From Atomic Orbitals to Energy Bands

A step-by-step introduction to the electronic structure of solids and electrical conductivity

Stephan Bökelmann

Abstract

Why does copper conduct electric current while diamond does not? And why can silicon be switched between insulating and conducting by tiny amounts of foreign atoms? This text answers these questions in five steps, all of which follow from a single equation – the Schrödinger equation: (1) A single atom possesses discrete energy levels (orbitals). (2) When two atoms are brought together, each level splits into a bonding and an antibonding molecular orbital. (3) When $N \approx 10^{23}$ atoms are brought together, the levels broaden into energy bands, between which band gaps may lie. (4) The Pauli principle fills these bands from the bottom up; the Fermi level describes how far. (5) Current flows precisely when there are partially filled bands. All calculations are carried out explicitly and in small steps, and the original historical papers are cited as sources. The only prerequisites are school mathematics (derivatives, complex numbers) and the basic idea of the wave function.

Contents

1 Introduction: the common thread

Let us first state plainly what electrical conductivity *is* on the microscopic level: a material conducts if and only if it contains **mobile charge carriers** – in a solid, electrons that are free to move through the entire crystal. This immediately raises the central question of this text. Every atom consists of protons and electrons, so *every* material is full of electrons – why, then, are some electrons mobile and others not? The distinction is subtle, because in a sense all electrons move: even an electron firmly bound to its atom is in constant quantum motion within its orbital. But this motion is *confined* – the electron circulates around one nucleus and, on average, transports no charge anywhere. A mobile electron is different in kind: it can *leave* its atom and pass on to the next one, and from there to the next, wandering through the whole crystal. Conductivity is therefore not a question of whether electrons are present or whether they move, but of whether the crystal offers them states in which this atom-to-atom travel is possible – and whether those states are available to be used.

Answering when and why such states exist and are usable is the program of this text. Electrical conductivity turns out not to be a standalone, mysterious material property; it is the macroscopic consequence of three microscopic facts:

1. The Schrödinger equation determines which energies an electron in a given potential is *allowed* to take at all (allowed states).

2. In a crystal made of very many atoms, these allowed energies form *bands*, separated by forbidden *band gaps*.

3. The Pauli principle and the temperature determine which of these states are actually *occupied*.

The band states of point 2 are precisely the atom-to-atom travel states described above: as we will compute in Section 5, every state in a band stretches across the entire crystal, with the electron distributed over all atoms at once. Whether an electron in such a state actually contributes to a current is then a question of point 3: current can only flow if electrons can move into free states in their energetic neighborhood. Whether that is possible depends solely on *how the bands are occupied*. A convenient measure for this is the *Fermi level*: it decides nothing – it is a characteristic quantity that we compute from the number of electrons and the supply of states, and from whose position relative to the bands we can directly read off the conduction behavior. We now develop this line of reasoning from the ground up.

2 The tool: the time-independent Schrödinger equation

2.1 What the equation says

An electron is described by a wave function $\psi(\mathbf{r})$; $|\psi(\mathbf{r})|^2$ is the probability density of finding the electron at position $\mathbf{r}$.

This identification deserves a moment of honesty, because it is *not* a theorem: it is a postulate, added to the theory by Max Born in 1926 [9] (famously, the square instead of the magnitude appears only in a footnote added in proof). The Schrödinger equation itself says nothing about what ψ *means*; Born's rule is the bridge between the theoretical object ψ and what a detector registers, and it has survived a century of experiments. There are, however, strong reasons why – if ψ is to carry a probability interpretation at all – it must be exactly $|\psi|^2$: first, ψ itself is complex and oscillates through negative values, while $|\psi|^2$ is automatically non-negative; second, the Schrödinger equation conserves $\int |\psi|^2\, dV$ in time (and no other power of $|\psi|$), exactly as a total probability of 1 requires; third, interference experiments show that *amplitudes* add, not probabilities – the double-slit pattern is $|\psi_1 + \psi_2|^2$, in direct analogy to optics, where the intensity is the square of the field amplitude. One might ask whether the logic could be run in reverse: the probability density is what is measurable – can one reconstruct ψ from it? Not uniquely: the measured $|\psi|^2$ is blind to the *phase* of ψ, and the phase is physically essential (it carries the momentum and all capacity for interference). Reconstructing a quantum state from data therefore requires measuring several complementary distributions, not just the position density. The wave function remains the primary theoretical object; Born's rule connects it to observation.

States with a sharp energy E satisfy the *time-independent Schrödinger equation*

$$\underbrace{-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r})}_{\text{kinetic energy}} + \underbrace{V(\mathbf{r})\,\psi(\mathbf{r})}_{\text{potential energy}} = E\,\psi(\mathbf{r}). \tag{1}$$

One can read it like a balance sheet: curvature of the wave function (kinetic part) plus potential yields the total energy. The word *curvature* deserves a closer look, because it is the key to reading the equation. Mathematically it refers to the second derivative $\nabla^2\psi$ (in one dimension simply ψ''): the first derivative gives the slope of the graph of ψ, the second tells how fast that slope changes – how strongly the graph *bends*. A straight line has zero curvature; a rapidly oscillating wave bends back and forth violently and has large curvature. Now look at where the curvature sits in the equation: the kinetic term is $-\frac{\hbar^2}{2m}\nabla^2\psi$, so **strong bending means high kinetic energy**. This is the wave version of a familiar fact: a short wavelength must bend on a short scale (large curvature), and by de Broglie's relation $p = h/\lambda$ a short wavelength means large momentum, hence large kinetic energy $p^2/2m$. This one idea will do a lot of work in this text – for example, it is why squeezing an electron into a smaller box raises its energy, and why every additional node (a point where ψ must cross zero and therefore bend more) pushes a state's energy upward.

Two boundary conditions are crucial, because they enforce the quantization:

- ψ must be finite and continuous everywhere;
- $\int |\psi|^2\, dV = 1$ – by Born's rule, the integral of $|\psi|^2$ over any volume is the probability of finding the electron *within that volume*; taken over all of space it is the probability of finding the electron anywhere at all, and since the electron must be *somewhere*, this total must equal 1. For the integral to stay finite, ψ must vanish at infinity.

Who showed that the energies must be discrete? The idea that nature is quantized has a longer history than the Schrödinger equation, and it began as an admitted mathematical trick. In 1900, Max Planck could only fit the measured blackbody radiation spectrum by assuming that energy is exchanged between radiation and matter in packets $E = h\nu$ [1] – an assumption he himself later called an "act of desperation" and regarded for years as a formal device without

physical reality. It was Albert Einstein who in 1905 promoted the trick to physics: the photoelectric effect shows that light itself behaves as if it consisted of such quanta [2]. Even that was far from standard knowledge at the time – the light-quantum hypothesis was rejected by most leading physicists (including Planck) and became broadly accepted only after the Compton experiments of 1923. In 1913, Niels Bohr combined the quantum idea with Rutherford's atom and *postulated* that atoms possess only discrete energies, in order to explain the spectral lines of hydrogen [4] – again by an ad-hoc rule (quantized angular momentum) that could not be justified from anything deeper. The missing link came in 1924 from Louis de Broglie: matter has wave character, with wavelength $\lambda = h/p$ [3]. Schrödinger, who knew de Broglie's idea through Einstein's papers on gas theory, deliberately searched for the wave equation belonging to these matter waves – so the $\hbar$ in his equation is Planck's constant, inherited from a quarter century of quantum history.

What changed in 1926 is the *status* of the discreteness. In his paper *"Quantisierung als Eigenwertproblem"* ("Quantization as an Eigenvalue Problem") [5], Schrödinger *derived* it: where Planck and Bohr had to put the quantization *in* by hand, in Schrödinger's treatment it comes *out* – the allowed energies are the *eigenvalues* of equation (1), just as the allowed tones of a guitar string are the eigenfrequencies of the string. The discreteness is not an additional assumption; it follows solely from the requirement that the wave function satisfy the two boundary conditions above. With that, the trick character of the quantum hypothesis disappeared. That is exactly what we now verify in the simplest example.

2.2 Warm-up exercise: the particle in a box

Let an electron be trapped in a one-dimensional box of width L (Figure 1):

$$V(x) = \begin{cases} 0, & 0 < x < L, \\ \infty & \text{otherwise.} \end{cases} \tag{2}$$

The infinitely high walls need unpacking, because two separate arguments hide in them. First: *why is $\psi = 0$ outside the box?* Outside, the potential energy is $V = \infty$. If the electron had any probability of being there ($|\psi|^2 > 0$), the energy balance of equation (1) would pick up an infinite contribution $V|\psi|^2$ at that spot – but we are looking for states of *finite* energy E. The only way to keep the balance finite is for the probability to vanish wherever the potential is infinite: $\psi = 0$ everywhere outside. (For a wall of finite height this argument would fail, and indeed ψ then leaks slightly into the wall – the infinite wall is the idealized limit in which the leakage is squeezed to zero.) Second: *why also directly at the walls?* Because ψ must be continuous: $|\psi|^2$ is a probability density, and a density that jumps at a point would mean the probability of finding the electron changes abruptly between two places that lie infinitesimally close together – moreover, a jump in ψ would produce an infinite curvature there, hence infinite kinetic energy. So the value of ψ just inside the wall must match the value just outside, and the outside value is 0. Together: $\psi(0) = \psi(L) = 0$.

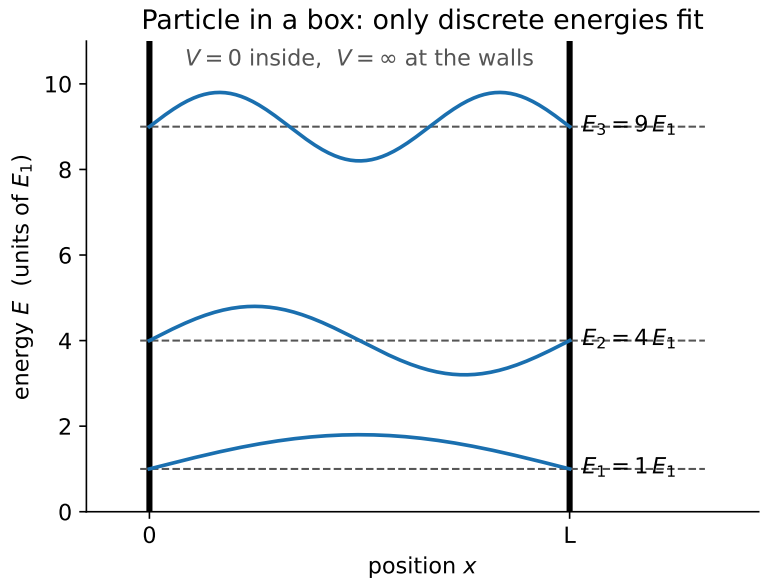

Figure 1: The potential box with the three lowest eigenstates. The box only accommodates standing waves whose half wavelength is an integer fraction of L – hence the discrete energies.

Step by step

Step 1 – equation inside the box. There, $V = 0$. The Schrödinger equation becomes one-dimensional and reads

$$-\frac{\hbar^2}{2m}\,\psi''(x) = E\,\psi(x).$$

We bring it into standard form by multiplying both sides by $-2m/\hbar^2$ and introducing the abbreviation $k^2 \equiv 2mE/\hbar^2$:

$$\psi''(x) = -k^2\,\psi(x).$$

In words: we seek a function whose second derivative reproduces the function itself with a negative prefactor.

Step 2 – general solution. Sine and cosine have this property: $(\sin kx)'' = -k^2 \sin kx$ and $(\cos kx)'' = -k^2 \cos kx$. The most general solution is the superposition

$$\psi(x) = A\sin(kx) + B\cos(kx)$$

with as yet undetermined constants A, B – the same equation and the same solution as for the harmonic oscillator of mechanics.

Step 3 – first boundary condition. At the left edge we must have $\psi(0) = 0$. Inserting $x = 0$:

$$\psi(0) = A\underbrace{\sin 0}_{=0} + B\underbrace{\cos 0}_{=1} = B \overset{!}{=} 0.$$

So the cosine part drops out: $\psi(x) = A\sin(kx)$.

Step 4 – second boundary condition. At the right edge we must have $\psi(L) = 0$:

$$A\sin(kL) = 0.$$

$A = 0$ is ruled out (then ψ would be zero everywhere – no electron). What remains is $\sin(kL) = 0$, and the sine vanishes only at the points $0, \pi, 2\pi, 3\pi, \ldots$ Thus

$$kL = n\pi \quad\Longrightarrow\quad k_n = \frac{n\pi}{L}, \qquad n = 1, 2, 3, \ldots$$

Here is where the quantization happens: not every wavenumber is compatible with the boundary conditions, but only countably many. ($n = 0$ would again give $\psi \equiv 0$; negative n yield the same function with opposite sign, hence no new state.)

Step 5 – energies. We solve the definition of k from Step 1 for E, $E = \hbar^2 k^2/2m$, and insert k_n:

$$\boxed{E_n = \frac{\hbar^2 \pi^2}{2mL^2}\, n^2, \qquad n = 1, 2, 3, \ldots}$$

Step 6 – normalization. One constant is still undetermined: the amplitude A. The boundary conditions cannot fix it, because the Schrödinger equation is linear – any multiple of a solution is again a solution. What fixes A is the *probability interpretation*: $|\psi(x)|^2$ is the probability density of finding the electron at x, and since the electron is certainly somewhere in the box, the probabilities must add up to one,

$$\int_0^L |\psi(x)|^2\, dx = A^2 \int_0^L \sin^2\!\Big(\frac{n\pi x}{L}\Big) dx = A^2\, \frac{L}{2} \overset{!}{=} 1 \quad \Longrightarrow \quad A = \sqrt{\frac{2}{L}}.$$

(The integral is $L/2$ because $\sin^2$ oscillates symmetrically between 0 and 1 and therefore averages to $1/2$ over full half-periods.) The energies do not depend on this step – but without it, $|\psi|^2$ would not be a probability density, and quantities such as expectation values, which we will need from Section 4 on, would come out wrong by an arbitrary factor.

Two observations that we will encounter again:

- The energies are *discrete* – a direct consequence of the confinement, not of any additional assumption.
- The *larger* the box ($L \uparrow$), the *closer* together the levels move. This is worth seeing explicitly: the spacing between two neighboring levels is

  $$E_{n+1} - E_n = \frac{\hbar^2 \pi^2}{2mL^2}\big[(n+1)^2 - n^2\big] = \frac{\hbar^2 \pi^2}{2mL^2}\,(2n+1) \;\propto\; \frac{1}{L^2}.$$

  Doubling L therefore shrinks every level spacing by a factor of four. Numbers make the point drastic: for an electron in a box of atomic size, $L = 1\,\text{nm}$, the ground state lies at $E_1 = \hbar^2\pi^2/2mL^2 \approx 0.38\,\text{eV}$ – level spacings of this order are easily observable. In a box of crystal size, $L = 1\,\text{cm}$, the same formula gives $E_1 \approx 3.8 \times 10^{-15}\,\text{eV}$, fourteen orders of magnitude smaller and far below any measurable resolution. A macroscopic crystal is a gigantic box – its levels lie so densely that they act like a continuum. That is exactly what a *band* will be.

Key point

Confinement ⇒ discrete levels (Schrödinger 1926: energies are eigenvalues). A larger "box" ⇒ more densely spaced levels. Bands are nothing other than extremely densely spaced levels of a macroscopically large system.

3 The single atom: discrete orbitals

3.1 The Coulomb potential

In the atom, the "box" is not a rectangle but the potential funnel of the nucleus with charge Ze:

$$V(r) = -\frac{Ze^2}{4\pi\varepsilon_0\, r}. \tag{3}$$

To save writing, from now on we use *atomic units* ($\hbar = m_e = e = 4\pi\varepsilon_0 = 1$). In these units, 1 unit of energy (Hartree) = 27.2 eV, 1 unit of length (Bohr radius a_0) = 0.529 Å. The potential then reads simply $V(r) = -Z/r$. Figure 2 shows a one-dimensional cut through this funnel together with the energy levels that we will compute shortly; Figure 3 shows the two-dimensional picture.

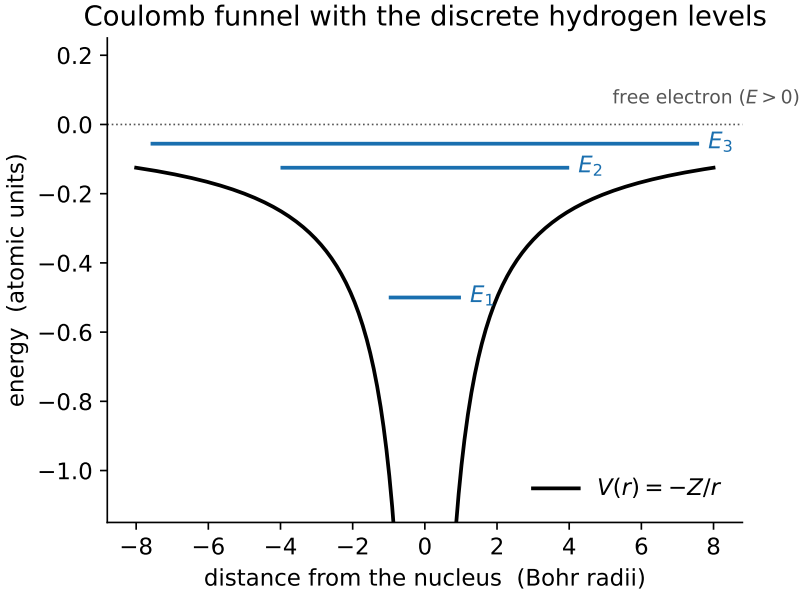

Figure 2: One-dimensional cut through the Coulomb funnel. Unlike the box, the walls are **soft** and become thinner towards the top – which is why the levels move closer together towards the top rather than farther apart. Above $E = 0$ the electron is free (continuous spectrum).

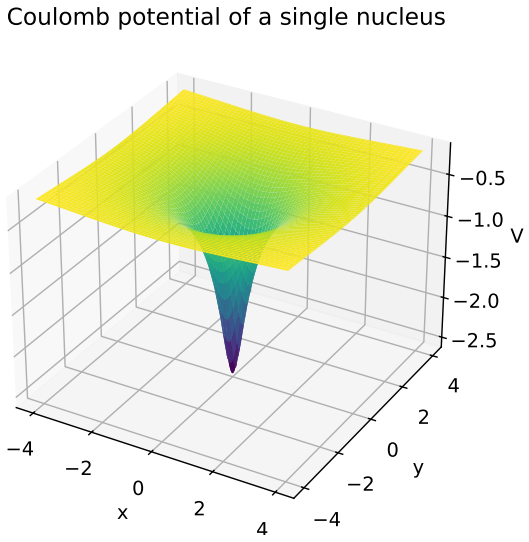

Figure 3: The potential funnel of a single nucleus in two dimensions. The electron is trapped in it just as in the box – only with soft walls.

3.2 How to find the ground state

For the hydrogen atom ($Z = 1$) one finds the ground state not by guessing but through two systematic considerations: first we look at how ψ must behave *far away from the nucleus*; this fixes the form of the solution. Then we insert this form into the full equation; this fixes the energy.

Step by step

Step 1 – Laplacian for spherically symmetric functions. First, which Laplacian are we starting from? The atom is three-dimensional, so the operator in the Schrödinger equation is the full Cartesian one,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Because the potential $-Z/r$ depends only on the distance r, it is convenient to rewrite this operator in spherical coordinates (r, θ, φ), where it takes the form

$$\nabla^2\psi = \frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial\psi}{\partial r}\Big) + (\text{terms with derivatives in } \theta, \varphi).$$

The ground state has no reason to single out any direction; we therefore seek a ψ that depends only on r. For such a function all angular derivatives vanish, and the remaining radial part can be compactly rewritten (as one checks by carrying out both derivatives) as

$$\nabla^2\psi = \frac{1}{r^2}\frac{d}{dr}\Big(r^2\frac{d\psi}{dr}\Big) = \frac{1}{r}\frac{d^2}{dr^2}(r\psi),$$

and with the substitution $u(r) \equiv r\,\psi(r)$ the Schrödinger equation becomes the ordinary differential equation

$$-\frac{1}{2}\,u''(r) - \frac{Z}{r}\,u(r) = E\,u(r).$$

Step 2 – behavior for large r. Far away from the nucleus, Z/r is negligibly small, and what remains is

$$u'' \approx -2E\,u = \kappa^2 u, \qquad \kappa \equiv \sqrt{-2E}$$

(bound states have $E < 0$, so κ is real). The solutions are $e^{+\kappa r}$ and $e^{-\kappa r}$. The growing solution violates normalizability – the electron would be "at infinity". So the wave function must decay like $e^{-\kappa r}$ for large r. This motivates the simplest possible ansatz:

$$u(r) = r\,e^{-\kappa r} \qquad (\text{the factor } r, \text{ so that } \psi = u/r \text{ stays finite at the origin}).$$

Step 3 – inserting the ansatz. Differentiating twice with the product rule:

$$u' = (1 - \kappa r)\,e^{-\kappa r},$$
$$u'' = -\kappa\,e^{-\kappa r} + (1 - \kappa r)(-\kappa)\,e^{-\kappa r} = (\kappa^2 r - 2\kappa)\,e^{-\kappa r}.$$

Inserted into the equation from Step 1 (and canceling $e^{-\kappa r}$):

$$-\frac{1}{2}(\kappa^2 r - 2\kappa) - Z = E\,r.$$

Step 4 – comparing coefficients. This equation must hold for *every* r. So the r-parts and the constant parts must each match separately:

$$r\text{-part:} \quad -\tfrac{1}{2}\kappa^2 = E \quad (\text{automatically satisfied, since } \kappa^2 = -2E),$$
$$\text{constant part:} \quad \kappa - Z = 0 \quad \Longrightarrow \quad \kappa = Z.$$

The ansatz therefore works only for exactly one value of κ – and thus for exactly one energy:

$$\boxed{\psi_{1s} \propto e^{-Zr}, \qquad E_1 = -\frac{\kappa^2}{2} = -\frac{Z^2}{2} \overset{Z=1}{=} -\tfrac{1}{2}\ \text{a.u.} = -13.6\ \text{eV}.}$$

This is exactly the measured ionization energy of hydrogen – the first great confirmation of Schrödinger's equation [5].

3.3 The higher levels

Why are there further bound states besides the ground state, and why precisely with the energies $E_n \propto 1/n^2$? The answer comes from the same mechanism as before, only more systematically: instead of the factor r we allow a whole polynomial in front of the exponential and see when the boundary conditions can be satisfied. (We stick to spherically symmetric states, $\ell = 0$; for $\ell > 0$ the calculation proceeds analogously with the additional term $\ell(\ell+1)/2r^2$ in the potential.)

Step by step

Step 1 – ansatz with a polynomial. We set

$$u(r) = f(r)\, e^{-\kappa r}, \qquad f(r) = a_1 r + a_2 r^2 + a_3 r^3 + \cdots$$

(f starts at r^1, so that $\psi = u/r$ stays finite at the origin).
Step 2 – inserting. By the product rule, $u'' = (f'' - 2\kappa f' + \kappa^2 f)\, e^{-\kappa r}$. Inserted into $-\frac{1}{2}u'' - \frac{Z}{r}u = Eu$ and simplified using $\kappa^2 = -2E$, what remains is

$$f'' - 2\kappa f' + \frac{2Z}{r} f = 0.$$

Step 3 – comparing coefficients power by power. Inserting the series and sorting by powers of r yields, for the coefficient of r^{k-1}, the *recursion formula*

$$a_{k+1} = \frac{2\,(\kappa k - Z)}{k\,(k+1)}\, a_k.$$

It determines all further coefficients from a_1 – the series can therefore be constructed for every κ. So where is the problem?
Step 4 – the series must terminate. For large k the recursion behaves like $a_{k+1} \approx \frac{2\kappa}{k} a_k$ – exactly the construction rule of the exponential series of $e^{2\kappa r}$. If the series does not terminate, $f \sim e^{2\kappa r}$ grows, and $u = f\, e^{-\kappa r} \sim e^{+\kappa r}$ blows up: not normalizable. The only way out: a numerator of the recursion must become zero, so that from there on all coefficients vanish. That happens at $\kappa k = Z$ for some $k = n$:

$$\kappa = \frac{Z}{n} \quad \Longrightarrow \quad \boxed{E_n = -\frac{\kappa^2}{2} = -\frac{Z^2}{2n^2}, \qquad n = 1, 2, 3, \ldots}$$

Once again the boundary condition (normalizability) enforces the quantization.
Step 5 – check against the first two cases. $n = 1$: termination already at $k = 1$, so $f = a_1 r$ and $\kappa = Z$ – exactly the ground state from above. $n = 2$: $\kappa = Z/2$, and the recursion gives

$$a_2 = \frac{2(\kappa \cdot 1 - Z)}{1 \cdot 2}\, a_1 = (\kappa - Z)\, a_1 = -\frac{Z}{2} a_1,$$

hence $u \propto r\,(1 - \frac{Z}{2} r)\, e^{-Zr/2}$ and thus

$$R_{2s}(r) \propto \Big(1 - \frac{Zr}{2}\Big) e^{-Zr/2}.$$

This function has a zero at $r = 2/Z$ – the excited state possesses a *node*. The pattern – **higher energy means more nodes** – is already familiar from the box, and we will see it again for molecules and bands.

Summarized for hydrogen ($Z = 1$):

n	E_n (a.u.)	E_n (eV)	Orbitals
1	$-1/2$	-13.6	$1s$
2	$-1/8$	-3.40	$2s$, $2p$
3	$-1/18$	-1.51	$3s$, $3p$, $3d$

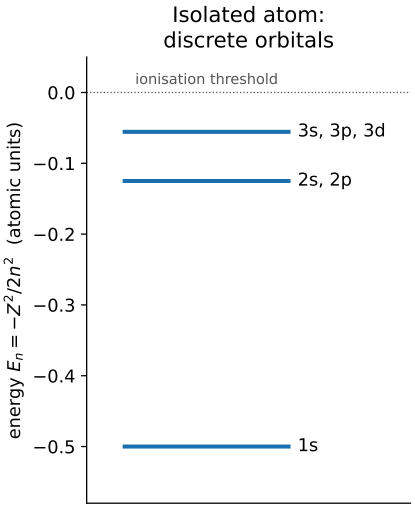

Figure 4: Discrete energy levels of an isolated atom. Each line is an allowed solution of the Schrödinger equation.

3.4 The Pauli principle

How many electrons fit into one orbital? The answer was given by Wolfgang Pauli in 1925 – even before Schrödinger's equation – from the analysis of atomic spectra: *two electrons can never agree in all quantum numbers* [6]. For this, Pauli had to postulate a then-mysterious "two-valuedness" of the electron, which was interpreted in the same year by Uhlenbeck and Goudsmit as an intrinsic angular momentum (*spin*) with the two orientations $\uparrow, \downarrow$ [7]. The practical consequence:

Key point

Each orbital holds at most **two** electrons (spin up/down). This counting rule – not the Coulomb force – will in the end decide between metal and insulator.

4 Two atoms: bonding and antibonding states

4.1 The potential changes

If a second nucleus approaches to a distance R, the two funnels add up:

$$V_{\mathrm{mol}}(\mathbf{r}) = -\frac{Z}{|\mathbf{r}-\mathbf{R}_A|} - \frac{Z}{|\mathbf{r}-\mathbf{R}_B|}. \tag{4}$$

Between the nuclei a region emerges in which *both* nuclei attract the electron (Figures 5 and 6).

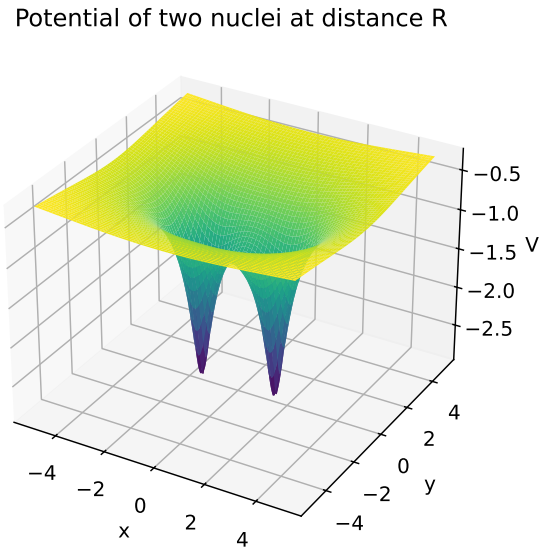

Figure 5: Potential of two nuclei at distance R.

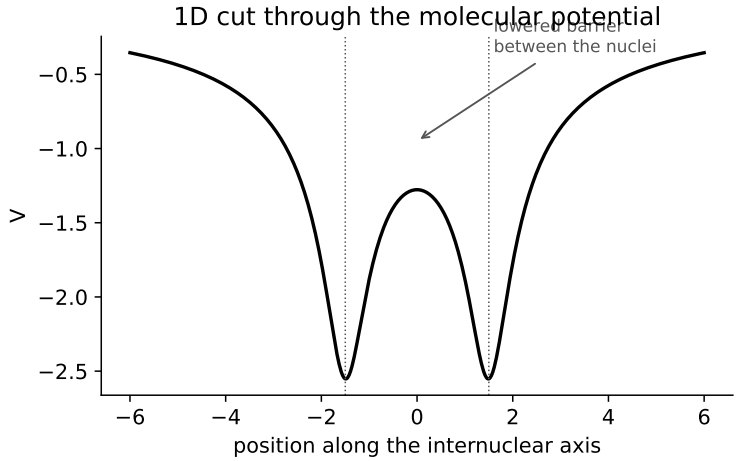

Figure 6: Cut along the internuclear axis: the potential barrier between the nuclei is lowered.

4.2 Ansatz: linear combination of atomic orbitals (LCAO)

Solving the Schrödinger equation exactly in the two-center potential is laborious. The decisive physical insight is due to Heitler and London (1927), who were the first to explain the covalent bond of the H_2 molecule quantum mechanically [12]; the formulation used here in terms of *molecular orbitals as linear combinations of atomic orbitals* (LCAO-MO) goes back to Lennard-Jones and Mulliken [15, 14].

The idea: if the electron sits close to nucleus A, it sees almost only that nucleus's funnel – there the solution should look approximately like the atomic orbital ψ_A. Analogously for nucleus B. A good approximate ansatz for the total solution is therefore the superposition

$$\psi = c_A\,\psi_A + c_B\,\psi_B \tag{5}$$

with initially free coefficients c_A, c_B. Since both nuclei are identical, $|\psi|^2$ must not distinguish between A and B; this forces $|c_A| = |c_B|$, and exactly two possibilities remain:

$$\psi_+ = \frac{1}{\sqrt{2(1+S)}}\,(\psi_A + \psi_B) \qquad \text{(bonding)}, \tag{6}$$

$$\psi_- = \frac{1}{\sqrt{2(1-S)}}\,(\psi_A - \psi_B) \qquad \text{(antibonding)}. \tag{7}$$

Let us remember the structure of equation (5) well – in Section 5.2 we will apply the *same* ansatz, with N instead of two terms, to the entire crystal.

4.3 The overlap integral

In the normalization factors the quantity

$$S(R) = \int \psi_A(\mathbf{r})\,\psi_B(\mathbf{r})\,dV \tag{8}$$

appears, the *overlap integral*. The integrand is the *product* of both orbitals: it is appreciably different from zero only where *both* wave functions are large at the same time – that is, in the zone between the nuclei (Figure 7, left). S thus measures how strongly the orbitals interpenetrate:

- $R \to 0$: the orbitals lie on top of each other, the integral becomes the normalization integral $\int \psi_A^2\,dV = 1$, so $S \to 1$.
- $R \to \infty$: nowhere are both orbitals large at the same time, so $S \to 0$.

For two 1s orbitals the integral can be computed in closed form (in elliptic coordinates; we quote the result):

$$S(R) = e^{-R}\Big(1 + R + \frac{R^2}{3}\Big) \qquad (R \text{ in Bohr radii}). \tag{9}$$

At the bond length of the H_2 molecule, $R = 1.4\,a_0$, this gives $S \approx 0.75$ – the orbitals thus interpenetrate massively (Figure 7, right).

Where does the value $R = 1.4\,a_0$ come from? It is a *measured* quantity: $R = 0.741\,\text{Å} = 1.40\,a_0$. It was first determined in the 1920s from the *band spectrum* of molecular hydrogen: a rotating molecule has quantized rotational energies $E_J = \hbar^2 J(J+1)/2I$, so the spacing of the rotational lines in the spectrum directly yields the moment of inertia $I = \mu R^2$ (with the reduced mass μ) – and from I one obtains R. The classic analysis of the hydrogen band spectrum is due to Hori (1927) [13]; it provided exactly the number against which Heitler and London could test their theory in the same year [12].

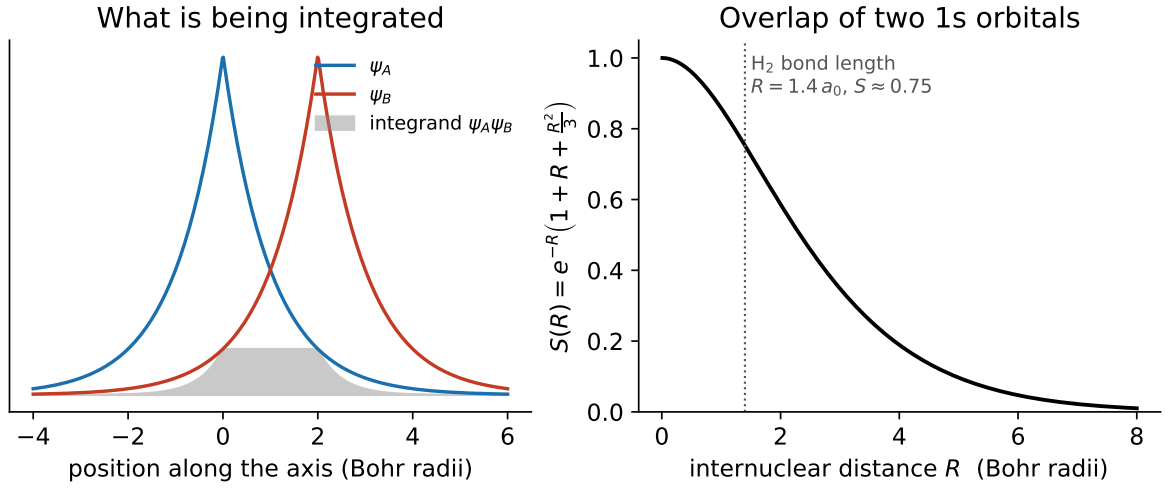

Figure 7: Left: the integrand $\psi_A\psi_B$ (gray) lives in the zone between the nuclei. Right: the overlap integral $S(R)$ of two 1s orbitals as a function of the internuclear distance.

4.4 The energy splitting – computed explicitly

How large are the energies of the two combinations? For a *normalized* wave function, the mean energy one would measure is the expectation value $E = \int \psi\,\hat{H}\,\psi\,dV$: at every point, $\hat{H}\psi$ contains

the local energy contribution, and the integral weights it with the probability amplitude ψ. Our combinations $\psi_A \pm \psi_B$, however, are *not* normalized yet – their norm $\int \psi^2\, dV$ is not 1. We could first normalize them by dividing by $\sqrt{\int \psi^2\, dV}$; inserting that normalized function into the expectation value produces the same factor twice (once for each ψ), which is exactly why the general formula is the quotient

$$E = \frac{\int \psi\, \hat{H}\, \psi\, dV}{\int \psi^2\, dV}. \tag{10}$$

The denominator simply repairs the missing normalization. We abbreviate the integrals that appear:

$$\alpha \equiv \int \psi_A\, \hat{H}\, \psi_A\, dV, \qquad \beta \equiv \int \psi_A\, \hat{H}\, \psi_B\, dV. \tag{11}$$

α is the energy expectation value of the unchanged atomic orbital ψ_A in the molecular Hamiltonian – approximately the old atomic level. β couples both orbitals and is therefore called the coupling or resonance integral; like S, its integrand lives in the overlap zone, β is negative and grows in magnitude with the overlap. By symmetry we also have $\int \psi_B \hat{H} \psi_A\, dV = \beta$ and $\int \psi_B \hat{H} \psi_B\, dV = \alpha$.

Step by step

Step 1 – numerator for $\psi_+ \propto \psi_A + \psi_B$. We multiply out; four terms arise:

$$\int (\psi_A + \psi_B)\hat{H}(\psi_A + \psi_B)\, dV = \underbrace{\alpha}_{A\text{–}A} + \underbrace{\beta}_{A\text{–}B} + \underbrace{\beta}_{B\text{–}A} + \underbrace{\alpha}_{B\text{–}B} = 2(\alpha + \beta).$$

Step 2 – denominator. In the same way, with $\int \psi_A \psi_B\, dV = S$ and $\int \psi_{A,B}^2\, dV = 1$:

$$\int (\psi_A + \psi_B)^2\, dV = 1 + S + S + 1 = 2(1 + S).$$

Step 3 – forming the quotient, and likewise for ψ_- (there the cross terms change sign):

$$\boxed{E_+ = \frac{\alpha + \beta}{1 + S}, \qquad E_- = \frac{\alpha - \beta}{1 - S}.}$$

Since $\beta < 0$, E_+ lies *below* and E_- *above* the atomic level α. For small overlap ($S \ll 1$) we have approximately $E_\pm \approx \alpha \pm \beta$: the level splits symmetrically by $2|\beta|$ (Figure 8). For large S the splitting becomes asymmetric: the antibonding state is raised more strongly ($1 - S$ in the denominator!) than the bonding state is lowered – this will become important shortly for He_2.

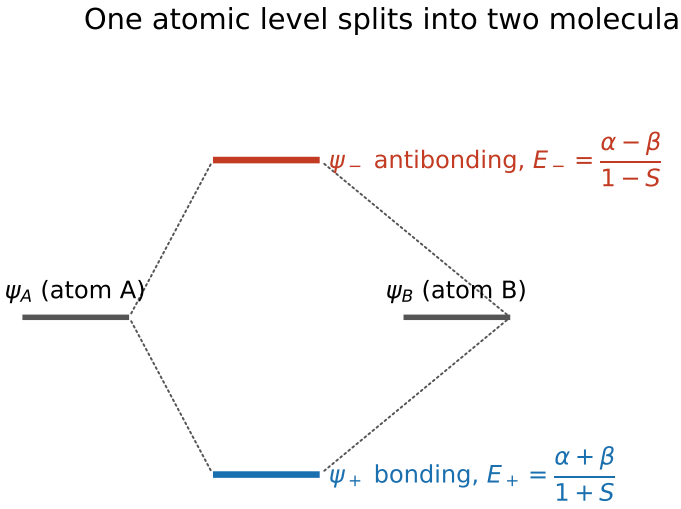

Figure 8: One atomic level becomes two molecular levels: bonding (lowered) and antibonding (raised).

4.5 Intuitive interpretation: nodes cost energy

Why does ψ_+ lie lower? In ψ_+ the wave functions add up between the nuclei – exactly where the potential is most attractive, a lot of electron density sits. In ψ_- they subtract: right in the middle a *nodal plane* with $|\psi|^2 = 0$ arises (Figure 9). The density is pushed out of the favorable region, and the additional curvature of the wave function costs kinetic energy on top – the same **more nodes = more energy** that we saw for the box and for the $2s$ orbital.

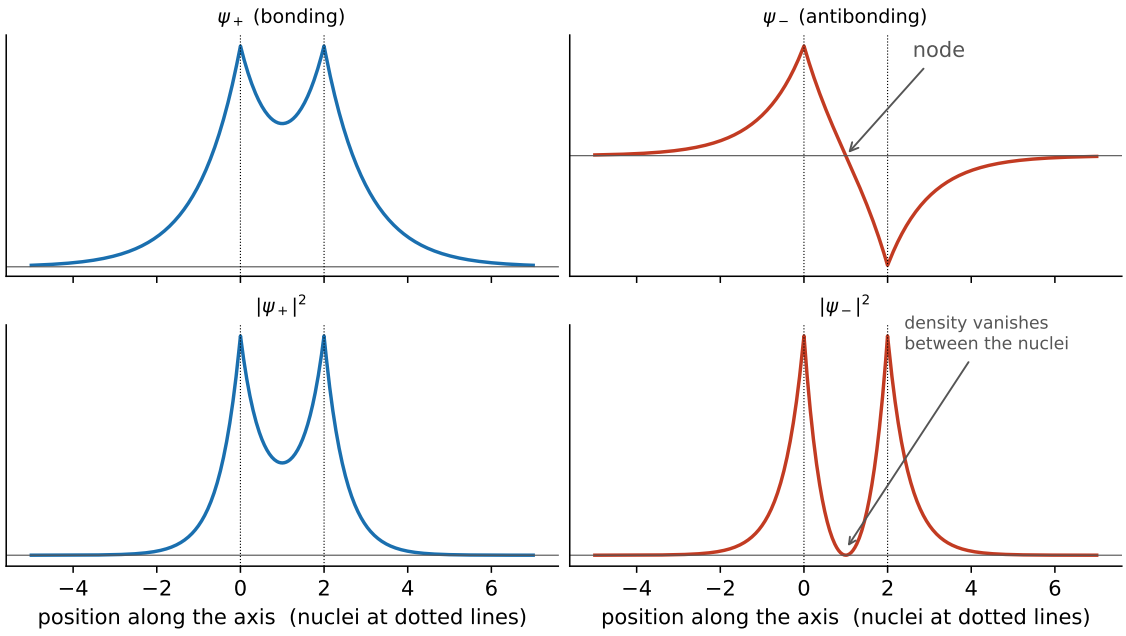

Figure 9: Bonding and antibonding wave functions together with the probability density. The antibonding orbital has a node between the nuclei.

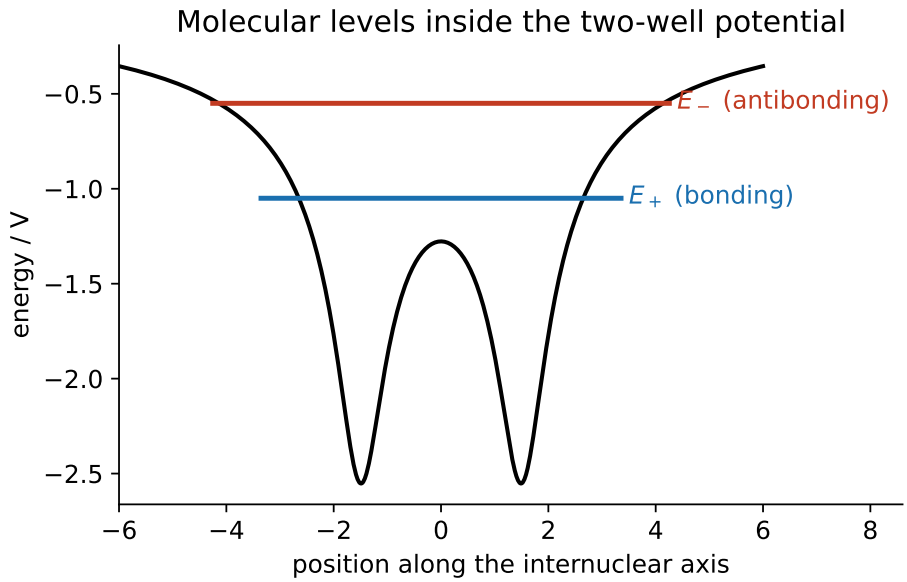

Figure 10: Both molecular levels, drawn into the one-dimensional molecular potential.

4.6 Occupation rules: who sits in which orbital?

Up to this point we have only computed *states*. Which of them are actually occupied is governed by the following, each with a clear historical source:

1. **Exclusion principle** (Pauli 1925 [6]): at most two electrons per orbital, with opposite spin [7].
2. **Energy minimization at** $T = 0$: the ground state of the total system fills the orbitals from the bottom up (the "Aufbau principle", systematically formulated by Bohr for the periodic table [8]).
3. **Statistics at** $T > 0$ (Fermi 1926 [10], Dirac 1926 [11]): at finite temperature, a state of energy E is occupied with probability $f(E) = \left[e^{(E-E_F)/k_BT} + 1\right]^{-1}$ – the *Fermi-Dirac distribution*, to which we return in Section 6.

An important remark on the logical status of these points: the third is *not* an independent rule of its own. Fermi and Dirac *derived* the distribution $f(E)$ by applying the general laws of thermal equilibrium (which state of a system is most probable at temperature T) *under the constraint of point 1* – counting only those configurations in which no state holds more than its allowed electrons. The exclusion principle is the physical input; the Fermi-Dirac distribution is its statistical consequence. Point 2 is in turn contained in point 3 as the limit $T \to 0$: there, $f(E)$ becomes a sharp step – everything below E_F occupied, everything above empty – which is exactly filling from the bottom up. The genuinely independent ingredients are thus only the exclusion principle and ordinary thermodynamics.

Applied to our molecules:

- **H_2**: 2 electrons → both in ψ_+ (spin up/down). Both gain energy ⇒ stable bond.
- **He_2**: 4 electrons → two in ψ_+, two *must*, because of Pauli, go into ψ_-. How large is S here? Helium's $1s$ orbitals are more compact than hydrogen's, because the electron feels an effective nuclear charge $Z_{\text{eff}} \approx 1.7$; the overlap formula then reads $S(R) = e^{-Z_{\text{eff}}R}(1 + Z_{\text{eff}}R + \frac{(Z_{\text{eff}}R)^2}{3})$, which at a hypothetical bond length of $R \approx 1.5\,a_0$ gives $S \approx 0.45$ – far from negligible. The asymmetry of the splitting (denominators $1 \mp S$) therefore matters: the two electrons in ψ_- lose *more* energy than the two in ψ_+ gain ⇒ no bond.

In general one defines the *bond order*

$$\mathrm{BO} = \frac{n_{\mathrm{bonding}} - n_{\mathrm{antibonding}}}{2}. \tag{12}$$

The table lists it for the classic diatomics, together with the measured *dissociation energy* D_e – the energy required to pull the molecule apart into two free atoms, i.e. the depth of the molecular binding. If the bond-order picture is right, larger BO should mean larger D_e – and it does:

Molecule	bonding e^-	antibonding e^-	BO	Finding
H_2	2	0	1	stable
He_2	2	2	0	not bound
N_2	8	2	3	$D_e \approx 9.9\,\mathrm{eV}$
O_2	8	4	2	$D_e \approx 5.2\,\mathrm{eV}$
F_2	8	6	1	$D_e \approx 1.6\,\mathrm{eV}$

Key point

The *existence* of the levels is provided by the Schrödinger equation; whether anything useful comes of it is decided by the *occupation* (Pauli + energy minimization + Fermi-Dirac statistics). This interplay is exactly the same principle that will later distinguish metals from insulators.

5 From molecule to solid: energy bands

5.1 The idea: N atoms $\Rightarrow$ N levels

With two atoms, each atomic level split into two molecular levels. With three atoms it becomes three, with four atoms four – and with $N \approx 10^{23}$ atoms, N levels arise, all squeezed into an energy window of finite width. The spacing of neighboring levels is then of the order of $10^{-23}\,\mathrm{eV}$ – immeasurably small. The levels practically merge into a continuum: an **energy band** (Figure 11). The quantum-mechanical theory of electrons in the periodic potential of a crystal was founded in 1929 by Felix Bloch [17].

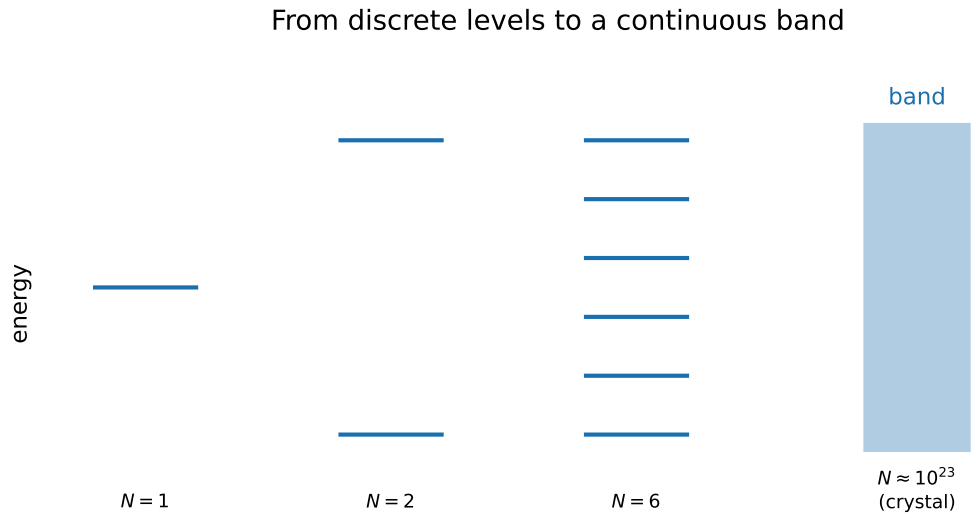

Figure 11: From $N = 1$ via $N = 2$ and $N = 6$ to $N \to \infty$: the discrete level broadens into the band.

5.2 The chain of N atoms: the LCAO ansatz, take three

We now verify this claim by calculation – using the *same* ansatz as for the molecule. As a reminder: for H_2 we wrote $\psi = c_A\psi_A + c_B\psi_B$ (equation 5). For a chain of N identical atoms at spacing a we generalize literally:

$$\psi = \sum_{j=1}^{N} c_j\,\psi_j, \tag{13}$$

where ψ_j is the orbital at atom j and the coefficients c_j specify with what weight and sign each atom contributes.

It helps to recall what distinguished the two molecular solutions: nothing but their coefficients. In ψ_+ both orbitals entered with the *same* sign, $(c_A, c_B) \propto (1, 1)$ – the wave function has no node between the atoms. In ψ_- they entered with *opposite* signs, $(c_A, c_B) \propto (1, -1)$ – one node between the atoms. Each solution was therefore characterized by a *pattern of signs* along the (two-atom) chain, and the energy grew with the number of nodes in that pattern. For N atoms we expect the same: each solution will be one particular pattern $(c_1, c_2, \ldots, c_N)$, ranging from all atoms in phase (no nodes, lowest energy) to alternating signs from atom to atom (a node in every bond, highest energy), with all gradations in between. Our task is to find which patterns actually solve the Schrödinger equation, and with which energies. As before, let $\alpha = \langle\psi_j|\hat{H}|\psi_j\rangle$ and $\beta = \langle\psi_j|\hat{H}|\psi_{j\pm1}\rangle$ (coupling only between nearest neighbors; we neglect the overlap integral S for simplicity). This model is called the *tight-binding model* (model of strongly bound electrons) [17].

Step by step

Step 1 – equation for the coefficients. We insert the ansatz into $\hat{H}\psi = E\psi$ and project onto the orbital at atom j (i.e., we multiply by ψ_j and integrate). Of the sum, only three terms survive – the atom itself and its two neighbors:

$$\alpha\, c_j + \beta\,(c_{j-1} + c_{j+1}) = E\, c_j.$$

Why does this appear? It is nothing but the Schrödinger equation $\hat{H}\psi = E\psi$, viewed from the perspective of atom j: the right side is the total energy E times the weight c_j that the solution gives to atom j; the left side says where this energy contribution comes from – the atom's own level α, weighted with c_j, plus the coupling β to whatever amplitude sits on the two neighboring atoms. Atom j notices the rest of the chain *only* through c_{j-1} and c_{j+1}. That is one such equation per atom, so N coupled equations – the same structure as for a chain of coupled pendulums.

Step 2 – wave ansatz. For chains of identical coupled elements, trying a wave that advances by a constant phase from element to element is the standard tool of mechanics – it solves the vibrating chain and the coupled pendulums alike. For electrons in the periodic potential of a crystal, this ansatz was introduced by Felix Bloch in 1929 [17]; solutions of this form are called *Bloch waves* in his honor. We try

$$c_j = e^{\,\mathrm{i}kja},$$

a plane wave running from atom to atom with wavenumber k. Inserting:

$$\alpha\, e^{\mathrm{i}kja} + \beta\big(e^{\mathrm{i}k(j-1)a} + e^{\mathrm{i}k(j+1)a}\big) = E\, e^{\mathrm{i}kja}.$$

Step 3 – canceling. We factor out $e^{\mathrm{i}kja}$ and divide by it:

$$\alpha + \beta\big(e^{-\mathrm{i}ka} + e^{+\mathrm{i}ka}\big) = E.$$

The j-dependence has disappeared – so the ansatz solves all N equations simultaneously. With Euler's formula $e^{\mathrm{i}x} + e^{-\mathrm{i}x} = 2\cos x$ it follows that

$$\boxed{E(k) = \alpha + 2\beta\cos(ka).}$$

Step 4 – which k are allowed? As for the particle in a box, the ends of the chain turn the continuum into discrete values: there are exactly N allowed k-values, uniformly distributed in the interval $-\pi/a < k \leq \pi/a$ (the *Brillouin zone*). N atoms, N states – the bookkeeping works out.
Step 5 – consistency check with the molecule. At $k = 0$ all $c_j = 1$: all atoms oscillate in phase, nowhere a sign change – the maximally *bonding* combination, energy $\alpha + 2\beta$ (minimum, since $\beta < 0$). At $k = \pi/a$ we have $c_j = e^{i\pi j} = (-1)^j$: the sign alternates from atom to atom, a node in every bond – maximally *antibonding*, energy $\alpha - 2\beta$ (maximum). The band interpolates continuously between the two extremes that we know from the molecule (Figure 12).

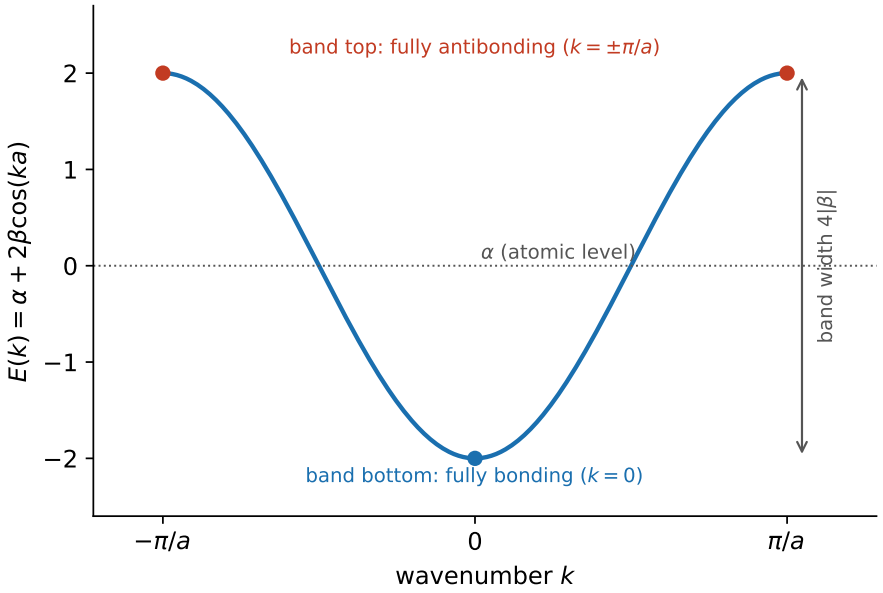

Figure 12: The dispersion relation $E(k) = \alpha + 2\beta\cos(ka)$ of the atomic chain: bonding at the bottom, antibonding at the top, bandwidth $4|\beta|$.

The result confirms and sharpens the picture:

- The N allowed energies lie between $\alpha - 2|\beta|$ and $\alpha + 2|\beta|$: the band has **width** $4|\beta|$. Strong overlap ⇒ wide band.
- From *each* atomic level ($1s$, $2s$, $2p$, …) a separate band emerges. If the atomic levels lie far apart and the bands are narrow, a forbidden region remains between the bands: the **band gap** E_g.

Key point

A band is a "smeared-out" atomic level: lower edge = bonding, upper edge = antibonding, width = $4|\beta|$ (overlap of the orbitals). A band made of N atoms contains N orbitals and, by Pauli, holds $2N$ electrons.

5.3 Density of states

Within a band the states are not distributed uniformly over energy. The *density of states* $D(E)$ specifies how many states are present per energy interval (Figure 13). For what follows it suffices that $D(E) > 0$ within a band and $D(E) = 0$ in the band gap.

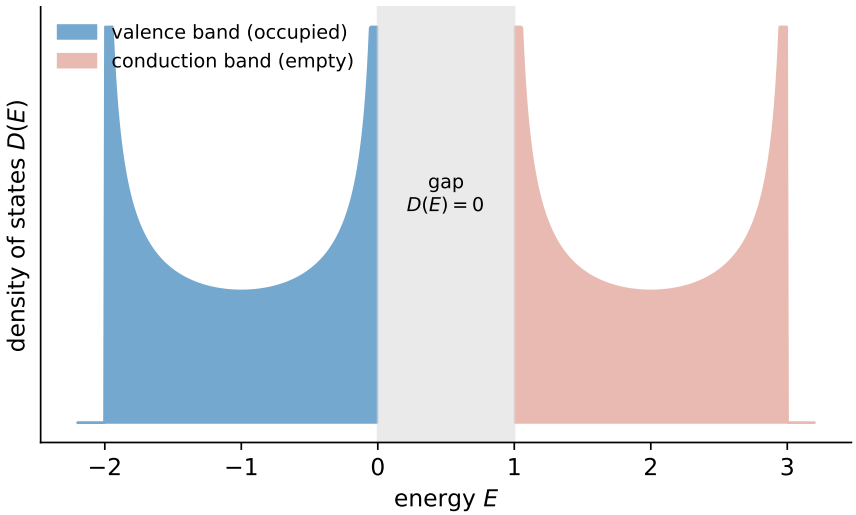

Figure 13: Density of states $D(E)$: many states in the bands, none in the gap.

6 The Fermi level: who sits where?

6.1 Filling up according to Pauli

A crystal with N atoms and ν valence electrons per atom must accommodate νN electrons. According to the occupation rules from Section 4.6, the states are filled from the lowest energy upwards, each with at most two electrons. The energy of the last state occupied at $T = 0$ is called the **Fermi energy** E_F.

Once more, clearly, because it is often phrased incorrectly: E_F is not a cause but a *derived measure*. The physics lies in the occupation; E_F summarizes it in a single number. But because the occupation and the position of E_F determine each other uniquely, one may use E_F as a convenient *read-off criterion* – and that is exactly how we use it in what follows.

A simple count with far-reaching consequences:

- $\nu = 1$ (e.g. sodium, one $3s$ electron): the $3s$ band offers room for $2N$ electrons, but is filled with $N \Rightarrow$ the band is *half full*. E_F lies in the middle of the band.
- Even number of electrons per cell and no band overlap: all bands are either completely full or completely empty. E_F lies *in the gap* between the highest full band and the lowest empty band.

6.2 Finite temperature: the Fermi distribution

At $T > 0$ the Fermi-Dirac distribution holds [10, 11]

$$f(E) = \frac{1}{e^{(E-E_F)/k_BT} + 1}. \tag{14}$$

At $E = E_F$ we have $f = \frac{1}{2}$; the softening of the occupation edge has a width of a few k_BT ($k_BT \approx 0.025\,\mathrm{eV}$ at room temperature). This number – 25 millielectronvolts – is worth memorizing: it is the yardstick against which all band gaps must be measured.

6.3 Numerical example: the Fermi energy of copper

In the simplest model (free electron gas: conduction electrons as particles in a macroscopic box, Sommerfeld 1928 [16]), counting the states up to the density n (electrons per volume) yields:

$$E_F = \frac{\hbar^2}{2m}\left(3\pi^2 n\right)^{2/3}. \tag{15}$$

Copper contributes one conduction electron per atom. To use this formula we need the density n of *mobile* electrons – and for copper that number is far from obvious, because a copper atom brings along 29 electrons. Their configuration is $[\mathrm{Ar}]\,3d^{10}\,4s^1$, and in the crystal the three groups of electrons play very different roles:

- *The 18 core electrons* (the argon shell) occupy compact inner orbitals close to the nucleus. Neighboring atoms are simply too far away for these orbitals to overlap: their coupling β is essentially zero, so by our band formula their "bands" have essentially zero width – they remain sharp atomic levels, completely filled, and their electrons stay with their atom.
- *The ten* $3d$ *electrons* reach out further; their orbitals do overlap moderately, and they form genuine bands. But five d orbitals per atom provide, by Pauli, room for exactly ten electrons per atom – and copper delivers exactly ten. The d bands are therefore *completely filled*, and a completely filled band, as Section 7.1 will show, carries no current.
- *The single* $4s$ *electron* lives in the outermost, most extended orbital. It overlaps strongly with the neighbors (large $|\beta|$, wide band), and the resulting band offers room for $2N$ electrons while only N arrive: the band is *half filled*.

Of the 29 electrons per atom, exactly one therefore ends up in a partially filled band. This is the conduction electron, and hence n equals the density of atoms.

Step by step

Step 1 – electron density n. Copper has mass density $\rho = 8.96\,\mathrm{g/cm^3}$ and molar mass $M = 63.5\,\mathrm{g/mol}$. The atomic density is therefore

$$n_{\mathrm{atoms}} = \frac{\rho\, N_A}{M} = \frac{8.96\times10^{3}\,\mathrm{kg/m^3}\times 6.02\times10^{23}\,\mathrm{mol^{-1}}}{63.5\times10^{-3}\,\mathrm{kg/mol}} \approx 8.5\times10^{28}\,\mathrm{m^{-3}}.$$

With one conduction electron per atom, $n = n_{\mathrm{atoms}}$.
Step 2 – intermediate quantity $(3\pi^2 n)^{2/3}$.

$$3\pi^2 n \approx 29.6\times8.5\times10^{28}\,\mathrm{m^{-3}} \approx 2.5\times10^{30}\,\mathrm{m^{-3}} \;\Rightarrow\; (3\pi^2 n)^{2/3} \approx 1.86\times10^{20}\,\mathrm{m^{-2}}.$$

Step 3 – inserting.

$$E_F = \frac{(1.055\times10^{-34}\,\mathrm{Js})^2}{2\times9.11\times10^{-31}\,\mathrm{kg}}\times 1.86\times10^{20}\,\mathrm{m^{-2}} \approx 1.1\times10^{-18}\,\mathrm{J} \approx 7\,\mathrm{eV}.$$

The literature value is $E_F(\mathrm{Cu}) = 7.0\,\mathrm{eV}$ – the simple model does astonishingly well.

The degenerate electron gas. An electron system in this state – occupation dictated by the Pauli principle rather than by temperature – is called **completely degenerate**. (The word has nothing to do with "degenerate energy levels"; here it labels a regime of quantum statistics.) What lies behind the term: a classical gas, upon cooling, would let all particles sink into the lowest state. An electron gas cannot do that, because Pauli allows only two electrons per state; the electrons are forced to stack up to E_F, and even at $T = 0$ the topmost ones carry the enormous kinetic energy of $\sim 7\,\mathrm{eV}$. One defines the *Fermi temperature* $T_F = E_F/k_B \approx 8\times10^4\,\mathrm{K}$: only at temperatures of this order would the electron gas behave classically. At room temperature, $T/T_F \approx 0.004$ – only a wafer-thin window of $\sim k_B T$ around E_F is thermally redistributed, while the remaining $99.9\ldots\,\%$ of the electrons are frozen in the Pauli stack. A system this cold, measured against T_F, has practically the entire structure of its occupation determined by the Pauli principle and not by the temperature – and that is precisely the regime the word *degenerate* refers to.

7 Electrical conductivity

7.1 Why a full band does not conduct

The central, often surprising point: *having many electrons is not enough.* In every band, to each right-moving state ($+k$) there belongs a mirror-image left-moving one ($-k$) with the same energy – one sees this directly from our band formula, since the cosine is an even function:

$$E(-k) = \alpha + 2\beta\cos(-ka) = \alpha + 2\beta\cos(ka) = E(k). \tag{16}$$

If the band is *full*, $+k$ and $-k$ are occupied pairwise without exception; all electron currents cancel exactly. An electric field would like to shift the electrons towards slightly higher $+k$ values – but every target state is already occupied, and Pauli forbids double occupation. The electrons cannot move aside: **net current zero.** This explanation of why insulators exist at all is due to A. H. Wilson (1931) [18].

If, on the other hand, the band is only *partially* filled, free places lie directly above the occupied states. The field can make the occupation asymmetric (more $+k$ than $-k$) – current flows.

7.2 Can the conductivity be calculated?

Yes – and the calculation also answers the obvious question of whether "how hard it is for the field to reach free states" is the measure of conductivity. The answer is twofold, and the dividing line is instructive:

- *Whether* free neighboring states are reachable at all decides the *number of mobile charge carriers* n. In a metal the next free state costs arbitrarily little energy – the field has a trivially easy time. In a semiconductor the gap must first be overcome (thermally or by doping) – here lies the "difficulty", and it enters *exponentially.*
- *How much current* the available carriers then transport is limited not by the field but by *scattering*: collisions with lattice vibrations and defects slow the electrons down again after a mean collision time τ.

Both together yield the famous formula of Drude [19], which Sommerfeld later underpinned quantum mechanically [16]:

Step by step

Step 1 – balance of forces. A field $\mathcal{E}$ exerts the force $-e\mathcal{E}$ on each electron and accelerates it. On average, an electron is scattered after the collision time τ and "forgets" its extra velocity. In the steady state it therefore carries the mean *drift velocity*

$$v_D = \frac{\text{acceleration} \times \text{time}}{1} = \frac{e\mathcal{E}}{m}\,\tau.$$

The same result deserves a translation into the band picture, because there it shows *which* electrons actually carry the current. In k-space the occupied states of a metal fill a sphere of radius k_F, the *Fermi sphere* – the three-dimensional version of "all states from $-k_F$ to $+k_F$ are full". A field changes every electron's wavenumber at the rate $\hbar\,dk/dt = -e\mathcal{E}$ (this is Newton's second law, with $\hbar k$ playing the role of momentum), so between two collisions *the entire sphere* drifts by $\delta k = e\mathcal{E}\tau/\hbar$ to one side; scattering keeps it in this slightly displaced steady position. The sphere is then no longer centered at $k = 0$, and the pairwise $\pm k$ cancellation fails at the surface: on the leading side, a thin crescent of states is occupied whose mirror partners on the trailing side are now empty. Everything deeper inside the sphere still cancels pairwise – the net current is carried entirely by that uncompensated crescent of electrons near E_F. The displacement is tiny ($\delta k/k_F \sim 10^{-10}$ for ordinary fields),

which is why the response is linear in $\mathcal{E}$: Ohm's law.
Step 2 – current density. n carriers per volume with charge $-e$ and velocity v_D give

$$j = n\,e\,v_D = \frac{n\,e^2\,\tau}{m}\,\mathcal{E} \qquad \Longrightarrow \qquad \boxed{\sigma = \frac{n\,e^2\,\tau}{m}.}$$

Step 3 – numerical check for copper. With $n = 8.5 \times 10^{28}\,\mathrm{m}^{-3}$ and the collision time (known from measurements) $\tau \approx 2.5 \times 10^{-14}\,\mathrm{s}$:

$$\sigma = \frac{8.5 \times 10^{28} \times (1.60 \times 10^{-19})^2 \times 2.5 \times 10^{-14}}{9.11 \times 10^{-31}}\,\frac{1}{\Omega\mathrm{m}} \approx 6 \times 10^7\,\mathrm{S/m},$$

in good agreement with the measured value $\sigma_{\mathrm{Cu}} = 5.96 \times 10^7\,\mathrm{S/m}$.

The formula $\sigma = ne^2\tau/m$ sorts the material classes cleanly: between metals and insulators, τ/m varies only by small factors – what varies by more than twenty orders of magnitude is n. That is why the question "How hard is it for electrons to reach free states?" really is *the* question – but it acts through the carrier density n, not through the force exerted by the field.

7.3 Metal, semiconductor, insulator

With this, the classification is complete (Figure 14), historically first formulated by Wilson [18]:

- **Metal**: E_F lies in a partially filled band (or bands overlap). Free neighboring states exist at every temperature ⇒ good conductor.
- **Semiconductor**: full valence band, empty conduction band, separated by a *moderate* gap ($E_g \approx 0.1$–$3\,\mathrm{eV}$). Thermal excitation across the gap is rare but possible; doping helps massively.
- **Insulator**: the same situation with a *large* gap ($E_g \gtrsim 4\,\mathrm{eV}$). Thermal excitation is practically impossible.

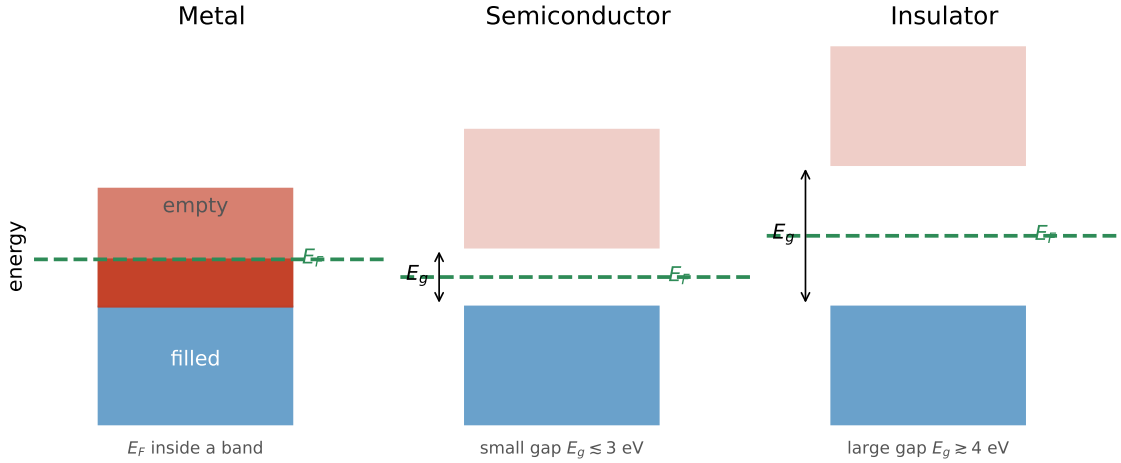

Figure 14: Band diagrams of a metal, a semiconductor, and an insulator. The difference between semiconductor and insulator is only the size of the gap.

The origin of the names *valence band* and *conduction band*. Both names describe the *role* of the states, not their origin in the calculation. *Valence band* (from Latin *valentia*, worth or

capacity): it arises from the bonding combinations of the valence orbitals – that is, from precisely those electrons which in chemistry determine the valence of an element and carry the covalent bonds. *Conduction band*: the next higher band, whose electrons – if any get into it – are freely mobile and conduct the current.

Here a clarification is needed. The often-heard statement "in the ground state the conduction band is empty" holds in this form only for the *ideal semiconductor or insulator at* $T = 0$. More precisely:

- At $T > 0$ the conduction band of a semiconductor is always occupied by some thermally excited electrons. For pure silicon this is not merely a theoretical statement but a measured fact: the intrinsic carrier density was determined from conductivity and Hall measurements on high-purity crystals by Morin and Maita in 1954 [20]; the modern precision value at 300 K is $n_i \approx 1.0 \times 10^{16}\,\mathrm{m}^{-3}$ ($1.0 \times 10^{10}\,\mathrm{cm}^{-3}$) [21]. The same order of magnitude follows from our own formula: with the effective densities of states of silicon, $N_c \approx 2.8 \times 10^{25}\,\mathrm{m}^{-3}$ and $N_v \approx 1.0 \times 10^{25}\,\mathrm{m}^{-3}$, one gets $n_i = \sqrt{N_c N_v}\, e^{-E_g/2k_BT} \approx 1.7 \times 10^{25} \times 3 \times 10^{-10}\,\mathrm{m}^{-3} \approx 5 \times 10^{15}\,\mathrm{m}^{-3}$. This is tiny compared to the $\sim 10^{29}\,\mathrm{m}^{-3}$ valence electrons – but not zero.
- In doped semiconductors the conduction band is populated even at low temperatures (Section 8).
- For metals the distinction loses its meaning: the partially filled band *is* simultaneously valence and conduction band – there, the "conduction band" is already partially occupied in the ground state.

Both bands are genuine, distinct solutions of the Schrödinger equation – the valence band of predominantly bonding, the conduction band of predominantly antibonding character – and not shifted copies of each other.

7.4 Numerical example: why the gap size acts so dramatically

For a *pure (intrinsic)* semiconductor, the density of thermally generated charge carriers scales as

$$n_i \propto e^{-E_g/2k_BT}. \tag{17}$$

Where does the factor 2 come from – and does it always hold? It does *not* hold for every system; it is a property of the intrinsic case: there, every excitation creates a *pair* (electron in the conduction band + hole in the valence band). From $n = p$ and the law of mass action $n\,p \propto e^{-E_g/k_BT}$, taking the square root yields equation (17) – the gap energy is, in a sense, shared between two particles. Equivalently: the Fermi level then lies (for similar densities of states of valence and conduction band) near the middle of the gap, and each electron only has to overcome half the gap starting from E_F. In general – in particular in the doped semiconductor, where E_F lies somewhere else entirely – one has instead

$$n \propto e^{-(E_c - E_F)/k_BT}, \tag{18}$$

with the conduction band edge E_c: the exponent always contains the distance between E_F and the band edge, and only in the intrinsic special case is this distance $\approx E_g/2$. (If the effective masses of electrons and holes are very different, E_F shifts somewhat away from the middle of the gap even in the intrinsic case – the correction is logarithmically small.)

For the intrinsic comparison at room temperature ($k_BT = 0.025\,\mathrm{eV}$):

Material	E_g	$e^{-E_g/2k_BT}$
Silicon	1.1 eV	$e^{-22} \approx 3 \times 10^{-10}$
Diamond	5.5 eV	$e^{-110} \approx 10^{-48}$

The difference of a factor of 5 in the gap becomes, in the exponent, a factor of $\sim 10^{38}$ in the charge carrier density. That is why silicon is a (weak) conductor and diamond, for all practical purposes, a perfect insulator – although both have the same crystal structure and the same number of valence electrons.

7.5 What an external field does

An applied voltage tilts the electrostatic potential and with it the band edges (Figure 15). *Existing* free charge carriers then drift downhill – quantitatively described by $\sigma = ne^2\tau/m$ from above. The tilting alone, however, creates no charge carriers: in an insulator, nothing happens beyond dielectric polarization.

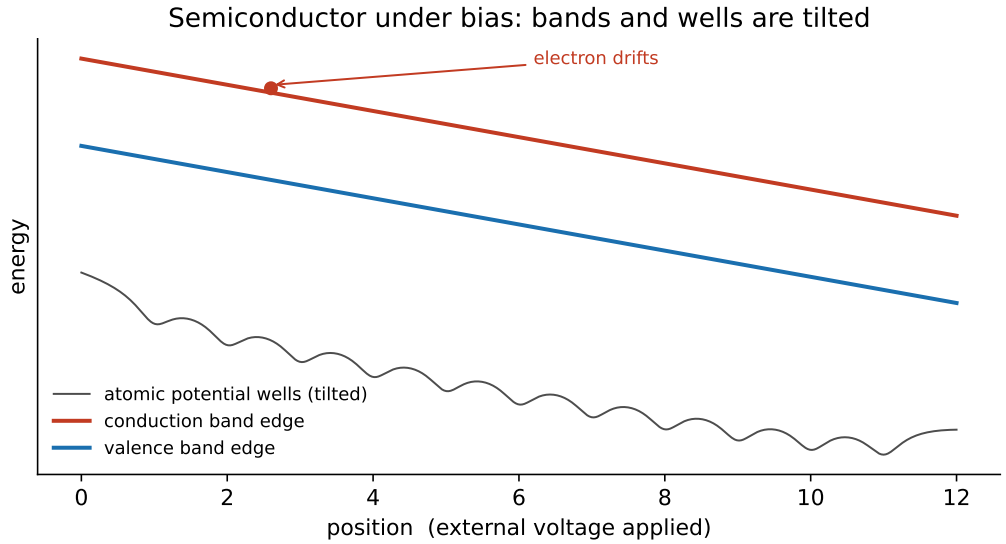

Figure 15: Semiconductor under an applied voltage: potential wells and band edges are tilted; existing charge carriers drift.

8 Doping: shifting the Fermi level deliberately

8.1 The principle

If one replaces isolated atoms in silicon (4 valence electrons) with phosphorus (5 valence electrons), one electron per foreign atom is left over that finds no bond. It is only weakly bound to the (singly positively acting) phosphorus ion – a "hydrogen atom inside the crystal".

8.2 Calculation: the binding energy of the donor electron

Step by step

Step 1 – starting point. For hydrogen we know the binding energy $E_1 = 13.6\,\mathrm{eV}$ and know from the derivation how it depends on the fundamental constants: in SI units,

$$E_1 = \frac{m\,e^4}{2(4\pi\varepsilon_0)^2\hbar^2} \;\propto\; \frac{m}{\varepsilon^2}.$$

Step 2 – what changes in the crystal. Exactly two ingredients:

- The Coulomb field is screened by the polarizability of the crystal: $\varepsilon_0 \to \varepsilon_r\varepsilon_0$ with $\varepsilon_r \approx 12$ for silicon. Because of $E \propto 1/\varepsilon^2$, this weakens the binding by a factor of $12^2 = 144$.

- The electron moves in the periodic potential like a particle with *effective mass* $m^* \approx 0.3\, m_e$ (the band curvature replaces the free mass). Because of $E \propto m$, this weakens the binding by a factor of 0.3.

Step 3 – applying the scaling.

$$E_D \approx 13.6\,\text{eV} \times \frac{m^*/m_e}{\varepsilon_r^2} = 13.6\,\text{eV} \times \frac{0.3}{144} \approx 0.03\,\text{eV} = 30\,\text{meV}.$$

The measured value for phosphorus in silicon is 45 meV – the rough estimate hits the right order of magnitude. For comparison: $k_B T \approx 25\,\text{meV}$ at room temperature.

The donor level thus lies only $\sim k_B T$ below the conduction band edge: already room temperature ionizes practically all donors, and their electrons populate the conduction band. The Fermi level thereby moves up close beneath the conduction band edge (*n-type*). Mirror-symmetrically, acceptors (B, Al: only 3 valence electrons) create shallow empty levels just above the valence band; they capture valence electrons and leave behind mobile *holes* (*p-type*).

Key point

Doping does not change the bands, but the *occupation* – and thus the carrier density n in $\sigma = ne^2\tau/m$. Parts-per-million fractions of foreign atoms change the conductivity by many orders of magnitude – the foundation of all semiconductor technology.

9 Outlook: further conduction mechanisms

Band conduction in crystalline solids is the most important, but not the only mechanism:

- **Hopping/polaron conduction** in disordered and organic solids (electrons hop between localized states),
- **Ionic conduction** in electrolytes and solid-state ionic conductors,
- **Ballistic conduction** in nanostructures (no collisions over the length of the sample),
- **Percolation** in composite materials,
- **Superconductivity** below a critical temperature.

All of these ultimately rest on the same foundation: the quantum-mechanical permission or prohibition of states and the statistics of their occupation.

10 Summary

The common thread once more, in five steps:

1. **Confinement quantizes** (Schrödinger 1926). The particle in a box and the Coulomb potential of the atom yield discrete energies; for hydrogen $E_n = -13.6\,\text{eV}/n^2$.
2. **Coupling splits** (Heitler–London 1927). Two overlapping orbitals yield a bonding level (E_+, density between the nuclei) and an antibonding level (E_-, node between the nuclei); splitting $\approx 2|\beta|$.

3. **Many couplings yield bands** (Bloch 1929). The LCAO ansatz with N terms yields $E(k) = \alpha + 2\beta\cos(ka)$: a band of width $4|\beta|$, bonding at the bottom, antibonding at the top, with possible gaps between the bands.

4. **Pauli fills the bands** (Pauli 1925, Fermi/Dirac 1926). The number of valence electrons determines the occupation; the Fermi level summarizes it as a single characteristic number and lets us read off the conduction behavior: E_F in a band or in a gap?

5. **Current needs free neighboring states** (Wilson 1931). Full bands do not conduct ($\pm k$ pairs); partially filled bands conduct with $\sigma = ne^2\tau/m$. Temperature, doping, and fields only shift charge carriers and thus the occupation – the bands themselves come from the Schrödinger equation.

The entire phenomenology of metals, semiconductors, and insulators thus follows from a single equation – the Schrödinger equation – together with the Pauli principle, which fills its solutions up to the Fermi level.

References

[1] M. Planck, *Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum*, Verhandlungen der Deutschen Physikalischen Gesellschaft **2**, 237–245 (1900).

[2] A. Einstein, *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*, Annalen der Physik **17**, 132–148 (1905).

[3] L. de Broglie, *Recherches sur la théorie des quanta*, Annales de Physique **10**(3), 22–128 (1925).

[4] N. Bohr, *On the Constitution of Atoms and Molecules*, Philosophical Magazine **26**, 1–25 (1913).

[5] E. Schrödinger, *Quantisierung als Eigenwertproblem (Erste Mitteilung)*, Annalen der Physik **79**, 361–376 (1926).

[6] W. Pauli, *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*, Zeitschrift für Physik **31**, 765–783 (1925).

[7] G. E. Uhlenbeck, S. Goudsmit, *Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons*, Die Naturwissenschaften **13**, 953–954 (1925).

[8] N. Bohr, *Der Bau der Atome und die physikalischen und chemischen Eigenschaften der Elemente*, Zeitschrift für Physik **9**, 1–67 (1922).

[9] M. Born, *Zur Quantenmechanik der Stoßvorgänge*, Zeitschrift für Physik **37**, 863–867 (1926).

[10] E. Fermi, *Zur Quantelung des idealen einatomigen Gases*, Zeitschrift für Physik **36**, 902–912 (1926).

[11] P. A. M. Dirac, *On the Theory of Quantum Mechanics*, Proceedings of the Royal Society A **112**, 661–677 (1926).

[12] W. Heitler, F. London, *Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik*, Zeitschrift für Physik **44**, 455–472 (1927).

[13] T. Hori, *Über die Analyse des Wasserstoffbandenspektrums im äußersten Ultraviolett*, Zeitschrift für Physik **44**, 834–854 (1927).

[14] R. S. Mulliken, *The Assignment of Quantum Numbers for Electrons in Molecules I*, Physical Review **32**, 186–222 (1928).

[15] J. E. Lennard-Jones, *The Electronic Structure of Some Diatomic Molecules*, Transactions of the Faraday Society **25**, 668–686 (1929).

[16] A. Sommerfeld, *Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik*, Zeitschrift für Physik **47**, 1–32 (1928).

[17] F. Bloch, *Über die Quantenmechanik der Elektronen in Kristallgittern*, Zeitschrift für Physik **52**, 555–600 (1929).

[18] A. H. Wilson, *The Theory of Electronic Semi-Conductors*, Proceedings of the Royal Society A **133**, 458–491 (1931).

[19] P. Drude, *Zur Elektronentheorie der Metalle*, Annalen der Physik **306**, 566–613 (1900).

[20] F. J. Morin, J. P. Maita, *Electrical Properties of Silicon Containing Arsenic and Boron*, Physical Review **96**, 28–35 (1954).

[21] A. B. Sproul, M. A. Green, *Improved value for the silicon intrinsic carrier concentration from 275 to 375 K*, Journal of Applied Physics **70**, 846–854 (1991).