Study of the Wigner function boundary conditions at different barrier heights

Andrea Savio and Alain Poncet
Lyon Institute of Nanotechnology - INSA Lyon
7 avenue Jean Capelle, 69621 Villeurbanne, France
Telephone: (33) 0472 437036
Email: andrea.savio@insa-lyon.fr

***Abstract*—In this work, we compute the Wigner function from wavefunctions obtained by solving the Schrödinger Equation. Our goal is to investigate issues that we encounter when simulating devices by solving the Wigner transport equation, namely the numerical discrepancies between the boundaries and the neighboring regions. In this paper, we focus on how the boundary conditions are affected by the barrier height in single- and double-barrier devices.**

I. Introduction

As the size of electronic devices is shrunk in order to attain higher integration densities, quantum phenomena begin to affect transport significantly, and thus simulators must be able to handle them. Solving the Wigner Transport Equation (WTE) [1] is one of the possible approaches in simulating quantum transport, which has been studied extensively, among others, by Kluksdahl [2], [3], Frensley [4]–[6], Jensen [7]–[9], Tsuchiya [10], Biegel [11] and Yamada [12].

The WTE is derived from the Schrödinger Equation (SE) and the Density Matrix Function (DMF). The SE defines the wavefunction of an electron that has a certain energy and moves though a given potential profile. The DMF is an operator that correlates the wavefunction at different points in the real space and weighs it across the wave vector space by a Fermi-Dirac Distribution (FDD). By combining the SE and the DMF, and then applying a Fourier Transform (FT), the WTE is obtained [13]. The solution of the WTE is called the Wigner Function (WF), which is a distribution of the wavefunctions across the physical and wave vector spaces.

In our previous work, we observe that implementing WTE solvers can be problematic. One key difficulty is in applying adequate boundary conditions. When any differential equation is solved numerically, the solution in the neighborhood of the system boundaries is expected to be consistent with the boundary conditions themselves. In the case of the WTE, when applying the boundary WF profiles commonly used in literature [5], large discrepancies appear instead. Our ongoing goal is to determine whether these discrepancies are caused by the application of inadequate numerical parameters to the WTE solver, such as insufficient grid densities, or whether it is the boundary conditions that are incorrect. In order to investigate these issues, we attempt a different approach in [14], [15], where we compute the WF directly from the wavefunctions, rather than by solving the WTE; we apply this method to single- and double-barrier structures. In this paper we go further with this approach, as we investigate specifically how the WF at the system boundaries varies with the barrier height.

II. Implementation details

We begin by casting a number of wavefunctions into the devices. By default, 250 wave vectors are cast into each contact, and they are linearly spaced over a 1 eV energy range from the conduction band upwards. For each incident wave vector, we compute the wavefunctions across the whole structure by solving the SE analytically with the transfer matrix method; we show in [14] that an analytical solution is needed to ensure accuracy in the DMF computation for the purposes of this study, though it does not allow for self-consistency and requires a flat potential in the contact, barrier and well regions. Afterwards, we compute the DMF at nodes located on a two-dimensional grid in real space (x, y) by numerically evaluating the following expression:

$$\begin{aligned}&\rho\left(x+\frac{y}{2},x-\frac{y}{2}\right)\\&=\frac{2m^*_{\text{Emitter}}k_BT}{h^2}\int_0^\infty \psi\left(x+\frac{y}{2}\right)\overline{\psi\left(x-\frac{y}{2}\right)}f_{\text{FD}}(E(k))\,dk\\&+\frac{2m^*_{\text{Collector}}k_BT}{h^2}\int_{-\infty}^0 \psi(x+\frac{y}{2})\overline{\psi\left(x-\frac{y}{2}\right)}f_{\text{FD}}(E(k))\,dk\end{aligned} \tag{1}$$

In this formula, ψ designates the wavefunction, h the Planck constant, k_B the Boltzmann constant, T the temperature, which is set at 300 K in all simulations, and m^* the carriers' effective mass at either contact. The first integral term in the formula represents wavefunctions incident at the structure's emitter contact, with a positive wave vector k; the second term represents wavefunctions incident at the collector, with a negative vector. The wavefunctions are weighed by a FDD integrated over transverse momenta:

$$f_{\text{FD}}=\ln\left[1+\exp\left(-\frac{E(k)-E_F}{k_BT}\right)\right] \tag{2}$$

$$E(k)=\frac{\hbar^2k^2}{2m^*_{\text{Contact}}}$$

978-1-4244-7700-5/10/$26.00 ©2010 IEEE

E is the energy of carriers incident at either the emitter or collector contacts, when a parabolic conduction band is assumed; E_F is the Fermi energy at a given contact, which is assumed to be at the same level as the conduction band, and $\hbar$ is the reduced Planck constant. The reason why a limit of 1 eV is set on the wavefunctions' energy is that beyond this value the FDD falls below the standard double IEEE 754 precision used in the computations: higher-energy wavefunctions are thus negligible to machine precision when evaluating the DMF integral. The WF is finally computed by applying a FT to the DMF:

$$f_W = \int_{-\infty}^{\infty} e^{-iky} \rho \left(x + \frac{y}{2}, x - \frac{y}{2} \right) \, dk \qquad (3)$$

In this formula, f_W is the WF. In the numerical implementation, the condition $\Delta y = 2\Delta x$ is applied to the x and y grids in real space in (1) and (3), where Δx and Δy are the respective grid spacings; this makes it possible to use the same nodes for the x and y grids. The wave vector space is meshed with two distinct k grids: one is for the wavefunctions and is used in the evaluation of the DMF integral in (1), while the other is used in the FT computation in (3). This paper focuses on the mesh properties of the first one, with the aim to measure the minimum number of wave vectors that is necessary to accurately compute the DMF and ultimately the WF. The second wave vector grid is kept unchanged in the various simulations presented here; it spans a range of 20 nm^{-1} on each side of the origin and has a resolution of 5000 nodes. Because the y and k grids in (3) mesh a FT, they must satisfy certain conditions in order to avoid aliasing: they have the same number of nodes N and their respective spacings are related by $\Delta y = 2\pi/(\Delta k\, N)$ [5].

We apply material parameters that are representative of silicon- and III-V-based structures: the default barrier heights are 1.5 eV and 0.3 eV respectively; the relative masses are uniform across the entire devices and are 0.5 and 0.067 respectively. In single-barrier structures, the default barrier thickness is 7 nm, in order to enhance high-barrier effects; in double-barrier ones, the barrier thickness is 1 nm and the well length is 3 nm. Contacts are 30 nm long in silicon devices and 60 nm long in III-V ones. The structures simulated in this paper are unbiased, with the conduction band level being aligned at the emitter and collector contacts.

In our previous work, when we compute the WF with the method presented above, we observe that the WF at the contacts is proportional to a FDD within working double numerical precision; this is consistent with the boundary conditions applied in literature to WTE solvers. However, in some configurations, the WF follows the FDD profile only up to a certain level, and then moves away from it and forms lobes. For instance, in the case of double-barrier structures, we find that lobes are not caused by numerical inaccuracies, but rather by the particular shape of the DMF, which we attribute to the transmission spikes that occur when the well resonates [15]. In this paper, we present novel results describing the effect of lowering the barrier height from its default value.

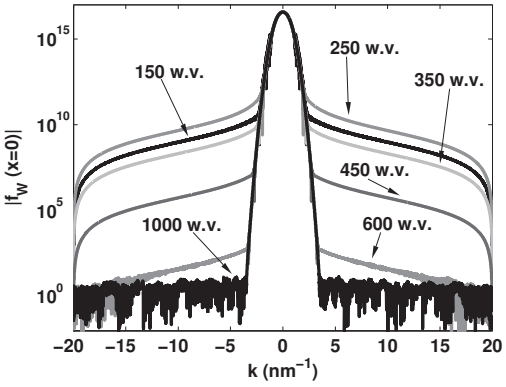

Fig. 1. Variation in the lobe height with the wave vector count for a single-barrier silicon structure with a height of 0.1 eV.

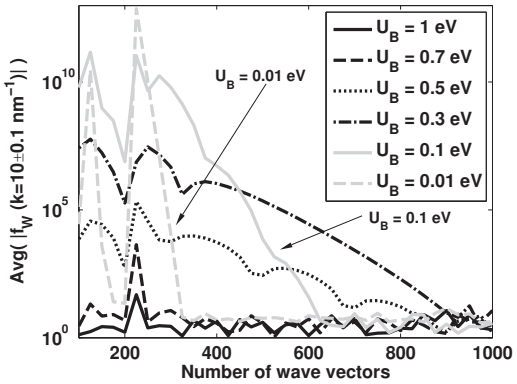

Fig. 2. Lobe height vs. wave vector count at different barrier heights U_B in a single-barrier silicon structure. The lobe height is taken as the average of ten linearly-spaced absolute WF values between k = 9.9 and 10.1 nm^{-1} (cf. Fig. 1).

III. Single-barrier structures

In general, default numerical parameters are sufficient to eliminate lobes in both silicon and III-V single barrier structures. However, as the barrier height is lowered, lobes might appear, and it becomes necessary to increase the wave vector count to eliminate them. This phenomenon is especially evident in silicon structures: for example, Fig. 1 displays the WF profiles obtained with a barrier height of 0.1 eV. For up to about 350 wave vectors, lobes appear at about 5 to 7 decades below the WF maximum. At higher wave vector counts, the lobe height drops off sharply. Finally, if about 1000 wave vectors are cast, the lobes disappear into the numerical noise level at about 15 orders of magnitude below the WF peak.

Fig. 2 summarizes how this behavior is affected by the barrier height in a silicon structure. The plots show the lobe height around $k = 10$ nm^{-1} at different wave vector counts. As Fig. 1 shows, when this height drops to a numerical value of about 10, it means that the lobes have fallen below the numerical noise threshold. This is indeed the case for a barrier height of 1 eV at about all wave vector counts. As the barrier height decreases, some interesting patterns are observed. First,

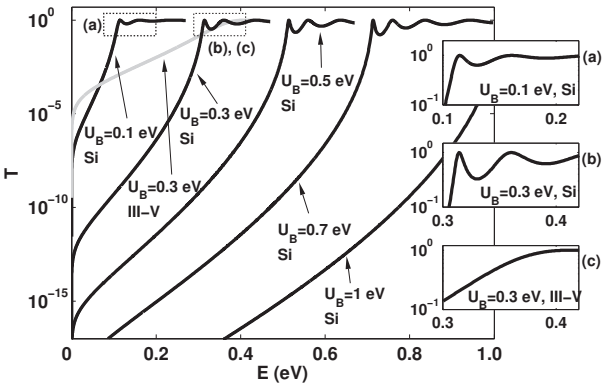

Fig. 3. Transmission coefficient vs. the energy of the wavefunctions at different barrier heights U_B in a single-barrier silicon structure. The transition regions where the structures begin to transmit the wavefunctions are magnified in the insets for barrier heights of 0.1 and 0.3 eV for silicon and 0.3 eV for III-V materials. The axes have the same scale in the three insets.

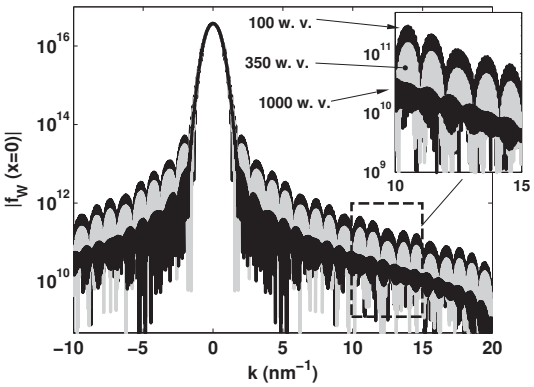

Fig. 4. Variation in the lobe height with the wave vector count for a double-barrier silicon structure with a height of 0.5 eV.

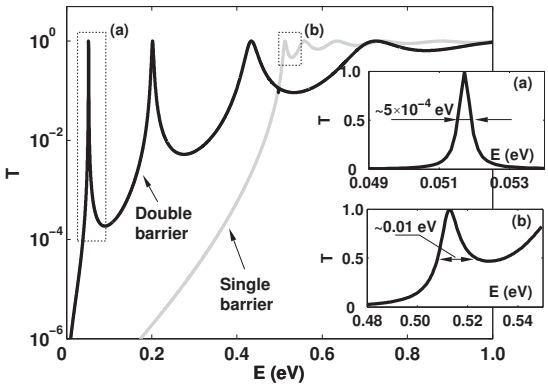

Fig. 5. Comparison of the transmission spectrum of a single- and a double-barrier structure with a barrier height of 0.5 eV. The insets magnify the first peak in the two spectra.

when the number of wavefunctions is less than 300, the height of the lobes rises quite consistently. Secondly, the number of wave vectors at which the lobes disappear rises for barrier heights greater than 0.3 eV, and then drops again. In the case of III-V structures, lobes only appear at wave vector counts equal to or lower than 100 at the default barrier height of 0.3 eV or lower.

Fig. 3 can give some insight into the behavior observed on silicon structures. It displays the transmission coefficient versus the wavefunction energy for silicon at different barrier heights. With barriers of 1 eV and higher, the structure is almost totally reflective. As the barrier lowers, the structure becomes transmissive in the upper wave vector range; also, in the reflective-to-transmissive transition region, the transmission coefficient oscillates. Because wavefunctions are weighed by a FDD, the transition region acquires an increasing weight when it shifts to smaller energy values as the barrier decreases: it becomes therefore important to have a finer wave vector meshing to capture the oscillations. However, as the barrier keeps falling, and specifically by the time it drops below 0.3 eV, the oscillations become less pronounced, and a smaller wave vector meshing is required.

In the case of III-V structures, the transition between the reflective and transmissive regions in the spectrum shows much smaller oscillations than for silicon. The structure has in fact a higher transmission coefficient because of the lower carrier mass. Also due to the lower mass, the FDD profile is narrower than that of silicon, thus the transition region has a comparatively lesser weight in the DMF integral at the same barrier height. These observations can explain why no lobes appear in a III-V WF profile, as opposed to a silicon one.

IV. Double-barrier structures

Fig. 4 shows the lobes generated in a silicon structure with a barrier of 0.5 eV. Contrarily to what observed on a single-barrier structure, the lobes remain at a constant height as the wave vector count increases. Our work in [15] shows that this is not due to the wave vector mesh, but rather to the combined effect of two factors: the first has to do with the y-axis of the DMF being meshed over a range that is too small, and the second with the occurrence of spikes in the transmission spectrum at the resonant energy levels in the well. Fig. 5 shows that these spikes are in fact quite different from the oscillations observed in single barrier structures. The figure displays on the same plot the transmission spectra of a single- and a double-barrier silicon structure with a barrier height of 0.5 eV, and magnifies the first peak in both spectra. For one thing, this peak occurs at a smaller energy for the double-barrier configuration, and thus has a greater weigh in the DMF integral. Also, the peak for the double-barrier structure has a width at half maximum of about 5×10^{-4} eV, while that for the single barrier is over ten times wider. With higher barriers, the width difference can be even bigger by orders of magnitude. In [15], we suggest that the very narrow transmission peaks displayed by resonant double-barrier structures cause the DMF profile at either contact to fall much more slowly than for single barriers; with the current grid definition parameters, the y-axis mesh in the DMF covers a range of a few hundred nanometers, which causes the DMF to be trimmed abruptly and generates the

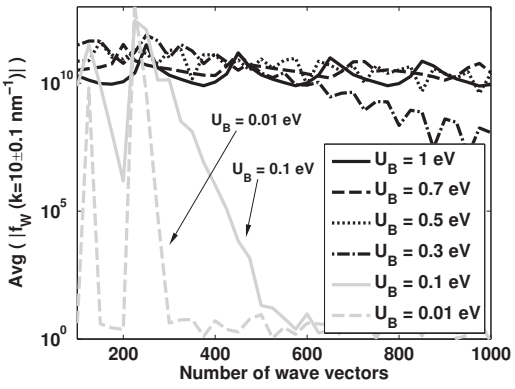

Fig. 6. Lobe height vs. wave vector count at different barrier heights U_B in a single-barrier silicon structure. The lobe height is taken as the average of ten linearly-spaced absolute WF values between k = 9.9 and 10.1 nm^{-1}.

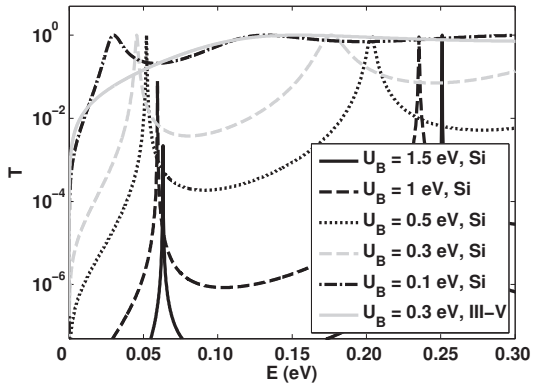

Fig. 7. Transmission coefficient vs. the energy of the wavefunctions at different barrier heights U_B in a double-barrier silicon structure.

lobes on the WF when the FT is computed. In order for the trimming to take place when the DMF profile has dropped to a numerically negligible level, the range of the y-grid would have to be extended by orders of magnitude, which is not feasible due to constraints in computational resources. This effect is not present in the single-barrier case, because the greater width of the transmission peaks causes the DMF profile to fall more rapidly.

Fig. 6 shows however that, as the barrier height decreases, the lobes do ultimately drop below numerical noise level, though this might require raising the wave vector count. Fig. 7 can provide an explanation for this behavior. It shows the transmission spectra at different barrier heights and illustrates how the width of the peaks increases as the barrier falls. Ultimately, at barrier heights below 0.5 eV, the well loses its distinct resonant properties, the width of the transmission peaks increases and the transmission spectrum, the DMF and the WF match those observed in equivalent single-barrier configurations.

On double-barrier III-V structures, the lobes in the WF profile are in general lower by about 5 decades than those on silicon devices. Fig. 7 shows that the transmission spectrum for a double-barrier III-V structure with a barrier height of 0.3 eV is in fact much smoother than the one for an equivalent silicon configuration. The effect of reducing the effective mass is thus similar to lowering the barrier height.

V. Conclusion

This study gives us insight into the minimum number of wave vectors required to compute the WF to a given numerical precision. As the barrier height is lowered, lobes appear in the WF profile at the contact boundaries. In single-barrier devices, these lobes are due to insufficient wave vector meshing at the transition between the reflective and transmissive regions, and can be removed by increasing the wave vector count. A mesh of up to 1000 nodes in each wave vector direction might be required. Lowering the effective mass can also eliminate the lobes. In contrast, in double-barrier devices with high barriers, the lobes cannot be significantly reduced by increasing the wave vector count. This implies very severe limitations in using the WTE to simulate resonant devices.

References

[1] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Phys. Rev.*, vol. 40, no. 5, pp. 749–759, 1932.

[2] N. Kluksdahl, W. Pötz, R. U., and F. D.K., "Wigner function study of a double quantum barrier resonant tunnelling diode," *Superlattices and Microstructures*, vol. 3, no. 1, pp. 41–45, 1987.

[3] N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, "Self-consistent study of the resonant-tunneling diode," *Phys. Rev. B*, vol. 39, no. 11, pp. 7720–7735, 1989.

[4] W. R. Frensley, "Transient response of a tunneling device obtained from the Wigner function," *Phys. Rev. Lett.*, vol. 57, no. 22, pp. 2853–2856, 1986.

[5] W. R. Frensley, "Wigner-function model of a resonant-tunneling semiconductor device," *Phys. Rev. B*, vol. 36, no. 3, pp. 1570–1580, 1987.

[6] W. R. Frensley, "Boundary conditions for open quantum systems driven far from equilibrium," *Rev. Mod. Phys.*, vol. 62, no. 3, pp. 745–791, 1990.

[7] K. L. Jensen and F. A. Buot, "Numerical simulation of transient response and resonant-tunneling characteristics of double-barrier semiconductor structures as a function of experimental parameters," *J. Appl. Phys.*, vol. 65, no. 12, pp. 5248–5250, 1989.

[8] K. L. Jensen and F. A. Buot, "The effects of scattering on current-voltage characteristics, transient response, and particle trajectories in the numerical simulation of resonant tunneling diodes," *J. Appl. Phys.*, vol. 67, no. 12, pp. 7602–7607, 1990.

[9] K. L. Jensen and A. K. Ganguly, "Numerical simulation of field emission and tunneling: A comparison of the Wigner function and transmission coefficient approaches," *J. Appl. Phys.*, vol. 73, no. 9, pp. 4409–4427, 1993.

[10] H. Tsuchiya, M. Ogawa, and T. Miyoshi, "Simulation of quantum transport in quantum devices with spatially varying effective mass," *IEEE Trans. Electron Devices*, vol. 38, no. 6, pp. 1246–1252, 1991.

[11] B. A. Biegel and J. D. Plummer, "Comparison of self-consistency iteration options for the Wigner function method of quantum device simulation," *Phys. Rev. B*, vol. 54, no. 11, pp. 8070–8082, 1996.

[12] Y. Yamada, H. Tsuchiya, and M. Ogawa, "Quantum transport simulation of silicon-nanowire transistors based on direct solution approach of the Wigner transport equation," *IEEE Trans. Electron Devices*, vol. 56, no. 7, pp. 1396–1401, 2009.

[13] A. Jüngel, *Transport equations for semiconductors*. Berlin: Springer, 2009.

[14] A. Savio and A. Poncet, "Study of the Wigner function computed by solving the Schrödinger equation," in *Proc. 4th International Conference on Quantum, Nano and Micro Technologies (ICQNM'10)*, 2010, pp. 59–64.

[15] A. Savio and A. Poncet, "Study of the Wigner function at the device boundaries in 1D single- and double-barrier structures," unpublished.