**Report**  
  
**Introduction**

Semiconductor devices have revolutionized the field of electronics, paving the way for advancements in various industries such as computing, telecommunications, healthcare, and energy. They are integral to many gadgets that are essential to modern life. MOS structures, due to their unique properties such as high electron mobility and low power consumption, have garnered significant attention in the electronics field. They are particularly relevant in the current digital era, where there is a growing demand for compact and power-efficient electronic devices. The aim of this presentation is to provide a comprehensive analysis of the tunneling phenomenon in ultra-thin oxide MOS structures. This quantum mechanical phenomenon can significantly impact the performance of MOS structures and is therefore crucial to understand for the further development of electronic devices.

A diagram of a waveform

Description automatically generated

Figure 1. Band model for thin MOS structure. The lower diagram represents Fowler-Nordheim tunneling.

Ongoing progress in Si nanotechnology has resulted in Si MOSFET prototypes with gate lengths well below 20 nm. As channel lengths decrease, the significance of quantum effects, such as size quantization and gate oxide tunneling, becomes more pronounced. Traditional CMOS transistors face performance limitations as dimensions and operating voltages decrease. To overcome these limitations while maintaining compatibility with conventional processes, continuous technological advancements are being pursued. A state-of-the-art CMOS technology illustrates key features, emphasizing the need for gate oxide thickness around 1.2-1.5nm for sub-100nm channel lengths. In ultra-thin gate oxide regimes, gate leakage current can significantly contribute to off-state leakage, impacting the stability of Very Large Scale Integration (VLSI) circuits. The occurrence of edge direct tunneling (EDT) between the source-drain extension and gate overlap further complicates the scenario. Understanding these dynamics is crucial for effective design and operation in advanced semiconductor technologies.

**Tunneling in metal–insulator–semiconductor structures**

A diagram of a metal and metal

Description automatically generated

Figure 2. Band diagram for a metal–oxide–semiconductor structure without applied voltage; the band bending is caused by the metal–semiconductor work function difference and the oxide fixed charges. Ec is the conduction band, Ei is the intrinsic Fermi level, Efs is the Fermi level in the semiconductor and Efm is the Fermi level in the metal.

The metal–oxide–p-type semiconductor system is depicted in Fig. 2 without applied voltage.

The illustration assumes a negative metal–semiconductor work function difference, common in MOS devices with an Al gate on a Si substrate. When a substantial positive voltage is applied to the metal compared to the substrate, the left side of the band diagram lowers. This voltage shift facilitates electron tunneling from the semiconductor's conduction band into the oxide's conduction band, forming an approximately triangular barrier (refer to Fig. 3(a)). The described tunneling through an approximately triangular potential barrier is identified as Fowler–Nordheim tunneling. This phenomenon occurs when a large positive voltage is applied, allowing electrons to tunnel through the triangular barrier. The simplicity of this tunneling process is captured in a relatively straightforward expression for current as a function of applied voltage.

A diagram of a voltage

Description automatically generated with medium confidence

Figure 3. (a) Tunneling through a triangular barrier into the insulator conduction band (Fowler–Nordheim tunneling). (b) Tunneling through a trapezoidal barrier (direct tunneling).

The band diagram's left-hand side lowering, induced by a large positive voltage on the metal, signifies a key aspect of the MOS device's operation. The voltage-induced shift in the band diagram enables the creation of the triangular barrier, facilitating electron tunneling from semiconductor to oxide. This process plays a crucial role in the device's functionality and is fundamental to understanding gate tunneling current in MOS devices. This provides a relatively simple expression detailing the current as a function of applied voltage during Fowler–Nordheim tunneling.

A math equation with numbers and symbols

Description automatically generated

where q is the electron charge, h is Planck’s reduced con- stant, mox is the electron effective mass in the insulator, /b is the barrier height at the semiconductor–oxide inter- face and Fox is the electric field across the oxide. The assumptions for the derivation of (1) are [8,9]:

• The electrons on the emitting electrode can be described as a free Fermi gas.

• The dependence of carrier availability for tunneling with respect to temperature is not taken into account.

• The potential barrier has triangular shape and barrier lowering due to image forces is neglected.

• The effect of the insulator can be described by a single effective mass.

• The tunneling probability takes into account only the component of the electron momentum in the direction normal to the surface.

This relationship captures the essence of how current flows because of the triangular barrier and the applied voltage. Understanding this current-voltage relationship is vital for comprehending the gate tunneling current in MOS devices, offering insights into their electrical behavior.

**Direct Tunnelling**

In early MOS structures, tunneling currents were observed under high applied fields, causing a reduction in one side of the potential barrier at the silicon–oxide interface. Thick oxides (approximately 70–1000 Å) were used, making direct tunneling into the other electrode's conduction band improbable. However, when oxide thickness is reduced to a few nanometers, direct tunneling at lower fields becomes significant, resulting in an approximately trapezoidal barrier (Fig. 3(b)). Fowler–Nordheim model's validity is limited when the potential across the insulator is lower than the potential barrier at the Si–SiO2 interface, as depicted in Fig. 5. The direct tunneling current through the energy gap of an insulator or semiconductor is expressed as a function of tunneling probability (P) and various parameters [20,21].

A black math equation

Description automatically generated with medium confidence

where P is the tunneling probability, E is the total electron energy, ky and kz are the wave vectors in the plane of the barrier (perpendicular to the tunneling direction), and f1 and f2 are the probabilities of occupation of the states on each side of the barrier, given by the Fermi–Dirac distribution functions; the integration limits for ky and kz are determined from momentum and energy conservation.

The tunneling probability involves integration over wave vectors, electron energy, and probabilities of states on each side of the barrier. The Fermi–Dirac distribution functions and the effective mass approximation are essential in formulating the current-voltage relationship [22]. The relationship is influenced by the electron effective mass, kinetic energy, temperature, and the 'supply' function derived from Fermi–Dirac distribution functions.

A black and white math symbols

Description automatically generated

where h is Planck’s constant, mx is the electron effective mass in the direction perpendicular to the barrier, Ex is the electron kinetic energy in the direction perpendicular to the barrier, k is the Boltzmann constant, T is the temperature, and S(Ex) is the ‘‘supply’’ function, which is derived from the integration of the Fermi–Dirac distribution function, where,

A black text with white text

Description automatically generated with medium confidence

where Efs and Efg are the electron Fermi levels in the semi-

conductor and the gate, respectively. Note that the applied voltage is implicitly present in this equation through the difference between the Fermi levels.

Determining tunneling probability involves solving Schrödinger's equation for the electron wave function. The Wentzel–Kramers–Brillouin (WKB) approximation is commonly used, assuming a 'slow' change in potential. The WKB approximation neglects reflections and interference of electron wave functions, which Gundlach addressed by solving Schrödinger's equation numerically, revealing periodic 'oscillations' in the current [26]. The classical turning points in a trapezoidal barrier, crucial for tunneling probability, are identified as the metal–insulator and insulator–semiconductor interfaces [24]. Various approximations are employed for modeling tunneling current, such as replacing Fermi–Dirac functions with Fermi level at low temperatures and assuming a degenerate inversion or accumulation layer [19]. These approximations can lead to overestimation of tunneling current, especially at low voltages with limited carrier availability. Alternative expressions, like the one proposed by Schuegraf et al. [19], aim to overcome the limitations of previous models but still face challenges in predicting correct trends, particularly for very thin oxides. Accurate modeling of direct tunneling in the moderate inversion or accumulation regime requires consideration of the relation between surface potential and applied voltage, posing analytical challenges [33]. Determining tunneling current involves solving Schroedinger's equation for electron wave function, considering periodic lattice potential and applying approximations like the Wentzel–Kramers–Brillouin or neglecting Fermi–Dirac functions for modeling, with an alternative expression to address limitations:

A math equations and numbers

Description automatically generated

1. \*\*Quantization Effects on Carrier Concentration:\*\*

- The derivation of the Fowler–Nordheim formula and calculations for direct tunneling assume a three-dimensional free Fermi gas model for electrons in the semiconductor.

- However, the electric field in the gate–channel region can create band bending, confining electrons to a narrow potential well near the semiconductor surface, leading to the formation of a 2-dimensional electron gas (2-DEG).

- Quantization of energy in the direction normal to the surface, as depicted in Fig. 9, results in a different carrier concentration distribution compared to the classical theory, impacting semiconductor capacitance estimation.

- The peak carrier concentration, influenced by quantization effects, shifts away from the surface, impacting the accuracy of capacitance predictions.

2. \*\*Threshold Voltage Shift due to Quantization:\*\*

- Quantization effects in the substrate contribute to an increase in the threshold voltage of MOS transistors.

- This impact is particularly significant for highly doped semiconductor substrates, as illustrated in Fig. 10.

- The shift in threshold voltage is a consequence of the quantization of electron energies in the accumulation/inversion layers, emphasizing the importance of considering quantum effects in transistor performance.

3. \*\*Modeling Electron Behavior in Quantized States:\*\*

- The quantization of electron energies in the accumulation/inversion layers necessitates a departure from the free electron gas model.

- Instead of a 'transmission probability,' the calculation of gate tunneling current involves the use of a 'carrier tunneling lifetime' for electrons in quantized energy levels, often referred to as quasi-bound states.

- Quantized energy levels are determined by solving Schrödinger's equation for the wave function of electrons, where the potential and carrier distribution influence each other in a coupled manner.

- The simultaneous solution of Schrödinger's equation, as described in Eq. (15), provides insight into the behavior of electrons in the potential well and their contribution to tunneling current.

4. \*\*Effective Masses and Sub-Bands in Semiconductor:\*\*

- Due to the band structure of the semiconductor, there exist different effective masses corresponding to various energy sub-bands.

- Schrödinger's equation is expressed in terms of these effective masses, emphasizing the complex nature of electron behavior in the quantized states.

- The expression (15) highlights the involvement of the charged density in the semiconductor and the electrical permittivity of the semiconductor in determining quantized energy levels.

- The consideration of multiple effective masses reflects the intricacies involved in accurately modeling the behavior of electrons in quantized states, adding another layer of complexity to the study of gate tunneling current in MOS devices.

1. \*\*Adjustment for Large Values in the Approximate Solution (22):\*\*

- The solution (22) provides an approximate solution for 'large' values of \(i\).

- For specific values (\(i=1, i=2, i=3\)), the term \((i-1/4)\) is replaced with precise constants (0.7587, 1.7540, 2.7525) [41].

- This adjustment enhances the accuracy of the solution for different values of \(i\), ensuring a more precise representation of the tunneling phenomenon.

2. \*\*Inconsistency of Infinite Potential Barrier Assumption:\*\*

- The assumption of an infinite potential barrier is not congruent with the existence of tunneling current.

- An infinite barrier suggests zero probability for the wave function to penetrate into the barrier, contradicting the occurrence of tunneling current.

- Using the actual barrier height as a boundary condition yields different discrete energies, especially pronounced for thinner oxides, as demonstrated in Fig. 11 [42].

3. \*\*Validity of Linear Potential Approximation:\*\*

- The linear potential approximation provided by (20) holds true only in weak inversion [41].

- For strong inversion or accumulation, alternative approximations or numerical solutions of wave and potential equations become necessary [41,43].

- In the accumulation region, a mix of quantized energy levels near the semiconductor surface, corresponding to bound electrons, and unbound electrons with wave functions extending into the bulk, complicates the determination of surface potential [43].

4. \*\*Quantized Energy Levels and Tunneling Current Calculation:\*\*

- Quantization of electron energies in accumulation/inversion layers requires a departure from the free electron gas model.

- Instead of a 'transmission probability,' a 'carrier tunneling lifetime' is used for calculating tunneling current from quantized levels [31].

- Quasi-bound states, allowing electrons to tunnel through the barrier, contribute to the surface potential, requiring a consistent solution to Schrödinger’s and Poisson’s equations.

- The determination of quasi-bound state lifetime involves solving Schrödinger's equation for the wave function and considering the potential barrier's influence on electron behavior.

5. \*\*Numerical Implementation and Consideration of Quasi-Bound States:\*\*

- Once energy levels (\(E\_{ij}\)) are known, the quasi-bound state lifetime (\(\tau\_{ij}\)) is calculated for each, and tunneling current is the sum of contributions from these states.

- The charge density (\(q\)) and potential (/) are interconnected, necessitating a simultaneous solution of Schrödinger’s and Poisson’s equations.

- Numerical implementation, often with self-consistent solutions, is common for calculating tunneling current from quasi-bound states.

- Compact modeling may account for carrier quantization by assuming 'barrier lowering' or omitting it based on self-compensating effects [46,29].

A diagram of a subband

Description automatically generated

Figure 4. Schematic band profile for a poly Si–SiO2–Si structure in inversion, showing the formation of sub-bands due to carrier confinement in a narrow potential well

A graph of a graph of electrons and electrons

Description automatically generated

Figure 5. Variation in the threshold voltage with respect to the classical theory caused by quantum-mechanical effects in the substrate for electron and hole inversion layers

5. \*\*Quantization Effects in Two-Dimensional Electron Gas (2-DEG):\*\*

- The electric field in the gate–channel region can create band bending strong enough to confine electrons to a narrow potential well close to the semiconductor surface, forming a 2-dimensional electron gas (2-DEG).

- Quantization effects in the substrate lead to a distribution of carrier concentration different from classical theories based on the assumption of energy continuum.

- This quantization influences the potential and, consequently, the carrier distribution, affecting MOS transistor characteristics such as capacitance and threshold voltage.

- Solving Schrödinger’s equation for the wave function of electrons in the quantized energy levels is necessary to accurately model the carrier tunneling lifetime in gate tunneling current calculations.  
  
A diagram of a graph

Description automatically generated with medium confidence

Figure 6: Schematic representation of the tunneling components in a Si/ SiO2/Si structure.

A close-up of a diagram

Description automatically generated

Table 1: Dominant current mechanisms for each tunneling current component.

6. \*\*Carrier Tunneling Lifetime and Quasi-Bound States:\*\*

- In gate tunneling current calculations, the contribution of each quantized energy level is expressed through a "carrier tunneling lifetime."

- Quasi-bound states, representing electrons tunneling through the barrier, are formed as electrons are not entirely bound inside the potential well.

- The determination of the tunneling lifetime involves solving Schrödinger's and Poisson's equations in a self-consistent manner.

- Numerical methods are often employed for implementing these calculations, considering the carrier quantization effects for compact modeling purposes.

7. \*\*Alternative Approaches for Tunneling Lifetime Calculation:\*\*

- The tunneling lifetime, critical for accurate gate tunneling current modeling, can be determined using semi-classical or fully quantum-mechanical approaches.

- Semi-classical methods involve considering quasi-bound states as point-like charged particles bouncing inside the well, providing a quasi-bound state lifetime.

- Fully quantum-mechanical alternatives treat the potential well and barrier as scattering centers, yielding complex energies and a resonance width related to the tunneling lifetime.

- The transverse-resonant method is one numerical approach to calculate the resonance width, offering insights into the energy broadening of quasi-bound states during tunneling.

A diagram of a gate and channel

Description automatically generated A diagram of a gate voltage

Description automatically generated

Figure 7: Gate tunneling current components for a MOS transistor. Measured gate current for an n+ Si/SiO2/p Si MOS capacitor with an n+ source, tox = 1.6 nm and W/L = 562/62 lm (solid line). The dashed and dotted lines show the calculated gate-overlap and gate– channel components, respectively [29]. The overlap current component is dominant at low positive and negative gate voltages (depletion–accumulation), even though the overlap area is much smaller than the channel area.

8. \*\*Incorporating Quantum Effects in Compact Modeling:\*\*

- For compact modeling purposes, it is common to consider the "barrier lowering" due to energy quantization effects, taking into account the difference between the first energy level and the conduction band.

- Despite the numerical complexity of incorporating quantum effects, certain approximations, such as assuming a linear potential variation with distance, allow for more manageable solutions in weak inversion.

- Quantum effects, especially carrier quantization, play a crucial role in determining the tunneling probability and, consequently, the gate tunneling current, impacting the performance of MOS devices.

- Balancing the need for accuracy with the computational efficiency required for compact modeling remains a challenge in the inclusion of quantum effects in gate tunneling current models.