J_p = p\mu_p \nabla E_{fp}/q, \tag{7}$$

where μ_n and μ_p are mobilities of electrons and holes, respectively. E_{fn} and E_{fp} are the quasi-Fermi levels for electrons and holes, respectively.

The importance of various basic equations in the laser simulation is schematically described in Fig. 1. For a semiconductor diode to lase, electrons and holes (J_n and J_p) are injected into active region (quantum wells). The drift-diffusion (DD) model is used to determine the electron and hole distribution of the injection process. When the injection level is high enough to cause population inversion, i.e., to cause the split of quasi-Fermi levels to be larger than the bandgap,the optical gain $g(n,p)$ becomes positive and light from spontaneous emission gets amplification as described by the rate and the wave equations. The stimulated emission, being proportional to the light intensity in the cavity, couples to the drift-diffusion equation when the light intensity is strong enough. These process clearly indicates that a selfconsistent solution of all the basic equations must be found to describe the laser operation properly.

The goal of the 2D simulation is to solve for the partial differential and integral equations over a segment of the laser device. Several important terms such as the mobility, recombination, and stimulated emission, are discussed or derived in more detail in the following Subsections.

2.2 Complex Dielectric Constant and Optical Gain

The complex dielectric constant at the optical frequencies appearing in Eq. (4) is related to the local optical gain function g by the following relation:

$$\epsilon = \epsilon_1 + i\epsilon_2 = [\overline{n} - \Delta\overline{n} - i(g-a)/2k_0]^2, \tag{8}$$

$$\Delta\overline{n} = \gamma_l(g-a)/2k_0, \tag{9}$$

where $\overline{n}$ is the real part of the refractive index under transparent condition. An overline is used here to avoid confusion with electron concentration n. γ_l is the line width enhancement factor relating the real and imaginary part of the refractive indices. All of $\overline{n}$, g and α are functions of position. g is dependent on the carrier density in the system and thus on all the equations.

The imaginary part of ϵ (which is proportional to the energy gain or loss in a material) is separated into two parts corresponding to the optical gain g and loss α:

$$\epsilon_2 = \epsilon_2^g + \epsilon_2^\alpha. \tag{10}$$

To a good approximation these two terms are proportional to g and α, respectively.

$$\epsilon_2^g = [\overline{n} - \Delta\overline{n}]g/k_0, \tag{11}$$

$$\epsilon_2^\alpha = [\overline{n} - \Delta\overline{n}]\alpha/k_0. \tag{12}$$

2.3 Stimulated Emission and Rate Equation

An excellent derivation of the basic rate equation for the amplitude and phase from the Maxwell's equations was given in Ref. [6]. In the 2D simulation described here the total light power or the total number of photons is of concern. To derive a rate equation for the total photon number, it is necessary to use the integral form of the Maxwell equation, i.e., the energy conservation relation,over the whole laser cavity. Following Ref. [1] the derivations of the stimulated emission rate and Eq. (5) are given as follows.

The optical power generation (because of interband transition, i.e., local gain) per unit volume can be expressed as:

$$power\ gen. = \sigma F^2 = -\epsilon_2^g \epsilon_0 \omega F^2. \tag{13}$$

where F is the root-mean-square of the electrical field at the optical frequencies. σ is the conductivity at the optical frequencies. The convention here is that negative ϵ_2 represents power generation. For convenience a constant γ is introduced such that the complex wave amplitude is related to the electric field by

$$|F|^2 = \gamma |W|^2. \tag{14}$$

To determine γ one uses the following basic relation:

$$S\hbar\omega = \int \epsilon_0 \epsilon_1 |F|^2 dv = \epsilon_0 \gamma \int \epsilon_1 |W|^2 dv = \epsilon_0 \gamma < \epsilon_1 >, \tag{15}$$

where the integration is over the whole cavity.

The power generation becomes

$$power\ gen. = -S\hbar\omega^2 \epsilon_2^g |W|^2 / < \epsilon_1 > . \tag{16}$$

For the stimulated emission the power generated due to interband transition converts into recombination of electron-hole pairs accompanied by the emission of a photon:

$$R_{st} = power\ gen./\hbar\omega = -\epsilon_2^g \omega |W|^2 S / < \epsilon_1 >, \tag{17}$$

The rate equation for the photon number in the laser cavity should be viewed as a statement of the conservation of energy, i.e.

$$\begin{aligned} &stimulated\ emission - internal\ loss - emitted\ power \\ &\quad + spontaneous\ emission = power\ increase \end{aligned} \tag{18}$$

where each term is in the unit of photon number per unit time (i.e. energy per ($\hbar\omega$) per second).The "power increase" in the above equation is the rate of photon number ($\partial S/\partial t$). Eq. (18) can be detailed as follows:

$$stimulated\ emission = \int R_{st} dv = \frac{c}{\beta_1} g_m S \tag{19}$$

where

$$g_m = \beta_1 k_0 \frac{< \epsilon_2^g >}{< \epsilon_1 >}. \tag{20}$$

where β_1 is the effective real index of the device (or solution of the wave equation).

According to Eq. (11), the modal gain defined above [Eq. (20)] can be approximated by

$$g_m \approx \int |W|^2 g dv \tag{21}$$

Note that the wave function W has been normalized such that the above integral yields the Γ factor if the local gain is uniform in the active region. Therefore the definition here reduces to the

$$g_m = \Gamma g \tag{22}$$

In a 2D simulation, there is no need to use the Γ factor here because the optical field distribution is computed and can be directly used in Eq. (20).

Similarly the internal loss can be written as

$$\alpha_{int} \approx \int |W|^2 \alpha dv \tag{23}$$

Define an effective loss coefficient α_m due to emitted power loss from the facet or from the gratings of a DFB, and the emitted power in Eq. (18) can be written in the same form as Eq. (19):

$$emitted\ power = \frac{c}{\beta_1} \alpha_m S \tag{24}$$

In the special case of a Fabry-Perot cavity,the emitted power loss equals the facet mirror loss and one has the following familiar formulae:

$$emitted\ power = S \frac{c}{\beta_1} \frac{1}{L} \ln\left(\frac{1}{r_m}\right). \tag{25}$$

The power due to spontaneous emission is simply the integration of the spontaneous emission rate over the cavity:

$$spontaneous\ emission = c_m \int R_{sp} dv, \tag{26}$$

Summation of the above terms yields the rate equation:

$$\frac{c}{\beta_1}(g_m - \alpha_{int} - \alpha_{em}) + S + c_m \int R_{sp} dv = \partial S / \partial t \tag{27}$$

The rate equation has averaged out most details of longitudinal distribution of the optical fields. Since the rate equation is essentially an average over the whole cavity, it is still valid even if the cross section of simulation is over the a segment along the longitudinal direction [but the 2D wave equation Eq. (4) needs to be modified].

The discussion here considers only one dominant longitudinal mode. The extension to multi-longitudinal mode is out of the scope of this chapter.

2.4 SRH and Auger Recombination

The most important non-radiative recombination mechanisms are the Shockley-Read -Hall (SRH) recombination and the Auger recombination. SRH recombination involves recombination centers (also called deep level traps) with energy levels located within the energy bandgap of the semiconductor. The Auger recombination involves high energy three-particle interactions and is proportional to the third order of the carrier concentrations [6].

To model the SRH recombination properly, one should use the following equation for the recombination rate [8]:

$$R_n^{tj} = c_{nj} n N_{tj}(1 - f_{tj}) - c_{nj} n_{1j} N_{tj} f_{tj}, \tag{28}$$

$$R_p^{tj} = c_{pj} p N_{tj} f_{tj} - c_{pj} p_{1j} N_{jk}(1 - f_{tj}). \tag{29}$$

where n_{1j} is the electron concentration when the electron quasi-Fermi level coincides with the energy level, E_{tj} of the jth trap. Similar definition applies to p_{1j}.

Under transient condition the following trap dynamic equation is valid:

$$N_{tj} \frac{\partial f_{tj}}{\partial t} = R_n^{tj} - R_p^{tj}. \tag{30}$$

which is basically a mathematical statement that the change of the trap occupancy is caused by electron capture and emission from the trap

center. The capture coefficients c_{nj} and c_{pj} for electrons and holes relates to the life time of the carrier due to the j the recombination center by the following relations:

$$\frac{1}{\tau_{nj}} = c_{nj} N_{tj}, \tag{31}$$

$$\frac{1}{\tau_{pj}} = c_{pj} N_{tj}. \tag{32}$$

The capture coefficients can be further expressed as capture cross sections and thermal velocities:

$$c_{nj} = \sigma_{nj} \overline{v}_n, \tag{33}$$

$$\overline{v}_n = \sqrt{\frac{8kT}{\pi m_n}}, \tag{34}$$

$$c_{pj} = \sigma_{pj} \overline{v}_p, \tag{35}$$

$$\overline{v}_p = \sqrt{\frac{8kT}{\pi m_p}}. \tag{36}$$

Under steady state condition Eqs. (28) to (32) results in the more familiar Schockley-Read-Hall recombination formulae [7]:

$$R_n^{tj} = R_p^{tj} = \frac{pn - n_i^2}{\tau_{pj}(n + n_{1j}) + \tau_{nj}(p + p_{1j})} \tag{37}$$

Therefore a trap (or recombination center) is completely specified by its density N_{tj}, capture cross sections σ_{nj} and σ_{pj}, and energy level E_{tj}.

The Auger recombination rate is given by [4,6]:

$$R_{au} = (C_n n + C_p p)(np - n_i^2), \tag{38}$$

where the Auger coefficients C_n and C_p depends on the type of material simulated. A complete microstructural model of Auger recombination is very complicated [6] and is out of the scope of this chapter. Here it is assumed that both in the quantum well and in the bulk, Eq. (38) is valid and the coefficients C_n and C_p are empirical parameters.

2.5 Carrier Statistics

A semiconductor laser operates under high injection levels of electrons and holes and therefore the energy distribution of the carriers are highly degenerate. The Fermi-Dirac statistics should be used in this case [7]. The electron and hole concentrations with Fermi-Dirac distributions for a bulk material is given by:

$$n = N_c F_{1/2}\left(\frac{E_{fn} - E_c}{kT}\right), \tag{39}$$

$$p = N_v F_{1/2}\left(\frac{E_v - E_{fp}}{kT}\right), \tag{40}$$

where $F_{1/2}$ is the Fermi integral of order one-half. Here a parabolic band model is assumed.

For the convenience of numerical evaluation, the approximation proposed by Bednarczyk and Bednarczyk is used [9]:

$$F_{1/2}(x) \approx (e^{-x} + \xi(x))^{-1}, \tag{41}$$

$$\xi(x) = \frac{3}{4}\sqrt{\pi}[\nu(x)]^{3/8}, \tag{42}$$

$$\nu(x) = x^4 + 50 + 33.6x\left\{1 - 0.68\exp[-0.17(x+1)^2]\right\}. \tag{43}$$

This expression is accurate to within 0.4% of error in all ranges.

In the limit of low carrier concentration Eqs. (39) and (40) reduce to the familiar Boltzmann statistics:

$$n = N_c \exp\left(\frac{E_{fn} - E_c}{kT}\right), \tag{44}$$

$$p = N_v \exp\left(\frac{E_v - E_{fp}}{kT}\right), \tag{45}$$

In the case of quantum wells the carrier concentration is the sum of contributions from all subbands as follows.

$$n = \sum_j \rho_j^{x0} kT \ln[1 + e^{(E_{fn} - E_j)/kT}] + \textit{unconfined electrons}, \tag{46}$$

$$p = \sum_i \rho_i^{xo} kT \ln[1 + e^{(E_i - E_{fp})/kT}] + \textit{unconfined holes}, \tag{47}$$

where the label "j" and "i" are used to denote the subbands of the conduction band the valence band. ρ_i^{x0} and ρ_j^{x0} are the densities of states in the plane of the quantum well for the valence band and the conduction band, respectively.

2.6 Incomplete Ionization of Impurities

A semiconductor laser diode normally uses high density of dopants to define the pn-junction. Under high doping or high carrier injection levels, the quasi-Fermi levels are close to the impurity energy levels near the band edge. The dopants are in a state of partial ionization. To account for the incomplete ionization of shallow impurities in semiconductors, the occupancies f_D and f_A are used to describe the degree of ionization. It is assumed that the shallow impurities are in equilibrium with the local carriers, i.e., they share the same quasi-Fermi levels. Therefore the occupancy of the shallow impurities can be described by [7]:

$$f_D = \frac{1}{1 + g_d^{-1} \exp[(E_D - E_{fn})/kT]}, \tag{48}$$

$$f_A = \frac{1}{1 + g_a \exp[(E_A - E_{fp})/kT]}, \tag{49}$$

where the subscripts D and A are used to denote shallow donors and acceptors, respectively.

From the discussions in Subsection 2.4 the occupancy of a deep level trap can be determined through the trap dynamic equation, Eq. (30). In general the deep trap is not in equilibrium with the carriers (i.e., the trap does not share the same quasi-Fermi level as the carriers). From Equations (28), (29) and (30), one obtains the following expression for the trap occupancy under steady state condition:

$$f_{tj} = \frac{c_{nj} n + c_{pj} p_{1j}}{c_{nj}(n + n_{1j}) + c_{pj}(p + p_{1j})}. \tag{50}$$

The ionization states of the impurities are important because they affect the potential or energy band profile through the Poisson's equation.In the case of surface states or surface recombination centers, the treatment is similar. One can treat the surface states as deep levels traps located near the surface and they should be described in quantities per surface area. One can show that the deep trap models here

can be used to describe effects due to surface charge states as well as to surface recombination. For example, Fermi level pinning effect on a semiconductor surface can be modeled using this approach.

In a transient simulation the trap occupancy is a function of time depending on the trap capture rates as well as on the local carrier concentrations. Equation (30) should be used to determine the trap states in each transient time step.

2.7 Carrier Mobility

Mobility of carriers in a laser affects the following electrical behavior. It directly determines the speed at which the laser can be switched on and off. It affects the leakage current due to diffusion over the barrier. It also affects the potential drops (or profile) across the diode for a given current.

An accurate carrier mobility model should account for scattering mechanism in electrical transport. In general the mobility is a function of the electrical field and the dopant densities [7]. Excellent analytical mobility models have been described in Ref. [5]. The models in Ref. [5] are mainly for Silicon and GaAs but should be valid for other types of semiconductors with minor modifications.

For the field dependence the following equations can be implemented in a simulator [5]:

$$\mu_n = \frac{\mu_{0n}}{\left(1 + \left(\mu_{0n} F / v_{sn}\right)^{\beta_n}\right)^{1/\beta_n}}, \tag{51}$$

$$\mu_p = \frac{\mu_{0p}}{\left(1 + \left(\mu_{0p} F / v_{sp}\right)^{\beta_p}\right)^{1/\beta_p}}, \tag{52}$$

where F, v_{sn} and v_{sp} are used to denote the static electric field, saturation electron and hole velocities, respectively. These formulas describe a rather simple field dependence for the carrier velocity. The velocity increases with field linearly at low field with a constant mobility (μ_{0n} or μ_{0p}), but saturates to a fixed velocity (v_{sn} or v_{sp}) at the high field limit. The smoothness of this saturation is controlled by the parameter β_n (or β_p).

Many III-V compound semiconductor [(e.g., GaAs) exhibit negative differential resistance due to transition of carrier into band valleys

with lower mobility [7]. In such a case the following field dependence can be used [5]:

$$\mu_n = \frac{\mu_{0n} + (v_{sn}/F_{0n})(F/F_{on})^3}{1 + (F/F_{0n})^4}. \tag{53}$$

Experience shows that Eqs. (51) and (52) results in a stable solution. The field dependence in Eq. (53)can give kinks in the L-I curve for some devices because of the bistability associated with the negative differential resistance.

Besides the field dependence of the mobility, another important effect is the impurity dependence of the low field mobility [7]. Reference [5] has proposed the following empirical expression for this dependence:

$$\mu_{0n} = \mu_{1n} + \frac{(\mu_{2n} - \mu_{1n})}{1 + \left(\frac{N_D + N_A + \sum_j N_{tj}}{N_{rn}}\right)^{\alpha_n}}, \tag{54}$$

$$\mu_{0p} = \mu_{1p} + \frac{(\mu_{2p} - \mu_{1p})}{1 + \left(\frac{N_D + N_A + \sum_j N_{tj}}{N_{rp}}\right)^{\alpha_p}}. \tag{55}$$

where parameters such as μ_{1n} and μ_{2n} are fitting parameters from experimental data. Physically these dependence takes into account the effect of impurity scattering to mobility.

3. Discretization

To solve the device equations on a computer, one must discretize them on a simulation grid. That is, the continuous functions of the PDE's are represented by vectors of function values at the nodes, and the differential operators are replaced by suitable difference operators. Instead of solving for four unknown functions plus the photon number, one solves for $4N + 1$ unknown variables from $4N + 1$ unknown discretized equations, where N is the grid number.

Two methods of discretization are discussed in this section. The first one is the finite difference method (FDM) and the second is the finite element method (FEM). The FEM method commonly used in

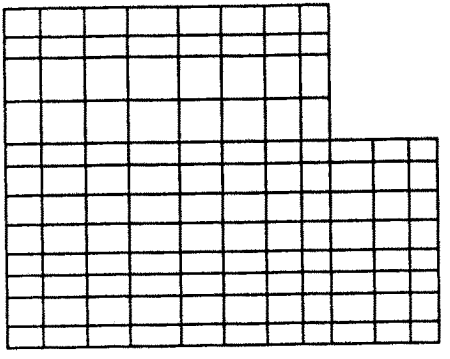
electronic simulator is actually a finite box method on a FEM mesh [10,5]. It will be shown that the two methods are equivalent mathematically and physically. FDM is simple to implement and easy to understand. The key advantage of FEM is that it is capable of handling arbitrary geometry. An additional advantage is that it is more convenient to save mesh points in 2D simulations.

3.1 Generation of Mesh

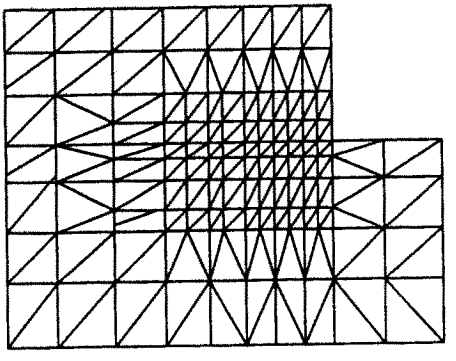
To discretize the equations in 2D the first step is to generate a suitable mesh. For finite difference method the mesh consists of perpendicular lines within a rectangular domain. To implement a mesh generator for FDM one can control each dimension independent of the other. Therefore from the point of view of mesh generation the 2D problem is actually two 1D problems in FDM (Figure 3(a)). In FEM the mesh points are fully 2D in nature and are more difficult to generate and handle. One practical approach [5] is to start with a FDM mesh. The triangulation are done after the some of the mesh lines are terminated in areas where only coarse mesh is required. Similarly one can start with a coarse FDM mesh and put additional mesh lines in areas where fine mesh is required (Figure 3).

It is often necessary to refine a mesh at some stage of a simulation. The reason is that to determine an optimal mesh distribution one needs to know the solution of the equations. But to find the solution of the equations one needs know the mesh first. Therefore a practical approach is to start with an initial mesh and proceed with a solution. The mesh is refined according to the solution and one starts a new solution based on the new mesh. In practice the criteria to refine a mesh region is the difference of a physical quantity between adjacent points [5].

For FDM the refinement is simple because one only needs to add mesh lines in the two directions separately. For FEM one possible scheme is to add smaller triangles within an existing triangle. This method is simple to implement since the refinement procedures do not disturb the existing elements. However, it is easy to see that such a simple procedure can generate a large amount of undesirable obtuse triangles even if the old ones are not obtuse triangles. One solution to this problem is to shift the mesh points near the refinement area to minimize the amount of new obtuse triangles [5]. The effect of obtuse triangles on discretized equations are discussed later in this section.



(a)



(b)

Figure 3. Schematics of (a) finite difference mesh and (b) finite element mesh.

Another method of mesh refinement in FEM is described as follows. One starts with a coarse FDM mesh, and add mesh lines in the refinement area when necessary. The triangulation is done after the refinement in the FDM environment. When further refinement is needed, one abandons the FEM and go back to the FDM environment in which refinement can be done by adding new mesh lines. This method seems to avoid the generation of new obtuse triangles more easily. The only draw back is that one has to be able to switch between FDM and FEM mesh environment constantly, which requires some additional development and computation effort.

3.2 Discretization of Equations

Common to all the basic equations (except the rate equation)the spatial differential operator has the same form as follows:

$$\nabla \cdot \mathbf{J}(\mathrm{x}, \mathrm{y}) = \mathrm{G}(\mathrm{x}, \mathrm{y}, \mathrm{t}) \tag{56}$$

where $\mathbf{J}$ is a vector (e.g., electron current) and G is a scalar function of position and time. In the finite difference (FDM) approximation Eq. (56) can be simply approximated by a five-point difference equation (see also Figure 4):

$$\frac{J_{x2} - J_{x1}}{(d_{x1} + d_{x2})/2} + \frac{J_{y2} - J_{y1}}{(d_{y1} + d_{y2})/2} = G(x_i, y_j, t) \tag{57}$$

where J_{x1} is the x-component of vector $\mathbf{J}$ evaluated at the mid-point between (x_i, y_j) and (x_{i-1}, y_j). Similar definition applies to all the other three components. In the case of FEM Eq. (56) can be discretized using similar ideas. The first step is to convert Eq. (56) into an integral equation as follows.

$$\int_{surf} \mathbf{J} \cdot \mathrm{ds} = \int_{\mathrm{volumn}} \mathrm{Gdv} \tag{58}$$

which can be applied to the small volume surrounding point k (see Figure 5). Then, one obtains the discretized form for FEM:

$$\sum_{k_1} J_{kk_1} h_{k_1} = A_k G(x_k, y_k, t), \tag{59}$$

where k_1 is the neighboring point of k and J_{kk_1} is the projection of $\mathbf{J}$ on the axis of $\mathbf{d}_{\mathrm{kk}_1}$ evaluated at the mid-point. A_k is the area surrounding point k.

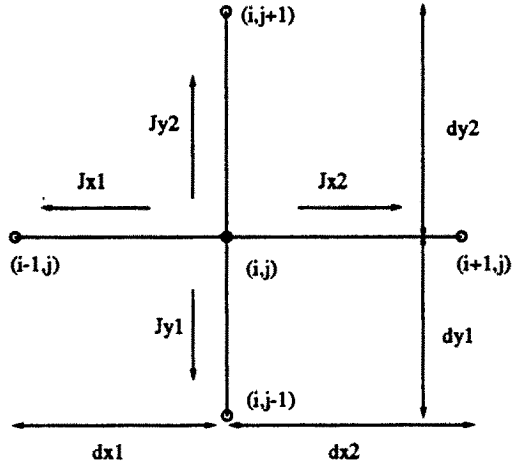


Figure 4. Schematics of the five point difference.

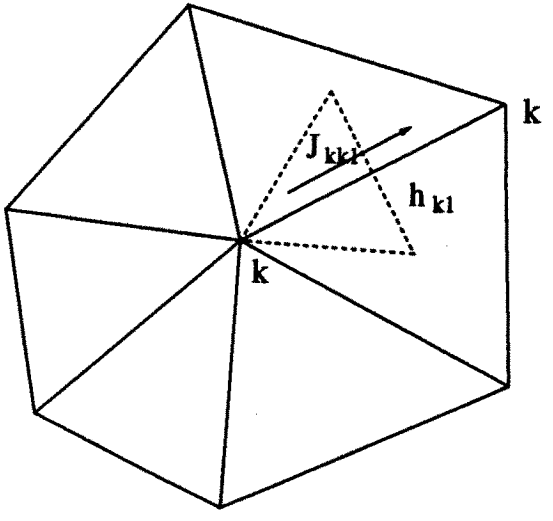


Figure 5. Schematics of the finite box method.

Having discretized Eq. (56) the next major task is to find a proper expression for the components of vector J. For the Poisson's equation and the wave equation, the vector can be simply be expressed in terms of finite difference of the potential and wave amplitude, respectively. In the case of continuity equation, the carrier fluxes must be evaluated with care; the classic finite difference formulas are modified as first demonstrated by Scharfetter and Gummel (SG formulas) in 1969 [11] for silicon devices. The original SG formulas are not applicable for lasers since bandgap variation and Fermi statistics are not taken into account.

A modified SG formulae is given for the x-component of electron current here (the formulas for hole current are completely analogous). One starts with the assumption that the semiconductor material under consideration does not have abrupt junctions (which are to be discussed later), and the material parameters such as bandgap, effective masses and doping can only vary slowly from point to point in the mesh (either FEM or FDM mesh). Define a coefficient γ_F to account for the difference between Fermi statistics and Boltzman statistics as follows.

$$\gamma_F(\mathbf{X}) = \ln[\mathrm{F}_{1/2}(\mathbf{X})/\exp(\mathbf{X})]. \tag{60}$$

Then the electron concentration can be written as

$$n = \exp(E_{fn}/kT + \eta), \tag{61}$$

$$\eta = \frac{kT}{q}\ln(N_c) - E_c/kT + \gamma_F, \tag{62}$$

where η is a function of position. One further assumes that η varies linearly from point to point (in the x-direction for example):

$$\eta = \eta_a x + \eta_b. \tag{63}$$

One obtains an approximate expression for the J_n between x_i and x_{i+1} as follows:

$$J_n = \frac{\mu_n}{q}\exp(E_{fn}/kT + \eta_a x + \eta_b)\frac{dE_{fn}}{dx}. \tag{64}$$

The basic assumption of the original SG formulae is that the current flux between two mesh points is constant and that the mobility is also constant between the points. If the same assumption is applied

to the above equation that J_n is a constant between two mesh points, then the single variable differential equation of Eq. (64) is integrable between the two points. The solution to Eq. (64) can be arranged in the following form:

$$J_n = \mu_n \frac{kT}{q} \frac{n_{i+1} B(\eta_{i+1}) - n_i B(\eta_i)}{x_{i+1} - x_i}, \tag{65}$$

where B is the Bernouli function defined as follows.

$$B(x) = \frac{x}{\exp(x) - 1}. \tag{66}$$

This completes the discussion on spatial discretization.

3.3 Boundary Conditions

In a 2D laser simulation, there are two types of boundary conditions solved for the drift-diffusion equations. The first is the Ohmic contact at the electrodes and the second is the hetero-interface between different material regions.

The basic assumption for the Ohmic contact is charge neutrality and infinity surface recombination velocity. The condition of charge neutrality determines the position of the quasi-Fermi levels at the equilibrium condition [4]. The condition of infinite recombination velocity forces the electron and hole quasi-Fermi levels to coincide with each other and to be fixed at the position at equilibrium. In another word the Ohmic contact is always at the equilibrium condition [4].

The boundary condition for heterointerface can be simulated with a thermionic emission model. The electron at one side of the heterointerface emits to the other side of the interface with a thermal velocity. The current across the interface can be expressed as [12]:

$$J_{hn} = \gamma_{hn} \overline{v}_{bn}^{therm} (n_b - n_{b0}), \tag{67}$$

$$J_{hp} = \gamma_{hp} \overline{v}_{bp}^{therm} (p_b - p_{b0}), \tag{68}$$

where n_b is the electron density at the barrier side ("b" is used to label barrier) and n_{b0} is its value when the quasi-Fermi levels are the same on both sides. γ_{hn} is a scaling factor to account for effects such as tunneling. Note that the boundary current is such that the current flow is zero when the quasi-Fermi levels are flat. The 1-dimensional thermal velocity is defined as [12]:

$$\overline{v}_{bn}^{therm} = \sqrt{\frac{kT}{2m_{bn}\pi}}, \tag{69}$$

$$\overline{v}_{bp}^{therm} = \sqrt{\frac{kT}{2m_{bp}\pi}}. \tag{70}$$

Note that the boundary condition only applies to abrupt junction. For graded junction, the usual current expression in Eqs. (6) and (7) should be used.

3.4 Time Domain Discretization

The most simple method to discretize the differential equations in time is the backward Euler method. Its basic assumption is that for a differential equation of the form:

$$\frac{\partial S}{\partial t} = G(t), \tag{71}$$

the discretized equation is the following:

$$\frac{S(t+\Delta t) - S(t)}{\Delta t} = G(t+\Delta t). \tag{72}$$

This method has the advantage that it is easy to implement and highly stable. In the limit of infinite time the basic equations simply reduce to the steady state equations. The backward Euler method is recommended not just because of its simplicity but also because of its compatibility with discretization scheme of the trap dynamic equations [Eq. (30)].

In a transient state the trap occupancy of a deep level trap depends on many factors such as the history of the occupancy, and electron and hole carrier concentrations. The trap dynamic equation must be discretized to accurately simulate the trap occupancy as a function of time. A discretized expression of the trap dynamic equation was described by Ref. [13]:

$$f_{tj}(t+\Delta t) = \frac{N_{tj} f_{tj}(t) + \left(\frac{n(t+\Delta t)}{\tau_{tj}} + \frac{p_{1j}}{\tau_{pj}}\right)\Delta t}{N_{tj} + \left(\frac{n(t+\Delta t)+n_{1j}}{\tau_{nj}} + \frac{p(t+\Delta t)+p_{1j}}{\tau_{1j}}\right)\Delta t}. \tag{73}$$

where Δt is the time step used in the transient simulation. The above expression can be obtained from Eq. (30) together with Eq. (72).

Once all the equations are discretized in terms of solution of the previous time step, one can treat the solution of the present time step using the same method as for steady state solution, i.e., one can linearize the discretized equations and use Newton's method to solve them.

Strictly speaking the trap dynamic equation should be considered a separate differential equation in addition to the basic equations. However the trap dynamic equation does not depend on spatial variables explicitly. After the equation is discretized in the form of Eq. (73) the trap occupancy appearing in the basic equations can be replaced in terms of the carrier concentration of the present time step and the previous trap occupancy. This essentially eliminates the trap dynamic equation indirectly.

4. Solving the Discretized Equations

4.1 Coupling Between Equations

To determine the optimal numerical approach used for solving the discretized equations, one has to consider the relation between varies equations first. The equations should be solved in a decoupled manner if at all possible since the difficulty or the computation time increases as the square of the number of equations.

In a realistic semiconductor laser, the coupling between the first three equations, which govern the electrical behavior, is usually strong based on experience generated from well established electron device models used in the electronic industry over the years. Since the stimulated emission, which depends on the photon density, is the dominant recombination mechanism beyond lasing threshold, the rate equation is also strongly coupled to the first three equations. In contrast the wave equation is relatively decoupled from the other equations because the optical field distribution for a particular lateral mode is relatively insensitive to the bias conditions of the laser as long as the lasing wavelength does not change too much as a function of bias. This observation makes it possible to decouple the solution of the wave equation from the others.

From the argument above it is concluded that the Newton's method should be used to solve the first three equations plus the rate equation in a completely self-consistent manner. The wave equation is solved and updated separately. Experience shows that this approach is good as long as the updating of the wave equation is frequent enough.

4.2 Newton's method

In Newton's method, all of the variables in the problem are allowed to change during each iteration, and all of the coupling between variables is taken into account. Due to this the Newton algorithm is very stable, and the solution time is nearly independent of bias conditions. The basic algorithm is the generalization of the Newton-Raphson method for the root of a single equation. It can be expressed as follows.

$$F_v^j(V^{j1}, E_{fn}^{j1}, E_{fp}^{j1}) = 0, \tag{74}$$

$$F_n^j(V^{j1}, E_{fn}^{j1}, E_{fp}^{j1}, S) = 0, \tag{75}$$

$$F_p^j(V^{j1}, E_{fn}^{j1}, E_{fp}^{j1}, S) = 0, \tag{76}$$

$$F_S(V^j, E_{fn}^j, E_{fp}^j, S) = 0, \tag{77}$$

Where j runs from 1 to N, $j1$ includes j itself plus its surrounding mesh points. Equations (74) to (77) represent a total number of $3N+1$ equations which is sufficient to solve for $3N+1$ variables: (V^1 , E_{fn}^1 , E_{fp}^1 , V^2 , E_{fn}^2 , E_{fp}^2 , ..., V^N , E_{fn}^N , E_{fp}^N ,S).

Once the equations are discretized in the above form, standard Newton techniques can be used to solve for them. These involve the evaluation of the Jacobian matrix to linearize Eqs. (74) to (77), linear solution of the linearized equations (involving LU factorization of the matrix) and nonlinear iteration to get the final solution. Since the Jacobian matrix is a sparse matrix, sparse matrix solution techniques are used to improve the computation speed.

Acceleration of a Newton iteration can be achieved using the Newton-Richardson method, which only refactors the Jacobian matrix when necessary. When it is not necessary to factorize, iterative method using the previous factorization is employed. The iterative method is extremely fast provided the previous factorization is reasonable. Frequently the Jacobian need only be factorized only once or twice per bias point using Newton-Richardson method, as opposed to the twenty

to thirty times required in the conventional Newton method. The decision to refactor is made on the basis of the decrease per step of the maximum error of both the equation residues and the variable differences.

Special treatment is needed for the discretized rate equation since unphysical situation can arise in a Newton procedure [27]. Consider the rate equation again:

$$-\frac{c}{\beta_1}F_g S + c_m \int R_{sp} dv = 0, F_g = \alpha_{em} + \alpha_{int} - g_m. \tag{78}$$

where α_{em} is the power loss due to emission of light. During a normal Newton iteration procedure F_g could become less or equal to zero before it converges to a physical solution. This causes the photon number S to become infinity and interrupt the Newton solver. To overcome this difficulty, a damping function can be used for F_g:

$$F_g = \alpha_{em}+\alpha_{int}-g_m \simeq \frac{1}{2}\left[\sqrt{(\alpha_{em}-g_m)^2+\delta^2}+\alpha_{int}+\alpha_{em}-g_m\right]. \tag{79}$$

where δ is a sufficiently small number. This function well approximates F_g near the physical solution F_g^0, and prevents the equations from being trapped in the unphysical infinity.

4.3 Solution for Wave Equation

As soon as the solution from the Newton method described in the above section is found, one is ready to solve the wave equation, since the new solution of electron and hole concentrations allows one to obtain a new dielectric constant distribution to be used in the wave solution update.

The scalar wave equation needs special treatment since the its discretized form is a complex eigenvalue problem. After discretization the wave equation Eq. (4) can be written in the following form (in a FDM):

$$W_{ij} = G(W_{i-1j}, W_{i+1j}, W_{ij-1}, W_{ij+1}). \tag{80}$$

where all variables are treated as complex variables. An adaptive, real relaxation parameter ω_a is used in a SOR method:

$$W_{ij}^{k+1} = W_{ij}^{k} + \omega_a[G(W_{i-1j}^{k+1}, W_{i+1j}^{k}, W_{ij-1}^{k+1}, W_{ij+1}^{k}) - W_{ij}^{k}] \qquad (81)$$

where k denotes the k th iteration and ω_a is determined according to the convergence behavior of W. Similarly for the complex eigenvalue β, a real parameter ω_β is introduced in the following SOR relation:

$$\beta_{k+1}^2 = \beta_k^2 + \omega_\beta \left(\epsilon + k_0^{-2} \int W^{*k+1} \nabla^2 W^{k+1} dv - \beta_k^2 \right). \qquad (82)$$

where $\int$ and ∇^2 are evaluated by sum and five point operator of finite difference, respectively. This approach normally guarantees the solution of the fundamental lateral mode.

4.4 Initial Guess

As is common to all Newton's solvers, the initial guess is important for the final solution. Several types of initial guess solutions are possible. The first is the simple charge neutral assumption used to obtain the first (equilibrium) bias point. For the wave equation, the guess solution is a delta function at the center of the active region. This is the starting point of any device simulation. Any later solution with applied bias needs an initial guess of some type, obtained by modifying one or two previous solution(s). When only one previous solution is available, the solution currently loaded is used as the initial guess, modified by setting the applied bias at the contact points.

When two previous solutions with two different bias are available, it is possible to obtain a better initial guess by interpolating the two previous solutions to the present one according to the two previous bias points and the present bias [5].

In principle the Newton nonlinear iteration should always converge as long as the initial guess is close enough to the solution. The closer the initial guess is to the solution, the fewer nonlinear iterations are required to reach convergence. One can take this effect as a way of automatically controlling the bias in a simulation. For example if the present bias step takes too few number of iterations to reach convergence, one can increase the bias step; On the other hand, if convergence for the present bias step takes too many iterations or simply

fail to converge, one can reduce the bias step. In the limit of zero bias step, no iteration is needed to reach convergence.

To conclude the discussion on techniques on solving the drift-diffusion equations, it should be pointed out that good simulation packages (see, for example, Ref. [5]) of electronic devices using advanced numerical techniques have generated large amount of experience in dealing with practical numerical problems. Consultation of previous works in solving drift-diffusion equations is highly recommended for semicondutor laser modeling.

5. Quantum Well Gain Function

Optical gain in a semiconductor diode under high forward bias condition makes it possible to amplify the light in a laser. The optical gain also makes its modeling unique compared with modeling of other electronic devices. This section gives a semi-classical derivation of the gain function for a quantum well system. The effect of strain will also be discussed.

Throughout this section it is assumed that all the quantum levels for the sub-bands of Γ, L, light holes and heavy holes can be computed from well known formulas in quantum mechanics for a square quantum well [4]. No further details will be discussed regarding the subband levels in the quantum wells.

5.1 A semi-classical derivation

A rigorous derivation of the optical gain involves theories of interaction of a solid with quantized radiation fields [6,15]. Getting the algebra right in the derivation is not trivial, since one has to deal with factors like 4π, c and $\hbar$, depending on which unit system one chooses. In fact the factor appearing in the gain has caused confusion in the literature before [16]. For the purpose of getting the correct factor in the gain expression, a semi-classical derivation (in MKS unit) is given in this Subsection. The derivation only assumes basic knowledge of elementary quantum mechanics [14] and classical electrodynamics [18]. Therefore, it is easy to follow for those readers who are not familiar with theories of field quantization.

The first step is to express the Hamiltonian in terms of the momentum matrix element. Consider a small volume of solid (i.e., small compared with the wavelength of the light) under the radiation of a lightwave with an oscillating electric field $F(t)$:

$$F(t) = F_0 cos(\omega t). \tag{83}$$

The Hamiltonian of an electron in the solid can be written as

$$H(t) = H_0 cos(\omega t) = qxF_0 cos(\omega t), \tag{84}$$

where x is the displacement operator of the electron. The radiation causes the electron to make a transition from state $|j>$ (in the conduction band) to $|i>$ (in the valence band). The basic relation:

$$[x, H] = \frac{i\hbar M}{m_0}, \tag{85}$$

results in

$$x_{ij} =< i|x|j >= \frac{< i|M|j >}{im_0\omega} = \frac{M_{ij}}{im_0\omega}, \tag{86}$$

where M is the momentum operator. Equations (84) and (86) yields the following:

$$|H_{0ij}|^2 = | < i|H_0|j > |^2 = \left(\frac{qF_0 M_{ij}}{m_{0\omega}}\right)^2. \tag{87}$$

The second step is to find the relation between the optical gain and the interband optical transition probability. Classical electrodynamics gives the power gain (or loss) associated with the light wave as follows [18]:

$$\begin{aligned} power &= \frac{1}{2}\omega\epsilon_0\epsilon_2 F_0^2 \\ &= \frac{1}{2}c\epsilon\overline{n}g_{ij}F_0^2. \end{aligned} \tag{88}$$

where the relation between gain (g_{ij}) and the imaginary part of the dielectric constants has been used. On the other hand the optical power

can be written as the the transition probability per unit volume (w_{ij}) times $\hbar\omega$:

$$power = \hbar\omega w_{ij} \tag{89}$$

Combining Eqs. (88) and (89) gives:

$$g_{ij} = \frac{2\hbar\omega w_{ij}}{c\epsilon_o \overline{n} F_0^2}. \tag{90}$$

The final step is to calculate the transition probability density w_{ij} which can be obtained from the Golden rules [14] in quantum mechanics:

$$w_{ij} = \int \frac{\pi}{2\hbar} |H_{oij}|^2 (f_j - f_i)\delta(E_{ij} - \hbar\omega)\rho_{ij} dE_{ij}, \tag{91}$$

$$= \frac{\pi}{2\hbar} |H_{oij}|^2 (f_j' - f_i')\rho_{ij}. \tag{92}$$

where ρ_{ij} is the reduced density of states, and f_i and f_j are the Fermi functions for the i th and the j th levels, respectively, and f_i' and f_j' are given by

$$f_i' = \left\{ 1 + \exp\left[(E_i^0 - \frac{m_{ij}}{m_i}(E - E_{ij}^0) - E_{fp})/kT \right] \right\}^{-1}, \tag{93}$$

$$f_j' = \left\{ 1 + \exp\left[(E_j^0 + \frac{m_{ij}}{m_j}(E - E_{ij}^0) - E_{fn}/kT \right] \right\}^{-1}. \tag{94}$$

where $E = \hbar\omega$. E_i^0 and E_j^0 are the maximum of the valence subband and the minimum of conduction subband, respectively. E_{ij}^0 is the difference between them.

Note the factor $f_j - f_i$ comes from the net probability of transition from j to i states. The probability of j to i transition is proportional to $f_j(1 - f_i)$ while this process is counter balanced by that of the opposite process (i to j transition) with a probability of $f_i(1 - f_j)$. Therefore the net transition probability has a factor:

$$f_j - f_i = f_j(1 - f_i) - f_i(1 - f_j). \tag{95}$$

Finally combination of Eqs. (92), (90), and (87) yields the desired gain expression:

$$g_{ij} = \frac{\pi q^2 M_{ij}^2 \rho_{ij}}{c\epsilon_0 \overline{n} m_0^2 \omega}(f_j' - f_i'). \tag{96}$$

Note that Eq. (96) is rather general. In the case of quantum well one can use 2D reduced density of states for ρ_{ij} :

$$\rho_{ij} = \frac{m_{ij}}{\pi \hbar t_{wq}}, \tag{97}$$

$$m_{ij} = \frac{m_i m_j}{m_i + m_j} \tag{98}$$

while in the case of bulk material, one can replace ρ_{ij} by the corresponding 3D reduced density of states.

There are several important characteristics one should notice about the important gain expression of Eq. (96) as a function of frequency.

- 1) For quantum well system, the density of state has a step function. This gives rise to the sharp steps in the unbroadened gain spectrum. The gain is zero when the frequency is below the bandgap.
- 2) The factor $f_j' - f_i'$ determines whether the gain is positive or negative (loss). One can show that the gain is positive only when the splitting of the quasi-Fermi levels is greater than the bandgap (E_{ij}^0). Physically it means that only under high injection condition, can the material exhibit stimulated emission (or gain).
- 3) The high frequency roll off of the gain spectrum is caused by the the sensitivity of $f_j' - f_i'$ to the frequency. The physical reason is that in the high frequency limit, the probability of finding an electron with high energy in the conduction band and a vacancy in the low energy of the valence band is too small to cause stimulated emission (or gain).

5.2 Selection Rules

The selection rules for the optical transition from the evaluation of the momentum matrix element. In general the momentum matrix element M_{ij}^2 can be expressed as the bulk material matrix element,

times the overlap integral of the conduction subband and valence subband states, and times a factor to take into account the polarization effects in a quantum well environment [19]. For example the matrix element of TE mode for an unstrained quantum well can be written as:

$$A_{hh} = \frac{3 + 3E_{ij}^0/E}{4}, \tag{99}$$

$$A_{lh} = \frac{5 - 3E_{ij}^0/E}{4} \qquad \text{for } E > E_{ij}^0, \tag{100}$$

$$A_{hh} = 3/2, A_{lh} = 1/2 \qquad \text{for } E \leq E_{ij}^0, \tag{101}$$

$$M_{hh} = A_{hh} O_{ij} M_0, \tag{102}$$

$$M_{lh} = A_{lh} O_{ij} M_0, \tag{103}$$

The selection rules come in because the overlap integral vanish for many pairs of transition states. For example, all even-to-odd and odd-to-even overlap integrals vanish. If the effective masses of the valence band and conduction band are identical, the only non-vanishing integral is the one with the same subband label in both the conduction subband and the valence subbands. However in general all transitions of the same parity should be considered.

5.3 Models of Gain Broadening

The intra-band scattering broadening significantly reduces the the local gain function and round off the sharp peaks in the gain spectrum. The most simple model for gain broadening is to use the following Lorentzian function in a convolution integral:

$$g(E) = \int g^0(E_1) F_l(E_1 - E) dE_1, \tag{104}$$

where g^0 is the gain function without broadening. The Lorentzian function is given by the expression:

$$F_l(E_1 - E) = \frac{1}{2\pi} \frac{\Gamma_0}{(\Gamma_0/2)^2 + (E_1 - E)^2}, \tag{105}$$

where Γ_0 is the half width of the broadened energy level and is assumed a constant.The assumption of constant Γ_0 means the whole

gain spectrum is broadened to the same degree across the whole spectrum. The advantage of this model is that it is relatively simple to implement and results in a most stable solution in a 2D simulation.

However the broadening across the spectrum should not be uniform according to the Landsberg's model [20–22]. The half width Γ is a complicated function of the transition energy which has its maximum value at the bottom of the band (or band edge) and decreases to zero as the energy approaches the quasi-Fermi level. Therefore the half width is a function of both the transition energy and the carrier density. One can use the following approximation according to Martin and Stormer [22]:

$$F_l(E_1 - E) = \frac{1}{2\pi} \frac{\Gamma(E_1)}{[\Gamma(E_1)/2]^2 + (E_1 - E)^2}, \tag{106}$$

$$\Gamma(E_1) = \Gamma_0 \left[1 - 2.229 \frac{E_1}{E_{fn} - E_{fp}} + 458 \left(\frac{E_1}{E_{fn} - E_{fp}} \right)^2 - 0.229 \left(\frac{E_1}{E_{fn} - E_{fp}} \right)^3 \right].$$

$$\text{for } E_{fn} - E_{fp} \geq 0 \;\&0\; \leq E_1 \leq E_{fn} - E_{fp} \tag{107}$$

For bulk material a Γ_0 value of 2 meV was found. For quantum wells the broadening is much larger (up to 30 meV). The implementation of this model is important for accurate modeling of the gain spectrum especially for quantum well. Usually the maximum broadening Γ_0 in the Landsberg model is material and process dependent and therefore, is treated as a fitting parameter.

Note that Eq. (107) is valid only when the peak gain is positive, i.e., only when the difference between the quasi-Fermi levels is greater than the bandgap of the quantum well subbands. The transition energy must also be between the band gap and the this difference. To extend the spectrum to other conditions, the formulas proposed by Zielinski et. al. [23] can be used:

$$g(E) = I_{conv}(g^0, E) \qquad \text{for } E \leq E_{fn} - E_{fp}, \tag{108}$$

$$g(E) = I_{conv}(g^0, E) + g^0(E) \qquad \text{for } E \geq E_{fn} - E_{fp}, \tag{109}$$

with

$$I_{conv}(g^0, E) = \sum_{i,j} \int_{E_{ij}}^{E_{fn}-E_{fp}} \frac{1}{2\pi} \frac{g^0(E_1)\Gamma(E_1)dE_1}{[\Gamma(E_1)/2]^2 + (E_1 - E)^2}$$

$$for E_{fn} - E_{fp} \geq E_{ij}, \tag{110}$$

$$I_{conv}(g^0, E) = 0$$

$$for E_{fn} - E_{fp} < E_{ij}. \tag{111}$$

The two-piece function is continuous across the spectrum because $g^0(E) = 0$ when $E = E_{fn} - E_{fp}$.

5.4 Strained Quantum Wells

Biaxial strain is introduced into the quantum well when the lattice constant is different from that of the barrier material.Inclusion of biaxial strain into the design of quantum-well semiconductor lasers provides an additional degree of freedom and produces some desirable effects such as a lower threshold current. The effects of strain have been conventionally investigated theoretically using a k.p description of the band structure [24]. This type of calculation is usually limited to the computation of optical gain and has limited capability in analyzing practical design issues, such as the optimization of the laser geometry.

Strain is known to cause the valence band of a III-V semiconductor to split into separate light hole (LH) and heavy hole (HH) bands which are strongly non-parabolic. A rigorous treatment of the strain effects on the band structure is rather complicated [24]. For example the gain function involves an integration over k-space. Hence a full treatment of the strain is difficult to include into a 2D model, the major difficulty being the non-parabolicity of the band structure. The concept of effective mass is no longer valid because of the non-parabolicity.

Recently a anisotropic effective mass model was proposed [25,26] in an attempt to revive the concept of effective mass. The basic idea is to fit the nonparabolic band structure with a parabolic anisotropic band structure model. The advantage is that once the effective masses are known from the fitting, all the formulas developed over the years for unstrained quantum wells can be borrowed directly with minor modifications. It was demonstrated that such an approach yields reasonable results and that the gain function for strained quantum wells is efficient and easy to implement.

To be more specific about the anisotropic effective mass approximation [25,26], the effective of strain is to make the parabolic band anisotropic, i.e., the effective valence band mass perpendicular to the quantum well plane is different from that parallel to the plane. The effective mass perpendicular to the plane (m_{vy}) determines the quantum subband levels (or quantum confinement effects) and the optical transition energies. The (2D) density of states and joint density of states of each subband depends on the effective mass in the plane (m_{vx}). More details about this approach can be found in Refs. [25,26].

5.5 Implementation of Gain Model in 2D Simulation

For the gain function to be included into the 2D model, it must be efficient to evaluate at any frequency and at any location in the device. The broadening integral in Eq. (105) does not have analytical solution. Numerical evaluation at every mesh point can be rather time consuming.

Reference [27] has proposed a scheme to obtain an approximate analytical solution:

$$F_l(x) = \frac{\pi^{-1}}{1+x^2} = \sum_{k=1}^{3} A_k e^{-B_k x}. \tag{112}$$

For the gain function the following approximation is made:

$$\begin{aligned} (1+e_A)^{-1} &\simeq 1 + \sum_{k=1}^{4} a_k e_{kA} \\ &\text{for} \quad A \leq 0 \\ &\simeq \sum_{k=1}^{4} a_k e^{-kA} \\ &\text{for} \quad A > 0 \end{aligned} \tag{113}$$

Equations (112) and (113) allow one to obtain an analytical expression for the broadened gain function and reduces computation burden.

The gain is a strong function of the wavelength and the question of what wavelength should be used arises naturally in a simulation. A simple choice is to locate the wavelength corresponding to the maximum gain. However in some DFB lasers, the lasing wavelength is forced

to locate off the gain peak by the DFB grating. In such a case the simulator should be flexible enough to fix the working wavelength of the simulator.

Another important term is the spontaneous recombination which dominates in GaAs/AlGaAs and InGaAs/AlGaAs laser systems below lasing threshold. The expression for spontaneous emission can be obtained from the gain function through Einstein relation [16,17]. The result for the unbroadened gain spectrum is given by

$$r_{sp}^{qw}(E) = \sum_{i=j}\left(\frac{2\pi}{\hbar}\right)|H_{ij}|^2 f_j'(1-f_i')D(E)p_{ij}, \tag{114}$$

where D(E) is the optical mode density in the material which has a refractive index of $\overline{n}$, given by [16,17],

$$D(E) = \frac{\overline{n}^3 E^2}{\pi^2\hbar^3 c^3}. \tag{115}$$

Equation (114) can not be used directly in the continuity equations because the device experience the total recombination rate integrated over all possible frequencies. Therefore the integrated spontaneous rate should be used:

$$R_{sp}^{qw} = \int_0^\infty r_{sp}^{qw}(E)dE. \tag{116}$$

A quantum well responds to the optical frequency with interband transitions which is the dominant contribution to the gain or loss. The secondary but also important loss mechanisms are the free carrier absorption and/or intervalence band absorption which can be written as [28–30]:

$$\alpha_{qw} = -\alpha_{fn}n - \alpha_{fp}p, \tag{117}$$

Note that these terms do not contribute to the radiation in the optical frequencies and should not appear in the stimulated recombination term.

To summarize all the gain/loss terms should be simple to evaluate and should be expressed in terms of the variables used in the global Jacobian matrices. For a self-consistent 2D solution all recombinations terms should be included in the full Newton scheme.

6. An Example of 2D Simulation

As an example of 2D simulation, the turn-on characteristics of a strained multiple quantum well is briefly described here. For detailed examples of CW simulation (light-current characteristics) of other types of laser obtained from a 2D model described in this chapter the reader is referred to Refs. [1,25,26].

Ridge waveguide structure is used for the 2D cross section. 3 μm is used for the ridge width. The structure and mesh generated are shown in Figs. 6 and 7, respectively. The structure of the devices is as follows. The compressively strained device has 5 quantum wells of thickness 8.7 nm with composition of x = 0.2 and $y = 0.85$ (4% compressive strain). The barrier is lattice matched with $y = 0.568$. The spacing between wells is 20 nm and the total thickness of the well/barrier region is 263 nm. A single step of InP is used for the GRIN confinement.

The gain function is modeled with a Landsberg type of broadening and the gain spectrum is shown in Fig. 8 together with experimental results [31]. The agreement with experiment is very reasonable.

The current distribution and optical field distribution is shown in Figs. 9 and 10 respectively. The band diagrams of the laser in equilibrium and under high injection condition are shown in Figs. 11 and 12. The distribution of local optical gain at a 1D section before and after the laser is turned on are shown in Figs. 13 and 14. Here the local optical gain is assumed to be a constant (loss) in the barrier region for simplicity. The distributed nature of the local gain is evident from these results.

For the purpose of illustrating the importance of the 2D simulation, a corresponding 1D simulation is also performed. The difference in the turn-on delay is shown in Fig. 15. Note that the turn on delay can be simply understood as the time it takes for the well to fill up to the lasing threshold density under the constant applied current. The model implemented is capable of producing large amount of useful simulation data to help understand the operation of of a complex laser device and to help optimize its design.

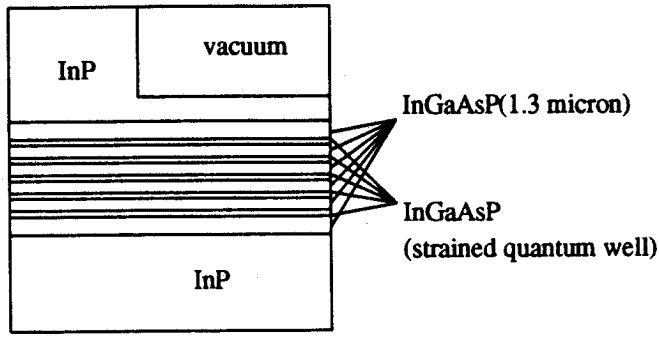


Figure 6. Schematics of a strained multiple quantum well laser.

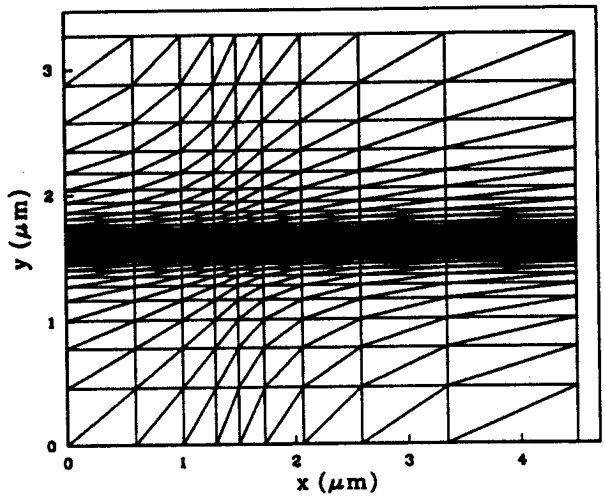


Figure 7. Mesh generated for the ridge waveguide laser.

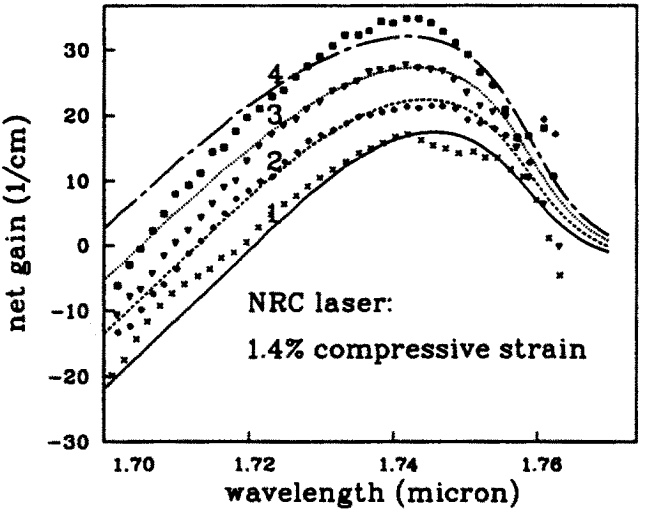


Figure 8. Experimental and theoretical gain spectrum for the MQW laser.

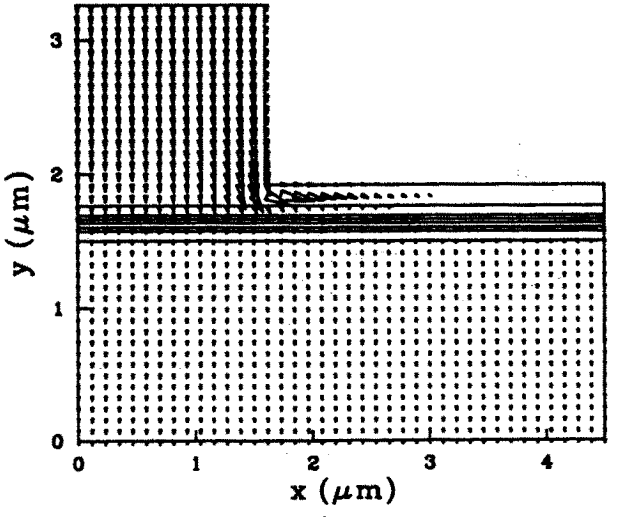


Figure 9. 2D current distribution.

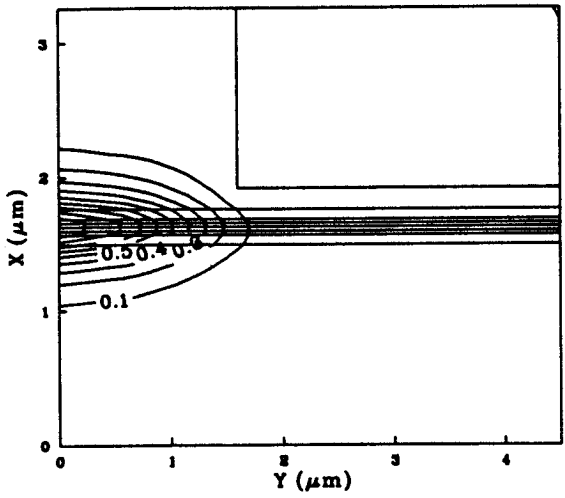


Figure 10. 2D optical field distribution.

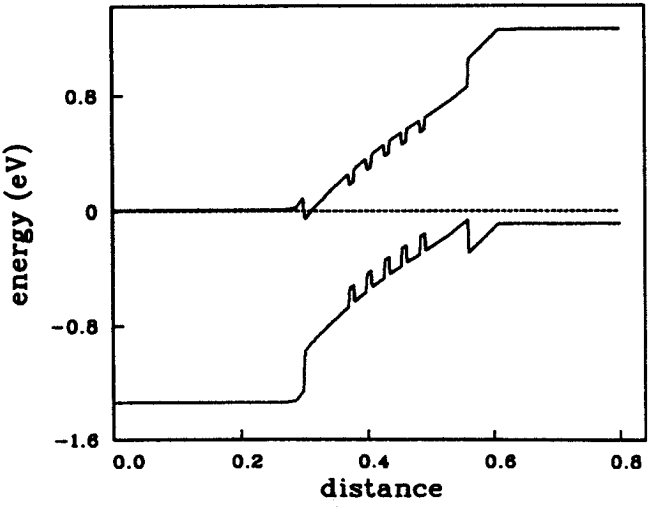


Figure 11. Band diagram at equilibrium.

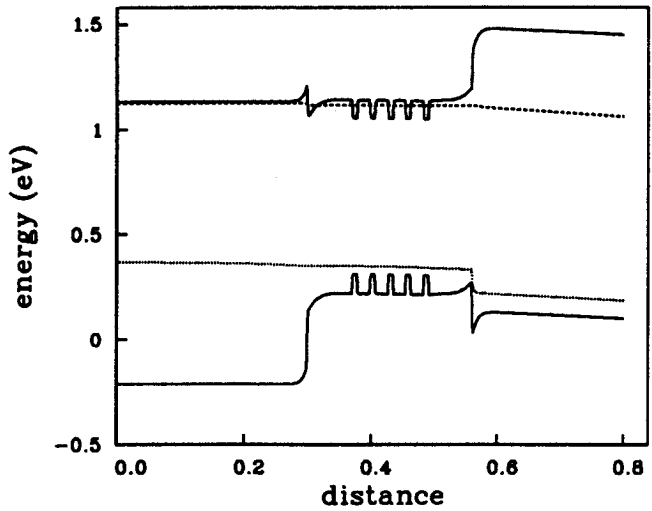


Figure 12. Band diagram under high current injection condition.

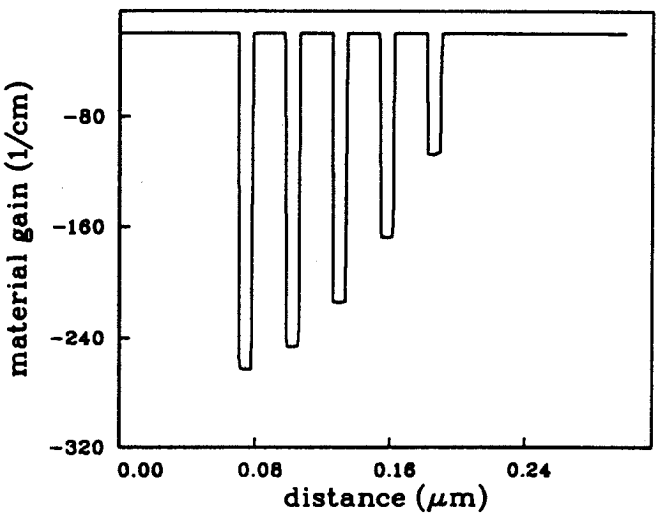


Figure 13. Material gain distribution before the laser turn-on.

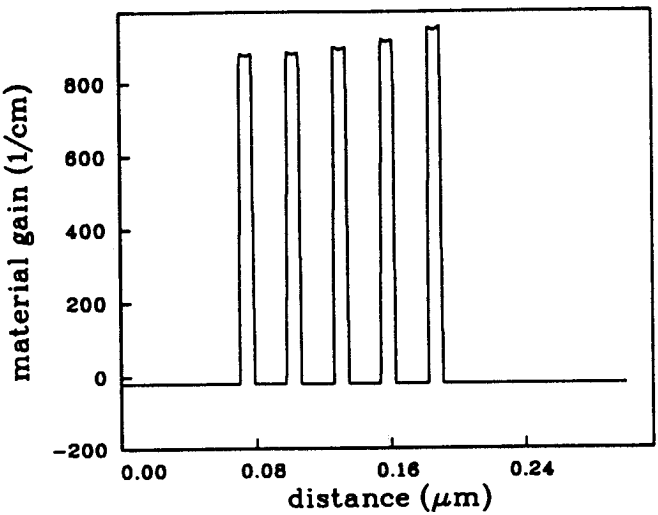


Figure 14. Material gain distribution after the laser turn-on.

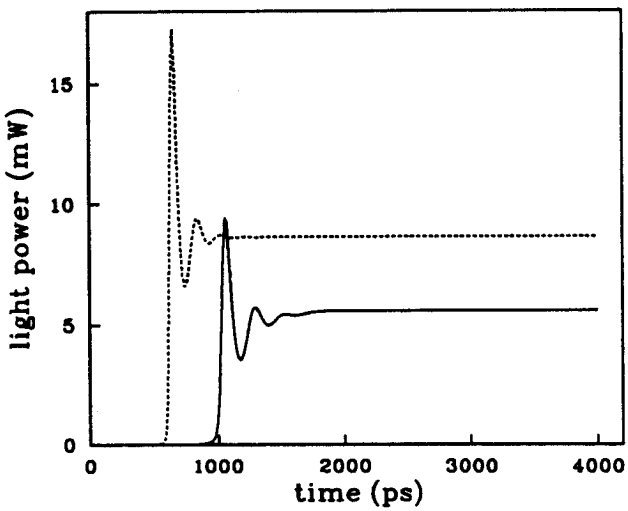


Figure 15. Turn-on characteristics of the laster. The solid line corresponds to the 2D simulation and the dash lines is the result of 1D model.

7. Summary

This chapter has described the key physical and numerical models necessary for the development of a 2D quantum well semiconductor laser simulator. A large part of the model is similar to the drift-diffusion model used in electronic devices. The major effort required is in combining the electronic device model with the optical model (i.e., rate equation, gain and wave equation) with efficiency and selfconsistency.

Getting the model developed and start producing reasonable results are only part of the effort in analysis and optimization of semiconductor lasers. Testing and calibration of the model for a specific device or process is also important in making useful predictions with the 2D model. It is hoped that the modeling technique described in this chapter helps the reader understand the basic principles of multiple dimensional semiconductor laser simulation.

Acknowledgment

The author wishes to thank Dr. S. P. MaAlister for managerial support of this work.

References

1. Li, Z.-M., K. M. Dzurko, A. Delage, and S. P. McAlister, "A self-consistent two-dimensional model of a GRIN-SCH SQW laser structure," *EEE J. Quantum Electron.,* Vol. 28, 792–802, 1992.
2. Song, G. H., K. Hess, T. Kerkhoven, and U. Ravaioli, "Two-dimen-
   sional simulation of quantum well lasers," *European Transactions on Telecommunicactions and Related Technologies,* Vol. 1, 375–381, 1990.
3. Ohtoshi, T., K. Y. Yamaguchi, C. N. Nagaoka, T. Uda, Y. Murayama, and N. Chinone, "A two-dimensional device simulator of semiconductor lasers," *Solid-State Electronics,* Vol. 30, 627–638, 1987.
4. Selberherr, S., *Analysis and Simulation of Semicondutor Devices,* Springer-Verlag, Wien, New York, 1984.
5. Pinto, M. R., C. S. Rafferty, and R. W. Dutton, *Pisces II User's Manual,* Stanford Electron. Labs., Stanford, CA, 1984.

6. Agrawal, G. P., and N. K. Dutta, *Long-Wavelength Semiconductor Lasers,* van Nostrand Reinhold, New York, 1986.
7. Sze, S. M., *Physics of Semiconductor Devices,* 2nd ed., John Wiley & Sons, 1981.
8. Smith, R. A., *Semiconductors,* 2nd ed., London, Cambridge University Press, 1978.
9. Bednarczyk, D., and J. Bednarczyk, *Phys. Lett.,* Vol. 64A, p. 409, 1978.
10. Varga, R. S., *Matrix Iterative Analysis,* Prentice-Hall, Engelwood Cliffs, NJ, 1962.
11. Scharfetter, D. L., and H. K. Gummel, "Large-singnal analysis of a silicon read diode oscillator," *IEEE Trans. Electron Devices,* Vol. ED-16, 64–77, 1969.
12. Li, Z.-M., S. P. McAlister, and C. M. Hurd, "Use of Fermi statistics in two-dimensional numerical simulation of heterojunction devices," *Semicond. Sci. Techn.,* Vol. 5, p. 408, 1990.
13. Son, I., and T.-W. Tang, "Modeling deep-level trap effects in GaAs MESFET'S', *IEEE Trans. Electron Devices,* Vol. 36, 632–640, 1989.
14. Blokhintsev, D. I., *Quantum Mechanics,* D. Reidel Publishing Company, Dordrecht-Holland, 1964.
15. Wooten, F., *Optical Properties of Solids,* Academic Press, New York and London, 1972.
16. Yan, R. H., S. W. Corzine, L. A. Coldren, and I. Suemune, "Corrections to the expression of gain in GaAs," *IEEE J. Quantum Electron.,* Vol. QE-26, 213–216, 1990.
17. Chinn, S. R., P. S. Zory, and A. R. Reisinger, "A model for GRIN-SCH-SQW diode lasers," *IEEE J. Quantum Electron.,* Vol. QE-24, 2191–2214, 1988.
18. Jackson, J. D., *Classical Electrodynamics,* 2nd ed., John Wiley & Sons, 1975.
19. Yamada, M., S. Ogita, M. Yamagishi, and K. Tabata, "Anisotropy and broadening of optical gain in a GaAs/AlGaAs multiquantum-well laser," *IEEE. J. Quantum Electron.,* Vol. QE-21, 640–645, 1985.
20. Landsberg, P. T., "A contribution to the theory of soft x-ray emmission bands of sodium," *Proc. Phys. Soc.,* Vol. A62, 806–816, 1949.

21. Landsberg, P. T., "Electron interaction effects on recombination spectra," *Phys. Status Solidi.*, Vol. 15, 623–626, 1966.
22. Martin, R. W., and H. L. Stormer, "On the low energy tail of the electron-hole drop recombination spectrum," *Solid State Comm.*, Vol. 22, 523–526, 1977.
23. Zielinski, E., H. Schweizer, S. Hausser, R. Stuber, M. H. Pilkuhn, and G. Weimann, "Systematics of laser operation in GaAs/-AlGaAs multiquantum well heterostructures," *J. Quantum Electron.*, Vol. QE-23, 969–976, 1987.
24. Chuang, S. L., "Efficient band-structure calculation of strained quantum-wells," *Phys. Rev. B,* Vol. 43, 9649–9661, 1991.
25. Li, Z.-M., M. Dion, S. P. McAlister, R. L. Williams, and G. C. Aers, "Incorporation of strain into a two-dimensional model of quantum well-well semiconductor lasers." to appear in Feb. 1993 issue of *IEEE J. Quantum Electron.*
26. Li, Z.-M., R. L. Williams, G. C. Aers, and S. P. McAlister, "A self-consistent two-dimensional model for strained InGaAs/AlGaAs quantum-well lasers," *Proceedings of the 1991 international semiconductor device research symposium,* December 4-6, 1991, Omni Charlottesville Hotel.
27. Li, Z.-M., K. M. Dzurko, and S. P. McAlister, "Two-dimensional modeling of quantum well semiconductor lasers," *COMPEL,* Vol. 10, 255–268, 1992.
28. Tsang, W. T., "Extremely low threshold, AlGa)As graded-index waveguide separate-confinement heterostructure lasers grown by molecular beam epitaxy," *Appl. Phys. Lett.*, Vol. 40, 217–219, 1982.
29. Fuchs, G., J. Horer, A. Hangleiter, V. Harle, and F. Scholz, R. W. Glew, L. Goldstein, "Intervalence band absorption in strained and unstrained InGaAs multiple quantum well structure," *Appl. Phys. Lett.*, Vol. 60, 231–233, 1992.
30. Huang, A., "Phonon-assisted intervalenceband absorption in semiconductor lasers," *Semicond. Sci. Technol.*, Vol. 5, 557–560, 1990.
31. Li, Z.-M., M. Davis, M. Dion, and S. P. McAlister "Material parameters for 2D modeling of quaternary quantum well lasers," *Proceedings of the 5th International Conf. on InP and Related Materials,* April 18-22, Paris, France, 1993.