Basic Semiconductor Physics, continued

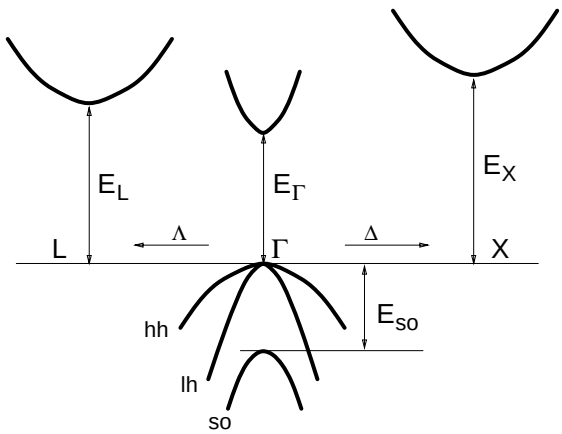


Figure. Important extremal points in the band structure of cubic semiconductors. The schematic picture (drawn not to scale) is appropriate for a direct-gap III-V semiconductor. For GaAs at room temperature the indicated energies are:
$E_\Gamma = 1.42\,eV$, $E_{\rm L} = 1.71\,eV$, $E_{\rm X} = 1.90\,eV$, $E_{\rm so} = 0.34\,eV$.
In silicon the lowest conduction band point is in the Δ direction, 85% of the way to X point. The indicated energies for silicon at 300 K are:
$E_\Gamma = 4.08\,eV$, $E_{\rm L} = 1.87\,eV$, $E_\Delta = 1.13\,eV$, $E_{\rm so} = 0.04\,eV$.
In Ge the lowest conduction band point is at L but the Γ point is not far away:
$E_\Gamma = 0.89\,eV$, $E_{\rm L} = 0.76\,eV$, $E_\Delta = 0.96\,eV$, $E_{\rm so} = 0.29\,eV$.

Bands: degenerate / nondegererate
isotropic / anisotropic
parabolic / nonparabolic

In the vicinity of a nondegenerate extremal point it is convenient to describe the dispersion relation $E_n(\mathbf{k})$ using *the effective mass tensor* $\mathbf{M}_n^{-1}$:

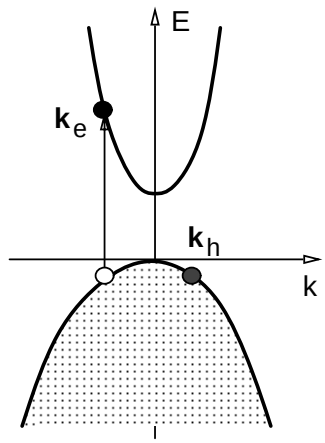
$$E_n(\mathbf{k}_0 + \mathbf{k}) - E_n(\mathbf{k}_0) = (\hbar^2/2)\,\mathbf{k}\cdot\mathbf{M}_n^{-1}\cdot\mathbf{k} \equiv \frac{\hbar^2}{2}\sum_{i,j=1}^{3}\left[\mathbf{M}_n^{-1}\right]_{ij} k_i\,k_j \quad ,$$

where the components of $\mathbf{M}_n^{-1}$ can be written down in terms of the free electron mass and parameters of the lattice potential (the periodic $V(x,y,z)$ characteristic of the point $\mathbf{k}_0$.

Concept of holes

Rules of the game

hole is an alternate description of a band with one missing electron



1. k (hole) = - k (missing electron)

2. $E_h(\mathbf{k}_h) = -E_e(\mathbf{k}_e)$

   energy counted downwards

3, $\mathbf{v}_h = \mathbf{v}_e$

   velocity of hole equals that of missing electron (take gradient of the band with respect to the relevant k vector)

4, $m_h = -m_e$

5. Equation of motion for a hole in electric and magnetic fields is that for a positive particle of charge + e

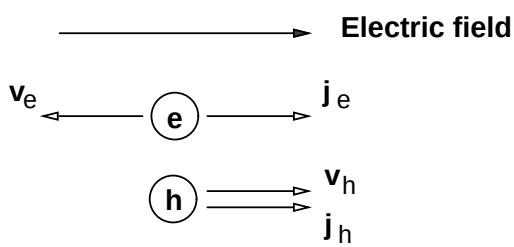


Figure. Concept of holes, illustration

In the vicinity of a nondegenerate band extremum, the surfaces of equal energy are ellipsoids, as evident from Eq. (1). The symmetric tensor $\mathbf{M}_n^{-1}$ has, most generally, six independent components. The coordinate axes can always be chosen so as to diagonalize this tensor, i.e. along the ellipsoid's principal directions:

$$\mathbf{M}_n^{-1} = \begin{bmatrix} 1/m_1 & 0 & 0 \\ 0 & 1/m_2 & 0 \\ 0 & 0 & 1/m_3 \end{bmatrix} .$$

In general, the energy ellipsoid is determined by six independent parameters: three diagonal values of $\mathbf{M}_n^{-1}$ and three directions of the principal axes. However, the number of parameters can be often reduced by symmetry considerations. The ellipsoid symmetry is determined uniquely by the symmetry of the extremal point $\mathbf{k}_0$. For an extremum located on a crystal symmetry axis, one of the principal directions of the energy ellipsoid coincides with the symmetry axis. If the latter is an axis of 3-fold, 4-fold, or 6-fold symmetry, then the ellipsoid is an ellipsoid of revolution ($m_1 = m_2 \equiv m_t$, $m_3 \equiv m_l$). If more than one such axis intersects at $\mathbf{k}_0$, then the ellipsoid anisotropy disappears and energy surfaces become spherical ($m_1 = m_2 = m_3 \equiv m$). Such is the situation in the conduction band at the Γ point of cubic semiconductors:

$$E(k) = \frac{\hbar^2 k^2}{2m} .$$

In silicon the conduction band minima are on the 4-fold rotation axes and the low-energy isoenergetic surfaces are ellipsoids of revolution, their long axes being along <100> directions. There are six symmetry related minima. Similar local minima exist in the conduction band of germanium, but the true conduction band minima in Ge are located at L points.

There are only four symmetry-related ellipsoids of constant energy in the vicinity of the conduction band edge of Ge. It is convenient to picture these ellipsoids as eight half-ellipsoids joined together on opposite faces by translations through suitable reciprocal lattice vectors. In each ellipsoid, the band curvature is least in the direction along the rotation axis and highest in the transverse directions. This means that the longitudinal mass is heavier than the transverse mass. The anisotropy is particularly high in Ge, $m_l/m_t = 20$, but it is also considerable in Si, where $m_l/m_t \approx 5$.

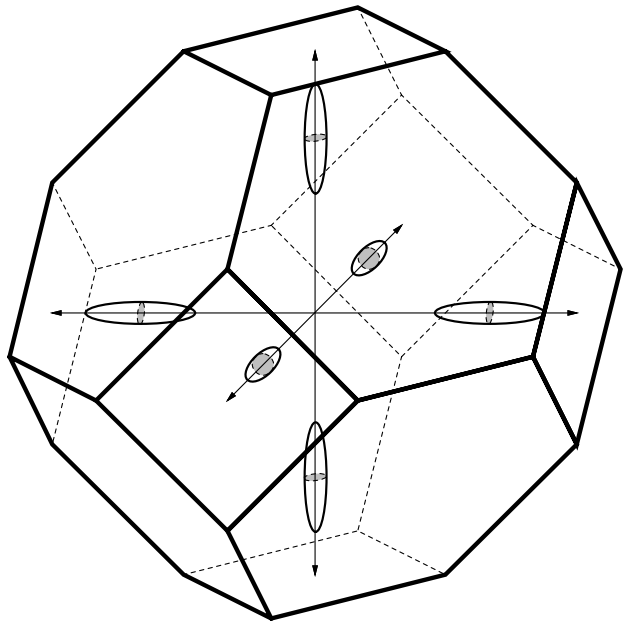


Figure . Surface of constant energy in the vicinity of the conduction band edge in silicon represents six ellipsoids of revolution, extended along <100> directions. The band curvature is least in the longitudinal direction (heavy mass) and highest in the transverse direction (light mass) The effective mass ratio $m_l/m_t \approx 5$.

If the band edge in Si were at the zone boundary rather than at a general point in the Δ direction ($\approx 85\%$ toward X), then there would be only three ellipsoids. Location away from the zone boundary of the conduction band edge in Si and the local Δ minimum in Ge is related to the inversion symmetry of the diamond structure.

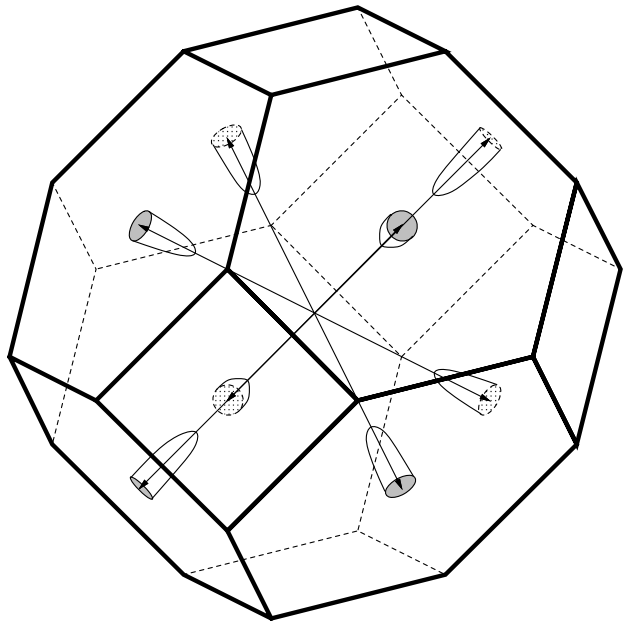


Figure: Surface of constant energy in the vicinity of the conduction band edge in germanium represents four ellipsoids of revolution, extended along <111> directions. The effective mass ratio is very large, $m_l/m_t = 20$, so that each ellipsoid really looks like a sausage.

In order to exhibit a full ellipsoid we would have to chose a primitive cell for which some L point would be internal. In the Brillouin zone picture, each ellipsoid is cut in two by the boundary and the equivalent half is shifted to the opposite face by a reciprocal lattice vector.

Band Electrons in External Electric Field.

For an electron initially belonging to a nondegenerate isotropic band n and subject to a static or slowly varying electric field (due to an impurity or an

$$\mathrm{i}\hbar \frac{\partial \psi}{\partial t} = H\psi \ , \qquad H = \frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}) - e\,U(\mathbf{r},t) \ ,$$

reduces to a simple "effective-mass" equation

$$\mathrm{i}\hbar \frac{\partial \Phi}{\partial t} = \tilde{H}\Phi \ , \qquad \tilde{H} = \frac{\mathbf{p}^2}{2m} - e\,U(\mathbf{r},t) \ ,$$

for a slowly varying (both in space and time) function Φ, called the *envelope function*. Equation (5) describes a quasi-particle of mass m moving in the external field only. The only remnant of the periodic potential is the effective mass m. The complete wave function in effective mass approximation is of the form

$$\psi(\mathbf{r},t) = \Phi(\mathbf{r},t)\,\psi_{n\mathbf{k}_0}(\mathbf{r},t) = \Phi(\mathbf{r},t)\,\mathrm{e}^{\mathrm{i}\mathbf{k}_0\cdot\mathbf{r}}\,u_{n\mathbf{k}_0}(\mathbf{r})\,\mathrm{e}^{-\mathrm{i}E_0 t/\hbar} \ ,$$

where $E_0 = E(\mathbf{k}_0)$. This function represents the band-edge Bloch function modulated by a large-scale envelope function Φ.

In the absence of an external field ($U \to 0$) the full wave function (6) must assume the Bloch form and the envelope function, therefore, becomes a plane wave, $\Phi \to \exp[\mathrm{i}(\mathbf{k}\cdot\mathbf{r} - \varepsilon t/\hbar)]$, where $\varepsilon(\mathbf{k}) = E(\mathbf{k}) - E_0$. If the external fields are sufficiently smooth and slow varying,

---

Smooth means they vary little over distances of order the lattice constant a and *slow* means the frequency of their variation with time is much lower than $\Delta E/\hbar$, where ΔE is the vertical (i.e. for the same $\mathbf{k}_0$) separation between nearest bands.

---

then the plane-wave expansion of a true wave function will get its main contribution from small wavevectors $\mathbf{k}$, in the vicinity of $\mathbf{k}_0$. It is this difference in the length scale of characteristic variations which permits to factor ψ into a product of the Bloch and the envelope functions, where the former is determined only by the periodic potential and the latter describes the entire effect of the external potential.

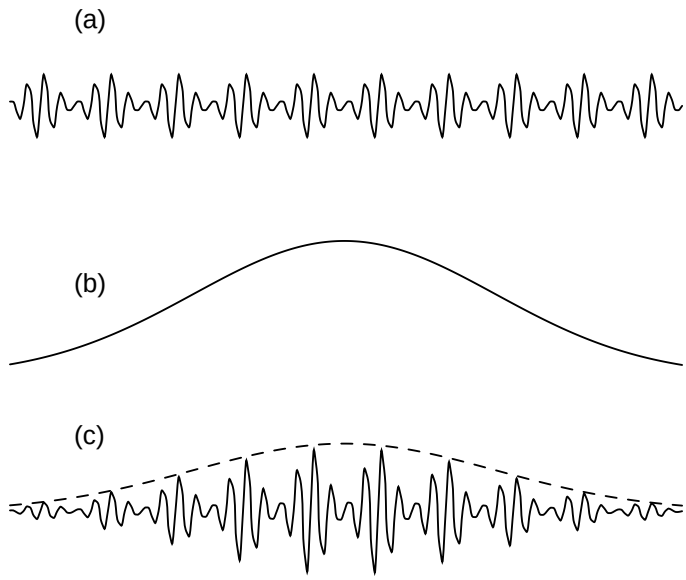


Figure: Schematic illustration of the Bloch and envelope wave functions.

(*a*) Bloch function $u(\mathbf{r})$ at $\mathbf{k}_0 = 0$ varies rapidly within the crystal unit cell.

(*b*) Envelope function $F(\mathbf{r})$ varies little on the unit cell scale.

(*c*) Complete wave function $\psi(\mathbf{r})$ of a localized electron in the effective mass approximation.

In the presence of a uniform electric field the effective mass method needs a clarification. The exact time-dependent Schrödinger equation,

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi \ , \qquad H = H_0 + e\,\mathbf{F}\cdot\mathbf{r} \ ,$$

where $H_0 = p^2/m_0 + V(\mathbf{r})$ is the crystal hamiltonian, can be written in the crystal momentum representation. To do this, we expand $\psi(\mathbf{r},t)$ in a complete set of Bloch states:

$$\psi(\mathbf{r},t) = \sum_{n\,\mathbf{k}} B_n(\mathbf{k},t)\,\psi_{n\mathbf{k}}(\mathbf{r}) \ ,$$

where

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\,u_{n\mathbf{k}}(\mathbf{r})$$

are eigenfunctions of the unperturbed hamiltonian H_0. Coefficients $B_n(\mathbf{k},t)$ form a crystal-momentum representation of the time-dependent wave function.

In the presence of a uniform electric field we face two conceptual difficulties, associated with the phenomena illustrated in the Figure. First "difficulty" is associated with interband transitions. States near the band edge acquire a finite lifetime due the possibility of *tunneling* across the bandgap. The idea that electrons can tunnel through the forbidden gap of an insulator in a strong electric field was first suggested by Zener (1934) and the phenomenon is called the Zener tunneling. Theoretical description of Zener tunneling requires a model that includes states of both the conduction and valence bands. It results in a slight broadening of the sharp band edges; however, moderate electric fields, $F \lesssim 10^6$ V/cm, do not change the band structure too much.

In the absence of interband transitions one finds

$$B_n(\mathbf{k},t) = B_n(\mathbf{k} - eFt/\hbar, 0)$$

$$\frac{d\,\hbar\mathbf{k}}{dt} = e\,\mathbf{F}$$

$\mathbf{k}$ varies linearly in time and hence the electron motion is *periodic* with a frequency

$$f = eF/\hbar b \ .$$

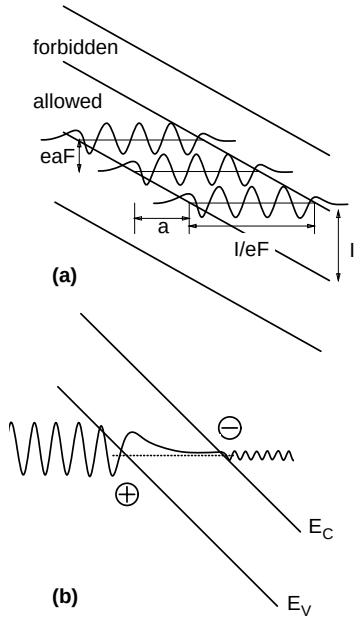


Figure. Schematic illustration of band electrons in a uniform electric field.

(*a*) If interband transitions are suppressed, the allowed energy bands break into a ("Stark") ladder of equally spaced ("Wannier") levels. To observe this phenomenon it necessary that the energy separation between adjacent levels be larger than their broadening $\hbar/\tau$ due to scattering or, equivalently, that the scattering time τ be longer than the time necessary for an electron to traverse the allowed band. This requirement can be met in superlattices (Chap. III) but hardly in real solids.

(*b*) Electrons can tunnel across the bandgap, leaving behind a hole in the valence band. This phenomenon is called the Zener tunneling. The envelope of a wave function exponentially decays in the forbidden band while rapidly oscillating in the allowed bands. In experimentally attainable fields of order 10^6 V/cm only the edges of the allowed and forbidden bands are slightly broadened by the field, by an amount of order $m^{-1/3}(eF\hbar)^{2/3}$.

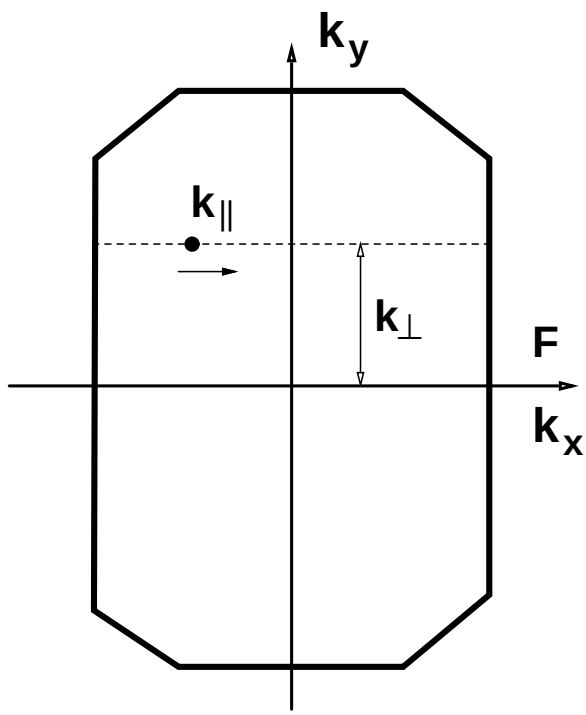


Figure. Band electrons in an electric field **F** move so that the wavevector $\mathbf{k}_{||}$ is varying along a straight line parallel to the field **F**. On crossing the Brillouin zone boundary, the wave-vector $\mathbf{k}_{||}(t)$ re-emerges at the equivalent point on the opposite side of the zone and continues its variation in the same direction. If **F** is parallel to a reciprocal lattice vector, electrons perform an oscillatory motion along the segment $\Delta\mathbf{k}_{||}$ (dashed line), subject to an occasional Zener tunneling into another band. In the absence of tunneling, the electron velocity component in the direction of **F** changes sign at every crossing. If **F** is not aligned with a reciprocal lattice vector, the motion is aperiodic.

The distance electron travels during one cycle is of the order of I/eF (where I is the width of the allowed band in which the electron is moving). This indicates that the wave function, is built of functions with **k** vectors running through the entire band. If interband tunneling is forbidden, the energy levels will be quantized, as can be expected when the classical motion is periodic.

The band breaks into discrete energy levels, called the Wannier levels, equally spaced by the energy intervals eaF. Each of the levels corresponds to a wave function, centered on a particular site and extending in the direction of FF over approximately I/eaF lattice sites. These assertions are consequences of the fact that a translation by one lattice period changes the hamiltonian by the constant amount eaF.

Nevertheless, the effective mass method can be applied to such problems as electric conductivity in semiconductors without a contradiction. The periodic motion due to the Bragg reflection of electrons has never been observed in the bulk crystal, despite many attempts at detecting such an effect. The reason for this failure and the saving grace of the effective mass method is the limited and short time that electrons can travel without *collisions*. As an accelerated electron moves higher and higher in the band, very rapid collision processes, like optical phonon emission ($\tau \approx 10^{-13}$ sec) and impact ionization ($\tau \approx 10^{-14}$ sec), set in and return the energetic electron back to the bottom of the band. Even at the highest fields, the probability that an electron traverses the entire Brillouin zone and suffers a Bragg reflection is very low and can be neglected in transport studies. The situation is different in *superlattices* where the lattice period is large and the dimensions of the Brillouin zone are short.

In most cases, it is a reasonable approach to describe the electronic motion in a uniform electric field by the effective mass equation.

Density of states and the Fermi level

The phase space of a single particle in d-dimensions. It has $2d$ axes, corresponding to d coordinates and d momenta of the particle, Fig. 20. It is the basic tenet of quantum statistics that a hyper-volume $V^{(d)}$ (say, a hypercube $V^{(d)} = L^d P^d$) in the phase space contains N distinct states, where for a spin-$\frac{1}{2}$ particle

$$N = \frac{2\,V^{(d)}}{(2\pi\hbar)^d} \; .$$

The factor of 2 in the numerator comes from states of different spin. For a hyper-sphere of radius R, the hyper-volume equals $R^d\,\pi^{d/2}/\Gamma(1+d/2)$.

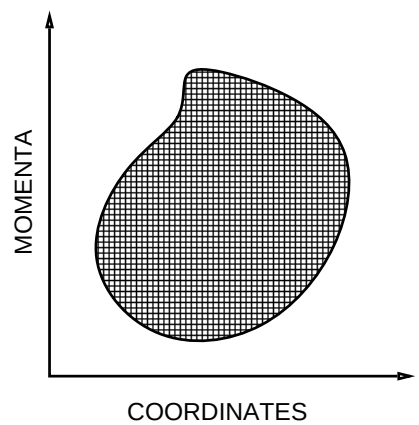
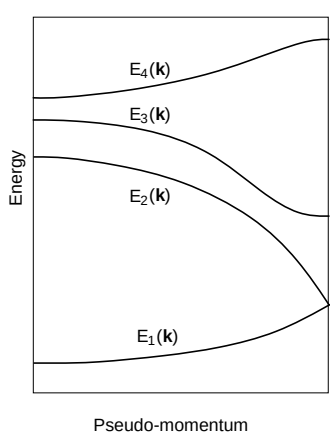


Figure: Hypervolume $L^d P^d$ in the phase space contains N cells of volume $(2\pi\hbar)^d$ and twice as many electronic states. ***In semiconductors*** the situation is similar, but instead of electron momentum one deals with its *crystal momentum* and the above number of states is ***per each band***.



For a free electron gas, the energy-dispersion relation is isotropic, $E = p^2/2m$, which allows a simple counting of states in spherical energy shells $dE = (2E/m)^{1/2}\,dp$:

$$\text{1D:} \quad dN = \frac{2L\,dp}{2\pi\hbar} = \left[\frac{\sqrt{m}}{\pi\hbar\sqrt{2E}}\right] L\,dE \;;$$

$$\text{2D:} \quad dN = \frac{2L^2\,2\pi p\,dp}{(2\pi\hbar)^2} = \left[\frac{m}{\pi\hbar^2}\right] L^2\,dE \;;$$

$$\text{3D:} \quad dN = \frac{2L^3\,4\pi p^2\,dp}{(2\pi\hbar)^3} = \left[\frac{(2m)^{3/2}\sqrt{E}}{2\pi^2\hbar^3}\right] L^3\,dE \;.$$

In these equations, the quantity in the brackets is the density of states $g(E) \equiv dN/dE$ (per unit length, area or volume of the electron gas). Convenient units for $g(E)$ are $[\mathrm{cm}^{-d}\ \mathrm{eV}^{-1}]$. Figure illustrates the $g(E)$ function for 1, 2, and 3 dimensional free-electron gases.

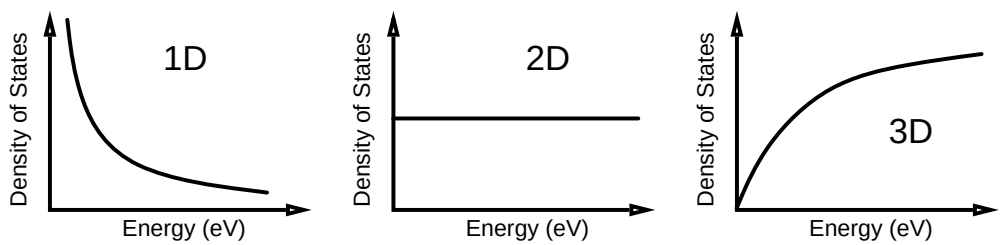


Figure: The density of states in these three plots is measured in different units: $g\ [\mathrm{eV}^{-1}\mathrm{cm}^{-d}]$.

The term Fermi level in semiconductor physics is synonymous with ‘‘chemical potential’’; it is defined by

$$n = \int_{E_0}^{\infty} dE\ g(E)\ f(E - E_F)\ ,$$

where $f(E) = [\exp(E/kT) + 1]^{-1}$ is the Fermi function.

Inversion layer in a silicon MOS structure

The schematic cross-section of a silicon MOS structure is illustrated on the next page along with the band-bending near the Si/SiO_2 interface under a sufficiently large positive gate bias. Let us look in more detail at the band structure of the 2DEG in an inversion layer on the {100} surface.

In a (roughly triangular) quantum well formed near the Si/SiO_2 interface under a positive gate bias, ellipsoids oriented differently with respect to the surface give rise to a quite different subband structure. Let us specify the actual {100} Si surface as a (100) crystal plane. Electrons in the two ellipsoids elongated in [100] direction possess the heavy mass m_l in z direction and an isotropic light mass m_t in any direction lying in the (100) plane. These electrons give rise to the subbands whose bottom-edge energies are denoted by E_n. The other four ellipsoids, whose longitudinal axes lie in the (100) plane, correspond to the light mass m_t in the [100] direction, and their subbands are denoted by E_n'.

Because the quantum-well energy levels scale with $1/m$, one has $E_0 < E_0'$, and so the inversion-layer electrons in their ground subband have an isotropic light mass. The order of the higher-lying subbands can be established only on the basis of self-consistent numerical calculations. The subband energies depend not only on the field and the background doping but also on the temperature, which affects the relative population of higher-lying subbands and, hence, the self-consistent field.

‡ Notation: equivalent crystallographic planes, e.g., (100), $(01\bar{0})$, etc., are collectively denoted by {100}. Similarly, equivalent (symmetry-related) directions in the reciprocal lattice, e.g., [100], [010], etc., are collectively referred to as the <100> direction.

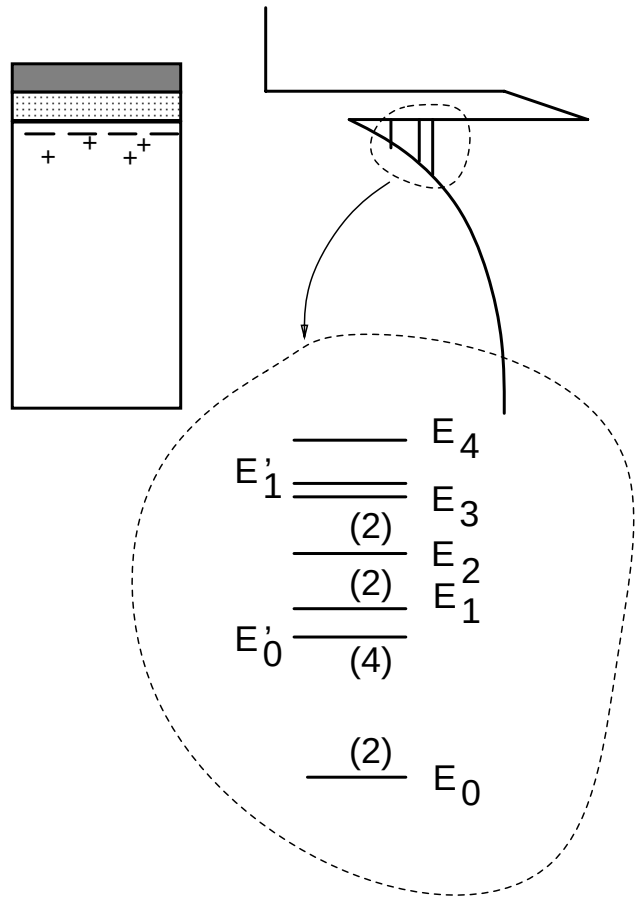


Figure: The order of the seven lowest subbands, calculated for an inversion layer at a Si-{100} surface with $n_S = 10^{12}\,\mathrm{cm}^{-2}$ in a lightly-doped ($N_A = 10^{15}\,\mathrm{cm}^{-3}$) p-type material at room temperature. The levels E_0' and E_1 are close in energy and, in fact, change their order at lower T and/or n_S.