A new finite difference scheme adapted to the one-dimensional Schrödinger equation

By Bernard J. Geurts[1], Philips Research Laboratories, P.O. Box 80 000, 5600 JA Eindhoven, The Netherlands

To the memory of my good friend Erik van Loon, taken from us so soon

1. Introduction

A number of different approaches exist for a direct numerical solution of the Schrödinger equation in one dimension. Most algorithms are hybrid in the sense that eigenfunctions and eigenvalues are determined separately and with different numerical approaches. Specially adapted (Numerov) schemes [1] exist for the eigenfunctions which require the eigenvalues to be known. Complementary (shooting) schemes using scaled Prüfer variables [2] find eigenvalues independent of the corresponding eigenfunctions. The combination of both approaches is needed for the full solution to the eigenproblem. Differentiability of the potential is generally an essential requirement. Techniques based on expansion of the eigenfunctions in a (plane wave) basis have also been employed [3, 4]. For applications with a (moderately) complex geometry a large number of basis functions must be included to cover the (wide) spread in length scales in such cases. Moreover, an error analysis for the eigenfunctions is virtually impossible. For complex geometries with possibly discontinuous potentials a more specialised treatment is required. Adopting a central space finite difference scheme as e.g. in ref. [5] or a finite element scheme [6] yields eigenfunctions and eigenvalues simultaneously but also implies a large number of mesh points in order to obtain high accuracy. Such approaches are also not well suited for discontinuous potentials. Since in many applications the potential entering the Schrödinger equation is to be determined selfconsistently, both high accuracy as well as a limited number of mesh points form essential requirements. In this paper we derive an adapted second order accurate finite difference

---

[1] Present address: Twente University, Department of Applied Mathematics, P.O. Box 217, 7500 AE Enschede, The Netherlands.

scheme for the Schrödinger equation based on analytic solutions to local linearisations. The resulting scheme is exact for piecewise constant potentials and masses. Both the *normalised* eigenfunctions as well as the corresponding eigenvalues are obtained simultaneously. The approximations made in the linearisations can be controlled directly and mesh-refinement can be employed iteratively. A consequence of the simultaneous determination of eigenvalues and (normalised) eigenfunctions is that the numerical problem to be solved is nonlinear whereas the Schrödinger equation itself is linear. Using a damped Newton iteration and continuation in the potential, results in a robust and flexible algorithm which applies to a great variety of problems.

Semiconducting heterostructures composed of a sequence of layers with different and finely tuned material properties have received considerable theoretical and experimental attention over the last years. With Molecular Beam Epitaxy (MBE) and other techniques, a detailed control over the heterostructure composition has become available and exciting new phenomena have been discovered and studied (e.g. the review by Ando et al. [7]). Most work has been concentrated on electron and hole gasses that are confined in one direction near one or more of the interfaces in the heterostructure [8]. Theoretically these problems have been studied mostly within the envelope function formalism. The simplest approximation in this class is formed by the Ben Daniel-Duke Hamiltonians. Both a position dependent piecewise constant effective mass as well as a discontinuous potential arise in these models. The direct numerical treatment of such Hamiltonians involves a scheme that takes these complications into account explicitly, in particular for cases in which a (large) number of widely different layers are incorporated in the structure.

In Section 2 we formulate the Ben Daniel-Duke Hamiltonians and prepare the derivation of the finite difference scheme given in Section 3. We also describe the full algorithm and show simulation results obtained for some simple model potentials. Section 4 is devoted to the Hartree approximation to the full many body problem and $Si/Si_x Ge_{1-x}$ heterojunctions will be studied numerically.

2. Sturm-Liouville form of the one dimensional Schrödinger equation

We introduce the Ben Daniel-Duke Hamiltonians in three dimensions and obtain the related one dimensional model problem for the eigenfunctions and eigenvalues in the growth direction of the heterostructure. Then we reformulate this Sturm-Liouville problem and obtain a system of first order equations governing the eigenfunctions, their normalisation as well as the corresponding eigenvalues. Some general properties of the eigenfunctions and eigenvalues will be discussed for use in the final algorithm.

A quantum mechanical description of properties of heterostructures is based on the Schrödinger equation. Rather than adopting a detailed 'crystal formulation' we use the envelope function formalism [9]. Specifically we consider Hamiltonians of the Ben Daniel-Duke type [10] and a potential depending on only one of the coordinates:

$$\left\{ -\frac{\hbar^2}{2m(z)}[\partial_{xx} + \partial_{yy}] - \frac{\hbar^2}{2}\partial_z\left(\frac{1}{m(z)}\partial_z\right) + U(z)\right\}\Psi(z, y, z) = E\Psi(x, y, z)$$

(2.1)

in which $\hbar$ is Planck's constant divided by 2π, $m(z)$ the position dependent effective mass, Ψ the wave function, E the energy and U the potential. For convenience we scale all variables, i.e. $x \to x/\lambda$ etc. in which λ is a reference length scale. Furthermore, we introduce

$$p(z) = \frac{m_0}{m(z)}\,; \quad \mathscr{E} = \frac{\hbar^2}{2|m_0|\lambda^2}$$

(2.2)

in which m_0 is a reference mass which may be positive (in case we consider electrons) or negative (in case we consider holes). Moreover, $p(z) > 0$ for all z and $\mathscr{E}$ is the energy scale. The problem as stated in eqn. (2.1) describes 'free' particles in the (x, y)-directions with a z-dependent effective mass and a nontrivial problem in the z-direction sensing the potential U. With the frequently used Ansatz

$$\Psi(x, y, z) \sim e^{i(k_x x + k_y y)}f(z)$$

(2.3)

we obtain a solution to (2.1) if f satisfies

$$\left\{ -d_z(p(z)d_z) + \text{sign}(m_0)U(z)\right\}f(z) = \left\{\text{sign}(m_0)E - p(z)k_\perp^2\right\}f(z)$$

(2.4)

in which $k_\perp^2 = k_x^2 + k_y^2$ and we incorporated the energy scale $\mathscr{E}$ into the definition of U and E. Thus the eigenproblem for f is coupled to the state in the (x, y)-directions, specified by $k_\perp^2$. At $k_\perp^2$ given, this problem is fully specified and an eigenvalue spectrum $\{E_l(k_\perp^2)\}$; $l = 0, 1, , \ldots,$ can be determined together with the corresponding eigenfunctions. In most cases this coupling is quite small [11] and the relevant values for $k_\perp$ are contained in a small neighborhood around $k_{\perp,\text{ref}}$ which may be assumed equal to zero without loss of generality. The eigenvalue spectrum can then quite accurately be approximated by $E_l(k_\perp^2) = E_l(0) + \text{sign}(m_0)p_l k_\perp^2$, i.e. a parabolic band centered on $E_l(0)$. Various levels of approximation exist for $\{p_l\}$ [8], but we will restrict ourselves to the simplest case $p_l = 1$ for all l. The full problem is then determined by the eigenproblem at $k_\perp = 0$ which has the form

$$\left\{ -d_z(p(z)d_z) + W(z)\right\}f(z) = v_z f(z)$$

(2.5)

in which we put $W(z) = \text{sign}(m_0)U(z) - W_{z,-} \geq 0$ for all z and $v_z = \text{sign}(m_0)E(0) - W_{z,-}$ where $W_{z,-} = \min_z \text{sign}(m_0)U(z)$. Notice that at any given $k_\perp$ the resulting problem (2.4) also has this structure, with slight redefinitions of the potential. We require the eigenfunctions to be normalised, i.e. $\|f\| = 1$ and consider the total system to be put in a large box with infinite walls yielding $f(L_z^{(-)}) = f(L_z^{(+)}) = 0$ where $L_z^{(\pm)}$ denotes the z-positions of the boundaries. Thus we arrive at a standard Sturm-Liouville problem. The eigenfunctions and eigenvalues have a number of important properties, of which we will mention two for later use in the numerical algorithm. First, the eigenvalues are non-degenerate and positive; $0 \leq v_{z,0} < v_{z,1} < v_{z,2} < \cdots$. Second, the eigenfunctions have a number of *distinct* real roots as a function of z; in particular the eigenfunction f_l has exactly l distinct roots between $L_z^{(-)}$ and $L_z^{(+)}$.

In order to formulate an algorithm which determines the *normalised* eigenfunctions and the corresponding eigenvalues simultaneously, we introduce

$$u_1 = f; \quad u_2 = pd_z f; \quad u_3 = v_z; \quad u_4 = \int_{L_z^{(-)}}^{z} f^2 \, dz \qquad (2.6)$$

with which the full problem can be written as

$$d_z u_1(z) = \frac{1}{p(z)} u_2(z)$$

$$d_z u_2(z) = [W(z) - u_3(z)]u_1(z)$$

$$d_z u_3(z) = 0 \qquad (2.7)$$

$$d_z u_4(z) = u_1^2(z)$$

and the corresponding boundary conditions are

$$u_1(L_z^{(-)}) = u_1(L_z^{(+)}) = u_4(L_z^{(-)}) = u_4(L_z^{(+)}) - 1 = 0. \qquad (2.8)$$

Furthermore, one may readily show that all solution components u_j are at least continuous for all z, even in case W and/or p show discontinuities. Notice that the above problem is nonlinear. In spite of the fact that the Schrödinger equation is linear in Ψ, the simultaneous determination of the eigenfunctions and the eigenvalues as well as the normalisation of the eigenfunctions renders the total problem nonlinear. Thus an iterative approach is required in solving this problem with which all quantities of interest can be obtained in a single consistent scheme. The numerical formulation will be subject of the next section.

3. Derivation of the finite difference scheme

The Sturm-Liouville problem formulated in the previous section will be discretised on a mesh. Local problems obtained by approximating the potential as constant in between two mesh points will be introduced and treated analytically. The corresponding solutions will be used to derive a finite difference scheme which is exact for these solutions. Thus exponential and oscillatory behavior in the solution components is incorporated explicitly. The treatment of discontinuities in the potential and the effective mass will be discussed in detail and the full algorithm is described. Finally, some simulation results based on this algorithm are shown and the performance of the full scheme investigated.

Let $\{z_j\}$ denote a mesh on $[L_z^{(-)}, L_z^{(+)}]$ and put $u_{i,j} = u_i(z_j)$ etc. The most general two point scheme consistent with (2.7) for intervals in which p and W are continuous can be written as

$$u_{1,j} - u_{1,j-1} - (c_{1,j}u_{2,j} + d_{1,j}u_{2,j-1}) = 0$$

$$u_{2,j} - u_{2,j-1} - (c_{2,j}u_{1,j} + d_{2,j}u_{1,j-1}) = 0$$

$$u_{3,j} - u_{3,j-1} = 0 \tag{3.1}$$

$$u_{4,j} - u_{4,j-1} - (c_{4,j}u_{1,j}^2 + d_{4,j}u_{1,j-1}^2 + e_{4,j}u_{1,j}u_{1,j-1}) = 0$$

in which consistency demands that the 'mesh constants' satisfy: $p_{j-1/2}(c_{1,j} + d_{1,j})$ and $c_{4,j} + d_{4,j} + e_{4,j}$ are asymptotically equal to $z_j - z_{j-1}$ as well as $c_{2,j} + d_{2,j} \sim (W - u_3)(z_j - z_{j-1})$ in the limit $z_j \to z_{j-1}$ in which we put $p_{j-1/2} = (p(z_j) + p(z_{j-1}))/2$. We consider the first two equations in (2.7) separately and introduce $\alpha(z) = (W(z) - u_3(z))/p(z)$ assuming u_3 to be known for the moment. Approximating $\alpha(z)$ by $\alpha_{j-1/2} = (\alpha_j + \alpha_{j-1})/2$ which is second order accurate in between z_{j-1} and z_j we can readily solve for u_1 and u_2 on this interval. We obtain

$$u_1(z) = a_1 e^{\sqrt{\alpha_{j-1/2}}z} + a_2 e^{-\sqrt{\alpha_{j-1/2}}z}; \quad \alpha_{j-1/2} > 0$$

$$= a_1 \cos(\sqrt{|\alpha_{j-1/2}|}z) + a_2 \sin(\sqrt{|\alpha_{j-1/2}|}z); \quad \alpha_{j-1/2} < 0 \tag{3.2}$$

and similarly for u_2

$$u_2(z) = p_{j-1/2}\sqrt{\alpha_{j-1/2}}a_1 e^{\sqrt{\alpha_{j-1/2}}z} - p_{j-1/2}\sqrt{\alpha_{j-1/2}}a_2 e^{-\sqrt{\alpha_{j-1/2}}z}; \quad \alpha_{j-1/2} > 0$$

$$= -p_{j-1/2}\sqrt{|\alpha_{j-1/2}|}a_1 \sin(\sqrt{|\alpha_{j-1/2}|}z)$$

$$+ p_{j-1/2}\sqrt{|\alpha_{j-1/2}|}a_2 \cos(\sqrt{|\alpha_{j-1/2}|}z); \quad \alpha_{j-1/2} < 0. \tag{3.3}$$

We require the discretisation (3.1) to be exact for these solutions. Insertion into the first two equations in (3.1) yields $c_{1,j} = d_{1,j}$ and $c_{2,j} = d_{2,j}$ implying

second order accuracy. Moreover, after some algebra we find explicitly

$$c_{1,j} = \frac{1}{p_{j-1/2}\sqrt{\alpha_{j-1/2}}} \tanh(\Delta_j/2); \quad \alpha_{j-1/2} > 0$$

$$= \frac{1}{p_{j-1/2}\sqrt{|\alpha_{j-1/2}|}} \tan(\Delta_j/2); \quad \alpha_{j-1/2} < 0 \tag{3.4}$$

in which we put $\Delta_j = \sqrt{|\alpha_{j-1/2}|}(z_j - z_{j-1})$. For the second equation we obtain in a similar way $c_{2,j} = p_{j-1/2}^2 \alpha_{j-1/2} c_{1,j}$. The last equation in (3.1) is somewhat more complicated. One may use (3.2) to integrate u_4 over the interval $[z_{j-1}, z_j]$, recalling the definitions in (2.6). Collecting like terms, i.e. those proportional to a_1^2, a_2^2 and $a_1 a_2$, we find $c_{4,j} = d_{4,j}$ and in full detail

$$c_{4,j} = \frac{1}{2\sqrt{\alpha_{j-1/2}}} \frac{\sinh(2\Delta_j) - 2\Delta_j}{\cosh(2\Delta_j) - 1}; \quad \alpha_{j-1/2} > 0$$

$$= \frac{1}{2\sqrt{|\alpha_{j-1/2}|}} \frac{\sin(2\Delta_j) - 2\Delta_j}{\cos(2\Delta_j) - 1}; \quad \alpha_{j-1/2} < 0 \tag{3.5}$$

and

$$e_{4,j} = \frac{1}{\cosh(\Delta_j)}(z_j - z_{j-1} - 2c_{4,j}) \tag{3.6}$$

for $\alpha_{j-1/2} > 0$. A similar expression with cosh replaced by cos is found in case $\alpha_{j-1/2} < 0$. As can be gathered from the expressions in (3.4–6) the values of Δ_j are restricted in case $\alpha_{j-1/2} < 0$. In particular, $\Delta_j < (\pi/2)$ in order to avoid singularities in the mesh-constants appearing in (3.1). In case $\alpha_{j-1/2} > 0$ there is no such restriction. Thus, oscillatory behavior in the solutions locally, requires a certain minimal resolution, related to mesh spacing as well as the value of the eigenvalues relative to the potential, within this scheme.

We next consider the corresponding expressions in case $\Delta_j \to 0$ and establish consistency of the above scheme as well as the connection of both the $\alpha_{j-1/2} > 0$ and < 0 regions. One readily derives using small argument expansions of the relevant functions [12]

$$c_{1,j} = \frac{1}{2p_{j-1/2}}(z_j - z_{j-1})\left[1 \pm \frac{\Delta_j^2}{6}\left(1 \pm \frac{\Delta_j^2}{10}\right)\right] \tag{3.7}$$

where $\alpha_{j-1/2} > 0$ corresponds to $(-)$ and $\alpha_{j-1/2} < 0$ to $(+)$. For the mesh-constants in the last equation in (3.1) we find

$$c_{4,j} = \frac{1}{3}(z_j - z_{j-1})\frac{1 \pm (\Delta_j^2/5)}{1 \pm (\Delta_j^2/3)} \tag{3.8}$$

where now $(+)$ corresponds to $\alpha_{j-1/2} > 0$ and $(-)$ to $\alpha_{j-1/2} < 0$. Thus, in the extreme limit $c_{1,j} \sim (z_j - z_{j-1})/2p_{j-1/2}$, $c_{2,j} \sim p_{j-1/2}\alpha_{j-1/2}(z_j - z_{j-1})/2$

and $c_{4,j} = e_{4,j} \sim (z_j - z_{j-1})/3$ which directly shows the consistency of the scheme. It is also obvious that the regions in which $\alpha_{j-1/2}$ changes sign are smoothly connected through these expansions. For Δ_j not approaching this continuum limit the scheme represents finite mesh corrections in accordance with the analytic solution for constant potential. We hence obtain a second order accurate, consistent discretisation scheme which is exact for regions in which the potential and the effective mass are constant. In most problems this is certainly not the case. Then, we may construct (iteratively) a mesh such that the 'constant potential' approximation is locally valid and useful, automatically increasing the resolution in case the local variations in the potential are large. We return to this momentarily and consider the exact treatment of discontinuities in W and/or p next.

If W and/or p are discontinuous, the above derivation is no longer valid. However, if a discontinuity arises at some z^*, one may integrate the equations (2.7) locally around z^* and obtain 'jump-conditions' for the solution components. It is readily shown that for all solution components $u_i(z^*_-) = u_i(z^*_+)$ where $z^*_\pm$ denotes the limit from above (below) to z^*. The numerical implementation which represents this property exactly can be obtained by introducing a double mesh point at the position z^* and likewise a double set of unknowns, i.e. $u_{i,j}^\pm$. The 'left' unknowns $(u_{i,j}^-)$ are coupled to $u_{i,j-1}$ and likewise the 'right' unknowns $(u_{i,j}^+)$ are coupled to $u_{i,j+1}$ through the 'normal' scheme (3.1). There is no ambiguity as to the value of $\alpha_{j\pm1/2}$ to be used in these intervals. The coupling between the 'left' and 'right' unknowns at z^* is given by the continuity requirement of all solution components. In this way some extra 'bookkeeping' is introduced into the scheme and the positions of the discontinuities need to be known explicitly in the definition of a particular problem.

The total discretisation implies a system of coupled, nonlinear algebraic equations for the unknowns $\{u_{i,j}\}$ on a mesh $\{z_j\}$. This can be solved with a damped Newton algorithm. In fact, by introducing $\mathcal{U} = [u_{1,0}, u_{4,0}, u_{2,0}, u_{3,0}, \ldots, u_{1,j}, u_{2,j}, u_{3,j}, u_{4,j}, \ldots, u_{2,N}, u_{3,N}, u_{1,N}, u_{4,N}]$, where N denotes the total number of mesh intervals, the total system of equations can be expressed as $\mathcal{F}(\mathcal{U}) = 0$ with a suitable $\mathcal{F}$ representing the discretisation (3.1) and the boundary conditions (2.6). With $\mathcal{U}^{(k)}$ denoting the k-th iterand in the Newton scheme, the damped Newton iteration can be written as

$$[\mathcal{D}_\mathcal{U}\mathcal{F}(\mathcal{U}^{(k)})]\xi^{(k)} = \mathcal{F}(\mathcal{U}^{(k)})$$
$$\mathcal{U}^{(k+1)} = \mathcal{M}(\mathcal{U}^{(k)}, \xi^{(k)}) \tag{3.9}$$

in which $\mathcal{D}_\mathcal{U}\mathcal{F}$ is the Jacobian matrix, $\xi^{(k)}$ the 'pure' Newton update for the solution and $\mathcal{M}$ the damping and restriction operator. If no damping or restriction would be applied we would have $\mathcal{U}^{(k+1)} = \mathcal{U}^{(k)} - \xi^{(k)}$. However,

the domain of convergence for this iteration scheme can be extended by including simple damping and restriction on the updates $\zeta^{(k)}$. In fact $0 \le u_4 \le 1$ so any update taking u_4 outside this region is restricted. Likewise, since the eigenvalues are distinct and ordered, we may have a good estimate for an interval containing u_3 and updates are restricted to this interval. Damping is also included in a simple way. Updates larger than some values, relative to a characteristic scale of the particular solution component, are limited to some maximum. In this way the convergence behavior of the Newton scheme is enhanced considerably.

The Newton update equations (3.9) have a particularly simple 'band' structure since we formulated a two-point discretisation rule. In fact

$$
\mathscr{D}_{\mathscr{U}}\mathscr{F} =
\begin{bmatrix}
I & & & & & \\
 & S_{1,0} & & & 0 & \\
 & & \ddots & & & \\
 & & & S_{j,j-1} & & \\
 & 0 & & & \ddots & \\
 & & & & & S_{N,N-1} & \\
 & & & & & & I
\end{bmatrix}
\tag{3.10}
$$

where I denotes a 2×2 unit matrix and $S_{j,j-1}$ a 4×8 matrix, given by the Jacobian of the left hand side of (3.1). The matrices $S_{1,0}$ and $S_{N,N-1}$ are obtained from the general form of $S_{j,j-1}$ through a permutation of some columns, consistent with $\mathscr{U}$ which shows different definitions for the first and last mesh points compared to the other mesh points. Derivatives of the mesh-constants with respect to the unknowns $\{u_{i,j}\}$ are required in setting up the $S_{j,j-1}$ which involves some extra calculation. No essential difficulties are encountered and hence these expressions will not be given. The simple structure of $\mathscr{D}_{\mathscr{U}}\mathscr{F}$ implies that a direct solution method can be formulated, thus enhancing the accuracy and speed. The full iteration typically converges in less than 10 steps, with damping and restriction included.

Continuation in the potential W is required in some cases. Since the eigenproblem has many different eigensolutions, and the full problem is nonlinear, it may happen that some of the different estimated eigenfunctions in an initial guess converge to one and the same eigenfunction for the problem with potential W. Whether such an event occurred can easily be checked by determining the number of roots of the solutions generated. This can be resolved by approaching the eigenproblem with potential W through a sequence of intermediate problems with potentials

$$
W_a = (1 - a)W_{\text{ref}} + aW
\tag{3.11}
$$

in which the continuation parameter a is increased from 0 to 1 and W_{ref} is a reference potential for which the full solution is known. For example, if W is a fixed potential we may take $W_{\text{ref}} = 0$ and as an initial guess the analytic

solution for a particle in a box. A step in a may be rejected if convergence fails or if the converged solutions do not have the proper number of roots. Also if $\Delta_j > (\pi/2)$ a smaller step in a is required. The number of steps in a may vary considerably. If the initial guess is adequate, a single step is needed but if the guess corresponding to W_{ref} is widely different from the solution at potential W many more steps (up to ≈ 20) may be needed. Typically, if the length scales in W_{ref} differ significantly from W a larger number of intermediate a-values is needed. The sequence of a-values is generated as follows. We start with $a = 1$ and on rejection divide a by a 2. This is repeated until a step at some $\tilde{a}$ is accepted and we proceed with doubling the increments in a whenever possible. A sequence of rejection and acceptance of intermediate solutions follows, gradually taking $a \to 1$. Usually, the last few steps in the sequence are all accepted consecutively and the total number of intermediate potentials is mostly limited.

The full algorithm can now be summarised as follows. Given a solution at potential W_{ref}, set up the Newton update equations and perform iterations until convergence at potential W. If convergence fails or the number of roots in the solutions is not adequate, take a smaller step in a, i.e. solve an intermediate problem 'closer' to W_{ref}. On proper convergence, check if the mesh spacings are adequate. So check if $\Delta_j \leq \varepsilon_s(\pi/2)$ in the relevant region $(\alpha_{j-1/2} < 0)$, and if the relative mesh variation of the potential is within some tolerance. Here ε_s denotes the 'mesh-safety' parameter introduced to prevent $\Delta_j > (\pi/2)$. Typically ε_s was taken equal to $1/2$. If the mesh is found too coarse locally we insert new points $z_{j-1/2} = (z_j + z_{j-1})/2$ and repeat the calculation at a fixed a. This can be done for each of the eigensolutions separately; each eigenproblem may be put on its own mesh. The above is repeated until $a = 1$ in which case we have obtained the desired solutions and automatically refined the mesh depending on the actual solution. Thus the linearisation errors can be controlled directly.

We next describe the operation of this algorithm using some simple test problems. We restrict ourselves to piecewise constant potentials and effective masses in order to obtain some insight in the physics of 'bare' heterostructures, i.e. cases in which we disregard many body effects. The latter will be incorporated in the next section, using the Hartree approximation. First we consider effects related to discontinuities in the effective mass. Then we show eigenfunctions for some potential well problems.

In Fig. 1 we plotted the ground state eigenfunction at zero potential, showing discontinuities in the derivative at the positions where m is discontinuous. We consider electrons (i.e. $\text{sign}(m_0) = 1$) for the moment. In this example we divided the total structure into two different regions. For $|z|$ smaller than some value (in the Figure 250 Å) we used a mass m_1 and put the mass equal to m_2 elsewhere. For the eigenfunctions only the ratio m_1/m_2 is relevant. The derivative shows a definite discontinuity; $d_z f(z_+^*) =$
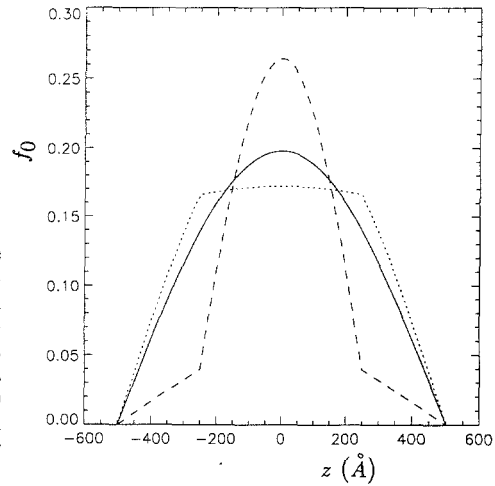
Figure 1
Ground state eigenfunction, showing the influence of discontinuities in the effective mass at zero potential. We used a structure with a total length of 1000 Å and show results for a symmetric effective mass distribution. For $|z| < 250$ Å the mass is the same as in the outer region (full curve), or it is 0.1 times that mass (dotted curve) or taken 10 times that mass (dashed curve). We used $m_0 = 0.1 m_e$ the electron rest mass, and an energy scale $\mathscr{E} = 0.1$ eV.

$m(z^*_+) d_z f(z^*_-)/m(z^*_-)$, which is easily recognised in the Figure. If $m_1/m_2 > 1$ the electrons get concentrated more in the inner region, whereas $m_1/m_2 < 1$ implies a flatter and lower eigenfunction in this region. If $m_1 > m_2$ the 'kinetic' energy is lower for the inner region, thus favoring electrons to move to the corresponding high mass segment of the system. The scheme represents these effects exactly; at the discontinuities the unknowns are doubly defined.

Figure 2 shows the first two eigenfunctions for a symmetric well, as well as the mesh-points used in the calculation. As initial mesh-points we used all discontinuity points of the potential plus all midway points. Remeshing was performed and the final mesh is consistent with $\Delta_j < \varepsilon_s (\pi/2)$ in which ε_s was put equal to $1/2$. Notice that with this setting, roughly 5 points are needed
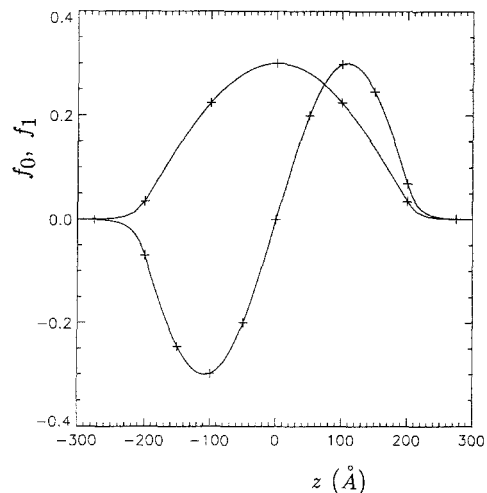
Figure 2
Ground state and first excited state for a symmetric well of width 400 Å and 0.15 eV depth in a structure of total length 1000 Å, with the same scale parameters as in Fig. 1. The mesh points used in the calculation are shown by 'crosses', using a mesh-safety parameter of 0.5. Reconstruction of the eigenfunctions was applied for the Figure.

per 'oscillation' in the eigenfunctions. In the regions where $\alpha_{j-1/2} > 0$ only 3 points are needed in this case. After completion of the calculation the solution was reconstructed by matching (3.2–3) to the numerical solution obtained on the coarse mesh. This reconstruction was used for plotting. We also studied linearly and quadratically varying potentials and obtained high accuracy results with a very limited number of mesh points. Typically up to 50 points are required for $\mathcal{O}(10^{-8})$ accuracy in the eigenvalues, for the first few eigenfunctions. The computational effort is reduced considerably compared to other numerical approaches. Figure 3 displays the first three eigenfunctions for a more complicated well structure. Notice that higher excited states have longer tails into barrier type material. Such structures can be analysed using the new scheme with great ease and accuracy; some seconds are required in order to obtain fully converged results on a common workstation. Moreover, potentials which are not piecewise constant but vary smoothly yield similar results; a limited number of mesh points gives very accurate numerical predictions. The refinement of a given mesh, based on a direct control of the relative variation of the potential per mesh interval, automatically increases the numerical resolution in regions of rapid variation, maintaining relatively wide intervals in regions where the potential varies slowly and $\alpha > 0$. The asymmetry in the mesh-intervals due to such refinements does not lead to numerical problems.

4. Selfconsistent Hartree calculations for a SiGe heterojunction

We formulate the Hartree equations governing the eigenproblem in $Si_x Ge_{1-x}$ heterojunctions. The potential in this case is determined partly by
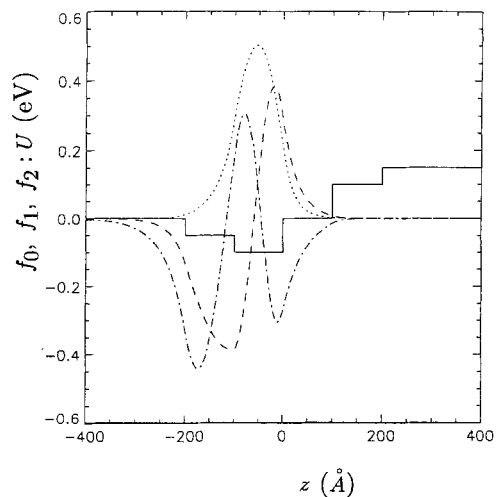

Figure 3
A more complicated well structure showing the first three eigenfunctions, with the same scale parameters as in Fig. 1. The potential in eV is shown as a full curve, the groundstate as a dotted curve, first excited state as a dashed curve and second excited state as a chain-dotted curve.

the electrostatic potential which is governed by Poisson's equation. In turn the electrostatic potential is fixed if the charge distribution is given. The latter quantity is related to the eigenfunctions and eigenvalues through Fermi-Dirac statistics. The full problem requires an iterative approach to obtain the selfconsistent solution.

In the Hartree approximation the many body problem governing the behavior of the electrons/holes in the structure is treated by the incorporation of the electrostatic potential with which the particles interact [8]. The potential entering the Schrödinger equation can be written as

$$U(z) = U_b(z) - G(z) \tag{4.1}$$

in which U_b denotes a background potential defined by the (flat) conduction (valence) band in case we consider electrons (holes); U_b is treated as a piecewise constant function. The other term in (4.1) represents the electrostatic potential energy; $G = (|e|\Phi)/\mathscr{E}$ where $|e|$ is the elementary charge and Φ the electrostatic potential. One readily shows that G is governed by

$$d_z(s(z)d_z G(z)) = -\gamma g(z) \tag{4.2}$$

in which $s(z)$ denotes the z-dependent relative permittivity of the structure measured in some typical scale for this quantity $\tilde{\varepsilon}_r$ and

$$\gamma = \frac{2|m_0||e|^2\lambda^4\tilde{\mathscr{N}}_0}{\varepsilon_0\tilde{\varepsilon}_r\hbar^2} \tag{4.3}$$

where $\tilde{\mathscr{N}}_0$ denotes a reference particle density and ε_0 the permittivity of the vacuum. Moreover, g denotes the charge distribution in units $\tilde{\mathscr{N}}_0$ in the structure. Including doping by donors and acceptors, this charge distribution can be written as

$$g(z) = n_d^+(z) - n_a^-(z) - \text{sign}(m_0)n_S(z) \tag{4.4}$$

where n_d^+ and n_a^- are the scaled ionised donor and acceptor concentrations respectively. The charge distribution related to the Schrödinger solution (n_S) obeys Fermi-Dirac statistics. After an elementary calculation one may show that

$$n_S(z) = (\pi\beta\lambda^3\tilde{\mathscr{N}}_0)^{-1}\sum_{l=0}^{\infty}\ln(1+\alpha_l)|f_l(z)|^2 \tag{4.5}$$

in which $\beta = \mathscr{E}/(k_B T)$ with k_B Boltzmann's constant and T the temperature. The eigenvalue spectrum is incorporated into this expression through

$$\alpha_l = \exp[\beta(\text{sign}(m_0)\mu - v_{z,l})]. \tag{4.6}$$

So $\{\alpha_l\}$ depend on the eigenvalues as well as on the chemical potential μ. For the ionised donor and acceptor concentrations we adopt a simple model

(see e.g. Sze [13]) valid for 'bulk'-like situations. Thus

$$n_d^+(z) = n_d(z)\left[1 - \left\{1 + \frac{1}{2}\exp(t_d(z))\right\}^{-1}\right] \tag{4.7}$$

and

$$n_a^-(z) = n_a(z)[1 + 4\exp(t_a(z))]^{-1} \tag{4.8}$$

where n_d and n_a denote the 'chemical' donor and acceptor concentrations, put into the heterostructure in a controlled way during fabrication. The effect of degeneracy in the donor and acceptor levels has been included, viz. the factors $1/2$ and 4 appearing here. Strain in the Si/SiGe layers may remove this degeneracy. Finally,

$$t_d(z) = \beta\left\{U(z) - E_d(z) - \mu - \frac{1}{2}(\text{sign}(m_0) - 1)E_g(z)\right\} \tag{4.9}$$

$$t_a(z) = \beta\left\{U(z) + E_a(z) - \mu - \frac{1}{2}(\text{sign}(m_0) + 1)E_g(z)\right\} \tag{4.10}$$

in which E_d (E_a) denotes the donor (acceptor) binding energy measured relative to the conduction (valence) band edge and E_g denotes the energy gap between conduction and valence bands. In the calculations these functions are piecewise constant. At low temperatures, i.e. high β-values, the ionised doping concentrations vary rapidly in regions where $t_{a,d}$ has a root. In the limit $T \to 0$ the ionised profiles tend to step functions, i.e. full or no ionisation depending on the sign of $t_{a,d}$.

In order to complete the model, we briefly discuss the constraints and boundary conditions to be used in the calculation. The reference point for the potential energy is taken at $z = L_z^{(-)}$ and hence $G(L_z^{(-)}) = 0$. Furthermore, the total charge distribution is required neutral, that is the integral of g over the structure is zero. In particular this implies the following condition from which the chemical potential μ can be determined:

$$\sum_{l=0}^{\infty} \ln(1 + \alpha_l) = \text{sign}(m_0)\pi\beta\lambda^3\tilde{N}_0 \int_{L_z^{(-)}}^{L_z^{(+)}} dz(n_d^+(z) - n_a^-(z)) \tag{4.11}$$

where use is made of the normalisation of the eigenfunctions. Furthermore, we require the electric field, or more precisely $d_z G(L_z^{(\pm)})$ to be zero. Thus, the value of G at $z = L_z^{(+)}$ is determined iteratively such that the derivative of G is zero. Obviously, these equilibrium conditions can only be analysed if a full solution of the Schrödinger equation is available, which on the other hand itself is determined by the solution G. The coupled system of equations can only be treated iteratively in order to obtain the selfconsistent solution.

The numerical procedure adopted for solving the Schrödinger-Poisson problem can be summarised as follows. With $U = U_b$ the solution to the Schrödinger equation is determined. With the corresponding eigenfunctions, eigenvalues and potential the chemical potential μ is found from (4.11) and g specified. Solving the Poisson equation generates a solution G which is used to set up the new potential U with which the iteration proceeds. This process is repeated until two subsequent iterands for the potential differ less than some tolerance (typically $\mathcal{O}(10^{-6})$ relative error was found adequate). At high doping concentrations and/or low temperatures mixing of iterands to the final potential must be applied in order to damp the charge-oscillations. In such cases a small change in the potential may cause a large shift in the position(s) at which $t_{a,d}$ has a root, i.e. the position of the 'tail' in the ionised doping profiles depends sensitively on small variations in the potential-iterands. Typically the potential used in the next step in the iteration was built up of $\approx 10-25$ percent of the potential corresponding to the new G plus $\approx 75-90$ percent of the old potential used in the previous iteration. A large number of iterations is required in order to obtain full selfconsistency in the solution (typically ≈ 100 steps are required starting with the background potential as initial guess). In the initial stages of this iteration process continuation between the old and new potentials is required. Closer to convergence only one continuation step is needed. A full calculation takes up several minutes in a typical situation.

Before turning to a discussion of some simulation results based on the Hartree equation we sketch the numerical treatment of Poisson's equation (4.2) for piecewise constant $s(z)$. Introducing $v_1 = G$ and $v_2 = sd_zG$ we may write

$$d_z v_1 = v_2/s; \quad d_z v_2 = -\gamma g \tag{4.12}$$

On the interval $[z_{j-1}, z_j]$ we approximate g linearly, matching on g_{j-1} and g_j. The resulting system of equations approximating (4.12) on $[z_{j-1}, z_j]$ can be solved easily and a similar analysis as given in Section 3 for the Schrödinger equation yields for intervals on which s and g are continuous

$$v_{1,j} - v_{1,j-1} - \frac{1}{2s_j}(z_j - z_{j-1})\{v_{2,j} + v_{2,j-1}\} = \frac{\gamma}{12s_j}(z_j - z_{j-1})^2(g_j - g_{j-1})$$

$$v_{2,j} - v_{2,j-1} = -\frac{\gamma}{2}(z_j - z_{j-1})(g_j + g_{j-1}) \tag{413}$$

which is second order accurate and exact for linearly varying charge distributions. The boundary conditions imply $v_{1,0} = 0$ and $v_{1,N} = G_+$ in which G_+ must be adjusted such that $v_{2,0} = 0$. In regions where s and/or g are discontinuous we may again define the unknowns doubly and apply continuity of both v_1 and v_2. The mesh used contains all different points

used in the Schrödinger equation plus some additional points at discontinuities in g. In addition, remeshing is applied such that regions in which the derivative of g varies rapidly are treated with increased resolution. In particular regions close to the tail in the ionised doping profiles and near the interfaces at which the electron/hole gasses are found are treated with greater precision.

With this formulation, we can calculate various properties directly, as a function of temperature and material parameters. We apply this model to a simple p-Si/$Si_x Ge_{1-x}$ heterojunction as shown in Fig. 4. This structure consists of a highly doped p-Si layer (acceptor dopant Boron), a 'spacer' layer of intrinsic Silicon with a width d and an intrinsic layer of $Si_x Ge_{1-x}$ with x denoting the fraction of Si in the Ge-matrix. We also sketched the potential in order to make the connection with the ionised doping profile and introduced the valence band offset at $z = 0$ denoted by Δ. The intersection point of the chemical potential and the potential shifted by an amount E_a corresponds to the 'tail' in the ionised doping concentration. A detailed account of such devices can be found e.g. in ref. [14]. One easily recognizes the potential 'well' in which the 2DHG (two-dimensional hole gas) is bound; since the effective mass is negative the holes are bound in a region where the potential is maximal, rather than in a minimum as is the case for electrons. In order to obtain a general impression of the potential and the eigenfunctions we plotted the first two eigenfunctions and the potential for a typical case in Fig. 5. The ground state (dotted curve) is sharply bound and the distance over which the 2DHG spreads is $\approx 250\ \text{Å}$ for this case since the weight of the first excited state in the contribution to the total charge density is negligible here. Notice that the first excited state is unbound, roughly speaking. Typically at low temperatures the ground state generates the dominant contribution, higher excited states contribute significantly only at higher temperatures.
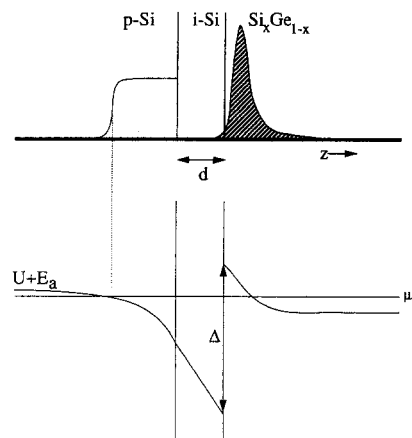

Figure 4
Model Si/$Si_x Ge_{1-x}$ structure used in the Hartree calculations. The ionised acceptor profile and the 'two dimensional' hole gas (shaded curve) are drawn. Below a sketch of the potential is given.

Figure 5
Plot of the first two eigenfunctions (ground state dotted, first excited state dashed) and the potential (full curve) for a sample SiGe heterojunction. We set the temperature $T = 10$ K, used a structure of total length 1000 Å, $m_0 = 0.32 m_e$ and $\mathscr{E} = 0.1$ eV. The reference particle density $\tilde{\mathscr{N}}_0 = 1 \times 10^{23}$ m^{-3} and the valence band offset at the interface $\Delta = 0.148$ eV. Only acceptor doping was taken into account; the doping is uniform for $z < 0$ and equal to $\tilde{\mathscr{N}}_0$. The acceptor binding energy used is $E_a = 0.03$ eV. The spacer width $d = 0$ and the weighing of old and new potentials was 10 percent. Finally, the scale for the relative permittivity $\tilde{\varepsilon}_r = 12.5$. These parameters correspond roughly to $x = 0.2$.
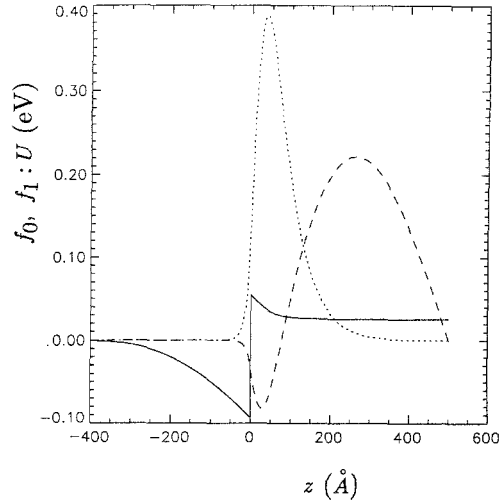
The temperature dependence of the 2DHG and the ionised acceptors is presented in Fig. 6. As the temperature is increased the ionised acceptor profile extends further into the p-Si layer and becomes smoother. At high temperatures the tail in the profile extends to the left hand side numerical boundary. The corresponding 2DHG profiles also become more extended and the area under these curves, corresponding to the sheet concentration increases rapidly as temperature increases. One observes that the first excited state becomes gradually more important as T increases and contributions of this state can easily be recognised as extra structure in the tail of the 2DHG in the high temperature results. The selfconsistent potential corresponding to these temperature results is shown in Fig. 7. Since more
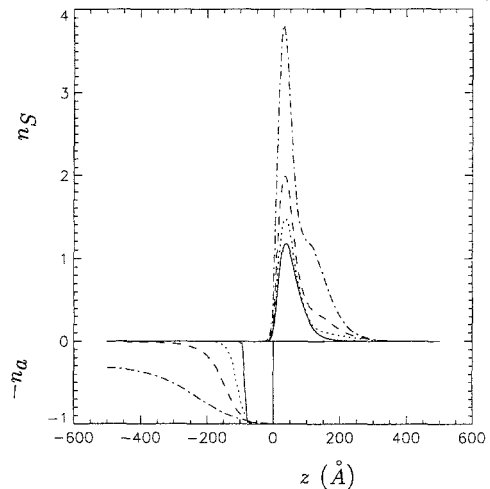
Figure 6
The ionised acceptor and hole density profile scaled with $\tilde{\mathscr{N}}_0$ as a function of temperature at the same parameters as in Fig. 5. For clarity the acceptor profile is drawn negative. Results are shown at $T = 10$ K (full curves); $T = 50$ K (dotted curve); $T = 100$ K (dashed curves) and $T = 250$ K (chain dotted curves). In these calculation only the first two eigenfunctions were included.
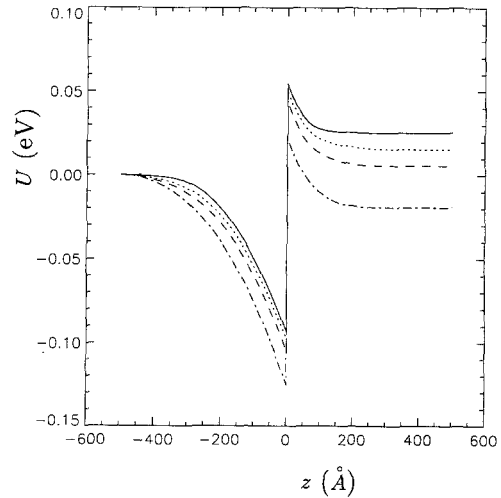
Figure 7
The selfconsistent potential as a function of temperature for the same cases as in Fig. 6 and similar curve labelling.

acceptors are ionised in the p-Si layer, the minimum shifts to lower values as temperature increases. The valence band offset Δ is however fixed and hence there is a maximum value for the temperature above which no 2DHG can be supported. One observes that the maximum in the potential decreases and at some T-value it becomes zero. At that point the model breaks down and a spurious potential well at the left hand boundary comes into existence. At higher doping concentrations the limiting temperature decreases quite rapidly.

An important property of the 2DHG is the sheet concentration which can be measured directly. This sheet concentration is defined as

$$N_s = (\pi \beta \lambda^2)^{-1} \sum_{l=0}^{\infty} \ln(1 + \alpha_l) \tag{4.14}$$

i.e. denotes the equivalent areal concentration of holes contained in the 2DHG profile. This quantity varies quite rapidly with temperature and also depends quite strongly on the value for the acceptor binding energy E_a used. In Fig. 8 we show N_s at two doping concentrations showing the temperature dependence as well as the influence of variations in the binding energy. An increase in the doping concentration results in an increase in the sheet concentration; at high doping concentrations a saturation comes into play, i.e. the increase in N_s is smaller than the increase in the doping concentration. Furthermore, increasing the binding energy implies a decrease in the sheet concentration. We also studied the effects related to the spacer layer. An increase in the width of the spacer corresponds to a sharp decrease in the sheet concentration. Typically a spacer is introduced to enhance mobility of holes in the (x, y)-directions, at the expense, however, of having a lower concentration of charge carriers.
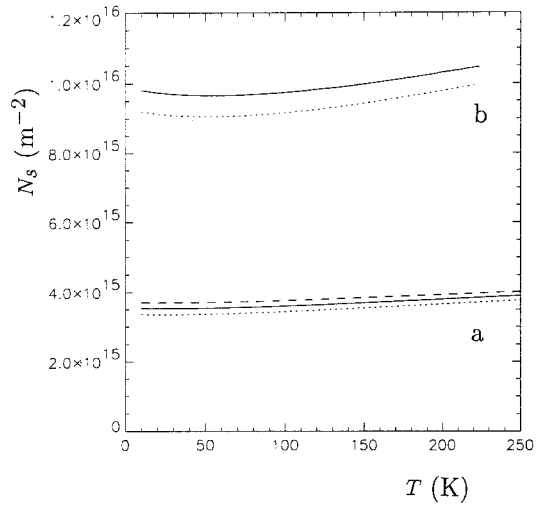
Figure 8
The sheet concentration as a function of temperature showing the influence of variations in the acceptor binding energy at two different acceptor-doping concentrations. Case a corresponds to a maximal doping of 1×10^{23} m^{-3} and case b to 1×10^{24} m^{-3}. The full curves denote $E_a = 0.03$ eV, the dotted curves correspond with $E_a = 0.04$ eV and the dashed curves represent $E_a = 0.02$ eV.

The new integration algorithm for the Schrödinger equation has proven to result in a flexible and robust tool with which various properties of a wide class of heterostructures can be analysed with minimal computational effort. A direct connection of these predictions with experimental results will be considered elsewhere and inclusion into a device simulation code is being investigated presently. Moreover, the influence of applied electric and magnetic fields can also be incorporated into the scheme in a rigorous, locally exact manner. These field effects and the inclusion of exchange and correlation into the potential are studied in delta-doped structures and more complicated MBE-grown well structures. C-V measurements and transport properties will be investigated in relation to the various levels of approximation. Preliminary results indicate a fair agreement of simulation results and experimental data.

References

[1] L. Gr. Ixaru and M. Rizea, J. Comp. Phys., *73*, 306 (1987).
[2] P. B. Bailey, J. SIAM Appl. Math., *14*, 242 (1966).
[3] T. Ando and S. Mori, J. Phys. Soc. Japan, *47*, 1518 (1979).
[4] T. Ando, J. Phys. Soc. Japan, *51*, 3900 (1982).
[5] I. H. Tan, G. L. Snider, L. D. Chang and E. L. Hu, J. Appl. Phys., *68*, 4071 (1990).
[6] K. Inoue, H. Sakaki, J. Yoshino and T. Hotta, J. Appl. Phys., *58*, 4277 (1985).
[7] T. Ando, A. B. Fowler and F. Stern, Rev. Mod. Phys., *54*, 437 (1982).
[8] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures*, Les éditions de physique 1988.
[9] M. F. H. Schuurmans and G. W.'t Hooft, Phys. Rev., *B31*, 8041 (1985).
[10] D. J. Ben Daniel and C. B. Duke, Phys. Rev., *98*, 368 (1955).

[11] T. Ando, J. Phys. Soc. Japan, *51*, 3893 (1982).
[12] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover 1972.
[13] S. M. Sze, *Physics of Semiconductor Devices*, Wiley 1981.
[14] R. People, IEEE, *QE-22*, 1696 (1986).

Abstract

We present a new discretisation scheme for the Schrödinger equation based on analytic solutions to local linearisations. The scheme generates the normalised eigenfunctions and eigenvalues simultaneously and is exact for piecewise constant potentials and effective masses. Highly accurate results can be obtained with a small number of mesh points and a robust and flexible algorithm using continuation techniques is derived. An application to the Hartree approximation for SiGe heterojunctions is discussed in which we solve the coupled Schrödinger-Poisson model problem selfconsistently.