Semiconductor Devices: Lectures 5&6

12. Semiconductor Heterojunctions

The semiconductor junctions considered so far have been *homojunctions*, e.g. *Si-Si*; we now consider the properties of *heterojunctions*.

(a) Energy Band Alignment

Electron affinity χ: Energy required to remove an electron from the conduction band and take it to the vacuum

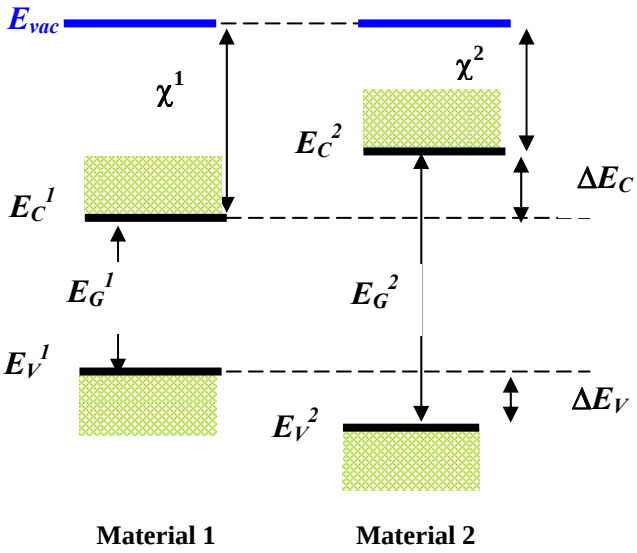
Work function ϕ: Energy required to remove an electron from the Fermi energy and take it to the vacuum.

The values of χ and ϕ are dominated by the bulk cohesion of the atoms, but are affected by the surface phenomena:

- reconstruction, where the surface atoms rearrange in the surface plane
- relaxation, where the atoms move slightly away from their bulk positions
- surface electronic states
- impurities at the surface
- dipole layer due to charge leakage out of the surface

The conduction (valence) band offset ΔE_C (ΔE_V) is the change in the conduction (valence) band at the heterojunction. A simple (Anderson's) rule states that the vacuum levels should line up as shown below, so that:

$$\Delta E_C = \chi^1 - \chi^2$$
$$\Delta E_V = E_g^1 - E_g^2 - \Delta E_C$$



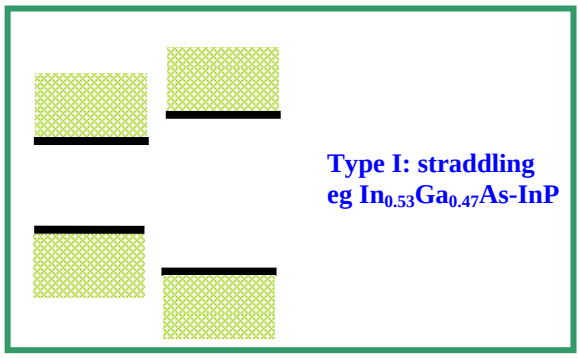
This rule is often inaccurate as χ contains the surface contributions listed above, which are different from those at an interface. The band offsets can be obtained from microscopic bandstructure calculations, but for most practical purposes they are treated as empirical parameters.

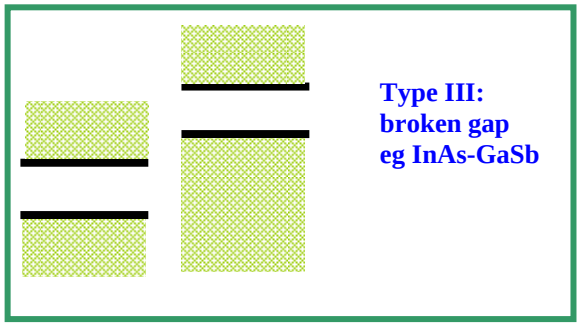
One of the most commonly used heterojunctions is that formed by GaAs and $Ga_{1-x}Al_xAs$. The bandgap of the latter is given approximately by Vegard's law:

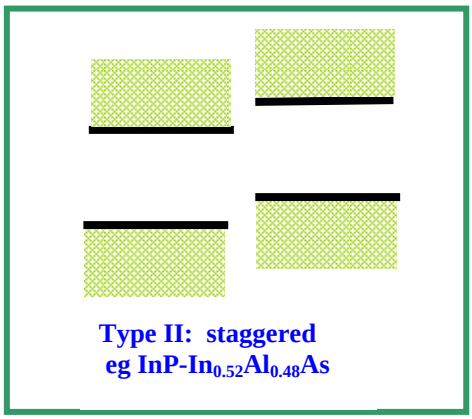
$$E_G(\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}) = xE_G(\mathrm{AlAs}) + (1-x)E_G(\mathrm{GaAs})$$

which simply assumes a linear interpolation between the values of the two constituent materials. One of the major advantages of this material combination is that the lattice parameter of the alloy is very close to that of GaAs for a wide range of x, so that there is very little strain at the interface (see (c) below). The band offset ratio $\Delta E_C / \Delta E_V$ is approximately 60/40.

Various types of band alignment are observed:

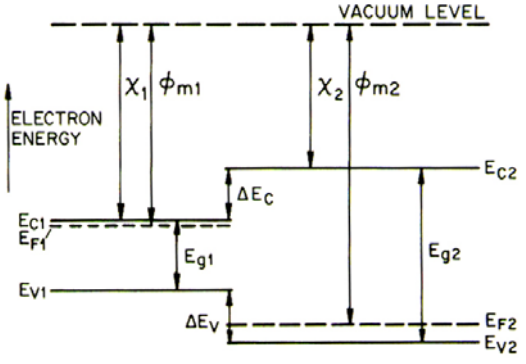




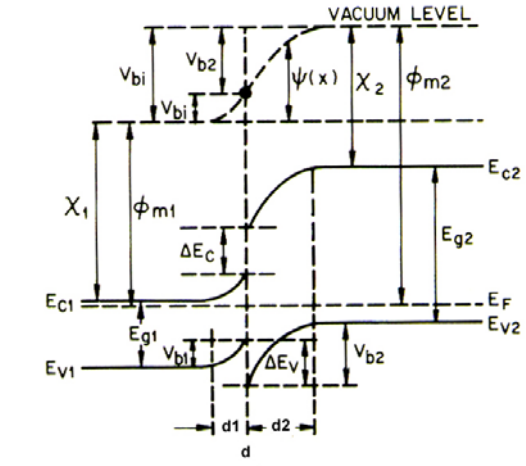


These examples of band alignment show how the potential barrier across a *pn* junction may be increased (Type I), electronic states can be made "spatially indirect" (Type II), or semi-metallic behaviour can be produced due to overlapping conduction and valence bands (Type III).

(b) P-N Heterojunction Diode



(a)



(b)

(a) Energy level structure for the isolated semiconductors, the narrower gap material being n-type, and the larger gap material p-type.

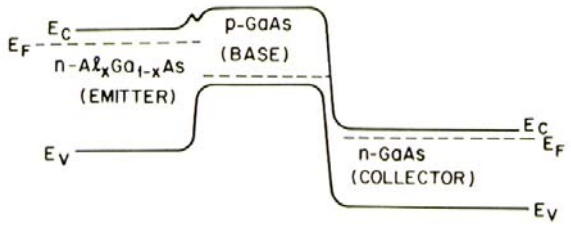
(b) When the junction is formed the Fermi level is the same on both sides, and the vacuum level is parallel to the band edges and is *continuous*. The built-in (or contact) potential is the sum of the potentials in both materials: $V_{bi}=V_{b1}+V_{b2}$. The depletion layer widths are calculated as before, except that the boundary condition is no longer continuity of electric field E at the interface but the displacement, i.e. $\varepsilon_1 E_1=\varepsilon_2 E_2$:

$$d_1=\left[\frac{2N_{A2}\varepsilon_0\varepsilon_1\varepsilon_2(V_{bi}-V)}{qN_{D1}(\varepsilon_1 N_{D1}+\varepsilon_2 N_{A2})}\right]^{1/2}$$

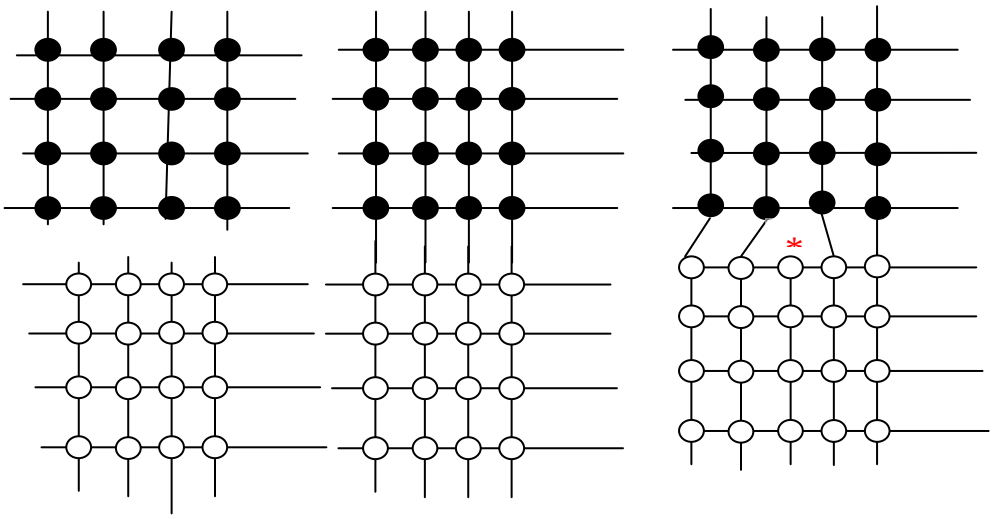
$$d_2=\left[\frac{2N_{D1}\varepsilon_0\varepsilon_1\varepsilon_2(V_{bi}-V)}{qN_{A2}(\varepsilon_1 N_{D1}+\varepsilon_2 N_{A2})}\right]^{1/2}$$

This reverts to the usual expression for a *p-n* homojunction when $\varepsilon_1=\varepsilon_2$.

In the heterojunction bipolar transistor shown below the wider gap material forming the emitter is n-type, while the base-collector junction is homotype. There are several important potential advantages of this device:
(i) higher emitter efficiency (larger β) due to the base-emitter hole current being reduced due to the enhanced potential barrier, (ii) reduced base resistance since the base can now be heavily doped, (iii) frequency response is better because of (i) and (ii).



(c) Strain due to lattice mismatch: e.g. $In_xGa_{1-x}As/GaAs$



(a) Separate layers in equilibrium; (b) Thin layer of $In_xGa_{1-x}As$ on GaAs; (c) Thicker layer of $In_xGa_{1-x}As$

(a) In equilibrium, $In_xGa_{1-x}As$ has a larger lattice constant than GaAs. The GaAs substrate is so thick that it can not be significantly distorted. In (b), the $In_xGa_{1-x}As$ layer is thin, so it strains to conform to the lattice constant of the GaAs in the plane of the heterojunction. In (c), the strain has relaxed due to a misfit dislocation at the heterointerface, shown by an asterisk. This releases the elastic energy, which is proportional to the thickness.

Strain can be highly beneficial in devices. It can re-inforce the changes in bandstructure and so controls ΔE_C and ΔE_V which need to be as large as possible. Strained-layer lasers have a low threshold current and high efficiency for electrical → optical power conversion. Strain has a pronounced effect on valence band states: the hole mass can be made significantly lighter so that the hole mobility is enhanced.

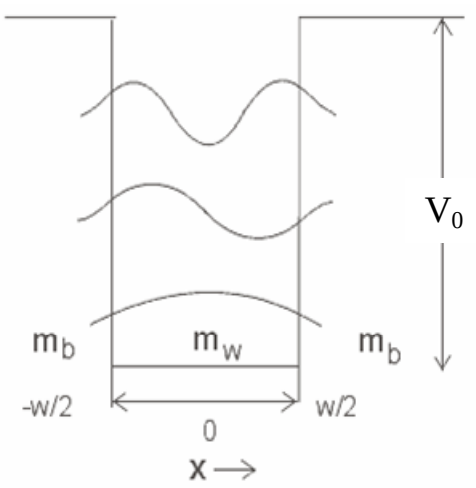
13. Quantum Well Structures

As the active layer thickness in a *double* heterostructure becomes close to the electron de-Broglie wavelength (about 10nm for GaAs) quantum effects become apparent. Quatum wells are important in many device applications, e.g. semiconductor lasers - where the wavelength of the emitted light can be controlled through adjustment of the energy levels within the well by selecting the well width:

Solving the Schrödinger equation for a quantum well with finite potential obtains values of the energy levels within the well:

$$\left(-\frac{\hbar^2}{2m_w}\frac{\partial^2}{\partial^2 x}-V_0\right)\psi_n(x)=\varepsilon_n\psi_n(x) \quad \text{for } |x|<w/2$$

$$\left(-\frac{\hbar^2}{2m_b}\frac{\partial^2}{\partial^2 x}\right)\psi_n(x)=\varepsilon_n\psi_n(x) \quad \text{for } |x|>w/2$$



The continuity conditions at the interfaces are that ψ_n and $\frac{1}{m}\frac{\partial\psi}{\partial x}$ should be continuous (the latter condition includes the effective mass in order that particle current is continuous).
Then:

$$\begin{aligned}
\psi_n(x) &= A\cos kx && \text{for } |x| < w/2 \\
&= B\exp\left[-K\left(x-\frac{w}{2}\right)\right] && \text{for } x > w/2 \\
&= B\exp\left[+K\left(x+\frac{w}{2}\right)\right] && \text{for } x < -w/2
\end{aligned}$$

} even solutions

$$\begin{aligned}
\psi_n(x) &= A\sin kx && \text{for } |x| < w/2 \\
&= B\exp\left[-K\left(x-\frac{w}{2}\right)\right] && \text{for } x > w/2 \\
&= B\exp\left[+K\left(x+\frac{w}{2}\right)\right] && \text{for } x < -w/2
\end{aligned}$$

} odd solutions

where $\varepsilon_n=\frac{\hbar^2k^2}{2m_w}-V_0$, $\varepsilon_n=-\frac{\hbar^2K^2}{2m_b}$, **for** $-V_0<\varepsilon<0$

Even solutions:

$$A\cos\left(\frac{kw}{2}\right)=B$$

$$\frac{k}{m_w}A\sin\left(\frac{kw}{2}\right)=\frac{KB}{m_b}$$

$$\therefore \frac{k}{m_w}\tan\left(\frac{kw}{2}\right)=\frac{K}{m_b}$$

Odd solutions:

$$A\sin\left(\frac{kw}{2}\right)=B$$

$$\frac{k}{m_w}A\cos\left(\frac{kw}{2}\right)=-\frac{KB}{m_b}$$

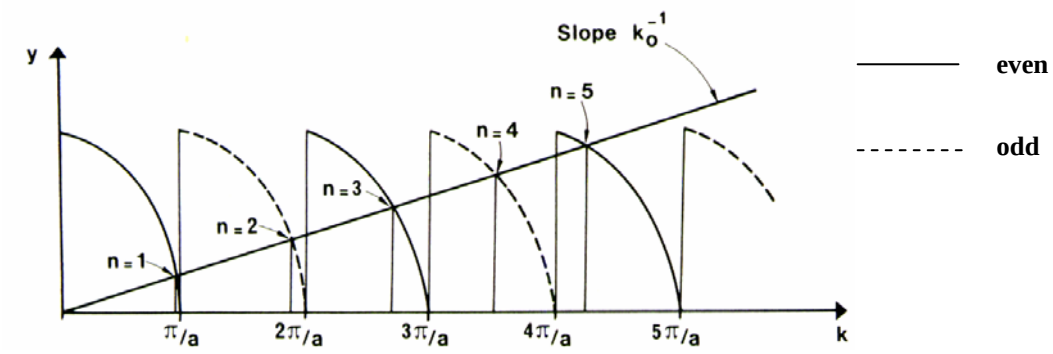
$$\therefore \frac{k}{m_w}\cot an\left(\frac{kw}{2}\right)=-\frac{K}{m_b}$$

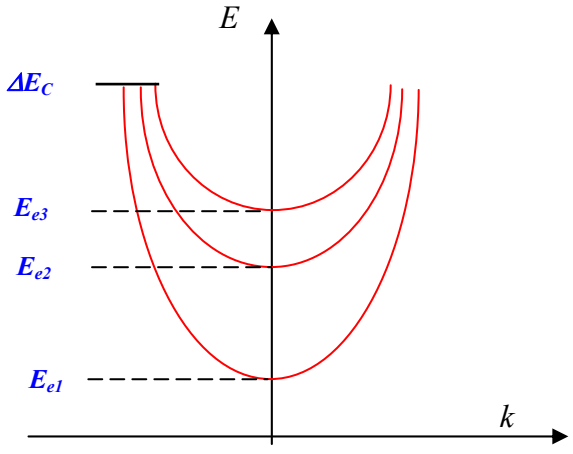
The equations can be solved numerically or graphically. However, in the limit of $m_w=m_b$ the solution simplifies to:

$$cos\frac{kw}{2}=\frac{k}{k_0} \qquad \textbf{for} \quad tan\left(\frac{kw}{2}\right)>0$$

$$sin\frac{kw}{2}=\frac{k}{k_0} \qquad \textbf{for} \quad tan\left(\frac{kw}{2}\right)<0$$

where $k_0^2=\frac{2mV_0}{\hbar^2}$. The number of bound states is: $1+Int\left[\left(\frac{2mV_0w^2}{\pi^2\hbar^2}\right)^{\frac{1}{2}}\right]$ so that there is always one bound state.





There is a very familiar limiting case, viz. *infinitely high potential barriers*, $V_0 \rightarrow \infty$, $k_0 \rightarrow \infty$; then there are an infinite number of bound states of the form:

$$\psi \sim sin(kx)$$

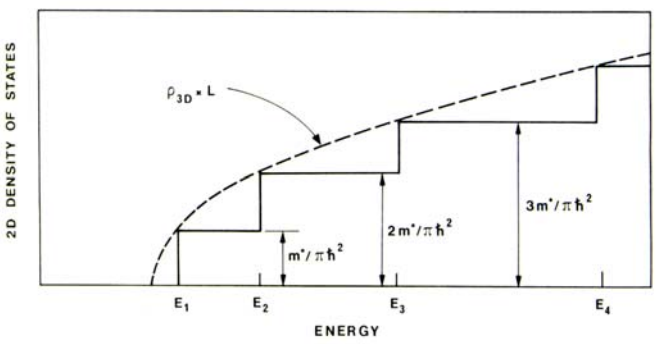
with $k=n\pi/w$. The total energy for the n^{th} level in an infinitely deep quantum well is:

$$E_n=\frac{\pi^2\hbar^2n^2}{2m^*w^2}+\frac{\hbar^2k_x^{\,2}}{2m^*}+\frac{\hbar^2k_y^{\,2}}{2m^*}$$

The two-dimensional density of states per unit area per unit energy, $N(E)$, is given by:

$$N(k)dk=\frac{2}{(2\pi)^2}(2\pi kdk)$$

$$\Rightarrow \qquad N(E)dE=\frac{m^*}{\pi\hbar^2}dE$$



The density of states per unit area n_{2d} :

$$n_{2d} = \sum_j n_{2dj}$$

where, for the jth subband,

$$n_{2dj} = \frac{m^*}{\pi\hbar^2}\int_{\varepsilon_j}^{\infty} f(E)dE$$

$$= \frac{m^*}{\pi\hbar^2}\int_{\varepsilon_j}^{\infty} \frac{dE}{exp\{(E-E_F)/k_BT\}+1}$$

$$= \frac{m^* k_B T}{\pi\hbar^2} ln\left[exp\{(E_F-\varepsilon_j)/k_BT\}+1\right]$$

c.f. 3D nondegenerate system, i.e. high T and low doping levels,

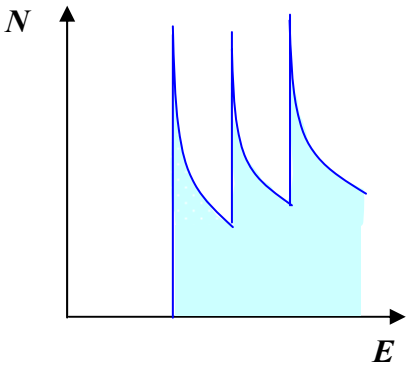
$$n_{3d} = \frac{m^{*3/2}\sqrt{2}}{\pi^2\hbar^3}\int_0^{\infty} f(E)E^{1/2}dE$$

$$\approx \frac{m^{*3/2}\sqrt{2}}{\pi^2\hbar^3}\int_0^{\infty} E^{1/2}\,exp\{-(E-E_F)/k_BT\}dE$$

$$= N_C\,exp\{E_F/k_BT\} \quad ; N_C = 2\left(\frac{m^* k_B T}{2\pi\hbar^2}\right)^{3/2}$$

14. Quantum Wires and Quantum Dots

So far we have considered spatial confinement in one direction. However, it is also possible to produce semiconductor structures that are confined in two and three directions. Consider a quantum wire with square cross-section a confined within infinite potential barriers; the energy states are:



$$E_n = \frac{\pi^2\hbar^2 n_1^2}{2m^* a^2} + \frac{\pi^2\hbar^2 n_2^2}{2m^* a^2} + \frac{\hbar^2 k_y^2}{2m^*}$$

Given that $N(k)\mathrm{d}k = \frac{2}{(2\pi)}(2)\mathrm{d}k$ the density of states per unit area per unit energy is:

$$N(E)dE = \frac{1}{\pi}\left(\frac{2m^*}{\hbar^2}\right)^{1/2} E^{-1/2}dE$$

For (n_1,n_2) subbands occupied

$$N(E) = \sum N(E)_{n_1,n_2} \; ; N(E)_{n_1,n_2} = \frac{\sqrt{2m^*}}{\pi\hbar}\frac{\Theta\left(E-\varepsilon_{n_1,n_2}\right)}{\sqrt{E-\varepsilon_{n_1,n_2}}}$$

Similarly the energy levels of a cubic quantum box or dot are:

$$E_n = \frac{\pi^2\hbar^2 n_1^2}{2m^* a^2} + \frac{\pi^2\hbar^2 n_2^2}{2m^* a^2} + \frac{\pi^2\hbar^2 n_3^2}{2m^* a^2}$$

Such a structure is analogous to an artificial atom with discrete energy levels, i.e. the electrons are totally confined, and the density of states is:

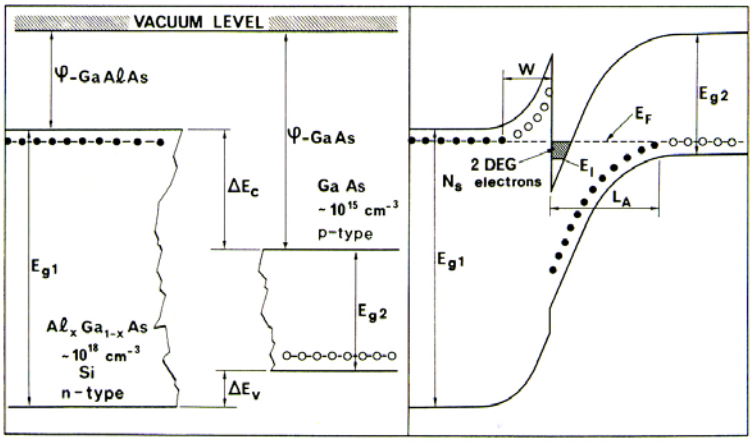
$$N(E) = \sum N(E)_{n_1 n_2 n_3} \; ; \; N(E)_{n_1 n_2 n_3} = \delta\left(E - \varepsilon_{n_1 n_2 n_3}\right)$$

Semiconductor quantum wires and dots give enhanced radiative electron-hole recombination: this can be seen as arising from the strong wavefunction overlap. Lasers with threshold currents <1mA have been produced, more than an order of magnitude improvement over quantum well lasers.

Mesoscopic effects occur in low dimensional structures. These are anomalies, e.g. large fluctuations, in transport properties observed at low temperatures, due to the discrete charge on an electron. These effects are associated with the intrinsically small size of the structure, which only contains a few carriers, e.g: *Coulomb Blockade* - electrostatic energy required to add one electron to a quantum dot is $e^2/(2C)$ where C is dot capacitance, this can easily exceed k_BT at low temperatures, and indeed even at room temperature with a nanometre-sized dot (see section **16** on SETs).

15. Modulation Doping: MODFET

Doping semiconductors to introduce carriers has the major disadvantage that the charged donors or acceptors scatter the carriers through their Coulomb interaction. This scattering reduces the mobility and (lifetime) broadens energy levels. The reduced coherence is serious in quantum devices which require the interference between electron waves.

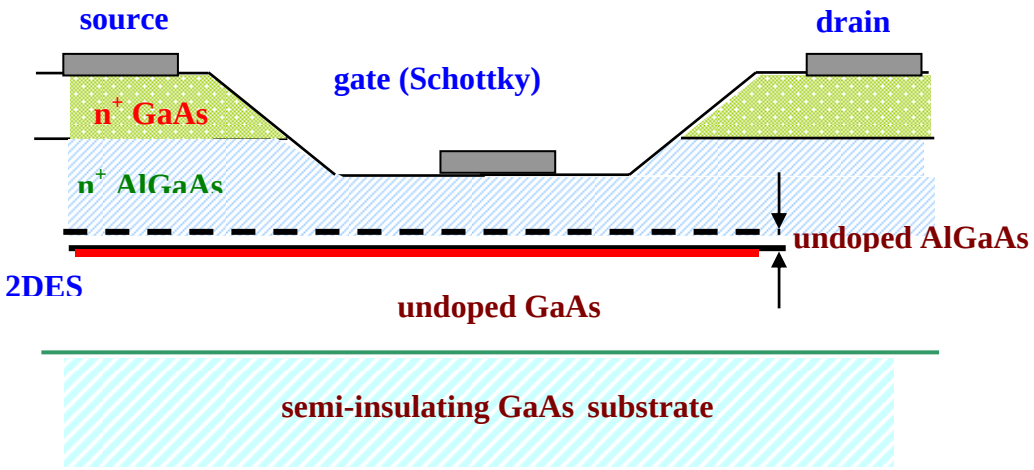


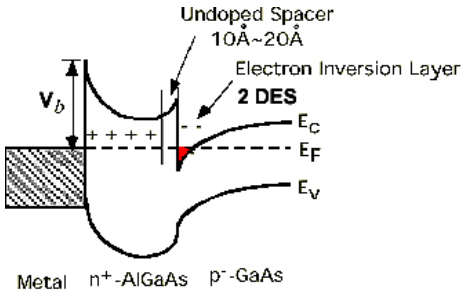
The solution to this problem is *modulation doping* where impurities are placed in the barrier region, but the carriers migrate to the quantum confined region. In this example electrons transfer from donors in the AlGaAs barrier to the GaAs layer until equilibrium is reached. The carriers are now in a high purity region, remote from the ionised donors. This effect can be further enhanced by introducing an undoped spacer layer in the barrier between the doped region and the interface. In the example shown electrons are confined in a narrow region close to the interface, i.e. they are spatially confined, and so form a two-dimensional electron system (2DES). The confining potential is approximately a triangular quantum well, and assuming that the AlGaAs potential barrier is infinite, the eigenvalues are:

$$E_n = \left(\frac{\hbar^2}{2m^*}\right)^{\frac{1}{3}} \left(\frac{3}{2}\pi q F\right)^{\frac{2}{3}} \left(n + \frac{3}{4}\right)^{\frac{2}{3}} \qquad n = 0,1,2....$$

where F is the electric field on the GaAs side of the interface.

The very high electron mobilities that can be obtained by modulation doping are used in FETs and BJTs to enhance their speed. A schematic modulation doped FET (MODFET) structure is shown below. Electron transport occurs through the 2DES, and is controlled by the gate voltage. The combination of short channel length (~100nm) and high mobility (~100$m^2V^{-1}s^{-1}$) allows switching speeds of >1THz to be obtained. The optimum material at present is AlInAs/InGaAs.



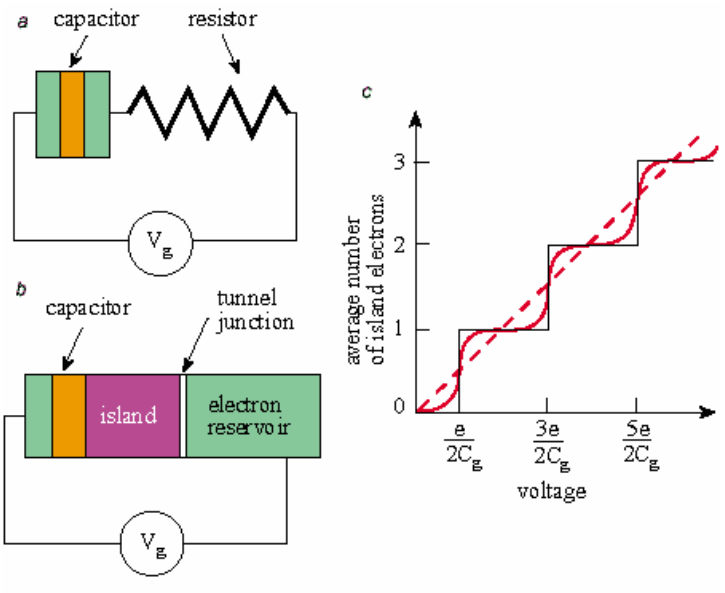


Structures similar to this one are used to study low temperature magneto-transport properties of electrons in two-dimensional systems. With very thick spacer layers ~300nm, mobilities ~2x10^4$m^2V^{-1}s^{-1}$ can be achieved. The quantum Hall effect and fractional quantum Hall effect are observed when a strong magnetic field is applied normal to the 2 DES.

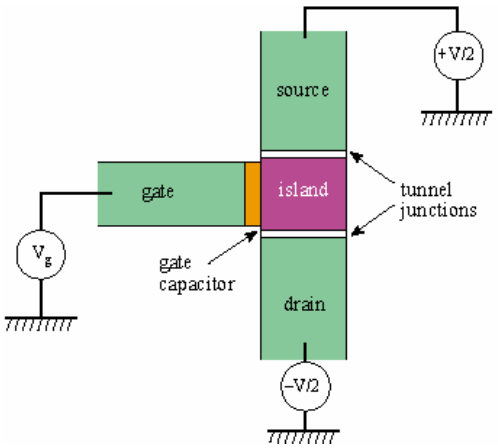
16. Single Electron Transistor (SET)

As the size of transistors continues to shrink, a question naturally arises: will the quantum nature of electrons become important in determining how the devices are built? In other words, what will happen when a transistor is reduced to the size of a few atoms or a single molecule? The single-electron transistor is such a device that exploits quantum tunnelling to control and measure the movement of single electrons. Charge does not flow continuously in this device but in a quantized way. Consequently, SETs have potential for very high density data storage.

When a capacitor is charged through a resistor, the charge on the capacitor is proportional to the applied voltage and shows no sign of quantization. When a tunnel junction replaces the resistor, a conducting island is formed between the junction and the capacitor plate. In this case the average charge on the island increases in steps as the voltage is increased. The steps are sharper for more resistive barriers and at lower temperatures:



The single-electron tunnelling (SET) transistor consists of a gate electrode that electrostatically influences electrons travelling between the source and drain electrodes. However, the electrons in the SET transistor need to cross two tunnel junctions that form an isolated conducting electrode or "island" (or quantum dot). The source, drain and island are usually obtained by "cutting" regions in the 2DES formed at the interface between GaAs and AlGaAs. The conducting regions are defined by metallic electrodes patterned on the top semiconducting layer. Negative voltages applied to these electrodes deplete the electron gas just beneath them, and the depleted regions can be made sufficiently narrow to allow tunnelling between the source, island and drain. Moreover, the electrode that shapes the island can be used as the gate electrode. Electrons passing through the island charge and discharge it, and the relative energies of systems containing 0 or 1 extra electrons depends on the gate voltage. At a low source-drain voltage, a current will only flow through the SET transistor if these two charge configurations have the same energy.



The current flowing in a single-electron transistor increases with the bias voltage between the source and drain, and varies periodically with the gate voltage. For low bias voltages, current flows when the charge on the gate capacitor is a half-integer multiple of *e*, but is suppressed for integer multiples of *e*. Each time an electron is added to the gate, an electron tunnels into the island, which sets the field in the gate capacitor back to its initial value. Peaks in the conductance are observed for half-integer multiples of *e*, and minima are seen at integer multiples of *e*. For bias voltages larger than e/C, conduction occurs independently of the gate voltage.