TUNNELING

DRAGICA VASILESKA AND GERHARD KLIMECK

QUANTUM EFFECTS

- Quantum-mechanical space quantization
- Tunneling
- Quantum Interference

In all but the smallest devices quantum-mechanical space quantization effects and tunneling play dominant role and they can be captured with quantum correction models

Quantum interference dominates the operation of resonant tunneling diodes and fully quantum transport approaches are needed to treat this device.

TREATMENT OF TUNNELING

- WKB Approximation
- Transfer Matrix Approach
 - Piece-Wise Constant Potential Barrier Approximation
 - Piece-Wise Linear Potential Barrier Approximations

WENTZEL-KRAMERS-BRILLOUIN (WKB) APPROXIMATION

IMPORTANT APPLICATIONS IN WHICH WKB APPROXIMATION IS USED

- Tunneling Breakdown in normal diodes (reverse biased diode)
- Tunnel (Esaki) diode (forward + reverse bias)
- Scanning Tunneling Microscope
- Gate Leakage in MOSFET Devices

A. BREAKDOWN MECHANISMS IN A DIODE

- Junction breakdown can be due to:
 - tunneling breakdown
 - avalanche breakdown
- One can determine which mechanism is responsible for the breakdown based on the value of the breakdown voltage V_{BD} :
 - ♦ $V_{BD} < 4E_g/q \rightarrow$ tunneling breakdown
 - ♦ $V_{BD} > 6E_g/q \rightarrow$ avalanche breakdown
 - ♦ $4E_g/q < V_{BD} < 6E_g/q \rightarrow$ both tunneling and avalanche mechanisms are responsible

• Tunneling breakdown occurs in heavily-doped *pn*junctions in which the depletion region width *W* is about 10 nm.





• Tunneling current (obtained by using WKB approximation):



- $F_{cr} \rightarrow$ average electric field in the junction
- The critical voltage for tunneling breakdown, V_{BR}, is estimated from:

 $I_t(V_{BR}) \propto 10 I_S$

• With $T \uparrow$, $E_g \lor$ and $I_t \uparrow$.

B. TUNNEL (ESAKI) DIODE

Leo Esaki







Nobel Prize in Physics 1973

(ESAKI) TUNNEL DIODE (TD)

- Simplest tunneling device
- Heavily-doped pn junction
 - Leads to overlap of conduction and valence bands
- Carriers are able to tunnel inter-band
- Tunneling goes exponentially with tunneling distance
 - Requires junction to be abrupt







Indirect materials require phonons to tunnel, thus reducing the probability of a tunneling event

C. SCANNING TUNNELING MICROSCOPE

revolution of tunnelling: Scanning Tunnelling Microscope



D. WKB APPROXIMATION EXPLAINED

- The Wentzel-Kramers-Brillouin (WKB) approximation is a "semiclassical calculation" in quantum mechanics in which the wavefunction is assumed an exponential function with amplitude and phase that slowly varies compared to the de Broglie wavelength, λ, and is then semiclassically expanded
- While Wentzel, Kramers and Brillouin developed this approach in 1926, earlier in 1923 Harold Jeffreys had already developed a more general method of approximating linear, second-order differential equations (the Schrodinger equation is a linear second order differential equation)

WKB APPROXIMATION EXPLAINED, CONT'D

- While technically this is an "Approximate Method" not an "Exact solution" to the Schrodinger equation, it is very close to simple plane wave solutions that we discussed while describing transmission coefficient calculation in piece-wise constant potential barriers
- The WKB method is most often applied to 1D problems but can be applied to 3D Spherically Symmetric problems as well (see Bohm 1951)
- The WKB approximation is especially useful in deriving the tunnel current in a tunnel diode

BASIC IDEA OF THE METHOD

• The WKB approximation states that since in a constant potential, the wavefunction solutions of the Schrodinger equation are of the form of simple plane waves, then

$$\Psi(x) = Ae^{\pm ikx}, \quad k = 2\pi / \lambda = \sqrt{\frac{2m(E-U)}{\hbar^2}}$$

 Now, if the potential U=U(x) changes slowly with x, the solution of the Schrodinger equation can also be written of the general form

$$\Psi(x) = A e^{i\phi(x)}$$

where $\phi(x) = xk(x)$.

- For the **constant potential case**, $\phi(x)=\pm kx$ so the phase changes linearly with x

- In a **slowly varying potential** $\phi(x)$ should vary slowly from the linear case $\pm kx$

BASIC IDEA OF THE METHOD, CONT'D

 For the two cases, E>U and E<U, let k(x) be defined as (so we only have to solve the problem once)

$$k(x) = \sqrt{\frac{2m(E - U(x))}{\hbar^2}}, \qquad E > U(x)$$
$$k(x) = -i\sqrt{\frac{2m(U(x) - E)}{\hbar^2}} = -i\kappa(x), \quad E < U(x)$$

WENTZEL-KRAMERS-BRILLOUIN (WKB) APPROXIMATION

Starting from the 1D Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x) + U(x)\psi(x) = E\psi(x)$$

 And substituting the general solution for slowly-varying potentials, one gets the following differential equation

$$i\frac{\partial^2\phi}{\partial x^2} - \left(\frac{\partial\phi}{\partial x}\right)^2 + k^2(x) = 0$$

WENTZEL-KRAMERS-BRILLOUIN (WKB) APPROXIMATION

- The WKB approximation assumes that the potentials are slowly varying in space
- Then the 0th order approximation assumes

$$\frac{\partial^2 \phi}{\partial x^2} = 0, \quad \frac{\partial \phi_0}{\partial x} = \pm k(x) \to \phi_0(x) = \pm \int k(x) dx + C_0$$
$$\to \psi(x) = \exp\left[\pm i \int k(x) dx + C_0\right]$$

Wentzel-Kramers-Brillouin (WKB) Approximation

If a higher order solution is required, then we solve

$$i\frac{\partial^2\phi}{\partial x^2} - \left(\frac{\partial\phi}{\partial x}\right)^2 + k^2(x) = 0 \longrightarrow \frac{\partial\phi}{\partial x} = \pm\sqrt{k^2(x) + i\frac{\partial^2\phi}{\partial x^2}}$$

Then the 1th order approximation assumes

$$\frac{\partial \phi}{\partial x} = \pm \sqrt{k^2(x) \pm i \frac{\partial k}{\partial x}}$$
$$\rightarrow \psi(x) = \exp\left[\pm i \int \sqrt{k^2(x) \pm i \frac{\partial k}{\partial x}} dx + C_1\right]$$

Wentzel-Kramers-Brillouin (WKB) Approximation

1. In order to apply the WKB approximation we only need to know the shape of the potential since

$$U(x) \to k(x) \to \phi(x) \to \psi(x) = \exp\left[\pm \int \sqrt{k^2(x) \pm i \frac{\partial k}{\partial x}} dx + C_1\right]$$

2. For slowly varying U(x) the first order and the zero order approximation give almost the same result as

$$\left|\frac{\partial}{\partial x}k(x)\right| \ll \left|k^2(x)\right|$$

Wentzel-Kramers-Brillouin (WKB) Approximation

3. The WKB approximation breaks down where E~U (classical turning points) in which case the wavevector k(x) approaches zero but the derivative does not and there in fact the argument in (2) does not hold

$$\left. \frac{\partial}{\partial x} k(x) \right| << \left| k^2(x) \right|$$

Under these circumstances, connection formulas must be applied to tie together regions on each side of the classical turning point.

E. EXAMPLE: GATE LEAKAGE



GATE LEAKAGE

- For sub-micrometer devices, due to smaller oxide thickness, there is significant conductance being measured on the gate contact. The finite gate current gives rise to the following effects:
 - Negative => degradation in the device operating characteristics with time due to oxide charging; larger off-state power dissipation
 - Positive => non-volatile memories utilize the gate current to program and erase charge on the "floating contact" – FLASH, FLOTOX, EEPROM
- There are two different types of conduction mechanisms to the insulator layer:
 - ♦ Tunneling: <u>Fowler-Nordheim</u> or <u>direct</u> tunneling process
 - ♦ Hot-carrier injection: <u>lucky electron model</u> or <u>Concannon model</u>

Electron is emitted into the oxide when it gains sufficient energy to overcome the insulator/semiconductor barrier.

- Similar to the lucky electron model, but assumes non-Maxwellian high energy tail on the distribution function.
- Requires solution of the energy balance equation for carrier temperature.

TUNNELING CURRENTS

 Three types of tunneling processes are schematically shown below (courtesy of D. K. Schroder)



- For $t_{ox} \ge 40$ Å, Fowler-Nordheim (FN) tunneling dominates
- For $t_{ox} < 40$ Å, direct tunneling becomes important
- $I_{dir} > I_{FN}$ at a given V_{ox} when direct tunneling active
- For given electric field: I_{FN} independent of oxide thickness
 I_{dir} depends on oxide thickness

SIGNIFICANCE OF GATE LEAKAGE

- As oxide thickness decreases, the gate current becomes more important. It eventually dominates the off-state leakage current (I_D at V_G = 0 V)
- The drain current I_D as a function of technology generation is shown below (courtesy of D. K. Schroder)



FOWLER-NORDHEIM TUNNELING



- The difference between the Fermi level and the top of the barrier is denoted by Φ_B
- According to WKB approximation, the tunneling coefficient through this triangular barrier equals to:

$$T \propto \exp\left[-2\int_{0}^{a}\gamma(x)dx\right]$$
 where: $\gamma(x) = \sqrt{\frac{2m^{*}}{\hbar^{2}}(\Phi_{B} - eEx)}$

$$V(Z) = E_G \left(1 - \frac{Z}{W}\right)$$

 E_G

The attenuation of the barrier is thus

$$2\alpha = 2 \int_{0}^{Z_{0}} \sqrt{\frac{2m}{\hbar^{2}}} \left[V(Z) - E_{Z} \right] dZ$$

$$= 2 \int_{0}^{Z_{0}} \sqrt{\frac{2m}{\hbar^{2}}} \left[E_{G} \left(1 - \frac{Z}{W} \right) - E_{Z} \right] dZ$$

$$= 2 \int_{0}^{Z_{0}} \sqrt{\frac{2mE_{G}}{\hbar^{2}}} \left[1 - \frac{Z}{W} - \frac{E_{Z}}{E_{g}} \right] dZ$$

$$= 2 \int_{0}^{Z_{0}} \sqrt{\frac{2mE_{G}}{\hbar^{2}}} \sqrt{1 - \frac{Z}{W} - \frac{E_{Z}}{E_{g}}} d\left(-\frac{Z}{W} \right) (-W)$$

$$= (-2W) \int_{0}^{Z_{0}} \sqrt{\frac{2mE_{G}}{\hbar^{2}}} \sqrt{1 - \frac{Z}{W} - \frac{E_{Z}}{E_{g}}} d\left(-\frac{Z}{W} \right)$$

$$= (-2W) \sqrt{\frac{2mE_{G}}{\hbar^{2}}} \left[\left(1 - \frac{Z}{W} - \frac{E_{Z}}{E_{g}} \right)^{3/2}} \right]_{0}^{Z_{0}}$$

$$= -\frac{4W}{3} \sqrt{\frac{2mE_{G}}{\hbar^{2}}} \left[\left(1 - \frac{Z_{0}}{W} - \frac{E_{Z}}{E_{g}} \right)^{3/2} - \left(1 - \frac{E_{Z}}{E_{g}} \right)^{3/2} \right]$$

$$\int_{E_{G}-E_{Z}} V(z)$$

$$\int_{Z} J(z)$$

$$Transmission coefficient:$$

$$T = e^{-2\alpha}$$

where

$$1 - \frac{Z_0}{W} - \frac{E_Z}{E_g} = 1 - \frac{E_g - E_Z}{E_g} - \frac{E_Z}{E_g}$$
$$= \frac{E_g - E_Z}{E_g} - \frac{E_g - E_Z}{E_g}$$

Hence

$$\alpha = \frac{4W}{3} \sqrt{\frac{2mE_G}{\hbar^2}} \left(1 - \frac{E_Z}{E_g}\right)^{3/2}$$
$$\approx \frac{4W}{3} \sqrt{\frac{2mE_G}{\hbar^2}} \left(1 - \frac{3}{2}\frac{E_Z}{E_g}\right)$$

$$T(E_Z) = e^{-2\alpha}$$

$$= \underbrace{e^{-\frac{4W}{3}\sqrt{\frac{2mE_G}{\hbar^2}}}_{T_0}}_{T_0} e^{\frac{2W}{E_g}\sqrt{\frac{2mE_G}{\hbar^2}}}_{E_Z}E_Z$$

$$= T_0 e^{E_Z/E_0}$$

$$E_0 = \frac{E_g}{2W} \sqrt{\frac{\hbar^2}{2mE_g}}$$
$$= \sqrt{\frac{\hbar^2 E_g^2}{2mE_g}} \frac{1}{2W}$$
$$= \frac{1}{2W} \sqrt{\frac{\hbar^2 E_g}{2m}}$$

FOWLER-NORDHEIM TUNNELING

 The final expression for the Fowler-Nordheim tunneling coefficient is:

$$T \propto \exp\left[-\frac{4\sqrt{2m^*}\Phi_B^{3/2}}{3eE\hbar}\right]$$

- Important notes:
 - The above expression explains tunneling process only qualitatively because the additional attraction of the electron back to the plate is not included
 - Due to surface imperfections, the surface field changes and can make large difference in the results



Calculated and experimental tunnel current characteristics for ultra-thin oxide layers.

(M. Depas et al., *Solid State Electronics*, Vol. 38, No. 8, pp. 1465-1471, 1995)

TRANSFER MATRIX APPROACH

TUNNELING: TRANSFER MATRIX APPROACH

Within the Transfer Matrix approach one can assume to have either

- Piece-wise constant potential barrier
- Piecewise-linear potential barrier

PIECE-WISE CONSTANT POTENTIAL APPROXIMATION

Piece-Wise Constant Potential Barrier (PCPBT Tool) installed on the nanoHUB



The Approach at a Glance



The Approach, Continued



total transfer matrix:

$$\binom{C}{D} = T \binom{A}{B}$$



Slide property of Sozolenko.

PIECE-WISE LINEAR POTENTIAL APPROXIMATION

PIECE-WISE LINEAR APPROXIMATION

- This algorithm is implemented in ASU's code for modeling Schottky junction transistors (SJT)
- It approximates real barrier with piece-wise linear segments for which the solution of the 1D Schrodinger equation leads to Airy functions and modified Airy functions
- Transfer matrix approach is used to calculate the energydependent transmission coefficient
- Based on the value of the energy of the particle E, T(E) is looked up and a random number is generated. If r<T(E) than the tunneling process is allowed, otherwise it is rejected.

The Approach at a Glance

 1D Schrödinger equation:

 $-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + V(x)\psi = E\psi$

• Solution for piecewise linear potential:

 $\Psi_i = C_i^{(1)} A_i(\xi) + C_i^{(2)} B_i(\xi)$

 The total transmission matrix:

 $M_T = M_{FI} M_1 M_2 \dots M_{N-1} M_{BI}$

• **T(E):**
$$T = \frac{k_{N+1}}{K_0} \frac{1}{\left| m_{11}^T \right|^2}$$



Simulation Results for Gate Leakage in SJT





T. Khan, D. Vasileska and T. J. Thornton, "Quantum-mechanical tunneling phenomena in metalsemiconductor junctions", *NPMS 6-SIMD 4*, November 30-December 5, 2003, Wailea Marriot Resort, Maui, Hawaii.

TSU-ESAKI FORMULA FOR THE CURRENT CALCULATION

DERIVATION OF THE TSU-ESAKI FORMULA



Energy barrier with two electrodes which can be used to describe the ECB or HVB processes.

ASSUMPTIONS

- Effective mass approximation. The different masses corresponding to the band structure of the considered material are lumped into a single value for the effective mass. This is denoted by m_{eff} in the electrodes and m_{diel} in the dielectric layer.
- Parabolic bands. The dispersion relation in semiconductors is approximated by

$$E = \frac{\hbar^2 k^2}{2m_{eff}} = \frac{\hbar^2}{2m_{eff}} \left(k_x^2 + k_y^2 + k_z^2\right)$$

• Conservation of parallel momentum. Only transitions in the k_x direction are considered, the parallel wavevector $k_{\parallel} = k_y e_y + k_z e_z$ is not altered by the tunneling process.

CURRENT CALCULATION

 The net tunneling current from Electrode 1 to Electrode 2 can be written as the net difference between current flowing from Side 1 to Side 2 and vice versa:

$$J = J_{1 \to 2} - J_{2 \to 1}$$

• The current density through the two interfaces depends on the perpendicular component of the wavevector k_x , the transmission coefficient T_c , the perpendicular velocity v_x , the density of states g_c and the distribution function at both sides of the barrier:

$$dJ_{1\to 2} = qT_c(k_x)v_xg_1(k_x)f_1(E)[1 - f_2(E)]dk_x$$

$$dJ_{2\to 1} = qT_c(k_x)v_xg_2(k_x)f_2(E)[1 - f_1(E)]dk_x$$

 In this expression it is assumed that the transmission coefficient only depends upon the momentum perpendicular to the interface. The density of states g(k_x) is:

$$g(k_x) = \int_0^\infty \int_0^\infty g(k_x, k_y, k_z) dk_y dk_z$$

 Where g(k_x,k_y,k_z) denotes the three-dimensional density of states in momentum space. Considering the quantized wavevector components within a cube of side L yields for the density of states within the cube:

$$g(k_x, k_y, k_z) = \frac{2}{L^3} \frac{1}{\Delta k_x \Delta k_y \Delta k_z} = \frac{1}{4\pi^3}, \quad \Delta k_i = \frac{2\pi}{L}$$

 For the parabolic dispersion relation, the velocity and energy components in the tunneling direction obey:

$$v_x = \frac{1}{\hbar} \frac{\partial E}{\partial k_x} = \frac{\hbar k_x}{m_{eff}}, \text{ and } v_x dk_x = \frac{1}{\hbar} dE_x$$

• Hence, the expressions for the current density become:

$$dJ_{1\to2} = \frac{q}{4\pi^{3}\hbar} T_{c} \left(E_{x}\right) dE_{x} \int_{0}^{\infty} \int_{0}^{\infty} f_{1}(E) [1 - f_{2}(E)] dk_{y} dk_{z}$$
$$dJ_{2\to1} = \frac{q}{4\pi^{3}\hbar} T_{c} \left(k_{x}\right) dE_{x} \int_{0}^{\infty} \int_{0}^{\infty} f_{2}(E) [1 - f_{1}(E)] dk_{y} dk_{z}$$

 Using polar coordinates for the parallel wavevector components

$$k_{\rho} = \sqrt{k_y^2 + k_z^2}$$

• The current density evaluates to:

$$J_{1\to2} = \frac{4\pi m_{eff} q}{h^3} \int_{E_{min}}^{E_{max}} T_c(E_x) dE_x \int_{0}^{\infty} f_1(E) [1 - f_2(E)] dE_{\rho}$$
$$J_{2\to1} = \frac{4\pi m_{eff} q}{h^3} \int_{E_{min}}^{E_{max}} T_c(E_x) dE_x \int_{0}^{\infty} f_2(E) [1 - f_1(E)] dE_{\rho}$$

The total energy is sum of longitudinal part E_x and transverse part E_{ρ} .

• Evaluating the difference, the net current through the interface equals:

$$J = J_{1 \to 2} - J_{2 \to 1} = \frac{4\pi m_{eff} q}{h^3} \int_{E_{min}}^{E_{max}} T_c(E_x) dE_x \int_{0}^{\infty} [f_1(E) - f_2(E)] dE_{\rho}$$

• This expression is usually written as an integral over the product of two independent parts which only depend upon the energy perpendicular to the interface: The transmission coefficient $T_c(E_x)$ and the supply function $N(E_x)$

$$J = \frac{4\pi m_{eff} q}{h^3} \int_{E_{min}}^{E_{max}} T_c(E_x) N(E_x) dE_x$$

TSU-ESAKI FORMULA

- The expression in the previous slide is known as the Tsu-Esaki formula.
- The supply function describes the difference in the supply of carriers at the interfaces of the dielectric layer. Following the definition of the current, the supply function is given by:

$$N(E_{x}) = \int_{0}^{\infty} [f_{1}(E) - f_{2}(E)] dE_{\rho}$$

The occupancy functions f₁ and f₂ are defined near the interfaces. Since the exact shape of these distributions is usually not known, approximate shapes are commonly used. Furthermore, it is assumed that the distributions are isotropic.

SUPPLY FUNCTION

 In equilibrium, the energy distribution function of electrons and holes is given by the FERMI-DIRAC statistics

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_f}{k_B T}\right)}$$

• Which can be derived from statistical mechanics. Separating the longitudinal and the transverse energies $E=E_x+E_\rho$, and splitting the integral $N(E_x)=\xi_1(E_x)-\xi_2(E_x)$, the values of ξ_1 and ξ_2 become:

$$\xi_{i} = \int_{0}^{\infty} f_{i}(E) dE_{\rho} = \int_{0}^{\infty} \frac{1}{1 + \exp\left(\frac{E_{x} + E_{\rho} - E_{fi}}{k_{B}T}\right)} dE_{\rho}, \quad i=1,2$$

SUPPLY FUNCTION

• The last expression can be integrated analytically using:

$$\int \frac{dx}{1 + \exp(x)} = \ln\left(\frac{1}{1 + \exp(-x)}\right) + C$$

• Then the total supply function is:

$$N(E_x) = k_B T \ln \left(\frac{1 + \exp\left(\frac{E_{f1} - E_x}{k_B T}\right)}{1 + \exp\left(\frac{E_{f2} - E_x}{k_B T}\right)} \right)$$