



Investigation of Quantum Tunnelling Effects in Nanoscale Semiconductor Devices

¹Leocadia Rafizadeh

¹Department of Computer Science and Engineering, Mindoro International School of Engineering and Management, Philippines

¹leocadia.rafizadeh@misem-ph.org

Peer Review Information

Submission: 20 April 2021

Revision: 06 May 2021

Acceptance: 21 May 2021

Keywords

quantum tunneling,
nanoscale MOSFET, WKB
approximation, NEGF,
quantum transport

Abstract

As semiconductor dimensions approach the sub-10-nm regime, the wave nature of carriers becomes a primary determinant of device behavior. This paper investigates direct quantum tunneling in nanoscale semiconductor devices using a methodology that combines the Wentzel-Kramers-Brillouin (WKB) approximation for rapid tunneling estimates with the Landauer and nonequilibrium Green's function (NEGF) formalisms as the transport framework. A double-gate silicon MOSFET is used as the reference structure, and the dependence of source-to-drain tunneling on channel length, barrier height, and transport effective mass is examined. For a 0.25 eV off-state barrier and an electron transport mass of 0.19 m_0 , the calculated WKB transmission increases from 2.30×10^{-12} at 12 nm to 1.32×10^{-4} at 4 nm, corresponding to an increase of approximately 5.7×10^7 in the tunneling leakage index. Lower effective mass further increases transmission, demonstrating the trade-off between high carrier velocity and leakage control in aggressively scaled channels. The results support the use of quantum-transport-aware design, material selection, and electrostatic barrier engineering as essential strategies for nanoscale transistor development.

Introduction

Continued transistor scaling has moved semiconductor electronics from a regime in which carrier motion can be treated largely by classical drift-diffusion theory to one in which quantum confinement, ballistic transport, and tunneling must be included explicitly. When the channel length and barrier width become comparable to the electron de Broglie wavelength, carriers can penetrate classically forbidden regions instead of being completely blocked by the channel potential barrier. In an ultra-short MOSFET this direct source-to-drain tunneling (DSDT) contributes to the off-state leakage current, weakens gate control, and can ultimately limit further reduction of the physical gate length [4], [10]. The same underlying tunneling physics can also be exploited

deliberately, as in tunnel field-effect transistors (TFETs), where band-to-band tunneling is used to obtain steep switching characteristics and potentially reduce supply voltage [5].

Accurate modeling of nanoscale devices therefore requires a transport description that preserves the wave nature of electrons and treats open contacts correctly. The nonequilibrium Green's function formalism provides such a framework and has become a standard basis for coherent and scattering-aware quantum transport calculations [1]-[3]. NEGF can be coupled self-consistently with Poisson's equation, while atomistic or effective-mass Hamiltonians can be selected according to the required level of physical detail. Large-scale simulation platforms such as NEMO5

demonstrate that quantum electrostatics, band structure, and transport can be integrated within a multiscale numerical environment [6]. At the same time, the WKB approximation remains valuable for rapid interpretation because it makes the exponential dependence of tunneling on barrier width, barrier height, and effective mass transparent [4].

The objective of this methodology paper is to investigate the principal factors controlling tunneling in nanoscale semiconductor devices and to present a compact simulation procedure suitable for an academic study. A double-gate silicon MOSFET is selected because strong gate coupling minimizes conventional short-channel effects and allows the effect of direct quantum penetration to be isolated. The study first defines the device electrostatics and an off-state barrier, then calculates the WKB transmission over channel lengths from 12 to 4 nm. The transport-effective-mass dependence is also evaluated to illustrate why low-mass channel materials, although attractive for high injection velocity, may experience stronger source-to-drain tunneling. The WKB calculations are interpreted within the Landauer/NEGF framework so that the simplified results remain connected to more rigorous quantum transport theory.

Literature Review

The theoretical foundation for modern nanoscale quantum transport modeling was strengthened by Lake et al. [1], who formulated nonequilibrium Green's function transport in a localized orbital basis and treated charging, open contacts, band structure, and scattering within a unified approach. For transistor-scale problems, direct real-space solutions can be computationally expensive. Venugopal et al. [2] demonstrated that mode-space representations can reproduce quantum transport in thin-body nanoscale transistors with much lower computational cost when the transverse modes are weakly coupled. Luisier et al. [3] later analyzed coupled-mode effects and showed that nonuniform confinement can require careful treatment of mode coupling for accurate charge and current calculations. These studies established NEGF-Poisson simulation as a key reference method for nanoscale transistor analysis.

As physical gate lengths decreased, source-to-drain tunneling emerged as a specific scaling concern. Raffay et al. [4] examined silicon, strained silicon, germanium, and GaAs double-gate devices and demonstrated that the tunneling probability depends exponentially on transport

effective mass. Their analysis highlighted an important materials trade-off: a smaller effective mass can improve ballistic drive current while simultaneously increasing DSDT leakage. This observation is directly relevant to the present work, where effective mass is treated as a controlled parameter in the WKB model. In parallel, Ionescu and Riel [5] reviewed TFET operation and showed how quantum tunneling can be used constructively for energy-efficient switching, indicating that tunneling is not merely a parasitic phenomenon but a mechanism that can be engineered.

The expansion of computational nanoelectronics also improved the practicality of rigorous transport calculations. Steiger et al. [6] introduced NEMO5 as a parallel multiscale tool capable of combining strain, electronic structure, Poisson solutions, and NEGF transport. Ryu [7] emphasized that quantum confinement and scattering mechanisms become increasingly important in sub-10-nm silicon nanowires, illustrating the need to move beyond bulk mobility concepts as channel cross-sections shrink. For alternative channel materials, Quhe et al. [8] used first-principles quantum transport simulations to study sub-5-nm phosphorene transistors, while Pan et al. [9] applied a similar approach to monolayer silicene MOSFETs. Both studies demonstrate that atomically thin channels can provide excellent electrostatics, but their performance remains strongly dependent on quantum transport and material band structure.

Research up to 2021 increasingly focused on extending electrostatic control while acknowledging the tunneling limit. Pandey et al. [10] specifically modeled DSDT in extremely scaled silicon transistors with negative-capacitance gate stacks and reported that improved electrostatics can suppress tunneling over part of the scaling range, although severe tunneling eventually limits the benefit. A comprehensive 2021 review by Quhe et al. [11] summarized theory and experiments for sub-10-nm two-dimensional transistors and emphasized the role of effective mass, atomistic quantum transport, and tunneling-aware design. Taken together, the literature indicates that channel length, barrier shape, effective mass, electrostatic control, and the selected quantum transport model are the dominant variables for studying tunneling at nanoscale dimensions.

Methodology

1. Device Structure and Assumptions

The reference device is an n-type, symmetric double-gate silicon MOSFET with heavily doped

source and drain regions and a thin undoped channel. The double-gate geometry is chosen to provide strong electrostatic control and to reduce ambiguity between classical short-channel leakage and direct quantum tunneling. The physical channel length L_g is swept from 12 nm to 4 nm. The silicon body thickness is fixed at 3 nm, the equivalent oxide thickness is taken as 0.7 nm, temperature is 300 K, and the drain bias

is 0.5 V. For the baseline off-state calculation, the effective conduction-band barrier above the injecting carrier energy is represented by 0.25 eV and the transport effective mass is taken as 0.19 m_0 , a value representative of a silicon transverse conduction-band mass. The barrier is treated as an effective one-dimensional barrier for the rapid WKB parameter sweep; the more complete NEGF procedure is defined below for validation.

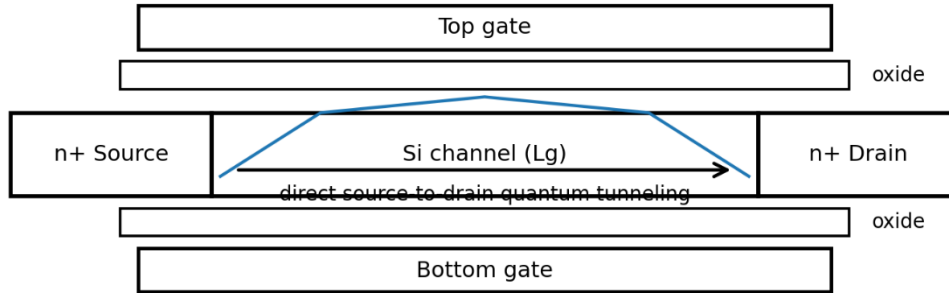


Figure 1. Reference double-gate nanoscale MOSFET and direct source-to-drain tunneling path.

Table 1. Baseline device and simulation parameters.

Parameter	Baseline value	Purpose
Channel length, L_g	4-12 nm	Tunneling-length sweep
Si body thickness	3 nm	Strong quantum confinement
Equivalent oxide thickness	0.7 nm	Double-gate electrostatic control
Drain bias, V_d	0.5 V	Finite source-drain energy window
Temperature	300 K	Room-temperature operation
Barrier height, Φ	0.25 eV	Off-state effective barrier
Transport mass, m^*	0.19 m_0	Silicon baseline; varied in sensitivity test

2. WKB Tunneling Model

The first calculation stage uses the WKB approximation because it provides a direct measure of how rapidly tunneling increases as the barrier becomes thinner. For an electron of energy E incident on a one-dimensional barrier $U(x)$, the transmission probability is approximated by [4]:

$$T(E) \approx \exp[-2 \int_{x_1}^{x_2} \sqrt{2 m^* [U(x) - E]} / \hbar dx]$$

For the baseline sweep the barrier is approximated as rectangular with effective height $\Phi = U - E = 0.25$ eV and width equal to L_g . The expression then reduces to $T \approx \exp[-2 L_g \sqrt{2 m^* \Phi} / \hbar]$. Although a real MOSFET barrier is not exactly rectangular, this approximation is useful for a methodology paper because it isolates the exponential scaling trend and agrees with the qualitative behavior reported in detailed DSDT studies [4], [10]. A second sweep varies m^* at a fixed 5 nm channel length to quantify the material sensitivity.

3. Landauer and NEGF Transport Framework

The WKB transmission is connected to device current through the Landauer picture. In a ballistic channel the current is written as:

$$I = (2q/h) \int T(E) [f_S(E) - f_D(E)] dE$$

where f_S and f_D are the Fermi-Dirac distributions of the source and drain. For a full device simulation, the transmission should be calculated from the retarded Green's function rather than imposed analytically. Following the NEGF formulations used in nanoscale transistor studies [1]-[3], the open-system retarded Green's function is:

$$GR(E) = [E I - H - \Sigma_S(E) - \Sigma_D(E)]^{-1}$$

The source and drain broadening matrices are $\Gamma_{S,D} = i(\Sigma_{S,D} - \Sigma_{S,D}^\dagger)$, and the quantum-mechanical transmission is $T(E) = \text{Tr}[\Gamma_S GR \Gamma_D GR]$. The carrier density obtained from the lesser Green's function is then returned to Poisson's equation, and the electrostatic potential is updated iteratively until the potential and charge converge. This self-

consistent NEGF-Poisson loop is the recommended high-fidelity validation stage for future work or for comparison with a device simulator such as NEMO5 [6]. In the present compact study, the WKB calculation is used for the numerical parameter sweep, while NEGF defines the rigorous transport framework against which the simplified model should be interpreted.

Step-by-Step Simulation Procedure

Step 1: Define the double-gate device geometry, temperature, oxide thickness, channel length, and source-drain bias.

Step 2: Establish the off-state conduction-band barrier and choose the transport effective mass for the channel material.

Step 3: Calculate WKB transmission T for each channel length from 12 to 4 nm.

Step 4: Repeat the calculation at $L_g = 5$ nm for $m^* = 0.25, 0.19, 0.12$, and $0.067 m_0$ to measure material sensitivity.

Step 5: Convert the tunneling trend into a leakage index by normalizing each transmission value to the 12 nm baseline.

Step 6: Interpret the transmission curves with the Landauer current relation and the NEGF transmission definition.

Step 7: For high-fidelity validation, solve Poisson and NEGF equations self-consistently and compare the energy-resolved transmission with the WKB estimate.

4. Evaluation Metrics and Validation Criteria

Three quantitative indicators are used to evaluate the tunneling behavior. The first is the energy-independent WKB transmission at the selected off-state barrier, which provides the most direct measure of wave-function penetration. The second is the logarithmic transmission, $\log_{10}(T)$, because tunneling spans many orders of magnitude and is therefore more meaningful on a logarithmic scale. The third is a

relative leakage index, defined as $T(L_g)/T(12 \text{ nm})$, which reports how strongly scaling amplifies the tunneling component without implying a complete device current model. For the effective-mass study, the same geometry and barrier height are retained so that only the mass-dependent part of the exponential is changed. This controlled comparison isolates the material contribution from electrostatic and geometric effects.

A higher-fidelity NEGF implementation should be validated using both numerical and physical checks. The Poisson-NEGF loop should be iterated until the maximum potential change between successive iterations is below approximately 10^{-5} to 10^{-4} V, while the calculated source and drain currents should agree within numerical tolerance to verify current conservation. The energy grid must be sufficiently fine to resolve sharp transmission features, and the simulated contact extensions should be long enough for the electrostatic potential to approach its reservoir value. Finally, WKB and NEGF results should be compared at the same barrier profile. Agreement is expected for slowly varying barriers, whereas deviations are anticipated near abrupt interfaces, resonant states, strong mode coupling, or atomistic band-structure effects [2], [3], [6]. These checks make the proposed workflow reproducible and clarify the domain in which the semi-analytical results can be trusted.

Results

The WKB sweep shows the expected exponential growth of tunneling as the channel barrier becomes thinner. At 12 nm, the calculated transmission is only 2.30×10^{-12} for the baseline silicon mass and barrier. Reducing L_g to 8 nm raises the transmission to 1.74×10^{-8} , while a 5 nm channel gives 1.42×10^{-5} . At 4 nm the transmission reaches 1.32×10^{-4} . Thus, reducing the tunneling length from 12 to 4 nm increases the leakage index by about 5.7×10^7 . This strong growth is consistent with the literature identifying DSDT as a major constraint for deeply scaled silicon devices [4], [10].

Table 2. Calculated WKB transmission and relative tunneling leakage index for the silicon baseline.

Lg (nm)	Transmission T	log10(T)	Relative leakage index
12	2.30e-12	-11.64	1.00e+00
10	2.00e-10	-9.70	8.70e+01
8	1.74e-08	-7.76	7.57e+03
6	1.52e-06	-5.82	6.59e+05
5	1.42e-05	-4.85	6.15e+06
4	1.32e-04	-3.88	5.74e+07

Effective mass produces a similarly strong effect because it appears inside the WKB exponential. At $L_g = 5$ nm, increasing the mass to $0.25 m_0$ reduces T to 2.74×10^{-6} , whereas the silicon baseline $m^* = 0.19 m_0$ gives 1.42×10^{-5} . Reducing m^* to $0.12 m_0$ increases T to 1.40×10^{-4} , and a low mass of $0.067 m_0$ gives 1.32×10^{-3} . Therefore, a low transport mass may

increase the tunneling probability by roughly two orders of magnitude compared with the silicon baseline at the same barrier dimensions. This result explains the material trade-off reported for alternative-channel MOSFETs [4] and reinforces the 2021 conclusion that effective mass is a critical design parameter in sub-10-nm transistors [11].

Table 3. Effective-mass sensitivity at $L_g = 5$ nm and $\Phi = 0.25$ eV.

Transport mass	WKB transmission T	Relative to Si baseline
0.250 m_0	2.74e-06	0.19 x
0.190 m_0	1.42e-05	1.00 x
0.120 m_0	1.40e-04	9.90 x
0.067 m_0	1.32e-03	93.24 x

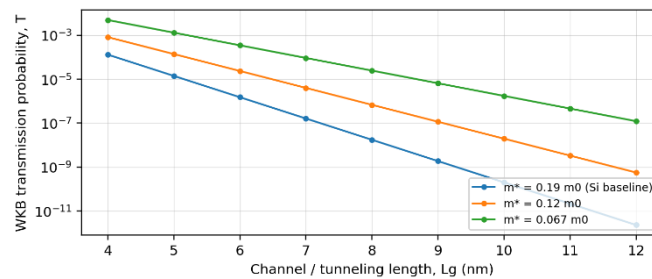


Figure 2. WKB transmission versus tunneling length for three representative transport effective masses.

Conclusion and Discussion

This paper presented a compact methodology for investigating quantum tunneling effects in nanoscale semiconductor devices using a silicon double-gate MOSFET as the reference structure. The method combines a WKB parameter sweep with the Landauer current relation and the NEGF formalism. The WKB stage provides rapid physical insight, while NEGF supplies the rigorous open-boundary quantum transport framework needed for full self-consistent simulation. The calculated results show that direct source-to-drain tunneling increases exponentially as channel length decreases: for a 0.25 eV barrier and $m^* = 0.19 m_0$, transmission rises from 2.30×10^{-12} at 12 nm to 1.32×10^{-4} at 4 nm. The corresponding relative tunneling index increases by approximately 5.7×10^7 . The effective-mass sweep further shows that low-mass materials can produce substantially larger tunneling at the same geometry, confirming the central trade-off between high carrier injection velocity and off-state leakage. The main implication is that classical electrostatic scaling rules are insufficient below the deep-nanometer regime. Device optimization must explicitly consider wave-mechanical penetration, quantum confinement, the energy-dependent transmission spectrum, and the material band

structure. The literature up to 2021 supports this conclusion across silicon MOSFETs, nanowires, two-dimensional channels, and tunneling devices [4]-[11]. The present numerical values should be interpreted as semi-analytical baseline estimates rather than replacements for atomistic NEGF-Poisson simulation. Future work can improve the study by extracting a realistic non-rectangular barrier from a two-dimensional Poisson solution, including multiple subbands and valley degeneracy, adding phonon and surface-roughness self-energies, and benchmarking WKB transmission directly against a self-consistent NEGF calculation. Even with these simplifications, the methodology clearly demonstrates why quantum tunneling is a fundamental constraint and a design opportunity in nanoscale semiconductor devices.

References

R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," *Journal of Applied Physics*, vol. 81, no. 12, pp. 7845-7869, 1997. DOI: 10.1063/1.365394.

R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, and D. Jovanovic, "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches," *Journal of Applied Physics*, vol. 92, no. 7, pp. 3730-3739, 2002. DOI: 10.1063/1.1503165.

M. Luisier, A. Schenk, and W. Fichtner, "Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism," *Journal of Applied Physics*, vol. 100, 043713, 2006. DOI: 10.1063/1.2244522.

Q. Rafhay, R. Clerc, G. Pananakakis, and G. Ghibaudo, "Impact of source to drain tunneling on the Ion/Ioff trade-off of alternative channel material MOSFETs," *Solid-State Electronics*, vol. 52, pp. 1474-1481, 2008. DOI: 10.1016/j.sse.2008.06.035.

M. Ionescu and H. Riel, "Tunnel field-effect transistors as energy-efficient electronic switches," *Nature*, vol. 479, pp. 329-337, 2011. DOI: 10.1038/nature10679.

S. Steiger, M. Povolotskyi, H.-H. Park, T. Kubis, and G. Klimeck, "NEMO5: A parallel multiscale nanoelectronics modeling tool," *IEEE Transactions on Nanotechnology*, vol. 10, no. 6, pp. 1464-1474, 2011. DOI: 10.1109/TNANO.2011.2166164.

H. Ryu, "A multi-subband Monte Carlo study on dominance of scattering mechanisms over carrier transport in sub-10-nm Si nanowire FETs," *Nanoscale Research Letters*, vol. 11, article 36, 2016. DOI: 10.1186/s11671-016-1249-4.

R. Quhe et al., "Simulations of quantum transport in sub-5-nm monolayer phosphorene transistors," *Physical Review Applied*, vol. 10, 024022, 2018. DOI: 10.1103/PhysRevApplied.10.024022.

Y. Pan et al., "Sub-5-nm monolayer silicane transistor: A first-principles quantum transport simulation," *Physical Review Applied*, vol. 14, 024016, 2020. DOI: 10.1103/PhysRevApplied.14.024016.

N. Pandey, G. Pahwa, and Y. S. Chauhan, "Addressing source to drain tunneling in extremely scaled Si-transistors using negative capacitance," *Solid-State Electronics*, vol. 186, 108189, 2021. DOI: 10.1016/j.sse.2021.108189.

R. Quhe et al., "Sub-10 nm two-dimensional transistors: Theory and experiment," *Physics Reports*, vol. 938, pp. 1-72, 2021. DOI: 10.1016/j.physrep.2021.07.006.