Efficient Calculation of Quasi-Bound State Tunneling in CMOS Devices

M. Karner*, A. Gehring†, H. Kosina*, and S. Selberherr*
* Institute for Microelectronics, TU Wien, Gußhausstraße 27–29/E360, 1040 Wien, Austria
Phone: +43-1-58801/36016, Fax: +43-1-58801/36099, E-mail: karner@iue.tuwien.ac.at
†AMD Saxony, Wilschdorfer Landstrasse 101, D-01109 Dresden, Germany

Abstract— **For an accurate description of direct tunneling in CMOS devices under inversion conditions lifetime based approaches rely on the precise determination of quasi-bound states (QBS).**

We study the calculation of QBS in resonant tunneling diode (RTD) structures and MOS inversion layers by the perfectly matched layer (PML) method. Introducing a complex coordinate stretching allows to apply artifical absorbing layers at the boundaries. The QBS appear as the eigenvalues of a linear non-Hermitian Hamiltonian where the QBS lifetimes are directly related to the imaginary part of the eigenvalues. The PML formalism has been compared to the established quantum transmitting boundary method where a computationally demanding scanning procedure yields the desired lifetimes. The PML method proves as an elegant, numerical stable, and efficient method to calculate QBS lifetimes.

I. INTRODUCTION

The continuous progress in the development of semiconductor devices within the last years goes hand in hand with down-scaling the device feature size. Since the device feature sizes approach the wave length of free electrons, the influence of quantum mechanical effects gain importance. Especially quantum mechanical tunneling has significant effects on the characteristics of state-of-the-art microelectronic devices. A major, if not the dominant, source of tunneling electrons in the inversion layers of MOS-structures and resonant-tunneling electrons is represented by quasi-bound states (QBS) [1] [2]. Since continuum based models, like the frequently used TSU-ESAKI formula [3], do not take account these effects, a life time based approach which requires the estimation of the QBS, is compulsory.

II. CALCULATION OF QBS

Whenever electrons are confined, or partially confined in movement, this gives rise to bound and quasi bound states. For an accurate description of the tunneling process, models which takes this into account are necessary.

Depending on the shape of the potential well, type-one QBS, which are related to energies where the system is fully open, and type-two QBS can occur as this is pointed out in Fig. 1.

Within our simulation framework, the QBS are obtained from the single particle, time-independent effective mass

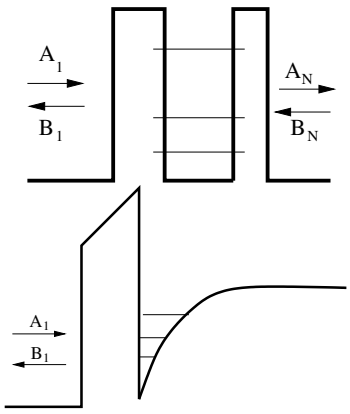

Figure 1: The upper figure shows type-one QBS which occur in an open system while in the lower figure the system is half open at the energy levels of the QBS which are therefore of type-two.

SCHRÖDINGER equation:

$$-\frac{\hbar^2}{2}\nabla\cdot\left(\tilde{m}^{-1}\nabla\Psi(\mathbf{x})\right)+V(\mathbf{x})\Psi(\mathbf{x})=\mathcal{E}\Psi(\mathbf{x})\,. \qquad (1)$$

In a first approximation the energy levels of the QBS can be estimated by the eigenvalues of the closed system Hamiltonian. Since closed boundaries are assumed, no information about the broadening and the associated QBS life times is available. Also, bound states cannot carry any current, since they fulfill the relation:

$$\Psi\nabla\Psi^*-\Psi^*\nabla\Psi=0\,. \qquad (2)$$

A semi-classical approximation based on corrected closed-boundary eigenvalues using a classical formulation of the escape time (life time) has also been reported [4] but using the closed-boundary eigenvalues for the calculation of open-boundary QBS lifetimes seems to be questionable.

A. QUANTUM TRANSMITTING BOUNDARY METHOD

A widely used method to apply open boundary conditions to (1) is the quantum transmitting boundary method (QTBM) [5]. Starting point is the discretized Hamiltonian of the closed system, which is augmented with traveling waves at the boundary points of the simulation region in order to obtain open boundary conditions. Since the entries of the Hamiltonian

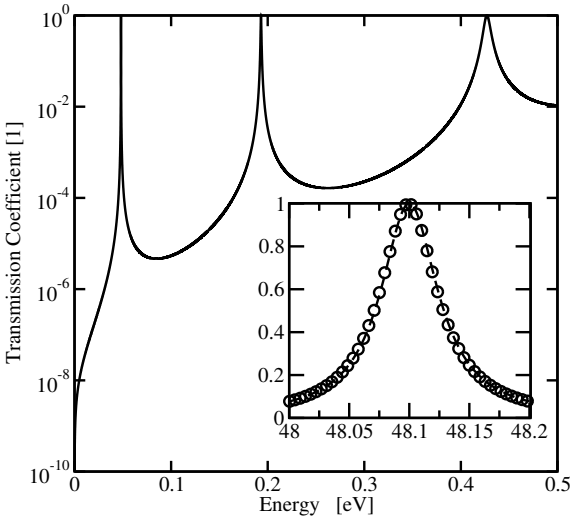

Figure 2: The transmission coefficient of the RTD computed by means of the QTB method. The inset compares the first resonance with a Lorentzian function.

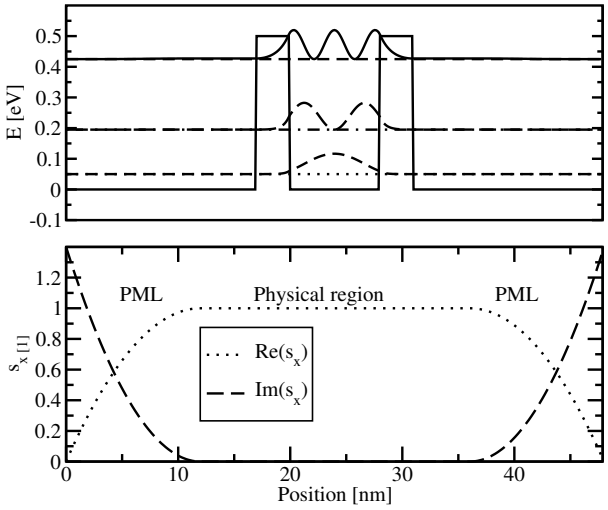

Figure 3: The upper figure shows the energy levels and the squared wavefunctions of the RTD as obtained by the PML method. The complex stretching function is depicted in the lower figure.

at the associated points depends exponential by on the energy, the Hamiltonian of the system becomes non-linear.

Therefore, the lifetime broadening of the QBS has to be determined by a computationally intensive scanning procedure. For type-one QBS which might occur in an unbiased RTD structure as sketched in Fig. 1, the QBS follow from the resonances of the transmission coefficient.

The transmission coefficient features a Lorentzian shape at the resonances as shown in Fig. 2. The full-width half maximum (FWHM) is related to the QBS lifetime by

$$\tau_i = \hbar/\mathrm{FWHM}_i \, . \tag{3}$$

However, the situation is more complex for the MOS-inversion layer. Since it is a half open system for the energy range of interest (containing the type-two QBS) the transmission coefficient is zero. In these situations the traditional approach is based on a computationally intensive scanning of the derivative of the phase of the reflection coefficient [6]. A numerically more stable method which detects the peaks of the resonance coefficient has been outlined in [7].

While these methods are feasible for situations with weak confinement (low or thin barriers), they are hardly applicable to energy barriers of MOS capacitors. Energy resolutions in the peV regime are necessary to accurately resolve the full-width half maximum (FWHM) value necessary to calculate the QBS lifetime, which is infeasible for everyday application.

B. PERFECTLY MATCHED LAYER METHOD

Recently, a method based on absorbing boundary conditions (called the Perfectly Matched Layer (PML) method) for SCHRÖDINGER's equation has been applied for band structure calculations in III-V heterostructure devices [8].

Within this work the PML formalism, which is often used in electromagnetics, has been applied to determine the lifetime broadening of QBS in a resonant tunneling diode structure and MOS inversion layers.

The basic principle is to add non-physical absorbing layers at the boundary of the simulation region (physical region). This procedure prevents reflections at the boundary of the physical region. The artifical absorbing layers allow the application of Dirichlet boundary conditions, and the QBS are determined by the eigenvalues of the non-Hermitian, linear Hamiltonian of the system. The resulting eigenenergies and the corresponding wavefunctions for the RTD described in [8] are displayed in Fig. 3.

The absorbing property of the PML region is achieved by introducing stretched coordinates

$$\tilde{x} = \int_0^x s_x(\tau)\,\mathrm{d}\tau \tag{4}$$

in (1). The evaluation of the nabla operator ∇ in one dimension yields:

$$\frac{\partial}{\partial \tilde{x}} = \frac{1}{s_x(x)}\frac{\partial}{\partial x} \tag{5}$$

Within the PML region, the stretching function $s_x(x)$ is given as $s_x(x) = 1 + (\alpha + \imath\beta)x^n$, with $\alpha = 1$, $\beta = 1.4$, and $n = 2$, while it is unity in the physical region as displayed in Fig. 3.

Adding absorbing layers at the boundary of the physical simulation region, the Hamiltonian becomes non-Hermitian and admits complex eigenvalues $\mathcal{E} = \mathcal{E}_r + \imath\mathcal{E}_i$, where the QBS lifetimes are related to the imaginary parts of the eigenvalue as

$$\tau_i = \hbar/2\mathcal{E}_i. \tag{6}$$

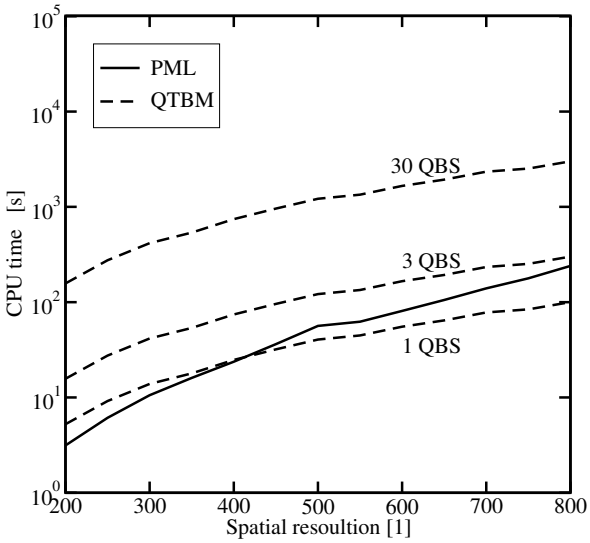

Figure 4: Comparison of CPU demand for the PML and QTB methods.

In contrast to the QTBM, the Hamiltonian of the system still remains linear. Furthermore, all QBS are calculated in one step and no iteration or scanning procedure is needed. Assuming a constant potential $V(z)$ within the PML regions, the wavefunctions can be written as a plane wave $\Psi(x) = \Psi_0 \exp(\imath \tilde{k_x} x)$ with the wave vector $\tilde{k_x} = k_x/s_x$. Considering two points in the PML region x_1, $x_2 = x_1 + \mathrm{d}x$ the wave vector at the point x_2 can be approximated as

$$k_x(x_2) \approx \frac{s_x(x_2)}{s_x(x_1)} k_x(x_1) = (1 + (\alpha + \imath\beta)\,\mathrm{d}x)\ . \qquad (7)$$

The parameter α scales the phase velocity of the plane wave, while β acts as a damping parameter. Since this damping coefficient is greater than zero within the absorbing region the envelope of the wavefunctions is pinned to zero as shown in Fig. 3. These parameters, as well as the thickness of the absorbing layer can be varied over a wide range with virtually no influence on the results as long as there are no reflections at the boundaries. However, for this goal the complex stretching function and its first derivative have to be continuous as shown in Fig. 3.

Using QTBM or assuming closed boundary conditions yields a superposition of two moving plane waves in opposite directions. In contrast, when using PML, there are no reflected waves. The wavefunction is a plane traveling wave with a constant envelope function. The QBS, however, are reproduced correctly.

C. COMPARISON

The PML formalism was applied to the AlGaAs-GaAs RTD studied in [8]. The resulting wave functions have been shown in Fig. 3. It also depicts the shape of the stretching function. Perfect agreement between PML and QTBM as well as with the results in [8] have been achieved as pointed out in Table. 1.

PML	E_{real} [meV]	E_{imag} [eV]	τ_{l} [ps]
1	48.1	-2.77×10^{-5}	11.84
2	193.1	-3.75×10^{-4}	0.876
3	426.9	-3.56×10^{-3}	0.0924

QTBM	E_{real} [meV]	FWHM [eV]	τ_{l} [ps]
1	48.1	5.0×10^{-5}	11.4
2	193.0	7.5×10^{-4}	0.877
3	427.0	7.0×10^{-3}	0.092

Closed	E_{real} [meV]
1	48.1
2	193.6
3	423.0

Table 1: The energy levels and the lifetime in the RTD structure. The PML method in the upper table is compared with the QTB method in the middle table. The eigenvalues of the closed system are also given.

To further justify the use of the PML method we compared the computational effort of the PML and QTBM approaches for the RTD structure. Fig. 4 shows the CPU time necessary to calculate 1, 3, and 30 quasi-bound states with the QTB and PML methods as a function of the spatial resolution. For the QTBM, an equidistant grid in energy space was used to determine the lifetime broadening of the QBS.

The PML which delivers all QBS at once, shows a stronger dependence on the spatial resolution. However, the demand on CPU time is almost always lower as compared to the QTB method, especially for stronger confined states as encountered in MOS structures.

III. CALCULATION OF GATE LEAKAGE CURRENTS

For the investigation of gate current leakage in MOS transistors the conduction band edge has been acquired from a self-consistent quantum-mechanical SCHRÖDINGER-POISSON solver. As a post-processing step, the QBS lifetimes have been evaluated using the PML formalism. Based on an accurate computation of the QBS lifetimes, the tunneling current can be calculated according to

$$J_{\mathrm{2D}} = \frac{\mathrm{k_B} T \mathrm{q}}{\pi \hbar^2} \sum_{i,\nu} \frac{g_\nu m_\parallel}{\tau_\nu(\mathcal{E}_{\nu,i}(m_\mathrm{q}))} \ln\left(1 + \exp\left(\frac{\mathcal{E}_\mathrm{F} - \mathcal{E}_{\nu,i}}{\mathrm{k_B} T}\right)\right)$$

where g_ν denotes the valley degeneracy, $m_\parallel$ the parallel mass, and m_q the quantization masses ($g = 2$: $m_\parallel = m_\mathrm{t}$, $m_\mathrm{q} = m_\mathrm{l}$ and $g = 4$: $m_\parallel = \sqrt{m_\mathrm{l} m_\mathrm{t}}$, $m_\mathrm{q} = m_\mathrm{t}$), and $\tau_\nu(\mathcal{E}_{\nu,i})$ is the lifetime of the quasi-bound state $\mathcal{E}_{\nu,i}$. For [100] silicon $\mathrm{m_l}$ = 0.916 * $\mathrm{m_0}$ and $\mathrm{m_t}$ = 0.196 * $\mathrm{m_0}$ where $\mathrm{m_0}$ is the mass of a free electron as pointed out in [9]. Fig. 5 shows the investigated MOS structure and some of the QBS wave functions considering the transversal mass.

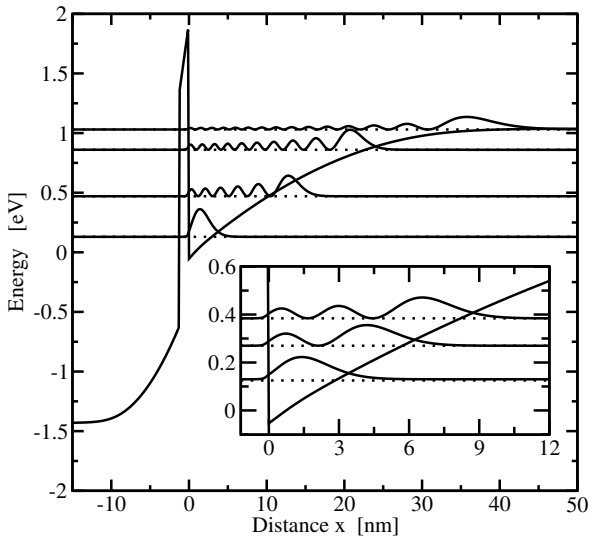

Figure 5: Some QBS in a MOS capacitor (n^+-insulator-p^+) with a doping of $N_{\rm A} = 1 \times 10^{17}\,{\rm cm}^{-3}$, dielectric thickness $t_{\rm diel} = 1.2\,{\rm nm}$, and a gate bias of 1.5 V.

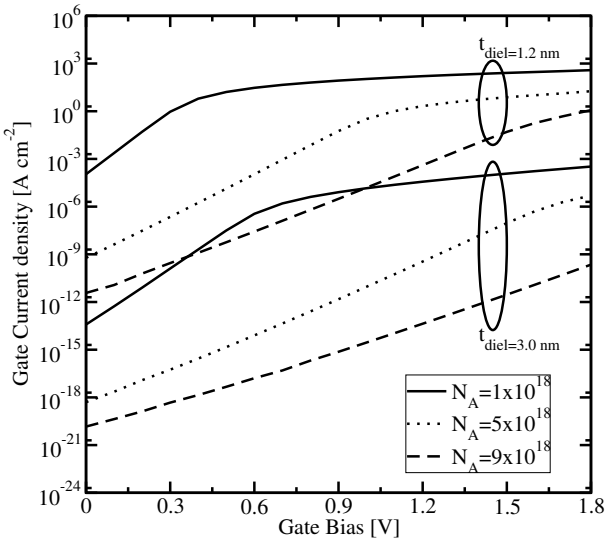

Figure 6: The gate current density for different bulk doping and dielectric thickness calculated from the QBS obtained by the PML method.

QBS	$\mathcal{E}_{\rm r}$ [eV]	N($E_{\rm r}$) [1]	$\tau_{\rm l}$ [ps]	$J_{\rm G}$ [A cm^{-2}]
1	0.14	6.3×10^{-3}	210	1.7×10^{6}
2	0.27	2.5×10^{-5}	160	8.6×10^{3}
3	0.38	3.4×10^{-7}	140	1.4×10^{2}
10	0.86	3.1×10^{-15}	56	3.2×10^{-6}
15	1.01	5.0×10^{-18}	93	3.1×10^{-9}

Table 2: The QBS of the MOS-capacitor for a gate bias of 1.5V, lifetimes, values of the supply function, and their contribution to the gate current density.

This procedure has been used to aquire the IV-characteristics of several MOS structures. For an n-MOS device with a bulk doping of $N_{\rm A}$=$1 \times 10^{17}{\rm cm}^{-3}$ some of the extracted quasi-bound states are listed in Tab. 2 together with their contribution to the total current density. The resulting gate current density is shown as a function of the dielectric thickness, doping, and the gate bias in Fig. 6.

IV. SUMMARY AND CONCLUSION

We presented a new method for the calculation of QBS lifetimes. In contrast to the traditional approach which requires a computationally very demanding scanning procedure, the QBS lifetimes appear as the complex eigenvalues of a non-Hermitian linear Hamiltonian. Since the equation which has to be solved is linear, highly efficient algorithms are available.

To compare and to calibrate the method, the new PML formalism and the established QTB method was applied to an RTD structure and perfect agreement was obtained.

Moreover, the PML approach was used to evaluate QBS in MOS inversion layers and the impact on direct tunneling through the dielectric layer which is for typical device parameters the dominant tunneling component. The PML formalism represents an efficient and numerically stable method to determine QBS. Furthermore, it is appropriate for integration in a device simulator for the investigation of direct tunneling phenomena.

ACKNOWLEDGMENT

This work has been partly supported by the European Commission, project SINANO, IST 506844 and from the special research project IR-ON (F25).

REFERENCES

[1] A. Gehring and S. Selberherr, in *Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices* (München, 2004), pp. 25–28.

[2] E. Cassan, P. Dollfus, S. Galdin, and P. Hesto, IEEE Trans.Electron Devices **48**, 715 (2001).

[3] R. Tsu and L. Esaki, Appl.Phys.Lett. **22**, 562 (1973).

[4] A. Dalla Serra *et al.*, IEEE Trans.Electron Devices **48**, 1811 (2001).

[5] C. L. Fernando and W. R. Frensley, J.Appl.Phys. **76**, 2881 (1994).

[6] E. Cassan, J.Appl.Phys. **87**, 7931 (2000).

[7] R. Clerc, A. Spinelli, G. Ghibaudo, and G. Pananakakis, J.Appl.Phys. **91**, 1400 (2002).

[8] S. Odermatt, M. Luisier, and B. Witzigmann, J.Appl.Phys. **97**, 046104 (2005).

[9] F. Stern, Phys.Rev.B **5**, 4891 (1972).