

Chapter 21

Electrical Conduction in Solids

First Edition

Hamzeh Khanpour

hamzeh.khanpour@agh.edu.pl

Faculty of Physics and Applied Computer Science

AGH University of Krakow

Physics 2

Physics for Technical and Engineering Students -
Volume II

Electricity, Magnetism, Waves, Optics, Relativity,
Quantum Physics, and Nuclear Physics

1 September 2026

Contents

Preface	iii
Conventions and notation	iv
21 Electrical Conduction in Solids	1
21.1 Why conduction in solids is a quantum problem	1
21.2 From isolated atoms to bands in a crystal	2
21.2.1 Allowed bands, forbidden gaps and occupancy	3
21.2.2 Connection to the Pauli exclusion principle	4
21.2.3 Real crystals and controlled impurities	4
21.3 Macroscopic transport: current density, conductivity and mobility	5
21.3.1 The Drude estimate and what it teaches	6
21.3.2 Mean free path, scattering and metallic resistance	8
21.3.3 Why a perfect periodic lattice does not itself cause resistance	8
21.4 Metals, insulators and semiconductors in the band picture	9
21.4.1 Metals	9
21.4.2 Insulators	9
21.4.3 Semiconductors	9
21.5 Intrinsic semiconductors: electrons and holes	11
21.5.1 Why a hole behaves like a positive carrier	11
21.5.2 Generation and recombination	11
21.6 Doping: controlling carrier populations	12
21.6.1 n-type material	13
21.6.2 p-type material	13
21.6.3 The Fermi level as a chemical-potential marker	14
21.7 Measuring carriers: the Hall effect as a bridge	14
21.7.1 Separating sample and contact resistance	16
21.8 The p–n junction: diffusion, drift and the depletion region	16
21.8.1 Built-in potential and equilibrium	17
21.9 Biasing the junction and rectification	18
21.9.1 The ideal diode equation	18
21.10 The junction diode in circuits	20
21.10.1 Full-wave rectification and protection	21
21.11 Light-emitting diodes and radiative recombination	22
21.11.1 Direct and indirect band gaps	22
21.12 Photodiodes and solar cells: running the junction optically	24
21.13 Transistors: using a third terminal to control current	25
21.13.1 Bipolar carrier-injection control	25
21.14 Temperature dependence: metals versus semiconductors	26
21.15 How semiconductor devices are fabricated: a physics-level overview	27
21.15.1 Oxidation, deposition and pattern transfer	27
21.15.2 Doping by diffusion and ion implantation	28
21.15.3 Contacts, passivation and packaging	28
21.16 Thermal effects, self-heating and electrostatic discharge	28
21.16.1 Thermal runaway as positive feedback	29
21.16.2 Electrostatic discharge	29

21.16.3	Interpreting specifications at their test conditions	29
21.16.4	Choosing a device model	30
21.17	Engineering and applied-science connections	30
21.17.1	Thermistors and temperature sensing	30
21.17.2	Radiation and particle detectors	31
21.17.3	Power electronics	31
21.17.4	Computing and digital logic	31
21.17.5	Photovoltaics and optical communication	31
21.18	Laboratory connection: measuring a diode I – V curve	31
21.18.1	Uncertainty and measurement limitations	32
21.19	Computational connection: diode modelling and parameter scans	33
21.20	Main chapter summary	34
21.21	Advanced Topics for Technical and Engineering Students	36
21.21.1	Crystal states and carrier dynamics	36
21.21.2	Quantum statistics and carrier populations	39
21.21.3	Nonequilibrium carriers, diffusion and recombination	46
21.21.4	Junction electrostatics and capacitance	49
21.21.5	Injection, the diode law and switching	54
21.21.6	Optical and radiation devices	56
21.21.7	Gate control and digital energy	59
21.21.8	Materials and alternative conduction regimes	62
21.21.9	Measurement, noise and model validation	65
21.22	Practice problems	68
21.23	Problem-solving workshop	72
21.23.1	Workshop 1: conductivity from several microscopic routes	72
21.23.2	Workshop 2: junction electrostatics and capacitance	72
21.23.3	Workshop 3: LED optical power from electrical input	72
21.23.4	Workshop 4: a photodiode receiver	73
21.23.5	Workshop 5: solar-cell maximum power	73
21.23.6	Workshop 6: digital switching and gate capacitance	73
21.24	Selected solutions to representative problems	73
21.25	Bridge to the next modern-physics chapter	77
A	Notation dictionary for Chapter 21	78
B	Compact formula sheet	80

Preface

Materials containing the same basic constituents can differ in conductivity by many orders of magnitude. Classical electric forces and circuit laws describe the response, but quantum states explain why mobile carriers and nearby unoccupied states are available in some solids and scarce in others. This chapter connects atomic levels and Pauli exclusion to bands, carrier populations, scattering and electronic devices.

The core develops bands, metals and semiconductors, electrons and holes, doping, transport measurements, p–n junctions, rectifiers, optical devices and qualitative transistor action. Worked examples connect microscopic parameters to measurable resistance, Hall voltage, diode current and photon energy. The clearly separated enrichment layer supplies state counting, effective mass, drift–diffusion, junction electrostatics and selected device models for strong undergraduate and early graduate students.

The chapter is intended for calculus-based technical and engineering Physics II. Chapters 6–7 provide the current and circuit prerequisites, Chapter 8 supplies magnetic-force and Hall ideas, and Chapters 18–20 supply matter waves, quantum states and the Pauli principle.

The Guideline

When reading any solid-state diagram, identify four things before calculating: **(1)** what physical quantity is on each axis, **(2)** which energy states are allowed, **(3)** which of those states are occupied, and **(4)** which carriers can actually change their state in response to an applied field. Most conceptual errors in this chapter come from confusing “an available state” with “an occupied state.”

Pedagogical boundary

This chapter is an introductory bridge between quantum physics and electronic materials. It does not attempt a full derivation of electronic band structure from Bloch’s theorem, a many-body treatment of electron–electron interactions, a microscopic theory of phonon scattering, or a device-technology treatment of modern integrated circuits. Those topics belong to dedicated courses in solid-state physics, condensed-matter physics and semiconductor-device engineering. Here they are introduced only to the extent needed to understand conduction, semiconductor doping and basic junction devices correctly.

Conventions and notation

Macroscopic electric quantities retain the notation used in the earlier electricity chapters. The electric field is $\mathbf{E}$, current density is $\mathbf{J}$, current is I , conductivity is σ , and resistivity is ρ . The symbol ρ therefore denotes *electrical resistivity* in this chapter unless a different meaning is stated explicitly; volume charge density, when needed, will be written ρ_q to avoid ambiguity.

The elementary positive charge is $e = 1.602 \times 10^{-19}$ C, so an electron has charge $q_e = -e$. Electron and hole number densities are denoted n and p , respectively, with SI unit m^{-3} . Carrier mobilities are μ_n and μ_p , with SI unit $\text{m}^2 \text{V}^{-1} \text{s}^{-1}$. The mean time between momentum-randomizing collisions in the simplest Drude model is τ .

Energy-band diagrams use E_C for the conduction-band edge, E_V for the valence-band edge and

$$E_g = E_C - E_V$$

for the band-gap energy. The Fermi energy or Fermi level is written E_F . In a simple equilibrium band diagram, the vertical axis represents electron energy; it is not a physical height in the crystal. When electrostatic potential varies in space, the band edges may bend as functions of position. Such bending is a compact representation of potential-energy variation, not a literal bending of a material layer.

The Boltzmann constant is $k_B = 1.381 \times 10^{-23}$ J K⁻¹ and Planck's constant is $h = 6.626 \times 10^{-34}$ J s. In semiconductor physics, electronvolts are often more convenient than joules. We shall use

$$1 \text{ eV} = e \times 1 \text{ V} = 1.602 \times 10^{-19} \text{ J}.$$

The thermal voltage

$$V_T = \frac{k_B T}{e}$$

is approximately 25.9 mV at $T = 300$ K.

A complete symbol dictionary and a formula sheet with the model conventions appear in Appendices A and B. In particular, N_D and N_A mean total dopant concentrations; N_D^+ and N_A^- denote their ionized parts. They coincide only when ionization is complete. The symbols I_0 in the core and I_S in enrichment describe the same ideal reverse-saturation scale. Depletion width W is distinct from MOSFET channel dimensions $W_{\text{ch}}, L_{\text{ch}}$.

Note

The word *hole* is a bookkeeping description of a missing electron in an otherwise almost filled valence band. A hole behaves in transport equations as a positively charged mobile carrier, but it is not a proton moving through the lattice and it is not an empty cavity in the material.

Note

The word *band gap* refers to an energy interval containing no allowed extended electronic states in the ideal band model. It is not a geometrical gap between atoms or layers in a crystal.

Chapter 21

Electrical Conduction in Solids

21.1 Why conduction in solids is a quantum problem

A circuit diagram treats a wire as a line, a resistor as a symbol and a diode as an arrowlike component with two terminals. That abstraction is extremely useful, but it hides the microscopic reason current can flow in one material and not in another. At the material level the decisive questions are: Which electronic energy states exist? Which are occupied? Are nearby unoccupied states available? How easily can carriers acquire a net drift velocity while scattering from the lattice, defects and impurities?

A classical free-electron model already explains one important feature of metals: an electric field can produce a small drift velocity superposed on rapid microscopic motion. Yet the classical model cannot explain why some solids with many electrons are excellent insulators, why carrier concentrations in semiconductors depend so strongly on temperature, or why a small amount of doping can create device behaviour such as rectification. Those facts require quantum mechanics and the Pauli exclusion principle.

When isolated atoms are brought together to form a crystal, their electronic wave functions overlap. The discrete atomic levels no longer remain exactly degenerate. Each atomic level splits into a very large number of closely spaced levels. For a macroscopic crystal containing of order 10^{23} atoms, these levels are so dense that we describe them as *bands* of allowed energy separated, in many materials, by intervals with no allowed states called *gaps*.

The electrical response depends not only on whether a band exists but also on whether it is full, empty or partially filled. An electron in a completely filled band cannot respond to a weak electric field in the same simple way as an electron near an unoccupied state, because the Pauli exclusion principle prevents two electrons with the same quantum state. The availability of nearby unoccupied states is therefore central to conduction.

Figure 21.1 connects these microscopic questions to the junction devices developed later in the chapter.

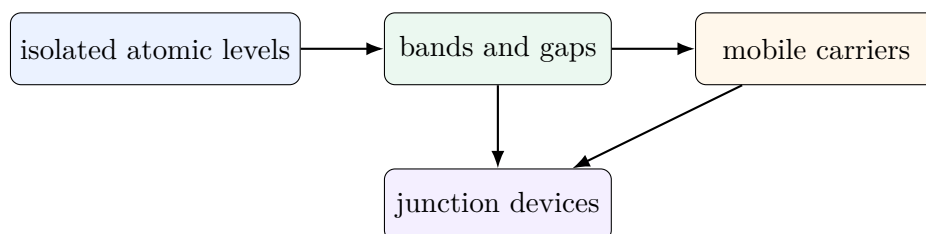


Figure 21.1: Conceptual roadmap. The chapter moves from atomic levels to energy bands, from bands to carrier transport, and from controlled carrier populations to semiconductor junction devices.

Learning goals

After studying the main chapter, the student should be able to:

- explain qualitatively why isolated atomic levels broaden into bands in a crystal;
- distinguish an allowed energy band from a band gap and an occupied state from an available state;
- classify idealized metals, insulators and semiconductors from band diagrams;
- use $\mathbf{J} = \sigma \mathbf{E}$ and the Drude expression $\sigma = ne^2\tau/m$ within their stated ranges of validity;
- explain intrinsic electron–hole generation and why semiconductor conductivity usually increases strongly with temperature;
- explain donor and acceptor doping and identify majority and minority carriers;
- use $\sigma = e(n\mu_n + p\mu_p)$ in a simple semiconductor transport model;
- explain how diffusion and drift create the depletion region of a p–n junction;
- interpret forward bias, reverse bias and rectification physically;
- use the ideal diode equation as an engineering model while recognizing its limitations;
- estimate an LED photon wavelength from a band gap and explain why radiative efficiency depends on material band structure;
- describe qualitatively how a transistor controls current using a third terminal;
- connect band and transport concepts to measurements of resistivity, Hall voltage, diode I – V curves, LEDs, solar cells and electronic circuits.

Key message

The most compact statement of the chapter is: *conductivity is controlled by both the number of mobile carriers and the quantum-mechanical availability of states into which they can move.* “Many electrons” does not by itself imply “good conductor.”

21.2 From isolated atoms to bands in a crystal

Consider first a single atom. In a bound-state quantum description, the electron can occupy only certain discrete energies. If we have two identical atoms very far apart, each atom has essentially the same set of atomic energy levels. As the atoms approach, their electron wave functions begin to overlap. The combined system can no longer be described as two completely independent atoms, and the degeneracy of the isolated atomic levels is lifted.

For N identical atoms, an atomic level can produce a cluster of order N closely spaced levels. The precise counting depends on spin, orbital degeneracy and the detailed crystal structure, but the key point is robust: when N is macroscopic, the spacing between neighbouring levels becomes extraordinarily small. On the energy scale of ordinary experiments the cluster is treated as a continuous allowed band.

The splitting is not equally strong for all atomic states. Core-electron wave functions are tightly localized near nuclei and overlap weakly between neighbouring atoms, so their bands tend to be narrow. Outer-electron states overlap more strongly and can form wider bands. The resulting band structure is determined by the periodic crystal potential and by the quantum-mechanical symmetry of the electronic states.

Figure 21.2 sketches how closely spaced crystal levels emerge from isolated atomic levels.

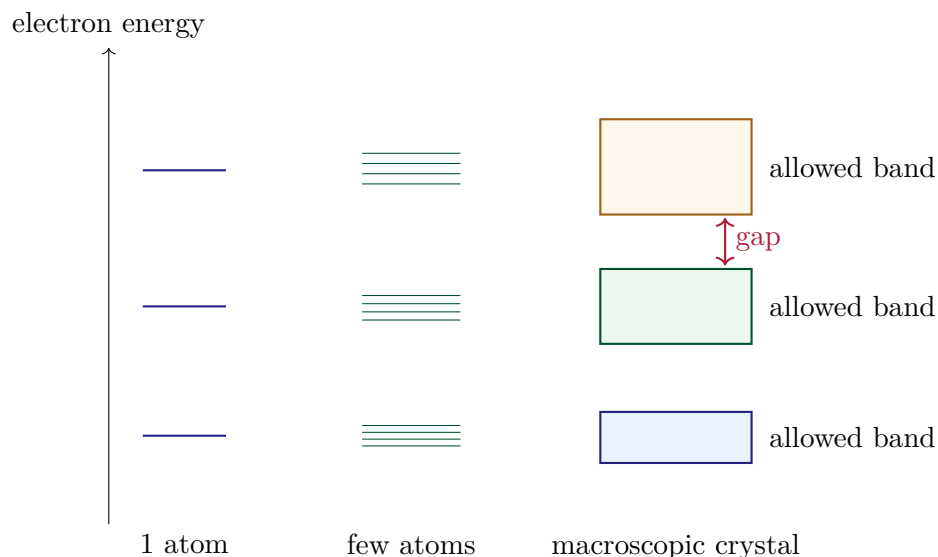


Figure 21.2: Schematic evolution from discrete atomic energy levels to dense allowed energy bands as many atoms form a crystal. The diagram is qualitative; real band structures depend on crystal momentum and direction in the lattice.

21.2.1 Allowed bands, forbidden gaps and occupancy

A band diagram contains two logically different pieces of information. First, the material's quantum structure determines which energies are allowed. Second, the available electrons determine which allowed states are occupied. It is possible to have an allowed band that is almost empty, almost full, exactly full in an idealized zero-temperature picture, or partially filled.

For electrical conduction, a partially filled band is especially important because occupied states can lie close in energy to unoccupied states. A weak external field can then perturb the electronic distribution and create a net current. By contrast, if the highest occupied band is completely filled and a large gap separates it from the next allowed band, a weak field has no nearby empty states into which the electrons can be promoted.

Why does a filled band give zero ordinary longitudinal current? In equilibrium, equal occupation of states with opposite velocities cancels their contributions. More generally, the velocity in a band is proportional to the gradient of its energy in wave-vector space. Summing this gradient over an entire filled, periodic band gives zero. A small field can rearrange a partially filled distribution near its occupied–empty boundary; it cannot create that boundary within a full band. This band-model statement does not exclude polarization, optical interband excitation or strong-field breakdown, and it is not a complete classification of correlated or topological solids.

The highest energy band that is substantially occupied in an ordinary semiconductor is called the *valence band*; the next higher band relevant for transport is called the *conduction band*. Their separation is the band gap E_g . The terms are most useful for semiconductors and insulators. In metals, bands may overlap or a band may be only partially filled, so the simple “full valence band plus empty conduction band” picture is not the best description.

Caution

A vertical band diagram is a diagram in *energy space*. The vertical distance labelled E_g is measured in joules or electronvolts, not metres. Students should never describe an electron as physically “jumping upward through a spatial gap.” The electron changes quantum state and energy.

Guided checks

1. If a band contains unoccupied states but no electrons, can it by itself carry current? Explain.
2. If a band is completely filled, why is “there are many electrons” not enough to guarantee conduction?
3. Why do the energy levels of a macroscopic crystal look continuous on an ordinary laboratory energy scale even though quantum energies remain quantized?
4. Which is a property of the material’s allowed-state structure: occupancy or band gap? Which depends directly on how many carriers are present and on temperature?

21.2.2 Connection to the Pauli exclusion principle

The Pauli exclusion principle states that no two identical fermions can occupy the same one-electron quantum state. In an independent-electron band picture, this means that when all states in a band are occupied, a small perturbation cannot simply move every electron into an already occupied neighbouring state. Current therefore depends on how the distribution of occupied states can be rearranged in response to a field.

This statement should not be overinterpreted. Electrons in a crystal are not literally stationary until an external voltage is applied. They occupy quantum states with well-defined statistical and dynamical properties. The net current vanishes in equilibrium because the occupied states are populated symmetrically so that their contributions cancel. Applying an electric field shifts the distribution slightly away from equilibrium; collisions and scattering oppose that shift and lead to a steady drift in ordinary resistive transport.

21.2.3 Real crystals and controlled impurities

The band picture is often introduced as though every solid were an ideal infinite crystal. Real engineering materials are more varied. A *single crystal* has long-range periodic order across the active region; a *polycrystalline* material consists of many crystalline grains separated by grain boundaries; an *amorphous* solid lacks long-range periodic order even though local bonding remains important. These structural distinctions can strongly affect transport because grain boundaries and disorder introduce additional scattering and localized electronic states.

For the present chapter, the ideal-crystal band picture should be understood as the clean reference model. Its success is practical: many technologically important semiconductor devices are fabricated from highly controlled crystalline materials, and deviations from ideal periodicity can then be treated as defects, interfaces or engineered perturbations.

Why purity and controlled impurities both matter

Electronics requires an apparently paradoxical combination: crystals must be very pure, but selected impurities are then introduced deliberately. Uncontrolled impurities create unpredictable carrier concentrations, traps and scattering. Controlled dopants, by contrast, are introduced at known concentrations and locations to set the intended majority-carrier population and to create device structures such as p–n junctions.

This distinction is central to semiconductor manufacturing. “Pure” in a device context does not mean that no foreign atom exists anywhere. It means that background contamination is low enough that the intentionally introduced dopants and designed interfaces dominate the electrical behaviour.

Guided checks

1. Why can grain boundaries change electrical transport even if each grain has the same chemical composition?
2. Why are controlled dopants useful while uncontrolled impurities are harmful?
3. Why does the simple band diagram remain useful even though a real device contains surfaces, interfaces and defects?

21.3 Macroscopic transport: current density, conductivity and mobility

Before distinguishing metals and semiconductors, it is useful to recall the macroscopic language of transport. In a homogeneous isotropic ohmic material under conditions where heating and nonlinear effects are negligible,

$$\mathbf{J} = \sigma \mathbf{E}, \quad (21.1)$$

where $\mathbf{J}$ is current density and σ is conductivity. Equivalently,

$$\mathbf{E} = \rho \mathbf{J}, \quad \rho = \frac{1}{\sigma}. \quad (21.2)$$

For a uniform wire of length L and cross-sectional area A , Eq. (21.1) leads to

$$R = \rho \frac{L}{A}, \quad (21.3)$$

provided the material properties and cross section are uniform and contact effects can be neglected.

The microscopic current density associated with one carrier species of number density n_c , charge q and average drift velocity $\mathbf{v}_d$ is

$$\mathbf{J} = n_c q \mathbf{v}_d. \quad (21.4)$$

For electrons, $q = -e$. Their drift velocity is opposite the conventional-current direction, so the product $q\mathbf{v}_d$ points along the conventional current.

In many materials and field ranges, drift speed is approximately proportional to electric field. Mobility is defined by the magnitude relation

$$v_d = \mu E. \quad (21.5)$$

For semiconductor electrons and holes, a convenient vector convention is

$$\mathbf{v}_n = -\mu_n \mathbf{E}, \quad (21.6)$$

$$\mathbf{v}_p = +\mu_p \mathbf{E}. \quad (21.7)$$

The corresponding current contributions both point in the direction of $\mathbf{E}$:

$$\mathbf{J} = e(n\mu_n + p\mu_p) \mathbf{E}. \quad (21.8)$$

Thus a simple two-carrier conductivity model is

$$\sigma = e(n\mu_n + p\mu_p). \quad (21.9)$$

Key message

Conductivity can change because the *carrier density* changes, because the *mobility* changes, or because both change. This is why the temperature dependence of a semiconductor can differ qualitatively from that of a metal.

The unit of conductivity follows directly: $[\sigma] = [J]/[E] = \text{A}/(\text{V m}) = \text{S m}^{-1}$. Likewise $[en\mu] = \text{C m}^{-3} \text{m}^2/(\text{V s})$ has the same unit. These checks help detect the common conversion error $1 \text{ cm}^{-3} = 10^6 \text{ m}^{-3}$.

For fixed temperature, the local rate of conversion of electrical work into heat is

$$p_{\text{heat}} = \mathbf{J} \cdot \mathbf{E} = \sigma E^2 = \rho J^2 \quad [\text{W m}^{-3}]. \quad (21.10)$$

This is a power density, not the hole concentration p . A finite steady drift therefore requires an energy supply even though the average drift kinetic energy stays constant: scattering transfers work to other degrees of freedom. In an anisotropic conductor the scalar formula becomes $\mathbf{E} \cdot \boldsymbol{\sigma} \mathbf{E}$, with a positive dissipative part.

21.3.1 The Drude estimate and what it teaches

The Drude model treats conduction electrons classically between scattering events. Suppose an electron experiences an electric field $\mathbf{E}$. Between collisions, Newton's second law gives

$$m \frac{d\mathbf{v}}{dt} = -e\mathbf{E}. \quad (21.11)$$

In the relaxation-time model, the averaged drift obeys

$$m \frac{d\mathbf{v}_d}{dt} = -e\mathbf{E} - \frac{m\mathbf{v}_d}{\tau}. \quad (21.12)$$

The final term removes directed momentum; random equilibrium motion has zero mean and is not represented by $\mathbf{v}_d$. For a constant field switched on at $t = 0$ with zero initial drift,

$$\mathbf{v}_d(t) = -\frac{e\tau}{m} \mathbf{E} (1 - e^{-t/\tau}). \quad (21.13)$$

After several relaxation times the steady field-induced drift is

$$\mathbf{v}_d = -\frac{e\tau}{m} \mathbf{E}. \quad (21.14)$$

Using $\mathbf{J} = -ne\mathbf{v}_d$ gives

$$\boxed{\sigma = \frac{ne^2\tau}{m}}. \quad (21.15)$$

The model also suggests a mobility

$$\mu = \frac{e\tau}{m}. \quad (21.16)$$

These formulas are valuable because they separate the role of carrier density from the role of scattering. A metal can have a huge electron density but still possess finite resistance because τ is finite. Raising temperature usually increases lattice vibration and reduces the momentum-relaxation time, which tends to reduce metallic conductivity.

The Drude model is not a complete theory of metals. It uses classical particle dynamics and ignores the Fermi–Dirac distribution, band structure and detailed scattering mechanisms. Surprisingly, several transport relations survive in modified form when the quantum picture is used, which is why Drude reasoning remains pedagogically useful.

Example 21.1: Estimating a Drude relaxation time

A metal has conductivity $\sigma = 5.8 \times 10^7 \text{ S m}^{-1}$ and conduction-electron density $n = 8.5 \times 10^{28} \text{ m}^{-3}$. Estimate the Drude relaxation time using the free-electron mass $m_e = 9.11 \times 10^{-31} \text{ kg}$.

Strategy. Rearrange Eq. (21.15):

$$\tau = \frac{m_e \sigma}{n e^2}.$$

Substitution gives

$$\begin{aligned} \tau &= \frac{(9.11 \times 10^{-31})(5.8 \times 10^7)}{(8.5 \times 10^{28})(1.602 \times 10^{-19})^2} \\ &\approx 2.4 \times 10^{-14} \text{ s}. \end{aligned}$$

Interpretation. The time is extremely short on human scales, yet an electron can undergo an enormous number of microscopic scattering events during one second of steady current. The result is an effective transport parameter, not a claim that every electron experiences identical periodically spaced collisions.

Dimensional check. Since σ has units $\text{A}/(\text{V m})$ and $e^2 n/m$ carries the complementary units, the result reduces to seconds as required.

Example 21.2: Mobility, conductivity and resistivity

A sample contains one dominant carrier species with density $n = 2.0 \times 10^{22} \text{ m}^{-3}$ and mobility $\mu = 0.14 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$. Find its conductivity and resistivity.

For one carrier species,

$$\sigma = n e \mu.$$

Hence

$$\begin{aligned} \sigma &= (2.0 \times 10^{22})(1.602 \times 10^{-19})(0.14) \\ &\approx 4.49 \times 10^2 \text{ S m}^{-1}, \end{aligned}$$

and

$$\rho = \frac{1}{\sigma} \approx 2.23 \times 10^{-3} \Omega \text{ m}.$$

This is far less conductive than a good metal but vastly more conductive than a high-quality insulator. The example illustrates why “semiconductor” is not a single fixed resistivity category: carrier concentration can be engineered over a wide range.

Method

For any conductivity problem, write the transport chain explicitly:

$$\text{field } \mathbf{E} \quad \longrightarrow \quad \text{drift velocity } \mathbf{v}_d \quad \longrightarrow \quad \text{current density } \mathbf{J}.$$

Then ask whether the field changes mobility, carrier density, or both. In ordinary metals at moderate fields, n is nearly fixed and scattering controls most of the temperature dependence. In semiconductors, n and p may change exponentially with temperature.

Guided checks

1. If the carrier density doubles while mobility stays fixed, what happens to σ ?
2. If scattering becomes more frequent so that τ decreases, what does the Drude model predict for ρ ?
3. Electrons drift opposite $\mathbf{E}$. Why does their conventional-current contribution point along $\mathbf{E}$?
4. Why should Eq. (21.1) be described as a material relation valid in a regime, rather than as a universal law of nature?

21.3.2 Mean free path, scattering and metallic resistance

The Drude relaxation time τ can be converted into a characteristic distance between momentum-randomizing events. If v is a representative carrier speed, the mean free path is roughly

$$\ell \sim v\tau. \quad (21.17)$$

For a degenerate metal, v_F is often the relevant microscopic speed scale. The conductivity can then be written schematically as

$$\sigma \sim \frac{ne^2\ell}{mv_F}. \quad (21.18)$$

This makes the geometrical meaning of increased scattering more visible: a shorter mean free path lowers conductivity.

Several mechanisms can contribute to momentum relaxation. Thermal lattice vibrations increase with temperature and generally strengthen electron–phonon scattering. Static defects and impurities provide a temperature-independent or weakly temperature-dependent background. In a simplified regime one sometimes summarizes independent scattering channels using Matthiessen’s rule,

$$\rho(T) \approx \rho_0 + \rho_{\text{ph}}(T), \quad (21.19)$$

where ρ_0 is the residual resistivity from static disorder and ρ_{ph} is the phonon contribution. This additive form is an approximation, not a fundamental identity.

Example 21.3: Mean free path from a relaxation time

An idealized metal has Fermi speed $v_F = 1.5 \times 10^6 \text{ m s}^{-1}$ and momentum-relaxation time $\tau = 2.0 \times 10^{-14} \text{ s}$. The mean free path estimate is

$$\ell \sim v_F\tau = (1.5 \times 10^6)(2.0 \times 10^{-14}) = 3.0 \times 10^{-8} \text{ m} = 30 \text{ nm}.$$

A device much longer than this scale is naturally described as diffusive. If a nanoscale conductor becomes comparable to or shorter than relevant mean free paths, ballistic or quasi-ballistic descriptions become increasingly important.

21.3.3 Why a perfect periodic lattice does not itself cause resistance

A common misconception is that electrons lose momentum simply because they continually “collide with atoms” in a crystal. In quantum mechanics, a perfectly periodic lattice is incorporated into the allowed crystal states themselves. A Bloch electron can propagate through a perfect periodic potential without being randomly scattered by every lattice site. Resistance arises from departures from perfect translational order and from interactions such as lattice vibrations, impurities, defects, boundaries and other excitations.

Caution

The classical picture of electrons as tiny balls ricocheting from immobile atoms is useful only as a rough mnemonic for a relaxation time. It should not be taken as a literal microscopic model of a perfect crystal.

21.4 Metals, insulators and semiconductors in the band picture

The band picture makes the broad electrical classification of crystalline solids conceptually simple. The categories are not defined by one universally sharp numerical resistivity boundary. Instead, they describe characteristic band structures and carrier populations.

21.4.1 Metals

A metal has accessible unoccupied states extremely close in energy to occupied states. This can occur because a band is only partially filled or because two bands overlap. An applied electric field can then alter the occupancy near the highest occupied energies and create a net current without requiring electrons to cross a large energy gap.

At ordinary temperatures the number of mobile electrons in a typical simple metal is enormous, often comparable to one or a few electrons per atom. The temperature dependence of resistivity is therefore dominated not by a large change in carrier number but by a change in scattering. As lattice vibrations grow with temperature, the average momentum-relaxation time generally decreases and resistivity generally rises over an ordinary temperature range.

21.4.2 Insulators

In an idealized insulator at low or moderate temperature, the highest occupied band is essentially full and the next available band is separated by a comparatively large energy gap. Thermal energy and ordinary electric fields are insufficient to promote a significant carrier population into the conduction band. The carrier density is therefore extremely small, so the conductivity is correspondingly tiny.

Real insulators are not perfectly nonconducting under all conditions. Strong electric fields can cause breakdown; defects and impurities can introduce states inside the gap; radiation can create carriers; surfaces can support leakage currents; and conductivity can become appreciable at sufficiently high temperature. The label “insulator” therefore describes a regime and material class, not an absolute prohibition on charge motion.

21.4.3 Semiconductors

A semiconductor resembles an insulator in having a valence band separated from a conduction band by a gap, but the gap is small enough that thermal excitation, optical excitation or controlled doping can create useful carrier populations. This sensitivity is the basis of semiconductor technology.

A crucial qualitative distinction follows. Increasing temperature usually increases lattice scattering, which by itself would reduce mobility. But in an intrinsic semiconductor, increasing temperature can create vastly more electron–hole pairs. The increase in carrier density can dominate the reduction in mobility, so the net conductivity rises strongly with temperature. That behaviour contrasts with the ordinary metallic trend.

Compare the nearby available states and gap sizes in Fig. 21.3.

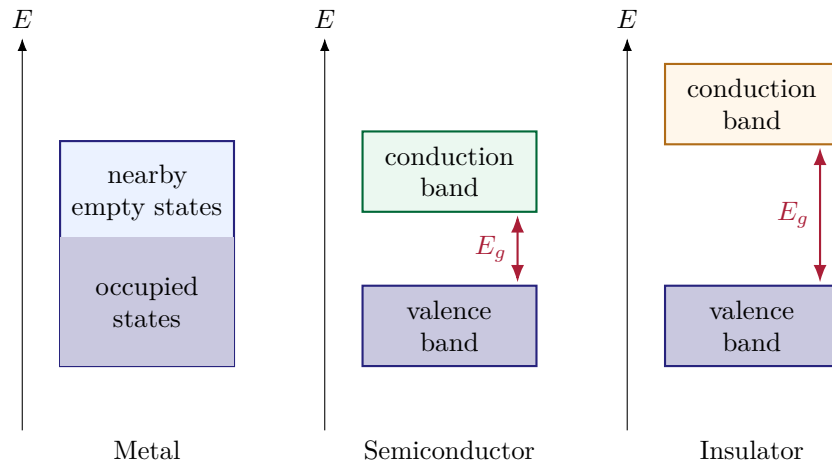


Figure 21.3: Idealized band pictures for a metal, semiconductor and insulator. The key distinctions are the availability of nearby states and the size of the gap separating occupied and unoccupied bands.

Example 21.4: Classifying idealized band diagrams

Three hypothetical materials have the following zero-temperature descriptions.

1. Material A has a highest band that is half filled.
2. Material B has a full valence band, an empty conduction band and $E_g = 1.1 \text{ eV}$.
3. Material C has a full valence band and an empty conduction band separated by $E_g = 6.0 \text{ eV}$.

Classify them qualitatively.

Solution. Material A is metallic because empty states exist arbitrarily close in energy within the same band. Material B is a plausible semiconductor: a gap of order an electronvolt allows thermally or optically generated carriers under realistic conditions. Material C is a strong insulator in ordinary conditions because the gap is much larger. The numerical boundary is not universal, so the classification is based on the physical scale and expected carrier population rather than on an exact gap threshold.

Caution

Do not define “metal” simply as “small resistance” and “insulator” simply as “large resistance.” Resistance depends on geometry. The material quantities are conductivity and resistivity, and the deeper microscopic distinction is the electronic state structure and carrier population.

Guided checks

1. Why can a partially filled band conduct even if no electron is promoted into a separate higher band?
2. Why can an insulator become conducting under a sufficiently strong field or intense radiation without ceasing to be an insulating material in its normal operating regime?
3. A semiconductor becomes more conductive when heated. Does this prove that its carriers scatter less at higher temperature? Explain.
4. Which is more directly connected to band structure: sample resistance or material conductivity?

21.5 Intrinsic semiconductors: electrons and holes

An *intrinsic semiconductor* is an idealized chemically pure semiconductor in which the mobile electrons and holes are generated by the material itself rather than intentionally introduced dopants. At absolute zero in the simplest band picture, the valence band is full and the conduction band is empty. At finite temperature, some electrons acquire enough energy to cross the band gap. Each promoted electron leaves behind an unoccupied valence-band state: a hole.

The creation process therefore produces carriers in pairs:

$$\text{valence electron} \longrightarrow \text{conduction electron} + \text{hole}. \quad (21.20)$$

In an ideal intrinsic material at equilibrium,

$$n = p = n_i, \quad (21.21)$$

where n_i is the intrinsic carrier concentration. The exact temperature dependence of n_i requires semiconductor statistics and will be developed in the advanced enrichment section. The most important qualitative result is that n_i grows rapidly with temperature, approximately with an exponential factor involving $E_g/(2k_B T)$.

21.5.1 Why a hole behaves like a positive carrier

Imagine a nearly full row of seats with one empty seat. If a person next to the empty seat moves into it, the empty seat appears to move in the opposite direction. A valence band behaves similarly, although the real quantum dynamics is more subtle than the seat analogy. When an electron moves from one valence state into a neighbouring empty state, the vacancy can be tracked as a quasiparticle with positive charge $+e$ and an effective mass determined by the band structure.

This description is not merely linguistic convenience. In semiconductor transport equations, a hole responds to electric and magnetic fields as though it were a mobile positive carrier. The total current density can therefore be written as the sum of electron and hole contributions, leading to Eq. (21.9).

Figure 21.4 tracks the electron and the vacancy created by one excitation event.

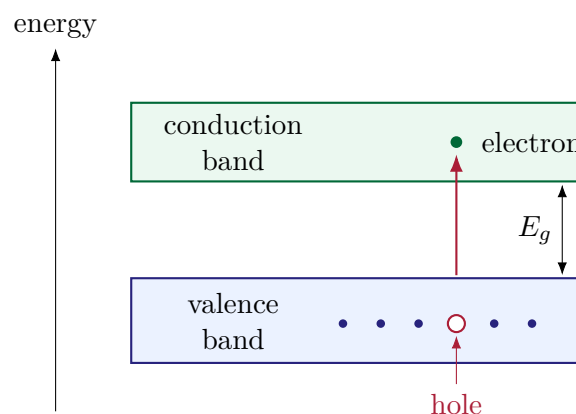


Figure 21.4: Thermal or optical excitation of a valence electron into the conduction band creates an electron–hole pair. The hole is the missing valence electron state, represented as a positive mobile carrier in transport theory.

21.5.2 Generation and recombination

Electron–hole pairs are continually generated and recombined at finite temperature. In equilibrium, the average generation rate equals the average recombination rate, so the carrier concentrations

remain statistically steady. Recombination occurs when a conduction-band electron transitions into an available valence-band state and annihilates a hole. The released energy may appear as a photon, as lattice vibration (phonons), or through more complicated multi-particle processes depending on the material and device.

The balance between generation and recombination becomes crucial in LEDs, photodiodes and solar cells. A device driven away from equilibrium can maintain electron and hole populations larger than their equilibrium values, allowing optical emission or photocurrent.

Example 21.5: Intrinsic conductivity from carrier density and mobility

At a certain temperature an intrinsic semiconductor has $n_i = 1.0 \times 10^{16} \text{ m}^{-3}$, electron mobility $\mu_n = 0.14 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ and hole mobility $\mu_p = 0.05 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$. Estimate the conductivity and resistivity.

Because the material is intrinsic, $n = p = n_i$. Therefore

$$\begin{aligned}\sigma &= en_i(\mu_n + \mu_p) \\ &= (1.602 \times 10^{-19})(1.0 \times 10^{16})(0.14 + 0.05) \\ &\approx 3.04 \times 10^{-4} \text{ S m}^{-1}.\end{aligned}$$

Thus

$$\rho \approx 3.29 \times 10^3 \Omega \text{ m}.$$

Even with mobilities that are not extraordinarily small, the material is weakly conducting because the intrinsic carrier density is tiny compared with a metal's conduction-electron density.

Key message

An intrinsic semiconductor is not electrically inactive. At nonzero temperature it contains a dynamic equilibrium population of electrons and holes. What makes the conductivity modest is the relatively small carrier concentration, not the complete absence of mobile charges.

Guided checks

1. Why does promoting one valence electron create two mobile-carrier descriptions rather than one?
2. If n_i increases by a factor of 100 while mobilities decrease by a factor of 2, by what factor does intrinsic conductivity change approximately?
3. Why does recombination not necessarily produce visible light?
4. In an intrinsic semiconductor, can n increase without p increasing under equilibrium thermal generation? Explain.

21.6 Doping: controlling carrier populations

The technological power of semiconductors comes from the ability to control carrier density deliberately. *Doping* means introducing a small concentration of impurity atoms into a semiconductor crystal so that extra electronic states appear near a band edge. The dopants are chosen to donate electrons to the conduction band or accept electrons from the valence band.

For a group-IV semiconductor such as silicon, a donor impurity may have one more valence electron than the host atom. Four electrons participate in covalent bonding while the additional electron is only weakly bound. Modest thermal energy can ionize the donor and place the extra electron in the

conduction band. The fixed donor ion is then positively charged, but the mobile electron remains in the crystal. The material as a whole remains approximately electrically neutral.

An acceptor impurity has one fewer valence electron than the host. It can accept an electron from the valence band, leaving behind a mobile hole. The ionized acceptor is negatively charged and fixed in the lattice, while the hole can move through the valence band.

21.6.1 n-type material

A semiconductor doped primarily with donors is called *n-type*. Electrons are the majority carriers and holes are the minority carriers. In a simplified regime where donors are fully ionized and intrinsic carriers are negligible compared with the dopant density,

$$n \approx N_D, \quad (21.22)$$

while $p \ll n$. The conductivity is then approximately

$$\sigma \approx eN_D\mu_n. \quad (21.23)$$

21.6.2 p-type material

A semiconductor doped primarily with acceptors is called *p-type*. Holes are the majority carriers and electrons are the minority carriers. Under analogous assumptions,

$$p \approx N_A, \quad (21.24)$$

and

$$\sigma \approx eN_A\mu_p. \quad (21.25)$$

The near-edge donor and acceptor levels in Fig. 21.5 explain why dopants can ionize without exciting an electron across the full band gap.

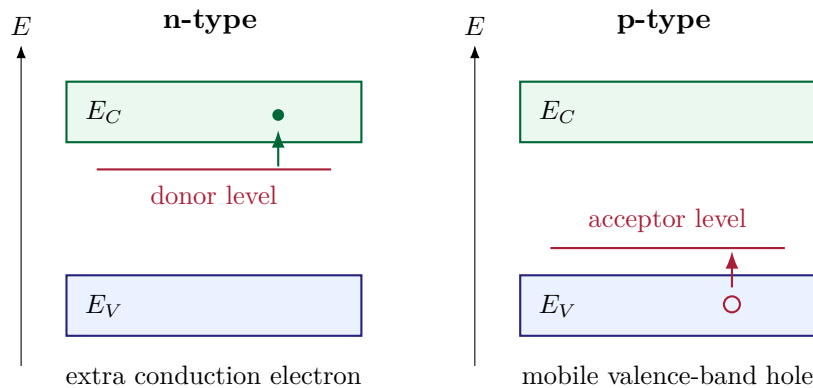


Figure 21.5: Simplified donor and acceptor pictures. Donor levels lie close to the conduction band and can supply electrons; acceptor levels lie close to the valence band and can create holes. The positions are schematic and not to scale.

Caution

Doping does *not* normally make the bulk semiconductor strongly charged. Ionizing a donor produces one fixed positive donor ion and one mobile electron; ionizing an acceptor produces one fixed negative acceptor ion and one mobile hole. Away from junction space-charge regions, these charges almost cancel macroscopically.

Example 21.6: Majority carriers after donor doping

A semiconductor is doped with donors at concentration $N_D = 4.0 \times 10^{21} \text{ m}^{-3}$. Assume the donors are fully ionized, acceptor compensation is negligible, and $N_D \gg n_i$. If $\mu_n = 0.12 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, estimate the conductivity due to majority carriers.

Under the stated assumptions,

$$n \approx N_D = 4.0 \times 10^{21} \text{ m}^{-3}.$$

Then

$$\begin{aligned} \sigma &\approx eN_D\mu_n \\ &= (1.602 \times 10^{-19})(4.0 \times 10^{21})(0.12) \\ &\approx 76.9 \text{ S m}^{-1}. \end{aligned}$$

The corresponding resistivity is

$$\rho \approx 1.30 \times 10^{-2} \Omega \text{ m}.$$

The calculation demonstrates how a dopant concentration far below the atomic density can still create a technologically useful carrier population.

21.6.3 The Fermi level as a chemical-potential marker

At this level, the Fermi level can be used as an energy marker that indicates how electronic states are statistically occupied. In an intrinsic semiconductor it lies near the middle of the gap in a simple symmetric picture. Donor doping shifts the equilibrium Fermi level toward the conduction band, while acceptor doping shifts it toward the valence band. A precise relation between E_F , carrier densities and temperature will be derived later in the advanced section.

The Fermi level is not itself necessarily an allowed electron state. In a semiconductor it may lie inside the forbidden band gap. Its importance is statistical: it sets the occupancy probability of allowed states.

Guided checks

1. A donor atom becomes positively ionized. Why does the crystal not acquire a large positive net charge?
2. Is an n-type semiconductor negatively charged overall? Explain carefully.
3. Why are holes called minority carriers in n-type material even though they can still control important processes?
4. If all donor atoms are ionized and N_D is doubled while mobility stays fixed, what happens approximately to conductivity?
5. Can the Fermi level lie inside an energy gap where there are no allowed one-electron states? Why is this not contradictory?

21.7 Measuring carriers: the Hall effect as a bridge

The Hall effect, introduced earlier in the study of magnetic forces, provides a direct experimental connection between transport equations and carrier type. Suppose a sample carries current density J_x while a magnetic field B_z is applied perpendicular to the current. Mobile carriers are deflected sideways until a transverse electric field E_y develops. In a simple one-carrier model, equilibrium of

transverse electric and magnetic forces gives

$$q(E_y - v_x B_z) = 0, \quad (\mathbf{v} \times \mathbf{B})_y = -v_x B_z. \quad (21.26)$$

Using $J_x = n_c q v_x$, one obtains the Hall coefficient

$$R_H = \frac{E_y}{J_x B_z} = \frac{1}{n_c q}. \quad (21.27)$$

For electron conduction $q = -e$, so the Hall coefficient is negative in this simple convention. For hole-dominated conduction it is positive. Energy-dependent scattering can introduce a Hall factor even in one-band transport, and multiband materials can be more complicated, but the one-carrier result is extremely useful pedagogically and experimentally.

If sample thickness is t and current is I , then $J_x = I/(wt)$ for sample width w , and we define the signed Hall voltage as $V_H = \phi(y = -w/2) - \phi(y = +w/2)$. Since $E_y = -d\phi/dy$, this convention gives

$$V_H = E_y w = R_H \frac{I B_z}{t}. \quad (21.28)$$

Thus a Hall measurement can determine both the sign of the dominant carrier and an estimate of carrier density.

Figure 21.6 fixes the axes and the order of the voltage leads before assigning a Hall sign.

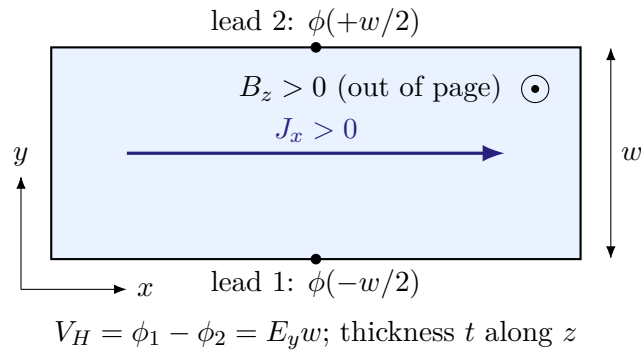


Figure 21.6: Hall measurement geometry and signed voltage convention. With this lead order, electron-only conduction gives $E_y < 0$ and $V_H < 0$, whereas hole-only conduction gives positive values. Swapping voltage leads reverses the displayed sign without changing the carriers.

Example 21.7: Carrier density from a Hall coefficient

A semiconductor sample has measured Hall coefficient $R_H = -3.1 \times 10^{-4} \text{ m}^3 \text{ C}^{-1}$. Use the one-carrier approximation to estimate the dominant carrier type and number density. The negative sign indicates electron-dominated conduction. From Eq. (21.27),

$$n = \frac{1}{e|R_H|} = \frac{1}{(1.602 \times 10^{-19})(3.1 \times 10^{-4})} \approx 2.0 \times 10^{22} \text{ m}^{-3}.$$

The result should be interpreted as an effective one-carrier density; if both electrons and holes contribute significantly, a more complete two-carrier Hall model is needed.

Key message

The Hall effect converts an abstract band-model statement – “electrons or holes are the dominant carriers” – into a laboratory measurement of a transverse voltage. This is a good example of how quantum-material concepts connect to ordinary voltmeters and magnets.

21.7.1 Separating sample and contact resistance

A resistance measurement is never made on an isolated abstract material. Wires, electrodes and interfaces connect the sample to instruments. If a two-terminal measurement applies current I and measures voltage V across the same contacts, the measured resistance can contain

$$R_{2T} = R_{\text{contact},1} + R_{\text{sample}} + R_{\text{contact},2} + R_{\text{leads}}. \quad (21.29)$$

For a low-resistance sample, contact contributions may dominate.

A four-terminal or Kelvin measurement separates current injection from voltage sensing. Two outer contacts carry current while two inner high-impedance voltage probes measure the potential drop. Because negligible current flows through the voltage leads, their contact voltage drops are greatly reduced in the measured result.

Figure 21.7 separates the current contacts from the two high-impedance voltage probes.

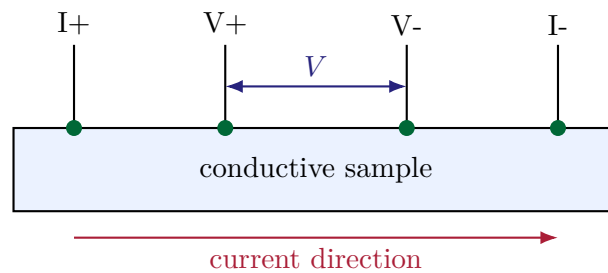


Figure 21.7: Conceptual four-terminal resistance measurement. Current and voltage contacts are separated so that contact and lead resistances have much less influence on the measured sample voltage.

Example 21.8: Resistivity from a four-terminal measurement

A uniform bar has voltage-probe separation $L = 8.0$ mm, width $w = 1.5$ mm and thickness $t = 0.40$ mm. A current $I = 5.0$ mA produces a measured voltage $V = 0.120$ V between the inner probes.

The cross-sectional area is

$$A = wt = (1.5 \times 10^{-3})(0.40 \times 10^{-3}) = 6.0 \times 10^{-7} \text{ m}^2.$$

Using $\rho = VA/(IL)$,

$$\rho = \frac{(0.120)(6.0 \times 10^{-7})}{(5.0 \times 10^{-3})(8.0 \times 10^{-3})} \approx 1.8 \times 10^{-3} \text{ } \Omega \text{ m}.$$

The result refers to the assumed uniform sample geometry. For thin films, irregular shapes or anisotropic materials, specialized geometrical analysis may be required.

21.8 The p–n junction: diffusion, drift and the depletion region

A p–n junction is formed within a single semiconductor crystal by creating neighbouring p-type and n-type regions. The boundary need not be atomically abrupt, but an abrupt-junction model is an excellent first approximation. The essential physics begins before any external voltage is applied.

On the n side, electrons are majority carriers. On the p side, holes are majority carriers. Because carrier concentrations vary sharply across the junction, electrons diffuse from the n side toward the p side and holes diffuse from the p side toward the n side. Near the junction they recombine with opposite carriers.

This diffusion leaves behind fixed ionized dopants: positively charged donor ions on the n side and negatively charged acceptor ions on the p side. These ions cannot move with ordinary current because they are part of the crystal lattice. Their uncovered charge forms a *space-charge region*, commonly called the *depletion region* because it is depleted of mobile majority carriers compared with the neutral bulk regions.

As illustrated in Fig. 21.8, the fixed charge creates an internal electric field directed from the positive donor side toward the negative acceptor side, that is, from n to p in the conventional abrupt-junction sketch. That field drives an electron drift contribution toward n and a hole drift contribution toward p, opposing the respective diffusion contributions. Each species has zero net current in thermal equilibrium; drift is not restricted to a special class of carriers. Thermal equilibrium is reached when the net current vanishes: diffusion and drift contributions balance.

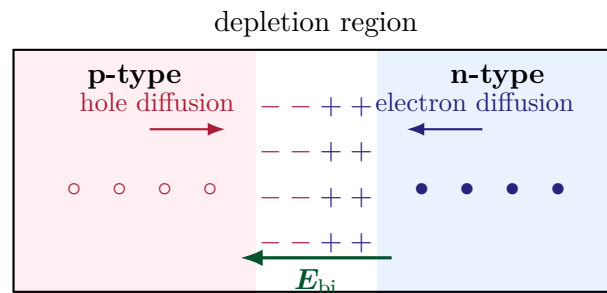


Figure 21.8: Qualitative abrupt p–n junction at equilibrium. Majority-carrier diffusion exposes fixed ionized dopants and creates a depletion region. The resulting built-in electric field opposes further diffusion. Open circles indicate holes and solid dots indicate electrons.

21.8.1 Built-in potential and equilibrium

The internal electric field corresponds to a built-in electrostatic potential difference V_{bi} across the depletion region. Majority carriers attempting to diffuse farther across the junction must effectively overcome an energy barrier associated with this potential. The exact value depends on doping concentrations, temperature and material parameters.

It is important to distinguish the built-in potential inside the junction from a directly measurable terminal voltage of an isolated equilibrium device. Contact potentials at external metal–semiconductor interfaces and equilibrium electrochemical-potential conditions prevent a simple voltmeter from reading V_{bi} as an open-circuit battery voltage. The built-in potential is nevertheless physically real and crucial to junction behaviour.

Caution

The depletion region is not empty space. It contains the semiconductor lattice and fixed ionized dopants. It is called “depleted” because the concentration of mobile majority carriers is strongly reduced there.

Guided checks

1. What physical gradient initially drives electrons and holes across a newly formed p–n junction?
2. Why does diffusion leave a fixed space charge near the junction?
3. In which direction does the built-in electric field point in the conventional p-left, n-right sketch of Fig. 21.8?
4. Why does equilibrium not require diffusion to stop completely at the microscopic level?

5. Why is the depletion region not literally devoid of atoms or charge?

21.9 Biasing the junction and rectification

An external voltage changes the barrier experienced by carriers at a p–n junction. In *forward bias*, the p side is connected to a higher electric potential than the n side. The external field partly opposes the built-in field, lowers the effective barrier and narrows the depletion region. Majority carriers can then cross the junction much more readily, producing a large current.

In *reverse bias*, the p side is connected to a lower electric potential than the n side. The applied field reinforces the built-in field, increases the effective barrier and widens the depletion region. Majority-carrier current becomes very small. A reverse current still exists because thermally generated minority carriers can be swept across the junction. At sufficiently large reverse voltage, breakdown mechanisms can produce a large current; that regime is not described by the simplest ideal diode equation.

This strong directional asymmetry is *rectification*. A junction diode can therefore convert an alternating-polarity source into a current that preferentially flows in one direction.

21.9.1 The ideal diode equation

A widely used idealized current–voltage relation is

$$I = I_0 \left[\exp \left(\frac{eV}{k_B T} \right) - 1 \right] = I_0 \left[\exp \left(\frac{V}{V_T} \right) - 1 \right], \quad (21.30)$$

where V is the applied junction voltage in the forward-positive convention and I_0 is a reverse-saturation current scale. In real diodes a more flexible empirical form often replaces V_T by $n_d V_T$, where n_d is an ideality factor typically of order unity. Series resistance, leakage, generation–recombination processes, high-level injection and breakdown all create deviations.

For forward voltages several thermal voltages above zero, the “ -1 ” term becomes negligible and

$$I \approx I_0 \exp \left(\frac{V}{V_T} \right). \quad (21.31)$$

The exponential sensitivity explains why a modest change in junction voltage can produce a large change in current.

The normalized plot in Fig. 21.9 shows both the forward exponential and the much smaller reverse saturation current.

Example 21.9: Current ratio for a small voltage increase

At $T = 300$ K, an ideal diode operates in the forward regime where the “ -1 ” term is negligible. By what factor does the current change if the forward voltage increases by 60 mV?

Using $I \propto \exp(V/V_T)$ and $V_T \approx 25.9$ mV,

$$\frac{I_2}{I_1} = \exp \left(\frac{\Delta V}{V_T} \right) = \exp \left(\frac{60}{25.9} \right) \approx 10.1.$$

Thus a voltage increase of only about 60 mV changes the ideal current by roughly one order of magnitude. This is why diode circuits require attention to temperature and current limiting.

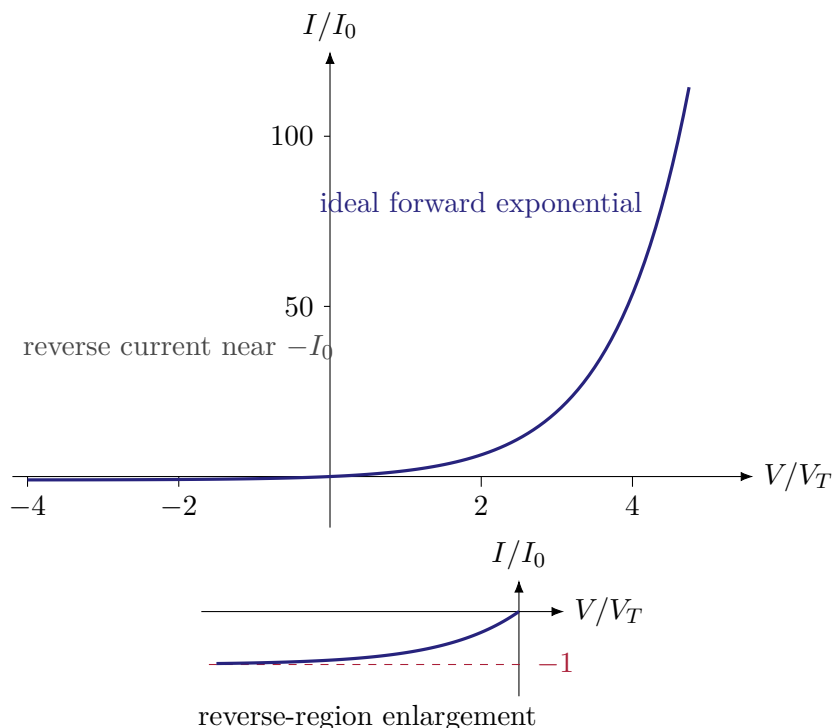


Figure 21.9: Calculated normalized ideal diode law $I/I_0 = \exp(V/V_T) - 1$. An enlarged reverse-current panel makes the nonzero saturation scale visible; it would be almost invisible on the forward-current scale. These plots exclude breakdown and series resistance.

Example 21.10: Finding a forward voltage in the ideal model

An ideal diode at 300 K has $I_0 = 2.0 \text{ nA}$. Estimate the forward voltage needed for $I = 2.0 \text{ mA}$. From Eq. (21.30), because $I \gg I_0$,

$$V \approx V_T \ln \left(\frac{I}{I_0} \right).$$

The current ratio is 10^6 , so

$$\begin{aligned} V &\approx (25.9 \text{ mV}) \ln(10^6) \\ &= (25.9 \text{ mV})(13.82) \\ &\approx 0.358 \text{ V}. \end{aligned}$$

This value is a consequence of the chosen ideal saturation current. It should not be interpreted as a universal “turn-on voltage” for all diodes.

Caution

A diode does not possess a perfectly sharp universal threshold voltage. Statements such as “a silicon diode turns on at 0.7 V” are circuit-level approximations tied to a particular current range, device, temperature and model. The microscopic junction current varies continuously with voltage.

Method

When interpreting a diode I – V curve, use four regimes: near-zero bias, forward conduction, reverse leakage/saturation, and reverse breakdown. Ask which physical mechanism dominates in each regime before selecting an equation. Never apply the ideal exponential law blindly across breakdown or high-current series-resistance regions.

21.10 The junction diode in circuits

At circuit level a diode is often approximated by progressively more realistic models. The *ideal-switch model* treats the forward-biased diode as a short circuit and the reverse-biased diode as an open circuit. A *constant-drop model* assigns a fixed forward voltage over a limited current range. The *exponential model* uses Eq. (21.30), possibly with ideality factor and series resistance. The correct model depends on the required accuracy.

A basic half-wave rectifier uses one diode in series with a load. During the half-cycle that forward-biases the diode, current flows through the load. During the opposite half-cycle the diode blocks most current. A smoothing capacitor placed across the load charges near waveform peaks and discharges through the load between peaks, reducing output ripple.

Trace the charging path and the parallel load–capacitor connection in Fig. 21.10.

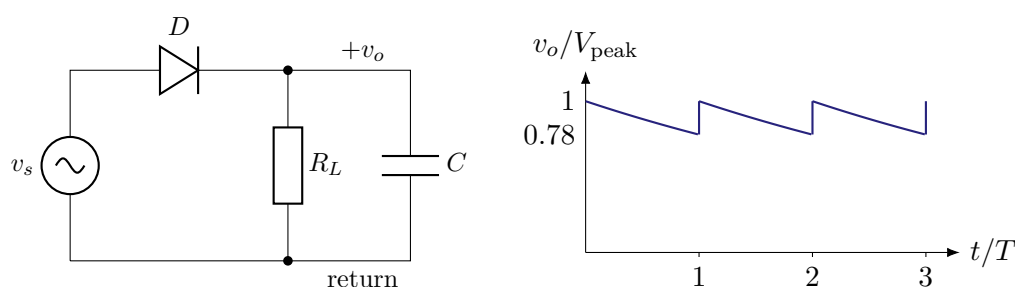


Figure 21.10: Connected half-wave rectifier: the load and smoothing capacitor are in parallel, with no wire bypassing either component. The idealized discharge plot uses $v_o/V_{\text{peak}} = \exp[-(t - t_{\text{peak}})/(R_L C)]$ between instantaneous peak recharge events, with $R_L C = 4T$. Real charging has finite duration and source impedance.

For a rough ripple estimate, if the load current is approximately constant between charging peaks,

$$\Delta V \approx \frac{I_L \Delta t}{C}. \quad (21.32)$$

This expression is an engineering approximation, not a detailed solution of the diode-capacitor circuit. It is useful for order-of-magnitude design.

Example 21.11: Sizing a smoothing capacitor

A half-wave rectifier is driven by a 50 Hz source and supplies an approximately constant load current of $I_L = 20$ mA. Estimate the capacitance required to limit the peak-to-peak droop between charging peaks to about 1.0 V.

For half-wave rectification, charging peaks are separated by approximately one source period,

$$\Delta t \approx \frac{1}{50} = 0.020 \text{ s}.$$

Using Eq. (21.32),

$$C \approx \frac{I_L \Delta t}{\Delta V} = \frac{(0.020)(0.020)}{1.0} = 4.0 \times 10^{-4} \text{ F} = 400 \mu\text{F}.$$

A real design must also check diode current, transformer/source resistance, capacitor tolerance, ripple-current rating and load variation.

21.10.1 Full-wave rectification and protection

A single diode can form a half-wave rectifier, but practical dc power supplies often use full-wave rectification so that both halves of an ac cycle contribute to the output. A bridge rectifier uses four diodes arranged so that current through the load has the same direction during both source polarities.

Trace the two conducting diode pairs in Fig. 21.11; each pair sends conventional current through the load in the same direction.

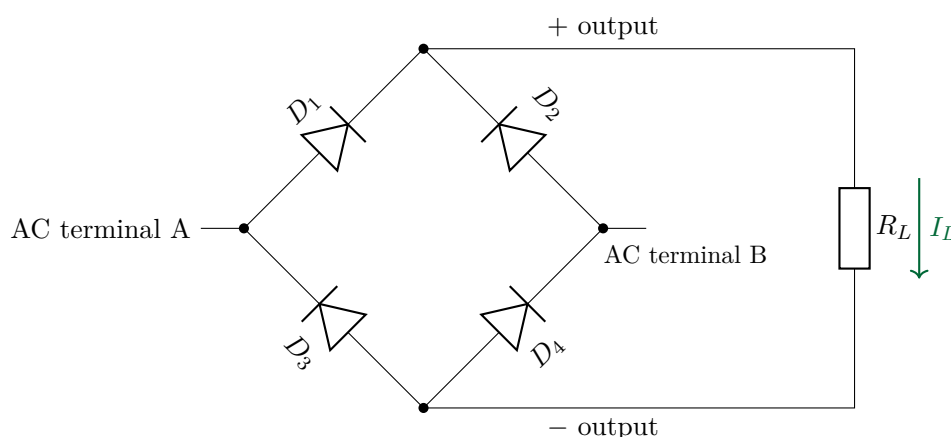


Figure 21.11: Four explicitly oriented diodes form a full-wave bridge. When A is positive relative to B, D_1 and D_4 conduct; when B is positive relative to A, D_2 and D_3 conduct. In both cases conventional load current runs from the positive output through R_L to the negative output. The AC source connects between A and B.

For a sinusoidal source of frequency f , full-wave rectification produces a ripple component with repetition frequency approximately $2f$. A smoothing capacitor can therefore discharge for a shorter interval between charging peaks than in a half-wave rectifier. A rough small-ripple estimate becomes

$$\Delta V \sim \frac{I_L}{2fC} \quad (21.33)$$

for a full-wave rectifier under approximately constant load current. This estimate neglects diode conduction angle, transformer impedance, source resistance and load variation.

Example 21.12: Half-wave versus full-wave smoothing

A 50 Hz source supplies a load drawing approximately $I_L = 40 \text{ mA}$. Estimate the capacitor required to keep ripple near $\Delta V = 0.50 \text{ V}$.

For half-wave rectification, the discharge interval is about $1/f$, so

$$C_{\text{half}} \sim \frac{I_L}{f\Delta V} = \frac{0.040}{(50)(0.50)} \approx 1.6 \text{ mF}.$$

For full-wave rectification,

$$C_{\text{full}} \sim \frac{I_L}{2f\Delta V} \approx 0.80 \text{ mF}.$$

The full-wave circuit can achieve the same rough ripple with about half the capacitance, although two bridge diodes are normally in the current path at a time and therefore contribute additional voltage loss and power dissipation.

Clipping, clamping and protection

A diode can also limit signal excursions. In a simple clipping circuit, the diode becomes strongly conducting when a node exceeds a chosen polarity and approximate threshold, diverting current through a resistor. This idea appears in input protection, over-voltage limiting and waveform shaping. Real protection design must respect maximum current, reverse voltage, surge energy, capacitance and switching speed.

21.11 Light-emitting diodes and radiative recombination

A forward-biased p–n junction can inject electrons and holes into the same region. When they recombine, the released energy is of order the band gap. In a semiconductor with efficient radiative recombination, a substantial fraction of that energy can emerge as photons. Such a junction is a light-emitting diode (LED).

The simplest photon-energy estimate is

$$E_\gamma = hf \approx E_g, \quad (21.34)$$

so

$$\lambda \approx \frac{hc}{E_g}. \quad (21.35)$$

Combining $[h] = \text{J s}$ with $[c] = \text{m s}^{-1}$ and $[E_g] = \text{J}$ gives a length, as required. If E_g is expressed in electronvolts and λ in nanometres, the convenient conversion $hc \approx 1240 \text{ eV nm}$ gives

$$\lambda(\text{nm}) \approx \frac{1240}{E_g(\text{eV})}. \quad (21.36)$$

The relation is approximate. Actual emission spectra have finite width; photons may be reabsorbed; contacts and packaging modify output; and the electronic band structure determines whether electron–hole recombination is direct and optically efficient.

21.11.1 Direct and indirect band gaps

Figure 21.12 adds the wave-vector information hidden by a vertical band cartoon.

In a direct-gap semiconductor, the conduction-band minimum and valence-band maximum occur at compatible crystal momentum, so an electron and hole can recombine by emitting a photon without requiring a large momentum transfer. In an indirect-gap semiconductor, a phonon is typically needed to conserve crystal momentum, making radiative recombination much less probable. This is why not every useful electronic semiconductor is an efficient LED material.

Figure 21.13 shows the energy released when a conduction electron fills a valence-band hole.

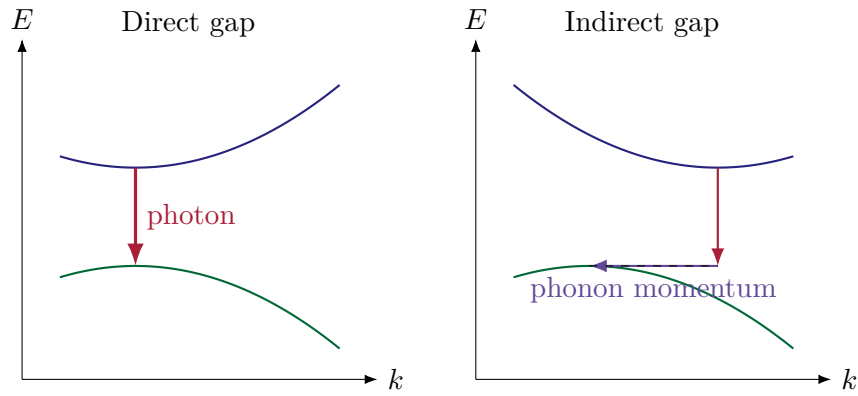


Figure 21.12: Schematic dispersions with equal minimum gaps. In a direct transition the photon carries negligible wave-vector change on this scale. An indirect transition also requires a phonon; the two arrows indicate energy/momentum bookkeeping, not a real intermediate state or a unique temporal sequence.

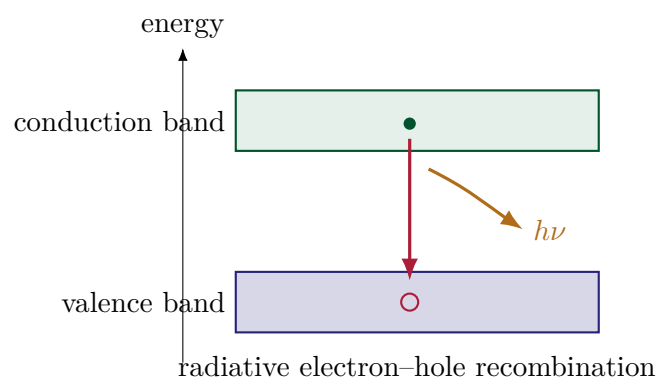


Figure 21.13: Simplified LED energy picture. Forward bias injects carriers; radiative recombination can release a photon with energy of order the band gap.

Example 21.13: Estimating LED wavelength from band gap

A direct-gap semiconductor has $E_g = 2.10$ eV. Estimate the characteristic emitted wavelength. Using Eq. (21.36),

$$\lambda \approx \frac{1240}{2.10} \text{ nm} \approx 590 \text{ nm}.$$

This wavelength lies in the visible region. The estimate identifies a characteristic optical scale, not an exact monochromatic line.

Caution

The formula $E_g \approx hc/\lambda$ is a photon-energy estimate, not a guarantee that every semiconductor with that band gap will be an efficient light emitter. The direct or indirect nature of the gap and nonradiative recombination pathways matter critically.

21.12 Photodiodes and solar cells: running the junction optically

The same junction physics can be used in reverse. A photon with sufficient energy can create an electron–hole pair. If generation occurs in or near a depletion region, the built-in electric field separates the carriers before they recombine. The separated carriers contribute to a measurable photocurrent.

A photodiode is commonly operated under reverse bias because the enlarged depletion region and strong electric field provide fast carrier collection and low junction capacitance. The measured current then contains a dark-current contribution plus a light-dependent photocurrent.

A photovoltaic solar cell uses a related effect but is operated to deliver electrical power to a load. Illumination shifts the diode I – V characteristic so that the device can operate in a quadrant where it supplies both current and voltage. Energy is not created by the junction: incident electromagnetic radiation supplies the input energy, and the device converts part of that optical energy into electrical energy.

Example 21.14: Simple photovoltaic power estimate

A small photovoltaic device, receiving 40 mW of optical power, delivers 0.45 V at 18 mA near its maximum-power operating point under a given illumination. Estimate the electrical output power.

$$P = VI = (0.45)(18 \times 10^{-3}) \approx 8.1 \times 10^{-3} \text{ W} = 8.1 \text{ mW}.$$

The specified incident power gives $\eta = P_{\text{out}}/P_{\text{in}} = 8.1/40 \approx 20\%$. Without the incident power, the electrical operating point alone would not determine efficiency. This comparison uses actual power at the operating point, not the product of the separate open-circuit voltage and short-circuit current.

Key message

LEDs and photodiodes illustrate microscopic reversibility of the same band picture: electrical injection can produce optical emission, while optical absorption can create mobile carriers and electrical current. Device efficiency, however, depends on many nonideal processes in addition to the basic energy bookkeeping.

21.13 Transistors: using a third terminal to control current

A diode has two terminals and primarily controls the direction and magnitude of current through one junction. A transistor introduces a third control terminal. The central engineering idea is that an input signal at a control terminal modulates current through a separately biased path. Appropriate bias and load conditions can turn this modulation into signal gain. This enables switching, amplification, digital logic and analog signal processing. The larger output energy comes from a bias supply; a transistor controls energy flow and does not generate energy from the input signal.

Two broad transistor families are common. A bipolar junction transistor (BJT) uses two nearby p–n junctions and relies on minority-carrier injection and transport through a thin base region. A field-effect transistor (FET) uses an electric field to control the carrier population and conductivity of a channel. Modern digital integrated circuits are dominated by metal–oxide–semiconductor field-effect transistors (MOSFETs), discussed more deeply in the advanced section.

At an introductory level, the transistor should not be treated as a mysterious ideal switch. Its operation follows the same chapter principles: energy bands determine available states, doping sets equilibrium carrier populations, electric fields change barriers or channel charge, and transport equations determine current.

Figure 21.14 locates the insulated gate, inversion channel and four physical terminals.

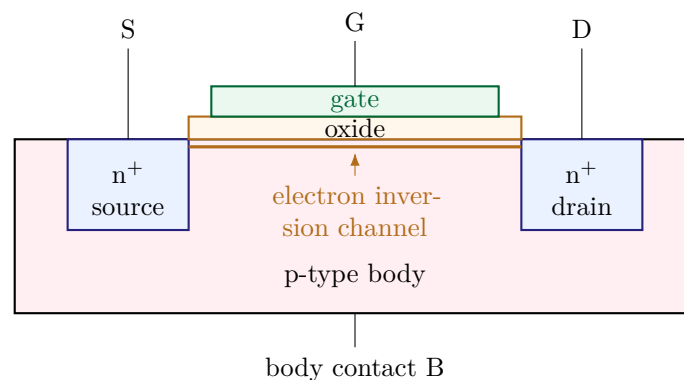


Figure 21.14: Schematic enhancement nMOS cross-section in the conducting state. Source and drain meet the semiconductor surface, oxide lies above the surface, and the inversion channel lies immediately below it. The body is a fourth physical terminal, often tied to the source in a three-terminal circuit model. Dimensions are not to scale.

Guided checks

1. What feature distinguishes a transistor conceptually from a two-terminal diode?
2. Why can an insulated MOSFET gate control current even though ideally almost no steady DC current flows into the gate?
3. Which chapter ideas are reused in transistor operation: band structure, doping, electric fields, carrier density, mobility, or all of them?

21.13.1 Bipolar carrier-injection control

The main transistor discussion emphasized the MOSFET because field-effect devices dominate modern digital electronics. A second fundamental transistor family is the bipolar junction transistor (BJT). An npn BJT contains an n-type emitter, thin p-type base and n-type collector. A pnp device reverses the carrier polarities.

In forward-active operation of an npn transistor, the base–emitter junction is forward biased and the base–collector junction is reverse biased. Electrons are injected from the emitter into the thin base. Because the base is thin and lightly doped, many electrons cross it without recombining and are swept into the collector by the reverse-biased junction field. A relatively small base current therefore accompanies a much larger collector current.

At an introductory circuit level one often writes

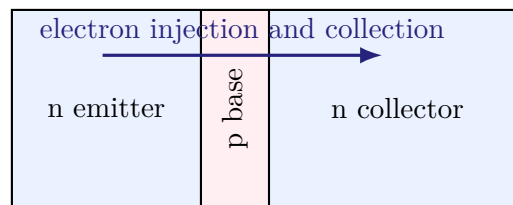
$$I_C \approx \beta I_B, \quad (21.37)$$

where β is a device- and operating-point-dependent current gain. This is not a fundamental constant. A more physically revealing relation connects collector current exponentially to base–emitter voltage over an appropriate active region,

$$I_C \approx I_S \exp\left(\frac{V_{BE}}{V_T}\right). \quad (21.38)$$

Real transistor models include Early effect, finite base resistance, high-injection effects, capacitances and temperature dependence.

Figure 21.15 emphasizes the thin base that allows injected minority carriers to reach the collector.



qualitative npn structure

Figure 21.15: Qualitative npn BJT structure. The base is drawn thin to emphasize that injected minority carriers can cross it and be collected. Real transistor geometry is optimized in three dimensions.

Guided checks

1. Which junction is forward biased in forward-active npn operation?
2. Why is a thin base useful for collector current?
3. Why should $I_C = \beta I_B$ be treated as an operating model rather than a universal identity?
4. Compare BJT carrier-injection control with MOSFET electrostatic gate control.

21.14 Temperature dependence: metals versus semiconductors

The opposite-looking temperature trends of metals and semiconductors provide a useful synthesis. Write conductivity schematically as

$$\sigma \sim \text{carrier density} \times \text{charge} \times \text{mobility}. \quad (21.39)$$

For a simple metal, the carrier density changes little with temperature over an ordinary range. Increasing temperature strengthens phonon scattering and usually lowers mobility, so resistivity increases.

For an intrinsic semiconductor, carrier mobility also generally decreases with increasing temperature because scattering becomes stronger. However, the carrier density increases very rapidly as more electrons are thermally excited across the band gap. The carrier-density increase usually dominates, so resistivity decreases strongly with temperature over the intrinsic regime.

A doped semiconductor can show more complicated behaviour. At very low temperature, dopants may not be fully ionized (the freeze-out regime). At intermediate temperature, the dopant density

controls the majority carriers (the extrinsic regime). At sufficiently high temperature, intrinsic generation overwhelms the dopant contribution (the intrinsic regime). The advanced section develops this sequence quantitatively.

Figure 21.16 compares the usual temperature trends while leaving material-specific scales unspecified.

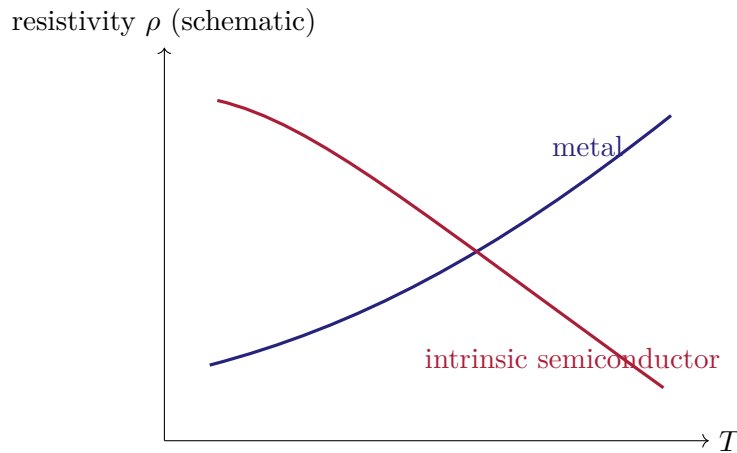


Figure 21.16: Qualitative contrast in ordinary temperature dependence. Metallic resistivity often increases as scattering strengthens, while intrinsic-semiconductor resistivity often decreases because carrier generation dominates. The curves are schematic and not universal laws over all temperatures.

Example 21.15: Separating carrier-density and mobility effects

Suppose heating a semiconductor from T_1 to T_2 increases both n and p by a factor of 40, while both mobilities fall by a factor of 2.5. Assume electron and hole contributions scale in the same way. Estimate the conductivity ratio.

Because $\sigma = e(n\mu_n + p\mu_p)$, every contribution gains a factor $40/2.5 = 16$. Therefore

$$\frac{\sigma_2}{\sigma_1} \approx 16, \quad \frac{\rho_2}{\rho_1} \approx \frac{1}{16}.$$

The example demonstrates why a semiconductor can become much more conductive even while individual carrier mobility decreases.

21.15 How semiconductor devices are fabricated: a physics-level overview

The electrical structures discussed in this chapter are produced by controlling material composition and geometry on small scales. A full fabrication course is far beyond our scope, but the basic sequence helps connect ideal band diagrams to real hardware.

21.15.1 Oxidation, deposition and pattern transfer

Thin insulating, conducting and semiconducting layers can be grown or deposited on a wafer. Photolithography then transfers a geometric pattern from a mask or digital exposure system into a light-sensitive resist. Selected regions are subsequently etched, implanted, doped or coated. Repetition builds a multilayer device structure.

21.15.2 Doping by diffusion and ion implantation

Dopants can be introduced thermally by diffusion or accelerated as ions and implanted into the surface. Subsequent thermal processing repairs lattice damage and activates dopants by placing them on electrically useful lattice sites. The resulting concentration is not generally a perfect step function; real profiles vary continuously with depth.

A simple diffusion model obeys Fick's law,

$$\mathbf{J}_{\text{dop}} = -D_{\text{dop}} \nabla N, \quad (21.40)$$

where N is dopant concentration and D_{dop} is a diffusion coefficient. This equation describes atomic dopant transport during processing, not the electrical current density of electrons and holes.

21.15.3 Contacts, passivation and packaging

Metal contacts must inject or collect carriers without adding unacceptable resistance or unwanted junction barriers. Surface passivation reduces electrically active interface states. Packaging provides electrical connections, mechanical protection and heat removal. Consequently, the measured device is always a system containing semiconductor, interfaces, metal contacts and thermal pathways.

Figure 21.17 connects these operations into an iterative fabrication sequence.

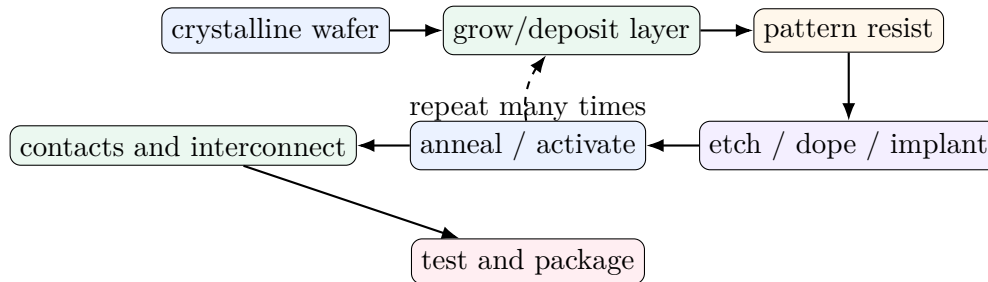


Figure 21.17: Highly simplified semiconductor-fabrication workflow. Modern integrated circuits require many repetitions and additional process steps; the purpose here is only to connect designed doping profiles and interfaces to physical fabrication.

Caution

Do not interpret textbook p–n junction drawings as two separately manufactured blocks simply pressed together. Practical junctions are normally created within a continuous semiconductor crystal by spatially controlled doping and processing.

21.16 Thermal effects, self-heating and electrostatic discharge

Electronic devices operate simultaneously as electrical and thermal systems. Electrical power dissipated in a device,

$$P = IV, \quad (21.41)$$

raises its temperature unless heat is removed. Because semiconductor parameters are temperature dependent, self-heating can change the very current that produces the heat.

A simple steady thermal model uses thermal resistance R_θ ,

$$T_j - T_a \approx PR_\theta, \quad (21.42)$$

where T_j is a junction temperature and T_a is an ambient or reference temperature. Thermal resistance is measured in KW^{-1} . This compact model hides spatial temperature gradients and transient heat capacity, but it is useful for first design estimates.

Example 21.16: Junction-temperature estimate

A device dissipates $P = 1.8 \text{ W}$ and has specified junction-to-ambient thermal resistance $R_\theta = 35 \text{ K W}^{-1}$ under a particular mounting condition. At ambient temperature $T_a = 25^\circ \text{C}$,

$$\Delta T \approx PR_\theta = (1.8)(35) = 63 \text{ K}.$$

The estimated junction temperature is therefore about 88°C . The calculation is only as good as the stated thermal boundary condition; airflow, heat sinks, board copper area and package mounting can substantially change the effective thermal resistance.

21.16.1 Thermal runaway as positive feedback

If increasing temperature increases current at fixed applied voltage, the dissipated power can rise, causing further heating. This positive feedback can lead to thermal runaway. Circuit designers use current limiting, negative feedback, heat sinking and temperature-aware control to keep devices inside a safe operating region.

21.16.2 Electrostatic discharge

A MOS gate oxide can be extremely thin. A static charge accumulated on a person or object can create a high electric field when discharged into a device terminal. Electrostatic-discharge (ESD) protection therefore uses controlled conducting paths, grounding procedures and on-chip protection structures to limit voltage and current transients.

Key message

Device reliability is not separate from the physics of this chapter. Band gaps, junction fields, leakage currents, thermal transport and breakdown mechanisms determine how much electrical and thermal stress a device can survive.

21.16.3 Interpreting specifications at their test conditions

Engineering data sheets translate microscopic device physics into guaranteed or typical terminal-level quantities. Learning to read them is an important bridge from textbook formulas to real components.

Diode quantities

Typical entries include maximum reverse voltage, forward voltage at specified current, reverse leakage current, junction capacitance, reverse-recovery time and maximum power dissipation. Each quantity is tied to a test condition. A statement such as “forward voltage V_F ” is incomplete without current and temperature.

LED quantities

LED sheets may list dominant or peak wavelength, luminous intensity or optical power, forward current, forward voltage, viewing angle and thermal resistance. Optical output is not determined by electrical current alone; temperature, package optics and device efficiency matter.

MOSFET quantities

Important quantities include drain-source voltage rating, continuous and pulsed current ratings, on-state resistance $R_{DS(\text{on})}$ at specified V_{GS} and temperature, threshold-voltage range, gate charge,

capacitances and thermal limits. Threshold voltage is not the gate voltage required for low-resistance power operation; it is usually defined at a small test current.

Method

When reading a data sheet, attach four questions to every number: *What is being measured? Under what bias and temperature? Is the value typical or guaranteed? What physical limitation does it represent?* This prevents a common engineering error: using a correct number outside its specified operating condition.

Guided checks

Suppose a MOSFET data sheet gives $R_{DS(on)} = 20 \text{ m}\Omega$ at $V_{GS} = 10 \text{ V}$ and $T = 25^\circ \text{C}$, but the circuit drives the gate only to 4.5 V at elevated temperature. Why is it unsafe to assume the same resistance? Which microscopic quantities could change with gate voltage and temperature?

21.16.4 Choosing a device model

The same device can be represented at different levels depending on the question.

Level	Typical representation	Best used for
Ideal switch	ON/OFF states	first circuit logic, rectifier direction, protection topology
Constant-drop diode	fixed V_F in conduction	rough power-supply and clipping estimates
Exponential diode	$I = I_S(e^{V/(n_d V_T)} - 1)$	moderate-current I - V physics and temperature trends
Diode with parasitics	exponential law plus R_s , leakage and capacitance	realistic transient and high-current circuit behaviour
Drift-diffusion device	Poisson + continuity + transport equations	spatial carrier and field profiles
Quantum transport	transmission amplitudes / Green functions	nanoscale coherent or quasi-ballistic devices

Take-home message

A model is not “good” because it has many parameters. It is good when its assumptions match the operating regime and it answers the engineering question with controlled error. Learning to move between ideal switch, compact equation and spatial transport model is a core skill in applied physics.

21.17 Engineering and applied-science connections

The band-and-carrier picture is not confined to discrete electronic components. It appears across sensing, communication, energy conversion, computation and scientific instrumentation.

21.17.1 Thermistors and temperature sensing

A negative-temperature-coefficient thermistor exploits thermally activated transport. Depending on its composition, activation may involve carrier generation, release from localized states or hopping; it need not follow the intrinsic-crystal band-gap law. Its resistance can change far more strongly than

that of an ordinary metallic wire. Thermistors therefore provide sensitive temperature measurement but also require calibration because the resistance–temperature relation is nonlinear.

21.17.2 Radiation and particle detectors

Ionizing radiation can create electron–hole pairs in semiconductor detector material. An applied electric field sweeps the carriers toward electrodes, producing a current pulse. The amount of collected charge is related to deposited energy. High-purity silicon and germanium detectors are important examples in nuclear and particle physics, while semiconductor photodetectors are ubiquitous in cameras, optical communication and spectroscopy.

21.17.3 Power electronics

Power diodes and transistors must handle large currents and voltages while limiting heat generation. Band gap, carrier mobility, breakdown field, thermal conductivity and device geometry all matter. Wide-band-gap semiconductors can offer high breakdown fields and high-temperature capability, but a realistic comparison requires material and manufacturing considerations beyond a single parameter.

21.17.4 Computing and digital logic

The modern processor is an immense network of field-effect transistors. At the logic level, a transistor is abstracted as a switch. At the device level, its switching action comes from electrostatic control of channel charge and from semiconductor transport. Shrinking transistor dimensions improves density and can improve switching speed, but it also makes leakage, quantum tunnelling, variability, contact resistance and heat removal increasingly important.

21.17.5 Photovoltaics and optical communication

Solar cells convert absorbed photon energy into electrical work by generating and separating carriers. LEDs and semiconductor lasers perform the complementary electrical-to-optical conversion. Photodiodes convert optical signals back into electrical signals. A single conceptual toolkit – bands, gaps, carriers, junction fields and recombination – therefore underlies an entire optical communication chain.

Key message

Semiconductor technology is powerful because electrical, optical and chemical inputs can all change carrier populations. The same material can therefore act as a switch, detector, emitter, sensor or energy converter depending on structure and bias.

21.18 Laboratory connection: measuring a diode I – V curve

A diode current–voltage experiment is a compact way to connect microscopic semiconductor ideas to laboratory measurement. A practical setup uses a controllable voltage source, a series resistor to limit current, a voltmeter across the diode and an ammeter or resistor-based current measurement. The series resistor is essential because forward current can rise rapidly.

A sensible measurement sequence is:

1. verify the diode polarity and current-limit setting before energizing the circuit;
2. increase forward bias in small steps while recording V_D and I ;
3. measure several reverse-bias points within the safe device rating;

4. plot I versus V_D on a linear scale to show rectification;
5. plot $\ln(I/I_{\text{ref}})$ versus V_D over a moderate forward-current region where $I > 0$ and series resistance is not yet dominant;
6. fit a straight line and compare the slope with an idealized exponential model;
7. identify where the data depart systematically from the model.

For the model

$$I \approx I_0 \exp\left(\frac{V}{n_d V_T}\right), \quad (21.43)$$

taking a logarithm of the dimensionless ratio I/I_{ref} , where I_{ref} is a fixed reference current such as 1 A, gives

$$\ln\left(\frac{I}{I_{\text{ref}}}\right) \approx \ln\left(\frac{I_0}{I_{\text{ref}}}\right) + \frac{V}{n_d V_T}. \quad (21.44)$$

A plot of $\ln(I/I_{\text{ref}})$ against V should therefore be approximately linear over a suitable range, with slope

$$m = \frac{1}{n_d V_T}. \quad (21.45)$$

The inferred ideality factor is

$$n_d = \frac{1}{m V_T}. \quad (21.46)$$

Method

A fit is not automatically a proof of a model. After fitting Eq. (21.44), inspect residuals. Curvature at low current can indicate leakage or recombination; curvature at high current can indicate series resistance and self-heating. The physically interesting information often lies in the *departure* from an ideal line.

21.18.1 Uncertainty and measurement limitations

Important uncertainty sources include voltage-meter resolution, current-sense resistor tolerance, temperature variation, contact resistance and self-heating. Because diode current is exponentially sensitive to voltage and temperature, a seemingly small uncertainty can produce a noticeable current change. A good report should therefore record ambient temperature and should avoid interpreting more significant digits than the instruments support.

Example 21.17: Ideality factor from a semilog slope

At approximately 300 K, a semilog plot of forward-diode data has slope $m = 22.0 \text{ V}^{-1}$ for $\ln(I/I_{\text{ref}})$ versus V . Estimate the ideality factor.

With $V_T = 25.9 \text{ mV}$,

$$n_d = \frac{1}{m V_T} = \frac{1}{(22.0)(0.0259)} \approx 1.76.$$

The value indicates that the measured region is not described by the most ideal $n_d = 1$ diffusion model. It does not by itself identify the detailed microscopic mechanism; additional analysis would be required.

21.19 Computational connection: diode modelling and parameter scans

A short numerical program can reveal the enormous dynamic range of the diode equation more effectively than a hand calculation. A robust computational activity should include both a direct $I(V)$ plot and a semilog plot. The following algorithm is language independent:

1. choose T , I_0 and optionally ideality factor n_d ;
2. compute $V_T = k_B T / e$;
3. create a voltage array covering modest reverse bias and forward bias;
4. evaluate

$$I(V) = I_0 \left[\exp \left(\frac{V}{n_d V_T} \right) - 1 \right];$$

5. plot I against V on linear axes;
6. for positive forward current, plot $\log_{10} I$ against V ;
7. repeat for several temperatures or ideality factors and compare slopes;
8. impose a series resistance numerically using the implicit relation

$$I = I_0 \left[\exp \left(\frac{V - I R_s}{n_d V_T} \right) - 1 \right]$$

and solve it iteratively;

9. compare the ideal and nonideal curves.

The series-resistance model is particularly instructive because it shows how a physically simple nonideality changes the high-current region. A Newton iteration or root-finding method can solve the implicit equation. This connects semiconductor physics to numerical methods, nonlinear equations and data fitting.

For forward terminal voltage $V \geq 0$ and $R_s > 0$, solve the inverse relation

$$F(I) = n_d V_T \ln(1 + I/I_S) + I R_s - V = 0. \quad (21.47)$$

It has $F'(I) = n_d V_T / (I + I_S) + R_s > 0$, so the forward root is unique and lies in $0 \leq I \leq V/R_s$. Bisection on this bracket is reliable and avoids evaluating a huge exponential. Stop when both the bracket width and voltage residual are small compared with the desired current and voltage accuracy. In code, functions such as `log1p` and `expm1` improve small-argument accuracy; their names indicate mathematical operations, not a required programming language. Unrestricted Newton iteration on the original exponential can overflow or step into $I < -I_S$.

Example 21.18: Separating ideality and series resistance in simulated data

At 300 K let a simulated device obey Eq. (21.47) with $I_S = 1.0 \text{ nA}$, $n_d = 1.5$ and $R_s = 20 \Omega$. Generate voltage from prescribed current to avoid an unnecessary root solve:

$$V = (1.5)(0.02585 \text{ V}) \ln \left(1 + \frac{I}{10^{-9} \text{ A}} \right) + (20 \Omega) I.$$

I (mA)	Junction voltage (V)	IR_s (V)	Terminal voltage (V)
0.10	0.446	0.002	0.448
1.00	0.536	0.020	0.556
10.0	0.625	0.200	0.825

Interpretation. At low current, each decade adds approximately $n_d V_T \ln 10 = 89.3$ mV to the junction voltage. At high current, the increasing resistor drop spoils a straight terminal-voltage semilog fit. Indeed $dV/d\ln(I/I_{\text{ref}}) \simeq n_d V_T + IR_s$; fitting the high-current region with $R_s = 0$ would overestimate ideality. These values are synthetic model outputs, not laboratory measurements.

Example 21.19: Temperature sensitivity of the exponential factor

Compare the thermal voltage at 250 K and 350 K.

$$V_T(250) = \frac{k_B(250)}{e} \approx 21.5 \text{ mV},$$

$$V_T(350) \approx 30.2 \text{ mV}.$$

For a fixed voltage, the simple exponential factor $\exp(V/V_T)$ therefore changes substantially with temperature. With fixed I_0 , the exponential factor decreases when temperature rises at fixed positive voltage. In a real diode the strong increase of $I_0(T)$ can reverse this trend. A scan holding I_0 fixed is therefore a controlled mathematical comparison, not a prediction of the measured device-temperature response.

Guided checks

1. Why is a semilog plot more informative than a linear plot for testing an exponential model over many decades of current?
2. Why can a series resistor produce curvature in a semilog I – V graph at high current?
3. If temperature is not controlled, which fitted diode parameters may appear to drift?
4. Why should reverse-breakdown voltages not be explored casually in a teaching lab without device ratings and current limiting?

21.20 Main chapter summary

Summary

- In a crystal, the discrete electronic levels of isolated atoms split into dense *allowed energy bands*. Intervals with no allowed states are *band gaps*.
- Conduction depends on the availability of nearby unoccupied states as well as on the number of electrons. A partially filled or overlapping band supports metallic conduction; a full band separated from the next empty band by a large gap gives insulating behaviour in ordinary conditions.
- Macroscopic ohmic transport is described by $\mathbf{J} = \sigma \mathbf{E}$. In the Drude model, $\sigma = ne^2\tau/m$, showing explicitly the roles of carrier density and momentum relaxation.
- A semiconductor has a band gap small enough that thermal excitation, light or doping can create useful mobile carriers. Intrinsic excitation creates electron–hole pairs, so $n = p = n_i$.

in an ideal intrinsic material.

- In the simple two-carrier transport model, $\sigma = e(n\mu_n + p\mu_p)$.
- Donor doping creates n-type material with electrons as majority carriers; acceptor doping creates p-type material with holes as majority carriers. Bulk doped material remains nearly charge neutral.
- At a p–n junction, majority carriers diffuse across the interface and expose fixed ionized dopants. The resulting depletion region and built-in electric field oppose further diffusion.
- Forward bias reduces the effective junction barrier and promotes current. Reverse bias increases the barrier and gives a small minority-carrier current until breakdown mechanisms become important.
- The ideal diode model $I = I_0[\exp(V/V_T) - 1]$ captures exponential forward conduction but does not describe all real-device regimes.
- LEDs use carrier injection and radiative recombination. A characteristic photon wavelength is estimated from $\lambda \approx hc/E_g$, with important material-dependent qualifications.
- Photodiodes and solar cells use light-generated electron–hole pairs and junction fields to produce electrical current or power.
- Transistors add a control terminal so that electric fields or carrier injection can regulate current. This is the microscopic foundation of switching and amplification.

Take-home message

The microscopic story of solid-state electronics can be compressed into a chain: *band structure determines available states; doping and temperature determine carrier populations; fields and scattering determine transport; junctions and gates control where those carriers can move.* Every diode, LED, photodiode and transistor in this chapter is an engineered use of that chain.

21.21 Advanced Topics for Technical and Engineering Students

The main chapter deliberately used a compact band picture and phenomenological transport laws. This enrichment section develops the same physics one level more deeply. It is intended for strong students who want to understand how carrier statistics, band curvature, diffusion, junction electrostatics and modern device models fit together. The goal is not to replace a dedicated solid-state physics course. Rather, it is to build a technically useful bridge from introductory quantum mechanics and electromagnetism to semiconductor-device physics.

Pedagogical boundary

The advanced material introduces effective mass, Fermi–Dirac statistics, nondegenerate carrier formulas, the mass-action law, drift–diffusion equations, the depletion approximation, a compact derivation of the ideal diode law, MOSFET square-law models, elementary quantum conductance and selected material-engineering ideas. Bloch’s theorem, reciprocal lattices, detailed phonon theory, many-body band-structure calculations and semiconductor fabrication process physics are mentioned only where they clarify the models; they are not developed formally.

Learning goals

After completing the enrichment section, a strong student should be able to:

- connect the free-electron Fermi gas to metallic carrier density and Fermi energy;
- interpret an $E(k)$ curve and define an effective mass from its curvature;
- use Fermi–Dirac and Maxwell–Boltzmann approximations to estimate semiconductor carrier populations;
- apply charge neutrality and the mass-action law to simple doped-semiconductor calculations;
- combine drift and diffusion currents and explain the Einstein relation;
- derive and use the abrupt-junction built-in potential and depletion-width formulas under stated assumptions;
- explain the physical origin and limitations of the Shockley diode equation;
- calculate junction capacitance and interpret its bias dependence;
- use simplified models for solar cells and long-channel MOSFETs;
- recognize where ballistic transport, thermoelectric effects and superconductivity require physics beyond the elementary Drude picture.

21.21.1 Crystal states and carrier dynamics

Bloch’s theorem: why a periodic crystal has propagating quantum states

The effective-mass and band pictures can be traced to a deeper result. If the electron potential energy U in the crystal is periodic,

$$U(\mathbf{r} + \mathbf{R}) = U(\mathbf{r}) \quad (21.48)$$

for any lattice translation $\mathbf{R}$, the one-electron Schrödinger equation admits solutions of Bloch form

$$\psi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}, \quad u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}). \quad (21.49)$$

The plane-wave factor describes propagation while the periodic function u carries the modulation imposed by the lattice. This is why a perfect periodic crystal does not simply behave like an array of independent obstacles that randomize electron motion at every atom.

At certain wave vectors, Bragg-like coupling between electron waves travelling in opposite directions can open energy gaps. A one-dimensional periodic-potential model such as the Kronig–Penney

model makes this explicit: only selected energy ranges satisfy the allowed-wave condition, producing bands separated by forbidden gaps.

Figure 21.18 isolates the spatial periodicity of the electron potential energy.

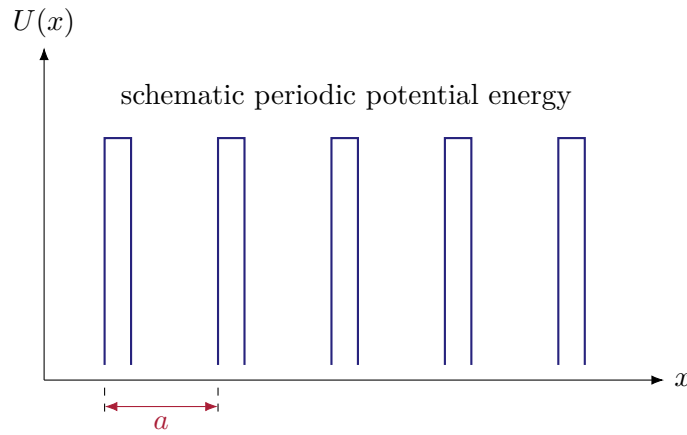


Figure 21.18: A schematic one-dimensional periodic electron potential energy. Simple models of this kind show how periodicity can produce allowed energy bands and forbidden gaps.

Pedagogical boundary

A formal derivation of Bloch's theorem, reciprocal lattices and Brillouin zones belongs to solid-state physics. Here the theorem is used only to replace the misleading “electron collides with every atom” picture by the correct idea that periodicity is already built into the crystal eigenstates.

Energy versus wave vector: the information hidden by a simple band diagram

A vertical band diagram such as those used in the main chapter tells us which energies are allowed, but it suppresses an important variable: the crystal wave vector $\mathbf{k}$. In a periodic solid the electron states are naturally organized into energy bands $E_n(\mathbf{k})$, where n labels the band. The function $E(\mathbf{k})$ is called a *dispersion relation*. It contains dynamical information because the group velocity of a wave packet in a band is

$$\mathbf{v}_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}). \quad (21.50)$$

For a one-dimensional band this reduces to

$$v_g = \frac{1}{\hbar} \frac{dE}{dk}. \quad (21.51)$$

A completely flat band therefore corresponds to small group velocity, while a strongly dispersive band can support rapidly moving wave packets.

Near a smooth band extremum, a Taylor expansion provides an especially useful approximation. Near the bottom of a conduction band at $k = k_0$,

$$E(k) \simeq E_C + \frac{\hbar^2 (k - k_0)^2}{2m_n^*}, \quad (21.52)$$

where the electron effective mass m_n^* is defined by the band curvature,

$$\frac{1}{m_n^*} = \frac{1}{\hbar^2} \left. \frac{d^2 E}{dk^2} \right|_{k_0}. \quad (21.53)$$

In three dimensions the effective mass can be a tensor; the scalar form above is sufficient for the isotropic introductory model.

Key message

The effective mass is not a statement that the electron's fundamental rest mass has changed. It is a compact way of describing how the electron responds to external forces while moving through a periodic crystal potential. The crystal modifies the relation between force, acceleration and wave-vector evolution.

Figure 21.19 links the sign of the band curvature to the conduction-electron and hole descriptions.

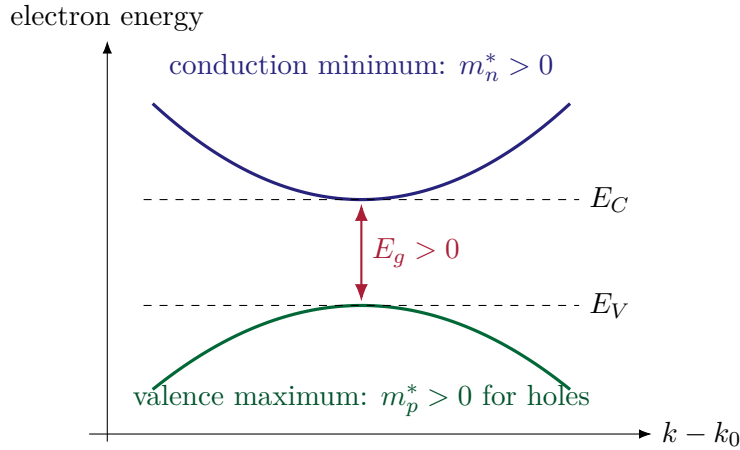


Figure 21.19: Local parabolic band approximations with the conduction minimum above the valence maximum. Conduction electrons have positive curvature; the valence-electron curvature is negative, and the missing-electron hole is assigned a positive mass. Energy and wave-vector scales are schematic.

For a uniform applied field in the semiclassical approximation,

$$\hbar \frac{d\mathbf{k}}{dt} = q\mathbf{E}, \quad \frac{dv_g}{dt} = \frac{qE}{\hbar^2} \frac{d^2E}{dk^2} = \frac{qE}{m^*}. \quad (21.54)$$

This explains why band curvature enters transport. A hole near a valence maximum is described by $E_h(k) \simeq \hbar^2(k - k_v)^2/(2m_p^*)$, with $1/m_p^* = -E_v''(k_v)/\hbar^2 > 0$; its electric charge is $+e$.

For a valence band the curvature near the top is negative. Describing the missing-electron states as positively charged holes avoids carrying a negative effective mass through elementary calculations. The resulting hole effective mass is defined as a positive parameter associated with the curvature near the valence-band maximum.

Guided checks

1. What does the slope dE/dk tell you about the group velocity?
2. Why can two materials with similar carrier densities have different conductivities even if their scattering times are similar?
3. Why is the effective mass generally a property of a particular band and direction rather than a universal constant of the material?

Anisotropic effective mass and conductivity tensors

In three dimensions, the curvature of $E(\mathbf{k})$ need not be the same in every direction. The inverse effective-mass tensor is

$$(m^{*-1})_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \partial k_j}. \quad (21.55)$$

Likewise, an anisotropic material need not obey a scalar conductivity law. The linear relation becomes

$$J_i = \sum_j \sigma_{ij} E_j, \quad (21.56)$$

where σ_{ij} is a second-rank tensor. Off-diagonal components can appear because of crystal symmetry or magnetic field.

This matters in layered materials, low-symmetry crystals and Hall measurements. It also clarifies why “the mobility” may depend on direction. In an isotropic introductory model, symmetry reduces σ_{ij} to $\sigma\delta_{ij}$ and the familiar vector equation $\mathbf{J} = \sigma\mathbf{E}$ is recovered.

Example 21.20: Current not parallel to the field in an anisotropic conductor

Consider a two-dimensional conductivity tensor

$$\boldsymbol{\sigma} = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} \text{ S m}^{-1}$$

and an electric field $\mathbf{E} = (3, 0) \text{ V m}^{-1}$. Then

$$\begin{pmatrix} J_x \\ J_y \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 12 \\ 3 \end{pmatrix} \text{ A m}^{-2}.$$

The current is not parallel to the applied field. A scalar conductivity would be insufficient to describe this material orientation.

Mobility, conductivity and scattering-time distributions

The elementary Drude formula uses one relaxation time τ . Real carriers occupy a range of energies, and scattering probability can depend on energy and direction. A more complete semiclassical transport theory averages velocity and relaxation time over the nonequilibrium distribution. Symbolically,

$$\sigma \sim e^2 \int v_x^2(E) \tau(E) \left[-\frac{\partial f}{\partial E} \right] g(E) dE. \quad (21.57)$$

This expression is schematic; numerical prefactors and tensor structure depend on dimensionality and band model. Its conceptual message is powerful: transport is weighted by the states whose occupation can be changed by the applied field, through the factor $-\partial f/\partial E$.

In a degenerate metal this weighting is concentrated near E_F . In a nondegenerate semiconductor it samples carriers in a thermal energy range above the band edge. This helps unify why Fermi-surface properties dominate metals while band-edge effective masses and scattering dominate many semiconductor transport problems.

21.21.2 Quantum statistics and carrier populations

The free-electron Fermi gas in a metal

The Drude model treats conduction electrons classically between scattering events. A first quantum improvement keeps the idea of nearly free electrons but imposes quantum state counting and the Pauli exclusion principle. For an ideal three-dimensional free-electron gas at zero temperature, all single-particle states are occupied up to the Fermi wave number k_F . In a normalization volume $\mathcal{V}$ with periodic boundaries, each wave-vector state occupies volume $(2\pi)^3/\mathcal{V}$ in k space. The occupied

sphere has volume $4\pi k_F^3/3$. Multiplying by spin degeneracy two and dividing by $\mathcal{V}$ gives

$$n = \frac{2}{(2\pi)^3} \frac{4\pi k_F^3}{3}.$$

Thus the number density is

$$n = \frac{k_F^3}{3\pi^2}, \quad k_F = (3\pi^2 n)^{1/3}. \quad (21.58)$$

The corresponding Fermi energy and Fermi speed are

$$E_F = \frac{\hbar^2 k_F^2}{2m} = \frac{\hbar^2}{2m} (3\pi^2 n)^{2/3}, \quad v_F = \frac{\hbar k_F}{m}. \quad (21.59)$$

These quantities set the characteristic quantum scales of a degenerate electron gas.

The phrase *Fermi energy* is most transparent at $T = 0$: it is the energy of the highest occupied state. At nonzero temperature the chemical potential plays the role denoted by E_F in many introductory treatments; only states within an energy range of order $k_B T$ around that level change their occupancy appreciably.

Example 21.21: Fermi energy from a specified electron density

Suppose an idealized monovalent metal has conduction-electron density

$$n = 8.5 \times 10^{28} \text{ m}^{-3}.$$

Estimate k_F , E_F and v_F in the free-electron model.

Solution. First,

$$k_F = (3\pi^2 n)^{1/3} \approx 1.36 \times 10^{10} \text{ m}^{-1}.$$

Then

$$E_F = \frac{\hbar^2 k_F^2}{2m_e} \approx 1.13 \times 10^{-18} \text{ J} \approx 7.0 \text{ eV}.$$

Finally,

$$v_F = \frac{\hbar k_F}{m_e} \approx 1.6 \times 10^6 \text{ m s}^{-1}.$$

The Fermi speed is much larger than the ordinary drift speed produced by laboratory electric fields. Current is not created because all electrons suddenly start moving at the drift speed; instead, an already fast quantum distribution is shifted slightly out of equilibrium.

Caution

Do not identify the Fermi speed with the drift speed. The Fermi speed characterizes occupied quantum states in the metal. The drift speed associated with a modest current is typically many orders of magnitude smaller and represents a small asymmetry of the electronic distribution.

Density of states and Fermi–Dirac occupancy

For a three-dimensional parabolic band with effective mass m^* , the density of available states per unit volume per unit energy has the square-root form

$$g(E) = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_0}, \quad E \geq E_0, \quad (21.60)$$

where E_0 is a conduction-band minimum and m^* is a density-of-states mass. Spin degeneracy two is included. For a valence-band maximum replace $\sqrt{E - E_0}$ by $\sqrt{E_V - E}$ and use a positive hole

density-of-states mass. The precise density of states of a real crystal may have a much richer shape; Eq. (21.60) is the local effective-mass approximation.

The probability that a one-electron state of energy E is occupied in thermal equilibrium is the Fermi–Dirac function

$$f(E) = \frac{1}{\exp[(E - E_F)/(k_B T)] + 1}. \quad (21.61)$$

It has three useful features:

- $f(E_F) = 1/2$ at any nonzero temperature;
- states far below E_F are very likely to be occupied;
- states far above E_F are unlikely to be occupied.

The square-root density of states follows from the same counting. The number per volume below energy E is

$$\frac{\mathcal{N}(< E)}{\mathcal{V}} = \frac{1}{3\pi^2} \left[\frac{2m^*(E - E_C)}{\hbar^2} \right]^{3/2}.$$

Differentiating with respect to E gives Eq. (21.60); $g(E)$ has unit $\text{m}^{-3} \text{J}^{-1}$. The number of occupied states is different from the number available: one must weight by $f(E)$.

The carrier concentration is obtained by multiplying the available-state density by the occupancy probability and integrating over the appropriate band.

Read Fig. 21.20 in units of $k_B T$ to see which energy range is thermally broadened.

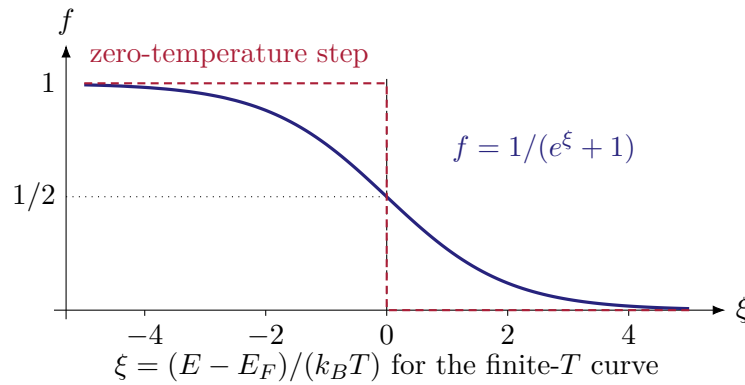


Figure 21.20: Calculated Fermi–Dirac occupation as a function of dimensionless energy offset. The dashed step is the $T \rightarrow 0$ limiting occupation, overlaid for comparison. Thermal occupancy changes over a few $k_B T$, while $f(E_F) = 1/2$ for finite temperature.

If $E - E_F \gg k_B T$, the “+1” in the denominator is negligible and the Fermi–Dirac distribution reduces to the Maxwell–Boltzmann form

$$f(E) \simeq \exp \left[-\frac{E - E_F}{k_B T} \right]. \quad (21.62)$$

This nondegenerate approximation is extremely useful for ordinary semiconductor calculations, but it fails when the Fermi level lies too close to or inside a band.

Method

Before using a semiconductor carrier formula, identify its statistical regime. The familiar exponential expressions for n and p assume nondegenerate statistics and a parabolic-band effective-mass model. Heavy doping can push the semiconductor into a degenerate regime where the full Fermi–Dirac integrals are required.

Carrier concentrations and effective densities of states

For a nondegenerate semiconductor in thermal equilibrium, integration of the parabolic-band density of states with the Maxwell–Boltzmann tail gives

$$n = N_C \exp\left[-\frac{E_C - E_F}{k_B T}\right], \quad (21.63)$$

$$p = N_V \exp\left[-\frac{E_F - E_V}{k_B T}\right], \quad (21.64)$$

where the effective densities of states are

$$N_C = 2 \left(\frac{2\pi m_n^* k_B T}{h^2} \right)^{3/2}, \quad N_V = 2 \left(\frac{2\pi m_p^* k_B T}{h^2} \right)^{3/2}. \quad (21.65)$$

For example, putting $\epsilon = E - E_C$ in $n = \int_{E_C}^{\infty} g(E) f(E) dE$ gives

$$n = \frac{1}{2\pi^2} \left(\frac{2m_n^*}{\hbar^2} \right)^{3/2} e^{-(E_C - E_F)/(k_B T)} \int_0^{\infty} \epsilon^{1/2} e^{-\epsilon/(k_B T)} d\epsilon.$$

The integral equals $(k_B T)^{3/2} \sqrt{\pi}/2$. Using $h = 2\pi\hbar$ yields the stated N_C , which has the dimensions of an inverse volume. The hole integral weights valence states by $1 - f(E)$ and gives N_V .

The factor 2 accounts for spin in this single-valley isotropic model. Multiple equivalent valleys multiply the state count; anisotropic or multiple valence bands require the appropriate density-of-states masses. These masses need not equal the transport masses inferred from mobility.

Multiplying Eqs. (21.63) and (21.64) eliminates E_F :

$$np = N_C N_V \exp\left(-\frac{E_g}{k_B T}\right). \quad (21.66)$$

For an intrinsic semiconductor, $n = p = n_i$, so

$$n_i = \sqrt{N_C N_V} \exp\left(-\frac{E_g}{2k_B T}\right), \quad (21.67)$$

and therefore

$$np = n_i^2. \quad (21.68)$$

Equation (21.68) is the *mass-action law* for a nondegenerate semiconductor in thermal equilibrium.

Example 21.22: Temperature sensitivity of intrinsic carrier concentration

Consider a simplified semiconductor whose effective densities of states vary only through the usual $T^{3/2}$ factor and whose band gap may be treated as constant at $E_g = 1.10$ eV over the temperature interval of interest. Estimate the ratio $n_i(350 \text{ K})/n_i(300 \text{ K})$.

Using Eq. (21.67),

$$\frac{n_i(T_2)}{n_i(T_1)} = \left(\frac{T_2}{T_1} \right)^{3/2} \exp\left[-\frac{E_g}{2k_B} \left(\frac{1}{T_2} - \frac{1}{T_1} \right)\right].$$

With $T_1 = 300 \text{ K}$ and $T_2 = 350 \text{ K}$, the exponential factor dominates and the ratio is approximately 26.3. Thus a modest temperature increase can produce an order-of-magnitude change in intrinsic carrier concentration. Real materials also have temperature-dependent band gaps and mobilities, so a precise device calculation requires more detailed data.

Example 21.23: A large gap can still give measurable intrinsic carriers

Consider a hypothetical material at 300 K with $E_g = 1.10$ eV, $N_C = 2.8 \times 10^{25} \text{ m}^{-3}$ and $N_V = 1.0 \times 10^{25} \text{ m}^{-3}$. How can it have carriers when $k_B T \simeq 0.02585$ eV $\ll E_g$?

Strategy. Use state counting and the thermal tail, not a rule that every carrier has energy $k_B T$. From Eq. (21.67),

$$n_i = \sqrt{(2.8 \times 10^{25})(1.0 \times 10^{25})} \exp \left[-\frac{1.10}{2(0.02585)} \right] \approx 9.6 \times 10^{15} \text{ m}^{-3}.$$

The population is tiny compared with the available-state scale, yet a cubic millimetre contains about 9.6×10^6 electron–hole pairs. The factor $E_g/2$ comes from the square root of the equilibrium product np ; exciting a single electron across the gap still requires energy E_g . Thermal energy is a distribution scale, not an upper limit on individual excitation energies.

The intrinsic Fermi level E_i follows by setting $n = p$. The result is close to midgap but is shifted if the electron and hole effective densities of states differ:

$$E_i = \frac{E_C + E_V}{2} + \frac{k_B T}{2} \ln \left(\frac{N_V}{N_C} \right). \quad (21.69)$$

Charge neutrality, ionized dopants and the location of the Fermi level

A macroscopic semiconductor cannot accumulate a large uncompensated bulk charge without generating an enormous electrostatic energy. In equilibrium its interior is therefore approximately charge neutral. If N_D^+ and N_A^- are the concentrations of ionized donors and acceptors,

$$p + N_D^+ = n + N_A^-. \quad (21.70)$$

In the common introductory limit of complete ionization, $N_D^+ \simeq N_D$ and $N_A^- \simeq N_A$.

For an n-type semiconductor with donors greatly exceeding both acceptors and intrinsic carriers, charge neutrality gives $n \simeq N_D - N_A$. The mass-action law then gives

$$p \simeq \frac{n_i^2}{n}. \quad (21.71)$$

Likewise, for a strongly p-type semiconductor,

$$p \simeq N_A - N_D, \quad n \simeq \frac{n_i^2}{p}. \quad (21.72)$$

These results quantify the meaning of majority and minority carriers.

Example 21.24: Majority and minority concentrations with compensation

At a particular temperature, take $n_i = 1.0 \times 10^{16} \text{ m}^{-3}$. A sample contains fully ionized donors $N_D = 5.0 \times 10^{21} \text{ m}^{-3}$ and acceptors $N_A = 1.0 \times 10^{21} \text{ m}^{-3}$. Estimate n and p .

Assume nondegenerate statistics in addition to complete ionization. The condition $N_D - N_A \gg n_i$ establishes the extrinsic n-type limit, but does not by itself establish nondegeneracy. Then

$$n \simeq N_D - N_A = 4.0 \times 10^{21} \text{ m}^{-3}.$$

Then

$$p = \frac{n_i^2}{n} = \frac{(1.0 \times 10^{16})^2}{4.0 \times 10^{21}} = 2.5 \times 10^{10} \text{ m}^{-3}.$$

The crystal remains nearly neutral: the very small minority-hole concentration does not imply a net bulk charge of $-en$ because positively charged ionized donor sites provide the compensating fixed charge.

Using Eq. (21.63), the shift of the Fermi level from the intrinsic level can be written

$$n = n_i \exp\left(\frac{E_F - E_i}{k_B T}\right), \quad p = n_i \exp\left(-\frac{E_F - E_i}{k_B T}\right). \quad (21.73)$$

Donor doping moves E_F upward toward E_C ; acceptor doping moves it downward toward E_V .

Guided checks

1. If donor doping increases n by a factor of 10^4 at fixed temperature, by what factor does p change under the mass-action law?
2. What physical assumption is hidden in writing $N_D^+ = N_D$?
3. Why can the Fermi level move through the band gap even though the gap contains no allowed one-electron states?

Advanced guided derivation: exact compensated carrier densities

The exact fully ionized nondegenerate equilibrium solution is useful enough to derive in the main mathematical flow. With

$$\Delta N = N_D - N_A, \quad (21.74)$$

charge neutrality and mass action give

$$n - p = \Delta N, \quad np = n_i^2. \quad (21.75)$$

Eliminating p yields

$$n^2 - \Delta N n - n_i^2 = 0.$$

The physical root is

$$n = \frac{\Delta N + \sqrt{\Delta N^2 + 4n_i^2}}{2}, \quad (21.76)$$

while

$$p = \frac{-\Delta N + \sqrt{\Delta N^2 + 4n_i^2}}{2}. \quad (21.77)$$

For numerical work, compute the majority root using the positive sum, then obtain the minority concentration from $np = n_i^2$. For example, when $\Delta N > 0$, compute n directly and set $p = n_i^2/n$. This avoids catastrophic cancellation in $\sqrt{\Delta N^2 + 4n_i^2} - \Delta N$ when $\Delta N \gg n_i$.

These equations avoid piecewise “n-type versus p-type” logic and are convenient for numerical parameter scans through the compensation point $N_D = N_A$.

Example 21.25: Smooth crossover through exact compensation

Let $n_i = 1.0 \times 10^{16} \text{ m}^{-3}$.

If $\Delta N = +1.0 \times 10^{18} \text{ m}^{-3}$, Eq. (21.76) gives $n \approx 1.0001 \times 10^{18} \text{ m}^{-3}$ and $p \approx 1.0 \times 10^{14} \text{ m}^{-3}$.

If $\Delta N = 0$, both concentrations equal $n_i = 10^{16} \text{ m}^{-3}$.

If $\Delta N = -1.0 \times 10^{18} \text{ m}^{-3}$, the roles reverse: holes become the majority carriers with $p \approx 10^{18} \text{ m}^{-3}$. The exact solution is continuous through compensation.

Freeze-out, extrinsic and intrinsic temperature regimes

The statement “an n-type semiconductor has $n \approx N_D$ ” is not universally valid. A doped semiconductor can pass through three qualitative temperature regimes:

1. **Freeze-out:** at sufficiently low temperature, a significant fraction of dopants are not ionized. The free-carrier concentration is below the dopant concentration.
2. **Extrinsic regime:** dopants are nearly fully ionized and intrinsic generation remains small. Here $n \approx N_D - N_A$ for n-type material or $p \approx N_A - N_D$ for p-type material.
3. **Intrinsic regime:** at sufficiently high temperature, thermally generated electron–hole pairs dominate over the dopant-provided carriers, so $n \approx p \approx n_i$.

This three-regime picture explains why a plot of conductivity versus inverse temperature can contain more than one approximately activated region.

Figure 21.21 separates incomplete ionization, dopant-controlled density and intrinsic generation.

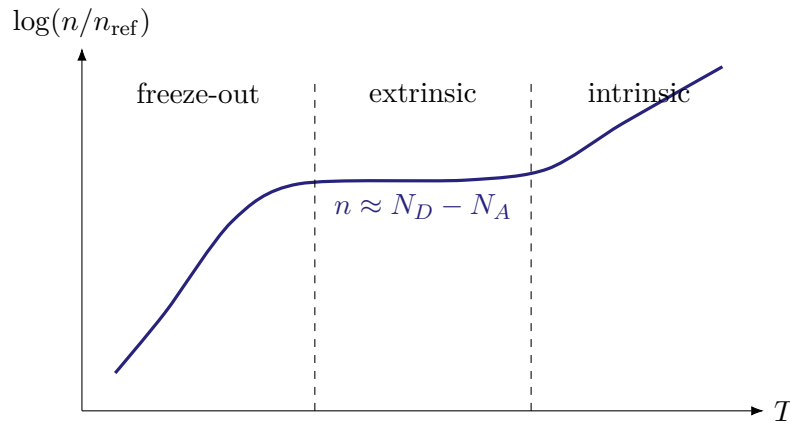


Figure 21.21: Qualitative carrier-concentration regimes for a doped semiconductor, with fixed positive reference density n_{ref} . The precise transition temperatures depend strongly on the material and dopant concentrations.

Mobility also varies with temperature because different scattering processes become important. Consequently, conductivity reflects the product of carrier population and mobility; its temperature dependence cannot be inferred from carrier concentration alone.

Advanced measurement case: Arrhenius plots and activation energies

Many carrier-generation or defect processes have approximately activated temperature dependence,

$$X(T) = X_0 T^m \exp\left(-\frac{E_a}{k_B T}\right), \quad (21.78)$$

where E_a is an activation energy and m describes a prefactor. Choose a fixed reference Y_{ref} with the same units as X/T^m . Taking a dimensionless logarithm gives

$$\ln\left(\frac{X/T^m}{Y_{\text{ref}}}\right) = \ln\left(\frac{X_0}{Y_{\text{ref}}}\right) - \frac{E_a}{k_B} \frac{1}{T}. \quad (21.79)$$

Thus a plot of $\ln[(X/T^m)/Y_{\text{ref}}]$ versus $1/T$ is linear if the model is valid, with slope $-E_a/k_B$. Changing the reference units shifts the intercept and leaves the activation-energy slope unchanged.

For intrinsic carrier concentration, $m = 3/2$ in the simplest effective-mass model and $E_a = E_g/2$ if the band gap is treated as constant. For dopant ionization or trap processes, a different activation

energy can appear. Multiple linear regions can therefore indicate different physical regimes rather than “bad data.”

Caution

A straight line on an Arrhenius plot does not by itself prove a unique microscopic mechanism. Different processes can produce similar activated slopes over limited temperature ranges. Independent measurements and physical consistency checks are needed.

21.21.3 Nonequilibrium carriers, diffusion and recombination

From drift to drift–diffusion transport

A spatially uniform carrier density can carry drift current in an electric field. A spatially nonuniform carrier density can also produce current through diffusion. For one-dimensional nondegenerate transport, the conventional current densities are written

$$J_n = en\mu_n E + eD_n \frac{dn}{dx}, \quad (21.80)$$

$$J_p = ep\mu_p E - eD_p \frac{dp}{dx}, \quad (21.81)$$

where D_n and D_p are diffusion coefficients. The opposite signs of the gradient terms reflect the combination of particle flow direction and carrier charge.

To derive the Einstein relation, write $E = -d\phi/dx$ and $E_C(x) = E_C^{(0)} - e\phi(x)$ for a homogeneous semiconductor. At equilibrium E_F and T are constant, so the Boltzmann density gives

$$\frac{dn}{dx} = \frac{en}{k_B T} \frac{d\phi}{dx} = -\frac{enE}{k_B T}.$$

Substituting in $J_n = 0$ gives $\mu_n - eD_n/(k_B T) = 0$. The analogous hole result follows from $dp/dx = +epE/(k_B T)$. Therefore

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{k_B T}{e} = V_T \quad (21.82)$$

for nondegenerate carriers.

Example 21.26: Diffusion coefficient from mobility

An electron mobility is measured as $\mu_n = 0.12 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ at $T = 300 \text{ K}$. Estimate the diffusion coefficient in the nondegenerate model.

At 300 K, $V_T = k_B T/e \approx 0.0259 \text{ V}$. Therefore

$$D_n = \mu_n V_T = 0.12 \times 0.0259 \approx 3.1 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}.$$

The Einstein relation is a fluctuation–dissipation connection: mobility describes field-driven response, while diffusion describes random thermal spreading.

Key message

The p–n junction is fundamentally a drift–diffusion system. Majority-carrier concentration gradients drive diffusion across the interface; the resulting space charge creates an electric field that drives an opposing drift process. Equilibrium is reached when the net current vanishes, not when microscopic motion stops.

Continuity equations, recombination and carrier lifetime

Carrier populations can change because carriers flow into or out of a region, because electron–hole pairs are generated, and because they recombine. In one dimension, a convenient sign convention gives the electron and hole continuity equations

$$\frac{\partial n}{\partial t} = \frac{1}{e} \frac{\partial J_n}{\partial x} + G_n - R_n, \quad (21.83)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{e} \frac{\partial J_p}{\partial x} + G_p - R_p. \quad (21.84)$$

Here G and R are generation and recombination rates per unit volume. Under low-level injection, a simple relaxation model for excess minority carriers is

$$\frac{d\Delta n}{dt} = -\frac{\Delta n}{\tau_n}, \quad (21.85)$$

with solution

$$\Delta n(t) = \Delta n(0)e^{-t/\tau_n}. \quad (21.86)$$

The parameter τ_n is a carrier lifetime, conceptually distinct from the momentum relaxation time used in the Drude conductivity. A carrier can lose momentum many times before it recombines.

Combining diffusion with recombination introduces a diffusion length,

$$L_n = \sqrt{D_n \tau_n}, \quad L_p = \sqrt{D_p \tau_p}, \quad (21.87)$$

which estimates how far an excess minority-carrier population spreads before recombination substantially reduces it.

Caution

There are several different “times” in semiconductor physics. The momentum relaxation time controls mobility in a Drude-like description. The recombination lifetime controls how long excess carriers survive. The dielectric relaxation time controls how rapidly space charge can be screened. They should not be interchanged merely because each has units of seconds.

Quasi-Fermi levels out of equilibrium

At thermal equilibrium a single Fermi level describes the occupation statistics of the entire semiconductor. Under illumination, carrier injection or current flow, electron and hole populations can depart from mutual equilibrium. It is then useful to introduce separate *quasi-Fermi levels* F_n and F_p through

$$n = N_C \exp \left[-\frac{E_C - F_n}{k_B T} \right], \quad (21.88)$$

$$p = N_V \exp \left[-\frac{F_p - E_V}{k_B T} \right]. \quad (21.89)$$

In equilibrium $F_n = F_p = E_F$. Under nonequilibrium conditions they split, assuming carriers within each band have thermalized sufficiently to be described by their own chemical potential. Multiplying the two density expressions gives

$$np = n_i^2 \exp \left(\frac{F_n - F_p}{k_B T} \right). \quad (21.90)$$

Thus the equilibrium mass-action law must not be imposed on illuminated or injected populations. Positive quasi-Fermi separation corresponds to excess electron–hole population.

The quasi-Fermi-level separation is especially useful in optoelectronics. It measures the nonequilibrium carrier population in a way that is consistent with band diagrams. In an illuminated solar cell, terminal voltage is linked to the separation of electron and hole electrochemical potentials at the contacts.

Key message

A single flat Fermi level is an equilibrium concept. When a semiconductor is driven strongly by light or electrical injection, separate electron and hole quasi-Fermi levels provide a more useful statistical description.

Recombination mechanisms: radiative, defect-assisted and Auger

Electron–hole recombination is not one universal microscopic process. Three broad mechanisms are important in semiconductor physics.

Radiative recombination. An electron makes a transition from conduction-band states to an empty valence-band state and emits a photon. This process is central to LEDs and semiconductor lasers and is especially efficient in direct-gap materials.

Defect-assisted recombination. A defect level inside the band gap provides an intermediate state. An electron and hole can recombine through sequential capture events. This Shockley–Read–Hall mechanism often dominates in imperfect materials and at surfaces or interfaces.

Auger recombination. The energy released by recombination is transferred to a third carrier rather than to a photon. The energized carrier then thermalizes. Auger processes become important at high carrier densities in many materials.

A common phenomenological lifetime model writes

$$\frac{1}{\tau_{\text{eff}}} = \frac{1}{\tau_{\text{rad}}} + \frac{1}{\tau_{\text{SRH}}} + \frac{1}{\tau_{\text{Auger}}}, \quad (21.91)$$

when the separate channels can be treated as statistically independent first-order loss processes over the chosen operating range. Detailed recombination rates generally depend nonlinearly on carrier concentrations, so Eq. (21.91) is a conceptual simplification.

Guided checks

1. Which recombination mechanism is desired in an LED?
2. Why can crystal defects lower LED efficiency or solar-cell lifetime?
3. Why might Auger recombination become more important at very high injection levels?

Defects, traps and why real materials depart from ideal band diagrams

Perfect periodic crystals are models. Real solids contain vacancies, interstitials, impurities, dislocations, grain boundaries, surfaces and interfaces. Some defects create allowed electronic states inside the band gap. Such states can trap carriers, assist recombination, change conductivity or introduce slow time-dependent response.

In a simple Shockley–Read–Hall picture, recombination through a trap state proceeds in two steps: a carrier is captured into a defect state and later recombines with a carrier of the opposite type. The complete rate equation is beyond our present scope, but the qualitative consequence is essential: *a band gap does not imply an electronically empty energy interval in a real imperfect material.* Localized defect levels may lie within the gap.

This matters for:

- leakage current in diodes;
- lifetime and efficiency of LEDs and solar cells;
- threshold-voltage stability in transistors;
- radiation damage in detectors and electronics;
- persistent photoconductivity and sensor hysteresis.

Note

A “trap” is an electronic state associated with a defect or interface, not a literal geometric hole in the crystal. The spatial wave function may be localized, which is why the state can temporarily capture a carrier.

21.21.4 Junction electrostatics and capacitance

Abrupt p–n junction electrostatics and the built-in voltage

For an ideal abrupt junction with uniform acceptor concentration N_A on the p side and donor concentration N_D on the n side, the equilibrium carrier relations lead to the built-in potential

$$V_{\text{bi}} = V_T \ln \left(\frac{N_A N_D}{n_i^2} \right), \quad V_T = \frac{k_B T}{e}. \quad (21.92)$$

To obtain it, compare the majority-electron concentration $n_n \simeq N_D$ with the minority-electron concentration $n_p \simeq n_i^2/N_A$. A common equilibrium Fermi level and the energy shift $E_C(p) - E_C(n) = eV_{\text{bi}}$ imply $n_n/n_p = \exp(V_{\text{bi}}/V_T)$. Taking the logarithm gives Eq. (21.92).

This formula assumes nondegenerate statistics, thermal equilibrium and approximately complete dopant ionization. V_{bi} is an internal electrostatic potential difference across the depletion region; it is not a directly available terminal battery voltage at equilibrium because contact and electrochemical-potential effects must be treated consistently.

Example 21.27: Built-in potential of an idealized junction

At 300 K, consider an ideal abrupt junction with

$$N_A = 1.0 \times 10^{22} \text{ m}^{-3}, \quad N_D = 5.0 \times 10^{21} \text{ m}^{-3}, \quad n_i = 1.0 \times 10^{16} \text{ m}^{-3}.$$

Using $V_T = 0.0259 \text{ V}$,

$$V_{\text{bi}} = 0.0259 \ln \left(\frac{(1.0 \times 10^{22})(5.0 \times 10^{21})}{10^{32}} \right) \approx 0.70 \text{ V}.$$

The numerical value is illustrative because the assumed n_i and dopant concentrations define an idealized material model. The logarithm makes the built-in voltage less sensitive to doping than the carrier concentrations themselves.

In Fig. 21.22, the equilibrium Fermi level is spatially constant. The apparent bending of conduction and valence band edges through the depletion region represents the spatially varying electrostatic potential energy of electrons.

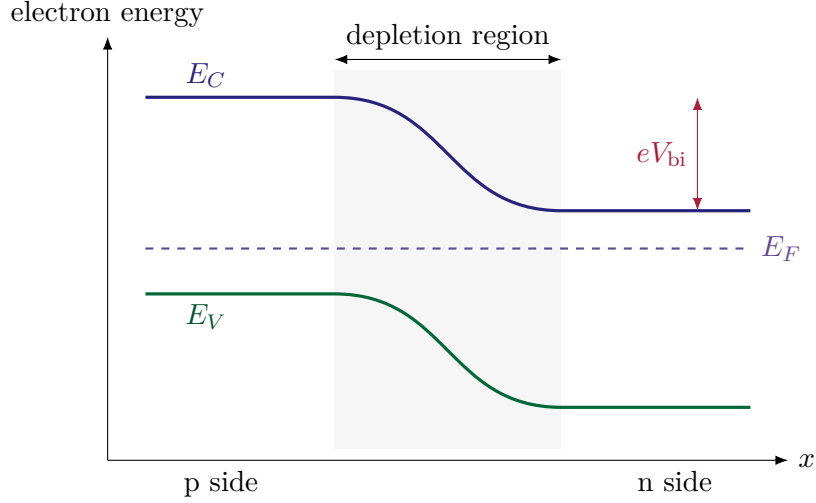


Figure 21.22: Equilibrium band bending for a nondegenerate homojunction, with p on the left and n on the right. The Fermi level stays inside both bulk gaps and is flat; E_C and E_V shift together by the electron potential energy $-e\phi$. The band gap stays constant. The indicated barrier height is eV_{bi} , and the drawing is schematic.

Depletion approximation and depletion width

The *depletion approximation* neglects mobile carriers inside a junction region of width $W = x_p + x_n$ and treats the uncovered ionized dopants as fixed uniform space charge. Charge neutrality of the depletion region requires

$$N_A x_p = N_D x_n. \quad (21.93)$$

Take p on $-x_p < x < 0$ and n on $0 < x < x_n$. With $E(-x_p) = E(x_n) = 0$, integrating $dE/dx = \rho_q/\epsilon_s$ gives

$$E(x) = \begin{cases} -eN_A(x + x_p)/\epsilon_s, & -x_p < x < 0, \\ eN_D(x - x_n)/\epsilon_s, & 0 < x < x_n, \\ 0, & \text{outside the depletion layer.} \end{cases} \quad (21.94)$$

The field is negative throughout the depletion region, consistent with its n-to-p direction. Its continuity at $x = 0$ gives Eq. (21.93). The electrostatic potential rises from p to n by the area under $-E$:

$$V_{bi} - V_a = \frac{e}{2\epsilon_s}(N_A x_p^2 + N_D x_n^2) = \frac{E_{\max} W}{2}. \quad (21.95)$$

Expressing x_p and x_n in terms of $W = x_p + x_n$ and charge balance then gives

$$W = \sqrt{\frac{2\epsilon_s}{e} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (V_{bi} - V_a)}, \quad (21.96)$$

where V_a is defined positive for forward bias in this formula. For reverse bias, $V_a < 0$ and W increases. The partition of the width is

$$x_n = \frac{N_A}{N_A + N_D} W, \quad x_p = \frac{N_D}{N_A + N_D} W. \quad (21.97)$$

Thus the depletion region extends farther into the more lightly doped side.

Example 21.28: Depletion width under zero bias and reverse bias

Take an ideal junction with $N_A = 1.0 \times 10^{22} \text{ m}^{-3}$, $N_D = 1.0 \times 10^{21} \text{ m}^{-3}$, $V_{bi} = 0.70 \text{ V}$ and $\epsilon_s = 11.7\epsilon_0$. Estimate W at zero bias and at $V_a = -5.0 \text{ V}$.

At zero bias,

$$W_0 = \sqrt{\frac{2(11.7\epsilon_0)}{e} \left(\frac{1}{10^{22}} + \frac{1}{10^{21}} \right) (0.70)} \approx 1.0 \text{ } \mu\text{m}.$$

Under 5 V reverse bias,

$$W_R = W_0 \sqrt{\frac{0.70 + 5.0}{0.70}} \approx 2.8 \text{ } \mu\text{m}.$$

Because N_D is ten times smaller than N_A , most of this width lies on the n side. The calculation illustrates why reverse-biased junctions can form useful field regions in photodiodes and particle detectors.

Guided checks

1. Which side of a strongly asymmetric junction contains most of the depletion width?
2. Why does reverse bias widen the depletion layer?
3. Which assumptions would fail first if the junction were extremely heavily doped or operated near breakdown?

Junction electric field and Poisson's equation

The electrostatic potential ϕ and electric field satisfy

$$\mathbf{E} = -\nabla\phi, \quad \nabla \cdot \mathbf{E} = \frac{\rho_q}{\epsilon_s}, \quad (21.98)$$

so that

$$\nabla^2\phi = -\frac{\rho_q}{\epsilon_s}. \quad (21.99)$$

In the one-dimensional depletion model, ρ_q is approximately $-eN_A$ on the depleted p side and $+eN_D$ on the depleted n side. Therefore $E(x)$ is piecewise linear, reaching its largest magnitude at the metallurgical junction. The potential is piecewise quadratic.

The peak field magnitude can be written in either-side form,

$$E_{\max} = \frac{eN_A x_p}{\epsilon_s} = \frac{eN_D x_n}{\epsilon_s}. \quad (21.100)$$

This connection between doping, depletion width and peak field is central to breakdown design and high-voltage semiconductor devices.

Note

The depletion approximation is a model of the space-charge region, not a claim that the mobile-carrier density is mathematically zero. It becomes useful because the fixed ionized-dopant charge dominates the net charge density over much of that region.

Figure 21.23 puts the charge, field and potential on one consistent position scale.

Junction capacitance and a useful measurement technique

The depletion charge magnitude on either side is $Q = eAN_D x_n = eAN_A x_p$. Since $x_n \propto W$ and $W \propto \sqrt{V_{bi} + V_R}$,

$$C_j = \frac{dQ}{dV_R} = \frac{Q}{2(V_{bi} + V_R)} = \frac{\epsilon_s A}{W}.$$

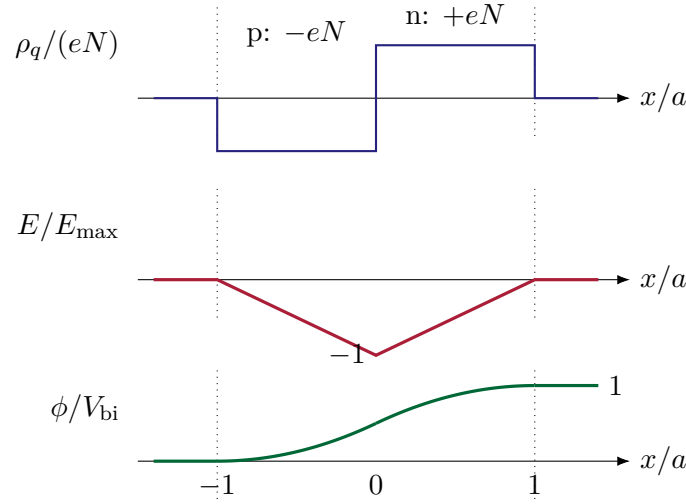


Figure 21.23: Calculated depletion profiles for a symmetric zero-bias junction, $N_A = N_D = N$ and $x_p = x_n = a$. Normalizations are $E_{\max} = eNa/\epsilon_s$ and $V_{bi} = eNa^2/\epsilon_s$. Potential increases from p to n, so electron band energies decrease, agreeing with Fig. 21.22. The charge discontinuities are the depletion approximation.

The derivative is taken with respect to reverse-voltage magnitude, making capacitance positive; it is a differential capacitance, not Q/V_R .

A reverse-biased depletion region contains separated fixed charges and behaves approximately like a voltage-dependent capacitor. With junction area A ,

$$C_j = \frac{\epsilon_s A}{W}. \quad (21.101)$$

For an abrupt junction, substituting the depletion-width formula shows

$$C_j \propto (V_{bi} - V_a)^{-1/2}. \quad (21.102)$$

Under reverse bias $V_R = -V_a > 0$,

$$\frac{1}{C_j^2} \propto V_{bi} + V_R. \quad (21.103)$$

Therefore a plot of $1/C_j^2$ against reverse-bias voltage can be approximately linear for a uniformly doped abrupt junction. More explicitly,

$$\frac{1}{C_j^2} = \frac{2}{e\epsilon_s A^2} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (V_{bi} + V_R). \quad (21.104)$$

For $N_A \gg N_D$, the measured slope $M = d(1/C_j^2)/dV_R$ gives $N_D \simeq 2/(e\epsilon_s A^2 M)$. The voltage-axis intercept is $-V_{bi}$, subject to the model assumptions, series resistance, parasitic capacitance and appropriate measurement frequency. The slope contains information about doping concentration, making capacitance–voltage measurement a standard diagnostic concept.

Example 21.29: Bias dependence of junction capacitance

If an abrupt junction has capacitance $C_0 = 80$ pF at zero bias and built-in potential $V_{bi} = 0.80$ V, estimate the capacitance at $V_R = 3.2$ V under the depletion model.

Because $C_j \propto (V_{bi} + V_R)^{-1/2}$,

$$\frac{C_R}{C_0} = \sqrt{\frac{V_{bi}}{V_{bi} + V_R}} = \sqrt{\frac{0.80}{4.00}} \approx 0.447.$$

Thus

$$C_R \approx 36 \text{ pF}.$$

The voltage-dependent junction capacitance is technologically useful in varactor devices, but real capacitance also contains package and parasitic contributions.

Computational mini-project: a self-consistent one-dimensional junction map

A useful bridge to numerical physics is to reconstruct the depletion-approximation electrostatics on a grid. The calculation can be completed with elementary arrays and cumulative integration; a full semiconductor device simulator is not required.

Model. Choose an abrupt junction at $x = 0$. Specify N_A , N_D , ε_s , V_{bi} and applied bias V_a . Calculate W , x_p and x_n . Define the charge density

$$\rho_q(x) = \begin{cases} -eN_A, & -x_p < x < 0, \\ +eN_D, & 0 < x < x_n, \\ 0, & \text{otherwise.} \end{cases} \quad (21.105)$$

Then solve numerically

$$\frac{dE}{dx} = \frac{\rho_q}{\varepsilon_s}, \quad \frac{d\phi}{dx} = -E. \quad (21.106)$$

Suggested algorithm.

1. Make a grid extending several depletion widths on each side.
2. Evaluate ρ_q at each grid point.
3. Integrate dE/dx numerically, choosing the integration constant so that E vanishes outside the depletion region.
4. Integrate $-E$ to obtain $\phi(x)$.
5. Verify that the potential rise across the depletion region equals approximately $V_{bi} - V_a$.
6. Repeat for several reverse biases and verify the predicted $W \propto \sqrt{V_{bi} + V_R}$ scaling.
7. Plot $\rho_q(x)$, $E(x)$ and $\phi(x)$ on separate axes.

Guided checks

A successful numerical junction map should pass three independent checks:

1. **Charge check:** the integrated negative and positive depletion charges are equal in magnitude.
2. **Field check:** the electric field returns to approximately zero outside the depletion region.
3. **Potential check:** the integrated field gives the required potential drop.

If one check fails, inspect units, grid spacing, sign convention and boundary conditions before refining the grid.

Advanced numerical project: coupled Poisson and carrier statistics

The depletion approximation can be relaxed by solving Poisson's equation together with equilibrium carrier statistics. In one dimension for a nondegenerate semiconductor,

$$\frac{d^2\phi}{dx^2} = -\frac{e}{\varepsilon_s} \left[p(\phi) - n(\phi) + N_D^+(x) - N_A^-(x) \right]. \quad (21.107)$$

At equilibrium the Fermi level is constant, while band edges shift with electrostatic potential. This makes n and p exponential functions of ϕ , so Eq. (21.107) is nonlinear.

A robust project can discretize the second derivative on a grid and solve the resulting nonlinear algebraic equations by Newton iteration or a damped fixed-point scheme. Start from the depletion-approximation potential as an initial guess. Compare the resulting smooth mobile-carrier profiles with the abrupt zero-mobile-carrier assumption. The objective is not industrial device accuracy; it is to see how self-consistency couples quantum statistics to electrostatics.

Method

For a nonlinear numerical device problem, monitor residual norms and physical invariants rather than trusting visual convergence. Check equilibrium current, charge neutrality far from the junction, potential drop, positivity of carrier concentrations and grid-spacing dependence.

21.21.5 Injection, the diode law and switching

Where the Shockley diode equation comes from

The ideal diode equation can be understood as a boundary-condition result for minority-carrier diffusion. Under forward bias V , the equilibrium barrier is reduced. In the nondegenerate approximation, minority-carrier concentrations at the depletion edges are multiplied by the Boltzmann factor

$$\exp\left(\frac{eV}{k_B T}\right) = \exp\left(\frac{V}{V_T}\right). \quad (21.108)$$

The injected excess minority carriers diffuse into the quasi-neutral regions and recombine over diffusion lengths L_n and L_p . For the long-neutral-region, low-injection model, define $n_{p0} = n_i^2/N_A$ and $p_{n0} = n_i^2/N_D$. Boundary injection gives

$$\Delta n_p(-x_p) = n_{p0}(e^{V/V_T} - 1), \quad \Delta p_n(x_n) = p_{n0}(e^{V/V_T} - 1).$$

In a field-free neutral p region, continuity and diffusion combine to give $D_n d^2\Delta n_p/dx^2 - \Delta n_p/\tau_n = 0$. Requiring the excess population to vanish far from the junction selects

$$\Delta n_p(x) = \Delta n_p(-x_p)e^{(x+x_p)/L_n}, \quad x < -x_p.$$

The analogous n-side profile is $\Delta p_n(x) = \Delta p_n(x_n)e^{-(x-x_n)/L_p}$ for $x > x_n$. Differentiating at the depletion edges gives

$$J_n(-x_p) = \frac{eD_n n_{p0}}{L_n}(e^{V/V_T} - 1), \quad J_p(x_n) = \frac{eD_p p_{n0}}{L_p}(e^{V/V_T} - 1).$$

Both conventional currents point from p to n under forward bias, even though the injected electrons travel the opposite way. Neglecting recombination in the depletion region, their sum is the terminal current density. Multiplication by junction area A gives

$$I = I_S \left[\exp\left(\frac{V}{V_T}\right) - 1 \right], \quad (21.109)$$

with ideal saturation current

$$I_S = eA \left(\frac{D_n n_i^2}{L_n N_A} + \frac{D_p n_i^2}{L_p N_D} \right) \quad (21.110)$$

for an abrupt homojunction with uniform nondegenerate dopings, complete ionization, low injection, negligible depletion-region recombination, ohmic contacts and neutral-region lengths much greater than L_n and L_p . V is the junction voltage; a terminal series-resistance drop must be removed before using it. Finite neutral-region lengths change the prefactor.

Equation (21.110) reveals why I_S can depend so strongly on temperature: it contains n_i^2 . It also explains why device geometry, diffusion and recombination matter even though the compact diode equation appears to contain only I_S , V and T .

A common empirical extension introduces an ideality factor n_d :

$$I \simeq I_S \left[\exp\left(\frac{V}{n_d V_T}\right) - 1 \right]. \quad (21.111)$$

This is useful over limited ranges but is not a universal law. Series resistance, recombination in the depletion region, high-level injection, tunnelling, self-heating and breakdown can all change the measured curve.

A deeper look at the diode equation through diffusion lengths

The ideal saturation current in Eq. (21.110) can be rewritten using $L = \sqrt{D\tau}$:

$$I_S = eA n_i^2 \left[\frac{\sqrt{D_n/\tau_n}}{N_A} + \frac{\sqrt{D_p/\tau_p}}{N_D} \right]. \quad (21.112)$$

This form shows several useful trends within the ideal model:

- larger intrinsic concentration increases I_S strongly through n_i^2 ;
- heavier doping of a neutral region suppresses the corresponding minority-carrier contribution;
- shorter minority-carrier lifetime can increase the spatial gradient needed to remove injected carriers and changes the saturation-current scale;
- device area scales the current linearly if current density is uniform.

The equation should not be extrapolated into high injection, strong recombination in the depletion region or degenerate doping.

Diode charge storage, diffusion capacitance and switching

A forward-biased diode stores excess minority carriers in the quasi-neutral regions. If the forward current changes rapidly, these stored carriers cannot disappear instantaneously. This produces a dynamic response beyond the static exponential I – V law.

The small-signal incremental conductance of an ideal forward-biased diode is

$$g_d = \frac{dI}{dV} \approx \frac{I}{n_d V_T}, \quad r_d = \frac{1}{g_d} \approx \frac{n_d V_T}{I}. \quad (21.113)$$

Stored minority charge can be represented by a *diffusion capacitance*

$$C_d = \frac{dQ_{\text{stored}}}{dV}, \quad (21.114)$$

which often grows strongly with forward current. Under reverse bias the depletion capacitance is usually the more important junction capacitance.

If a conducting diode is suddenly switched to reverse bias, current can temporarily flow in the reverse direction while stored minority charge is removed. This is the origin of reverse-recovery behaviour in ordinary p–n diodes. Fast switching circuits may therefore use specially designed fast-recovery, Schottky or wide-band-gap devices depending on voltage and current requirements.

Example 21.30: Small-signal diode resistance

At $T = 300$ K, a diode carries $I = 2.0$ mA with ideality factor $n_d = 1.5$. Using $V_T = 25.9$ mV,

$$r_d \approx \frac{n_d V_T}{I} = \frac{(1.5)(25.9 \times 10^{-3})}{2.0 \times 10^{-3}} \approx 19.4 \text{ } \Omega.$$

This is an incremental resistance for small ac variations around the operating point. It is not equal to the dc ratio V/I .

Schottky junctions as a contrasting rectifier concept

Rectification does not require a p–n junction. A metal–semiconductor contact can form an energy barrier called a Schottky barrier. Properly designed Schottky diodes conduct through majority carriers rather than relying on minority-carrier injection across a p–n junction. This can reduce stored charge and support fast switching.

A simplified thermionic-emission current law has the same exponential structure as a diode equation,

$$I \sim I_S \left[e^{eV/(n_d k_B T)} - 1 \right], \quad (21.115)$$

but the saturation-current physics and barrier formation are different. Real Schottky contacts also show image-force effects, interface states, series resistance and tunnelling.

This comparison is pedagogically useful: the terminal behaviour “rectifies current” does not uniquely determine the microscopic device mechanism.

21.21.6 Optical and radiation devices

Optical absorption, radiative recombination and direct versus indirect gaps

Photon absorption across a semiconductor band gap requires energy conservation,

$$hf \geq E_g, \quad (21.116)$$

but crystal momentum also matters. Photons in the optical range carry relatively little crystal momentum compared with the scale of the Brillouin zone. In a *direct-gap* semiconductor, the valence-band maximum and conduction-band minimum occur at the same crystal wave vector, so an interband transition can satisfy the momentum requirement without a phonon. In an *indirect-gap* semiconductor, a phonon is generally required to supply or absorb the missing crystal momentum.

This distinction helps explain why direct-gap materials are efficient light emitters: electron–hole recombination can readily produce a photon. It also explains why a band gap alone does not determine LED efficiency.

For an idealized direct transition, a characteristic emitted wavelength is

$$\lambda \approx \frac{hc}{E_g}. \quad (21.117)$$

The emitted spectrum has finite width, and heterostructure design, carrier confinement, temperature and nonradiative recombination all affect real devices.

Guided checks

1. Why is $hf \geq E_g$ necessary but not always sufficient for strong optical absorption?
2. Why does a material with a suitable band gap not automatically make an efficient LED?
3. In a photodiode, why can a wide depletion region improve collection of carriers generated inside that region?

Photodiode response and quantum efficiency

If incident optical power is P_{opt} and photon energy is hf , the incident photon rate is

$$\dot{N}_\gamma = \frac{P_{\text{opt}}}{hf}. \quad (21.118)$$

Define an external quantum efficiency η_q as the number of collected electron charges per incident photon. The photocurrent magnitude is then

$$I_{\text{ph}} = e\eta_q \frac{P_{\text{opt}}}{hf}. \quad (21.119)$$

The responsivity is

$$\mathcal{R} = \frac{I_{\text{ph}}}{P_{\text{opt}}} = \eta_q \frac{e}{hf} = \eta_q \frac{e\lambda}{hc}. \quad (21.120)$$

This simple expression assumes each collected photo-generated carrier pair contributes one electron charge to the external circuit and neglects internal gain mechanisms.

Example 21.31: Photodiode responsivity

A photodiode has external quantum efficiency $\eta_q = 0.75$ at wavelength $\lambda = 850$ nm. Estimate its responsivity.

Using Eq. (21.120),

$$\mathcal{R} = 0.75 \frac{e(850 \times 10^{-9} \text{ m})}{hc} \approx 0.51 \text{ A W}^{-1}.$$

Therefore 1.0 mW of optical power would ideally produce about 0.51 mA of photocurrent under the stated collection efficiency.

In fast optical receivers, reverse bias can reduce junction capacitance and speed carrier collection, but it can also increase dark current and noise. Detector design is therefore a multi-objective engineering problem.

Solar-cell equation, open-circuit voltage and fill factor

A simple illuminated-junction model combines a light-generated current I_L with the diode current:

$$I = I_L - I_S \left[\exp\left(\frac{V}{n_d V_T}\right) - 1 \right]. \quad (21.121)$$

The sign convention is chosen so that positive I represents useful current delivered to a load. At short circuit, $V = 0$ and $I_{\text{sc}} \approx I_L$. At open circuit, $I = 0$, so

$$V_{\text{oc}} = n_d V_T \ln\left(1 + \frac{I_L}{I_S}\right). \quad (21.122)$$

The maximum-power point (V_{mp}, I_{mp}) defines the fill factor

$$FF = \frac{V_{mp} I_{mp}}{V_{oc} I_{sc}}. \quad (21.123)$$

If incident optical power on the cell is P_{in} , the conversion efficiency is

$$\eta = \frac{V_{mp} I_{mp}}{P_{in}}. \quad (21.124)$$

Example 21.32: Open-circuit voltage in an idealized solar cell

At 300 K, an idealized cell has $n_d = 1.2$, $I_L = 35$ mA and $I_S = 2.0 \times 10^{-10}$ A. Estimate V_{oc} .

$$V_{oc} = (1.2)(0.0259) \ln \left(1 + \frac{0.035}{2.0 \times 10^{-10}} \right) \approx 0.59 \text{ V}.$$

This is a terminal voltage derived from a nonequilibrium illuminated device. It should not be confused with simply “measuring the built-in voltage” of an equilibrium p–n junction.

Caution

The built-in electrostatic potential of an equilibrium junction and the open-circuit voltage of an illuminated solar cell are different physical quantities. A voltmeter connected to an equilibrium junction does not simply read V_{bi} .

Radiation damage and semiconductor particle detectors

Reverse-biased semiconductor junctions are widely used as ionizing-radiation detectors because energy deposited in the material can create electron–hole pairs that are separated by the electric field. A simplified signal-charge estimate is

$$N_{eh} \approx \frac{E_{dep}}{w}, \quad Q \approx eN_{eh}, \quad (21.125)$$

where w is the mean energy required to create a collectable electron–hole pair. Importantly, w is generally larger than the band gap because deposited energy is also lost to lattice excitations and other processes.

Radiation can create lattice defects that act as generation–recombination centers or trapping sites. Detector performance can then change through increased leakage current, altered effective doping, reduced charge-collection efficiency and changes in depletion voltage. This is a direct bridge between semiconductor physics and nuclear/particle instrumentation.

Example 21.33: Signal charge from specified pair-creation energy

A hypothetical semiconductor detector receives an energy deposition of $E_{dep} = 120$ keV. If the mean pair-creation energy is specified as $w = 3.6$ eV, estimate the number of electron–hole pairs and the total collected charge if collection is complete.

$$N_{eh} = \frac{120 \times 10^3}{3.6} \approx 3.33 \times 10^4,$$

so

$$Q = eN_{eh} \approx (1.602 \times 10^{-19})(3.33 \times 10^4) \approx 5.3 \times 10^{-15} \text{ C}.$$

The charge is only a few femtocoulombs, illustrating why low-noise front-end electronics are essential in radiation detectors.

Advanced worked synthesis: from doping to a reverse-biased detector junction

Example 21.34: A multi-stage detector-junction calculation

An idealized abrupt semiconductor junction at 300 K has

$$N_A = 2.0 \times 10^{22} \text{ m}^{-3}, \quad N_D = 5.0 \times 10^{20} \text{ m}^{-3}, \quad n_i = 1.0 \times 10^{16} \text{ m}^{-3},$$

with $\varepsilon_s = 11.7\varepsilon_0$ and area $A = 1.0 \text{ mm}^2$. Estimate (a) the built-in potential, (b) the depletion width at $V_R = 20 \text{ V}$, (c) the fraction of the depletion width on the n side, and (d) the junction capacitance.

(a) Built-in potential.

$$V_{\text{bi}} = 0.0259 \ln \left(\frac{(2.0 \times 10^{22})(5.0 \times 10^{20})}{10^{32}} \right) \approx 0.66 \text{ V}.$$

(b) Reverse-biased depletion width. With $V_a = -V_R$,

$$W = \sqrt{\frac{2\varepsilon_s}{e} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (V_{\text{bi}} + V_R)} \approx 7.4 \text{ } \mu\text{m}.$$

(c) Width partition.

$$\frac{x_n}{W} = \frac{N_A}{N_A + N_D} \approx 0.976.$$

Thus nearly the entire depletion region lies in the lightly doped n side.

(d) Junction capacitance. Because $A = 1.0 \times 10^{-6} \text{ m}^2$,

$$C_j = \frac{\varepsilon_s A}{W} \approx 1.4 \times 10^{-11} \text{ F} = 14 \text{ pF}.$$

Interpretation. Light doping permits a relatively wide depletion region, which is useful for collecting charge deposited by radiation and lowers capacitance, benefiting electronic noise and timing. The same design must also withstand the peak electric field and maintain sufficiently low leakage current.

21.21.7 Gate control and digital energy

MOS capacitors and field-effect control

The conceptual core of a MOSFET is already present in a metal–oxide–semiconductor (MOS) capacitor. A conducting gate is separated from a semiconductor by a thin insulator. Changing the gate voltage changes the electrostatic potential near the semiconductor surface and therefore changes the local carrier population without requiring direct gate current through an ideal oxide.

For a p-type substrate, increasingly positive gate voltage can move the surface through three qualitative regimes:

1. **Accumulation** for negative gate bias: holes are attracted toward the surface.
2. **Depletion**: holes are repelled, exposing ionized acceptors.
3. **Inversion**: sufficiently strong positive gate bias creates an electron-rich surface layer, effectively an n-type channel within a p-type substrate.

The gate therefore acts electrostatically. This is the key to the extraordinarily low static input current of an ideal insulated-gate transistor.

Figure 21.24 locates the inversion layer at the semiconductor–oxide interface.

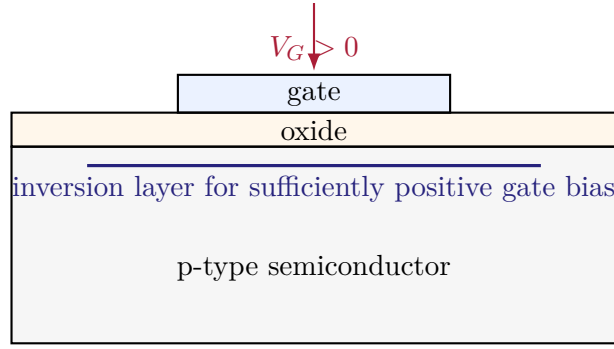


Figure 21.24: Schematic MOS capacitor in inversion. The drawing emphasizes electrostatic carrier control; real devices require detailed oxide, interface and work-function physics.

Long-channel MOSFET model and its limits

In an ideal n-channel enhancement MOSFET, a gate voltage above a threshold V_{th} creates a conducting inversion channel between source and drain. In the gradual-channel approximation, the inversion sheet-charge magnitude at local channel potential $v(x)$ is $|Q_{inv}(x)| \simeq C_{ox}[V_{GS} - V_{th} - v(x)]$. The drift current is $I_D = \mu_n W_{ch} |Q_{inv}(x)| dv/dx$. Integrating from source ($v = 0$) to drain ($v = V_{DS}$) over length L_{ch} gives

$$I_D L_{ch} = \mu_n W_{ch} C_{ox} \int_0^{V_{DS}} (V_{GS} - V_{th} - v) dv.$$

Thus the drain current in the linear region is

$$I_D = \mu_n C_{ox} \frac{W_{ch}}{L_{ch}} \left[(V_{GS} - V_{th}) V_{DS} - \frac{V_{DS}^2}{2} \right], \quad (21.126)$$

for $0 \leq V_{DS} < V_{GS} - V_{th}$, where C_{ox} is oxide capacitance per unit area and W_{ch}/L_{ch} is the channel width-to-length ratio. At $V_{DS} = V_{GS} - V_{th}$, the inversion-charge approximation reaches zero at the drain end (pinch-off). Additional drain voltage falls mainly beyond that end in the ideal long-channel description; carriers continue reaching the drain, so saturation does not mean zero current. In the ideal saturation regime,

$$I_D \simeq \frac{1}{2} \mu_n C_{ox} \frac{W_{ch}}{L_{ch}} (V_{GS} - V_{th})^2, \quad (21.127)$$

for $V_{DS} \gtrsim V_{GS} - V_{th}$.

Example 21.35: Ideal long-channel MOSFET current

An idealized n-channel MOSFET has

$$\mu_n C_{ox} \frac{W_{ch}}{L_{ch}} = 2.0 \text{ mA V}^{-2}, \quad V_{th} = 0.80 \text{ V}.$$

For $V_{GS} = 2.0 \text{ V}$ and $V_{DS} = 2.0 \text{ V}$, the overdrive is $V_{GS} - V_{th} = 1.2 \text{ V}$, and V_{DS} is large enough for the ideal saturation condition. Thus

$$I_D \simeq \frac{1}{2} (2.0 \text{ mA V}^{-2}) (1.2 \text{ V})^2 = 1.44 \text{ mA}.$$

This is a pedagogical long-channel result, not a precision model for a modern nanometre-scale transistor.

Modern transistors exhibit channel-length modulation, mobility degradation, velocity saturation, drain-induced barrier lowering, source/drain series resistance, leakage, tunnelling and self-heating. The square-law model remains valuable because it exposes the idea of transconductance: a voltage at one terminal controls current between two others.

Caution

In this section V_{th} denotes the MOSFET threshold voltage. It is deliberately distinguished from the thermal voltage $V_T = k_B T / e$. Both are measured in volts but arise from different physics.

MOS capacitance–voltage behaviour and interface quality

The MOS capacitor introduced earlier can be studied experimentally by measuring capacitance as a function of gate voltage. In accumulation, the semiconductor surface supplies abundant majority carriers, so the measured capacitance can approach the oxide capacitance. In depletion, the space-charge region acts like an additional series capacitance and the total capacitance falls. In inversion, the response depends on measurement frequency because minority carriers may or may not follow the applied ac signal quickly enough.

For oxide thickness t_{ox} and permittivity ϵ_{ox} , the ideal oxide capacitance per unit area is

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}. \quad (21.128)$$

Thinner oxide increases gate electrostatic control but also raises electric field for a given voltage and increases sensitivity to tunnelling and reliability constraints.

Charges trapped in the oxide or at the semiconductor–oxide interface can shift a measured capacitance–voltage curve and degrade transistor characteristics. C–V measurements are therefore a window into both electrostatics and material quality.

CMOS logic: complementary field-effect devices

Complementary metal–oxide–semiconductor (CMOS) logic combines n-channel and p-channel MOSFETs so that, in an ideal static logic state, one transistor is off and there is no direct dc path from supply to ground. This dramatically reduces static power compared with logic families that require continuous current.

A conceptual CMOS inverter has one pMOS pull-up device and one nMOS pull-down device. A low input turns the pMOS on and the nMOS off, driving the output high. A high input reverses the roles and drives the output low. During transitions both devices can conduct briefly and load capacitances must be charged or discharged.

The dynamic energy scale per full $0 \rightarrow V_{DD} \rightarrow 0$ switching event is of order

$$E_{sw} \sim C_L V_{DD}^2, \quad (21.129)$$

leading to

$$P_{dyn} \sim \alpha C_L V_{DD}^2 f, \quad (21.130)$$

where α is the average number of $0 \rightarrow 1$ charging events per clock cycle. Charging draws $C_L V_{DD}^2$ from a fixed-voltage supply, stores half in the capacitor and dissipates half in the charging path; discharge dissipates the stored half. This definition avoids a factor-of-two ambiguity in the activity factor. This scaling explains why supply-voltage reduction is such a powerful lever for energy-efficient computing.

Figure 21.25 separates gate control from the supply-to-output current paths.

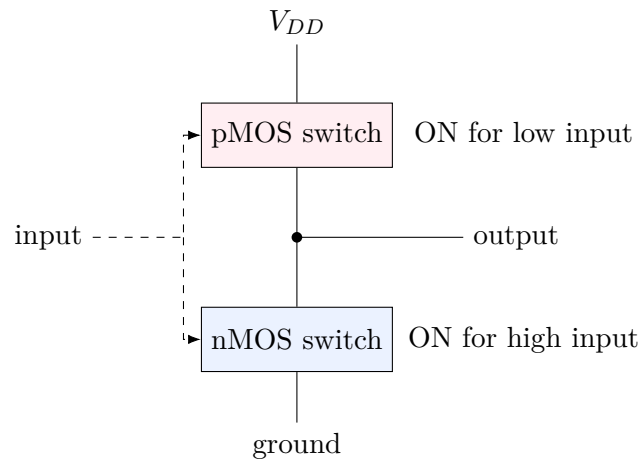


Figure 21.25: Functional CMOS inverter diagram. Solid paths carry supply and load current; dashed arrows convey gate control and are not electrical shorts to source or drain. Complementary switch states connect the output to the appropriate rail. This is a block representation, not a transistor circuit symbol.

Scaling limits in modern transistors

Making a transistor smaller can reduce switching charge and allow more devices per chip, but electrostatic and quantum limitations become more severe. Short channels weaken gate control because source and drain fields penetrate more strongly into the channel. Thin oxides improve capacitive control but increase electric field and quantum tunnelling. High current densities create heating and electromigration concerns in interconnects.

These constraints motivate multiple engineering strategies: three-dimensional gate geometries, high-permittivity dielectrics, strained channels, new semiconductor materials, improved thermal paths and architectural changes that reduce energy per operation. The microscopic physics of bands and carriers therefore remains directly connected to computer-system performance.

Key message

Transistor scaling is not simply geometric miniaturization. Every dimension change alters electrostatics, capacitance, scattering, leakage, heat flow and eventually quantum transport.

21.21.8 Materials and alternative conduction regimes

Band engineering and heterojunctions

A semiconductor device need not be made from one uniform material. Joining two semiconductors with different band gaps and electron affinities produces a *heterojunction*. Band offsets can confine electrons or holes, guide recombination into selected regions, reduce leakage or create high-mobility channels. This design philosophy is called *band engineering*.

A qualitative heterojunction band diagram contains discontinuities in E_C and E_V . Their values depend on the material pair and interface. Device engineers exploit these offsets in light-emitting diodes, laser diodes, high-electron-mobility transistors, infrared detectors and multijunction solar cells.

Figure 21.26 makes the conduction- and valence-band discontinuities explicit.

Band engineering is a useful reminder that “the band gap” is not the only device-relevant material parameter. Effective masses, mobilities, dielectric response, band offsets, defect levels, optical transition strengths and thermal properties can all matter.

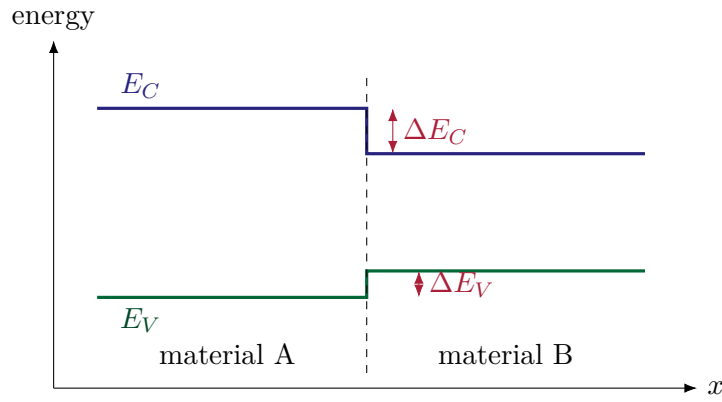


Figure 21.26: Qualitative band offsets at a heterojunction. The offset values and band bending are material- and interface-dependent; the figure shows only the basic design idea.

Breakdown, field concentration and power-device design

Under sufficiently strong reverse bias, a p–n junction leaves the regime of the ideal reverse-saturation current. Two idealized breakdown mechanisms are commonly distinguished:

- **Zener tunnelling:** in a very strongly doped, narrow depletion region, a strong electric field can enable interband tunnelling.
- **Avalanche breakdown:** carriers accelerated by the field can gain enough energy between collisions to create additional electron–hole pairs through impact ionization, producing carrier multiplication.

Which mechanism dominates depends on material, doping, junction profile and temperature.

Power semiconductor design often deliberately lowers the doping in a drift region so the depletion layer can spread over a larger distance, reducing the peak field for a given blocking voltage. This improves voltage capability but raises on-state resistance. The result is a fundamental engineering trade-off between conduction loss, blocking voltage, switching speed, device area and thermal management.

Key message

High-voltage device design is an electrostatics problem and a transport problem at the same time. Doping controls both the available mobile carriers and how space charge supports an electric field under reverse bias.

Thermoelectric transport: coupling heat and charge

A temperature gradient can drive an electrical response. In the Seebeck effect, an open-circuit temperature difference ΔT produces a voltage difference approximately described by

$$\Delta V = -S\Delta T, \quad (21.131)$$

Here $\Delta T = T_{\text{hot}} - T_{\text{cold}}$ and $\Delta V = \phi_{\text{hot}} - \phi_{\text{cold}}$ define the sign convention for S , the Seebeck coefficient. The sign of S often contains information about the dominant carrier type, though a rigorous interpretation can require band and scattering details.

Thermoelectric generators convert heat-flow conditions into electrical power; Peltier devices use electrical current to pump heat. A commonly used dimensionless measure of thermoelectric material performance is

$$ZT = \frac{S^2 \sigma T}{\kappa}, \quad (21.132)$$

where κ is thermal conductivity. High electrical conductivity, a large Seebeck coefficient and low thermal conductivity are all desirable, but these properties are not independent. Materials engineering is therefore a compromise rather than a single-parameter optimization.

Example 21.36: Simple Seebeck voltage estimate

A thermoelectric element has an effective Seebeck coefficient $S = 200 \mu\text{V K}^{-1}$ over a small temperature interval. If its ends differ in temperature by $\Delta T = 50 \text{ K}$, the open-circuit voltage magnitude is

$$|\Delta V| = |S|\Delta T = (200 \times 10^{-6})(50) = 1.0 \times 10^{-2} \text{ V}.$$

A practical module connects many thermoelectric legs electrically and thermally to obtain useful voltage, current and heat-flow characteristics.

From diffusive transport to ballistic quantum conductance

The Drude and drift–diffusion pictures assume that scattering is important over the device length. If a conductor becomes shorter than relevant mean free paths, transport can approach the ballistic regime. For elastic transport between ideal reservoirs in linear response, at zero temperature or with transmissions nearly constant over the thermal energy window, a compact quantum description is provided by the Landauer formula

$$G = \frac{2e^2}{h} \sum_n \mathcal{T}_n, \quad (21.133)$$

where $\mathcal{T}_n$ is the transmission probability of transport channel n . The factor

$$G_0 = \frac{2e^2}{h} \approx 7.75 \times 10^{-5} \text{ S} \quad (21.134)$$

corresponds to the conductance quantum for a spin-degenerate channel in this convention.

Example 21.37: Two quantum channels

A short coherent conductor supports two spin-degenerate channels with transmissions $\mathcal{T}_1 = 0.90$ and $\mathcal{T}_2 = 0.40$. Its Landauer conductance is

$$G = G_0(0.90 + 0.40) \approx (7.75 \times 10^{-5})(1.30) \approx 1.01 \times 10^{-4} \text{ S}.$$

The corresponding two-terminal resistance is about 9.9 k Ω in this idealized description. This resistance is not interpreted as ordinary bulk resistivity times length divided by area; contacts and quantum transmission are essential parts of the model.

This perspective becomes relevant in nanoscale electronics, quantum point contacts, nanowires and mesoscopic systems. It also shows that “conductivity” is not always the most natural material descriptor when the device is comparable to microscopic transport lengths.

Superconductivity as a qualitatively different conduction state

A superconductor below its critical temperature can exhibit zero dc electrical resistance and expulsion of magnetic flux from its bulk in the Meissner state. Zero resistance alone is not a complete definition: the magnetic response distinguishes superconductivity from a hypothetical perfect classical conductor.

Conventional superconductivity is a many-body quantum state involving correlated electron pairs and an energy gap in the excitation spectrum. In the weak-coupling Bardeen–Cooper–Schrieffer

(BCS) limit,

$$2\Delta(0) \approx 3.52k_B T_c, \quad (21.135)$$

where $\Delta(0)$ is the zero-temperature superconducting gap and T_c is the critical temperature. This relation is not universal for all superconductors.

Real superconductors also have critical magnetic fields and critical current densities. Engineering applications therefore require cryogenics, thermal stability, magnetic design and mechanical protection in addition to the microscopic physics.

Example 21.38: BCS gap scale

For a hypothetical weak-coupling conventional superconductor with $T_c = 9.0$ K,

$$\Delta(0) \approx 1.76k_B T_c \approx 1.76(8.617 \times 10^{-5} \text{ eV K}^{-1})(9.0 \text{ K}) \approx 1.36 \text{ meV}.$$

The small energy scale helps explain why low temperatures are required to protect the paired state against thermal excitations.

Pedagogical boundary

High-temperature superconductivity, microscopic pairing mechanisms in unconventional materials, vortex matter and Josephson-device physics belong to more advanced condensed-matter courses. The present connection is included to emphasize that not all conduction phenomena can be reduced to independent carriers scattering through a lattice.

21.21.9 Measurement, noise and model validation

Experimental parameter extraction from transport data

A measurement rarely returns “mobility” or “carrier density” directly. Instead, voltages, currents, dimensions, magnetic fields and temperatures are measured and combined through a model. For a rectangular sample of length L , cross-sectional area A , current I and voltage drop V in the ohmic regime,

$$\rho = \frac{VA}{IL}, \quad \sigma = \frac{IL}{VA}. \quad (21.136)$$

In a one-carrier Hall model,

$$R_H = \frac{E_y}{J_x B_z} = \frac{1}{nq}. \quad (21.137)$$

Combining $|R_H|$ with conductivity yields a Hall mobility estimate

$$\mu_H \simeq |R_H|\sigma \quad (21.138)$$

when the single-carrier model is appropriate.

Example 21.39: Combining resistivity and Hall data

A sample has measured conductivity $\sigma = 250 \text{ S m}^{-1}$ and Hall coefficient $R_H = -3.0 \times 10^{-4} \text{ m}^3 \text{ C}^{-1}$. In the one-carrier model,

$$n = \frac{1}{e|R_H|} \approx 2.08 \times 10^{22} \text{ m}^{-3},$$

and

$$\mu_H = |R_H|\sigma = (3.0 \times 10^{-4})(250) = 7.5 \times 10^{-2} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}.$$

The negative Hall coefficient indicates electron-dominated transport under this sign convention. In multicarrier materials, the same algebra can be misleading because electrons and holes contribute simultaneously.

Two-carrier Hall effect: when $R_H = 1/(nq)$ fails

The one-carrier Hall coefficient is exceptionally useful because it directly converts a voltage measurement into a carrier-density estimate. In a semiconductor where both electrons and holes contribute appreciably, the situation changes. In the weak-field classical transport model, the Hall coefficient is

$$R_H = \frac{p\mu_p^2 - n\mu_n^2}{e(p\mu_p + n\mu_n)^2}. \quad (21.139)$$

The conductivity remains

$$\sigma = e(p\mu_p + n\mu_n). \quad (21.140)$$

The squared mobilities in the numerator mean that the Hall sign need not simply identify the carrier type with the larger concentration. A lower-density carrier with much higher mobility can dominate the Hall response.

Example 21.40: Hall sign in a two-carrier semiconductor

Let

$$\begin{aligned} n &= 5.0 \times 10^{20} \text{ m}^{-3}, & p &= 2.0 \times 10^{21} \text{ m}^{-3}, \\ \mu_n &= 0.15 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}, & \mu_p &= 0.030 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}. \end{aligned}$$

Although holes are four times more numerous, compare the Hall numerator terms:

$$\begin{aligned} p\mu_p^2 &= (2.0 \times 10^{21})(0.030)^2 = 1.8 \times 10^{18}, \\ n\mu_n^2 &= (5.0 \times 10^{20})(0.15)^2 = 1.125 \times 10^{19}. \end{aligned}$$

The electron term is larger, so $R_H < 0$ in this simplified model. The example shows why Hall data should not be interpreted with $R_H = \pm 1/(en)$ unless one carrier type clearly dominates.

Uncertainty propagation and model adequacy

Suppose the conductivity of a bar-shaped sample is inferred from

$$\sigma = \frac{IL}{VA}.$$

If the uncertainties are independent and small, a first-order fractional-uncertainty estimate is

$$\left(\frac{u_\sigma}{\sigma}\right)^2 \simeq \left(\frac{u_I}{I}\right)^2 + \left(\frac{u_L}{L}\right)^2 + \left(\frac{u_V}{V}\right)^2 + \left(\frac{u_A}{A}\right)^2. \quad (21.141)$$

This statistical propagation is only one part of measurement quality. Systematic effects can dominate: contact resistance, sample heating, geometric nonuniformity, probe misalignment, instrument offsets and non-ohmic behaviour all create model error.

For an exponential diode model, direct least-squares fitting in linear current can overweight the highest-current points. Fitting $\ln(I/I_{\text{ref}})$ versus V can be useful in the exponential region, but only after excluding data close to the noise floor, reverse-current regime and series-resistance-dominated regime. A physically meaningful fit therefore requires *regime selection*, not merely an optimizer.

Method

A good device-data workflow has four layers: (1) inspect the raw I – V or R – T curve; (2) identify a model-valid regime; (3) fit parameters with uncertainty estimates; (4) inspect residuals and test whether the extracted parameters remain stable when the fit window is changed.

Electrical noise: thermal noise and shot noise

A device can produce the correct mean current and still be unsuitable as a sensor because of fluctuations. Two basic noise scales connect naturally to this chapter.

A resistor at temperature T exhibits Johnson–Nyquist voltage noise. Over bandwidth Δf , the mean-square open-circuit noise voltage is

$$\langle v_n^2 \rangle = 4k_B T R \Delta f. \quad (21.142)$$

The corresponding root-mean-square voltage is

$$v_{n,\text{rms}} = \sqrt{4k_B T R \Delta f}. \quad (21.143)$$

For statistically independent charge-transfer events producing average current I , ideal shot-current noise over bandwidth Δf has

$$i_{n,\text{rms}} = \sqrt{2eI\Delta f}. \quad (21.144)$$

Real semiconductor devices may also contain generation–recombination noise and $1/f$ noise.

Example 21.41: Shot-noise scale in a photodiode

A photodiode carries total dc current $I = 10 \mu\text{A}$ in a measurement bandwidth $\Delta f = 100 \text{ kHz}$. The ideal shot-noise current is

$$i_{n,\text{rms}} = \sqrt{2eI\Delta f} \approx \sqrt{2(1.602 \times 10^{-19})(10^{-5})(10^5)} \approx 5.7 \times 10^{-10} \text{ A}.$$

This is about 0.57 nA rms. Receiver design compares such noise contributions with the signal current and amplifier noise rather than considering current magnitude alone.

Caution

Noise formulas require a bandwidth. Quoting a noise voltage or current without specifying the frequency range is generally incomplete.

Advanced engineering case study: noise-limited optical detection

Consider a photodiode with responsivity $\mathcal{R}$ receiving optical power P . The signal current is

$$I_{\text{sig}} = \mathcal{R}P.$$

If the total average detector current contributing to shot noise is I_{tot} and the effective transimpedance resistance is R , two simplified current-noise contributions referred to the input are

$$i_{\text{shot}} = \sqrt{2eI_{\text{tot}}\Delta f}, \quad (21.145)$$

$$i_{\text{thermal}} \sim \sqrt{\frac{4k_B T \Delta f}{R}}. \quad (21.146)$$

A crude signal-to-noise estimate is

$$\text{SNR} \sim \frac{I_{\text{sig}}}{\sqrt{i_{\text{shot}}^2 + i_{\text{thermal}}^2 + i_{\text{amp}}^2}}, \quad (21.147)$$

Here Δf is the one-sided equivalent noise bandwidth; the terms are rms values and are assumed uncorrelated. The shot-noise term assumes independent Poisson arrival events, and the resistor term uses the classical white-noise limit. The quantity i_{amp} represents amplifier noise referred to the input over the same bandwidth.

The formula exposes several competing design choices. Increasing bandwidth increases noise. Increasing reverse bias can reduce capacitance and improve speed but may raise dark current. Increasing transimpedance gain raises signal voltage but can reduce bandwidth and affect stability. Device design and circuit design are inseparable.

Computational mini-project: fitting a real-looking diode curve

A second mini-project illustrates why model hierarchy matters. Generate synthetic current values with a series-resistance model

$$I = I_S \left\{ \exp \left[\frac{V - IR_s}{n_d V_T} \right] - 1 \right\}, \quad (21.148)$$

which is implicit because I appears on both sides. Solve it numerically at each applied voltage, add modest random measurement noise, and then attempt three fits:

1. ideal diode with $n_d = 1$ and $R_s = 0$;
2. ideality-factor model with fitted n_d but $R_s = 0$;
3. full chosen model with fitted I_S , n_d and R_s .

Compare residuals and parameter stability. Students should observe that an apparently excellent straight-line semilog fit over a narrow range can still fail to predict the high-current regime.

Note

When generating synthetic data for coursework, state clearly that the data are simulated. Do not present fabricated points as measurements from a real device or laboratory.

21.22 Practice problems

State your physical model, convert units before substitution, and interpret each result. All material parameters specified below are local exercise inputs, not handbook values. The conceptual set addresses the core; problems using carrier statistics, depletion widths, capacitance and MOSFET current models require the indicated enrichment material. Computational data must be labelled simulated. The workshops add six multi-stage applications.

Conceptual questions

- C1.** A crystal has a completely filled valence band and an empty conduction band separated by a large energy gap. Explain why the large number of electrons in the filled band does not by itself imply high conductivity.
- C2.** A student says, “Electrons in a metal move slowly because the measured drift velocity is slow.” Explain what is wrong with this statement using the distinction between drift velocity and microscopic electronic velocities.

- C3.** What is a hole? Explain why treating a hole as a positively charged quasiparticle is useful even though no proton moves through the valence band.
- C4.** Why does donor doping not normally leave an n-type sample with a large net negative charge?
- C5.** Why can a semiconductor become more conductive when heated while an ordinary metal can become less conductive over the same temperature interval?
- C6.** Explain how a Hall-voltage sign measurement can reveal the dominant carrier sign in a simple one-carrier model.
- C7.** What creates the electric field inside the depletion region of a p–n junction? Identify which charges are mobile and which are fixed in the idealized picture.
- C8.** Why does forward bias increase diode current while reverse bias generally suppresses it before breakdown?
- C9.** Why can a direct-gap semiconductor be a more efficient light emitter than an indirect-gap semiconductor with a comparable band gap?
- C10.** Explain the difference between the built-in potential of an equilibrium p–n junction and the open-circuit voltage of an illuminated solar cell.
- C11.** Explain how a transistor controls a separately powered current path and where amplification energy comes from.

Analytical and numerical problems

- A1.** A metal has carrier density $n = 6.0 \times 10^{28} \text{ m}^{-3}$ and relaxation time $\tau = 2.5 \times 10^{-14} \text{ s}$. Using the free-electron mass, calculate its Drude conductivity and resistivity.
- A2.** A conductor carries current density $J = 1.8 \times 10^6 \text{ A m}^{-2}$. The carrier density is $8.0 \times 10^{28} \text{ m}^{-3}$ and the carriers have charge $-e$. Calculate the drift-velocity magnitude. Compare the result with a characteristic electronic speed of 10^6 m s^{-1} .
- A3.** A semiconductor has $n = 2.0 \times 10^{21} \text{ m}^{-3}$, $p = 5.0 \times 10^{17} \text{ m}^{-3}$, $\mu_n = 0.10 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $\mu_p = 0.040 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$. Find the electron contribution, hole contribution and total conductivity.
- A4.** A material has $N_C = 2.5 \times 10^{25} \text{ m}^{-3}$ at 300 K and $E_C - E_F = 0.20 \text{ eV}$. Using the nondegenerate approximation, estimate n . Take $k_B T = 0.0259 \text{ eV}$.
- A5.** A compensated semiconductor has fully ionized donors $N_D = 7.0 \times 10^{21} \text{ m}^{-3}$ and acceptors $N_A = 2.0 \times 10^{21} \text{ m}^{-3}$, with $n_i = 1.0 \times 10^{16} \text{ m}^{-3}$. Estimate the majority and minority carrier concentrations in the extrinsic regime.
- A6.** A one-carrier sample has Hall coefficient $R_H = +4.0 \times 10^{-4} \text{ m}^3 \text{ C}^{-1}$. Identify the carrier sign and calculate the carrier density. For the same sample take $\sigma = 320 \text{ S m}^{-1}$ and infer the one-carrier Hall mobility.
- A7.** An ideal diode has saturation current $I_S = 5.0 \times 10^{-12} \text{ A}$ at 300 K. Using $V_T = 25.9 \text{ mV}$, calculate the current at $V = 0.50 \text{ V}$ and at $V = 0.55 \text{ V}$. Find the ratio.

- A8.** In a forward-bias region with $I \gg I_S$, a diode current changes from 0.10 mA to 1.0 mA when the voltage increases by 90 mV at 300 K. Estimate the ideality factor.
- A9.** An LED material has effective radiative transition energy $E = 2.10$ eV. Estimate the corresponding wavelength using $hc \approx 1240$ eV nm. State the approximate visible color region.
- A10.** A photodiode has external quantum efficiency 0.65 at 900 nm. Estimate the responsivity using $\mathcal{R} = \eta_q e \lambda / (hc)$.
- A11.** A solar cell has $V_{oc} = 0.62$ V, $I_{sc} = 35$ mA, $V_{mp} = 0.50$ V and $I_{mp} = 31$ mA. Calculate the fill factor. If incident power is 100 mW, calculate the conversion efficiency.
- A12.** At 300 K, an abrupt junction has $N_A = 8.0 \times 10^{21}$ m⁻³, $N_D = 2.0 \times 10^{21}$ m⁻³ and $n_i = 1.0 \times 10^{16}$ m⁻³. Calculate V_{bi} using $V_T = 25.9$ mV. For this same junction take $\epsilon_s = 12\epsilon_0$ and find the zero-bias depletion width and its p/n partition.
- A13.** A junction capacitance is 120 pF at zero bias and the built-in potential is 0.75 V. Assuming an abrupt-junction law, estimate the capacitance at $V_R = 6.0$ V.
- A14.** A semiconductor has electron mobility 0.085 m² V⁻¹ s⁻¹ at 300 K. Estimate D_n from the Einstein relation. If excess minority electrons have lifetime 4.0 μ s, use your diffusion coefficient to calculate their diffusion length.
- A15.** An idealized MOSFET has $\mu_n C_{ox}(W_{ch}/L_{ch}) = 1.5$ mA V⁻² and $V_{th} = 0.70$ V. Find the saturation current for $V_{GS} = 1.8$ V. Then set $V_{DS} = 0.30$ V, verify the linear-region condition and calculate the current at that bias.
- A16.** A ballistic conductor has three spin-degenerate channels with transmissions 0.95, 0.60 and 0.20. Calculate its Landauer conductance and resistance using $G_0 = 7.75 \times 10^{-5}$ S.

Computational and data-analysis problems

- N1.** Generate synthetic forward-diode data from Eq. (21.111) with a chosen I_S and n_d , add Gaussian voltage or current noise, and recover the parameters from a semilog fit. Repeat for several fit windows and discuss parameter stability. Repeat with a declared nonzero series resistance and solve the implicit current equation. Compare recovered parameters, residuals and fit-window dependence.
- N2.** Plot $n(E_F)$ and $p(E_F)$ across a 1.0 eV gap for three temperatures using chosen constant N_C and N_V . Mark the intrinsic Fermi level and verify numerically that np is independent of E_F in the nondegenerate model.
- N3.** Use $n_i(T) \propto T^{3/2} \exp[-E_g/(2k_B T)]$ with a fixed hypothetical E_g to produce an Arrhenius-style plot of $\ln[(n_i/T^{3/2})/Y_{ref}]$ versus $1/T$, with a declared fixed reference Y_{ref} . Show how the slope can be used to estimate E_g .
- N4.** Build a spreadsheet or script that calculates $\sigma = e(n\mu_n + p\mu_p)$ for a compensated semiconductor as N_D is scanned across N_A . Use charge neutrality plus the mass-action law rather than abruptly switching between n-type and p-type approximations.
- N5.** Numerically plot the depletion charge density, electric field and electrostatic potential of an abrupt p-n junction under the depletion approximation. Verify charge neutrality and the

integrated potential drop. Repeat for several reverse biases and test the predicted square-root laws for width and peak-field magnitude.

- N6.** For the long-channel MOSFET model, make a family of I_D versus V_{DS} curves for several V_{GS} . Mark the transition from the linear formula to the ideal saturation formula and discuss continuity at the boundary.

Engineering and applied-science problems

- E1.** A temperature sensor uses a semiconductor resistor whose resistance falls by a factor of 20 between two calibration temperatures. Explain why a simple constant temperature coefficient like that used for a metal may be inadequate. Propose a calibration strategy based on measured data rather than assuming linearity.
- E2.** Design a simple LED current-limiting calculation for a 5.0 V supply, an LED with an assumed forward voltage of 2.0 V at the operating point, and target current 10 mA. Choose a resistor and calculate its power dissipation. Explain why the LED should not be modeled as a fixed resistor.
- E3.** A half-wave rectifier drives a load current of approximately 25 mA. Estimate the filter capacitance required to limit the capacitor discharge ripple to 0.50 V over 20 ms. State why this estimate is conservative or approximate.
- E4.** A reverse-biased detector diode has capacitance 20 pF. It is connected to an amplifier input with 5 pF parasitic capacitance. By what factor does the total input capacitance change if detector redesign reduces its capacitance to 10 pF? Discuss why parasitics eventually limit the benefit.
- E5.** An optical link uses a photodiode with responsivity 0.60 A W^{-1} . If the receiver requires at least $60 \text{ }\mu\text{A}$ signal current, estimate the minimum optical power at the detector. Name at least three real effects omitted by this link-budget estimate.
- E6.** A high-voltage diode designer lowers the doping of a drift region. Explain qualitatively why this can improve reverse-voltage capability but increase conduction losses. Identify which equations or ideas from this chapter support each side of the trade-off.
- E7.** A radiation detector receives 1.0×10^5 events per second, each producing on average 2.0×10^4 collected electrons. Calculate the average detector current. Explain why pulse-processing electronics still sees discrete signals even though the average current is small.

Challenge problems

- H1.** Starting from $n = N_C e^{-(E_C - E_F)/k_B T}$ and $p = N_V e^{-(E_F - E_V)/k_B T}$, derive both the mass-action law and the intrinsic Fermi-level expression Eq. (21.69).
- H2.** For a fully ionized compensated semiconductor, solve the two equations $n - p = N_D - N_A$ and $np = n_i^2$ exactly for n and p . Show that your result smoothly approaches the intrinsic limit when $N_D = N_A$ and the majority-carrier limit when $|N_D - N_A| \gg n_i$.
- H3.** Derive the abrupt-junction depletion-width formula Eq. (21.96) from Poisson's equation. Begin from the piecewise charge density, integrate once to find $E(x)$, integrate again to obtain the potential drop, and use charge neutrality $N_A x_p = N_D x_n$. Then differentiate depletion charge to derive $C_j = \epsilon_s A/W$ and the linear $1/C_j^2$ law.

- H4.** Consider a diode model $V = n_d V_T \ln(I/I_S) + IR_s$ in strong forward conduction. Derive an expression for differential resistance $r_d = dV/dI$. Show how a plot of r_d versus $1/I$ can in principle separate R_s from $n_d V_T/I$.
- H5.** Starting from $I = I_L - I_S(e^{V/(n_d V_T)} - 1)$, derive the transcendental condition that defines the maximum-power point of an ideal solar cell. Explain why a numerical root finder is a natural solution method.
- H6.** Show that the ideal long-channel MOSFET linear-region expression Eq. (21.126) equals the saturation expression Eq. (21.127) at the boundary $V_{DS} = V_{GS} - V_{th}$.
- H7.** In the Landauer formula, suppose N channels all have equal transmission $\mathcal{T}$. Derive the resistance and discuss the perfect-transmission limit. Why does a finite two-terminal quantum resistance not contradict the absence of ordinary dissipative scattering inside an ideal ballistic channel?

21.23 Problem-solving workshop

The following workshop problems are longer than ordinary exercises. They are designed for auditorium classes, group work or graded homework. Each begins with a modelling decision before numerical calculation.

21.23.1 Workshop 1: conductivity from several microscopic routes

A rectangular sample has length $L = 12$ mm, width $w = 2.0$ mm and thickness $t = 0.50$ mm. A four-terminal measurement gives $V = 0.240$ V at $I = 4.0$ mA. A separate Hall measurement, interpreted in a one-carrier model, gives $R_H = -6.0 \times 10^{-4}$ m³ C⁻¹.

Tasks. (a) Determine R , ρ and σ . (b) Estimate the carrier density and sign. (c) Infer the mobility from $\sigma = ne\mu$. (d) If the effective mass is specified as $m^* = 0.30m_e$, infer a Drude-like momentum relaxation time $\tau = m^*\mu/e$. (e) Identify at least four assumptions in this inference chain that could fail in a real sample.

21.23.2 Workshop 2: junction electrostatics and capacitance

An abrupt junction has $N_A = 1.0 \times 10^{22}$ m⁻³, $N_D = 2.0 \times 10^{21}$ m⁻³, $n_i = 1.0 \times 10^{16}$ m⁻³, $\varepsilon_s = 11.7\varepsilon_0$ and area $A = 0.50$ mm² at 300 K.

Tasks. (a) Calculate V_{bi} . (b) For $V_R = 0, 1, 5, 10, 20$ V calculate W , x_p , x_n , E_{max} and C_j . (c) Plot $1/C_j^2$ against V_R . (d) Explain why the most lightly doped side determines most of the depletion width. (e) State the limits of the depletion approximation.

21.23.3 Workshop 3: LED optical power from electrical input

An LED operates at $I = 20$ mA and $V = 2.2$ V, emitting light near 620 nm. Its measured optical output is 8.0 mW.

Tasks. (a) Calculate electrical input power and wall-plug efficiency. (b) Calculate photon energy and emitted photon rate. (c) Calculate the electron-flow rate I/e and estimate the external photon-per-electron ratio. (d) Explain why this ratio is not determined solely by the band gap. (e) List thermal and optical loss channels.

21.23.4 Workshop 4: a photodiode receiver

At 850 nm, a photodiode has quantum efficiency $\eta_q = 0.70$, junction capacitance $C_j = 12$ pF and dark current $I_d = 5$ nA at its chosen reverse bias. It is connected to an amplifier with input capacitance $C_{in} = 3$ pF and a feedback resistance $R_f = 100$ k Ω .

Tasks. (a) Calculate responsivity. (b) Find photocurrent for $P_{opt} = 50$ μ W. (c) Estimate the simple RC time scale $R_f(C_j + C_{in})$. (d) Explain why this is not a complete bandwidth model of a real transimpedance amplifier. (e) Discuss qualitatively how reverse bias trades capacitance against dark current and breakdown margin.

21.23.5 Workshop 5: solar-cell maximum power

Use the ideal illuminated-diode model with $T = 300$ K, $n_d = 1.2$, $I_L = 0.50$ A and $I_S = 2.0 \times 10^{-9}$ A. Numerically generate the I - V curve from 0 to V_{oc} , calculate output power $P = IV$, and locate the maximum-power point. Report V_{oc} , I_{sc} , V_{mp} , I_{mp} and fill factor. Then add a series resistance through the implicit relation $V_{terminal} = V_{diode} - IR_s$ and explore how R_s reduces fill factor.

21.23.6 Workshop 6: digital switching and gate capacitance

A logic node presents an effective capacitance C_L that is charged from 0 to supply voltage V_{DD} and later discharged. Derive the energy drawn from the supply during charging, show that only $\frac{1}{2}C_L V_{DD}^2$ remains stored in the capacitor, and identify where the remaining energy is dissipated in a resistive charging path. For random activity factor α and switching frequency f , motivate the dynamic-power scaling

$$P_{dyn} \sim \alpha C_L V_{DD}^2 f.$$

Explain why lowering supply voltage is so powerful for reducing digital switching power, while noting that transistor speed, noise margins and leakage impose design constraints.

21.24 Selected solutions to representative problems

These solutions illustrate the expected style: identify the model, carry units, perform a limiting or consistency check, and interpret the result. They are not intended to remove the value of solving the remaining problems independently.

Solution to A1: Drude conductivity

Using

$$\sigma = \frac{ne^2\tau}{m_e},$$

we obtain

$$\sigma = \frac{(6.0 \times 10^{28})(1.602 \times 10^{-19})^2(2.5 \times 10^{-14})}{9.11 \times 10^{-31}} \approx 4.2 \times 10^7 \text{ S m}^{-1}.$$

Therefore

$$\rho = \sigma^{-1} \approx 2.4 \times 10^{-8} \text{ } \Omega \text{ m}.$$

The scale is characteristic of a good metallic conductor, which is consistent with the chosen large carrier density and short but finite momentum-relaxation time.

Solution to A4: carrier statistics

The nondegenerate conduction-band expression is

$$n = N_C \exp\left[-\frac{E_C - E_F}{k_B T}\right].$$

Using the energy ratio directly in electronvolts,

$$\frac{E_C - E_F}{k_B T} = \frac{0.20}{0.0259} \approx 7.72.$$

Therefore

$$n = (2.5 \times 10^{25}) e^{-7.72} \approx 1.1 \times 10^{22} \text{ m}^{-3}.$$

The result is much smaller than N_C , which is internally consistent with the nondegenerate approximation: the conduction band is only sparsely occupied.

Solution to A5: compensated doping

In the extrinsic fully ionized regime,

$$n \simeq N_D - N_A = 5.0 \times 10^{21} \text{ m}^{-3}.$$

Then

$$p = \frac{n_i^2}{n} = \frac{10^{32}}{5.0 \times 10^{21}} = 2.0 \times 10^{10} \text{ m}^{-3}.$$

The sample is strongly n-type. The subtraction $N_D - N_A$ rather than N_D alone reflects compensation by acceptors.

Solution to A6: Hall density and mobility

The positive Hall coefficient implies positive dominant carriers in the one-carrier sign convention, so the sample is hole dominated. The concentration is

$$p = \frac{1}{e R_H} = \frac{1}{(1.602 \times 10^{-19})(4.0 \times 10^{-4})} \approx 1.56 \times 10^{22} \text{ m}^{-3}.$$

For this same sample the mobility estimate is

$$\mu_H = |R_H| \sigma = (4.0 \times 10^{-4})(320) = 0.128 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}.$$

The unit follows because m^3/C times $\text{A}/(\text{V m})$ is $\text{m}^2/(\text{V s})$. These two calculations are reliable only if one carrier type dominates and the simple Hall factor is close to unity. The two-carrier formula in the advanced section shows how the interpretation changes otherwise.

Solution to A7: exponential diode response

For $I \gg I_S$,

$$I \approx I_S e^{V/V_T}.$$

At $V_1 = 0.50 \text{ V}$,

$$I_1 \approx 5.0 \times 10^{-12} e^{0.50/0.0259} \approx 1.2 \times 10^{-3} \text{ A}.$$

At $V_2 = 0.55 \text{ V}$,

$$\frac{I_2}{I_1} = e^{(0.05/0.0259)} \approx 6.9,$$

so $I_2 \approx 8.3 \text{ mA}$. A modest voltage change can therefore produce a large current change until series resistance and other nonidealities become important.

Solution to A8: extracting ideality factor

In the forward exponential regime,

$$\frac{I_2}{I_1} = \exp\left(\frac{V_2 - V_1}{n_d V_T}\right).$$

Here $I_2/I_1 = 10$ and $\Delta V = 0.090$ V, so

$$n_d = \frac{\Delta V}{V_T \ln 10} = \frac{0.090}{(0.0259)(2.3026)} \approx 1.51.$$

The result is compatible with a nonideal forward-current mechanism over the selected interval. It should not be interpreted as a universal constant of the diode; an experimental ideality factor can vary with current and temperature.

Solution to A10: photodiode responsivity

At wavelength $\lambda = 900$ nm and quantum efficiency $\eta_q = 0.65$,

$$\mathcal{R} = \eta_q \frac{e\lambda}{hc}.$$

Using $hc/e \approx 1240$ V nm,

$$\mathcal{R} \approx 0.65 \frac{900}{1240} \text{ A W}^{-1} \approx 0.472 \text{ A W}^{-1}.$$

The result assumes one collected electron charge per successfully collected photo-generated pair and no internal avalanche gain.

Solution to A11: fill factor and efficiency

The fill factor is

$$\text{FF} = \frac{(0.50)(0.031)}{(0.62)(0.035)} \approx 0.714.$$

The maximum electrical power is

$$P_{\text{mp}} = 0.50 \times 0.031 = 15.5 \text{ mW}.$$

For $P_{\text{in}} = 100$ mW,

$$\eta = \frac{15.5}{100} = 0.155 = 15.5\%.$$

The fill factor compares the maximum-power rectangle with $V_{\text{oc}}I_{\text{sc}}$; efficiency compares output power with incident optical power. They have different denominators and should not be confused. These are specified exercise values, not a performance claim about a commercial cell.

Solution to A12: junction barrier and width

With $V_T = 0.0259$ V,

$$V_{\text{bi}} = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) = 0.0259 \ln\left(\frac{(8.0 \times 10^{21})(2.0 \times 10^{21})}{10^{32}}\right) \approx 0.67 \text{ V}.$$

The formula is logarithmic in doping concentration and assumes nondegenerate equilibrium statistics.

For the junction of A12,

$$N_A = 8.0 \times 10^{21} \text{ m}^{-3}, \quad N_D = 2.0 \times 10^{21} \text{ m}^{-3},$$

with $V_{bi} \approx 0.67$ V and $\varepsilon_s = 12\varepsilon_0$. At zero external bias,

$$W = \sqrt{\frac{2\varepsilon_s}{e} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) V_{bi}}.$$

Using the unrounded built-in potential gives

$$W \approx 7.4 \times 10^{-7} \text{ m} = 0.74 \text{ } \mu\text{m},$$

The fraction on the n side is

$$\frac{x_n}{W} = \frac{N_A}{N_A + N_D} = \frac{8}{10} = 0.80.$$

Thus about 80% of the depletion region lies in the n side because it is the more lightly doped region.

Solution to A13: depletion capacitance

For an abrupt junction,

$$\frac{C_R}{C_0} = \sqrt{\frac{V_{bi}}{V_{bi} + V_R}}.$$

With $C_0 = 120$ pF, $V_{bi} = 0.75$ V and $V_R = 6.0$ V,

$$C_R = 120 \sqrt{\frac{0.75}{6.75}} \text{ pF} = 120 \left(\frac{1}{3} \right) \text{ pF} = 40 \text{ pF}.$$

The simple inverse-square-root law follows from a uniform abrupt doping profile. Graded junctions have a different voltage dependence.

Solution to A15: two MOSFET operating regimes

For $V_{GS} = 1.8$ V and $V_{th} = 0.70$ V, the overdrive is

$$V_{ov} = 1.10 \text{ V}.$$

In saturation,

$$I_D = \frac{1}{2} (1.5 \text{ mA V}^{-2}) (1.10)^2 \approx 0.908 \text{ mA}.$$

For $V_{DS} = 0.30$ V $< V_{ov}$, the linear-region model applies:

$$I_D = (1.5 \text{ mA V}^{-2}) \left[(1.10)(0.30) - \frac{0.30^2}{2} \right] \approx 0.428 \text{ mA}.$$

The two answers correspond to different drain biases and therefore different operating regions.

Solution to A16: quantum conductance

The summed transmission is

$$\sum_n \mathcal{T}_n = 0.95 + 0.60 + 0.20 = 1.75.$$

Thus

$$G = (7.75 \times 10^{-5}) (1.75) \approx 1.36 \times 10^{-4} \text{ S},$$

and

$$R = G^{-1} \approx 7.37 \text{ k}\Omega.$$

The result is naturally expressed through transmission channels rather than a bulk resistivity.

Solution to E2: LED current limiting

The resistor must drop approximately

$$V_R = 5.0 - 2.0 = 3.0 \text{ V}$$

at the target current $I = 10 \text{ mA}$. Hence

$$R = \frac{V_R}{I} = \frac{3.0}{0.010} = 300 \text{ } \Omega.$$

A nearby standard value can be selected after considering tolerance and desired current. The resistor power is

$$P_R = I^2 R = (0.010)^2 (300) = 0.030 \text{ W}.$$

A resistor rated comfortably above this value would be appropriate. The LED is not modeled as a fixed resistor because its current–voltage relation is strongly nonlinear and temperature dependent.

Solution to E7: average detector current

Each event produces collected charge

$$Q_{\text{event}} = e(2.0 \times 10^4) \approx 3.20 \times 10^{-15} \text{ C}.$$

At event rate 10^5 s^{-1} ,

$$I_{\text{avg}} = Q_{\text{event}} \times \text{rate} \approx 3.2 \times 10^{-10} \text{ A} = 0.32 \text{ nA}.$$

The small average current does not mean that individual events are unobservable. Low-noise electronics integrates each short charge pulse, and the time structure contains information that is lost in the average.

Solution to H4: differential resistance

In strong forward conduction, write

$$V = n_d V_T \ln\left(\frac{I}{I_s}\right) + I R_s.$$

Differentiate with respect to I :

$$r_d \equiv \frac{dV}{dI} = \frac{n_d V_T}{I} + R_s.$$

Thus a plot of r_d against $1/I$ is predicted to be linear in the regime where n_d and R_s are approximately constant. The intercept gives R_s and the slope gives $n_d V_T$. Experimentally, differentiation amplifies noise, so smoothing or direct nonlinear fitting is usually preferable.

21.25 Bridge to the next modern-physics chapter

Electrical conduction in solids demonstrates how quantum mechanics becomes engineering. The next modern-physics topics change the physical scale again: from electrons moving in periodic solids to the structure, stability and transformations of atomic nuclei. Several ideas will recur. Energy levels and conservation laws remain central; exponential laws reappear in radioactive decay; semiconductor junctions become practical detectors of ionizing radiation; and careful distinction between microscopic events and macroscopic measured signals becomes even more important.

The detector example developed in this chapter therefore provides a natural bridge. Nuclear radiation deposits energy in matter, the semiconductor converts that deposited energy into electron–hole pairs, internal electric fields collect the charge, and electronics converts the charge pulse into data. The device is a meeting point of quantum physics, solid-state physics, electromagnetism and nuclear measurement.

Chapter A

Notation dictionary for Chapter 21

Symbol	Meaning	Typical SI unit
$\mathbf{J}$	conventional current-density vector	A m^{-2}
$\mathbf{E}$	electric-field vector	V m^{-1}
σ	electrical conductivity	S m^{-1}
ρ	electrical resistivity	$\Omega \text{ m}$
n	mobile electron concentration	m^{-3}
p	mobile hole concentration	m^{-3}
n_i	intrinsic carrier concentration	m^{-3}
N_D	donor concentration	m^{-3}
N_A	acceptor concentration	m^{-3}
N_D^+	ionized donor concentration	m^{-3}
N_A^-	ionized acceptor concentration	m^{-3}
e	positive elementary-charge magnitude	C
q	signed carrier charge	C
$\mathbf{v}_d$	drift velocity	m s^{-1}
μ_n, μ_p	electron and hole mobilities	$\text{m}^2 \text{V}^{-1} \text{s}^{-1}$
τ	momentum-relaxation time in a Drude-like model	s
D_n, D_p	electron and hole diffusion coefficients	$\text{m}^2 \text{s}^{-1}$
τ_n, τ_p	minority-carrier lifetimes in a recombination model	s
L_n, L_p	diffusion lengths	m
E_C	conduction-band edge energy	J or eV
E_V	valence-band edge energy	J or eV
$E_g = E_C - E_V$	band-gap energy	J or eV
E_F	Fermi level / equilibrium chemical-potential energy	J or eV
E_i	intrinsic Fermi level	J or eV
N_C, N_V	effective densities of states	m^{-3}
$f(E)$	Fermi–Dirac occupation probability	dimensionless
$g(E)$	density of states per volume per energy	$\text{m}^{-3} \text{J}^{-1}$
m_n^*, m_p^*	electron and hole effective masses	kg
k	wave-vector magnitude in an $E(k)$ model	m^{-1}
k_B	Boltzmann constant	J K^{-1}
$h, \hbar$	Planck constant and reduced Planck constant	J s
$V_T = k_B T / e$	thermal voltage	V
V_{bi}	p–n-junction built-in potential	V
V_a	applied junction bias, positive for forward bias in the stated convention	V
V_R	reverse-bias magnitude, $V_R = -V_a > 0$	V
W	total depletion width	m
x_p, x_n	depletion widths on p and n sides	m
ε_s	semiconductor permittivity	F m^{-1}
I_S	ideal diode saturation-current parameter	A

Symbol	Meaning	Typical SI unit
n_d	diode ideality factor	dimensionless
C_j	junction capacitance	F
η_q	external quantum efficiency of a photodetector	dimensionless
$\mathcal{R}$	optical responsivity	A W^{-1}
I_L	light-generated current in simple solar-cell model	A
V_{oc}	solar-cell open-circuit voltage	V
I_{sc}	solar-cell short-circuit current	A
FF	solar-cell fill factor	dimensionless
C_{ox}	MOS oxide capacitance per unit area	F m^{-2}
V_{th}	MOSFET threshold voltage	V
V_{GS}, V_{DS}	MOSFET gate-source and drain-source voltages	V
W_{mch}/L_{mch}	MOSFET channel width-to-length ratio	dimensionless
$\mathcal{T}_n$	quantum transmission probability of channel n	dimensionless
$G_0 = 2e^2/h$	spin-degenerate conductance quantum in the stated convention	S
S	Seebeck coefficient	V K^{-1}
κ	thermal conductivity	$\text{W m}^{-1} \text{K}^{-1}$
ZT	thermoelectric figure of merit	dimensionless
Δ	superconducting excitation-gap parameter	J or eV
T_c	superconducting critical temperature	K

Note

The same letter can have different meanings in different branches of physics. In this chapter p normally denotes hole concentration, while $\mathbf{p}$ would conventionally denote momentum in mechanics. The context and vector notation must therefore remain explicit.

Chapter B

Compact formula sheet

This appendix collects equations used in the chapter. It is not a substitute for the assumptions stated in the main text.

Formula summary

Macroscopic transport

$$\begin{aligned} \mathbf{J} &= \sigma \mathbf{E}, & \rho &= \frac{1}{\sigma}, \\ \mathbf{J} &= n_c q \mathbf{v}_d, & \mu &= \frac{|v_d|}{E}, \\ \sigma_{\text{Drude}} &= \frac{n e^2 \tau}{m}, & \sigma_{\text{semi}} &= e(n \mu_n + p \mu_p). \end{aligned}$$

Free-electron Fermi gas

$$\begin{aligned} k_F &= (3\pi^2 n)^{1/3}, \\ E_F &= \frac{\hbar^2 k_F^2}{2m}, \\ v_F &= \frac{\hbar k_F}{m}. \end{aligned}$$

Band curvature and statistics

$$\begin{aligned} v_g &= \frac{1}{\hbar} \frac{dE}{dk}, \\ \frac{1}{m^*} &= \frac{1}{\hbar^2} \frac{d^2 E}{dk^2}, \\ f(E) &= \frac{1}{e^{(E-E_F)/(k_B T)} + 1}. \end{aligned}$$

Nondegenerate semiconductor statistics

$$\begin{aligned} n &= N_C e^{-(E_C - E_F)/(k_B T)}, \\ p &= N_V e^{-(E_F - E_V)/(k_B T)}, \\ N_C &= 2 \left(\frac{2\pi m_n^* k_B T}{h^2} \right)^{3/2}, \\ N_V &= 2 \left(\frac{2\pi m_p^* k_B T}{h^2} \right)^{3/2}, \\ n_i &= \sqrt{N_C N_V} e^{-E_g/(2k_B T)}, \\ np &= n_i^2, \\ p + N_D^+ &= n + N_A^-. \end{aligned}$$

Drift–diffusion and recombination

$$\begin{aligned}
J_n &= en\mu_n E + eD_n \frac{dn}{dx}, \\
J_p &= ep\mu_p E - eD_p \frac{dp}{dx}, \\
\frac{D_n}{\mu_n} &= \frac{D_p}{\mu_p} = \frac{k_B T}{e} = V_T, \\
L_n &= \sqrt{D_n \tau_n}, & L_p &= \sqrt{D_p \tau_p}.
\end{aligned}$$

Abrupt p–n junction

$$\begin{aligned}
V_{bi} &= V_T \ln \left(\frac{N_A N_D}{n_i^2} \right), \\
W &= \sqrt{\frac{2\varepsilon_s}{e} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (V_{bi} - V_a)}, \\
N_A x_p &= N_D x_n, \\
x_n &= \frac{N_A}{N_A + N_D} W, & x_p &= \frac{N_D}{N_A + N_D} W, \\
E_{\max} &= \frac{e N_A x_p}{\varepsilon_s} = \frac{e N_D x_n}{\varepsilon_s}, \\
C_j &= \frac{\varepsilon_s A}{W}.
\end{aligned}$$

Diode and optoelectronics

$$\begin{aligned}
I &= I_S \left(e^{V/V_T} - 1 \right), \\
I &\simeq I_S \left(e^{V/(n_d V_T)} - 1 \right), \\
E_\gamma &= hf = \frac{hc}{\lambda}, \\
\mathcal{R} &= \eta_q \frac{e\lambda}{hc}.
\end{aligned}$$

Idealized illuminated junction

$$\begin{aligned}
I &= I_L - I_S \left(e^{V/(n_d V_T)} - 1 \right), \\
V_{oc} &= n_d V_T \ln \left(1 + \frac{I_L}{I_S} \right), \\
FF &= \frac{V_{mp} I_{mp}}{V_{oc} I_{sc}}, \\
\eta &= \frac{V_{mp} I_{mp}}{P_{in}}.
\end{aligned}$$

Ideal long-channel nMOS model

$$\begin{aligned}
I_D &= \mu_n C_{ox} \frac{W_{ch}}{L_{ch}} \left[(V_{GS} - V_{th}) V_{DS} - \frac{V_{DS}^2}{2} \right], \\
I_{D,sat} &\simeq \frac{1}{2} \mu_n C_{ox} \frac{W_{ch}}{L_{ch}} (V_{GS} - V_{th})^2.
\end{aligned}$$

Selected enrichment formulas

$$\begin{aligned}
G &= \frac{2e^2}{h} \sum_n \mathcal{T}_n, \\
\Delta V &= -S \Delta T, \\
ZT &= \frac{S^2 \sigma T}{\kappa}, \\
2\Delta(0) &\approx 3.52 k_B T_c \quad (\text{weak-coupling conventional BCS limit}).
\end{aligned}$$

Bibliography

- [1] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics*, 12th ed., Vol. 2, Wiley. Chapter 41 (especially sections 41.1–41.3) provides the primary sequence and a standard calculus-based treatment of bands, semiconductors, p–n junctions and transistor concepts.
- [2] R. A. Serway and J. W. Jewett, Jr., *Physics for Scientists and Engineers with Modern Physics*, 9th ed., Brooks/Cole Cengage Learning. Chapter 43, sections 43.4–43.8, provides complementary background on free-electron theory, band theory, semiconductor devices and superconductivity.
- [3] S. J. Ling, J. Sanny, and W. Moebs, *University Physics, Volume 2*, OpenStax, Rice University. Chapter 9, especially sections 9.2–9.3, supplies the prerequisite transport comparison; Volume 2 does not contain a semiconductor band/device chapter. The electricity-and-magnetism chapters provide background on current, resistance, microscopic conduction and superconductors.