Chapter 4

Numerical Aspects of Quantum Transport Simulation using the Wigner-Poisson Model

The equations of motion of the Wigner function discussed in chapter 3 can be compactly written as follows:

$$\frac{\partial f(q,k)}{\partial t} = \mathcal{L}f(q,k) \quad -\infty < q < +\infty, -\frac{\pi}{2\Delta} \le k \le \frac{\pi}{2\Delta} \tag{4.1}$$

where $\mathcal{L}$ is the Liouville superoperator and Δ is the spacing of the lattice on which the equation is solved. In a finite system $[0, L]$, boundary conditions are imposed and the equation takes the following form:

$$\frac{\partial f(q,k)}{\partial t} = \mathcal{L}f(q,k) + b \quad 0 < q < L \tag{4.2}$$

where b incorporates the boundary contribution. The boundaries of the simulation domain are placed in regions far from the active region of the device where the solution is known with a high degree of confidence. The boundary conditions are typically imposed a few thermal wavelengths away from the active region of the device so that interference effects vanish at the boundaries. The goal is to convert the integro-differential equation 4.2 into a set of linear algebraic equations and use direct methods to invert the resulting matrix equation.

An accurate numerical model is necessary to obtain useful knowledge. In the past, the issue of numerical stability has been considered the most important, the fidelity to the physical model receiving much less attention. This

is perhaps because the problems inherent with past approaches are not obvious in the commonly simulated AlGaAs/GaAs heterostructures. As will be demonstrated in this chapter, the standard upwind schemes for the drift term, chosen from numerical stability considerations, are completely inadequate in the study of AlAs/GaAs heterostructures. The tradeoff between numerical stability and physical accuracy makes the discretization of the Wigner equation of motion a major challenge. After a discussion of the inadequacy of the upwind schemes, differencing schemes that greatly improve accuracy while maintaining numerical stability are presented. We present, for the first time, the proper numerical treatment when the effective-mass changes with position in the heterostructure.

4.1 Discrete Wigner-Poisson model: general considerations

The equation of motion is reduced to a set of linear algebraic equations for the Wigner function f_{jn} on a discrete phase-space:

$$\{q_j | q_j = j\Delta; j = 1, 2, \ldots, N_q\}$$
$$\left\{k_n | k_n = \frac{\pi}{2N_k\Delta}(2n - 1 - N_k); n = 1, 2, \ldots, N_k\right\}$$
$$\mathcal{L}f = \sum_{j'n'} \mathcal{L}_{jn;j'n'} f_{j'n'} = -b_{jn} \qquad (4.3)$$

Δ is the mesh spacing in position. N_q and N_k are the number of position and momentum nodes respectively. $\mathcal{L}_{jn;j'n'}$ is the discrete Liouville superoperator. This structure for the discrete phase-space has been widely used (Buot & Jensen, 1990; Frensley, 1990; Klukshdahl et al., 1989; Mains & Haddad, 1988).

b_{jn} is the boundary contribution which depends on $\mathcal{L}$ and the boundary conditions specified for the Wigner function. At the boundaries of the simulation domain, the potential is constant and the Wigner function is the Fermi-Dirac distribution. So far, the discussion has been for transport in one dimension. However, the momentum space is two dimensional (assuming cylin-

drical symmetry about the direction of current flow). Integrating over the transverse momenta, we get

$$
\begin{aligned}
f(0,k) &= \frac{m^*(0)k_B T}{\pi \hbar^2} \ln \left\{ 1 + e^{-\frac{\mathcal{E}(0,k) - \mu_0}{k_B T}} \right\} \\
f(L,k) &= \frac{m^*(L)k_B T}{\pi \hbar^2} \ln \left\{ 1 + e^{-\frac{\mathcal{E}(L,k) - \mu_L}{k_B T}} \right\}
\end{aligned} \tag{4.4}
$$

As discussed in Section 3.2 the transverse degrees of freedom can be integrated out only if the Wigner function in the longitudinal momentum is not coupled to that in the transverse momentum.

How the boundary distribution couples into the simulation domain depends on $\mathcal{L}$ at the boundary. Typically, ohmic contact behavior is simulated at the boundaries. Electrons with an equilibrium distribution are injected into the device regardless of the behavior interior of the device. The distribution of the electrons exiting the device at the boundary is not specified. If the boundaries are far from the active region of the device, and if dissipation is included in the contact regions, the exiting distribution will also resemble the equilibrium form.

The choice of the points in k-space is very important. The point $k = 0$ cannot be included because $\mathcal{L}$ then becomes singular. The integrals in 3.21 can be accurately estimated using the discrete sine and cosine transforms. To understand the k-mesh spacing, consider the discrete Wigner function

$$
f_{jn} = \sum_{j'=-N_k/2}^{N_k/2} \rho_{j+j',j-j'} e^{-2\imath k_n j' \Delta} \tag{4.5}
$$

and the discrete potential kernel corresponding to $\mathcal{V}(q, k - k')$

$$
\mathcal{V}_{j,n-n'} = -\Im \left\{ \sum_{j'=j-N_k/2}^{j+N_k/2} v_{j'} e^{-2\imath(j-j')(n-n')\Delta\Delta_k} \right\} \tag{4.6}
$$

which are invertible only if the Fourier completeness condition $N_k \Delta \Delta_k = \pi$ is satisfied. This condition must also be met if the discrete transport model is to

satisfy current continuity. In the continuum, the integral of the potential kernel over k vanishes due to the odd symmetry. For the discrete potential kernel to vanish after summation (i.e., $\sum_n \mathcal{V}_{j,n-n'} = 0$), the Fourier completeness condition must be satisfied (Frensley, 1990). Only then will the equation of continuity hold in the discrete model.

An important point is that the use of the discrete Fourier transform introduces the necessary periodicity in momentum. While the $\mathcal{E}(k)$ dispersion of a free-electron is not periodic, in a crystal it is. The Wigner function, and its equation of motion must also possess this periodicity in momentum. The use of a discrete sine and cosine transforms assures this periodicity.

To understand the limits on the summation in Eq. 4.5, consider the density matrix of a free electron

$$\rho_{j+j',j-j'} = e^{-j'^2/\ell^2} \quad \ell^2 = \frac{\hbar^2}{2m^* k_B T \Delta^2} \tag{4.7}$$

where T is the temperature, m^* is the effective mass, and Δ is the lattice constant. The density matrix falls off rapidly with increasing $|j'|$. At room temperature $\ell \sim 9$ in GaAs. This sets a safe choice of $N_k = 6\ell \approx 54$. A larger N_k is typically used to improve the accuracy in estimating the electron density which is evaluated as the integral of the Wigner function in k-space. An important reason for choosing a larger value for N_k is that it reduces the dissipative effects inherent on a discrete phase-space. In the absence of any physical dissipation, electron trajectories are reversible. However, after coarse graining, trajectories are not reversible. This is demonstrated in Fig. 4.1. This phenomena is similar to arithmetic on a finite precision computer where reversing the operations does not always yield the starting values.
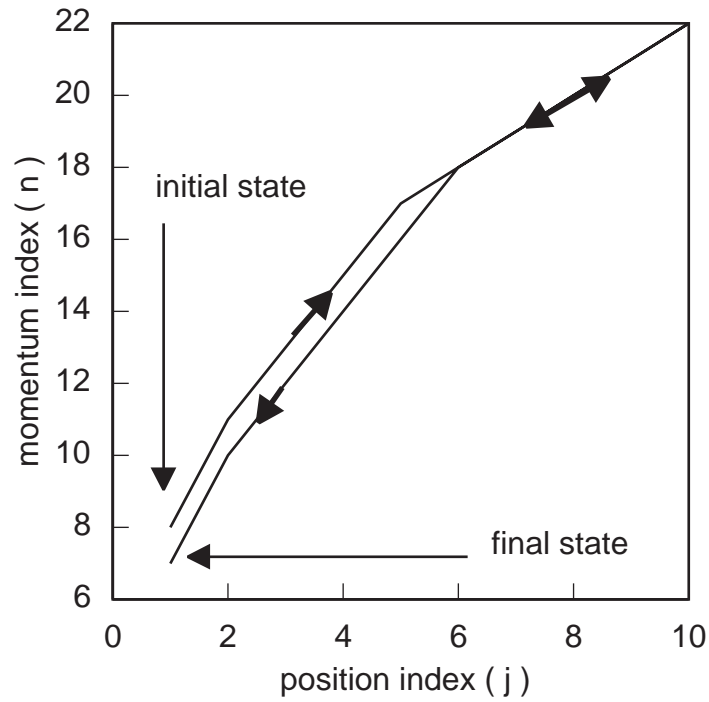
Figure 4.1: Trajectories on a discrete mesh are not reversible. Shown here is a trajectory and its time-reversed pair under the influence of a constant force: $\hbar^2 k \, dk/dx = m^* F$. $F = 1MV$, $N_k = 40$ $\Delta = 5.6533\text{Å}$ and $\Delta_k = \pi/(N_k \Delta)$. At each position the momentum is reassigned to the nearest node in $k - space$ given by $n = int[k/\Delta_k + 1/2]$.

4.2 Transport in a parabolic energy band

Before discussing the discretization of the various terms in the equation of motion, the term

$$\mathcal{T}f(q) = -\frac{\hbar k}{m^*_{GaAs}}\frac{\partial f}{\partial q}$$

is separated from Eq. 3.20. In the absence of any effective-mass variations in the device, $\mathcal{T}f(q)$ is the drift term. Discretization of the drift term presents a non-trivial problem and will be discussed separately. Ignoring the drift term for the moment, each term in Eq. 3.20 is approximated by using centered differences for the derivatives. Examining Eq. 3.33 it is clear that this is the most consistent discretization. The integrals in momentum are evaluated by expanding the distribution function in k-space using a delta-function basis ($f(q,k) = \Delta_k \delta(k - k_n)f_{jn}$).

The following discrete forms for the various terms in Eq. 3.20 are obtained:

$$\frac{1}{\pi}\int_{BZ} dk'\, k'\frac{\partial}{\partial q}\left[f(q,k')\int dr\left(\frac{1}{m^*(r)} - \frac{1}{m^*_{GaAs}}\right)\cos[2(k-k')(q-r)]\right]$$

$$\rightarrow \sum_{n'=1}^{N_k}\frac{k_{n'}}{N_k}\left(\frac{\mathcal{M}^e_{j+1n-n'}f_{j+1n'} - \mathcal{M}^e_{j-1n-n'}f_{j-1n'}}{2\Delta}\right) \quad (4.8)$$

$$\frac{1}{4\pi}\int_{BZ} dk'\frac{\partial^2}{\partial q^2}\left[f(q,k')\int dr\frac{\sin[2(k-k')(q-r)]}{m^*(r)}\right]$$

$$\rightarrow \sum_{n'=1}^{N_k}\left(\frac{\mathcal{M}^o_{j+1n-n'}f_{j+1n'} - 2\mathcal{M}^o_{jn-n'}f_{jn'} + \mathcal{M}^o_{j-1n-n'}f_{j-1n'}}{4N_k\Delta^2}\right) \quad (4.9)$$

$$\frac{2}{\pi}\int_{BZ} dk'\, f(q,k')\int dr\left(v(r) + \frac{k'^2}{2m^*(r)}\right)\sin[2(k-k')(q-r)]$$

$$\rightarrow \sum_{n'=1}^{N_k}\frac{k_{n'}^2\mathcal{M}^o_{jn-n'} + 2\mathcal{V}_{jn-n'}}{N_k}f_{jn'} \quad (4.10)$$

$$\left.\frac{\partial f}{\partial t}\right|_{Coll.} \rightarrow \frac{1}{\tau_f}\left(\frac{f^{eq}}{\sum_{n''}f^{eq}_{jn''}} - 1\right)f_{jn} \quad (4.11)$$

The matrices $\mathcal{M}^e$, $\mathcal{M}^o$ and $\mathcal{V}$ are evaluated using fast sine and cosine transforms (Gullapalli & Neikirk, 1994) with N_k chosen to be a power of 2.

$$
\begin{aligned}
\mathcal{M}^e_{jn-n'} &= \frac{(-1)^{n'-n}}{2}\left(\frac{1}{m^*_{j+N_k/2}}+\frac{1}{m^*_{j-N_k/2}}-\frac{2}{m^*_{GaAs}}\right)+\left(\frac{1}{m^*_j}-\frac{2}{m^*_{GaAs}}\right) \\
&\quad +\sum_{j'=1}^{N_k/2-1}\left(\frac{1}{m^*_{j+j'}}+\frac{1}{m^*_{j-j'}}-\frac{2}{m^*_{GaAs}}\right)\cos\left[\frac{\pi(n'-n)j'}{N_k/2}\right] \\
\mathcal{M}^o_{jn-n'} &= \sum_{j'=1}^{N_k/2-1}\left(\frac{1}{m^*_{j+j'}}-\frac{1}{m^*_{j-j'}}\right)\sin\left[\frac{\pi(n'-n)j'}{N_k/2}\right] \\
\mathcal{V}_{jn-n'} &= \sum_{j'=1}^{N_k/2-1}(v_{j+j'}-v_{j-j'})\sin\left[\frac{\pi(n'-n)j'}{N_k/2}\right]
\end{aligned}
\tag{4.12}
$$

The symmetry properties of $\mathcal{M}^e$, $\mathcal{M}^o$ and $\mathcal{V}$ greatly reduce the task of filling the corresponding matrices. For each j it is necessary to compute only one half of the first row ($n=1, n'=1,2,\ldots,N_k/2+1$).

Summarizing, the matrix elements of the Liouville superoperator are:

$$
\begin{aligned}
\mathcal{L}_{jn;j'n'} &= \mathcal{T}_{jn;j'n'}-\frac{k_{n'}\mathcal{M}^e_{j'n-n'}}{N_k}\left(\frac{\delta_{j'j+1}-\delta_{j'j-1}}{2\Delta}\right) \\
&\quad -\frac{\mathcal{M}^o_{j'n-n'}}{4N_k}\left(\frac{\delta_{j'j+1}-2\delta_{j'j}+\delta_{j'j-1}}{\Delta^2}\right) \\
&\quad +\frac{k_{n'}^2\mathcal{M}^o_{j'n-n'}+2\mathcal{V}_{j'n-n'}}{N_k}\delta_{j'j}+\frac{\delta_{j'j}}{\tau_f}\left(\frac{f^{eq}}{\sum_{n''}f^{eq}_{jn''}}-\delta_{n'n}\right)
\end{aligned}
\tag{4.13}
$$

4.3   Transport in a tight-binding energy band

For reasons already discussed in the last chapter, a consistent discrete model must also assume a bandstructure that is consistent with the discrete lattice. This is the tight-binding energy band. Solving for the bandstructure on a lattice with spacing 2Δ yields the band structure

$$
\mathcal{E}(k)=\frac{\hbar^2}{4m^*\Delta^2}\left[1-\cos(2k\Delta)\right],
$$

and the Wigner function must be solved on a lattice with half the spacing (Δ). This ensures, correctly, that the periodicity of the discrete Wigner function is

the same as that of the bandstructure. As in the case of the parabolic energy band, we first separate the uniform effective-mass drift term $\mathcal{T}f(q)$:

$$\mathcal{T}f(q) = -\frac{\hbar \sin(2k\Delta)}{2m_{GaAs}^*\Delta}\frac{f(q+\Delta,k) - f(q-\Delta,k)}{2\Delta}$$

Following the procedure outlined in the previous section, the discrete Liouville superoperator for a tight binding energy band is (Gullapalli & Neikirk, 1994),

$$\begin{aligned}
\mathcal{L}_{jn;j'n'} &= \mathcal{T}_{jn;j'n'} - \frac{\sin\left(2k_{n'}\Delta\right)\mathcal{M}_{j'n-n'}^e}{2N_k\Delta}\left(\frac{\delta_{j'j+1} - \delta_{j'j-1}}{2\Delta}\right) \\
&\quad - \frac{\cos\left(2k_{n'}\Delta\right)\mathcal{M}_{j'n-n'}^o}{4N_k}\left(\frac{\delta_{j'j+1} - 2\delta_{j'j} + \delta_{j'j-1}}{\Delta^2}\right) \\
&\quad + \frac{\left(\frac{1-\cos(2k_{n'}\Delta)}{2\Delta^2}\right)\mathcal{M}_{j'n-n'}^o + 2\mathcal{V}_{j'n-n'}}{N_k}\delta_{j'j} \\
&\quad + \frac{\delta_{j'j}}{\tau_f}\left(\frac{f^{eq}}{\sum_{n''}f_{jn''}^{eq}} - \delta_{n'n}\right)
\end{aligned} \tag{4.14}$$

where the matrices $\mathcal{M}^e$, $\mathcal{M}^o$ and $\mathcal{V}$ are as in 4.12. The approach can be easily extended to include the other Fourier components of the energy band structure.

4.4 Discretization of the drift term

The drift term in a tight-binding energy band is discrete in position, and in a centered difference form. Since the tight-binding equation of motion reduces to the equation of motion in a parabolic band in the limit $\Delta \to 0$, it is clear that in a parabolic energy band too, centered differencing the drift term is most appropriate. However, the centered difference form for the drift term is unsuitable for numerical simulation. An important aspect in numerical modeling is stability. Consider a semiconductor device initially in steady state and subject to some disturbance. When the disturbance is removed, the device will return to the original steady state. The time evolution of the initial disturbance $\delta f(0)$ is governed by the equation of motion and is generally given as $\delta[f](t) = e^{[\lambda]t}\delta[f](0)$ where $[\lambda]$ is the eigen-value spectrum of the discrete Liouville superoperator. For the disturbance to decay, and the system to return

stably to steady state, $[\lambda]$ must not have any positive real part. In the absence of any collisions, if the centered difference form of the drift term is retained, all eigenvalues of the Liouville superoperator have zero real part. This is due to the Hermiticity of the Liouvillian. Due to inevitable numerical round-off errors, this will actually lead to exponentially growing disturbances. Even the presence of disspative processes such as electron-phonon collisions, which serve to damp disturbances, does not always lead to a stable numerical model (Frensley, 1990). Even under steady state conditions, the boundary value problem cannot be solved if there exists a near zero eigenvalue. Incorporation of an arbitrary fraction of the spurious mode $\cos(j\pi)$ leads to zone-to-zone oscillations and renders the solution meaningless.

To realize a physical model, it is necessary to incorporate some mechanism in the numerical model to dissipate disturbances or errors. A way out of the difficulty is to use upwind differencing instead of the centered difference. Two upwind differencing schemes are discussed next and their utility for simulating the Wigner equation of motion investigated.

4.4.1 First order upwind/downwind discretization

The simplest way to obtain a stable numerical model is to use upwind differencing to approximate the drift term $\mathcal{T}f(q)$:

$$\mathcal{T}f(q) = -\frac{\hbar \sin(2k\Delta)}{2m^*\Delta^2} \begin{cases} f(q+\Delta) - f(q) & \text{for } k < 0 \\ f(q) - f(q-\Delta) & \text{for } k > 0 \end{cases} \qquad (4.15)$$

That the upwind difference scheme introduces dissipation and numerical stability can be understood by considering the quation of motion in the absence of any potential:

$$\frac{\partial f}{\partial t} = \mathcal{T}f(q)$$

Under steady state $f(q + \Delta) = f(q) = f(q - \Delta)$. Therefore, a perturbation δf_j at grid point j evolves according to:

$$\frac{\partial f_j}{\partial t} = -\frac{\hbar \sin(2k\Delta)}{2m^* \Delta^2} \begin{cases} -\delta f_j & \text{for } k < 0 \\ \delta f_j & \text{for } k > 0 \end{cases}$$

The solution is

$$\delta f_j(k) = e^{-\frac{\hbar |\sin(2k\Delta)|}{2\Delta_q^2 m^*} t}$$

the perturbation decaying with a time constant $\tau = 2\Delta^2 m^* / \hbar |\sin(2k\Delta)|$.

The utility of the upwind scheme for the simulation of resonant tunneling diodes is discussed next. Most commonly simulated is a $Al_{0.3}Ga_{0.7}As/GaAs$ diode, composed of 30 Å AlGaAs barriers sandwiching a 50 Å GaAs quantum well. Ignoring the self-consistent potential, the applied bias is usually dropped linearly across the double barrier quantum well structure. In this study the contact regions are assumed to be doped n-type at $2 \times 10^{18} cm^{-3}$. Room temperature (300K) is assumed in all calculations.

Calculations using 4.15 result in the $J - V$ curve shown in Fig. 4.2. Also shown in the figure are the curves due to Tsuchiya et al. (1991). Calculations with the parabolic band structure also yield similar results. This is to be expected since in the energy range of interest, the tight-binding and parabolic bands are similar. As discussed in Section 2.2.2, higher the effective-mass in the barriers, the narrower the resonances. Consequently the current density decreases with increasing effective-mass in the barrier. While our model correctly predicts this behavior, the model of Tsuchiya et al. (1991) yields the absurd result that the peak current increases with increasing barrier effective-mass. The problem with the work of Tsuchiya et al. (1991) is the rather unwieldy form in which they write the equation of motion, which in turn encourages a very bad numerical model. Appendix A outlines how their equation can be rewritten in a form that affords accurate and elegant discretization.
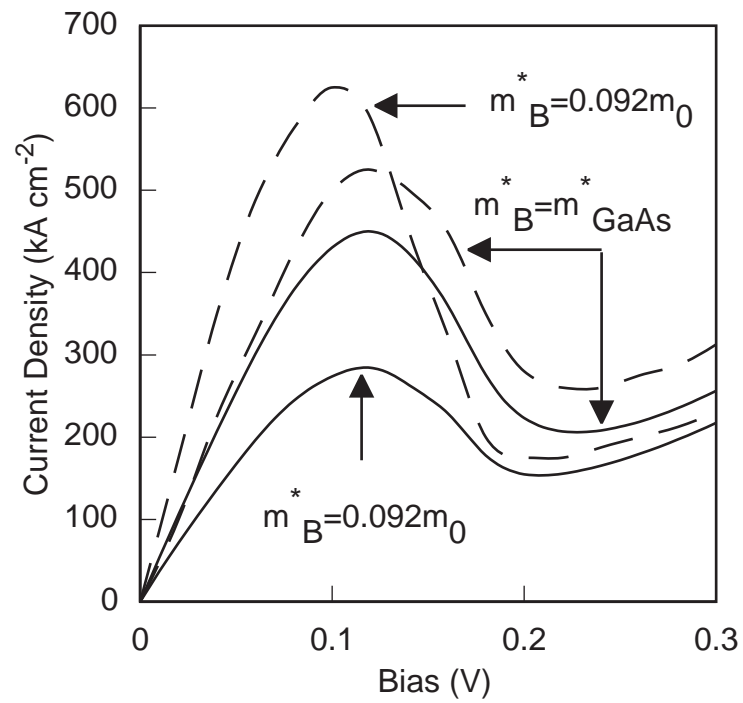
Figure 4.2: Calculated $J - V$ curve of the AlGaAs/GaAs double barrier diode. Collisions are ignored. $N_q = 80$ and $N_k = 64$. Solid lines are the results of this work and the light dashed lines are due to Tsuchiya.
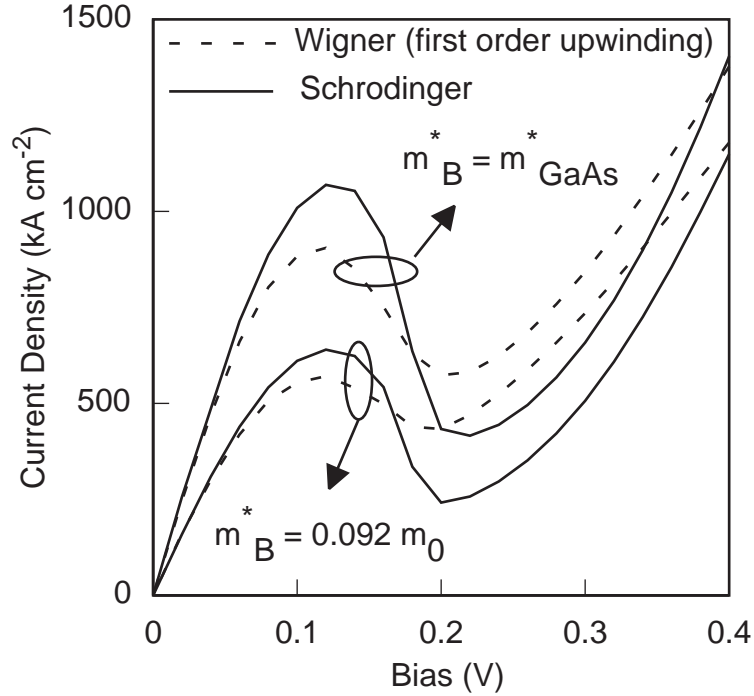
Figure 4.3: Calculated $J - V$ curve of the AlGaAs/GaAs double barrier diode using the Wigner and Schrödinger models.

The $J - V$ curve predicted by the Schrödinger and Wigner models are compared in Fig. 4.3 where a contact doping of $4 \times 10^{18} cm^{-3}$ is assumed. The Wigner and Schrödinger models predict similar characteristics. The Wigner calculation however predicts lower peak currents and higher valley currents when compared to the Schrödinger calculation. In the absence of collisions the two models should yield the same results since both models are based on the effective-mass Hamiltonian. Comparing results obtained using the two models can bring to light problems with Wigner function calculations that are not obvious.

Consider next the application of 4.15 in the simulation of the more useful AlAs/GaAs double barrier diodes. The AlAs/GaAs double barrier diode considered here is our baseline device. The tunneling structure consists of two 17 Å thick AlAs barriers sandwiching a 50 Å wide GaAs quantum well. On
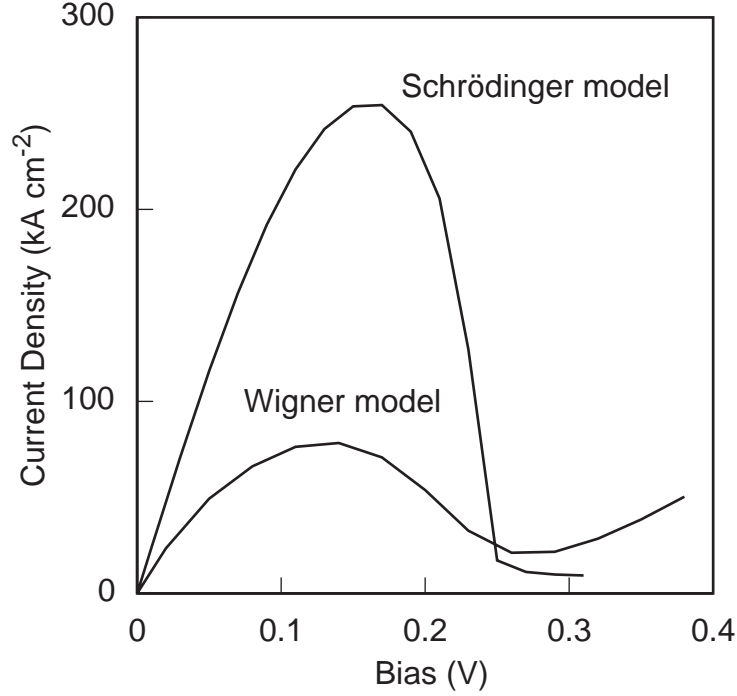
Figure 4.4: Calculated $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode under flat-band conditions and ignoring the higher effective-mass in $AlAs$. In the Wigner calculations collisions are ignored and $N_q = 314$, $\Delta = a$ and N_k=128, the boundaries of the simulation being placed $150a \sim 900$ Å away from the double barrier quantum well structure.

either side of the tunneling structure is a three step spacer layer consisting of 50 Å undoped GaAs closest to the barrier, 100 Å GaAs doped at $5 \times 10^{16} cm^{-3}$, and 100 Å GaAs doped at $6 \times 10^{17} cm^{-3}$. Thick GaAs contact regions doped at $4 \times 10^{18} cm^{-3}$ complete the device structure. The conduction band offset at the AlAs/GaAs heterojunction is taken to be $1eV$, $m^*_{GaAs} = 0.067m_0$, and $m^*_{AlAs} = 0.15m_0$.

If the requirement of self-consistency is ignored, and the applied potential is dropped across the tunneling structure, the resulting $J - V$ curve, ignoring the position dependence of effective-mass, is shown in Fig. 4.4. Also shown in the figure are the characteristics due to the Schrödinger model. The limitations of first order upwind differencing are beginning to surface. The
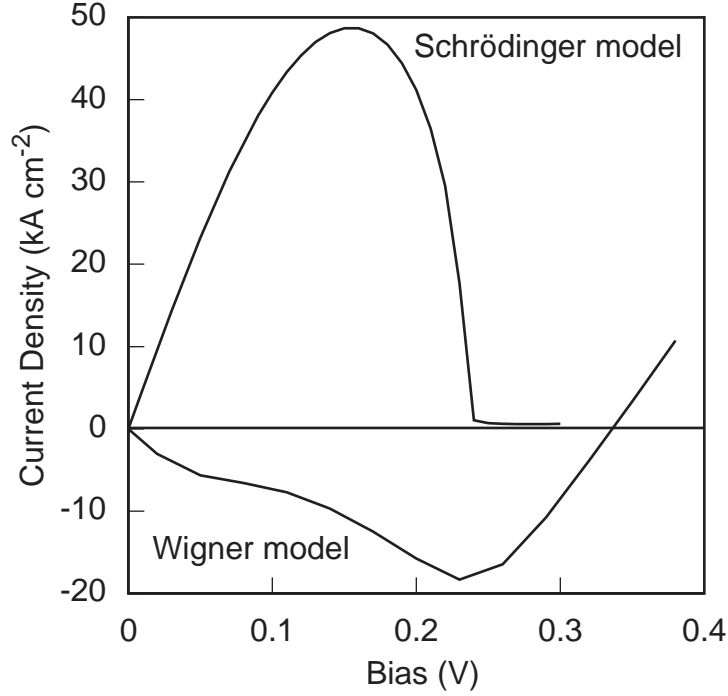
Figure 4.5: Calculated $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode under flat-band conditions and taking into account $m^*_{AlAs} = 0.15m_0$. In the Wigner calculations collisions are ignored and $N_q = 314$, $\Delta = a$ and $N_k{=}128$, the boundaries of the simulation being placed $150a \sim 900\,\text{Å}$ away from the double barrier quantum well structure.

peak current and the resonance in the current-voltage curve are severely degraded. The only other simulation of high barrier height devices is due to Mains and Haddad (1988). Their calculations, under flatband conditions gave peak current densities in $AlAs/In_{0.53}Ga_{0.47}As$ double barrier diodes that are in agreement with experimental values, even without the inclusion of the higher AlAs effective-mass. The calculations here show that since upwind differencing severely degrades the peak current, which is also the effect of the higher AlAs effective-mass, the agreement observed in the study of Mains and Haddad (1988) is fortuitous. When the non-uniformity of the effective-mass is considered, the upwind difference scheme yields completely meaningless results as shown in Fig. 4.5. The problem can be understood further by considering the
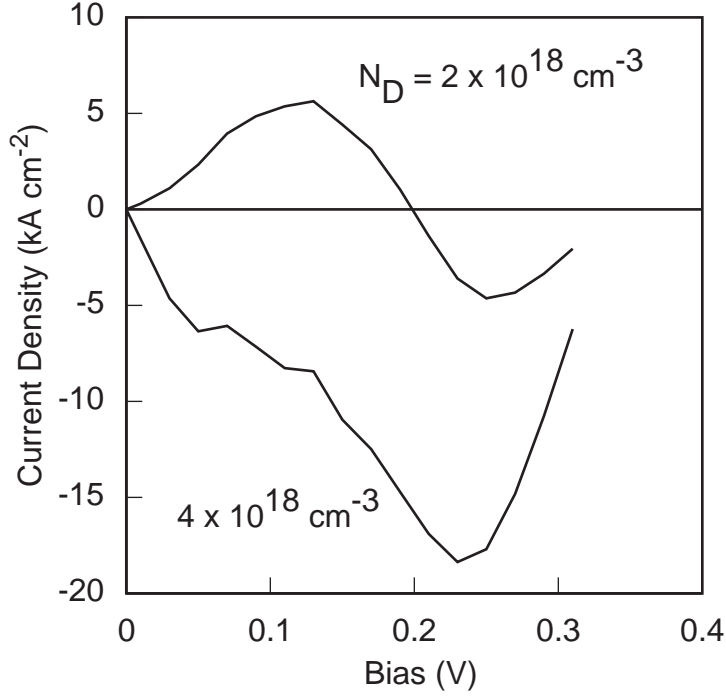
Figure 4.6: The effect of the contact dopant density on the calculated $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode under flat-band conditions and taking into account $m^*_{AlAs} = 0.15m_0$. $N_q = 214$, $\Delta = a$ and N_k=128, the boundaries of the simulation being placed $100a \sim 600\,\text{Å}$ away from the double barrier quantum well structure.

flatband calculation with a reduced contact doping. The result is shown in Fig. 4.6. As the contact doping is reduced, the electron density decreases, especially at high energy. As a result, the contribution of discretization errors in the Wigner trajectories are reduced. The situation therefore improves to some extent. However, the dependence of the Wigner trajectories on the applied bias is complex and the problem remains at higher applied voltages.

The possibility that the errors are due to the boundary being too close to the active region of the device has been investigated. The simulations have been done with increasing spacing between the boundary and the double barrier structure with the spacing being as large as 1600Å. The investigations have shown that the results are only weakly dependent on the position of the bound-

ary, a conclusion also reached by Jensen and Buot (1990). The real problem is the inherent flaw in the upwind differencing of the drift-term which surfaces when the barrier height is large. The stronger the barrier, the more nonlocal in momentum is its effect. Errors due to upwind differencing the drift-term in relatively low current carrying regions of momentum space can be strongly coupled to high current carrying regions. While this might not lead to major errors in the electron density, the error in the overall current can be large. In addition, since the effective-mass variation is also large in AlAs/GaAs devices, and since this term also adds non-locality in momentum space, the problem is even worse.

A more insightful view of the problem is obtained by considering what the upwind differencing implies about the bandstructure. Consider the drift term for a spatially uniform band structure $\mathcal{E}(k)$

$$\frac{\partial f_{jn}}{\partial t} = -\frac{2}{N_k} \sum_{j'=j-N_k/2}^{j+N_k/2} f_{j'n} \sum_{n'=1}^{N_k} \mathcal{E}_{n'} \sin\left[2\pi(j-j')(n-n')/N_k\right]$$

This is analogous to the term due to the potential. While the potential couples different momenta, the band structure couples different positions. The coefficients of $f_{j'n}$ are clearly the discrete sine transform of the band structure. For a real function, which the band structure is, the sine transform satisfies the relation $F(j-j') = -F^*(j'-j)$. This therefore leads to a centered difference form, with the number of terms given by twice the number of Fourier components. For the tight-binding band we have assumed, only one Fourier component exists, and we get the the centered difference form $f_{j+1n} - f_{j-1n}$. The upwind difference breaks this symmetry. If the coefficients in the upwind difference are taken to be the Fourier coefficients of the bandstructure, then the bandstructure cannot be real. The nature of the errors due to this may depend on the bandstructure assumed, and on the strength of the coupling in k space due to the potential and effective-mass variations. For the parabolic
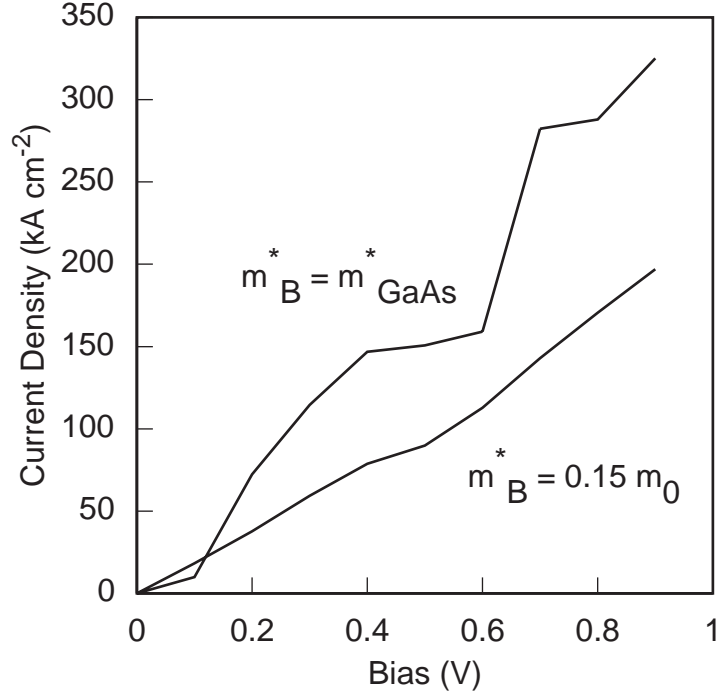
Figure 4.7: Calculated $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode. Collisions are ignored. $N_q = 268$, $\Delta = a$ and $N_k = 128$.

energy band, the problem of the negative currents at positive bias (for the case shown in Fig. 4.5) is not observed for the structure discussed here, the problem with the upwind difference manifesting itself in that no negative differential resistance is obtained.

The self-consistent $J-V$ curve of the baseline diode calculated using first order upwinding is shown in 4.7. Again, the results are clearly unsatisfactory. It is natural to expect the results to improve by increasing the density of the phase-space mesh. For instance, the approximation of the first order upwind difference to the centered difference improves with decreasing mesh-spacing. The density of the phase-space mesh was increased to study the effect on the results. With $\Delta = a, N_q = 268$ and $N_k = 128$ in the baseline mesh, two cases where investigated. In the first case, Δ is halved thereby doubling the mesh density in position and improving the accuracy of the finite difference

approximation to the drift term. To keep Δ_k the same N_k must be doubled. This increases the problem size (memory requirement) by a factor of 8. In the second case, N_k is doubled thus doubling the density of the k-space mesh. This increases the problem size by a factor of 4. Since the baseline k-mesh density is already high, there is no observable change in the $J - V$ curve curve with the finer k-mesh. Even with the finer spatial mesh, the results were essentially the same as that shown in Fig. 4.7. It is unlikely that increasing the mesh density even further will improve results. The problem does not lie in the coarseness of the mesh. In any case, the mesh density cannot be increased any further without suffering from shortage of workspace. A more elegant solution will be discussed shortly. It is also important to note that because the equation of motion and the spatial mesh follow directly from the band structure parameter Δ, it is clear that reducing the mesh-spacing will also imply a different bandstructure. Increasing the mesh density also makes the set of linear algebraic equations larger.

4.4.2   Second order upwind/downwind discretization

A second order upwind differencing scheme was investigated by Jensen and Buot (1991) which improves the accuracy of the Wigner trajectories in Al-GaAs/GaAs double barrier diodes. Consider a quadratic $f(q)$ passing through the points $(j\Delta, f_j)$, $((j + 1)\Delta, f_{j+1})$, and $((j + 2)\Delta, f_{j+2})$ for $k < 0$ and through the points $(j\Delta, f_j)$, $((j - 1)\Delta, f_{j-1})$, and $((j - 2)\Delta, f_{j-2})$ for $k > 0$. On taking the derivative of the quadratic at $q = j\Delta$, the second order upwind differencing results:

$$\mathcal{T} = -\frac{\hbar \sin(2k\Delta)}{4\Delta^2 m^*} \begin{cases} -f(q + 2\Delta) + 4f(q + \Delta) - 3f(q) & \text{for } k < 0 \\ f(q - 2\Delta) - 4f(q - \Delta) + 3f(q) & \text{for } k > 0 \end{cases} \qquad (4.16)$$

The $J - V$ curves of the baseline $AlAs/GaAs$ diode calculated using the second order upwind differencing are shown in Fig. 4.8. While there appears to
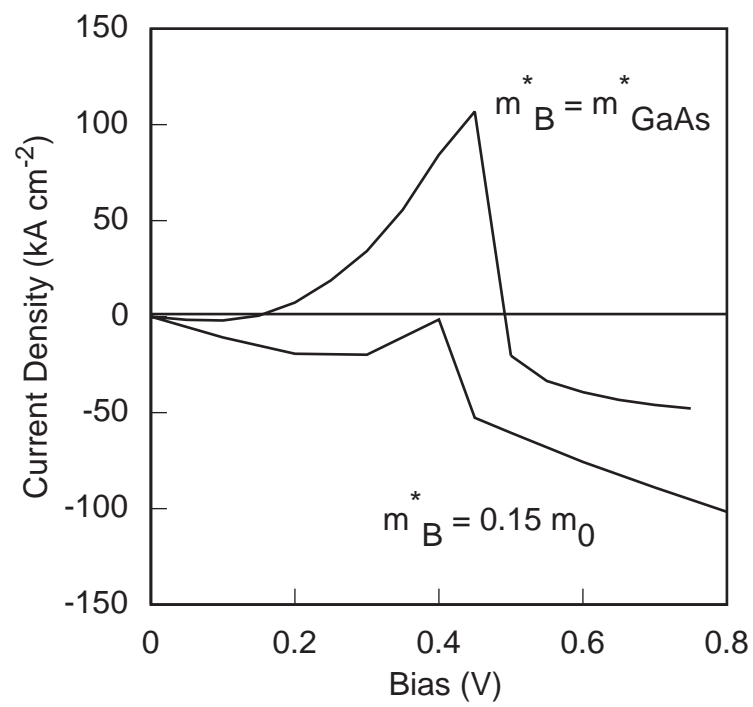
Figure 4.8: Calculated $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode. $N_q = 268$, $\Delta = a$ and $N_k = 128$.

be some improvement over first order upwind differencing, the results are still far from satisfactory. To understand the importance of properly discretizing the drift term we turn to the equation of motion in the tight-binding energy band.

4.4.3 Improved discretization schemes

Clearly, numerical stability cannot be the overriding factor in the development of a numerical model. In order to improve the fidelity of the numerical model to the physical model, hybrid discretization schemes are investigated next. Consider a second order polynomial $f(q)$ passing through the points $((j-1)\Delta, f_{j-1})$, $(j\Delta, f_j)$, and $((j+1)\Delta, f_{j+1})$. On taking the derivative of the polynomial we get:

$$\frac{\partial f}{\partial q} = \frac{f_{j+1} - f_{j-1}}{2\Delta} + \frac{f_{j+1} - 2f_j + f_{j-1}}{2\Delta}\left(\frac{q - j\Delta}{\Delta/2}\right) \tag{4.17}$$

Evaluating the derivative at $q = j\Delta$, yields the centered difference. Evaluating the derivative at $q = (j - 1/2)\Delta$ or $q = (j + 1/2)\Delta$ depending on the sign of k, yields first order upwind differencing. Within this framework, the centered and first order upwind schemes are special cases. Since the centered difference is most appropriate, the derivative can be evaluated close to $q = j\Delta$ but away from it so that some numerical dissipation is incorporated and a stable system obtained. This is summarized by the following approximation (Gullapalli & Neikirk, 1994):

$$f_{j+1n} - f_{j-1n} \to f_{j+1n} - f_{j-1n} - sgn\,(k_n)\delta(f_{j+1n} - 2f_{jn} + f_{j-1n}) \tag{4.18}$$

where $\delta = 2(q - j\Delta)/\Delta$ is a parameter that can be used to control the fidelity of the numerical model to the physical model. Using this scheme yields the results shown in Fig. 4.9. The result of including collisions in the relaxation-time approach with $\tau_f = 100fs$, and including effective-mass nonuniformity is shown in Fig. 4.10. Also shown in the figure is a typical measured room-temperature
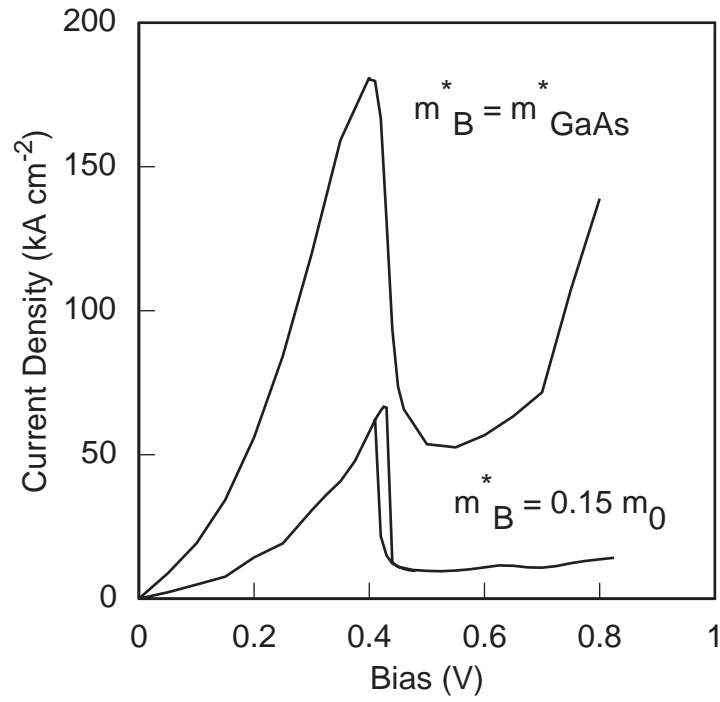
Figure 4.9: Self-consistent $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode. $N_q = 268$, $\Delta = a$ and $N_k = 128$. $\delta = 0.1$ and collisions are ignored.
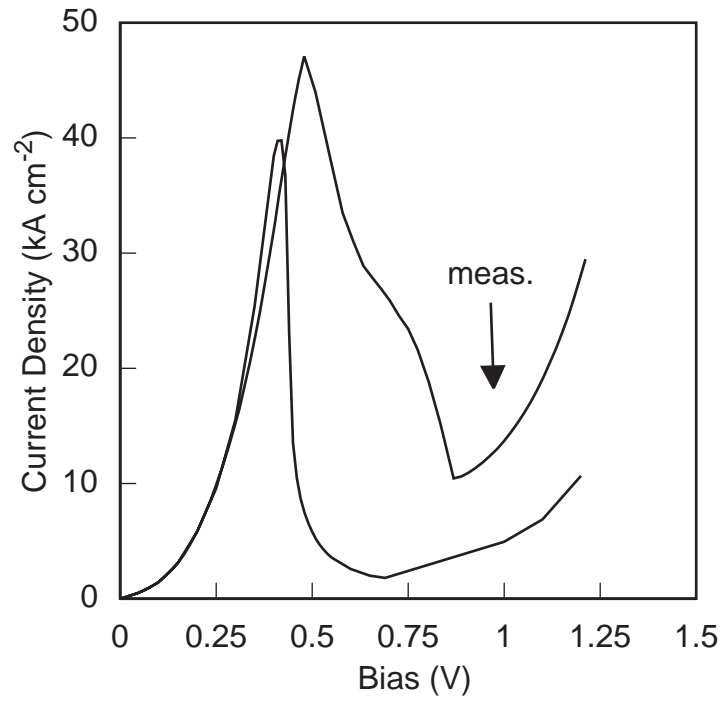
Figure 4.10: Self-consistent $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode. $N_q = 268$, $\Delta = a$ and $N_k = 128$. $\delta = 0.01$ and $\tau_f = 100 fs$. Also shown is the room-temperature measurement.

characteristic of the baseline diode. The agreement is good below the peak voltage. Phenomena such a , $- X$ tunneling are expected to play a major role in obtaining better agreement. While such effects can be included in the analysis as discussed in chapter 3, this work addresses the more urgent need to develop robust numerical techniques for the solution of the transport equation. δ cannot be made too small without running into numerical instability. It is interesting that there is a tradeoff between numerical stability and accuracy. What is perhaps surprising is that numerical stability is retained even with δ as low as 0.01.

Equation 4.18 can also be considered as a hybrid between the centered difference and the first order upwind schemes. In that spirit, a hybrid between the centered difference and the second order upwind schemes follows naturally. This results in (Gullapalli, Miller, & Neikirk, 1994):

$$f_{j+1n} - f_{j-1n} \rightarrow \frac{1}{1+\delta} \left[f_{j+1n} - f_{j-1n} + \delta \left\{ \begin{array}{ll} -f_{j+2n} + 4f_{j+1n} - 2f_{jn} & k_n < 0 \\ f_{j-2n} - 4f_{j-1n} + 3f_{jn} & k_n > 0 \end{array} \right. \right]$$
(4.19)

$\delta = 0$ yields the centered difference scheme and $\delta \rightarrow \infty$ yields second order upwind differencing. $\delta = 1/2$ yields a third order scheme. Using the third order scheme, the $J - V$ curve of the baseline device is shown in Fig. 4.11. While the improvement in the uniform effective-mass case is significant, when the effective-mass is spatially varying, the results are not satisfactory. Even letting δ take large values does not cure the problem of negative currents.

There might be some concern that the discretization schemes considered here are not transportive. Upwinding leads to transportive models: disturbances at a grid point propagate only in the direction of flow. However, since $\mathcal{V}$ strongly couples the upwind and downwind flows, the transportive property should not be of great concern here. In any case, the exact equation of motion is not transportive either.
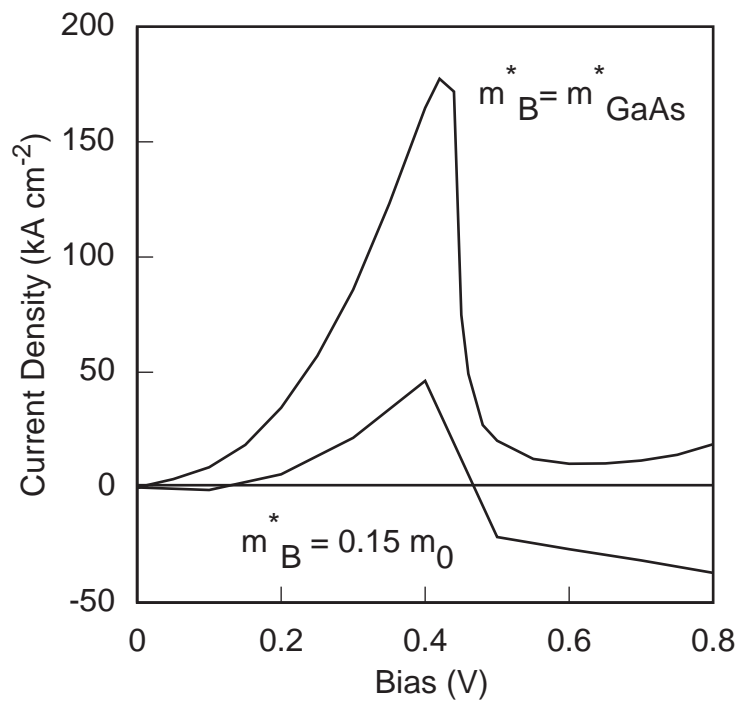
Figure 4.11: Self-consistent $J - V$ curve of the baseline $AlAs/GaAs$ double barrier diode. $N_q = 268$, $\Delta = a$ and $N_k = 128$. Collisions are ignored.

4.5 Current density in the discrete model

To define the current density in the discrete model, the equation of continuity is first written. Here, the equation is

$$\frac{\Delta_k}{2\pi}\frac{\partial}{\partial t}\sum_n f_{jn} = \frac{\partial n_j}{\partial t} = \frac{\Delta_k}{2\pi}\sum_{j'n'}\sum_n \mathcal{L}_{jn;j'n'}f_{j'n'} \qquad (4.20)$$

While the continuum continuity equation follows from the odd symmetry of the kernels $\mathcal{V}$ and $\mathcal{M}^o$, the discrete kernels vanish after summation over n only if the Fourier completeness condition is satisfied. The collision term also vanishes after the summation over n. Therefore the continuity equation is

$$\begin{aligned}
\frac{\partial n_j}{\partial t} &= -\frac{\Delta_k}{2\pi}\sum_{nj'n'}\mathcal{T}_{jn;j'n'}f_{j'n'}\\
&\quad - \sum_{nj'n'}\frac{\Delta_k \sin(2k_{n'}\Delta)\mathcal{M}^e_{j'n-n'}}{4\pi N_k\Delta}\left(\frac{\delta_{j'j+1}-\delta_{j'j-1}}{2\Delta}\right)
\end{aligned} \qquad (4.21)$$

where $\mathcal{T}$ is the discrete drift term.

For the discretization 4.18 the continuity equation is:

$$\begin{aligned}
\frac{\partial n_j}{\partial t} &= -\frac{\hbar\Delta_k}{8\pi m^*_{GaAs}\Delta^2}\sum_n \sin(2k_n\Delta)\left[(1\pm\delta)f_{j+1n}+(1\mp\delta)f_{jn}\right.\\
&\qquad\qquad\qquad\qquad\left. -(1\pm\delta)f_{jn}-(1\mp\delta)f_{j-1n}\right]\\
&\quad -\frac{\hbar\Delta_k}{8\pi N_k\Delta^2}\sum_{nn'}\sin(2k_{n'}\Delta)\left[\mathcal{M}^e_{j+1n-n'}f_{j+1n'}+\mathcal{M}^e_{jn-n'}f_{jn'}\right.\\
&\qquad\qquad\qquad\qquad\left. -\mathcal{M}^e_{jn-n'}f_{jn'}-\mathcal{M}^e_{j-1n-n'}f_{j-1n'}\right]\\
&= -\frac{\mathcal{J}_{j+1/2}-\mathcal{J}_{j-1/2}}{\Delta}
\end{aligned} \qquad (4.22)$$

where the top sign is used when $k_n < 0$ and the bottom when $k_n > 0$. Examining 4.22, the current density is defined as:

$$\begin{aligned}
\mathcal{J}_{j+1/2} &= \frac{\hbar}{8N_k\Delta^2}\sum_n \left[\frac{\sin(2k_n\Delta)\left[(1\pm\delta)f_{j+1n}+(1\mp\delta)f_{jn}\right]}{m^*_{GaAs}}\right.\\
&\qquad\qquad\left. +\frac{1}{N_k}\sum_{n'}\sin(2k_{n'}\Delta)\left[\mathcal{M}^e_{j+1n-n'}f_{j+1n'}+\mathcal{M}^e_{jn-n'}f_{jn'}\right]\right] \quad (4.23)
\end{aligned}$$

Following a similar procedure for the discretization 4.19, the definition for current follows:

$$
\begin{aligned}
\mathcal{J}_{j+1/2} \;=\; & \frac{\hbar}{8 N_k \Delta^2} \sum_n \left[\frac{\sin(2 k_n \Delta)}{(1+\delta) m^*_{GaAs}} \right. \\
& \times \left\{ \begin{array}{ll} (1+3\delta) f_{j+1n} + f_{jn} - \delta f_{j+2n} & k_n < 0 \\ (1+3\delta) f_{jn} + f_{j+1n} - \delta f_{j-1n} & k_n > 0 \end{array} \right. \\
& + \frac{1}{N_k} \sum_{n'} \sin(2 k_{n'} \Delta) \left. \left[\mathcal{M}^e_{j+1n-n'} f_{j+1n'} + \mathcal{M}^e_{jn-n'} f_{jn'} \right] \right]
\end{aligned}
\tag{4.24}
$$

4.6   Solution method

For completeness, the solution procedure adopted in this work is briefly outlined. Following the discretization discussed in the preceding sections, the set of equations $A_{N_q} f = b$ for the Wigner function can be naturally put into a block pentadiagonal system of the form:

$$
\begin{bmatrix}
A_{11} & A_{12} & A_{13} & & & \cdots & & 0 \\
A_{21} & A_{22} & A_{23} & \ddots & & & & \vdots \\
A_{31} & A_{32} & A_{33} & \ddots & \ddots & & & \\
& \ddots & \ddots & \ddots & \ddots & & \ddots & \\
& & \ddots & \ddots & \ddots & & \ddots & A_{N_q - 2 N_q} \\
\vdots & & & \ddots & \ddots & & \ddots & A_{N_q - 1 N_q} \\
0 & \cdots & & & A_{N_q N_q - 2} & A_{N_q N_q - 1} & A_{N_q N_q}
\end{bmatrix}
\begin{bmatrix}
f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{N_q - 2} \\ f_{N_q - 1} \\ f_{N_q}
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{N_q - 2} \\ b_{N_q - 1} \\ b_{N_q}
\end{bmatrix}
$$

If discretization 4.18 is used, we get a block tridiagonal system.

All the blocks are $N_k \times N_k$ and the f_i and b_i are in $\Re_k^N$. The solution of the Wigner transport equation, even in one dimension, can place an enormous burden on memory and computational resources. For example, solving the above equation using a band matrix algorithm requires $6 N_q N_k^2$ double precision storage. For typical values of $N_q = 200$ and $N_k = 128$ the workspace is 150MB. Adopting a block LU factorization algorithm reduces the storage required to

one-fourth of that required by a band algorithm. Comparing blocks in A_n and

$$
\begin{bmatrix} I & & & & \cdots & & 0 \\ & & & & & & \vdots \\ L_{21} & I & & & & & \\ L_{31} & L_{32} & I & & & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ \vdots & & & & & \ddots & \\ 0 & \cdots & & L_{nn-2} & L_{nn-1} & I \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} & A_{13} & & \cdots & & 0 \\ & U_{22} & U_{23} & \ddots & & & \vdots \\ & & U_{33} & \ddots & & & \\ & & & & \ddots & & \\ & & & & & \ddots & A_{n-2n} \\ \vdots & & & & & \ddots & U_{n-1n} \\ 0 & \cdots & & & & & U_{nn} \end{bmatrix}
$$

we obtain, since the U_{ii} are nonsingular, the following algorithm for the L_{ij} and U_{ij} :

$$U_{11}^T = A_{11}^T, \quad U_{12}^T = A_{12}^T, \quad y_1^T = b_1^T$$

Solve $\quad U_{11}^T \left[L_{21}^T | L_{31}^T \right] = \left[A_{21}^T | A_{31}^T \right]$

$$U_{22}^T = A_{22}^T - U_{12}^T L_{21}^T, \quad U_{23}^T = A_{23}^T - A_{13}^T L_{31}^T, \quad y_2^T = b_2^T - y_1^T L_{21}^T$$

for $\quad i = 3 : N_q$

$\quad$ Solve $\quad U_{i-1i-1}^T \left[L_{ii-1}^T | L_{i+1i-1}^T \right] = \left[A_{ii-1}^T - U_{i-2i-1}^T L_{ii-2}^T | A_{i+1i-1}^T \right]$

$\quad y_i^T = b_i^T - y_{i-1}^T L_{ii-1}^T - y_{ii-2}^T L_{ii-2}^T$

$\quad U_{ii}^T = A_{ii}^T - U_{i-1i}^T L_{ii-1}^T - A_{i-2i}^T L_{ii-2}^T$

$\quad U_{ii+1}^T = A_{ii+1}^T - A_{i-1i+1}^T L_{ii-1}^T$

end

Executing the back substitution,

$\quad$ Solve $\quad U_{N_q N_q} f_{N_q} = y_{N_q}$

$\quad$ Solve $\quad U_{N_q-1 N_q-1} f_{N_q-1} = y_{N_q-1} - U_{N_q-1 N_q} f_{N_q-1}$

for $\quad i = N_q - 2 : 1$

$\quad$ Solve $\quad U_{ii} f_i = y_i - U_{ii+1} f_{i+1} - A_{ii+2} f_{i+2}$

end

we obtain the Wigner function. A similar block oriented method has recently been published by Jensen and Ganguly (1993). The block LU factorization, while not being as robust as the band LU method, works very well for the simulation of the Wigner equation of motion. Whenever possible, the results have been checked against those obtained using band LU factorization with partial pivoting (see, for example, the code for band LU factorization in (Press et al., 1992)).

Block LU factorization is at least a factor of 4 faster than a band LU factorization. Also, there is opportunity for considerable reduction in work space required. At any stage only L_{ii-2}, L_{ii-1} and L_{i+1i-1} need to be stored. L_{ii-1} and L_{i+1i-1} are obtained using LU factorization. The forward substitution is carried out as soon as L_{ii-1} and L_{i+1i-1} are available. Storage is mainly required for $U_{ii}(N_k \times N_k)$ and $U_{ii+1}(N_k \times N_k/2)$. The reduced storage for U_{ii+1} is because one half of A_{ii+2} is zero and one half of A_{ii+1} does not need to be stored. A_{ii+1} and A_{ii+2} are easily recalculated as required during back substitution. Further improvements in the solution procedure may be possible by exploiting the odd-even symmetries of the blocks.

4.7   Summary

Traditionally, the use of the upwind difference scheme has been motivated by considerations of numerical stability. In the present problem, however, we have shown that it is more important to maintain the fidelity of the numerical model to the physical system. Upwind difference schemes for the drift term yield completely meaningless predictions for high band offset, AlAs/GaAs heterostructures where the bandstructure also varies significantly. The stronger the barrier, the more nonlocal in momentum is its effect. Errors due to upwind differencing the drift-term in relatively low current carrying regions of momentum space can be strongly coupled to high current carrying regions. While this might not lead to major errors in the electron density, the error in the overall current can be quite large especially near resonance, when the quantum effects can be particularly strong. When the barrier is small, the effect is not so important. In addition, since the effective-mass variation is also large in AlAs/GaAs devices, and since this term also adds non-locality in momentum space, the problem is even worse. We have presented improved hybrid difference schemes that greatly improve the accuracy of the numerical model while

maintaining numerical stability. An efficient algorithm for the solution of the transport equation, at least compared to the band LU factorization, has been presented. This algorithm allows the simulation of much longer devices with denser momentum grids.

While direct matrix solution has been adopted for the solution of the transport equation, such methods have serious drawbacks. In particular, the treatment is necessarily limited to one momentum direction, whereas processes such as electron-phonon scattering redistribute energy in all momentum directions. Also, including general bandstructures is computationally overburdening. Much of our understanding of semiclassical transport has been possible due to the Monte Carlo technique (Jacoboni & Lugli, 1989). Since the structure of the Wigner transport is very similar to the Boltzmann transport equation, it is conceivable that powerful methods such as the Monte Carlo and cellular-gas methods (Kometer & Zandler, 1992) can be adapted to solve the quantum transport problem. While these possibilities have not been investigated, the possibilities are intriguing.