Chapter 9

Two Dimensional Electron Gas, Quantum Wells & Semiconductor Superlattices

References:

- Ando, Fowler and Stern, *Rev. Mod. Phys.* 54 437 (1982).
- R.F. Pierret, *Field Effect Devices*, Vol. IV of Modular Series on Solid State Devices, Addison-Wesley (1983).
- B.G. Streetman, *Solid State Electronic Devices*, Series in Solid State Physical Electronics, Prentice-Hall (1980).

9.1 Two-Dimensional Electronic Systems

One of the most important recent developments in semiconductors, both from the point of view of physics and for the purpose of device developments, has been the achievement of structures in which the electronic behavior is essentially two-dimensional (2D). This means that, at least for some phases of operation of the device, the carriers are confined in a potential such that their motion in one direction is restricted and thus is quantized, leaving only a two-dimensional momentum or k-vector which characterizes motion in a plane normal to the confining potential. The major systems where such 2D behavior has been studied are MOS structures, quantum wells and superlattices. More recently, quantization has been achieved in 1-dimension (the quantum wires) and "zero"–dimensions (the quantum dots). These topics are further discussed in Chapter 10 and in the course on semiconductor physics (6.735J).

9.2 MOSFETS

One of the most useful and versatile of these structures is the metal-insulator-semiconductor (MIS) layered structures, the most important of these being the metal-oxide-semiconductor

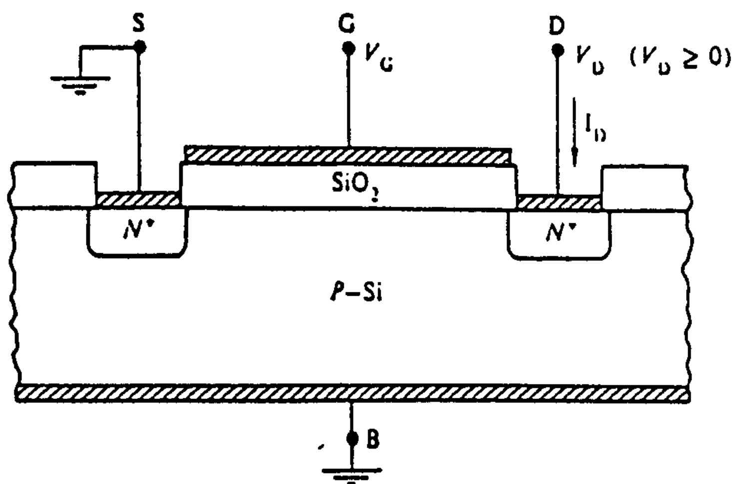


Figure 9.1: Cross-sectional view of the basic MOSFET structure showing the terminal designations and standard biasing conditions.

(MOS) structures. As shown in Fig. 9.1, the MOS device is fabricated from a substrate of usually moderately-doped p-type or n-type silicon which together with its grounded electrode is called the base and labeled B in the figure. On the top of the base is grown an insulating layer of silicon dioxide, followed by a metal layer; this structure is the gate (labeled G in the figure) and is used to apply an electric field through the oxide to the silicon. For the MOS device shown in the figure the base region is p-type and the source (S) and drain (D) regions are n-type. Measurements of the changes in the properties of the carriers in the silicon layer immediately below the gate (the conductance in the source-drain channel), in response to changes in the applied electric field at the gate electrode, are called field-effect measurements. As we show below, the field dramatically changes the conducting properties of the carriers beneath the gate. Use is made of this effect in the so-called metal-oxide-semiconductor field-effect transistor (MOSFET). To understand the operation of this device,we first consider the schematic energy band diagram of the MOS structure as shown in Fig. 9.2, for four different values of V_G, the gate potential relative to the substrate. For each V_G value, the diagram shows from left to right the metal (M) - oxide (O) - semiconductor (S) regions. In the semiconductor regions each of the diagrams show from top to bottom: the Si conduction band edge E_c, the "intrinsic" Fermi level for undoped Si as the dashed line, the Fermi level E_F in the p-type Si, and the valence band edge E_v. In each diagram, the central oxide region shows the valence band edge for the oxide. On the left hand side of each diagram, the Fermi level for the metal is shown and the dashed line gives the extension of the Si Fermi level. In the lower part of the figure, the charge layers of the interfaces for each case are illustrated.

We now explain the diagrams in Fig. 9.2 as a function of the gate voltage V_G. For $V_G = 0$ (the flat-band case), there are (ideally) no charge layers, and the energy levels of the metal (M) and semiconducting (S) regions line up to yield the same Fermi level (chemical potential). The base region is doped p-type. For a negative gate voltage ($V_G < 0$, the accumulation case), an electric field is set up in the oxide. The negative gate voltage

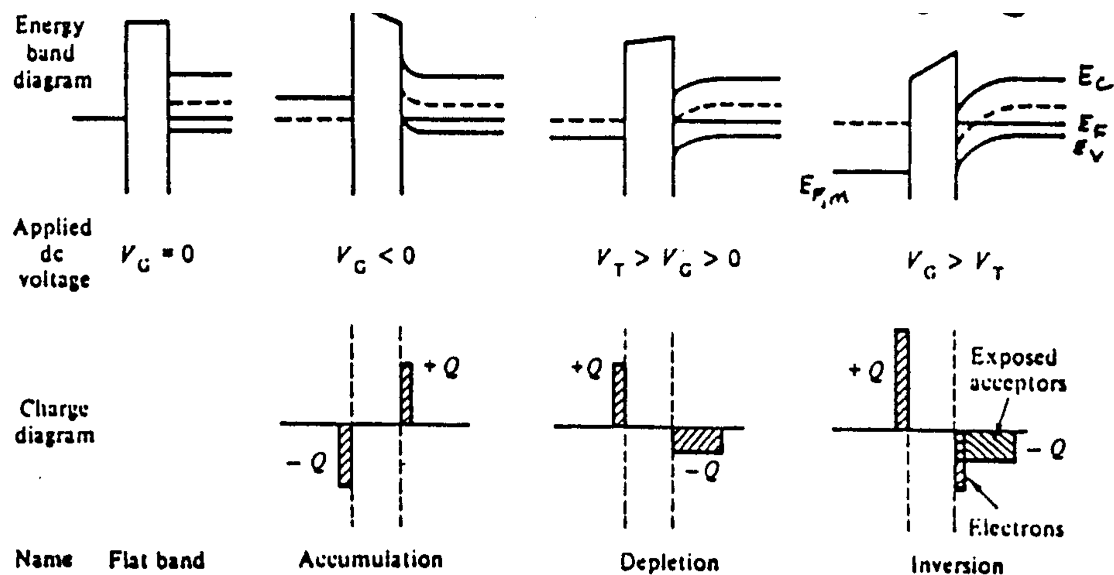


Figure 9.2: Energy band and block charge diagrams for a p–type device under flat band, accumulation, depletion and inversion conditions.

causes the Si bands to bend up at the oxide interface (see Fig. 9.2) so that the Fermi level is closer to the valence-band edge. Thus extra holes accumulate at the semiconductor-oxide interface and electrons accumulate at the metal-oxide interface (see lower part of Fig. 9.2). In the third (depletion) case, the gate voltage is positive but less than some threshold value V_T. The voltage V_T is defined as the gate voltage where the intrinsic Fermi level and the actual Fermi level are coincident at the interface (see lower part of Fig. 9.2). For the "depletion" regime, the Si bands bend down at the interface resulting in a depletion of holes, and a negatively charged layer of localized states is formed at the semiconductor-oxide interface. The size of this "depletion region" increases as V_G increases. The corresponding positively charged region at the metal-oxide interface is also shown. Finally, for $V_G > V_T$, the intrinsic Fermi level at the interface drops below the actual Fermi level, forming the "inversion layer", where mobile electrons reside. It is the electrons in this inversion layer which are of interest, both because they can be confined so as to exhibit two-dimensional behavior, and because they can be controlled by the gate voltage in the MOSFET (see Fig. 9.3)

The operation of a metal-oxide semiconductor field-effect transistor (MOSFET) is illustrated in Fig. 9.3, which shows the electron inversion layer under the gate for $V_G > V_T$ (for a p-type substrate), with the source region grounded, for various values of the drain voltage V_D. The inversion layer forms a conducting "channel" between the source and drain (as long as $V_G > V_T$). The dashed line in Fig. 9.3 shows the boundaries of the depletion region which forms in the p-type substrate adjoining the n^+ and p regions.

For $V_D = 0$ there is obviously no current between the source and the drain since both are at the same potential. For $V_D > 0$, the inversion layer or channel acts like a resistor, inducing the flow of electric current I_D. As shown in Fig. 9.3, increasing V_D imposes a reverse bias on the n^+-p drain-substrate junction, thereby increasing the width of the depletion region

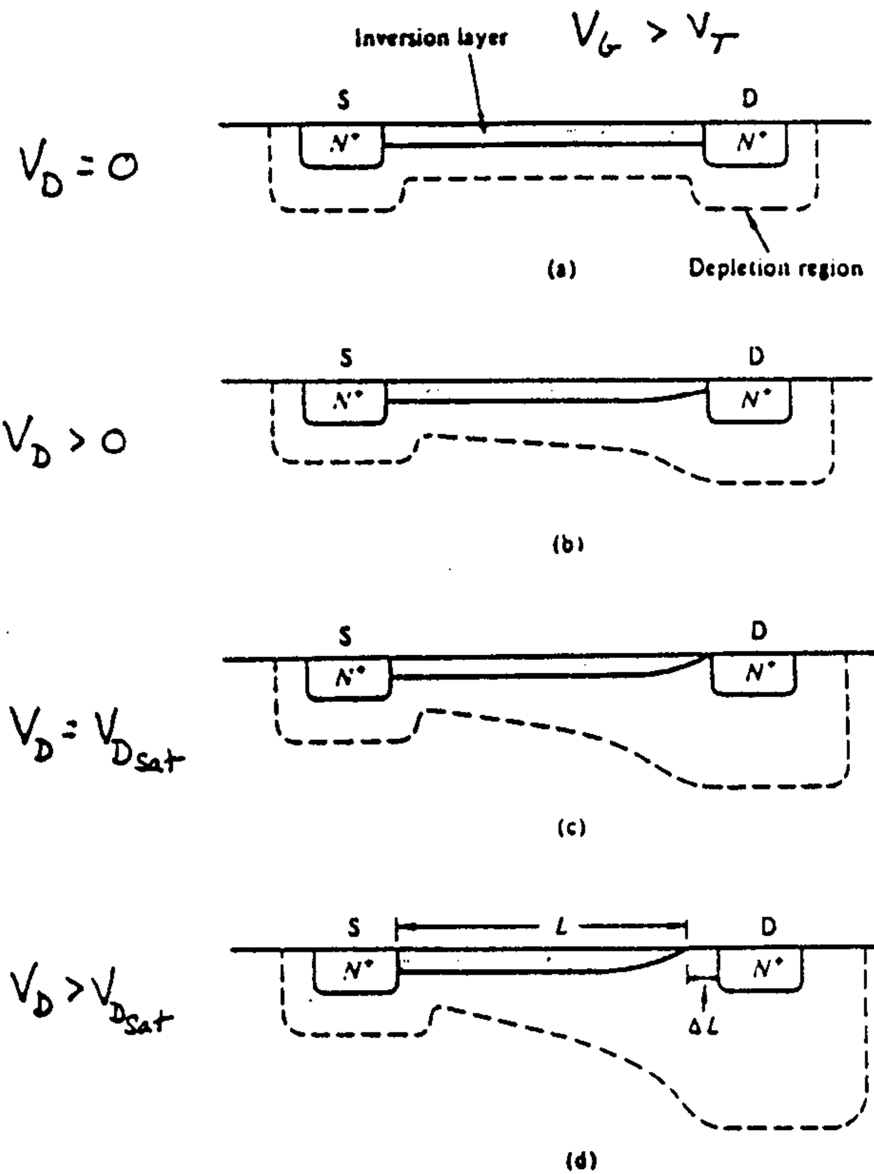


Figure 9.3: Visualization of various phases of $V_G > V_T$ MOSFET operation. (a) $V_D = 0$, (b) channel (inversion layer) narrowing under moderate V_D biasing, (c) pinch–off, and (d) post-pinch-off ($V_D > V_{Dsat}$) operation. (Note that the inversion layer widths, depletion widths, etc. are not drawn to scale.)

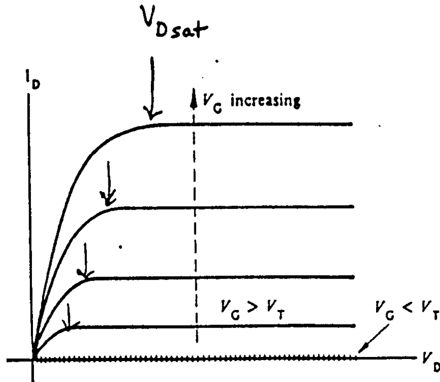


Figure 9.4: General form of the $I_D - V_D$ characteristics expected from a long channel ($\Delta L \ll L$) MOSFET.

and decreasing the number of carriers and narrowing the channel in the inversion layer as shown in Fig. 9.3. Finally as V_D increases further, the channel reaches the "pinched-off" condition V_{Dsat} shown in Fig. 9.3c. Further increase in V_D does not increase I_D but rather causes "saturation". We note that at saturation, $V_{Dsat} = V_G - V_T$. Saturation is caused by a decrease in the carrier density in the channel due to the pinch-off phenomena.

In Fig. 9.4 I_D vs. V_D curves are plotted for fixed values of $V_G > V_T$. We note that V_{Dsat} increases with increasing V_G. These characteristic curves are qualitatively similar to the curves for the bipolar junction transistor. The advantage of MOSFET devices lie in the speed of their operation and in the ease with which they can be fabricated into ultra-small devices.

The MOSFET device, or an array of a large number of MOSFET devices, is fabricated starting with a large Si substrate or "wafer". At each stage of fabrication, areas of the wafer which are to be protected are masked off using a light-sensitive substance called photoresist, which is applied as a thin film, exposed to light (or an electron or x-ray beam) through a mask of the desired pattern, then chemically developed to remove the photoresist from only the exposed (or, sometimes only the un–exposed) area. First the source and drain regions are formed by either diffusing or implanting (bombarding) donor ions into the p-type substrate. Then a layer of SiO_2 (which is an excellent and stable insulator) is grown by exposing the desired areas to an atmosphere containing oxygen; usually only a thin layer is grown over the gate regions and, in a separate step, thicker oxide layers are grown between neighboring devices to provide electrical isolation. Finally, the metal gate electrode, the source and drain contacts are formed by sputtering or evaporating a metal such as aluminum onto the desired regions.

9.3 Two-Dimensional Behavior

Other systems where two-dimensional behavior has been observed include heterojunctions of III-V compounds such as $GaAs/Ga_{1-x}Al_xAs$, layer compounds such as GaSe, $GaSe_2$ and related III-VI compounds, graphite and intercalated graphite, and electrons on the surface of liquid helium. The $GaAs/Ga_{1-x}Al_xAs$ heterojunctions are important for device

applications because the lattice constants and the coefficient of expansion of GaAs and $Ga_{1-x}Al_xAs$ are very similar. This lattice matching permits the growth of high mobility thin films of $Ga_{1-x}Al_xAs$ on a GaAs substrate.

The interesting physical properties of the MOSFET lie in the two-dimensional behavior of the electrons in the channel inversion layer at low temperatures. Studies of these electrons have provided important tests of modern theories of localization, electron-electron interactions and many-body effects. In addition, the MOSFETs have exhibited a highly unexpected property that, in the presence of a magnetic field normal to the inversion layer, the transverse or Hall resistance ρ_{xy} is quantized in integer values of e^2/h. This quantization is accurate to parts in 10^7 or 10^8 and provides the best measure to date of the fine structure constant $\alpha = e^2/hc$, when combined with the precisely-known velocity of light c. We will further discuss the quantized Hall effect later in the course (Part III).

We now discuss the two-dimensional behavior of the MOSFET devices in the absence of a magnetic field. The two-dimensional behavior is associated with the nearly plane wave electron states in the inversion layer. The potential $V(z)$ is associated with the electric field $V(z) = eEz$ and because of the negative charge on the electron, a potential well is formed containing bound states described by quantized levels. A similar situation occurs in the two–dimensional behavior for the case of electrons in quantum wells produced by molecular beam epitaxy. Explicit solutions for the bound states in quantum wells are given in §9.4. We discuss in the present section the form of the differential equation and of the resulting eigenvalues and eigenfunctions.

A single electron in a one-dimensional potential well $V(z)$ will, from elementary quantum mechanics, have discrete allowed energy levels E_n corresponding to bound states and usually a continuum of levels at higher energies corresponding to states which are not bound. An electron in a bulk semiconductor is in a three-dimensional periodic potential. In addition the potential causing the inversion layer of a MOSFET or a quantum well in GaAs/$Ga_{1-x}Al_xAs$ can be described by a one-dimensional confining potential $V(z)$ and can be written using the effective-mass theorem

$$[E(-i\vec{\nabla}) + \mathcal{H}']\Psi = i\hbar\left(\frac{\partial\Psi}{\partial t}\right) \tag{9.1}$$

where $\mathcal{H}' = V(z)$. The energy eigenvalues near the band edge can be written as

$$E(\vec{k}) = E(\vec{k}_0) + \frac{1}{2}\sum_{i,j}\left(\frac{\partial^2 E}{\partial k_i \partial k_j}\right)k_i k_j \tag{9.2}$$

so that the operator $E(-i\vec{\nabla})$ in Eq. 9.1 can be written as

$$E(-i\vec{\nabla}) = \sum_{i,j}\frac{p_i p_j}{2m_{i,j}} \tag{9.3}$$

where the p_i's are the operators

$$p_i = \frac{\hbar}{i}\frac{\partial}{\partial x_i} \tag{9.4}$$

which are substituted into Schrödinger's equation. The effect of the periodic potential is contained in the reciprocal of the effective mass tensor

$$\frac{1}{m_{ij}} = \frac{1}{\hbar^2}\frac{\partial^2 E(\vec{k})}{\partial k_i \partial k_j}\bigg|_{\vec{k}=\vec{k_0}} \tag{9.5}$$

where the components of $1/m_{ij}$ are evaluated at the band edge at $\vec{k}_0$.

If $1/m_{ij}$ is a diagonal matrix, the effective-mass equation $\mathcal{H}\Psi = E\Psi$ is solved by a function of the form

$$\Psi_{n,k_x,k_y} = e^{ik_x x} e^{ik_y y} f_n(z) \tag{9.6}$$

where $f_n(z)$ is a solution of the equation

$$-\frac{\hbar^2}{2m_{zz}} \frac{d^2 f_n}{dz^2} + V(z) f_n = E_{n,z} f_n \tag{9.7}$$

and the total energy is

$$E_n(k_x, k_y) = E_{n,z} + \frac{\hbar^2}{2m_{xx}} k_x^2 + \frac{\hbar^2}{2m_{yy}} k_y^2. \tag{9.8}$$

Since the $E_{n,z}$ energies (n=0,1,2,...) are discrete, the energies states $E_n(k_x, k_y)$ for each n value form a "sub-band". We give below (in §9.3.1) a simple derivation for the discrete energy levels by considering a particle in various potential wells (i.e., quantum wells). The electrons in these "sub–bands" form a 2D electron gas.

9.3.1 Quantum Wells and Superlattices

Many of the quantum wells and superlattices that are commonly studied today do not occur in nature, but rather are deliberately structured materials (see Fig. 9.5). In the case of superlattices formed by molecular beam epitaxy, the quantum wells result from the different bandgaps of the two constituent materials. The additional periodicity is in one–dimension (1–D) which we take along the z–direction, and the electronic behavior is usually localized on the basal planes (x–y planes) normal to the z–direction, giving rise to two–dimensional behavior.

A schematic representation of a semiconductor heterostructure superlattice is shown in Fig. 9.5 where d is the superlattice periodicity composed of a distance d_1, of semiconductor S_1, and d_2 of semiconductor S_2. Because of the different band gaps in the two semiconductors, potential wells and barriers are formed. For example in Fig. 9.5, the barrier heights in the conduction and valence bands are ΔE_c and ΔE_v respectively. In Fig. 9.5 we see that the difference in bandgaps between the two semiconductors gives rise to band offsets ΔE_c and ΔE_v for the conduction and valence bands. In principle, these band offsets are determined by matching the Fermi levels for the two semiconductors. In actual materials, the Fermi levels are highly sensitive to impurities, defects and charge transfer at the heterojunction interface.

The two semiconductors of a heterojunction superlattice could be different semiconductors such as InAs with GaP (see Table 9.1 for parameters related to these compounds) or a binary semiconductor with a ternary alloy semiconductor, such as GaAs with $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ (sometimes referred to by their slang names "Gaas" and "Algaas"). In the typical semiconductor superlattices the periodicity $d = d_1 + d_2$ is repeated many times (e.g., 100 times). The period thicknesses typically vary between a few layers and many layers (10Å to 500Å). Semiconductor superlattices are today an extremely active research field internationally.

The electronic states corresponding to the heterojunction superlattices are of two fundamental types–bound states in quantum wells and nearly free electron states in zone–folded energy bands. In this course, we will limit our discussion to the bound states in a single infinite quantum well. Generalizations to multiple quantum wells will be made subsequently.

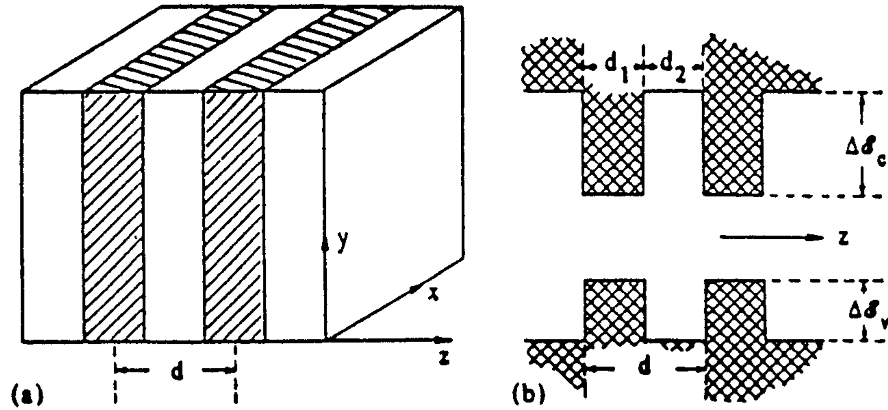


Figure 9.5: (a) A heterojunction superlattice of periodicity d. (b) Each superlattice unit cell consists of a thickness d_1 of material #1 and d_2 of material #2. Because of the different band gaps, a periodic array of potential wells and potential barriers is formed. When the band offsets are both positive as shown in this figure, the structure is called a type I superlattice.

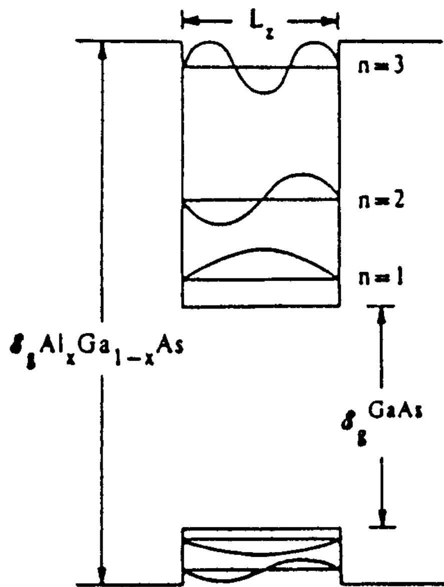


Figure 9.6: The eigenfunctions and bound state energies of an infinitely deep potential well used as an approximation to the states in two finite wells. The upper well applies to electrons and the lower one to holes. This diagram is a schematic representation of a quantum well in the GaAs region formed by the adjacent wider gap semiconductor $Al_xGa_{1-x}As$.

Table 9.1: Material parameters of GaAs, GaP, InAs, and InP.[1]

Property	Parameter (units)	GaAs	GaP	InAs	InP
Lattice constant	a(Å)	5.6533	5.4512	6.0584	5.8688
Density	$g(\text{g/cm}^3)$	5.307	4.130	5.667	4.787
Thermal expansion	$\alpha_{th}(\times 10^{-6}/^\circ\text{C})$	6.63	5.91	5.16	4.56
Γ point band gap	E_0(eV)	1.42	2.74	0.36	1.35
plus spin orbit	$E_0 + \Delta_0$(eV)	1.76	2.84	0.79	1.45
L point band gap	E_1(eV)	2.925	3.75	2.50	3.155
plus spin orbit	$E_1 + \Delta_1$(eV)	3.155	...	2.78	3.305
Γ point band gap	$E_0{}'$(eV)	4.44	4.78	4.44	4.72
Δ axis band gap	E_2(eV)	4.99	5.27	4.70	5.04
plus spin orbit	$E_2 + \delta$(eV)	5.33	5.74	5.18	5.60
Gap pressure coefficient	$\partial E_0/\partial P(\times 10^{-6}\text{eV/bar})$	11.5	11.0	10.0	8.5
Gap temperature coefficient	$\partial E_0/\partial T(\times 10^{-4}\text{eV}/^\circ\text{C})$	−3.95	−4.6	−3.5	−2.9
Electron mass	m^*/m_0	0.067	0.17	0.023	0.08
light hole	$m_{\ell h}{}^*/m_0$	0.074	0.14	0.027	0.089
heavy hole	$m_{hh}{}^*/m_0$	0.62	0.79	0.60	0.85
spin orbit hole	$m_{so}{}^*/m_0$	0.15	0.24	0.089	0.17
Dielectric constant: static	ϵ_s	13.1	11.1	14.6	12.4
Dielectric constant: optic	ϵ_∞	11.1	8.46	12.25	9.55
Ionicity	f_1	0.310	0.327	0.357	0.421
Polaron coupling	α_F	0.07	0.20	0.05	0.08
Elastic constants	$c_{11}(\times 10^{11}\text{dyn/cm}^2)$	11.88	14.120	8.329	10.22
	$c_{12}(\times 10^{11}\text{dyn/cm}^2)$	5.38	6.253	4.526	5.76
	$c_{44}(\times 10^{11}\text{dyn/cm}^2)$	5.94	7.047	3.959	4.60
Young's modulus	$Y(\times 10^{11}\text{dyn/cm}^2)$	8.53	10.28	5.14	6.07
	P	0.312	0.307	0.352	0.360
Bulk modulus	$B(\times 10^{11}\text{ dyn/cm}^2)$	7.55	8.88	5.79	7.25
	A	0.547	0.558	0.480	0.485
Piezo–electric coupling	$e_{14}(\text{C/m}^2)$	−0.16	−0.10	−0.045	−0.035
	$K_{[110]}$	0.0617	0.0384	0.0201	0.0158
Deformation potential	a(eV)	2.7	3.0	2.5	2.9
	b(eV)	−1.7	−1.5	−1.8	−2.0
	d(eV)	−4.55	−4.6	−3.6	−5.0
Deformation potential	Ξ_{eff}(eV)	6.74	6.10	6.76	7.95
Donor binding	G(meV)	4.4	10.0	1.2	5.5
Donor radius	a_B(Å)	136	48	406	106
Thermal conductivity	$\kappa(\text{watt/deg} - \text{cm})$	0.46	0.77	0.273	0.68
Electron mobility	$\mu_n(\text{cm}^2/\text{V} - \text{sec})$	8000	120	30000	4500
Hole mobility	$\mu_p(\text{cm}^2/\text{V} - \text{sec})$	300	–	450	100

[1]Table from *J. Appl. Physics* 53, 8777 (1982).

9.4 Bound Electronic States

From the diagram in Fig. 9.5 we see that the heterojunction superlattice consists of an array of potential wells. The interesting limit to consider is the case where the width of the potential well contains only a small number of crystallographic unit cells ($L_z < 100$ Å), in which case the number of bound states in the well is a small number.

From a mathematical standpoint, the simplest case to consider is an infinitely deep rectangular potential well. In this case, a particle of mass m^* in a well of width L_z in the z direction satisfies the free particle Schrödinger equation

$$-\frac{\hbar^2}{2m^*}\frac{d^2\psi}{dz^2} = E\psi \tag{9.9}$$

with eigenvalues

$$E_n = \frac{\hbar^2}{2m^*}\left(\frac{n\pi}{L_z}\right)^2 = \left(\frac{\hbar^2\pi^2}{2m^*L_z^2}\right)n^2 \tag{9.10}$$

and the eigenfunctions

$$\psi_n = A\sin(n\pi z/L_z) \tag{9.11}$$

where n = 1,2,3.... are the plane wave solutions that satisfy the boundary conditions that the wave functions in Eq. 9.11 must vanish at the walls of the quantum wells ($z = 0$ and $z = L_z$).

We note that the energy levels are not equally spaced, but have energies $E_n \sim n^2$, though the spacings $E_{n+1} - E_n$ are proportional to n. We also note that $E_n \sim {L_z}^{-2}$, so that as L_z becomes large, the levels become very closely spaced as expected for a 3D semiconductor. However when L_z decreases, the number of states in the quantum well decreases, so that for a well depth E_d it would seem that there is a critical width ${L_z}^c$ below which there would be no bound states

$${L_z}^c = \frac{\hbar\pi}{(2m^*E_d)^{\frac{1}{2}}}. \tag{9.12}$$

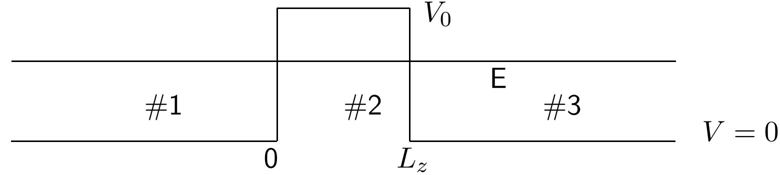
An estimate for ${L_z}^c$ is obtained by taking $m^* = 0.1m_0$ and $E_d = 0.1$ eV to yield ${L_z}^c = 61$Å. There is actually a theorem in quantum mechanics that says that there will be at least one bound state for an arbitrarily small potential well. More exact calculations considering quantum wells of finite thickness have been carried out, and show that the infinite well approximation gives qualitatively correct results.

The closer level spacing of the valence band bound states in Fig. 9.6 reflects the heavier masses in the valence band. Since the states in the potential well are quantized, the structures in Figs. 9.5 and 9.6 are called quantum well structures.

If the potential energy of the well V_0 is not infinite but finite, the wave functions are similar to those given in Eq. 9.11, but will have decaying exponentials on either side of the potential well walls. The effect of the finite size of the well on the energy levels and wave functions is most pronounced near the top of the well. When the particle has an energy greater than V_0, its eigenfunction corresponds to a continuum state $\exp(ik_zz)$.

In the case of MOSFETs, the quantum well is not of rectangular shape as shown in Fig. 9.7, but rather is approximated as a triangular well. The solution for the bound states in a triangular well cannot be solved exactly, but can only be done approximately, as for example using the WKB approximation described in §9.6.

Figure 9.7: Schematic of a potential barrier.



9.5 Review of Tunneling Through a Potential Barrier

When the potential well is finite, the wave functions do not completely vanish at the walls of the well, so that tunneling through the potential well becomes possible. We now briefly review the quantum mechanics of tunneling through a potential barrier. We will return to tunneling in semiconductor heterostructures after some introductory material.

Suppose that the potential V shown in Fig. 9.7 is zero ($V = 0$) in regions #1 and #3, while $V = V_0$ in region #2. Then in regions #1 and #3

$$E = \frac{\hbar^2 k^2}{2m^*} \tag{9.13}$$

$$\psi = e^{ikz} \tag{9.14}$$

while in region #2 the wave function is exponentially decaying

$$\psi = \psi_0 e^{-\beta z} \tag{9.15}$$

so that substitution into Schrödinger's equation gives

$$\frac{-\hbar^2}{2m^*}\beta^2\psi + (V_0 - E)\psi = 0 \tag{9.16}$$

or

$$\beta^2 = \frac{2m^*}{\hbar^2}(V_0 - E). \tag{9.17}$$

The probability that the electron tunnels through the rectangular potential barrier is then given by

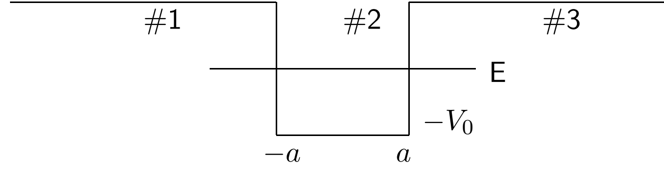
$$\mathcal{P} = \exp\left\{-2\int_0^{L_z}\beta(z)dz\right\} = \exp\left\{-2\left(\frac{2m^*}{\hbar^2}\right)^{\frac{1}{2}}(V_0 - E)^{\frac{1}{2}}L_z\right\} \tag{9.18}$$

As L_z increases, the probability of tunneling decreases exponentially. Electron tunneling phenomena frequently occur in solid state physics.

9.6 Quantum Wells of Different Shape and the WKB Approximation

With the sophisticated computer control available with state of the art molecular beam epitaxy systems it is now possible to produce quantum wells with specified potential profiles $V(z)$ for semiconductor heterojunction superlattices. Potential wells with non–rectangular

Figure 9.8: Schematic of a rectangular well.



profiles also occur in the fabrication of other types of superlattices (e.g., by modulation doping). We therefore briefly discuss (in the recitation class) bound states in general potential wells.

In the general case where the potential well has an arbitrary shape, solution by the WKB (Wentzel–Kramers–Brillouin) approximation is very useful (see for example, Shanker, "Principles of Quantum Mechanics", Plenum press, chapter 6). According to this approximation, the energy levels satisfy the Bohr–Sommerfeld quantization condition

$$\int_{z_1}^{z_2} p_z dz = \hbar\pi(r + c_1 + c_2) \tag{9.19}$$

where $p_z = (2m^*[E - V])^{\frac{1}{2}}$ and the quantum number r is an integer $r = 0, 1, 2, \ldots$ while c_1 and c_2 are the phases which depend on the form of $V(z)$ at the turning points z_1 and z_2 where $V(z_i) = E$. If the potential has a sharp discontinuity at a turning point, then $c = 1/2$, but if V depends linearly on z at the turning point then $c = 1/4$.

For example for the infinite rectangular well (see Fig. 9.8)

$$V(z) = 0 \quad \text{for} \;\mid z \mid < a \quad \text{(inside the well)} \tag{9.20}$$

$$V(z) = \infty \quad \text{for} \;\mid z \mid > a \quad \text{(outside the well)} \tag{9.21}$$

By the WKB rules, the turning points occur at the edges of the rectangular well and therefore $c_1 = c_2 = 1/2$. In this case p_z is a constant, independent of z so that $p_z = (2m^*E)^{\frac{1}{2}}$ and Eq. 9.19 yields

$$(2m^*E)^{\frac{1}{2}} L_z = \hbar\pi(r + 1) = \hbar\pi n \tag{9.22}$$

where $n = r + 1$ and

$$E_n = \frac{\hbar^2\pi^2}{2m^*{L_z}^2} n^2 \tag{9.23}$$

in agreement with the exact solution given by Eq. 9.10. The finite rectangular well shown in Fig. 9.8 is thus approximated as an infinite well with solutions given by Eq. 9.10.

As a second example consider a harmonic oscillator potential well shown in Fig. 9.9, where $V(z) = m^*\omega^2 z^2/2$. The harmonic oscillator potential well is typical of quantum wells in periodically doped (nipi which is n-type; insulator; p-type; insulator) superlattices. In this case

$$p_z = (2m^*)^{1/2}\left(E - \frac{m^*\omega^2}{2} z^2\right)^{\frac{1}{2}}. \tag{9.24}$$

The turning points occur when $V(z) = E$ so that the turning points are given by $z = \pm(2E/m^*\omega^2)^{\frac{1}{2}}$. Near the turning points $V(z)$ is approximately linear in z, so the phase

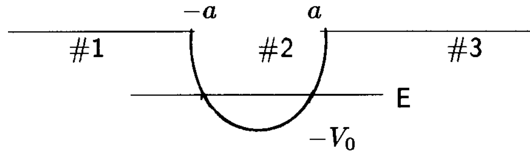
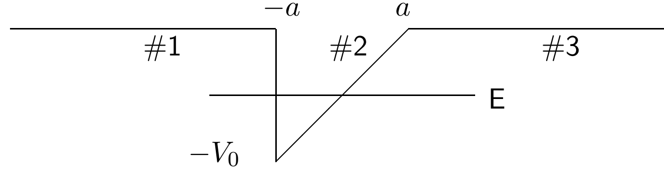


Figure 9.9: Schematic of a harmonic oscillator well.

Figure 9.10: Schematic of a triangular well.



factors become $c_1 = c_2 = \frac{1}{4}$. The Bohr–Sommerfeld quantization thus yields

$$\int_{z_1}^{z_2} p_z dz = \int_{z_1}^{z_2} (2m^*)^{\frac{1}{2}} \left(E - \frac{m^*\omega^2}{2} z^2\right)^{\frac{1}{2}} dz = \hbar\pi(r + \frac{1}{2}). \tag{9.25}$$

Making use of the integral relation

$$\int \sqrt{a^2 - u^2}\, du = \frac{u}{2}\sqrt{a^2 - u^2} + \frac{a^2}{2}\sin^{-1}\frac{u}{a} \tag{9.26}$$

we obtain upon substitution of Eq. 9.26 into 9.25:

$$(2m^*)^{\frac{1}{2}} \left(\frac{m^*\omega^2}{2}\right)^{\frac{1}{2}} \left(\frac{E_r}{m^*\omega^2}\right)\pi = \frac{E_r\pi}{\omega} = \hbar\pi(r + \frac{1}{2}) \tag{9.27}$$

which simplifies to the familiar relation for the harmonic oscillator energy levels:

$$E_r = \hbar\omega(r + \frac{1}{2}) \quad \text{where} \quad r = 0, 1, 2... \tag{9.28}$$

another example of an exact solution. For homework, you will use the WKB method to find the energy levels for an asymmetric triangular well. Such quantum wells are typical of the interface of metal–insulator–semiconductor (MOSFET) device structures (see Fig. 9.10).

9.7 The Kronig–Penney Model

We review here the Kronig–Penney model which gives an explicit solution for a one–dimensional array of finite potential wells shown in Fig. 9.11. Starting with the one dimensional Hamiltonian with a periodic potential (see Eq. 9.7)

$$-\frac{\hbar^2}{2m^*}\frac{d^2\psi}{dz^2} + V(z)\psi = E\psi \tag{9.29}$$

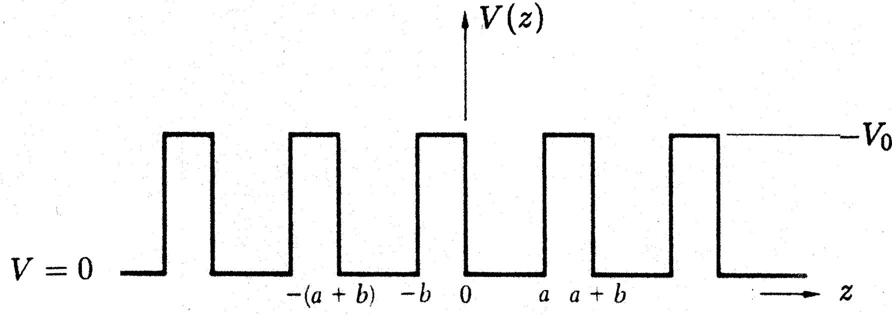


Figure 9.11: Kronig–Penney square well periodic potential

we obtain solutions in the region $0 < z < a$ where $V(z) = 0$

$$\psi(z) = Ae^{iKz} + Be^{-iKz} \tag{9.30}$$

$$E = \frac{\hbar^2 K^2}{2m^*} \tag{9.31}$$

and in the region $-b < z < 0$ where $V(z) = V_0$ (the barrier region)

$$\psi(z) = Ce^{\beta z} + De^{-\beta z} \tag{9.32}$$

where

$$\beta^2 = \frac{2m^*}{\hbar^2}[V_0 - E]. \tag{9.33}$$

Continuity of $\psi(z)$ and $d\psi(z)/dz$ at $z = 0$ and $z = a$ determines the coefficients A, B, C, D. At $z = 0$ we have:

$$\begin{aligned} A + B &= C + D \\ iK(A - B) &= \beta(C - D) \end{aligned} \tag{9.34}$$

At $z = a$, we apply Bloch's theorem (see Fig. 9.11), introducing a factor $\exp[ik(a + b)]$ to obtain $\psi(a) = \psi(-b)\exp[ik(a + b)]$

$$\begin{aligned} Ae^{iKa} + Be^{-iKa} &= (Ce^{-\beta b} + De^{\beta b})e^{ik(a+b)} \\ iK(Ae^{iKa} - Be^{-iKa}) &= \beta(Ce^{-\beta b} - De^{\beta b})e^{ik(a+b)}. \end{aligned} \tag{9.35}$$

These 4 equations (Eqs. 9.34 and 9.35) in 4 unknowns determine A, B, C, D. The vanishing of the coefficient determinant restricts the conditions under which solutions to the Kronig–Penney model are possible, leading to the algebraic equation

$$\frac{\beta^2 - K^2}{2\beta K}\sinh\beta b \sin Ka + \cosh\beta b \cos Ka = \cos k(a + b) \tag{9.36}$$

which has solutions for a limited range of β values.

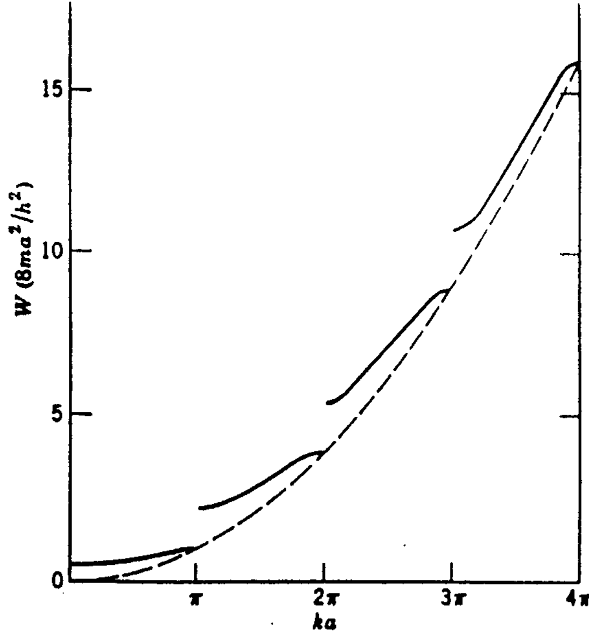


Figure 9.12: Plot of energy vs. k for the Kronig–Penney model with $P = 3\pi/2$. (After Sommerfeld and Bethe.)

Normally the Kronig–Penney model in the textbooks is solved in the limit $b \to 0$ and $V_0 \to \infty$ in such a way that $[\beta^2 ba/2] = P$ remains finite. The restricted solutions in this limit lead to the energy bands shown in Fig. 9.12.

For the superlattice problem we are interested in solutions both within the quantum wells and in the continuum. This is one reason for discussing the Kronig–Penny model. Another reason for discussing this model is because it provides a review of boundary conditions and the application of Bloch's theorem. In the quantum wells, the permitted solutions give rise to narrow bands with large band gaps while in the continuum regions the solutions correspond to wide bands and small band gaps.

9.8 3D Motion within a 1–D Rectangular Well

The thin films used for the fabrication of quantum well structures (see §9.4) are very thin in the z–direction but have macroscopic size in the perpendicular x–y plane. An example of a quantum well structure would be a thin layer of GaAs sandwiched between two thicker $Al_xGa_{1-x}As$ layers, as shown in the Fig. 9.5. For the thin film, the motion in the x and y directions is similar to that of the corresponding bulk solid which can be treated by the conventional 1–electron approximation and the Effective Mass Theorem. Thus the potential can be written as a sum of a periodic term $V(x, y)$ and the quantum well term $V(z)$. The electron energies thus are superimposed on the quantum well energies, the periodic solutions obtained from solution of the 2–D periodic potential

$$E_n(k_x, k_y) = E_{n,z} + \frac{\hbar^2(k_x^2 + k_y^2)}{2m^*} = E_{n,z} + E_\perp \tag{9.37}$$

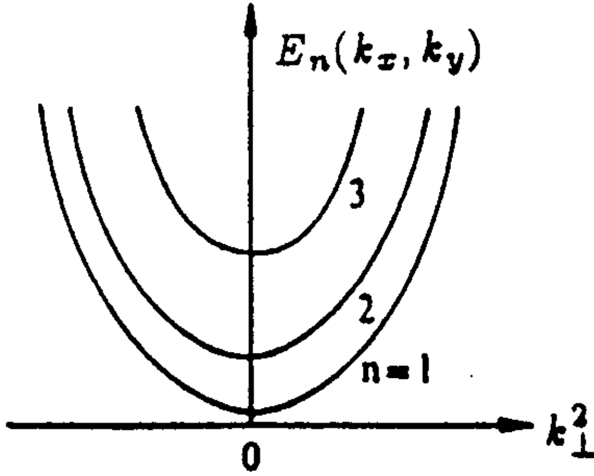


Figure 9.13: Subbands associated with bound states for the 2D electron gas.

in which the quantized bound state energies $E_{n,z}$ are given by Eq. 9.10. A plot of the energy levels is given in Fig. 9.13. At $(k_x, k_y) = (0,0)$ the energy is precisely the quantum well energy E_n for all n. The band of energies associated with each state n is called a subband.

Of particular interest is the density of states for the quantum well structures. Associated with each two–dimensional subband is a constant density of states, as derived below. From elementary considerations the number of electrons per unit area in a 2–dimensional circle is given by

$$N_{2D} = \frac{2}{(2\pi)^2}\pi k_\perp^2 \tag{9.38}$$

where $k_\perp^2 = k_x^2 + k_y^2$ and

$$E_\perp = \frac{\hbar^2 k_\perp^2}{2m^*} \tag{9.39}$$

so that for each subband the density of states $g_{2D}(E)$ contribution becomes

$$\frac{\partial N_{2D}}{\partial E} = g_{2D}(E) = \frac{m^*}{\pi\hbar^2}. \tag{9.40}$$

If we now plot the density of states corresponding to the 3D motion in a 1–D rectangular well, we have $g_{2D}(E) = 0$ until the bound state energy E_1 is reached, when a step function contribution of $(m^*/\pi\hbar^2)$ is made. The density of states $g_{2D}(E)$ will then remain constant until the minimum of subband E_2 is reached when an additional step function contribution of $(m^*/\pi\hbar^2)$ is made, hence yielding the staircase density of states shown in Fig. 9.14. Two generalizations of Eq. 9.40 for the density of states for actual quantum wells are needed, as we discuss below. The first generalization takes into account the finite size L_z of the quantum well, so that the system is not completely two dimensional and some k_z dispersion must occur. Secondly, the valence bands of typical semiconductors are degenerate so that coupling between the valence band levels occurs, giving rise to departures from the simple parabolic bands discussed below.

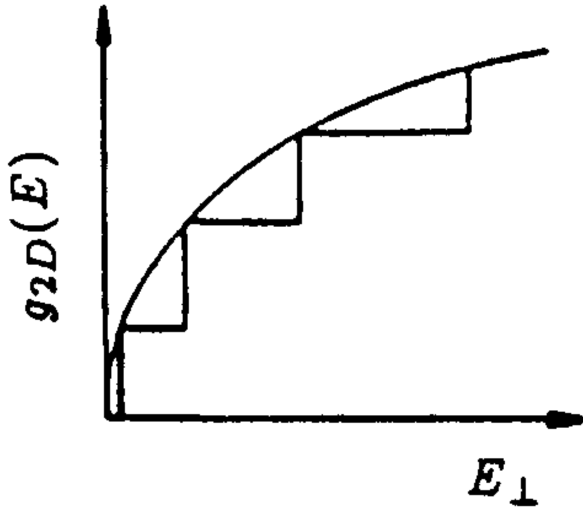


Figure 9.14: Two dimensional density of states $g_{2D}(E)$ for rectangular quantum well structures.

A comparison between the energy dispersion relation $E(\vec{k})$ and the density of states $g(E)$ in two dimensions and three dimensions is shown in Fig. 9.15 together with a quasi two–dimensional case, typical of actual quantum well samples. In the quasi–two dimensional case, the $E(\vec{k})$ relations exhibit a small degree of dispersion along k_z, leading to a corresponding width in the steps of the density of states function shown in Fig. 9.15(b).

A generalization of the simple 2D density of states in Fig. 9.14 is also necessary to treat the complex valence band of a typical III-V compound semiconductor. The $E(\vec{k})$ diagram (where $k_\perp$ is normal to k_z) for the heavy hole and light hole levels can be calculated using $\vec{k} \cdot \vec{p}$ perturbation theory to be discussed later in the course.

The most direct evidence for bound states in quantum wells comes from optical absorption measurements (to be discussed later in the course) and resonant tunneling effects which we discuss below.

9.9 Resonant Tunneling in Quantum Wells

Resonant tunneling (see Fig. 9.18) provides direct evidence for the existence of bound states in quantum wells. We review first the background material for tunneling across potential barriers in semiconductors and then apply these concepts to the resonant tunneling phenomenon.

The carriers in the quantum well structures are normally described in terms of the effective mass theorem where the wave functions for the carriers are given by the one electron approximation. The effective mass equation is written in terms of slowly varying wavefunctions corresponding to a slowly varying potential which satisfies Poisson's equation when an electric field is applied (e.g., a voltage is imposed across the quantum well structure).

Further simplifications that are made in treating the tunneling problem include:

1. The wavefunctions for the tunneling particle are expanded in terms of a single band

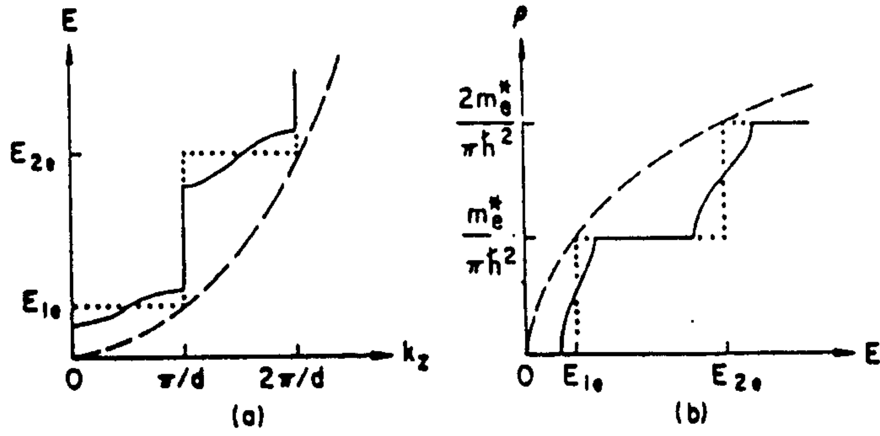


Figure 9.15: Schematic diagrams of (a) energy dispersion and (b) density of states. Indicated are the two–dimensional (dotted), three–dimensional (dashed), and intermediate (solid) cases.

on either side of the junction.

2. Schrödinger's equation is separated into two components, parallel and perpendicular to the junction plane, leading to a 1–dimensional tunneling problem.

3. The eigenstates of interest have energies sufficiently near those of critical points in the energy band structure on both sides of the interface so that the simplified form of the effective mass theorem can be used.

4. The total energy, E, and the momentum parallel to the interface or perpendicular to the layering direction, $k_\perp$, are conserved in the tunneling process. Since the potential acts only in the z–direction, the 1–dimensional Schrödinger equation becomes:

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dz^2} + V(z) - E\right]\psi_e = 0 \tag{9.41}$$

where $V(z)$ is the electrostatic potential, and ψ_e is an envelope function. The wave function ψ_e is subject, at an interface $z = z_1$ (see Fig. 9.16), to the following boundary conditions that guarantee current conservation:

$$\psi_e(z_1^-) = \psi_e(z_1^+) \tag{9.42}$$

$$\frac{1}{m_1}\left.\frac{d}{dz}\psi_e\right|_{z_1^-} = \frac{1}{m_2}\left.\frac{d}{dz}\psi_e\right|_{z_1^+} \tag{9.43}$$

The current density for tunneling through a barrier becomes

$$J_z = \frac{e}{4\pi^3\hbar}\int dk_z d^2k_\perp f(E)T(E_z)\frac{dE}{dk_z} \tag{9.44}$$

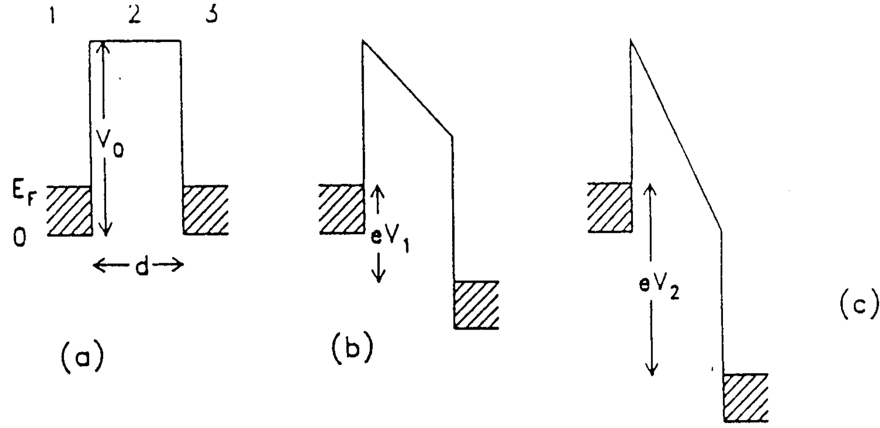


Figure 9.16: Rectangular–potential model (a) used to describe the effect of an insulator, 2, between two metals, 1 and 3. When a negative bias is applied to 1, electrons, with energies up to the Fermi energy E_F, can tunnel through the barrier. For small voltages, (b), the barrier becomes trapezoidal, but at high bias (c), it becomes triangular.

where $f(E)$ is the Fermi–Dirac distribution, and $T(E_z)$ is the probability of tunneling through the potential barrier. Here $T(E_z)$ is expressed as the ratio between the transmitted and incident probability currents.

If an external bias V is applied to the barrier (see Fig. 9.16), the net current flowing through it is the difference between the current from left to right and that from right to left. Thus, we obtain:

$$J_z = \frac{e}{4\pi^3\hbar}\int dE_z d^2k_\perp[f(E) - f(E+eV)]T(E_z) \tag{9.45}$$

where E_z represents the energy from the k_z component of crystal momentum, i.e., $E_z = \hbar^2 k_z^2/(2m)$. Since the integrand is not a function of $k_\perp$ in a plane normal to k_z, we can integrate over $d^2k_\perp$ by writing

$$dk_x dk_y = d^2k_\perp = \frac{2m}{\hbar^2}dE_\perp \tag{9.46}$$

where $E_\perp = \hbar^2 k_\perp^2/(2m)$ and after some algebra, the tunneling current can be written as,

$$\begin{aligned} J_z &= \tfrac{em}{2\pi^2\hbar^3}\left[eV\int_0^{E_F-eV} dE_z T(E_z) + \int_{E_F-eV}^{E_F} dE_z(E_F-E_z)T(E_z)\right] \quad \text{if } eV \le E_F \\ J_z &= \tfrac{em}{2\pi^2\hbar^3}\int_0^{E_F} dE_z(E_F-E_z)T(E_z) \quad \text{if } eV \ge E_F \end{aligned} \tag{9.47}$$

(see Fig. 9.16 for the geometry of the model) which can be evaluated as long as the tunneling probability through the barrier is known. We now discuss how to find the tunneling probability.

An enhanced tunneling probability occurs for certain voltages as a consequence of the constructive interference between the incident and the reflected waves in the barrier region between regions 1 and 3. To produce an interference effect the wavevector $\vec{k}$ in the plane

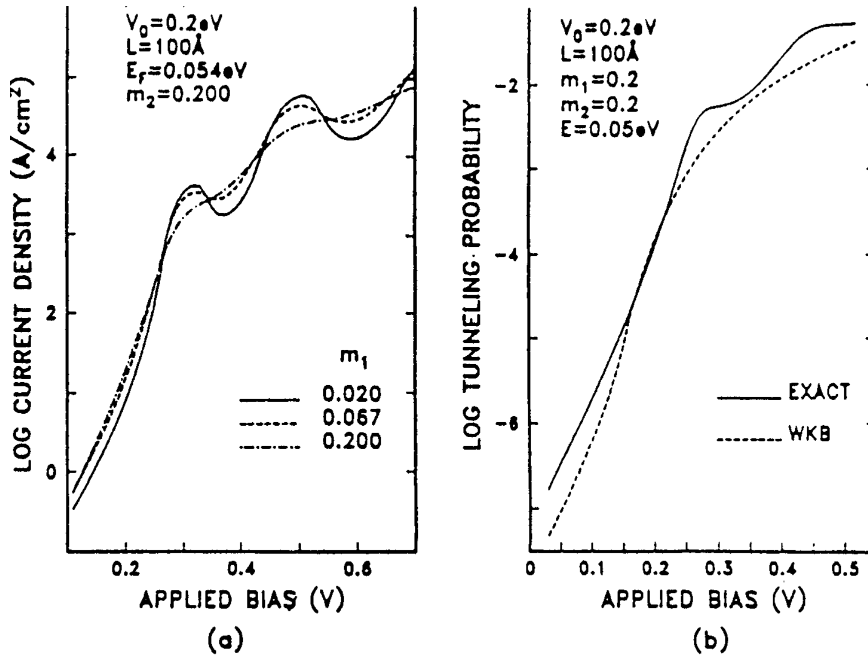


Figure 9.17: (a) Tunneling current through a rectangular barrier (like the one of Fig. 9.16a) calculated as a function of bias for different values of m_1, in the quantum well. (b) Comparison of an exact calculation of the tunneling probability through a potential barrier under an external bias with an approximate result obtained using the WKB method. The barrier parameters are the same as in (a), and the energy of an incident electron, of mass $0.2m_0$, is 0.05eV. (From the book of E.E. Mendez and K. von Klitzing, "Physics and Applications of Quantum Wells and Superlattices", NATO ASI Series, Vol. 170, p.159 (1987)).

wave solution e^{ikz} must have a real component so that an oscillating (rather than a decaying exponential) solution is possible. To accomplish this, it is necessary for a sufficiently high electric field to be applied (as in Fig. 9.16(c)) so that a virtual bound state is formed. As can be seen in Fig. 9.17a, the oscillations are most pronounced when the difference between the electronic mass at the barrier and at the electrodes is the largest. This interference phenomenon is frequently called resonant Fowler–Nordheim tunneling and has been observed in metal–oxide–semiconductor (MOS) heterostructures and in GaAs/$\mathrm{Ga_{1-x}Al_xAs}$/GaAs capacitors. Since the WKB method is semiclassical, it does not give rise to the resonant tunneling phenomenon, which is a quantum interference effect.

For the calculation of the resonant tunneling phenomenon, we must therefore use the quantum mechanical solution. In this case, it is convenient to use the transfer–matrix method to find the tunneling probability. In region (#1) of Fig. 9.16, the potential $V(z)$ is constant and solutions to Eq. 9.41 have the form

$$\psi_e(z) = A\exp(ikz) + B\exp(-ikz) \tag{9.48}$$

with

$$\frac{\hbar^2 k^2}{2m} = E - V. \tag{9.49}$$

When $E - V > 0$, then k is real and the wave functions are plane waves. When $E - V < 0$, then k is imaginary and the wave functions are growing or decaying waves. The boundary conditions Eqs. 9.42 and 9.43 determine the coefficients A and B which can be described by a (2×2) matrix R such that

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = R \begin{pmatrix} A_2 \\ B_2 \end{pmatrix} \tag{9.50}$$

where the subscripts on A and B refer to the region index and R can be written as

$$R = \frac{1}{2k_1 m_2} \begin{pmatrix} (k_1 m_2 + k_2 m_1) \exp[i(k_2 - k_1) z_1] & (k_1 m_2 - k_2 m_1) \exp[-i(k_2 + k_1) z_1] \\ (k_1 m_2 - k_2 m_1) \exp[i(k_2 + k_1) z_1] & (k_1 m_2 + k_2 m_1) \exp[-i(k_2 - k_1) z_1] \end{pmatrix} \tag{9.51}$$

and the terms in R of Eq. 9.51 are obtained by matching boundary conditions as given in Eqs. 9.42 and 9.43.

In general, if the potential profile consists of n regions, characterized by the potential values V_i and the masses m_i $(i = 1, 2, \ldots n)$, separated by $n - 1$ interfaces at positions z_i $(i = 1, 2, \ldots (n-1))$, then

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = (R_1 R_2 \ldots R_{n-1}) \begin{pmatrix} A_n \\ B_n \end{pmatrix}. \tag{9.52}$$

The matrix elements of R_i are

$$\begin{aligned}
(R_i)_{1,1} &= \left(\tfrac{1}{2} + \tfrac{k_{i+1} m_i}{2k_i m_{i+1}}\right) \exp[i(k_{i+1} - k_i) z_i] \\
(R_i)_{1,2} &= \left(\tfrac{1}{2} - \tfrac{k_{i+1} m_i}{2k_i m_{i+1}}\right) \exp[-i(k_{i+1} + k_i) z_i] \\
(R_i)_{2,1} &= \left(\tfrac{1}{2} - \tfrac{k_{i+1} m_i}{2k_i m_{i+1}}\right) \exp[i(k_{i+1} + k_i) z_i] \\
(R_i)_{2,2} &= \left(\tfrac{1}{2} + \tfrac{k_{i+1} m_i}{2k_i m_{i+1}}\right) \exp[-i(k_{i+1} - k_i) z_i]
\end{aligned} \tag{9.53}$$

where the k_i are defined by Eq. 9.49. If an electron is incident from the left (region #1) only a transmitted wave will appear in the last region $\#n$, and therefore $B_n = 0$. The transmission probability is then given by

$$T = \left(\frac{k_1 m_n}{k_n m_1}\right) \frac{|A_n|^2}{|A_1|^2}. \tag{9.54}$$

This is a general solution to the problem of transmission through multiple barriers. Under certain conditions, a particle incident on the left can appear on the right essentially without attenuation. This situation, called resonant tunneling, corresponds to a constructive interference between the two plane waves coexisting in the region between the barriers (quantum well).

The tunneling probability through a double rectangular barrier is illustrated in Fig. 9.18. In this figure, the mass of the particle is taken to be $0.067 m_0$, the height of the barriers is 0.3eV, their widths are 50Å and their separations are 60Å. As observed in the figures, for certain energies below the barrier height, the particle can tunnel without attenuation. These

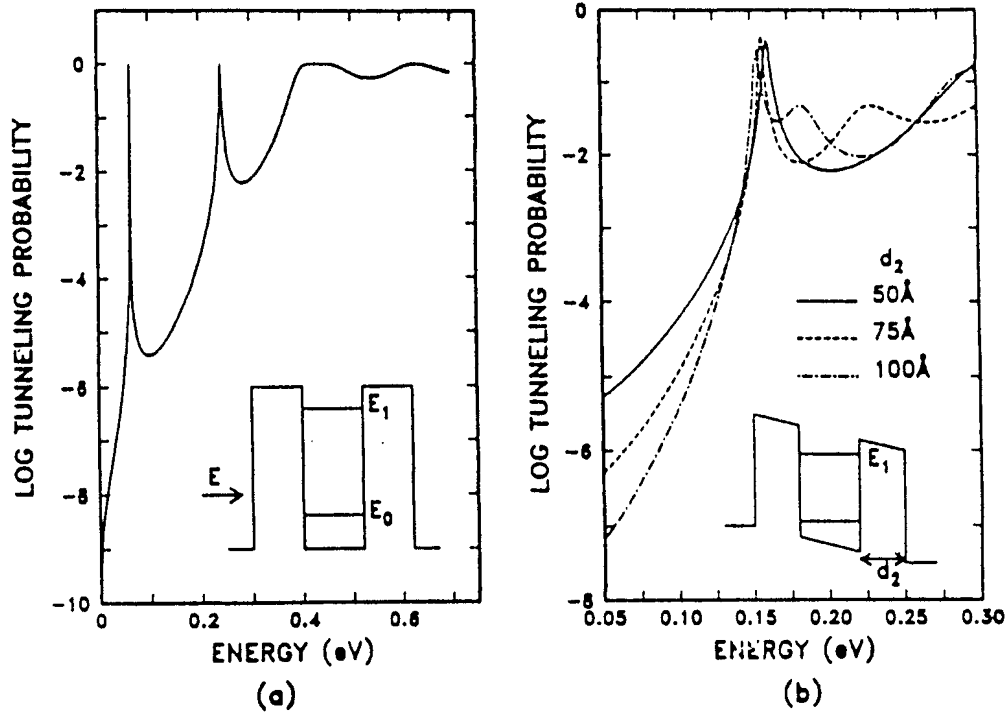


Figure 9.18: (a) Probability of tunneling through a double rectangular barrier as a function of energy. The carrier mass is taken to be $0.1m_0$ in the barrier and $0.067m_0$ outside, and the width of the quantum well is 60Å. (b) Tunneling probability through a double–barrier structure, subject to an electric field of 1×10^5V/cm. The width of the left barrier is 50Å, while that of the right barrier is varied between 50Å and 100Å. The peak at ~ 0.16eV corresponds to resonant tunneling through the first excited state (E_1) of the quantum well. The optimum transmission is obtained when the width of the right barrier is ~ 75Å.

energies correspond precisely to the eigenvalues of the quantum well; this is understandable, since the solutions of Schrödinger's equation for an isolated well are standing waves. When the widths of the two barriers are different (see Fig. 9.18b), the tunneling probability does not reach unity, although the tunneling probability shows maxima for incident energies corresponding to the bound and virtual states.

Chapter 10

Transport in Low Dimensional Systems

References:

- *Solid State Physics*, Volume **44**, Semiconductor Heterostructures and Nanostructures. Edited by H. Ehrenreich and D. Turnbull, Academic Press (1991).
- *Electronic transport in mesoscopic systems*, Supriyo Datta, Cambridge University Press, 1995.

10.1 Introduction

Transport phenomena in low dimensional systems such as in quantum wells (2D), quantum wires (1D), and quantum dots (0D) are dominated by quantum effects not included in the classical treatments based on the Boltzmann equation and discussed in Chapters 4-6. With the availability of experimental techniques to synthesize materials of high chemical purity and of nanometer dimensions, transport in low dimensional systems has become an active current research area. In this chapter we consider some highlights on the subject of transport in low dimensional systems.

10.2 Observation of Quantum Effects in Reduced Dimensions

Quantum effects dominate the transport in quantum wells and other low dimensional systems such as quantum wires and quantum dots when the de Broglie wavelength of the electron

$$\lambda_{\text{dB}} = \frac{\hbar}{(2m^*E)^{1/2}} \tag{10.1}$$

exceeds the dimensions of a quantum structure of characteristic length L_z ($\lambda_{\text{dB}} > L_z$) or likewise for tunneling through a potential barrier of length L_z. To get some order of magnitude estimates of the electron kinetic energies E below which quantum effects become important we look at Fig. 10.1 where a log-log plot of λ_{dB} vs E in Eq. 10.1 is shown for

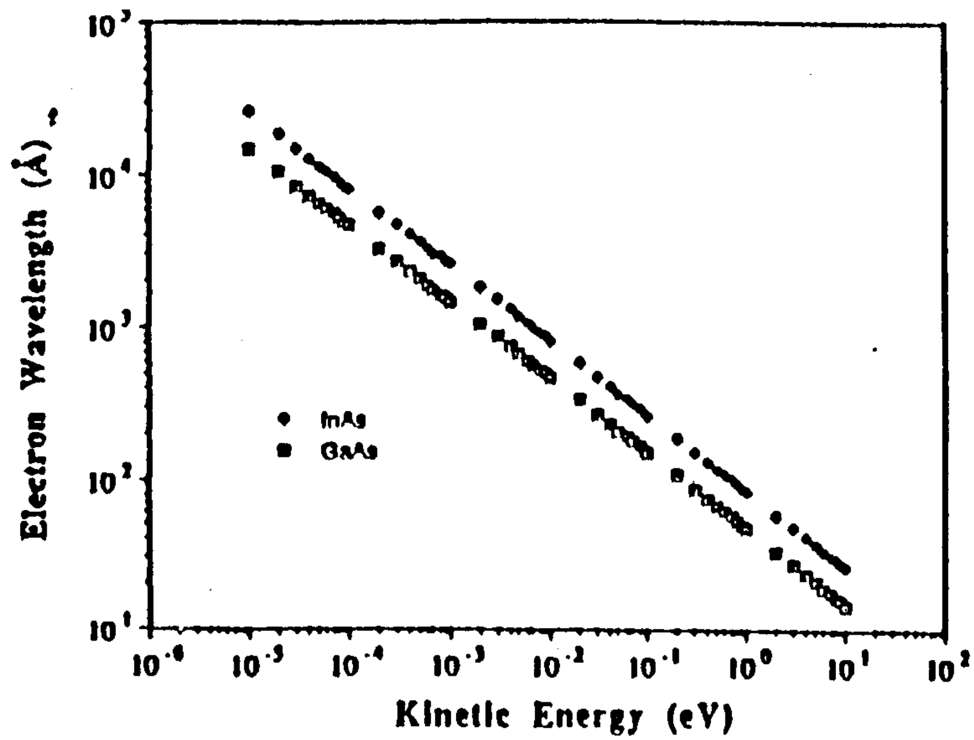


Figure 10.1: Plot of the electron de Broglie wavelength λ_{dB} vs the electron kinetic energy E for GaAs ($\square$) and InAs ($\diamond$).

GaAs and InAs. From the plot we see that an electron energy of $E \sim 0.1$ eV for GaAs corresponds to a de Broglie wavelength of 200 Å. Thus wave properties for electrons can be expected for structures smaller than $\lambda_{\rm dB}$.

To observe quantum effects, the thermal energy must also be less than the energy level separation, $k_BT < \Delta E$, where we note that room temperature corresponds to 25 meV. Since quantum effects depend on the phase coherence of electrons, scattering can also destroy quantum effects. The observation of quantum effects thus requires that the carrier mean free path be much larger than the dimensions of the quantum structures (wells, wires or dots).

The limit where quantum effects become important has been given the name of **mesoscopic physics**. Carrier transport in this limit exhibits both particle and wave characteristics. In this ballistic transport limit, carriers can in some cases transmit charge or energy without scattering.

The small dimensions required for the observation of quantum effects can be achieved by the direct fabrication of semiconductor elements of small dimensions (quantum wells, quantum wires and quantum dots). Another approach is the use of gates on a field effect transistor to define an electron gas of reduced dimensionality. In this context, negatively charged metal gates can be used to control the source to drain current of a 2D electron gas formed near the GaAs/AlGaAs interface as shown in Fig. 10.2. Between the dual gates shown on this figure, a thin conducting wire is formed out of the 2D electron gas. Controlling the gate voltage controls the amount of charge in the depletion region under the gates, as well as the charge in the quantum wire. Thus lower dimensional channels can be made in a 2D electron gas by using metallic gates. In the following sections a number of important applications are made of this concept.

10.3 Density of States in Low Dimensional Systems

We showed in Eq. 8.40 that the density of states for a 2D electron gas is a constant for each 2D subband

$$g_{\rm 2D} = \frac{m^*}{\pi\hbar^2}. \tag{10.2}$$

This is shown in Fig. 10.3(a) where the inset is appropriate to the quantum well formed near a modulation doped GaAs-AlGaAs interface. In the diagram only the lowest bound state is occupied.

Using the same argument, we now derive the density of states for a 1D electron gas

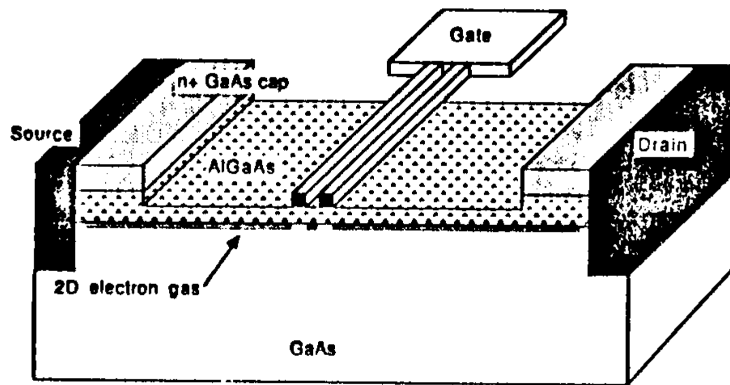
$$N_{\rm 1D} = \frac{2}{2\pi}(k) = \frac{1}{\pi}(k) \tag{10.3}$$

which for a parabolic band $E = E_n + \hbar^2k^2/(2m^*)$ becomes

$$N_{\rm 1D} = \frac{2}{2\pi}(k) = \frac{1}{\pi}\left(\frac{2m^*(E-E_n)}{\hbar^2}\right)^{1/2} \tag{10.4}$$

yielding an expression for the density of states $g_{\rm 1D}(E) = \partial N_{\rm 1D}/\partial E$

$$g_{\rm 1D}(E) = \frac{1}{2\pi}\left(\frac{2m^*}{\hbar^2}\right)^{1/2}(E-E_n)^{-1/2}. \tag{10.5}$$



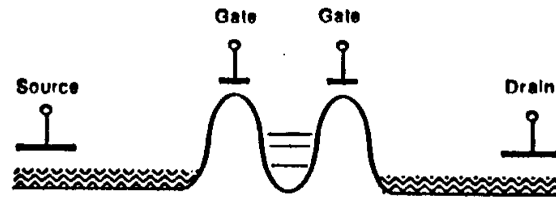


Figure 10.2: (a) Schematic diagram of a lateral resonant tunneling field-effect transistor which has two closely spaced fine finger metal gates; (b) schematic of an energy band diagram for the device. A 1D quantum wire is formed in the 2D electron gas between the gates.

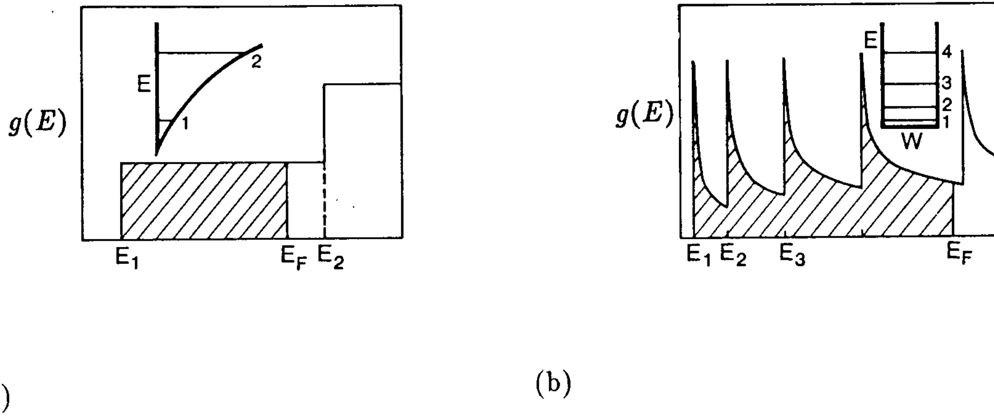


Figure 10.3: Density of states $g(E)$ as a function of energy. (a) Quasi-2D density of states, with only the lowest subband occupied (hatched). Inset: Confinement potential perpendicular to the plane of the 2DEG. The discrete energy levels correspond to the bottoms of the first and second 2D subbands. (b) Quasi-1D density of states, with four 1D subbands occupied. Inset: Square-well lateral confinement potential with discrete energy levels indicating the 1D subband extrema.

The interpretation of this expression is that at each doubly confined bound state level E_n there is a singularity in the density of states, as shown in Fig. 10.3(b) where the first four levels are occupied.

10.3.1 Quantum Dots

This is an example of a zero dimensional system. Since the levels are all discrete any averaging would involve a sum over levels and not an integral over energy. If, however, one chooses to think in terms of a density of states, then the DOS would be a delta function positioned at the energy of the localized state. For more extensive treatment see the review by Marc Kastner (Appendix D).

10.4 The Einstein Relation and the Landauer Formula

In the classical transport theory (Chapter 4) we related the current density $\vec{j}$ to the electric field $\vec{E}$ through the conductivity σ using the Drude formula

$$\sigma = \frac{ne^2\tau}{m^*}. \tag{10.6}$$

This equation is valid when many scattering events occur within the path of an electron through a solid, as shown in Fig. 10.4(a). As the dimensions of device structures become smaller and smaller, other regimes become important, as shown in Figs. 10.4(b) and 10.4(c).

To relate transport properties to device dimensions it is often convenient to rewrite the Drude formula by explicitly substituting for the carrier density n and for the relaxation time τ in Eq. 10.6. Writing $\tau = \ell/v_F$ where ℓ is the mean free path and v_F is the Fermi

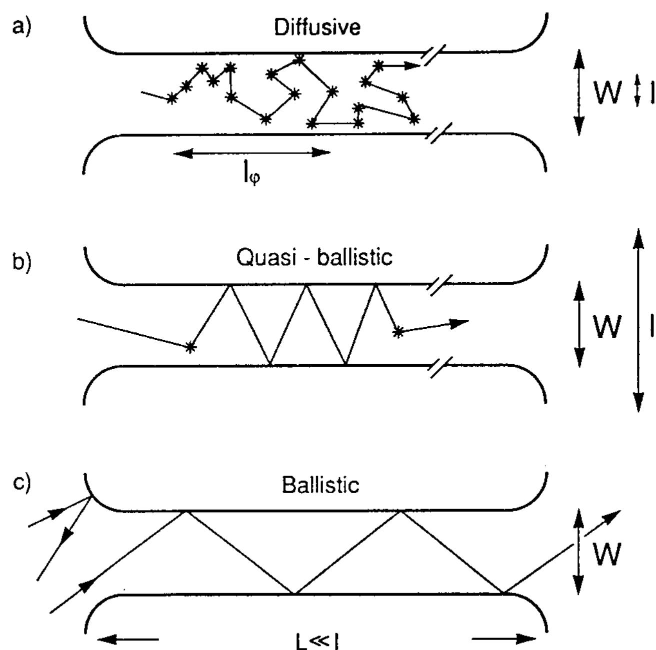


Figure 10.4: Electron trajectories characteristic of the diffusive ($\ell < W, L$), quasi-ballistic ($W < \ell < L$), and ballistic ($W, L < \ell$) transport regimes, for the case of specular boundary scattering. Boundary scattering and internal impurity scattering (asterisks) are of equal importance in the quasi-ballistic regime. A nonzero resistance in the ballistic regime results from backscattering at the connection between the narrow channel and the wide 2DEG regions. Taken from H. Van Houten et al. in "Physics and Technology of Submicron Structures" (H. Heinrich, G. Bauer and F. Kuchar, eds.) Springer, Berlin, 1988.

velocity, and writing $n = k_F^2/(2\pi)$ for the carrier density for a 2D electron gas (2DEG) we obtain

$$\sigma = \frac{k_F^2}{2\pi} e^2 \frac{\ell}{m^* v_F} = \frac{k_F^2 e^2 \ell}{2\pi\hbar k_F} = \frac{e^2}{h} k_F \ell \tag{10.7}$$

where e^2/h is a universal constant and is equal to $\sim (26\ \mathrm{k}\Omega)^{-1}$.

Two general relations that are often used to describe transport in situations where collisions are not important within device dimensions are the Einstein relation and the Landauer formula. We discuss these relations below. The Einstein relation follows from the continuity equation

$$\vec{j} = eD\vec{\nabla}n \tag{10.8}$$

where D is the diffusion coefficient and $\vec{\nabla}n$ is the gradient of the carrier density involved in the charge transport. In equilibrium the gradient in the electrochemical potential $\vec{\nabla}\mu$ is zero and is balanced by the electrical force and the change in Fermi energy

$$\vec{\nabla}\mu = 0 = -e\vec{E} + \vec{\nabla}n \frac{dE_F}{dn} = -e\vec{E} + \vec{\nabla}n/g(E_F) \tag{10.9}$$

where $g(E_F)$ is the density of states at the Fermi energy. Substitution of Eq. 10.9 into Eq. 10.8 yields

$$\vec{j} = eDg(E_F)e\vec{E} = \sigma\vec{E} \tag{10.10}$$

yielding the Einstein relation

$$\sigma = e^2 Dg(E_F) \tag{10.11}$$

which is a general relation valid for 3D systems as well as systems of lower dimensions.

The Landauer formula is an expression for the conductance G which is the proportionality between the current I and the voltage V,

$$I = GV. \tag{10.12}$$

For 2D systems the conductance and the conductivity have the same dimensions, and for a large 2D conductor we can write

$$G = (W/L)\sigma \tag{10.13}$$

where W and L are the width and length of the conducting channel in the current direction, respectively. If W and L are both large compared to the mean free path ℓ, then we are in the diffusive regime (see Fig. 10.4(a)). However when we are in the opposite regime, the ballistic regime, where $\ell > W, L$, then the conductance is written in terms of the Landauer formula which is obtained from Eqs. 10.7 and 10.13. Writing the number of quantum modes N, then $N\pi = k_F W$ or

$$k_F = \frac{N\pi}{W} \tag{10.14}$$

and noting that the quantum mechanical transition probability coupling one channel to another in the ballistic limit $|t_{\alpha,\beta}|^2$ is $\pi\ell/(2LN)$ per mode, we obtain the general Landauer formula

$$G = \frac{2e^2}{h} \sum_{\alpha,\beta}^{N} |t_{\alpha,\beta}|^2. \tag{10.15}$$

We will obtain the Landauer formula below for some explicit examples, which will make the derivation of the normalization factor more convincing.

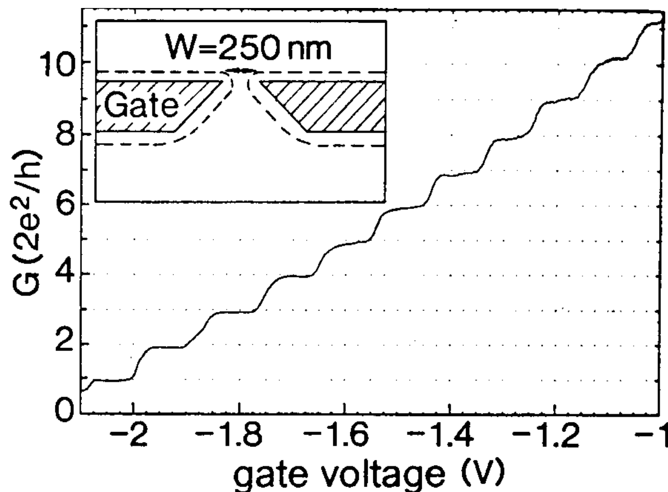


Figure 10.5: Point contact conductance as a function of gate voltage at 0.6 K, obtained from the raw data after subtraction of the background resistance. The conductance shows plateaus at multiples of $e^2/\pi\hbar$. Inset: Point-contact layout [from B.J. van Wees, et al., Phys. Rev. Lett. 60, 848 (1988)].

10.5 One Dimensional Transport and Quantization of the Ballistic Conductance

In the last few years one dimensional ballistic transport has been demonstrated in a two dimensional electron gas (2DEG) of a GaAs-GaAlAs heterojunction by constricting the electron gas to flow in a very narrow channel (see Fig. 10.5). Ballistic transport refers to carrier transport without scattering. As we show below, in the ballistic regime, the conductance of the 2DEG through the constriction shows quantized behavior with the conductance changing in quantized steps of $(e^2/\pi\hbar)$ when the effective width of the constricting channel is varied by controlling the voltages of the gate above the 2DEG. We first give a derivation of the quantization of the conductance.

The current I_x flowing between source and drain (see Fig. 10.2) due to the contribution of one particular 1D electron subband is given by

$$I_x = ne\delta v \tag{10.16}$$

where n is the carrier density (i.e., the number of carriers per unit length of the channel) and δv is the increase in electron velocity due to the application of a voltage V. The carrier density in 1D is

$$n = \frac{2}{2\pi}k_F = \frac{k_F}{\pi} \tag{10.17}$$

and the gain in velocity δv resulting from an applied voltage V is

$$eV = \frac{1}{2}m^*(v_F + \delta v)^2 - \frac{1}{2}m^* v_F^2 = m^* v_F \delta v + \frac{1}{2}m^*(\delta v)^2. \tag{10.18}$$

Retaining only the first order term in Eq. 10.18 yields

$$\delta v = eV/m^* v_F \tag{10.19}$$

so that from Eq. 10.16 we get for the source-drain current (see Fig. 10.5)

$$I_x = \frac{k_F}{\pi}\, e\, \frac{eV}{m^* v_F} = \frac{e^2}{\pi\hbar} V \tag{10.20}$$

since $\hbar k_F = m^* v_F$. This yields a conductance per 1D electron subband G_i of

$$G_i = \frac{e^2}{\pi\hbar} \tag{10.21}$$

or summing over all occupied subbands i we obtain

$$G = \sum_i \frac{e^2}{\pi\hbar} = \frac{ie^2}{\pi\hbar}. \tag{10.22}$$

Two experimental observations of these phenomena were simultaneously published [D.A. Wharam, T.J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J.E.F. Frost, D.G. Hasko, D.C. Peacock, D.A. Ritchie, and G.A.C. Jones, J. Phys. C: Solid State Phys. 21, L209 (1988); and B.J. van Wees, H. van Houten, C.W.J. Beenakker, J.G. Williamson, L.P. Kouvenhoven, D. van der Marel, and C.T. Foxon, Phys. Rev. Lett. 60, 848 (1988)]. The experiments by Van Wees et al. were done using ballistic point contacts on a gate structure placed over a two-dimensional electron gas as shown schematically in the inset of Fig. 10.5. The width W of the gate (in this case 2500 Å) defines the effective width W' of the conducting electron channel, and the applied gate voltage is varied in order to control the effective width W'. Superimposed on the raw data for the resistance vs gate voltage is a collection of periodic steps as shown in Fig. 10.5 after subtracting off the background resistance of 400 Ω.

There are several conditions necessary to observe perfect $2e^2/h$ quantization of the 1D conductance. One requirement is that the electron mean free path l_e be much greater than the length of the channel L. This limits the values of channel lengths to L <5,000 Å even though mean free path values are much larger, l_e=8.5 μm. It is important to note, however, that l_e=8.5 μm is the mean free path for the 2D electron gas. When the channel is formed, the screening effect of the 2D electron gas is no longer present and the effective mean free path becomes much shorter. A second condition is that there are adiabatic transitions at the inputs and outputs of the channel. This minimizes reflections at these two points, an important condition for the validity of the Landauer formula to be discussed later in this section. A third condition requires the Fermi wavelength $\lambda_F = 2\pi/k_F$ (or $k_F L > 2\pi$) to satisfy the relation $\lambda_F < L$ by introducing a sufficient carrier density (3.6×10^{11}cm^{-2}) into the channel. Finally, as discussed earlier, it is necessary that the thermal energy $k_B T \ll E_j - E_{j-1}$ where $E_j - E_{j-1}$ is the subband separation between the j and $j-1$ one dimensional energy levels. Therefore, the quantum conductance measurements are done at low temperatures (T <1 K).

The point contacts in Fig. 10.5 were made on high-mobility molecular-beam-epitaxy-grown GaAs-AlGaAs heterostructures using electron beam lithography. The electron density of the material is 3.6×10^{11}/cm^2 and the mobility is 8.5×10^5cm^2/V s (at 0.6 K). These values were obtained directly from measurements of the devices themselves. For the transport measurements, a standard Hall bar geometry was defined by wet etching. At a gate voltage of $V_g = -0.6$ V the electron gas underneath the gate is depleted, so that conduction takes place through the point contact only. At this voltage, the point contacts

have their maximum effective width $W'_{\max}$, which is about equal to the opening W between the gates. By a further decrease (more negative) of the gate voltage, the width of the point contacts can be reduced, until they are fully pinched off at $V_g = -2.2$ V.

The results agree well with the appearance of conductance steps that are integral multiples of $e^2/\pi h$, indicating that the conductance depends directly on the number of 1D subbands that are occupied with electrons. To check the validity of the proposed explanation for these steps in the conductance (see Eq. 10.22), the effective width W' for the gate was estimated from the voltage $V_g = -0.6$ V to be 3600 Å, which is close to the geometric value for W. In Fig. 10.5 we see that the average conductance varies linearly with V_g which in turn indicates a linear relation between the effective point contact width W' and V_g. From the 16 observed steps and a maximum effective point contact width $W'_{\max} = 3600$ Å, an estimate of 220 Å is obtained for the increase in width per step, corresponding to $\lambda_F/2$. Theoretical work done by Rolf Landauer nearly 20 years ago shows that transport through the channel can be described by summing up the conductances for all the possible transmission modes, each with a well defined transmission coefficient t_{nm}. The conductance of the 1D channel can then be described by the Landauer formula

$$G = \frac{e^2}{\pi\hbar} \sum_{n,m=1}^{N_c} |t_{nm}|^2 \tag{10.23}$$

where N_c is the number of occupied subbands. If the conditions for perfect quantization described earlier are satisfied, then the transmission coefficient reduces to $|t_{nm}|^2 = \delta_{nm}$. This corresponds to purely ballistic transport with no scattering or mode mixing in the channel (i.e., no back reflections).

A more explicit derivation of the Landauer formula for the special case of a 1D system can be done as follows. The current flowing in a 1D channel can be written as

$$I_j = \int_{E_i}^{E_f} e g_j(E) v_z(E) T_j(E) dE \tag{10.24}$$

where the electron velocity is given by

$$m^* v_z = \hbar k_z \tag{10.25}$$

and

$$E = E_j + \frac{\hbar^2 k_z^2}{2m^*} \tag{10.26}$$

while the 1D density of states is from Eq. 10.5 given by

$$g_j(E) = \frac{(2m^*)^{1/2}}{h(E - E_g)^{1/2}} \tag{10.27}$$

and $T_j(E)$ is the probability that an electron injected into subband j with energy E will get across the 1D wire ballistically. Substitution of Eqs. 10.25, 10.26 and 10.27 into Eq. 10.24 then yields

$$I_j = \frac{2e}{h} \int_{E_i}^{E_j} T_j(E) dE = \frac{2e^2}{h} T_j \Delta V \tag{10.28}$$

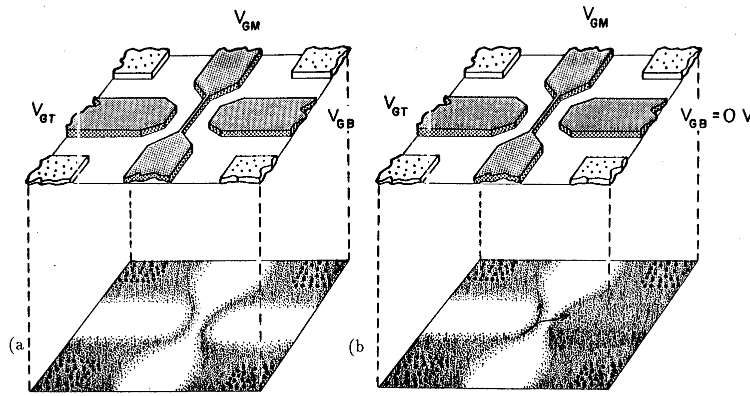


Figure 10.6: (a) Schematic illustration of the split-gate dual electron waveguide device. The top plane shows the patterned gates at the surface of the MODFET structure. The bottom plane shows the implementation of two closely spaced electron waveguides when the gates (indicated by V_{GT}, V_{GB} and V_{GM}) are properly biased. Shading represents the electron concentration. Also shown are the four ohmic contacts which allow access to the inputs and outputs of each waveguide. (b) Schematic of the "leaky" electron waveguide implementation. The bottom gate is grounded (V_{GB}=0) so that only one waveguide is in an "on" state. V_{GM} is fixed such that only a small tunneling current crosses it. The current flowing through the waveguide as well as the tunneling current (depicted by arrows) are monitored simultaneously.

where we have noted that the potential energy difference is the difference between initial and final energies $e\Delta V = E_f - E_i$. Summing over all occupied states j we then obtain the Landauer formula

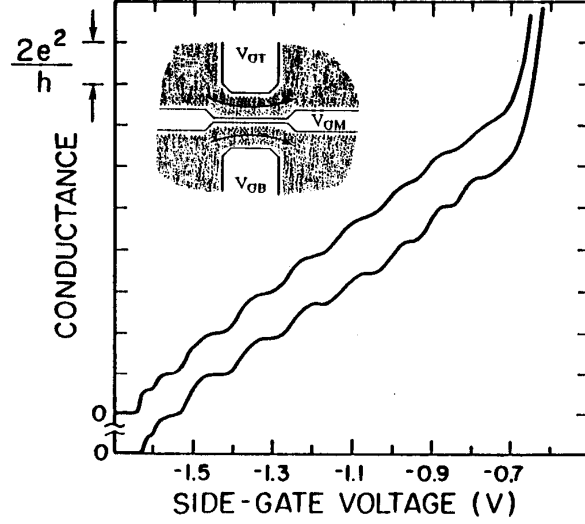
$$G = \frac{2e^2}{h} \sum_j T_j. \tag{10.29}$$

10.6 Ballistic Transport in 1D Electron Waveguides

Another interesting quantum-effect structure proposed and implemented by C. Eugster and J. del Alamo is a split-gate dual electron waveguide device shown in Fig. 10.6 [C.C. Eugster, J.A. del Alamo, M.J. Rooks and M.R. Melloch, Applied Physics Letters **60**, 642 (1992)]. By applying the appropriate negative biases on patterned gates at the surface, two electron waveguides can be formed at the heterointerface of a MODFET structure. As shown in Fig. 10.6, the two electron waveguides are closely spaced over a certain length and their separation is controlled by the middle gate bias (V_{GM}). The conductance of each waveguide, shown in Fig. 10.7, is measured simultaneously and independently as a function of the side gate biases (V_{GT} and V_{GB}) and each show the quantized $2e^2/h$ conductance steps. Such a device can be used to study 1D coupled electron waveguide interactions. An electron directional coupler based on such a structure has also been proposed [J. del Alamo and C. Eugster, Appl. Phys. Lett. **56**, 78 (1990)]. Since each gate can be independently accessed, various other regimes can be studied in addition to the coupled waveguide regime.

One interesting regime is that of a "leaky" electron waveguide [C.C. Eugster and J.A. del

Figure 10.7: Conductance of each waveguide in Fig. 10.6(a) as a function of side gate bias of an $L = 0.5\ \mu$m, $W = 0.3\ \mu$m split-gate dual electron waveguide device. The inset shows the biasing conditions (i.e., in the depletion regime) and the direction of current flow for the measurements.



Alamo, Physical Review Letters **67**, 3586 (1991)]. For such a scheme, one of the side gates is grounded so that the 2D electron gas underneath it is unaffected, as shown in Fig. 10.6(b). The middle gate is biased such that only a small tunneling current can flow from one waveguide to the other in the 2D electron gas. The other outer gate bias V_{GT} is used to sweep the subbands in the waveguide through the Fermi level. In such a scheme, there is only one waveguide which has a thin side wall barrier established by the middle gate bias. The current flowing through the waveguide as well as the current leaking out of the thin middle barrier are independently monitored. Figure 10.8 shows the $I - V_{GS}$ characteristics for the leaky electron waveguide implementation. As discussed earlier, conductance steps of order $2e^2/h$ are observed for the current flowing through the waveguide. However what is unique to the leaky electron experiment is that the tunneling current leaking from the thin side wall is monitored. As seen in Fig. 10.8, very strong oscillations in the tunneling current are observed as the Fermi level is modulated in the waveguide. We show below that the tunneling current is directly **tracing** out the 1D density of states of the waveguide.

An expression for the tunneling current I_{S2} flowing through the sidewall of the waveguide can be obtained by the following integral,

$$I_{S2} = e\sum_j \int_{-\infty}^{\infty} v_{\perp j}(E-E_j) g_{1D,j}(E-E_j) T_j(E-E_b) \Big[f(E-E_F-e\Delta V, T) - f(E-E_F, T) \Big] dE \tag{10.30}$$

where we have accounted for the contribution to the current from each occupied subband j. Here E_j is the energy at the bottom of the jth subband and E_b is the height of the tunneling energy barrier. The normal velocity against the tunneling barrier is $v_{\perp j}(E) = \hbar k_{\perp j}/m^*$. The transmission coefficient, $T_j(E - E_b)$, to first order, is the same for the different 1D electron subbands since the barrier height relative to the Fermi level E_F is fixed (see Fig. 10.9). The Fermi function, f, gives the distribution of electrons as a function of temperature and applied bias ΔV between the input of the waveguide and the 2DEG on

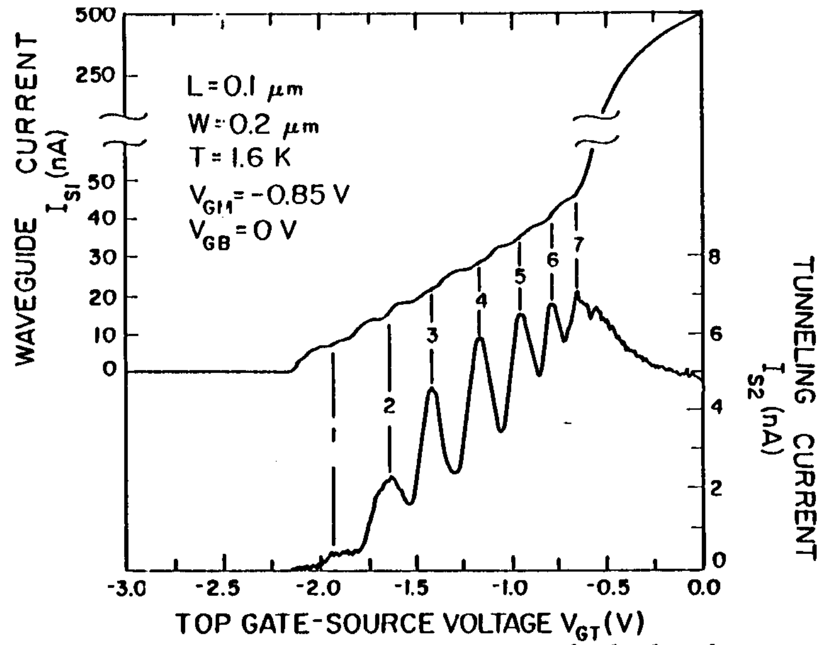


Figure 10.8: The waveguide current vs gate-source voltage characteristics of a leaky electron waveguide implemented with the proper biases for the device. For this case one of the side-gates in Fig. 10.6(b) is grounded so that only one leaky electron waveguide is on. I_{S1} is the current flowing through the waveguide and I_{S2} is the current tunneling through the thin middle side barrier. The bias voltage V_{DS} between the contacts is 100 μV.

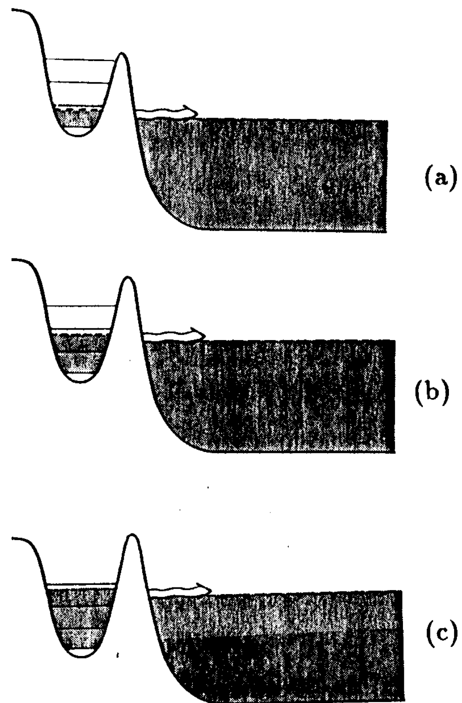


Figure 10.9: Cross-section of a leaky electron waveguide. Shaded regions represent electrons. The dashed line is the Fermi level and the solid lines depict the energy levels in the waveguide. The three figures are from top to bottom (a), (b), (c) at increasingly negative gate-source voltage V_{GT}.

the other side of the tunneling barrier (see Fig. 10.9).

For low enough temperatures and small ΔV, we can approximate Eq. 10.30 by

$$I_{S2} \cong \frac{e^2\hbar}{m^*}\sum_j k_{\perp j}T_j(E_F - E_b)g_{1D,j}(E_F - E_j)\Delta V \tag{10.31}$$

where $k_{\perp j}$ is a constant for each subband and has a value determined by the confining potential. As seen by Eq. 10.31, the tunneling current I_{S2} is proportional to $g_{1D,j}(E_F - E_j)$, the 1D density of states (see Fig. 10.10a). Increasing V_{GT} to less negative values sweeps the subbands through the Fermi level as shown in Fig. 10.9 and since $k_{\perp j}$ and $T_j(E_F - E_b)$ are constant for a given subband (see Fig. 10.10d), the 1D density of states $g_{1D,j}$ summed over each subband, can be extracted from measurement of I_{S2} (see Fig. 10.10e).

Figure 10.9 shows a schematic of the cross-section of the waveguide for three different values of the gate source voltage V_{GT}. The middle gate bias V_{GM} is fixed and is the same for all three cases. The parabolic potential in Fig. 10.9 is a result of the fringing fields and is characteristic of the quantum well for split-gate defined channels. By making V_{GT} more negative, the sidewall potential of the waveguide is raised with respect to the Fermi level. As seen in Fig. 10.9, this results in having fewer energy levels below the Fermi level (i.e., fewer occupied subbands). In Fig. 10.9a, which corresponds to a very negative V_{GT}, only the first subband is occupied. At less negative values of V_{GT}, the second subband becomes occupied (Fig. 10.9b) and then the third (Fig. 10.9c) and so on. There is no tunneling current until the first level has some carrier occupation. The tunneling current increases as a new subband crosses below the Fermi-level. The difference between E_F in the quantum well (on the left) and in the metallic contact (on the right) is controlled by the bias voltage between the input and output contacts of the waveguide V_{DS}. In fact, a finite voltage V_{DS} and a finite temperature gives rise to lifetime effects and a broadening of the oscillations in I_{S2} (see Fig. 10.10e).

A summary of the behavior of a leaky electron waveguide is shown in Fig. 10.10. Included in this figure are (a) the 1D density of states g_{1D}. The measurement of I_{S1} as shown in Fig. 10.8 is modeled in terms of $\sum k_{\parallel}g_{1D,j}$ and the results for $I_{S1}(V_{GT})$ which relate to the steps in the conductance can be used to extract g_{1D} as indicated in Figs. 10.10b and 10.10c. In contrast, Figs. 10.10d and 10.10e indicate the multiplication of $k_{\perp,j}$ and $g_{1D,j}$ to obtain $\sum k_{\perp,j}g_{1D,j}$ summed over occupied levels which is measured by the tunneling current I_{S2} in Fig. 10.8. The 1D density of states g_{1D} can then be extracted from either I_{S1} or I_{S2} as indicated in Fig. 10.10.

10.7 Single Electron Charging Devices

By making even narrower channels it has been possible to observe single electron charging in a nanometer field-effect transistor, shown schematically in Fig. 10.11. In studies, a metal barrier is placed in the middle of the channel and the width of the metal barrier and the gap between the two constricted gates are of very small dimensions.

The first experimental observation of single electron charging was by Meirav et al. [U. Meirav, M.A. Kastner and S.J. Wind, Phys. Rev. Lett. **65**, 771 (1990); see also M.A. Kastner, Physics Today, page 24, January 1993], working with a double potential barrier GaAs device, as shown in Fig. 10.12. The two dimensional gas forms near the GaAs-GaAlAs

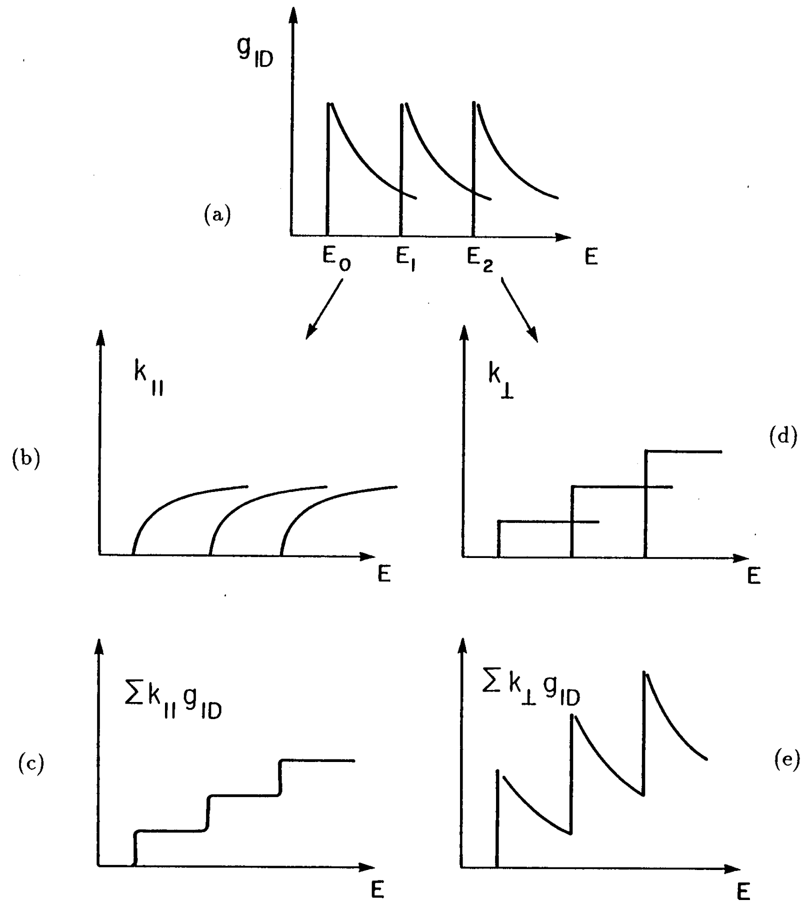
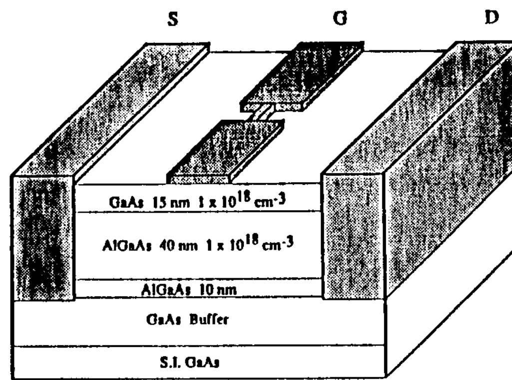
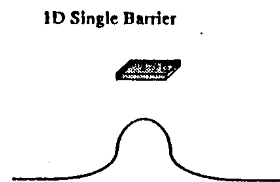
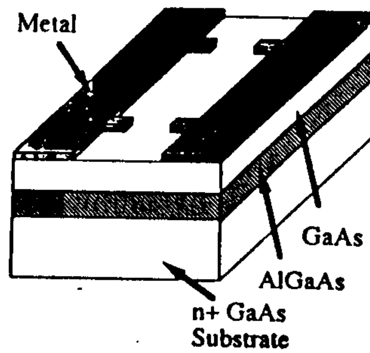


Figure 10.10: Summary of leaky electron waveguide phenomena. Top picture (a) represents the 1D density of states for the waveguide. The two left pictures (b) and (c) are for the current flowing through the waveguide ($k_{\parallel}$ is the wavevector along waveguide). Quantized conductance steps result from sweeping subbands through the Fermi level as I_{S1} in Fig. 10.8. The two right hand pictures (d) and (e) are for the tunneling current ($k_{\perp}$ is transverse wavevector). Oscillations in the tunneling current I_{S2} in Fig. 10.8 arise from sweeping each subband through the Fermi level.

Figure 10.11: A split gate nanometer field-effect transistor, shown schematically. In the narrow channel a 1D electron gas forms when the gate is biased negatively. The potential of the 1D barrier is shown.







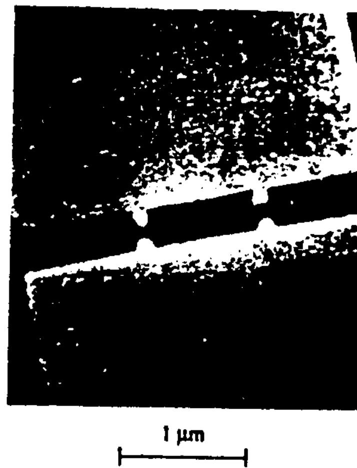
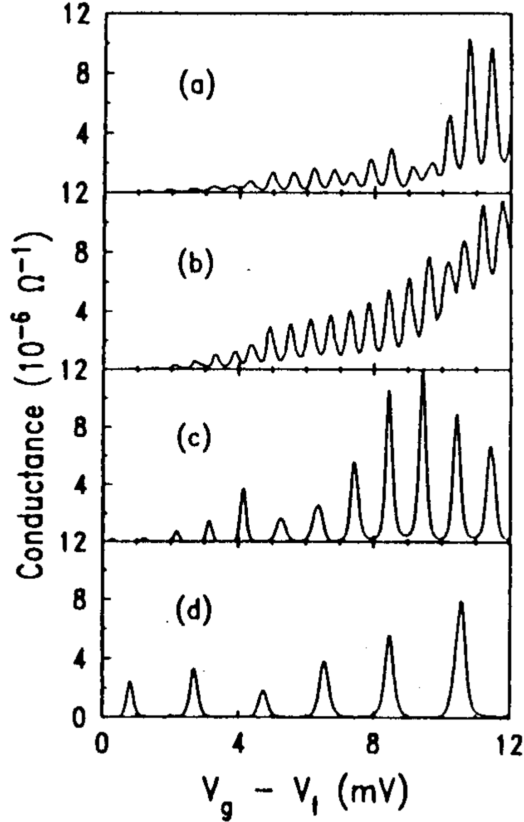


Figure 10.12: Schematic drawing of the device structure along with a scanning electron micrograph of one of the double potential barrier samples. An electron gas forms at the top GaAs-AlGaAs interface, with an electron density controlled by the gate voltage V_g. The patterned metal electrodes on top define a narrow channel with two potential barriers.

Figure 10.13: Periodic oscillations of the conductance vs gate voltage V_g, measured at $T \approx 50$mK on a sample-dependent threshold V_t. Traces (a) and (b) are for two samples with the same electrode geometry and hence show the same period. Traces (c) and (d) show data for progressively shorter distances between the two constrictions, with a corresponding increase in period. Each oscillation corresponds to the addition of a single electron between the barriers.



interface. Each of the constrictions in Fig. 10.12 is 1000 Å long and the length of the channel between constrictions is 1μm. When a negative bias voltage ($V_b \sim -0.5$V) is applied to the gate, the electron motion through the gates is constrained. At a threshold gate voltage of V_t, the current in the channel goes to zero. This is referred to as Coulomb blockade. If the gate voltage is now increased (positively) above V_t, a series of periodic oscillations are observed, as shown in Fig. 10.13. The correlation between the period of the conductance oscillations and the electron density indicates that a single electron is flowing through the double gated structure per oscillation. The oscillations in Fig 10.13 show the same periodicity when prepared under the same conditions [as in traces (a) and (b)]. As the length L_0 between the constrictions is reduced below 1μm, the oscillation period gets longer. To verify that they had seen single electron charging behavior, Meirav et al. fit the experimental lineshape for a single oscillation to the functional form for the conductance

$$G(\mu) \sim \frac{\partial F}{\partial E} \sim \cosh^{-2}\left[\frac{E-\mu}{2k_B T}\right]. \tag{10.32}$$

where μ is the chemical potential, $F(E-\mu, T)$ is the Fermi function and E is the single electron energy in the 2DEG.