# 1

# Fundamental Particles, Fundamental Forces, and Mathematical Tools

"Viribus Unitis"
[with united forces]
Emperor Franz Josef the First (and the last) (1830–1916)

"It is difficult to make predictions, especially about the future"
Lawrence P. "Yogi" Berra (1925–2015)

## 1.0 Introduction

This chapter presents several principles that underpin the chemical edifice; it is divided into four parts (Julius Caesar's Gaul was divided into only 3 parts). Part One (Sections 1.1–1.4) presents the fundamental physical forces whose interplay underlie Mendeleyeff's[1] periodic table of the chemical elements. Part Two (Sections 1.5–1.8) reviews essential mathematical methods. Part Three (Section 1.9) discusses electricity and magnetism. Part Four (Sections 1.10–1.18) discusses the meaning of quantum numbers, special relativity, general relativity, Feynman diagrams, physical optics, mathematical transforms, contour integration, and statistics.

The chapter concludes with eleven Special Topics: 1. Natural Linewidths; 2. Nuclear Shell Model; 3. Factoring a Polynomial; 4. Other Solutions of the Cubic Equation; 5. Solving a Quartic Equation; 6. Spherical Trigonometry; 7. Unit Systems for Electromagnetism; 8. Nonlinear Optics; 9. Conic Sections, Ellipses, and Planetary Motion; 10. Ellipsometry in Detail; 11: Standard Model, The Eightfold Way.

Within a traditional textbook, these concepts and methods would be presented as needed, but it seemed expedient to pull them together here: this is a deliberate effort to link mathematics, physics, and chemistry into a unified and coherent whole.

In this book, important equations are boxed. To save space at the cost of elegance, fractions are mostly replaced by negative exponents for the denominator. For emphasis, double parentheses are used for those equations that had been presented earlier in the text.

To relieve the reader from too much seriousness. Sidelines occasionally interrupt the narrative.

> **Sideline 1.1** The name "physics" derives from the Greek word φυσις (=nature, essence): early physicists like Newton[2] were called natural philosophers. The word "chemistry," through its Arabic precursor *alchimya*, derives from the Greek word χημι (=black Earth), a tribute to the Egyptians' embalming arts. Mathematics comes from the Greek μαθημα (=learning, study). Algebra comes from the Arabic *al-jabr* (=transposition [to the other side of an equation]). Calculus (as in "infinitesimal calculus") is the Latin word for a small pebble.

1 Dmitri Ivanovich Mendeleyeff (1834–1907).
2 Sir Isaac Newton (1642–1727).

*The Physical Chemist's Toolbox*, Second Edition. Robert M. Metzger.
© 2023 John Wiley & Sons, Inc. Published 2023 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/metzger/physchemtoolbox2

COPYRIGHTED MATERIAL

**Sideline 1.2** Michael Faraday[3] was asked "Of what use is electricity?" He supposedly answered "What is the use of a new-born baby?" This same answer was actually first given by Franklin[4] when witnessing the flight of the Montgolfier balloon over Versailles in 1785; this answer was repeated by Faraday in 1816 a propose of the discovery by Sir Humphry Davy[5] of several new chemical elements.

**PART ONE:** Fundamental Forces

## 1.1 Fundamental Forces, Elementary Particles, Nuclei, and Atoms

In a freshman chemistry class at the University of California, Los Angeles, in 1959, the instructor, Libby,[6] defined chemistry as "the study of electrons and what they do." However, he took pains to teach us that these electrons would have "nothing to do if atomic nuclei did not exist." So, it behooves us to talk briefly about nuclei, elementary particles, and the fundamental forces in the cosmos: this will put chemistry in its proper perspective.

**Sideline 1.3** Willard Libby told his physical chemistry class in 1960 "if you want to find out the science that is ten years old, read a book; if you want to find out what is five years old, read a review article; if you want to find out what is one year old, read a scientific article; if you want to find what is current today, pick up the telephone."

The four **fundamental forces**, their governing equations, the mediating particles, their relative magnitudes, and their ranges are listed in Table 1.1.

For some chemists and materials scientists, only the electrical force is considered essential, so much of Table 1.1 may seem esoteric. But stable atoms exist only if the relevant nuclei can survive: **chemists ignore nuclei at their own peril**. For some elementary particle experimentalists, most of everything "at lower energies" (below 1 GeV) is jocularly described as "chemistry." Astronomers try to observe and explain the whole universe, so they must worry about everything, starting with gravitation and ending in nuclear structure and chemistry! The known elementary particles are listed in Tables 1.2 and 1.3.

We now introduce the four fundamental forces.

**Force one**: The first (and weakest) force is Newton's force of universal gravitation (1687) [1.1]:

$$\boldsymbol{F}_{12} = -Gm_1M_2\boldsymbol{r}_{12}/r_{12}^3 \tag{1.1.1}$$

which describes the attractive force $\boldsymbol{F}_{12}$ between two bodies of masses $m_1$ and $M_2$ placed a distance $\boldsymbol{r}_{12}$ apart, where $G$ is the **constant of gravitation**. The largest visible objects in the universe (galaxies, stars, quasars, planets, satellites, comets) are

**Table 1.1** The fundamental forces.

| Force | Law | Equation | Mediating particle | Relative magnitude | Range |
|---|---|---|---|---|---|
| Gravitation | Newton's | $\boldsymbol{F}_{12} = -Gm_1M_2\boldsymbol{r}_{12}/r_{12}^3$ | Graviton | $10^{-39}$ | Infinite |
| Electricity | Coulomb's[a] | $\boldsymbol{F}_{12} = q_1q_2\boldsymbol{r}_{12}/4\pi\varepsilon_0 r_{12}^3$ | Photon | $10^{-2}$ | Infinite |
| Weak nuclear | — | — | Vector boson | $10^{-5}$ | $10^{-18}$ m |
| Strong nuclear | Inter-quark | — | Gluon[b] | 1 | $10^{-15}$ m |
| Strong nuclear | Inter-nucleon | — | Pion (gluon?)[b] | 1 | $10^{-15}$ m |

[a] Charles-Augustin de Coulomb (1736–1806).
[b] For nucleon–nucleon strong interactions within nuclei, pions (=two-quark particles; see below) may be the mediating particles: gluons are probably not involved directly, since the nucleons have no "color charge." The internucleon potential goes to zero beyond 1.7 fm = $1.7 \times 10^{-15}$ m.
*Source:* Adapted from Serway [1.3].

3 Michael Faraday (1791–1867).
4 Benjamin Franklin (1706–1790).
5 Sir Humphry Davy (1778–1829).
6 Willard Frank Libby (1908–1980).

**Table 1.2** Fundamental (quark, gluon, graviton, neutrino) and elementary (=fundamental plus two-quark and three-quark combinations) particles.

| Particle name | Symbol | Life-time $\tau$ (sec) | Relative mass $m'$ | Relative electric charge $Q$ | Spin $S$ | Parity $P$ | Isospin $T$ | $z$-component Isospin $T_z$ | Baryon Ba | Strangeness St | Charm Ch | Beauty By | Truth Tr | Composition |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Quarks (fundamental, with six "flavors" u, d, s, c, b, and t, but not observed as single particles):* | | | | | | | | | | | | | | |
| Up quark | $u$ | ∞? | ~4.6 | 2/3 | 1/2 | +1 | 1/2 | +1/2 | 1/3 | 0 | 0 | 0 | 0 | |
| Anti-up quark∗ | $\overline{u}$ | ∞? | ~4.6 | −2/3 | 1/2 | −1 | 1/2 | +1/2 | −1/3 | 0 | 0 | 0 | 0 | |
| Down quark | $d$ | ∞? | ~16 | −1/3 | 1/2 | +1 | 1/2 | −1/2 | 1/3 | 0 | 0 | 0 | 0 | |
| Anti-down quark∗ | $\overline{d}$ | ∞? | ~16 | 1/3 | 1/2 | −1 | 1/2 | −1/2 | −1/3 | 0 | 0 | 0 | 0 | |
| Charmed quark | $c$ | ∞? | ~2490 | 2/3 | 1/2 | +1 | 0 | 0 | 1/3 | 0 | +1 | 0 | 0 | |
| Anticharmed q∗ | $\overline{c}$ | ∞? | ~2490 | −2/3 | 1/2 | −1 | 0 | 0 | −1/3 | 0 | −1 | 0 | 0 | |
| Strange quark | $s$ | ∞? | ~200 | −1/3 | 1/2 | +1 | 0 | 0 | 1/3 | −1 | 0 | 0 | 0 | |
| Anti-strange q∗ | $\overline{s}$ | ∞? | ~200 | 1/3 | 1/2 | −1 | 0 | 0 | −1/3 | +1 | 0 | 0 | 0 | |
| Bottom quark | $b$ | ∞? | ~8480 | −1/3 | 1/2 | +1 | 0 | 0 | 1/3 | 0 | +1 | 0 | 0 | |
| Antibottom q∗ | $\overline{b}$ | ∞? | ~8480 | 1/3 | 1/2 | −1 | 0 | 0 | −1/3 | 0 | −1 | 0 | 0 | |
| Top quark | $t$ | ∞? | >3.4E5 | 2/3 | 1/2 | +1 | 0 | 0 | 1/3 | 0 | 0 | 0 | +1 | |
| Antitop quark∗ | $\overline{t}$ | ∞? | >3.4E5 | −2/3 | 1/2 | −1 | 0 | 0 | −1/3 | 0 | 0 | 0 | −1 | |
| *Fundamental interaction carriers (for gluons, combination of color [r, y, b] and anticolor):* | | | | | | | | | | | | | | |
| Photon | $\nu$ | ∞ | 0 | 0 | 1 | −1 | 1,0 | 1,0 | 0 | 0 | 0 | 0 | 0 | |
| Vector boson | $W^+$ | 3E-25 | 1.6E5 | 1 | 1 | −1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| Vector boson | $Z^0$ | 3E-25 | 1.8E5 | 0 | 1 | −1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Vector boson | $W^-$ | 3E-25 | 1.6E5 | −1 | 1 | −1 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | |
| Gluon1 | $\lvert r\overline{g}\rangle$ | ? | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Gluon2 | $\lvert r\overline{b}\rangle$ | ? | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Gluon3 | $\lvert g\overline{b}\rangle$ | ? | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Gluon4 | $\lvert g\overline{r}\rangle$ | ? | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Gluon5 | $\lvert b\overline{r}\rangle$ | ? | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Gluon6 | $\lvert b\overline{g}\rangle$ | ? | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Gluon7 | $(\lvert r\overline{r}\rangle - \lvert g\overline{g}\rangle)/\sqrt{2}$ | ? | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Gluon8 | $(\lvert r\overline{r}\rangle + \lvert g\overline{g}\rangle - 2\lvert b\overline{b}\rangle)/\sqrt{6}$ | 0 | 0 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

(*Continued*)

**Table 1.2** (Continued)

| Particle name | Symbol | Life-time $\tau$ (sec) | Relative mass $m'$ | Relative electric charge $Q$ | Spin $S$ | Parity $P$ | Isospin $T$ | $z$-component Isospin $T_z$ | Baryon Ba | Strangeness St | Charm Ch | Beauty By | Truth Tr | Composition |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Electron neutrino | $\nu_e$ | ∞ | 0 | 0 | 1/2 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Electron antineutrino∗ | $\overline{\nu}_e$ | ∞ | 0 | 0 | 1/2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Muon neutrino | $\nu_\mu$ | ∞ | 0 | 0 | 1/2 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Muon antineutrino∗ | $\overline{\nu}_\mu$ | ∞ | 0 | 0 | 1/2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Tau neutrino | $\nu_T$ | ∞ | 0 | 0 | 1/2 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Tau antineutrino∗ | $\overline{\nu}_T$ | ∞ | 0 | 0 | 1/2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Graviton | $G$ | 0? | 0 | 0 | 2 | ? | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Higgs boson | $H^0$ | 1.6E-22 | >224 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| *Leptons (elementary particles):* | | | | | | | | | | | | | | |
| Electron | $e^-$, $\beta^{-1}$ | ∞ | 1 | −1 | 1/2 | — | — | — | | 0 | — | | | |
| Positron∗ | $e^+$, $\beta^+$ | ∞ | 1 | +1 | 1/2 | — | — | — | | 0 | — | | | |
| Muon[a] | $\mu^-$ | 2.2E-6 | 237 | −1 | 1/2 | — | — | — | | 0 | — | | | |
| Positive muon∗ | $\mu^+$ | 2.2E-6- | 237 | +1 | 1/2 | — | — | — | | 0 | — | — | | |
| Tau | $\tau^-$ | <4E-13 | 3477 | −1 | 1/2 | — | — | — | | 0 | — | — | | |
| Positive Tau∗ | $\tau^+$ | <4E-13 | 3477 | +1 | 1/2 | — | — | — | | 0 | — | — | | |
| *Mesons (combinations of two quarks):* | | | | | | | | | | | | | | |
| Positive pion | $\pi^+$ | 2.6E-8 | 273 | 1 | 0 | −1 | 1 | +1 | 0 | 0 | 0 | 0 | 0 | $u\overline{d}$ |
| Negative pion∗ | $\pi^-$ | 3E-8 | 273 | −1 | 0 | −1 | 1 | −1 | 0 | 0 | 0 | 0 | 0 | $\overline{u}d$ |
| Neutral pion | $\pi^0$ | 8E-17 | 264 | 0 | 0 | −1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | $(u\overline{u}-d\overline{d})/\sqrt{2}$ |
| Neutral ρ meson | $\rho^0$ | 5E-24 | 1510 | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $(u\overline{u}+d\overline{d})/\sqrt{2}$ |
| Positive kaon | $K^+$ | 1.2E-8 | 493.8 | 1 | 0 | −1 | 1/2 | 1/2 | 0 | 1 | 0 | 0 | 0 | $u\overline{s}$ |
| Negative kaon∗ | $K^-$ | 1.2E-8 | 493.8 | −1 | 0 | −1 | 1/2 | −1/2 | 0 | −1 | 0 | 0 | 0 | $\overline{u}s$ |
| Neutral kaon | $K^0$ | 8.9E-8 | 493.8 | 0 | 0 | −1 | 1/2 | −1/2 | 0 | 1 | 0 | 0 | 0 | $d\overline{s}$ |
| Neutral kaon∗ | $\overline{K}^0$ | 5.2E-8 | 493.8 | 0 | 0 | −1 | 1/2 | 1/2 | 0 | −1 | 0 | 0 | 0 | $\overline{d}s$ |
| η meson | $\eta^0$ | 5E-19 | 549 | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $(u\overline{u}+d\overline{d}-2s\overline{s})/\sqrt{6}$ |
| η′ meson | $\eta'$ | 3E-21 | 958 | 0 | 0 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $(u\overline{u}+d\overline{d}+s\overline{s})/\sqrt{3}$ |

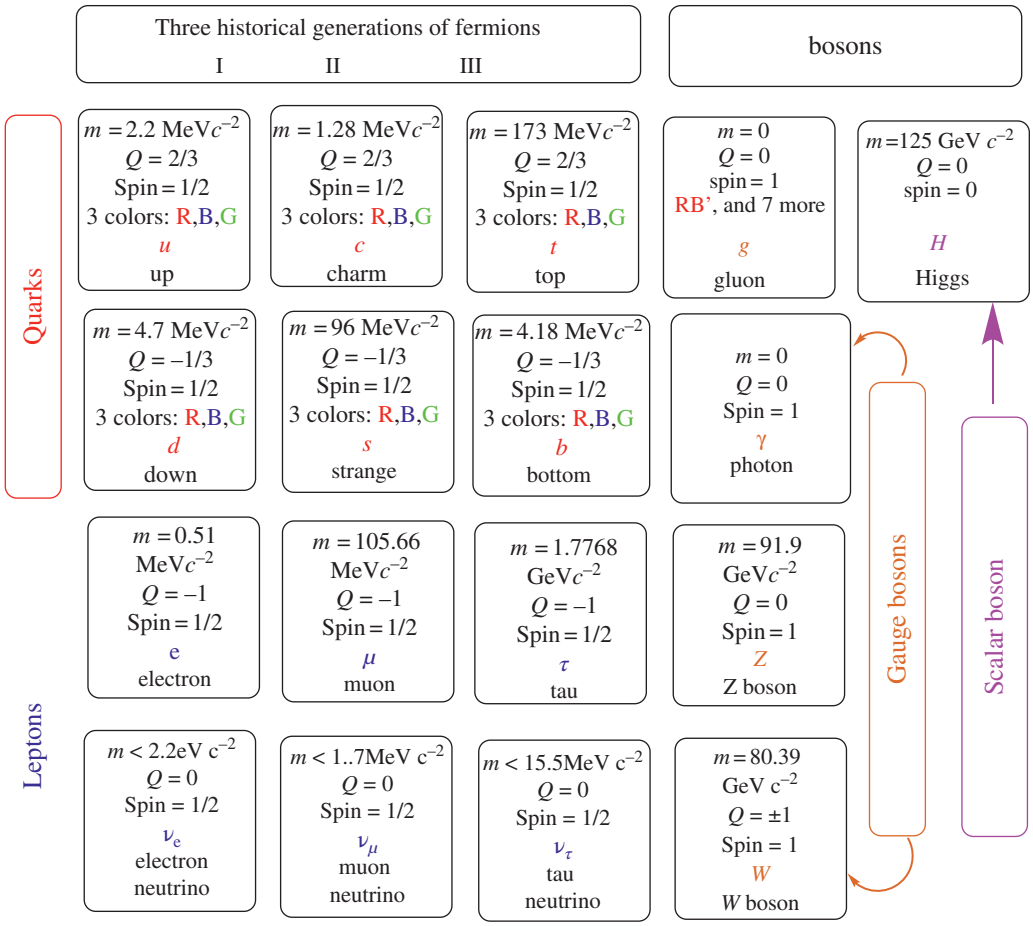
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Hadrons or Baryons (combinations of three quarks):* | | | | | | | | | | | | | | |
| Proton | p, ${}_1p^1$ | ∞ | 1838 | +1 | 1/2 | 1 | 1/2 | +1/2 | 1 | 0 | 0 | 0 | 0 | $u\bar{u}d$ |
| Anti-proton∗ | $p^-$ | ??? | 1838 | −1 | 1/2 | 1 | 1/2 | +1/2 | −1 | 0 | 0 | 0 | 0 | $u\bar{u}\bar{d}$ |
| Neutron | n, ${}_0n^1$ | 925 | 1839 | 0 | 1/2 | 1 | +1/2 | −1/2 | 1 | 0 | 0 | 0 | 0 | $ud\bar{d}$ |
| Lambda$^0$ | $\Lambda^0$ | 2.6E-10 | 1116 | 0 | 1/2 | +1 | 0 | 0 | 1 | −1 | 0 | 0 | 0 | $uds$ |
| Charm Lambda$^+$ | $\Lambda^{c+}$ | 2.0E-13 | 1116 | 1 | 1/2 | +1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | $udc$ |
| Bottom Lambda$^0$ | $\Lambda^{b0}$ | 1.4E-12 | 5622 | 0 | 1/2 | +1 | 0 | 0 | −1 | 0 | 0 | 1 | 0 | $udb$ |
| Sigma$^+$ | $\Sigma^+$ | 8E-11 | 1189 | 1 | 1/2 | +1 | 1 | 1 | 1 | −1 | 0 | 0 | 0 | $u\bar{u}s$ |
| Sigma$^0$ | $\Sigma^0$ | 6E-20 | 1192 | 0 | 1/2 | +1 | 1 | 0 | 1 | −1 | 0 | 0 | 0 | $uds$ |
| Sigma$^-$ | $\Sigma^-$ | 1.5E-10 | 1197 | −1 | 1/2 | +1 | 1 | −1 | 1 | −1 | 0 | 0 | 0 | $d\bar{d}s$ |
| Ch. Sigma$^{2+}$ | $\Sigma^{2+}$ | 3.0E-22 | 1200 | 2 | | +1 | 1 | −1 | 1 | −1 | 1 | 0 | 0 | $u\bar{u}c$ |
| Xi$^0$ | $\Xi^0$ | 2.9E-10 | 1315 | 0 | 1/2 | +1 | 1/2 | +1/2 | 1 | −2 | 0 | 0 | 0 | $us\bar{s}$ |
| Xi$^-$ | $\Xi^-$ | 1.6E-10 | 1321 | −1 | 1/2 | +1 | 1/2 | −1/2 | 1 | −2 | 0 | 0 | 0 | $ds\bar{s}$ |
| Omega$^-$ | $\Omega^-$ | 1.6E-10 | 1672 | −1 | 3/2 | +1 | 0 | 0 | 1 | −3 | 0 | 0 | 0 | $s\bar{s}b$ |

The masses $m'$ are rest masses, quoted relative to the mass of the electron, $m_e = 9.109\,389\,7 \times 10^{-31}$ kg, whose rest-mass energy is 0.510 999 06 MeV $c^{-2}$; the charges $Q$ are quoted relative to the absolute value of the charge on the electron, $e = 1.602\,176\,57 \times 10^{-19}$ C; lifetime is given as the known half-life (s) in the abbreviated format 2.34E-3, which translates as $\tau = 2.34 \times 10^{-3}$ s; the lifetime is given as infinite for those particles which have infinite lifetimes (except in particle-antiparticle collisions). "Quark comp." gives the composition of the elementary particle in terms of its presumed quark components. Not all unstable hadrons or mesons are listed. Antiparticles to other listed particles are indicated by asterisks (∗), but not all antiparticles are listed. The quantum numbers are for relative charge ($Q$), spin ($S$), parity ($P$), isospin ($T$), $z$-component of isospin ($T_z$), baryon number (Ba), strangeness (St), charm (Ch), beauty or bottomness (By), and truth or topness (Tr).

[a] The muons were first called mu-mesons, but are now better known as muons, leaving the π mesons, or pions, as the particles which may help "carry" the inter-nucleon strong force.

*Source:* Data from Serway [1.3].

**Table 1.3** Standard model of elementary particles.




Note that "gauge" is a misnomer for "phase."

held together by this weakest force, which may be transmitted by a mediating particle called the **graviton** (first detected in 2016) [1.2]. Its range extends to the whole universe. Masses are always positive.

**Force two:** it is the electrical force, which obeys Coulomb's law (1785) [1.4]:

$$\boldsymbol{F}_{12} = q_1 q_2 \boldsymbol{r}_{12} / 4\pi\varepsilon_0 r_{12}^3 \tag{1.1.2}$$

which describes the attractive (or repulsive) force $\boldsymbol{F}_{12}$ between two electrical charges $q_1$ and $q_2$ (positive or negative) placed $\boldsymbol{r}_{12}$ apart, where $\varepsilon_0$ is the electrical permittivity of vacuum. The fundamental electrical monopole (electron) is probably infinitely stable; the mediating particle for the electrical force (**photon**) is observed and well understood. Magnetism is usually due to moving electrical charges, but its monopole has never been seen, so magnetism is not really an independent force; atoms have magnetic properties, and in wires the gegenions of electrical currents are dealt with as if they were "stationary," yet the overall charge is zero: hence, *incredible dictu*, magnetism is a special relativistic effect. As explained in Section 2.9, electricity and magnetism are well described by Maxwell's[7] four field equations [1.5].

**Force three:** it is the "weak nuclear" or "Fermi"[8] force (1934) [1.6], which explains the "β decay" of unstable nuclei into stable ones and positron emission. The positron ($\beta^+$ or $e^+$) is the **antiparticle** of the electron ($e^-$);

**Force four:** it is the strongest force in the universe; it is the "**strong nuclear force**," which binds together the nuclei and the constituents of atomic nuclei, but has an extremely narrow range. Indirect experimental evidence exists for a mediating particle (**gluon**). Nucleons (neutrons, protons) and maybe also all nuclei consist of "elementary" particles

7 James Clerk Maxwell (1831–1879).
8 Enrico Fermi (1901–1954).

**Table 1.4** Fundamental constants (gram-mole = molar mass in grams).

| Name | Symbol | SI value |
|---|---|---|
| Gravitational constant | $G$ | $6.673\,84 \times 10^{-11}\ \mathrm{m^3\ kg^{-1}\ s^{-2}}$ |
| Rest mass of electron | $m_e$ | $9.109\,389\,7 \times 10^{-31}\ \mathrm{kg}$ |
| Electrical charge of electron | $e$ | $-1.602\,176\,57 \times 10^{-19}\ \mathrm{C}$ |
| Speed of light | $c$ | $2.997\,924\,58 \times 108\ \mathrm{m\ s^{-1}}$ |
| Planck's[a] constant of action | $h$ | $6.626\,070\,15 \times 10^{-34}\ \mathrm{J\ s}$ |
| Planck's reduced constant of action | $\hbar \equiv h/2\pi$ | $1.054\,571\,6 \times 10^{-34}\ \mathrm{J\ s}$ |
| Free electron $g$-factor | $g_e$ | 2.002 319 304 386 |
| Avogadro's[b] number | $N_A$ | $6.022\,140\,76 \times 10^{23}\ \mathrm{molecules(gram\text{-}mole)^{-1}}$ |
| Gas constant | $R$ | $8.314\,462\,2\ \mathrm{J\ mol^{-1}\ K^{-1}}$ |

[a] Max Planck (1858–1947).
[b] Lorenzo Romano Amedeo Carlo Bernadette Avogadro, Conte di Quaregna e Cerreto (1776–1856).
*Source:* Modified from Mohr et al. [1.7].

**Table 1.5** Other constants.

| Name | Symbol | Equation | SI value |
|---|---|---|---|
| Sommerfeld[a] fine-structure constant | $\alpha$ | $= e^2/2\varepsilon_0 ch$ | $1/137.035\,999 (= e^2/\hbar c$ in esu) |
| Gravitational acceleration at sea level at Equator | $g$ | | $9.780\,31\ \mathrm{m\ s^{-2}}$ |
| Boltzmann[b] constant | $k_B$ | $= R/N_A$ | $1.380\,648\,8 \times 10^{-23}\ \mathrm{J\ K^{-1}}$ |
| Electrical permittivity of vacuum | $\varepsilon_0$ | $= m_0^{-1} c^{-2}$ | $8.854\,187\,817 \times 10^{-12}\ \mathrm{F\ m^{-1}}$ |
| Magnetic permeability of vacuum | $\mu_0$ | $= 4\pi 10^{-7}$ | $1.256\,637\,061\,4 \times 10^{-6}\ \mathrm{N\ A^{-2}}$ |
| von Klitzing[c]–Landauer[d] quantum of resistance | $R_0$ | $= he^{-2}$ | 25 812.807 443 Ω |
| Magnetic flux quantum | $\Phi_0$ | $= h/2e$ | $2.067\,833\,76 \times 10^{-15}\ \mathrm{Wb}$ |

[a] Arnold Johannes Wilhelm Sommerfeld (1868–1951).
[b] Ludwig Boltzmann (1844–1906).
[c] Klaus von Klitzing (1940–).
[d] Rolf William Landauer (1927–1999).
*Source:* Modified from Mohr et al. [1.7].

called **quarks**, which have never been seen free (proton–proton scattering experiments do show that protons consist of "lumps": this may have been the first experimental evidence for quarks).

> **Sideline 1.4** The name "quark" comes from a sentence in Joyce's[9] *Finnegan Wake*; a free quark has never been isolated, but physicists have not looked in German grocery stores, where Quark is a well-known soft cheese!

Some important constants are given in Tables 1.4 and 1.5.

Other constants and unit conversions are given in Tables 1.6–1.12. The known nuclides are listed in Table X1 (please refer www.wiley/go/metzger/physchemtoolbox2). We next discuss in more detail gravity, the weak force, and the strong force.

The electrical force is delayed to Part Three (Section 1.9) until after a mathematical interlude is given in Part Two (Sections 1.5–1.8).

9 James Augustine Aloysius Joyce (1882–1941).

**Table 1.6** SI base units[a].

| Name | Symbol | Unit | Abbreviation |
|---|---|---|---|
| Length | $L$ | meter | m |
| Mass | $M$ | kilogram | kg |
| Time | $T$ | second | s |
| Electric current | $J$ | ampère | A |
| Temperature | $t$ | Kelvin | K |
| Amount of substance | | mole | mol |
| Luminous intensity | | candela | cd |

[a] Definitions: 1 m ≈ 1/40 000 000 of circumference of Earth (1790 definition) = 1 650 763.73 wavelengths of the orange-red line of $^{86}$Kr under specified conditions (1960 definition); 1 s = 1/86 400th of a mean Earth day; 1 kg ≈ mass of 1000 cm$^3$ of $H_2O$ at 4 °C and 1 atm (1790 definition) = mass of a standard Pt–Ir alloy cylinder in Sèvres, France; 1 A = constant current which, if maintained in two infinitely long wires of negligible cross section and held parallel 1 m apart in vacuum, would exert a force of $2 \times 10^{-7}$ N per m of length.; 1 mol = Avogadro's number of molecules = $6.022\,140\,76 \times 10^{23}$ molecules (gram-mole)$^{-1}$ = number of molecules per gram (NOT kilogram!); 1 cd = luminous intensity in the perpendicular direction of a surface of 1/600 000 of a square meter of a blackbody at 1773 °C and 1 atm (older definition) = luminous intensity in a given direction of a source of monochromatic radiation of frequency $5.40 \times 10^{14}$ Hz and that has a radiant intensity in that direction of 1/683 W sr$^{-1}$ (1979 definition).

**Table 1.7** Metric multipliers: tera, peta, exa come from tetra, penta, hexa with one letter removed; zetta, yotta come from zeta, iota, with one letter added; atto comes from Danish "atten" for 18; zepto comes from Latin "septem" or Greek "hepta" ($1000^{-7}$); and yocto comes from Greek "octo" ($1000^{-8}$) with one prefix letter added.

| | | | | | |
|---|---|---|---|---|---|
| deca | da | $\times 10^{+1}$ | deci | d | $\times 10^{-1}$ |
| hecto | h | $\times 10^{+2}$ | centi | c | $\times 10^{-2}$ |
| kilo | k | $\times 10^{+3}$ | milli | m | $\times 10^{-3}$ |
| mega | M | $\times 10^{+6}$ | micro | μ | $\times 10^{-6}$ |
| giga | G | $\times 10^{+9}$ | nano | n | $\times 10^{-9}$ |
| tera | T | $\times 10^{+12}$ | pico | p | $\times 10^{-12}$ |
| peta | P | $\times 10^{+15}$ | femto | f | $\times 10^{-15}$ |
| exa | E | $\times 10^{+18}$ | atto | a | $\times 10^{-18}$ |
| zetta | Z | $\times 10^{+21}$ | zepto | z | $\times 10^{-21}$ |
| yotta | Y | $\times 10^{+24}$ | yocto | y | $\times 10^{-24}$ |

**Table 1.8** **Atomic units** (using SI values for $e$, $m_e$, $h$, $c$).

| Quantity | Description | Symbol | Size (in SI units) |
|---|---|---|---|
| Charge | Charge of the electron | $e$ | $-1.602\,176\,57 \times 10^{-19}$ A s |
| Speed | Speed of light in vacuum | $c$ | $2.997\,924\,58 \times 10^{8}$ m s$^{-1}$ |
| Velocity | Velocity of electron in first Bohr orbit | $e^2/2\varepsilon_0 h$ | $2.1877 \times 10^{6}$ m s$^{-1}$ |
| Mass | Rest mass of the electron | $m_e$ | $9.109\,389\,7 \times 10^{-31}$ kg |
| Length | Radius of first Bohr orbit (1 b = $a_0$) | $10^7 (h/2\pi ec)^2 m_e^{-1}$ | $5.291\,772\,5 \times 10^{-11}$ m |
| Time | (From electron momentum in first Bohr orbit) | $2e_0^2 h^3/\pi m_e e^4$ | $2.4189 \times 10^{-17}$ s |
| Gravitation | Newton's constant of universal gravitation | $G$ | $6.673 \times 10^{-11}$ m$^3$ kg$^{-1}$ s$^{-1}$ |
| Energy | 2×H atom ionization energy (1 hartree) | $m_e e^4/4e_0^2 h^2$ | $4.3558 \times 10^{-18}$ J |
| Momentum | Electron momentum in first Bohr orbit | $m_e e^2/2e_0 h$ | $1.9926 \times 10^{-24}$ kg m s$^{-1}$ |

**Table 1.8** (Continued)

| Quantity | Description | Symbol | Size (in SI units) |
|---|---|---|---|
| Frequency | (From electron momentum in first Bohr orbit) | $\pi m_e e^4/2\varepsilon_0 h^3$ | $4.1341 \times 10^{16}\ s^{-1}$ |
| Electrical potential | (From electron momentum in first Bohr orbit) | $\pi^{1/2} m_e e^3/2\varepsilon_0^{3/2} h^2$ | 27.210 V |
| Electrical field strength | (From electron momentum in first Bohr orbit) | $\pi^2 m_e^2 e^4/\varepsilon_0^2 h^4$ | $5.142 \times 10^9$ V cm$^{-1}$ |
| $h$ | Planck's constant (of action) | $h$ | $6.626\,069\,0 \times 10^{-34}$ J s |
| $\varepsilon_0$ | Electron permittivity of vacuum (artificial) | $10^7/4\pi c^2$ | $8.853\,742\,338 \times 10^{-12}$ F m$^{-1}$ (use this number for SI unit conversions; otherwise units: s$^2$ m$^{-2}$) |
| $\mu_0$ | Magnetic permeability of vacuum (artificial) | $4\pi 10^{-7}$ | $1.256\ 637\ 061\ 4 \times 10^{-6}$ N A$^{-2}$ (use this for SI unit conversions; otherwise, pure number; also: H m$^{-1}$) |

**Table 1.9** Other derived constants (using SI values for $e$, $m_e$, $h$, $c$).

| Symbol | Definition | Numerical value |
|---|---|---|
| 1 amu | 1 atomic mass unit $= 10^{-3}/N_A \equiv (1/12)$ of mass of ${}_6C^{12}$ atom | $= 1.660\,540\,2 \times 10^{-23}$ kg |
| $\hbar$ | $\equiv h/2\pi$ | $= 1.054\ 571\ 635 \times 10^{-34}$ J s |
| $e'$ | $\equiv$ Charge on electron in cgs-esu $= 10ec$ | $= 4.803\,206\,799 \times 10^{-10}$ sC |
| 1 eV | $= (N_A e/R) = 11\,604.45$ K $= 10^{-2}\,(e/hc)$ | $= 8065.54$ cm$^{-1}$ $= 8.065\,54$ kK |
| | $= (hc/e)$ | $= 1.239\,842\,447 \times 10^{-6}$ m $= 12\,398.4$ Å $= 1239.84$ nm |
| | $= \{e\}$ | $= 1.602\,176\,57 \times 10^{-19}$ J |
| 1 hartree | $\equiv e'^2/a_0 = 27.211$ eV $= 4\pi 10^{-14}\, m_e e^4 c^4 h^{-2}$ | $= 4.3558 \times 10^{-18}$ J |
| 1 F | $\equiv N_A e$ | $= 96\ 485.309$ C mol$^{-1}$ |
| $\mu_B$ | $= 1$ bohr magneton | $= 9.274\,015\,4 \times 10^{-24}$ J T$^{-1}$ |
| | $\equiv e\underline{\hbar}/2m_e = eh/4\pi m_e$ | $= 9.274\,015\,4 \times 10^{-21}$ erg G$^{-1}$ |
| $G_0 \equiv 2e^2 h^{-1}$ | $= 1$ quantum of electrical conductance $= 1$/von Klitzing/Landauer's constant | $= 7.748\,091\,696 \times 10^{-5}$ S |
| $\Phi_0 \equiv h/2e$ | $= 1$ quantum of magnetic flux | $= 2.067\,833\,636 \times 10^{-15}$ Wb |
| $4\pi\varepsilon_0$ | $= 10^7 c^{-2}$ | $= 1.112\,650\,056 \times 10^{-10}$ F m$^{-1}$ (use this for SI unit conversions; otherwise: s$^2$ m$^{-2}$) |
| $\mu_0\varepsilon_0$ | $\equiv c^{-2}$ | $= 1.112\,650\,056 \times 10^{-17}$ s$^2$ m$^{-2}$ |
| $(\mu_0/\varepsilon_0)^{1/2}$ | $= \mu_0 c = 4\pi 10^{-7} c = Z_0$ (characteristic impedance of vacuum) | $= 376.730\,313\,461\ \Omega$ (for SI unit conversions; otherwise: m s$^{-1}$) |
| $\alpha$ | $\equiv$ Sommerfeld fine-structure constant | $= 7.297\ 353\ 080 \times 10^{-3}$ (pure #) in esu |
| | $\equiv e^2/\hbar c$ (using esu values for $e$,$\hbar$, $c$) | $= 1/137.035\,989\,5$ (pure #) in esu |
| | $\equiv e^2/4\pi\varepsilon_0 c\hbar$ (using their SI values) | $= 1/137.035\,989\,5$ (pure #) in SI |

**Table 1.10** Other conversion factors.

| Unit | Conversion | Expression in SI base units | Expression in other SI units |
|---|---|---|---|
| 1 Å(ngstrom)[a] | = 0.1 nm | $= 10^{-10}$ m(eters) | |
| 1 bohr | = 0.052 917 725 nm = 0.529 177 25 Å | $= 5.291\,772\,5 \times 10^{-11}$ m | |
| 1 K[b] = Kayser | = 1 wavenumber | | $= 1\ \mathrm{cm}^{-1}$ |
| 1 Hz | $= (1/2\pi)$ cycles $\mathrm{s}^{-1}$ | $= 1$ (cycle) $\mathrm{s}^{-1}$ | |
| 1 N = Newton | $= 10^5$ dyn | $= 1\ \mathrm{kg\ ms}^{-2}$ | $= 1\ \mathrm{J\ m}^{-1}$ |
| 1 dyn | $= (1\ \mathrm{abA})^2\mathrm{cm}^{-2}$ | | |
| 1 dyn | $= (1\ \mathrm{statC})^2\mathrm{cm}^{-2}$ | | |
| 1 Pa = Pascal[c] | = 0.007 500 616 Torr | $= 1\ \mathrm{kg\ m}^{-1}\ \mathrm{s}^{-2}$ | $= 1\ \mathrm{N\ m}^{-2}$ |
| 1 (standard) atm | = 101.325 kPa = 0.101 325 MPa | | = 760.000 0 Torr |
| 1 P = Poise[d] | = cgs unit of viscosity | $= 1\ \mathrm{g\ cm}^{-1}\ \mathrm{s}^{-1}$ | = 0.1 Pa s |
| 1 J = Joule | $= 10^7$ erg | $= 1\ \mathrm{kg\ m}^2\ \mathrm{s}^{-2}$ | = 1 Ws = 1 N m |
| 1 erg | $= 1\ \mathrm{statC}^2\ \mathrm{cm}^{-1}$ | $= 10^{-7}$ J | = 1 sC sV |
| 1 cal(orie) | = 1 g-cal | = 4.184 00 J | |
| 1 kcal | = 1 kg-cal | $= 4.184\,00 \times 10^3$ J | |
| 1 eV(olt) | $= 1.602\,176\,57 \times 10^{-19}$ J | $= 1.602\,176\,57 \times 10^{-12}$ erg | |
| 1 eV/molecule | = 96.485 kJ/mol | | |
| 1 hartree | = 27.211 eV | | |
| 1 C = Coulomb | | = 1 A s | = 1 C V s |
| 1 s(tat)C(coulomb) | $= 1\ \mathrm{g}^{1/2}\ \mathrm{cm}^{3/2}$ s(econd)$^{-1}$ | | |
| 1 abC(coulomb) | = 10 C | | |
| 1 V = Volt | | $= 1\ \mathrm{kg\ m}^2\ \mathrm{A}^{-1}\ \mathrm{s}^{-3}$ | $= 1\ \mathrm{W\ A}^{-1}$ |
| 1 s(tat)V(olt) | $= 1$ s(tat)C(oul) $\mathrm{cm}^{-1}$ | | |
| 1 abV(olt) | $= 10^8$ V | | |
| 1 Ohm = Ω | $= 1\ \mathrm{V\ A}^{-1}$ | $= 1\ \mathrm{kg\ m}^2\ \mathrm{A}^{-2}\ \mathrm{s}^{-3}$ | $= 1\ \mathrm{W\ A}^{-2}$ |
| 1 abΩ | $= 10^{-9}\,\Omega$ | | |
| 1 S = Siemens | $= 1\ \Omega^{-1}$ | $= 1\ \mathrm{kg}^{-1}\ \mathrm{m}^{-2}\ \mathrm{A}^2\ \mathrm{s}^3$ | $= 1\ \mathrm{A\ V}^{-1}$ |
| 1 F = Farad | $= 1$ C(oulomb)$^2$ J(oule)$^{-1}$ | $= 1\ \mathrm{kg}^{-1}\ \mathrm{m}^{-2}\ \mathrm{A}^2\ \mathrm{s}^4$ | $= 1\ \mathrm{C\ V}^{-1}$ |
| 1 H = Henry | $= 10^9$ emu of inductance | $= 1\ \mathrm{kg\ m}^2\ \mathrm{A}^{-2}\ \mathrm{s}^{-2}$ | $= 1\ \mathrm{Wb\ A}^{-1} = 1\ \mathrm{V\ s\ A}^{-1}$ |
| 1 W = Watt[e] | $= 1$ J(oule) $\mathrm{s}^{-1}$ | $= 1\ \mathrm{A\ V\ s}^{-1}$ | $= 1\ \mathrm{C\ V\ s}^{-2}$ |
| 1 W(att) | | $= 1\ \mathrm{kg\ m}^2\ \mathrm{s}^{-3}$ | $= 1\ \mathrm{J\ s}^{-1} = 1\ \mathrm{A\ V}$ |
| 1 Wb = Weber | | $= 1\ \mathrm{kg\ m}^2\ \mathrm{A}^{-1}\ \mathrm{s}^{-2}$ | = 1 V s |
| 1 T = Tesla | $= 10^4$ G(auss) | $= 1\ \mathrm{kg\ A}^{-1}\ \mathrm{s}^{-2}$ | $= 1\ \mathrm{Wb\ m}^{-2} = 1\ \mathrm{V\ s\ m}^{-2}$ |
| 1 Oe | $= 1\ \mathrm{g}^{1/2}\ \mathrm{cm}^{-1/2}$ s(econd)$^{-1}$ | | |
| 1 $\mathrm{Oe}^2\ \mathrm{cm}^{-3}$ | = 1 erg | $= 10^{-7}$ J | |
| 1 au time | $= 2.419 \times 10^{-17}$ s $= \left(2\varepsilon_0^2 h^3/\pi m_e e^4\right)$ | | |
| 1 au electric field | = electric field on electron 1 bohr radius from proton $= ea_0^{-2} = 1.7 \times 10^7\ \mathrm{sV\ cm}^{-1} = 5.142 \times 10^{11}\ \mathrm{V\ m}^{-1}$ | | |
| 1 au magnetic field | = (wave or Gaussian convention) $= a\hbar e^{-1}a_0^{-2}$ $= 1.716 \times 10^3$ T | | |
| 1 au magnetic field | = (Lorentz convention) $= 2.5 \times 10^5$ T | | |
| 1 L[f] | = 1 monolayer | | |
| 1 Da = Dalton | = 1 atomic mass unit = mass, in g, of $6.022 \times 10^{23}$ hydrogen atoms | | |

**Table 1.10** (Continued)

| Unit | Conversion | Expression in SI base units | Expression in other SI units |
|---|---|---|---|
| 1 dirac | = 1 word/h (after the taciturn Paul A.M. Dirac) | | |
| 1 helen[g] | = beauty that can launch 1000 ships (after Helen of Troy) | | |

[a] Anders Jonas Ångström (1814–1874).
[b] Heinrich Gustav Johannes Kayser (1853–1940).
[c] Blaise Pascal (1623–1682).
[d] Jean Léonard Marie Poiseuille (1799–1869).
[e] James Watt (1736–1819).
[f] Irving Langmuir (1881–1957).
[g] Helen of Sparta and Troy, daughter of Zeus and Leda; mythical (c. 1250 BC).

**Table 1.11** Definitions of the electric field $\boldsymbol{E}$, the (di)electric polarization $\boldsymbol{P}$, the electric displacement $\boldsymbol{D}$, the magnetic field $\boldsymbol{H}$, the magnetization $\boldsymbol{M}$, the magnetic induction or flux density $\boldsymbol{B}$, statement of the Maxwell equations, and of the Lorentz force equation in various systems of units; rat. = rationalized (no $4\pi$), unrat. = the explicit factor $4\pi$ is used in the definition of dielectric polarization and magnetization; $c$ = speed of light (using SI values for $e$, $m_e$, $h$, $c$) [1.29].

| System | Constitutive equations | Maxwell equations | Lorentz force | Units |
|---|---|---|---|---|
| cgs-esu (unrat.) | $\boldsymbol{D} = \boldsymbol{E} + 4\pi\boldsymbol{P}$ | $\nabla \bullet \boldsymbol{D} = 4\pi\rho \quad \nabla \times \boldsymbol{B} = 4\pi\boldsymbol{J} + \partial\boldsymbol{D}/\partial t$ | $\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}$ | $\boldsymbol{E}$ sV cm$^{-1}$<br>$\boldsymbol{P}$ sC cm$^{-2}$<br>$\boldsymbol{D}$ sC cm$^{-2}$ |
| | $\boldsymbol{H} = c^2\boldsymbol{B} - 4\pi\boldsymbol{M}$ | $\nabla \bullet \boldsymbol{B} = 0 \quad \nabla \times \boldsymbol{D} + \partial\boldsymbol{B}/\partial t = 0$ | | $\boldsymbol{H}$ sA cm$^{-1}$<br>$\boldsymbol{M}$ sA cm$^{-1}$<br>$\boldsymbol{B}$ statT |
| cgs-emu (unrat.) | $\boldsymbol{D} = c^{-2}\boldsymbol{E} + 4\pi\boldsymbol{P}$ | $\nabla \bullet \boldsymbol{D} = 4\pi\rho \quad \nabla \times \boldsymbol{B} = 4\pi\boldsymbol{J} + \partial\boldsymbol{D}/\partial t$ | $\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}$ | $\boldsymbol{E}$ abV cm$^{-1}$<br>$\boldsymbol{P}$ abC cm$^{-2}$ |
| | $\boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M}$ | $\nabla \bullet \boldsymbol{B} = 0 \quad \nabla \times \boldsymbol{D} + \partial\boldsymbol{B}/\partial t = 0$ | | $\boldsymbol{D}$ abC cm$^{-2}$<br>$\boldsymbol{H}$ Oe |
| | $\boldsymbol{P} = \chi_{\text{emu}}\boldsymbol{E}$<br>$\boldsymbol{B} = \mu_{\text{emu}}\boldsymbol{H}$<br>$\boldsymbol{M} = \chi_{\text{emu}}\boldsymbol{H}$ | | | $\boldsymbol{M}$ G<br>$\boldsymbol{B}$ G |
| Gaussian (cgs, unrat.) | $\boldsymbol{D} = \boldsymbol{E} + 4\pi\boldsymbol{P}$ | $\nabla \bullet \boldsymbol{D} = 4\pi\rho \quad \nabla \times \boldsymbol{B} = 4\pi c^{-1}\boldsymbol{J} + c^{-1}\partial\boldsymbol{D}/\partial t$ | $\boldsymbol{E} + c^{-1}\boldsymbol{v} \times \boldsymbol{B}$ | $\boldsymbol{E}$ sV cm$^{-1}$<br>$\boldsymbol{P}$ sC cm$^{-2}$<br>$\boldsymbol{D}$ sC cm$^{-2}$ |
| | $\boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M}$ | $\nabla \bullet \boldsymbol{B} = 0 \quad \nabla \times \boldsymbol{D} + c^{-1}\partial\boldsymbol{B}/\partial t = 0$ | | $\boldsymbol{H}$ Oe<br>$\boldsymbol{M}$ G<br>$\boldsymbol{B}$ G |
| Heaviside-Lorentz (cgs, rat.) | $\boldsymbol{D} = \boldsymbol{E} + \boldsymbol{P}$ | $\nabla \bullet \boldsymbol{D} = \rho \quad \nabla \times \boldsymbol{B} = c^{-1}\boldsymbol{J} + c^{-1}\partial\boldsymbol{D}/\partial t$ | $\boldsymbol{E} + c^{-1}\boldsymbol{v} \times \boldsymbol{B}$ | $\boldsymbol{E}$ sV cm$^{-1}$<br>$\boldsymbol{P}$ sC cm$^{-2}$<br>$\boldsymbol{D}$ sC cm$^{-2}$ |
| | $\boldsymbol{H} = \boldsymbol{B} - \boldsymbol{M}$ | $\nabla \bullet \boldsymbol{B} = 0 \quad \nabla \times \boldsymbol{D} + c^{-1}\partial\boldsymbol{B}/\partial t = 0$ | | $\boldsymbol{H}$ Oe<br>$\boldsymbol{M}$ G<br>$\boldsymbol{B}$ G |
| Atomic units (unrat.) (Hartree; Gauss convention) | $\boldsymbol{D} = \boldsymbol{E} + 4\pi\boldsymbol{P}$ | $\nabla \bullet \boldsymbol{D} = 4\pi\rho \quad \nabla \times \boldsymbol{B} = \alpha(4\pi\boldsymbol{J} + \partial\boldsymbol{D}/\partial t)$ | $\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}$ | $\boldsymbol{E}$ e b$^{-2}$<br>$\boldsymbol{P}$ e b$^{-2}$<br>$\boldsymbol{D}$ e b$^{-2}$ |
| | $\boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M}$ | $\nabla \bullet \boldsymbol{B} = 0 \quad \nabla \bullet \boldsymbol{D} + \alpha\partial\boldsymbol{B}/\partial t = 0$ | | $\boldsymbol{H}$ e b$^{-2}$<br>$\boldsymbol{M}$ e b$^{-2}$<br>$\boldsymbol{B}$ e b$^{-2}$ |

(*Continued*)

**Table 1.11** (Continued)

| System | Constitutive equations | Maxwell equations | Lorentz force | Units |
|---|---|---|---|---|
| Atomic units (unrat.) (Hartree; Lorentz convention) | $\boldsymbol{D} = \boldsymbol{E} + 4\pi\boldsymbol{P}$ | $\nabla \bullet \boldsymbol{D} = 4\pi\rho \; \nabla \times \boldsymbol{B} = 4\pi\boldsymbol{J} + \partial\boldsymbol{D}/\partial t$ | $\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}$ | $\boldsymbol{E}$ e b$^{-2}$<br>$\boldsymbol{P}$ e b$^{-2}$<br>$\boldsymbol{D}$ e b$^{-2}$ |
| | $\boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M}$ | $\nabla \bullet \boldsymbol{B} = 0 \; \nabla \bullet \boldsymbol{D} + \partial\boldsymbol{B}/\partial t = 0$ | | $\boldsymbol{H}$ e b$^{-2}$<br>$\boldsymbol{M}$ e b$^{-2}$<br>$\boldsymbol{B}$ e b$^{-2}$ |
| SI and SI′ (rat.) | $\boldsymbol{D} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P}$ | $\nabla \bullet \boldsymbol{D} = \rho \; \nabla \times \boldsymbol{B} = \boldsymbol{J} + \partial\boldsymbol{D}/\partial t$ | $\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}$ | $\boldsymbol{E}$ V m$^{-1}$<br>$\boldsymbol{P}$ C m$^{-2}$<br>$\boldsymbol{D}$ C m$^{-2}$ |
| | $\boldsymbol{H} = \mu_0^{-1}\boldsymbol{B} - \boldsymbol{M}$ | $\nabla \bullet \boldsymbol{B} = 0 \; \nabla \bullet \boldsymbol{D} + \partial\boldsymbol{B}/\partial t = 0$ | | $\boldsymbol{H}$ A m$^{-1}$<br>$\boldsymbol{M}$ A m$^{-1}$<br>$\boldsymbol{B}$ T |

For Hartree atomic units of magnetism, two conventions exist: (i) the "Gauss" or wave convention, which requires that $\boldsymbol{E}$ and $\boldsymbol{H}$ have the same magnitude for electromagnetic waves *in vacuo*; (ii) the Lorentz convention, which derives the magnetic field from the Lorentz force equation; the ratio between these two sets of units is the Sommerfeld fine-structure constant $\alpha = 1/137.035\,989\,5$.

**Table 1.12** Conversion factors (using SI values for $e$, $m_e$, $\hbar$, and $c$).

| Symbol | Name | SI | esu (→SI) | emu (→SI) | au (→SI) |
|---|---|---|---|---|---|
| $L$ | Length (do not confuse this with inductance L) | m | cm<br>$\rightarrow 10^{-2}$ | cm<br>$\rightarrow 10^{-2}$ | b $\equiv a_0$<br>$\rightarrow 5.292 \times 10^{-11}$<br>$= 10^7 (\hbar/ec)^2 m_e^{-1}$ |
| $m$ | Mass | kg | g<br>$\rightarrow 10^{-3}$ | g<br>$\rightarrow 10^{-3}$ | electron mass $\equiv m_e$<br>$\equiv 9.109 \times 10^{-31}$ |
| $t$ | Time | s | s | s | $\hbar$ (hartree)<br>$\rightarrow 2.419 \times 10^{-17}$<br>$= 2\varepsilon_0^2 h^3/\pi m_e e^4$ |
| $v$ | Speed | m s$^{-1}$ | cm s$^{-1}$<br>$\rightarrow 10^{-2}$ | cm s$^{-1}$<br>$\rightarrow 10^{-2}$ | $e^2/\hbar$<br>$\rightarrow 2.1877 \times 10^6$<br>$= e^2/2\varepsilon_0 h$ |
| $V$ | Volume (do not confuse this with V = Volt) | m$^3$ | cm$^3$<br>$\rightarrow 10^{-6}$ | cm$^3$<br>$\rightarrow 10^{-6}$ | b$^3$<br>$\rightarrow 1.482 \times 10^{-31}$<br>$= a_0^3$ |
| $F$ | Force | N(ewton) | dyn<br>$\rightarrow 10^{-5}$ | dyn<br>$\rightarrow 10^{-5}$ | hartree b$^{-1}$<br>$\rightarrow 8.238 \times 10^{-8}$<br>$= m_e e^4/a_0(4\pi\varepsilon_0\hbar)^2$ |
| $W$ | Work, energy | J(oule) | erg<br>$\rightarrow 10^{-7}$ | erg<br>$\rightarrow 10^{-7}$ | hartree<br>$\rightarrow 4.359 \times 10^{-18}$<br>$= e^2/4\pi\varepsilon_0 a_0$<br>$= m_e e^4/4\varepsilon_0^2 h^2$ |
| (The hartree is twice the ionization energy of the hydrogen atom) | | | | | |
| $A$ | Action (do not confuse this with A = Ampère) | J s | erg s<br>$\rightarrow 10^{-7}$ | erg s<br>$\rightarrow 10^{-7}$ | $\hbar$ = Planck constant/2π<br>$\rightarrow 1.054 \times 10^{-34}$<br>$= h/2\pi$ |
| $q$ | Electric charge | C(oulomb) | s(tat)C(oulomb)<br>$\rightarrow 3.336 \times 10^{-10}$<br>$= 10^{-1} c^{-1}$ | abC<br>$\rightarrow 10$ | Electron charge $\equiv e$<br>$\rightarrow 1.602 \times 10^{-19}$<br>$\equiv e$ |
| $\rho$ | Electric charge density | C m$^{-3}$ | sC cm$^{-3}$<br>$\rightarrow 3.336 \times 10^{-4}$<br>$= 10^5 c^{-1}$ | abC cm$^{-3}$<br>$\rightarrow 10^{-5}$ | e b$^{-3}$<br>$\rightarrow 1.081 \times 10^{12}$<br>$= e a_0^{-3}$ |
| $m_{el}$ | Electrical dipole moment | C m | sC cm<br>$\rightarrow 3.334 \times 10^{-8}$<br>$= 10 c^{-1}$ | abC cm<br>$\rightarrow 0.1$ | electron bohr<br>$\rightarrow 8.478 \times 10^{-30}$<br>$= e a_0$ |

**Table 1.12** (Continued)

| Symbol | Name | SI | esu (→SI) | emu (→SI) | au (→SI) |
|---|---|---|---|---|---|
| (NB: 1 Debye = $3.336 \times 10^{-30}$ C m = 4.803 sC cm = 2.541 e b = 4.803 e Å) | | | | | |
| $Q_{el}$ | Electrical quadrapole moment | C $m^2$ | sC $cm^2$ $\rightarrow 3.336 \times 10^{-14} = \rightarrow 10^{-3} = 10^{-5}c^{-1}$ | abC $cm^2$ $\rightarrow 10^{-3}$ | au $\rightarrow 4.487 \times 10^{-40} = ea_0^2$ |
| $I$ | Electric current | A(mpère) | statC $s^{-1}$ $\rightarrow 3.336 \times 10^{-10} = 10^{-1}c^{-1}$ | abA = Biot $\rightarrow 10$ | $e$/"au time" $\rightarrow 6.624 \times 10^{-3} = e^3/4\pi\varepsilon_0\hbar a_0$ |
| $J$ | Electric current density (volume) | A $m^{-2}$ | statC $s^{-1}$ $cm^{-2}$ $\rightarrow 3.336 \times 10^{-6} = 10^3c^{-1}$ | abA $cm^{-2}$ $\rightarrow 10^3$ | e $b^{-2}$("au time")$^{-1}$ $\rightarrow 2.365 \times 10^{18} = 1.28 \times 10^{-26}\pi^7 m_e^3 c^8 e^9 h^{-7}$ |
| $E$ | Electric field | V $m^{-1}$ | statV $cm^{-1}$ $\rightarrow 2.998 \times 10^4 = 10^{-4}c$ | abV $cm^{-1}$ $\rightarrow 1 \times 10^{-6}$ | e $b^{-2}$ $\rightarrow 5.142 \times 10^{11} = e/4\pi\varepsilon_0 a_0^{-2}$ |
| $P$ | (Di)electric polarization | C $m^{-2}$ | statC $cm^{-2}$ $\rightarrow 3.336 \times 10^{-6} = 10^3c^{-1}$ | abC $cm^{-2}$ $\rightarrow 10^{-3}$ | e $b^{-2}$ $\rightarrow 5.142 \times 10^{11} = ea_0^{-2}$ |
| $D$ | Electric displacement dielectric induction | C $m^{-2}$ | sC $cm^{-2}$ $\rightarrow 2.653 \times 10^{-7} = 10^3/4\pi c$ | abC $cm^{-2}$ $\rightarrow 10^{-3}$ | e $b^{-2}$ $\rightarrow 5.142 \times 10^{11} = ea_0^{-2}$ |
| $\varepsilon$ | Dielectric constant | pure # | pure # | pure # | pure # |
| $V$ | Electric potential (do not confuse this with V = volume) | V(olt) | statV = sC $cm^{-1}$ $\rightarrow 2.998 \times 10^2 = 10^{-6}c$ | abV $\rightarrow 10^{-8}$ | e $b^{-1}$ $\rightarrow 27.211\ 39 = e/4\pi\varepsilon_0 a_0$ |
| $\alpha$ | Linear electric polarizability | $m^3$ | $cm^3$ $\rightarrow 1.257 \times 10^{-5} = 4\pi 10^{-6}$ | $cm^3$ $\rightarrow 1.257 \times 10^{-5} = 4\pi 10^{-6}$ | $b^3$ (Gauss convention) $\rightarrow 1.862 \times 10^{-30} = 4\pi a_0^3$ |
| | | SI′: C $m^2$ $V^{-1}$ [SI′/SI = $\varepsilon_0$ = $8.854 \times 10^{-12}$] | | | au (Gauss convention) $\rightarrow 1.649 \times 10^{-41} = 4\pi\varepsilon_0 a_0^3 = e^2 a_0^2$/hartree |
| $\beta$ | First-order nonlinear electric hyperpolarizability | $m^4$ $V^{-1}$ | $cm^5$ $abC^{-1}$ $\rightarrow 4.192 \times 10^{-10} = 4\pi 10^{-2}c^{-1}$ | $cm^5$ $abC^{-1}$ $\rightarrow 2.998 \times 10^4 = 4\pi$ | au (Gauss convention) $\rightarrow 3.622 \times 10^{-42} = 16\pi^2\varepsilon_0 a_0^5 e^{-1}$ |
| | | SI′: C $m^3$ $V^{-2}$ [SI′/SI = $\varepsilon_0$ = $8.854 \times 10^{-12}$] | | | au (Gauss convention) $\rightarrow 3.206 \times 10^{-53} = e^3 a_0^3$/hartree$^2$ $= 16\pi^2 e_0^2 a_0^5 e^{-1}$ |
| $\gamma$ | Second-order nonlinear electric hyperpolarizability | SI: $m^5$ $V^{-2}$ | $cm^7$ $sC^{-2}$ $1.398 \times 10^{-14} = 4\pi 10^2 c^{-2}$ | $cm^7$ $abC^{-2}$ $\rightarrow 1.256 \times 10^{-5} = 4\pi 10^6$ | au (Gauss convention) $\rightarrow 7.042 \times 10^{-54} = 256\pi^3 e_0^2 a_0^7 e^{-2}$ |
| | | SI′: C $m^4$ $V^{-3}$ [SI′/SI = $\varepsilon_0$ = $8.854 \times 10^{-12}$] | | | au (Gauss convention) $\rightarrow 6.235 \times 10^{-65} = e^4 a_0^4$/hartree$^3$ $= 256\pi^3 e_0^3 a_0^7 e^{-2}$ |
| $\chi^{(1)}$ | First-order linear electrical susceptibility | SI: pure # | esu: pure # | emu: pure # | pure # |
| | | SI′: F $m^{-1}$ [SI′/SI = $\varepsilon_0$ = $8.854 \times 10^{-12}$] | | | |
| $\chi^{(2)}$ | Second-order nonlinear electrical susceptibility | SI: m $V^{-1}$ | sC $sV^{-2}$ $\rightarrow 4.192 \times 10^{-4} = 4\pi 10^4 c^{-1}$ | abC ab $V^{-2}$ $\rightarrow 5.171 \times 10^{-30} = 4\pi 10^{12} c^{-5}$ | e(e $b^{-1}$)$^{-2}$ $\rightarrow 2.444 \times 10^{-11} = 16\pi^2 e_0^2 a_0^2 e^{-1}$ |
| | | SI': C $V^{-2}$ [SI′/SI = $\varepsilon_0$ = $8.854 \times 10^{-12}$] | | | |

(*Continued*)

**Table 1.12** (Continued)

| Symbol | Name | SI | esu (→SI) | emu (→SI) | au (→SI) |
|---|---|---|---|---|---|
| $\chi^{(3)}$ | Third-order nonlinear electrical susceptibility | SI: $m^2\,V^{-2}$<br>SI′: C m $V^{-3}$<br>[SI′/SI = $\varepsilon_0$ = $8.854\times10^{-12}$] | cm sC $sV^{-3}$<br>$\rightarrow1.398\times10^{-8}$<br>$=4\pi10^{8}c^{-2}$ | cm abC $abV^{-3}$<br>$\rightarrow8.99\times10^{22}$<br>$=10^{6}c^{2}$ | e b(e $b^{-1}$)$^{-3}$<br>$\rightarrow4.208\times10^{-34}$<br>$=64\pi^{3}e_0^{3}a_0^{4}e^{-2}$ |
| $R$ | Electrical resistance | Ω | s $cm^{-1}$<br>$1.113\times10^{12}$<br>$=10^{5}c^{-2}$ | abΩ<br>$\rightarrow10^{-9}$<br>$=e^{2}a_0$ | au<br>$\rightarrow4.878\times10^{14}$<br>$=h^{-1}$ |
| $G$ | Electrical conductance | S = mho | cm $s^{-1}$<br>$8.988\times10^{11}$<br>$=10^{-5}c^{2}$ | abS<br>$\rightarrow10^{9}$ | au<br>$\rightarrow2.050\times10^{-15}$<br>$=e^{-2}ha_0^{-1}$ |
| $\rho$ | Electrical resistivity (volume) | Ω m | s<br>$\rightarrow8.985\times10^{9}$<br>$=10^{-7}c^{2}$ | abΩ cm<br>$\rightarrow10^{-11}$ | Quantum of resistivity<br>$\rightarrow2.5813\times10^{4}$<br>$=he^{-2}$ |
| $\sigma$ | Electrical conductivity (volume) | S $m^{-1}$ | $s^{-1}$<br>$\rightarrow1.113\times10^{-10}$<br>$=10^{7}c^{-2}$ | abΩ$^{-1}$ $cm^{-1}$<br>$\rightarrow10^{-7}$ | Quantum of conductance<br>$\rightarrow3.874\times10^{-5}$<br>$=e^{2}h^{-1}$ |
| $C$ | Electrical capacitance (do not confuse this with C = Coulomb) | F(arad) | statF = cm<br>$\rightarrow1.113\times10^{-12}$<br>$=10^{5}c^{-2}$ | abF<br>$\rightarrow10^{9}$ | b<br>$\rightarrow5.29\times10^{-11}$<br>$=a_0$ |
| $L$ | Electrical inductance | H(enry) | statH = $cm^{-1}\,s^{-2}$<br>$\rightarrow1.113\times10^{-8}$<br>$=10^{9}c^{-2}$ | abH<br>$\rightarrow10^{-9}$ | $b^{-1}$(au time)$^{-2}$<br>$\rightarrow3.229\times10^{43}$<br>$=10^{-7}\pi^{4}\varepsilon_0^{-4}m_e^{3}e^{10}c^{2}h^{-8}$ |
| $m$ | Magnetic dipole moment (loop; do not confuse this with $m$ = mass) | A $m^2$ = J $T^{-1}$ | statA $cm^2$<br>$\rightarrow3.34\times10^{-14}$<br>$=10^{-5}c^{-1}$ | abA $cm^2$<br>= erg $G^{-1}$<br>$\rightarrow10^{-3}$ | au (Gauss convention)<br>$\rightarrow1.354\times10^{-25}$<br>$=\alpha e\hbar m_e^{-1}$<br>au (Lorentz convention)<br>$\rightarrow1.855\times10^{-23}$<br>$=e\hbar m_e^{-1}$ |
| $j$ | Magnetic dipole moment (if $B=\mu_0H+J$, where $J=j\times$volume) | Wb m<br>= V s m<br>= J m $A^{-1}$ | statT $cm^3$<br>$\rightarrow3.77\times10^{11}$<br>$=12\pi10^{10}$ | emu<br>$\rightarrow1.26\times10^{-9}$<br>$=4\pi10^{-10}$ | au (Lorentz convention)<br>$\rightarrow1.855\times10^{-23}$<br>$=e\hbar/m_e$ |
| $\mu_B$ | Bohr magneton | J $T^{-1}$ | erg stat $T^{-1}$<br>$\rightarrow10^{-3}$ | erg $G^{-1}$<br>$\rightarrow10^{-3}$ | au (Gauss convention)<br>$\rightarrow6.769\times10^{-26}$<br>$=\alpha e\hbar/2m_e$<br>au (Lorentz convention)<br>$\rightarrow9.274\times10^{-24}$<br>$=e\hbar/2m_e$ |
| $\Phi$ | Magnetic flux | Wb = V s<br>= J $A^{-1}$ | statT $cm^2$<br>$\rightarrow2.998$<br>$=10^{-8}c$ | M(a)x(well)<br>$\rightarrow10^{-8}$<br>$=10^{-8}$ | au (Gauss convention)<br>$\rightarrow4.803\times10^{-18}$<br>$=\alpha\hbar e^{-1}=ec\times10^{-7}$<br>au (Lorentz convention)<br>$\rightarrow6.583\times10^{-16}$<br>$=\hbar e^{-1}$ |
| $B$ | Magnetic induction = magnetic flux density = magnetic field | T(esla)<br>=Wb $m^{-2}$<br>= J $A^{-1}\,m^{-2}$<br>= V s $m^{-2}$ | statT<br>$\rightarrow2.998\times10^{6}$<br>$=10^{-2}c$ | Gauss = abT<br>= Mx $cm^{-2}$<br>$\rightarrow10^{-4}$ | au (Gauss convention)<br>$\rightarrow1.716\times10^{3}$<br>$=\alpha\hbar e^{-1}\,a_0^{-2}$<br>au (Lorentz convention)<br>$\rightarrow2.351\times10^{5}$<br>$=\hbar e^{-1}\,a_0^{-2}$ |
| $H$ | Magnetic field strength = magnetizing force (do not confuse this with H = Henry) | A $m^{-1}$ | statA $cm*^{-1}$<br>$\rightarrow2.652\times10^{-10}$<br>$=10^{-1}/4\pi c$ | Oe = abA $cm^{-1}$<br>$\rightarrow79.577$<br>$*=10^{3}/4\pi$ | au (Gauss convention)<br>$\rightarrow1.716\times10^{3}=$<br>$=\alpha\hbar e^{-1}a_0^{-2}$<br>au (Lorentz convention)<br>$\rightarrow2.351\times10^{5}$<br>$=\hbar e^{-1}a_0^{-2}$ |

**Table 1.12** (Continued)

| Symbol | Name | SI | esu (→SI) | emu (→SI) | au (→SI) |
|---|---|---|---|---|---|
| $M$ | Volume magnetization (moment/volume) | A $m^{-1}$ | esu $cm^{-3}$<br>$\rightarrow 3.334\times10^{-8}$<br>$= 10\ c^{-1}$ | emu $cm^{-3}$<br>$\rightarrow 10^{3}$<br>$= a\hbar e^{-1}a_0^{-2}$ | au (Gauss convention)<br>$\rightarrow 1.716\times10^{3}$<br>au (Lorentz convention)<br>$\rightarrow 2.351\times10^{5}$<br>$= \hbar e^{-1}a_0^{-2}$ |
| $U$ | Magnetic potential difference magnetomotive force | A = A-turn | esu<br>$\rightarrow 2.654\times10^{-10}$<br>$= 1/4\pi c$ | Gb = Gilbert[a]<br>$\rightarrow 0.7958$<br>$= 10/4\pi$ | au (Gauss convention)<br>$\rightarrow 1.169\times10^{-21}$<br>$= \alpha e$<br>au (Lorentz convention)<br>$\rightarrow 1.602\times10^{-19}$<br>$= e$ |
| $J$, $I$ | Magnetic polarization intensity of magnetization (if $B = \mu_0 H + J$) | T = Wb $m^{-2}$ | esu $cm^{-3}$<br>$\rightarrow 3.767\times10^{7}$<br>$= 4\pi c10^{-2}$ | emu $cm^{-3}$<br>=erg/(Gg)<br>$\rightarrow 1.257\times10^{-3}$<br>$=4\pi10^{-4}$ | au (Gauss convention)<br>$\rightarrow 1.716\times10^{3}$<br>$= a\hbar e^{-1}a_0^{-2}$<br>au (Lorentz convention)<br>$\rightarrow 2.351\times10^{5}$<br>$= \hbar e^{-1}a_0^{-2}$ |
| $\chi$, $\kappa$ | (Volume) magnetic susceptibility $\chi \equiv M/\boldsymbol{H}\rightarrow$<br>iff $\boldsymbol{B} = \mu_0(\boldsymbol{H}+\boldsymbol{M})\rightarrow$ | pure # | pure #<br>➔12.57<br>$= 4\pi$ | pure #<br>➔12.57<br>$= 4\pi$ | au (Gauss convention)<br>$\rightarrow 9.17\times10^{-2}$<br>$= \alpha 4\pi$<br>au (Lorentz convention)<br>➔12.57<br>$= 4\pi$ |
| | $\chi \equiv \boldsymbol{M}/\boldsymbol{H}\rightarrow$<br>iff $B = \mu_0 H + M \rightarrow \rightarrow$ | SI″:H $m^{-1}$ | | $\rightarrow 1.579\times10^{-5}$<br>$= (4\pi)^2 10^{-7}$ | au (Gauss convention)<br>$\rightarrow 1.15\times10^{-7}$<br>$= \alpha(4\pi)^2 10^{-7}$<br>au (Lorentz convention)<br>$\rightarrow 1.579\times10^{-5}$<br>$= (4\pi)^2 10^{-7}$ |
| $\chi_\rho$, $\kappa_\rho$ | Mass magnetic susceptibility<br>$\chi_\rho \equiv \chi$/density<br>iff $\boldsymbol{B} = \mu_0(\boldsymbol{H}+\boldsymbol{M})\rightarrow$ | $m^3\ kg^{-1}$ | | $cm^3\ g^{-1}$<br>$\rightarrow 1.257\times10^{-2}$<br>$= 4\pi10^{-3}$ | au (Gauss convention)<br>➔0.5638<br>$= \alpha 4\pi m_e a_0^{-3}$<br>au (Lorentz convention)<br>➔77.24<br>$= 4\pi m_e a_0^{-3}$ |
| | iff $\boldsymbol{B} = \mu_0\boldsymbol{H} + \boldsymbol{M} \rightarrow \rightarrow$ | SI′ : $H$ $m^2$ $kg^{-1}$ | | $cm^3\ g^{-1}$<br>$\rightarrow 1.579\times10^{-8}$<br>$= (4\pi)^2 10^{-10}$ | au (Gauss convention)<br>$\rightarrow 7.085\times10^{-10}$<br>$= \alpha(4\pi)^2 10^{-7} m_e a_0^{-3}$<br>au (Lorentz convention)<br>$\rightarrow 9.707\times10^{-8}$<br>$= (4\pi)^2 10^{-7} m_e a_0^{-3}$ |
| $\chi_{mol}$, $\kappa_{mol}$ | Molar magnetic susceptibility<br>iff $\boldsymbol{B} = \mu_0(\boldsymbol{H}+\boldsymbol{M})\rightarrow$ | $m^3\ mol^{-1}$ | $cm^3\ mol^{-1}$<br>$\rightarrow 10^{-6}$ | emu $mol^{-1}$<br>$\rightarrow 1.257\times10^{-5}$<br>$= 4\pi10^{-6}$ | |
| | iff $\boldsymbol{B} = \mu_0\boldsymbol{H} + \boldsymbol{M} \rightarrow \rightarrow$ | SI″:H $m^2\ mol^{-1}$ | | emu $mol^{-1}$<br>$\rightarrow 1.579\times10^{-11}$<br>$= (4\pi)^2 10^{-13}$ | |
| $\mu$ | Magnetic permeability<br>$\mu \equiv \boldsymbol{B}/\boldsymbol{H}$ | H(enry)<br>$m^{-1}$ = Wb $A^{-1}$<br>$m^{-1}$ = N $A^{-2}$ | | pure #<br>$\rightarrow 1.257\times10^{-6}$<br>$= 4\pi10^{-7}$ | au (Gauss convention)<br>➔1<br>au (Lorentz convention)<br>➔1 |
| $W$ | Magnetic energy density<br>SI: $W \equiv (1/2\mu_0)\boldsymbol{B}^2$; cgs:<br>$W \equiv (1/8\pi)\boldsymbol{B}^2$ | J $m^{-3}$ | erg $cm^{-3}$<br>➔0.1 | hartree $b^{-3}$<br>$\rightarrow 3.402\times10^{14}$ | |

For brevity $a_0$ = Bohr radius = $5.292\times10^{-11}$ m = $4\pi\varepsilon_0\hbar^2 m^{-1}e^{-2}$. SI = Système International d'Unités = rationalized MKS = MKSC = MKSA = Giorgi system; SI′ = small variation on SI used by some magneticians. "Rationalization" implies putting factors of $4\pi$ into the Maxwell equations and removing them from the constitutive equations. Any quantity given in unrationalized cgs-esu ("esu"), unrationalized cgs-emu ("emu"), or unrationalized au units is converted into SI units by multiplying the quantity by whatever follows the arrow "➔." The au system used here is the "Hartree system" (not the Rydberg system, nor the Planck system, nor the astronomical unit system). For the Hartree au of magnetism, two conventions exist: (i) the "Gauss" or "wave" convention (inspired by cgs) requires that $E$ and $H$ in au have the same magnitude for electromagnetic waves *in vacuo*; (ii) the Lorentz convention, which derives the magnetic field from the Lorentz force equation; the ratio between these units is the Sommerfeld fine-structure constant $\alpha = 1/137.035\,989\,5$.

[a] William Gilbert (1544–1603).

*Source:* Jackson [1.29].

## 1.2 Force One: Gravitation

The gravitational force is the weakest force in nature, but it binds together the most massive bodies in the universe. The force is in N(ewtons) in the SI system, and in dynes in the cgs system (see Table 1.2). This force can be rewritten in terms of a vector **gravitational field** $\boldsymbol{\Phi}_1(\boldsymbol{r}_2)$ experienced by particle 2 at position $\boldsymbol{r}_2$, due to the existence of a particle 1 of mass $m_1$ at $\boldsymbol{r}_1$, and mediated by a continuous flow of virtual gravitons emanating from particle 1:

$$\boldsymbol{\Phi}_1(\boldsymbol{r}_2) = Gm_1\boldsymbol{r}_{12}r_{12}^{-3} \tag{1.2.1}$$

where $\boldsymbol{r}_{12} = \boldsymbol{r}_2 - \boldsymbol{r}_1$. (By definition, a process is virtual when it involves a temporary violation of energy conservation.) This gravitational field can be integrated once to yield the scalar **gravitational potential energy** $U_1$:

$$U_1 = Gm_1r_1^{-1} \tag{1.2.2}$$

where the field is the negative gradient of the potential $U_1$ evaluated at any "field point" $\boldsymbol{r}$, e.g. $\boldsymbol{r} = \boldsymbol{r}_2$ (except at the singular position $\boldsymbol{r} = \boldsymbol{r}_1$):

$$\boldsymbol{\Phi}_1 = -\nabla U_1 = -\boldsymbol{e}_r(\partial/\partial r)U_1 \tag{1.2.3}$$

where $\boldsymbol{e}_r$ is the unit vector in the radial direction. This potential energy is measured in joules, J (1 J ≡ 1 N m)[10] in the SI system, or in erg (1 erg ≡ 1 dyn cm) in the cgs system. We can also define the gravitational potential energy $U_{12}$ as the potential energy of the two-body system:

$$U_{12} = Gm_1M_2r_{12}^{-1} \tag{1.2.4}$$

**What is mass?** Six answers can be given:

1) as a proportionality constant between force and acceleration;
2) "**gravitational mass**" (Eq. (1.1.1));
3) "**inertial mass**";
4) Einstein[11] found an interdependence between **mass *m* and energy *E*** (1905):

$$E = mc^2 \tag{1.2.5}$$

within his special theory of relativity [1.8] (Section 1.11), which explained why the speed of light $c$ through vacuum is constant, validating the 1887 Michelson[12]–Morley[13] experiment [1.9], which found a zero seasonal dependence of $c$ and eliminated the need for a "luminiferous ether" for the propagation of light. Equation (1.2.5) allows for the interconversion of mass and energy: this is at the core of nuclear physics and stellar synthesis.

**(5)** Einstein's general theory of relativity [1.10] allows for a curvature of the space-time continuum, which can lead to a **geometrical understanding of mass** (Wheeler's[14] geometrodynamics [1.11]).
**(6) Quantum mass**: the rest masses of elementary particles cannot yet be predicted, but a Higgs[15] boson has finally been found in 2013: it may usher in a future theory to explain their masses.

Experimental searches are still continuing for free quarks and gluons.

Given the dearth of new results from high-energy scattering studies, high-energy physics has also turned to the cosmos for clues: the current ΛCDM (lambda-cold dark matter) model of "big bang cosmology" estimates that of all matter in the universe: (i) 23% is **dark matter** (Zwicky[16] proposed dark matter in 1933 to account for the gravitational stability of visible galaxies); (ii) 73% is **dark energy**; (iii) 5% is "normal" (visible) matter.

10 James Prescott Joule (1818–1889).
11 Albert Einstein (1879–1955).
12 Albert Abraham Michelson (1852–1931).
13 Edward Williams Morley (1838–1923).
14 John Archibald Wheeler (1911–2008).
15 Peter Higgs (1929–).
16 Fritz Zwicky (1898–1974).

A discussion of the second force (electrical AKA electromagnetic) is given in Section 1.9 after some necessary mathematical concepts and tricks are reviewed.

**Problem 1.2.1** Given the mass of the Earth $M_e = 5.977 \times 10^{24}$ kg and its mean radius $R_e = 6371$ km, verify that the acceleration due to mean gravity at sea level at the Equator is

$$g = GM_eR_e^{-2} = 9.780 \text{ m s}^{-2} \tag{1.2.6}$$

## 1.3 Force Three: Weak Force

The classic example of the weak (or Fermi) force is **β decay**, in which a free neutron ($_0n^1$) decays within a **half-life** $t^{1/2}$ of 10.2 minutes into a proton ($_1p^1$), an electron ($e^-$), and an electron antineutrino ($\overline{\nu}_e$):

$$_0n^1 \rightarrow {}_1p^1 + e + \overline{\nu} \quad t^{1/2} = 10.2 \text{ m} = 611 \text{ s} \tag{1.3.1}$$

As explained in Chapter 6, a sample with $N_0$ free neutrons at $t = 0$ will decay by first-order kinetics and will have only $N = N_0 \exp(-0.693\,15t/t^{1/2})$ neutrons at time $t$. The half-life $t^{1/2}$ in Eq. (1.3.1) is defined as the time needed for **half** of the sample (here, of free neutrons) to have undergone the stated reaction; a second, longer, measure of "endurance" is the **mean lifetime** $\tau$ which is defined by

$$\tau = t^{1/2}/\ln_e 2 = t^{1/2}/0.693\,15 = 1.44t^{1/2} \tag{1.3.2}$$

The weak force is $10^3$ times weaker than the electromagnetic force (Table 1.1). The particles carrying this interaction, the **W and Z vector bosons**, are surprisingly massive (see Tables 1.2 and 1.3).

The **positron**($\beta^+$) is the **antiparticle** of the electron ($e^-$ or $\beta^-$): same mass, same spin, but opposite charge. Antiparticles, first discussed in the 1930s, seemed to be like members of an alternate universe (antiprotons, etc.) which may be symmetrical to the particle universe. However, when a positron (often emitted in nuclear decay) encounters an electron, it annihilates very quickly to produce two photons of energy $h\nu = 0.511$ MeV each (pair elimination):

$$\beta^+ + \beta^- \rightarrow 2h\nu \tag{1.3.3}$$

This reaction can also be run in reverse (pair production) either in air or also when a photon interacts with a nucleus:

$$2\,h\nu \rightarrow \beta^+ + \beta^- \tag{1.3.4}$$

There are also short-lived and negatively charged antiprotons. In the 1930s, it was presumed that there are equal numbers of particles and antiparticles in the universe (which would annihilate each other upon mutual contact as in Eq. (1.3.3)). At present, the guesses are shifting to smaller amounts of antimatter, but nobody really knows.

The Fermi force has a very narrow range ($10^{-18}$ m). Many nuclei can undergo spontaneous β decay. Neutrinos (a name given by Fermi) carry little mass, occur with great abundance in the universe, but are very difficult to detect (they often pass through the mass of the Earth without deflection or detection). Huge tanks of liquid chlorinated hydrocarbons in deep mines have been successfully tested for neutrino-induced formation of a few atoms of radioactive argon.

Many artificial radioisotopes undergo β decay. For instance, in the upper atmosphere, $_6C^{14}$ is generated by cosmic rays (neutron bombardment of $_7N^{14}$) to yield a constant fraction of radioactive carbon in living matter by photosynthetic absorption of $_6C^{14}O_2(1.3 \times 10^{-12})$; it decays by β decay, with a half-life of 5730 years, into stable $_7N^{14}$: this is the basis of Libby's archeological carbon-14 dating method [1.12]:

$$_7N^{14} + {}_0n^1(\text{from cosic ray shower}) \rightarrow {}_6C^{14*}(\beta \text{ decay}) \rightarrow {}_7N^{14+} + \beta^- + \overline{\nu}_e \tag{1.3.5}$$

## 1.4 Force Four: Strong Force

Chemists are often content to ignore quarks and stay with the evidence accumulated by 1930 that electrically neutral atoms of atomic number $Z$ are made of $Z$ protons ($_1p^1$), $N$ neutrons ($_0n^1$), and $Z$ electrons ($e^-$ or $\beta^-$) (Table 1.13) and that these subatomic particles are characterized by relative mass $M$, charge $Q$, spin ($S$ for electron, $I$ for proton and neutron), and half-lives $\tau$. Furthermore $Q$, $S$, and $I$ are also "quantum numbers" (more about this later).

**Table 1.13** Simplified table of elementary particles used in low-level chemistry courses: $m_e = 9.109\,389\,7 \times 10^{-31}$ kg is the electron rest mass; $\varepsilon = -1.602\,176\,57 \times 10^{-19}$ C is the charge on the electron.

| Name | Symbol | Half-life $t^{1/2}$ (s) | Mass $M$ ($m_e$) | Charge $Q$ ($\lvert\varepsilon\rvert$) | Spin $S$ or $I$ |
|---|---|---|---|---|---|
| Electron | $e^-$ or $\beta^-$ | $\infty$ | 1 | −1 | 1/2 |
| Proton | ${}_1p^1$ | $\infty$ | 1838 | +1 | 1/2 |
| Neutron | ${}_0n^1$ | 611 | 1839 | 0 | 1/2 |

Physicists probed much further: between 1900 and 1960, a zoo of 100-odd stable and/or unstable elementary particles was discovered; the shortest-lived among them were called "resonances" (Table 1.2). Quarks (Tables 1.2 and 1.3) were proposed in 1964 by Zweig[17] and Gell-Mann[18] to establish order in this zoo. Within the nucleus, the inter-nucleon "strong" force was traditionally thought of as being mediated by pions (themselves combinations of two quarks). The nuclear "shell model" assigns quantum numbers to the protons and neutrons that have merged to form a certain nucleon. Certain "magic values" of these nuclear quantum numbers explain why certain nuclei are more stable (have longer lives) than others.

Efforts to unify all four forces of Table 1.1 into a **single grand unified theory** have failed. All of eighteenth and nineteenth century mathematical physics was based on continua, on the solution of second-order partial differential equations (PDEs), and on microscopic extensions of macroscopic Newtonian ideas of distance-dependent potentials. As we shall see, classical mechanics, electrodynamics, and quantum mechanics (in its wave mechanical formulation) all have potential energy functions $U(r)$ which are some function of the inter-particle distance $r$. This works well if the particles themselves are much smaller than the distances that typically separate them, and when experiments can test the distance dependence of the potentials directly.

What are we to do when the range of the strong forces are so short (Table 1.1), of the order of magnitude of the presumed particle sizes?

Much progress was made nevertheless. In 1960, electrical and weak forces were merged by Glashow[19] into the **electroweak theory**. Evolving in the 1960s and 1970s from the quark hypothesis, Glashow, Weinberg,[20] and Salam[21] explained nucleons and other particles (hadrons, baryons, and mesons) as unions of either three or two "quarks" each, with a new set of *ad hoc* "quantum numbers" (charge $Q$, spin [$S$ or $I$], parity [$P$], isospin [$T$], $z$-component of isospin [$T_z$], baryon number [Ba], strangeness [St], charm [Ch], beauty or bottomness [By], and truth or topness [Tr]) [1.3] (Tables 1.2 and 1.3) which fit within a finite special unitary group SU(3): all this was called **quantum chromodynamics (QCD)**. An internal relationship between these many seemingly "*ad hoc*" quantum numbers is the amended Gell–Mann–Nishijima[22] relation:

$$Q = T_z + (\mathrm{Ba} + \mathrm{St} + \mathrm{Ch} + \mathrm{By} + \mathrm{Tr})/2 \tag{1.4.1}$$

Adding the electroweak theory to QCD and explaining gravity by positing a Higgs field bathing the whole universe, the overall theory was named the Standard Model (outlined in Special Topic ST1.11): it unites three of the four fundamental forces (strong, weak, and electromagnetic) into a single theory, but has no force field, and addresses gravitation only by positing that moving particles acquire their masses by moving within this Higgs field. The discovery of the mediating Higgs boson in 2012 (with a surprisingly large mass) validated the concept of a universal and pervasive Higgs field: how exactly does it "create" masses? Remember that a universal luminiferous ether was discarded from electromagnetic theory more than a century ago!

**Sideline 1.5** We should always remember Newton's stern warning: "Physica, cave metaphysicam" (=physicists, beware of metaphysics) and Ockham's[23] "razor": "entia non sunt multiplicanda praeter necessitatem" (=fundamental assumed quantities should not be multiplied beyond necessity).

17 George Zweig (1937–).
18 Murray Gell-Mann (1929–2019).
19 Sheldon Glashow (1932–).
20 Steven Weinberg (1933–).
21 Mohammed Abdus Salam (1926–1996).
22 Kazuhiko Nishijima (1926–2009).
23 William of Ockham (c.1285–1348).

Already, the existence of spin angular momentum forces us to consider quantization without potentials (see Chapter 2). Perturbation expansions become funny, since the interaction is no longer smaller than a presumed over-riding field. When the particles are too close, or the potential wells are too deep, then the old tricks do not seem to apply. What to do? Instead of the four dimensions used in special relativity ($x$, $y$, $z$, and $ct$: see Section 1.11 below), **string theory** posits 17 dimensions, but has produced no experimentally verifiable prediction to date: is it useful? Is it necessary?

The seemingly *ex post facto* arguments and numerous new quantum numbers in the Standard Model are reminiscent of the situation in 1869, when chemical trends were used by Mendeleyeff to construct his epochal Periodic Table of chemical elements: the mathematical explanation and vindication for the Periodic Table came with the Schrödinger[24] equation of 1926.

**Sideline 1.6** Besides his chemical discovery, Mendeleyeff had helped his government in the design of weaponry. Mendeleyeff divorced his wife in 1882 and married a student: by the rules of the Russian Orthodox Church, he was a bigamist, and according to an Edict of the Russian Czar, only members of the Church in good standing could teach in Russian Universities. When apprised of the dilemma, Czar Alexander III[25] thought for a moment, then said: "Prof. Mendeleyeff may have two wives, but I only have one Mendeleyeff": Mendeleyeff kept his job!

**Sideline 1.7** When the Nobel Prizes were initiated in 1901, Mendeleyeff was still alive: he was the obvious candidate for the Chemistry prize: it did not happen because a jealous Svante Arrhenius[26] was lusting for his own prize (which he did get in 1916). Instead of Mendeleyeff, some nonentities were awarded the first few Chemistry Prizes. Many other deserving scientists were neglected: the names of Gilbert Newton Lewis[27] and Henry Eyring[28] come to mind. Therefore, this book will not highlight the Nobel prizes, which have honored many worthy recipients, but have bypassed so many others.

Present guesses about the origin of the Universe suggest that matter and energy were all concentrated in a subnanoscopic "singularity" at "$t = 0$," the beginning of the Universe (about 14 billion years ago, see Problem 1.13.1). Within a very short time (Planck time), the Big Bang explosion started the Universe, the four fundamental forces differentiated, and galaxies and the oldest stars and planets were born.

Tables 1.2 and 1.3 list the presently known **fundamental particles** (along with unobserved quarks and three kinds of neutrinos and antineutrinos), the **elementary particles**, and the observed particles (photon, vector bosons, Higgs boson, graviton) and unobserved particles (gluons, some bosons) that mediate the interactions (strong, electromagnetic, weak, and gravitational) between them. Here the term "fundamental" is used for the presumed building blocks, while the term "elementary" is reserved for the experimentally observed smallest constituents of matter. Recent work suggests that the three kinds of neutrinos shown in Table 1.2 have small but differing rest masses (of the order of $10^{-3} m_e$) and that neutrino oscillations (interconversions) between their three types are possible.

The **half-lives** $t^{1/2}$ or **lifetimes** $\tau$ can measured directly when $\tau \geq 1.0 \times 10^{-12}$ s or so. Shorter lifetimes ($t < 1.0 \times 10^{-12}$ s) are inferred from measuring the "natural" or Breit[29]–Wigner[30] or Lorentzian[31] linewidth $\Delta E$ of their energy $E$ (see Special Topic ST1.1) and then using the **Heisenberg**[32] **uncertainty principle** (discussed in Chapter 2).

Searches for individual quarks using high-energy accelerators have failed, up to rest-mass energies in vast excess of the masses of the stable known leptons and hadrons. Searches for quarks in minerals and sea water, potentially left over when hadrons and leptons first formed, and focused on their putative fractional electrical charge, have also failed. Therefore, an explanation of "**quark confinement**" has emerged: quarks are confined by twos and threes in a very deep potential well inside baryons and mesons (Figure 1.1), and are held together by forces so strong, that only maybe future high-energy accelerator experiments may (if ever) detect an individual quark. The sum of the estimated sum of the probable rest masses of one $u$, one $\overline{u}$, and one $d$ quark, namely $4.6 + 4.6 + 16.0 = \sim 25.2 m_e$ is far short of the rest mass of the proton ($1836 m_e$); the remaining $1811 m_e$ are tentatively ascribed to the Einstein mass energy (Eq. (1.2.5)) of the gluon interchanging between quarks.

24 Erwin Rudolf Josef Alexander Schrödinger (1887–1961).
25 Alexander III Alexandrovich Romanov, Czar of All the Russias (1845–1894).
26 Svante August Arrhenius (1859–1927).
27 Gilbert Newton Lewis (1875–1946).
28 Henry Eyring (1901–1981).
29 Gregory Breit (1899–1981).
30 Eugene Paul Wigner = Jenö Pál Wigner (1902–1995).
31 Hendrik Antoon Lorentz (1853–1928).
32 Werner Heisenberg (1901–1976).

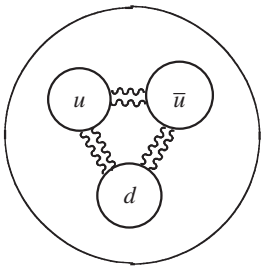



**Figure 1.1** This crude model tries to show the confinement of three quarks ("up" ($u$), "conjugate up" ($\bar{u}$), and "down" ($d$), with electrical charges +2/3, +2/3, and −1/3) inside a proton. The springs depict the mutual interactions, meditated by virtual gluons, which are (somehow) limited by the inter-quark potential to remain within the inside of the proton.

**Nuclei** of atoms can be thought of as super-dense combinations of "nucleons": $Z$ protons ($_1p^1$) and $N$ neutrons ($_0n^1$). Thus, nuclei have a **formal nucleon number** $A$ given by $A \equiv Z + N$ "nucleons," the experimental nuclear mass is $M(Z,A)$. For any chemical element, e.g. "Et," the isotope with $Z$ protons and $A$ protons + neutrons is designated as "$_ZEt^A$" or "Et $A$" or "Et-$A$." Table X1 (please refer www.wiley/go/metzger/physchemtoolbox2) lists most known nuclei.

The nuclear densities ($10^{14}$–$10^{15}$ g cm$^{-3}$ = $10^{17}$–$10^{18}$ kg m$^{-3}$) are so great that the nucleons are in intimate contact with each other, held together by the strong force, is attractive between 2.0 and 0.7 fm, but repulsive below 0.7 fm. It was thought in the 1930s that this internucleon force is mediated by pions; at present, gluons (in a **residual nuclear strong force**) must also be considered. Some distance-dependent potentials (e.g. Yukawa[33]) have been partially successful.

Nuclei are less massive than the sum of the masses of $Z$ free protons and $N$-free neutrons: this mass defect, or nuclear binding energy $\Delta E$, or mass loss, increases with atomic number, up to Fe. $\Delta E$ in atomic mass units (the mass of one $_6C^{12}$ nucleus is 12.000 atomic mass units) is given by

$$\Delta E(Z,A) \equiv -[M(Z,A) - 1.007\,825Z + 1.008\,665(A-Z)] \tag{1.4.2}$$

This binding energy is colossal, when compared to chemical bond energies, is released in stellar nucleosynthesis, and can be partially released in fission and fusion reactions (A-bombs, H-bombs, nuclear power plants: see Section 9.2). The quotient $\Delta E/A$ rises from He to Fe, allowing for exergonic nucleosynthesis in stars. However, $\Delta E/A$ declines beyond Fe (Figure 1.2): this explains why successive gravitational collapses of dying stars form first He stars, then C stars, then Ne stars, then finally Fe stars. When all nuclear fuel is depleted in an Fe star, an ultimate gravitational collapse into black holes must occur, if the star mass exceeds eight solar masses, i.e. $>1.6\times10^{31}$ kg; otherwise, the star will decay into a white dwarf.

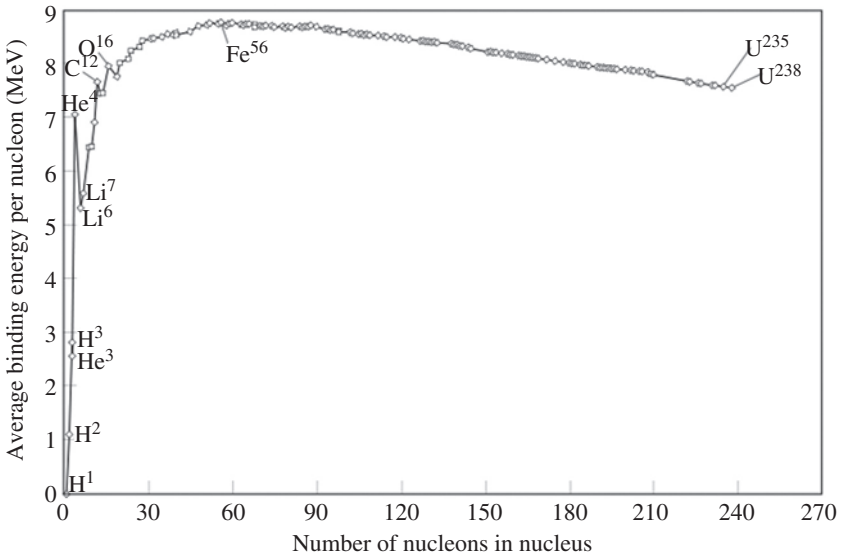



**Figure 1.2** Average binding energy per nucleon $\Delta E/A$ for stable nuclei, as a function of mass number $A$. *Source:* Nuclear Binding energy curve [1.13].

33 Hideki Ogawa Yukawa (1907–1981).

Some nuclei are stable: for low-$Z$ elements, $N \approx Z$ for stable nuclei (e.g. ${}_6C^{12}$, ${}_7N^{14}$, or ${}_8O^{16}$); for high-$Z$ elements, $N \approx 2Z$ is a predictor of stability (infinite half-life). As first suggested by Soddy[34] in 1913, several nuclei can share the same atomic number $Z$: they are **isotopes**. Some isotopes are stable. A large majority of the isotopes are radioactive and unstable (particularly for $Z > 92$) and decay into nuclear "daughters" with a bewildering range of half-lives $t^{1/2}$. Some high-$Z$ radioactive isotopes occur in nature: they are the decay products of other isotopes formed eons ago when the Earth was formed and decay progressively into lighter daughters, ultimately ending the chain of nuclear decays into an isotope of lead (Pb) (Section 9.2). Much experimental efforts on making high-$Z$ elements are driven by curiosity and by so far empty promises of possible "islands of stability" around $Z \approx 120$. Most other radioactive nuclei are formed by irradiating stable nuclei with energetic nuclear particles; natural or artificial radioactive nuclei ($N$, $Z$, $A$) can exhibit several decay modes into new nuclei ($N'$, $Z'$, $A'$) by

i) α-decay ($N' = N-4$, $Z' = Z-2$; $A' = A-4$; with emission of a ${}_2He^4$ nucleus, i.e. helium-4 dictation $({}_2He^4)^{2+}$, which is dominant for elements of atomic number greater than Pb);
ii) β-decay or electron emission ($N' = N-1$; $Z' = Z+1$; $A' = A$; this involves the weak force and the extra emission of an electron antineutrino);
iii) positron or $\beta^+$ decay ($N' = N+1$; $Z' = Z-1$; $A' = A$, emission of a positron and an electron neutrino; this also involves the weak force);
iv) γ decay: no changes in $N$ or $Z$;
v) electron capture ($N' = N+1$; $Z' = Z-1$; $A' = A$, emission of electron; this also involves the weak force);
vi) internal conversion, from a metastable nucleus to a more stable nucleus with no particle emission.

Several models were adopted to explain the structure of stable and radioactive nuclei. **The liquid drop model** assumes that protons and neutrons coalesce to form a liquid drop of high density (spherical, or prolate spheroidal, or oblate spheroidal). Weizsäcker's[35] semiempirical mass formula accounts fairly well for the masses $M(Z,A)$ (in atomic mass units) for stable nuclei:

$$\begin{aligned} M(Z,A) &= 1.007\,825Z + 1.008\,665(A-Z) - 0.016\,91A + 0.019\,11A^{2/3} + 0.000\,763Z^2A^{-1/3} \\ &\quad + 0.101\,75(Z-A/2)^2A^{-1} + (-1,\ 0,\ \text{or}\ +1)0.012A^{-1/2} \end{aligned} \tag{1.4.3}$$

However, assuming that all nuclei are drops of nuclear matter would predict all nuclei to be spherical (many are not) and cannot account for the relative stability of certain nuclei called "islands of (relative) nuclear stability" ($Z$ and/or $N = 2$, 8, 20, 28, 50, 82, 126, 184). A much better **nuclear shell model** was developed by Göppert-Mayer[36] and Jensen[37] (Special Topic ST1.2).

Very useful is a wall chart of all **nuclides** developed by the Knolls Atomic Power Laboratory of the General Electric Company in the 1950s and updated continuously since then (Table X1 [please refer www.wiley/go/metzger/physchemtoolbox2]) [1.14]. Figure 1.3 shows schematically how this chart of the nuclides is arranged.

**Problem 1.4.1** Table X1 (please refer www.wiley/go/metzger/physchemtoolbox2) shows that the mass of the radioactive isotope ${}_5B^{13}$ is 13.017 780 amu. It emits an electron $\beta^-$ and an electron antineutrino to yield the stable isotope ${}_6C^{13}$ of mass 13.003 354 8. Neglecting any energy carried away by the electron antineutrino, compute the energy emission. (**Answer**: 13.436 MeV)

**Problem 1.4.2** The free neutron undergoes β decay to a proton, an electron, and an electron antineutrino, with a half-life of 13.6 minutes. (a) But when you add a proton to the neutron to form a deuteron (D), there is no β decay. Why? (b) In contrast, tritium (T: 1 proton + 2 neutrons) does undergo β decay (half-life 12.32 years) to ${}_1He^3$. Why?

| | | |
|---|---|---|
| ${}_3Li^5$<br>d out | ${}_3Li^6$<br>p out | ${}_3Li^7$<br>$\beta^+$ out, EC |
| ${}_4Be^6$<br>n out | ${}_4Be^7$<br>Original nucleus | ${}_4Be^8$<br>n in |
| ${}_5B^7$<br>$\beta^-$ out | ${}_5B^8$<br>p in | ${}_5B^9$<br>d in |

**Figure 1.3** Diagrammatical representation of how an original nucleus (here ${}_4Be^7$) can be transformed into its neighbors by absorption or emission of nuclear particles (d = deuteron = ${}_1H^2$ = positively charged hydrogen-2). Full details are given in Table X1. EC = Electron capture.

34 Frederick Soddy (1877–1956).
35 Carl Friedrich Freiherr von Weizsäcker (1912–2007).
36 Maria Göppert-Mayer (1906–1972).
37 Johannes Hans Daniel Jensen (1907–1973).

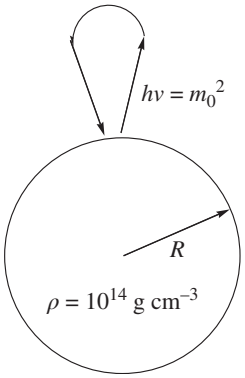



**Figure 1.4** A photon of rest energy $E = h\nu = m_0c^2$ cannot escape the gravitational pull of a cold black hole of radius $R$ and density $10^{14}$ g cm$^{-3}$.

**Problem 1.4.3** **Black hole** (Figure 1.4). A photon of total energy $E = h\nu = mc^2$ becomes unable to escape the gravitational potential of a massive spherical black hole (a term popularized by Wheeler) of density $\rho = 10^{14}$ g cm$^{-3}$ (the density of a light atomic nucleus) and radius $R$.

$$E = mc^2 = GmM/R = Gm\rho(4/3)\pi R^3/R \tag{1.4.4}$$

Find the radius $R$ of the black hole and its mass $M$. Then compare the mass of the black hole $M$ to the mass of our Sun, $1.985 \times 10^{30}$ kg. This resembles the Schwarzschild[38] radius $R$ that occurs in the spherically symmetric solution to Einstein's field equations for general relativity. However, the Schwarzschild radius is twice smaller. Read up on Hawking's[39] explanation of even horizons and how a black hole will keep absorbing everything until it either decreases its density or explodes.

**Problem 1.4.4** Assume that the charge on the proton becomes $+2|e|$ instead of $+|e|$, and that everything else stays the same. What would be the consequences for chemistry?

**Problem 1.4.5** The negative muon, or mu meson, $\mu^-$ is 207 times heavier than the electron and is a natural product of cosmic rays (which first produce pions that decay into muons), or it can be made in particle accelerators. It has a lifetime of 2.2 μs (decaying into an electron and two neutrinos). It can replace the electron in mu-mesic atoms and molecules. Would it produce a drastic modification in atomic and molecular structure?

**Problem 1.4.6** Positrons ($e^+$ or $\beta^+$) are "antielectrons" which can form bound states with electrons ($e^-$ or $\beta^-$, "negatrons" if you like) before they decay. "Positronium" is a hydrogen-like structure, where the positron replaces the positively charged proton, or hydrogen nucleus. Would this be like a hydrogen atom?

Whenever stable nuclei exist, stable chemical elements are possible. The **PERIODIC TABLE OF THE CHEMICAL ELEMENTS (Table 1.14)** was first organized by Mendeleyeff in 1869 [1.15] well before quantum mechanics and the modern theory of atomic structure, by using group analogies in chemical and physical properties.

**Sideline 1.8** In the 1780s, Lavoisier[40] first pinpointed the "irreducibility" of chemical elements (like hydrogen and oxygen) into simpler elements and their combination in chemical compounds (like water). In the early 1800s, Dalton[41] revived the ancient idea of indivisible "fundamental" atoms proposed by Leucippus[42] and his pupil Democritus.[43] Dalton also demonstrated two laws (of definite and multiple proportions): as a result, relative empirical formulas and separate and sometimes contrasting tables of relative atomic weights were established for a growing list of chemical compounds. For several decades, the molecular structure of water was erroneously assumed to be "HO." Avogadro's 1811 principle that at constant pressure and temperature equal volumes of gases contained an equal number of molecules was ignored until Cannizzaro's 1858 work,[44] circulated at the Karlsruhe conference of 1860, convinced the German chemists to finally take Avogadro's principle seriously: Shazaam: the molecular formula for water became $H_2O$, all relative scales rolled into one, and Mendeleyeff could then build his magnificent periodic table!

**Problem 1.4.7** We now have 118 chemical elements, but the elements beyond U (element 92) are produced in nuclear reactors (e.g. Pu) or by collisions of lighter elements with high-energy ions (e.g. Og was made by colliding Cf with calcium ions; only three or four Og atoms were produced!). By its placement, Og is a "noble gas" in Group 18 (albeit a very reactive one because its nucleus soon decays by emitting an α ray). Does this close down the search for stable transuranium elements?

**Problem 1.4.8** Why is iron Fe? Why is mercury Hg? Why is tungsten W?

**Problem 1.4.9** Which two elements were named after living scientists?

38 Karl Schwarzschild (1873–1916).
39 Stephen William Hawking (1942–2018).
40 Antoine-Laurent de Lavoisier (1743–1794).
41 John Dalton (1766–1844).
42 Leucippus of Elea (early 5th c. BC).
43 Democritus of Abdera (c. 460–370 BC).
44 Stanislao Cannizzaro (1826–1910).

**Table 1.14** The periodic table of the known chemical elements.

| Group 1 /1A/ | Group 2 /2A/ | Group 3 /3B/ | Group 4 /4B/ | Group 5 /5B/ | Group 6 /6B/ | Group 7 /7B/ | Group 8 /8B/ | Group 9 /8B/ | Group 10 /8B/ | Group 11 /1B/ | Group 12 /2B/ | Group 13 /3A/ | Group 14 /4A/ | Group 15 /5A/ | Group 16 /6A/ | Group 17 /7A/ | Group 18 /8A/ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Hydrogen** | | | | | | | | | | | | | | | | | **Helium** |
| 1 | | | | | | | | | | | | | | | | | 2 |
| **H** | | | | | | | | | | | | | | | | | **He** |
| 1.0079 | | | | | | | | | | | | | | | | | 4.0026 |
| **Lithium** | **Beryllium** | | | | | | | | | | | **Boron** | **Carbon** | **Nitrogen** | **Oxygen** | **Fluorine** | **Neon** |
| 3 | 4 | | | | | | | | | | | 5 | 6 | 7 | 8 | 9 | 10 |
| **Li** | **Be** | | | | | | | | | | | **B** | **C** | **N** | **O** | **F** | **Ne** |
| 6.941 | 9.0122 | | | | | | | | | | | 10.811 | 12.011 | 14.0067 | 15.9994 | 18.9984 | 20.1797 |
| **Sodium** | **Magnesium** | | | | | | | | | | | **Aluminum** | **Silicon** | **Phosphorus** | **Sulfur** | **Chlorine** | **Argon** |
| 11 | 12 | | | | | | | | | | | 13 | 14 | 15 | 16 | 17 | 18 |
| **Na** | **Mg** | | | | | | | | | | | **Al** | **Si** | **P** | **S** | **Cl** | **Ar** |
| 22.9898 | 24.3050 | | | | | | | | | | | 26.9815 | 28.0855 | 30.9738 | 32.066 | 35.4527 | 39.948 |
| **Potassium** | **Calcium** | **Scandium** | **Titanium** | **Vanadium** | **Chromium** | **Manganese** | **Iron** | **Cobalt** | **Nickel** | **Copper** | **Zinc** | **Gallium** | **Germanium** | **Arsenic** | **Selenium** | **Bromine** | **Krypton** |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
| **K** | **Ca** | **Sc** | **Ti** | **V** | **Cr** | **Mn** | **Fe** | **Co** | **Ni** | **Cu** | **Zn** | **Ga** | **Ge** | **As** | **Se** | **Br** | **Kr** |
| 39.0983 | 40.078 | 44.955912 | 47.867 | 50.9415 | 51.9961 | 54.938045 | 55.845 | 58.933195 | 58.6934 | 63.546 | 65.38 | 69.723 | 72.64 | 74.92160 | 78.96 | 79.904 | 83.798 |
| **Rubidium** | **Strontium** | **Yttrium** | **Zirconium** | **Niobium** | **Molybdenum** | **Technetium** | **Ruthenium** | **Rhodium** | **Palladium** | **Sliver** | **Cadmium** | **Indium** | **Tin** | **Antimony** | **Tellurium** | **Iodine** | **Xenon** |
| 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 |
| **Rb** | **Sr** | **Y** | **Zr** | **Nb** | **Mo** | **Tc** | **Ru** | **Rh** | **Pd** | **Ag** | **Cd** | **In** | **Sn** | **Sb** | **Te** | **I** | **Xe** |
| 85.4678 | 87.62 | 88.90585 | 91.224 | 92.90638 | 95.96 | (98) | 101.07 | 102.90550 | 106.42 | 107.8682 | 112.411 | 114.818 | 118.710 | 121.760 | 127.60 | 126.90447 | 131.293 |
| **Cesium** | **Barium** | **Lanthanum** | **Hafnium** | **Tantalum** | **Tungsten** | **Rhenium** | **Osmium** | **Iridium** | **Platinum** | **Gold** | **Mercury** | **Thallium** | **Lead** | **Bismuth** | **Polonium** | **Astatine** | **Radon** |
| 55 | 56 | 57 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 |
| **Cs** | **Ba** | **La*** | **Hf** | **Ta** | **W** | **Re** | **Os** | **Ir** | **Pt** | **Au** | **Hg** | **Tl** | **Pb** | **Bi** | **Po** | **At** | **Rn** |
| 132.9054511 | 137.327 | 138.90547 | 178.49 | 180.94788 | 183.84 | 186.207 | 190.23 | 192.217 | 195.084 | 196.966569 | 200.59 | 204.3833 | 207.2 | 208.98040 | (209) | (210) | (222) |
| **Francium** | **Radium** | **Actinium** | **Rutherfordium** | **Dubnium** | **Seaborgium** | **Bohrium** | **Hassium** | **Meitnerium** | **Darmstadtium** | **Röntgenium** | **Copernicium** | **Nihonium** | **Flerovium** | **Moscovium** | **Livermorium** | **Tennessine** | **Oganesson** |
| 87 | 88 | 89 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 |
| **Fr** | **Ra** | **Ac‡** | **Rf** | **Db** | **Sg** | **Bh** | **Hs** | **Mt** | **Ds** | **Rg** | **Cn** | **Nh** | **Fl** | **Mc** | **Lv** | **Ts** | **Og** |
| (223) | (226) | (227) | (267) | (268) | (271) | (272) | (270) | (276) | (281) | (280) | (285) | (284) | (289) | (288) | (293) | (294) | (294) |
| | | | **Cerium** | **Praseodymium** | **Neodymium** | **Protoactinium** | **Samarium** | **Europium** | **Gadolinium** | **Terbium** | **Dysprosium** | **Holmium** | **Erbium** | **Thulium** | **Ytterbium** | **Lutetium** | |
| | | | ***58** | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | |
| | | | **Ce** | **Pr** | **Nd** | **Pm** | **Sm** | **Eu** | **Gd** | **Tb** | **Dy** | **Ho** | **Er** | **Tm** | **Yb** | **Lu** | |
| | | | 140.116 | 140.90765 | 144.242 | (145) | 150.36 | 151.964 | 157.25 | 158.92535 | 162.500 | 164.93032 | 167.259 | 168.93421 | 173.054 | 174.9668 | |
| | | | **Thorium** | **Protactinium** | **Uranium** | **Neptunium** | **Plutonium** | **Americium** | **Curium** | **Berkelium** | **Californium** | **Einsteinium** | **Fermium** | **Mendelevium** | **Nobelium** | **Lawrencium** | |
| | | | **‡90** | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | |
| | | | **Th** | **Pa** | **U** | **Np** | **Pu** | **Am** | **Cm** | **Bk** | **Cf** | **Es** | **Fm** | **Md** | **No** | **Lr** | |
| | | | 232.03806 | 231.03588 | 238.02891 | (237) | (244) | (243) | (247) | (247) | (251) | (252) | (257) | (258) | (259) | (262) | |

The 18 groups (Group *n*) are the modern ones; the older grouping is given inside slashes, e.g. /7B/.

**Problem 1.4.10** Which 11 elements are named after [continents], {countries}, or (regions)?

**Problem 1.4.11** Which nine elements are named after cities?

**Problem 1.4.12** Which 16 elements are named after scientists?

**Sideline 1.9** The discovery of two new chemical elements, "alabamine" (element 8) and "virginium" (element 89) was published in 1930 in J. Am. Chem. Soc. by Prof. Fred Ellison (1882–1974), a physicist at Auburn Polytechnic Institute in Auburn, Alabama (now Auburn University), who claimed to have used "magneto-optical spectroscopy" (exploiting the time dependence of the Faraday effect?). His claims remained unverified. In 1953, Irving Langmuir labeled this work as a sad example of "pathological science" (pseudo-science based on incomplete observations, fundamental errors, and/or fanciful hypotheses). Other notorious examples of pathological science have been: Giovanni Schiaparelli's "Martian canals" (1877), Boris Deryagin's "polywater" (1961), and Stanley Pons and Martin Fleishman's "cold fusion" (1989).

**PART TWO:** Mathematical Interlude

## 1.5 Review of Mathematical Concepts

### 1.5.1 Mathematics and Statistics

We present here some necessary concepts. A recent massive review of such methods is given by Wolfram[45] [1.16], and his computing system, called Mathematica [1.17], an encyclopedic compendium of applied mathematics (and its for-profit website to ease the burden on the indolent student). This section reviews at warp speed, some mathematical concepts that should have been learned in an undergraduate science curriculum: functions, equations, and functionals; quadratic, cubic, and quartic equations; plane trigonometric functions; hyperbolic functions; differential and integral calculus; Taylor[46] and Maclaurin[47] series; differential forms; infinite series; partial fractions; ordinary differential equations (ODEs); partial derivatives; Lagrange's[48] method of undetermined multipliers; and PDEs.

### 1.5.2 Functions, Equations, and Functionals

When a **function** $y = f(x)$ is specified, e.g. $y = x^4 + 3\sin x + \tanh x$, then $x$ is the independent variable, and $y$ is the dependent variable, whose value is computed once numbers are assigned to the (one or more) independent variables for $f$. In other words, a function is a recipe for going from a variable ($x$) to a number ($f(x)$).

An **equation** is when the function is restricted by a definite value, it must obtain after evaluation, e.g. $x^3 + 3\sin x + \tanh x = 55$ means that we must "solve" the equation for $x$, (i.e. compute $x$) such that that value of $x$ will satisfy the given equation.

A **functional** $F[g] = 33$ means that the explicit functional form of $g$ is not known, or not knowable, but its use must yield the definite value of 33. In other words, a functional is a recipe for going from a function ($g$) to a number ($F[g]$).

The **factorization of a polynomia**l with integer coefficients is discussed in Special Topic ST1.3**.**

### 1.5.3 Quadratic, Cubic, and Quartic Equations

Algebraic equations with one variable $\lambda$ of order $n = 1$–4 can be solved explicitly.

**If $n = 2$**, the **quadratic equation**:

$$\lambda^2 + a\lambda + b = 0 \tag{1.5.1}$$

45 Stephen Wolfram (1959–).
46 Sir Brook Taylor (1685–1731).
47 Colin Maclaurin (1698–1746).
48 Joseph Louis Lagrange = Giuseppe Lodovico Lagrangia (1736–1813).

has two solutions:

$$\lambda_1 = -a/2 + (1/2)\left[a^2 - 4b\right]^{1/2}$$
$$\lambda_2 = -a/2 - (1/2)\left[a^2 - 4b\right]^{1/2} \tag{1.5.2}$$

If the discriminant $D \equiv a^2 - 4b < 0$, then both roots are complex; if $D = 0$, then the roots are real and degenerate (equal to each other). The solution for $n = 2$ was known to the Egyptians in the Middle Kingdom (c. 2160–1700 BC), to the Hindus (Brahmagupta[49] in 628 AD), and, in its geometrical form, to the ancient Greeks Euclid[50] and Diophantus.[51] The solution is generated by adding and subtracting a term, so that a perfect square can be isolated, and its square root taken.

A similar approach works for the cubic equation: If $\boldsymbol{n = 3}$, for the **cubic equation**:

$$\lambda^3 + a\lambda^2 + b\lambda + c = 0 \tag{1.5.3}$$

the solution was found by del Ferro[52] and Tartaglia,[53] published by Cardano[54] in 1545, and confirmed by Ferrari[55] and also by Viète.[56] Several forms of the solution are given in Special Topic ST1.4. A fool-proof solution is

$$\begin{aligned}
\lambda_1 &= -(1/3)a + 3^{-1}2^{-1/3}\left\{2a^3 - 9ab + 27c + \left[\left(2a^3 - 9ab + 27c\right)^2 - 4\left(a^2 - 3b\right)^3\right]^{1/2}\right\}^{1/3} \\
&\quad + 3^{-1}2^{-1/3}\left\{2a^3 - 9ab + 27c - \left[\left(2a^3 - 9ab + 27c\right)^2 - 4\left(a^2 - 3b\right)^3\right]^{1/2}\right\}^{1/3} \\
\lambda_2 &= -(1/3)a - \left(1 + \mathrm{i}3^{1/2}\right)3^{-1}2^{-4/3}\left\{2a^3 - 9ab + 27c + \left[\left(2a^3 - 9ab + 27c\right)^2 - 4\left(a^2 - 3b\right)^3\right]^{1/2}\right\}^{1/3} \\
&\quad - \left(1 - \mathrm{i}3^{1/2}\right)3^{-1}2^{-4/3}\left\{2a^3 - 9ab + 27c - \left[\left(2a^3 - 9ab + 27c\right)^2 - 4\left(a^2 - 3b\right)^3\right]^{1/2}\right\}^{1/3} \\
\lambda_3 &= -(1/3)a - \left(1 - \mathrm{i}3^{1/2}\right)3^{-1}2^{-4/3}\left\{2a^3 - 9ab + 27c + \left[\left(2a^3 - 9ab + 27c\right)^2 - 4\left(a^2 - 3b\right)^3\right]^{1/2}\right\}^{1/3} \\
&\quad - \left(1 + \mathrm{i}3^{1/2}\right)3^{-1}2^{-4/3}\left\{2a^3 - 9ab + 27c - \left[\left(2a^3 - 9ab + 27c\right)^2 - 4\left(a^2 - 3b\right)^3\right]^{1/2}\right\}^{1/3}
\end{aligned} \tag{1.5.4}$$

The **quartic equation,** solved by Ferrari, is discussed in Special Topic ST1.5.

For polynomial equations in a single variable of degree $\boldsymbol{n}$ greater than 4 ($\boldsymbol{n > 4}$), Abel[57] showed in 1824 that there can be no general closed-form solution [1.18]. When $n > 4$, numerical methods must be used.

**Planar trigonometric functions**: In a right plane triangle with right angle $\gamma = 90^\circ$, $\sin\ \alpha \equiv A/C$ where $A$ is the side opposite to the angle $\alpha$, and $C$ is the hypotenuse; $\cos\ \alpha \equiv B/C$; $\tan\ \alpha \equiv A/B = \sin\ \alpha/\cos\ \alpha$; $\sin\ \beta \equiv B/C$; $\cos\ \beta \equiv A/C$; $\alpha + \beta = 90^\circ$; $\sec\ \alpha \equiv 1/\cos\ \alpha$; $\mathrm{cosec}\ \alpha \equiv 1/\sin\ \alpha$; $\mathrm{cotan}\ \alpha \equiv 1/\tan\ \alpha$; $\sin^2\ \alpha + \cos^2\ \alpha = 1$; $\sin(-x) = -\sin\ x$; $\cos(-x) = \cos\ x$; $\tan(-x) = -\tan(x)$; thus, $\sin\ x$ and $\tan\ x$ are **odd functions** of $x$, while $\cos\ x$ is an even function of $x$; look out for the **principal value branches** of these periodic functions: ($x = 0$–$\pi$ rad for $\cos\ x$, but $-\pi/2$ to $+\pi/2$ radians for $\sin\ x$ and $\tan\ x$); addition formulas:

$$\sin(x \pm y) = \sin\ x \cos\ y \pm \cos\ x \sin\ y \tag{1.5.5}$$
$$\cos(x \pm y) = \cos\ x \cos\ y \mp \sin\ x \sin\ y \tag{1.5.6}$$
$$2\sin\ x \cos\ y = \cos(x + y) + \sin(x - y) \tag{1.5.7}$$

49 Brahmagupta (598–668).
50 Euclid of Alexandria (fl[oruit] c. 300 BC).
51 Diophantus (c. 200–284).
52 Scipione del Ferro (1495–1528).
53 Niccolò Fontana, nicknamed "Il Tartaglia" = "The Stutterer" (1499–1557).
54 Girolamo Cardano (1501–1576).
55 Lodovico Ferrari (1522–1565).
56 François Viète (1540–1603).
57 Niels Henrik Abel (1802–1829).

$$2\cos\ x\cos\ y = \cos(x+y) + \cos(x-y) \tag{1.5.8}$$

$$2\sin\ x\sin\ y = \cos(x-y) - \cos(x+y) \tag{1.5.9}$$

$$\cos^2 x = (1/2) + (1/2)\cos\ 2x \tag{1.5.10}$$

In some countries, tan $x$ is written tg $x$, and cotan $x$ is written cotg $x$.
**Inverse planar trigonometric functions**: if $x = \cos\ y$, then $y = \cos^{-1}\ x$. **Careful**: $\cos^{-1}\ x$ is **not** the same as $1/\cos x$!!!
**Hyperbolic functions**:

$$\exp(x) \equiv e^x \text{ where } e \equiv 2.718\,281\,828 \text{ is an } \textbf{irrational} \text{ number } (\pi \text{ is also irrational}) \tag{1.5.11}$$

$$\sinh\ x \equiv (1/2)[\exp(x) - \exp(-x)] \tag{1.5.12}$$

$$\cosh\ x \equiv (1/2)[\exp(x) + \exp(-x)] \tag{1.5.13}$$

$$\tanh\ x \equiv \sinh\ x/\cosh\ x = [\exp(x) - \exp(-x)]/[\exp(x) + \exp(-x)] \tag{1.5.14}$$

$$\text{cotanh}\ x \equiv \cosh\ x/\sinh\ x = [\exp(x) + \exp(-x)]/[\exp(x) - \exp(-x)] \tag{1.5.15}$$

In the last three centuries, many other special functions have been invented, used, tabulated, and computer-coded for navigation, ballistics, physics, chemistry, and engineering: here a listing of their names must suffice: Bessel,[58] confluent hypergeometric, Coulomb, elliptic, error (see below), Euler,[59] exponential integral, gamma (see below), hypergeometric, Jacobian,[60] Legendre,[61] Mathieu,[62] Neumann,[63] parabolic cylinder, Riemann,[64] spheroidal, Struve,[65] theta, Weierstrass,[66] and zeta functions, plus the orthogonal polynomials (Chebyshev,[67] Hermite,[68] Jacobi, Laguerre,[69] Legendre) [1.19–1.22].
**Differential calculus** was developed independently by Newton and Leibniz.[70]
**Derivatives**: The general definition of a derivative of a function $y(x)$ is

$$dy/dx = \lim_{\Delta x \to 0}[\Delta y/\Delta x] = \lim_{\Delta x \to 0}[[y(x+\Delta x) - y(x)]/\Delta x] \tag{1.5.16}$$

Here are some derivatives:

$$(d/dx)x^n = nx^{n-1} \tag{1.5.17}$$

$$(d/dx)x^{-n} = -nx^{-n-1} \tag{1.5.18}$$

$$(d/dx)\sin\ x = \cos\ x \tag{1.5.19}$$

$$(d/dx)\cos\ x = -\sin\ x \tag{1.5.20}$$

$$(d/dx)\tan\ x = \sec^2 x = 1/\cos^2 x \tag{1.5.21}$$

$$(d/dx)\text{cotan}\ x = -\text{cosec}^2\ x \tag{1.5.22}$$

$$(d/dx)\sec\ x = \sec\ x\tan\ x \tag{1.5.23}$$

$$(d/dx)\text{cosec}\ x = -\text{cosec}\ x\,\text{cotan}\ x \tag{1.5.24}$$

$$(d/dx)e^x \equiv (d/dx)\exp(x) \equiv \exp(x) = e^x \tag{1.5.25}$$

58 Friedrich Wilhelm Bessel (1784–1846).
59 Leonhard Euler (1707–1783).
60 Carl Gustav Jacob Jacobi (1804–1851).
61 Adrien-Marie Legendre (1752–1833).
62 Émile Léonard Mathieu (1835–1890).
63 Carl Gottfried Neumann (1832–1925).
64 Georg Friedrich Bernhard Riemann (1826–1886).
65 Karl Hermann Struve (1854–1920).
66 Karl Theodor Wilhelm Weierstrass (1815–1897).
67 Pafnuty Lvovich Chebyshev (1821–1894).
68 Charles Hermite (1822–1901).
69 Edmond Nicolas Laguerre (1834–1886).
70 Gottfried Wilhelm von Leibniz (1646–1716).

$$(\mathrm{d}/\mathrm{d}x)\cosh\ x = \sinh\ x \tag{1.5.26}$$

$$(\mathrm{d}/\mathrm{d}x)\sin^{-1}\ x = \left[1-x^2\right]^{-1/2} \tag{1.5.27}$$

$$(\mathrm{d}/\mathrm{d}x)\cos^{-1}\ x = -\left[1-x^2\right]^{-1/2} \tag{1.5.28}$$

$$(\mathrm{d}/\mathrm{d}x)\tan^{-1}\ x = \left[1+x^2\right]^{-1} \tag{1.5.29}$$

$$(\mathrm{d}/\mathrm{d}x)\mathrm{cotan}^{-1}\ x = -\left[1+x^2\right]^{-1} \tag{1.5.30}$$

**Maxima, minima, and saddle points** of functions are found by differentiating them and setting the derivative to zero. For a function $f(x)$, if for some $x = a[\mathrm{d}f(x)/\mathrm{d}x]_{x=a} = 0$ and also $[\mathrm{d}^2f(x)/\mathrm{d}x^2]_{x=a} > 0$, then $f(x)$ has a minimum at $x = a$. If instead $[\mathrm{d}f(x)/\mathrm{d}x]_{x=a} = 0$ and also $[\mathrm{d}^2f(x)/\mathrm{d}x^2]_{x=a} < 0$, then $f(x)$ has a maximum at $x = a$. If $[\mathrm{d}f(x)/\mathrm{d}x]_{x=a} = 0$ and $[\mathrm{d}^2f(x)/\mathrm{d}x^2]_{x=a} = 0$, then we have neither max nor a min. If for some $x = a[\mathrm{d}f(x)/\mathrm{d}x]_{x=a} = 0$, and $[\mathrm{d}^2f(x)/\mathrm{d}x^2]_{x=a} = 0$, but $[\mathrm{d}^3f(x)/\mathrm{d}x^3]_{x=a} \neq 0$, then we have a "saddle point."

**Differential operators:** $(\mathrm{d}/\mathrm{d}x)f(x) \equiv \mathrm{d}f/\mathrm{d}x \equiv f'(x)$; $(\mathrm{d}/\mathrm{d}x)(\mathrm{d}/\mathrm{d}x)f(x) = (\mathrm{d}^2f/\mathrm{d}x^2) \equiv f''(x)$; $(\mathrm{d}/\mathrm{d}x)^nf(x) \equiv (\mathrm{d}^nf/\mathrm{d}x^n) = f^{(n)}(x)$.

**Chain rule:**

$$\mathrm{d}(uv)/\mathrm{d}x = u(\mathrm{d}v/\mathrm{d}x) + v(\mathrm{d}u/\mathrm{d}x) \tag{1.5.31}$$

**Problem 1.5.1** Prove Eq. (1.5.19) by using the chain rule.

**Problem 1.5.2** Evaluate $(\mathrm{d}^2/\mathrm{d}x^2)\cosh([1+x^2]^{1/2})$.

**Problem 1.5.3** Evaluate $(\mathrm{d}^2/\mathrm{d}z^2)\exp[-(1/2)z^2]$.

**Problem 1.5.4** Evaluate $(\mathrm{d}^3/\mathrm{d}y\,\mathrm{d}x^2)y\ \ \exp[-(1/2)yz^2]$ (these smacks of partial derivatives!).

**Problem 1.5.5** Evaluate $(\mathrm{d}^2/\mathrm{d}x^2)\sin^{-1}\ x = [1-x^2]^{-1/2}$.

### 1.5.4 Partial Derivatives

If a function like $f(x,y,z)$ depends on two or more mutually independent variables (here $x$, $y$, and $z$), then the derivatives of that function can be evaluated with respect to each variable in turn. Thus, one writes

$$(\partial f/\partial x)_{y,z} \tag{1.5.32}$$

to mean the explicit derivative of $f$ with respect to the $x$ variable only, holding the other two variables ($y$ and $z$) fixed. These two suffixes are often omitted when their identity is clear in context, but they are rarely omitted in thermodynamics (Chapter 3) because of frequent and easily confusing switches in independent thermodynamic variables. One can also make multiple differentiations, e.g.

$$\left(\partial^2f/\partial x^2\right)_{y,z} \equiv \left(\partial\left[(\partial f/\partial x)_{y,z}\right]/\partial x\right)_{y,z} \tag{1.5.33}$$

and also

$$\left(\partial^2f/\partial x\,\partial y\right) \equiv \left(\partial\left[(\partial f/\partial x)_{y,z}\right]/\partial y\right)_{z,x} \tag{1.5.34}$$

For nonpathological functions the order of differentiation is immaterial, so that:

$$\left(\partial^2f/\partial x\,\partial y\right) \equiv \left(\partial^2f/\partial y\,\partial x\right) \tag{1.5.35}$$

(the Euler condition) but this is **not** always true. A useful way to visualize first partial derivatives is to consider a snow-covered mountain whose height function $z = f(x, y)$ depends on the orthogonal cardinal directions $x$ and $y$ (say $x$ = East and $y$ = South). Deciding **which slope to ski down on** depends on the values of $(\partial z/\partial x)$ and $(\partial z/\partial y)$ in various compass headings, or combinations thereof, i.e. how $z$ depends on both $x$ and $y$.

**Problem 1.5.6** For the equation

$$f(x,y,z) = x^2 + 2yz - 2y^2 - z^2 = 11$$

find $(\partial f/\partial z)_x$ the partial derivative of $f$ with respect to $z$, with $x$ held constant.

**Problem 1.5.7** It is important to distinguish between an ordinary derivative and a partial derivative. Evaluate $(\mathrm{d}/\mathrm{d}x)[2x + 3xy + y^3]$, where $y$ is also an (implicit) function of $x$, and also $(\partial/\partial x)[2x + 3xy + y^3]$.

**Problem 1.5.8** If $w = uv$, where $u$ and $v$ are independent variables, evaluate $\mathrm{d}w$.

**Problem 1.5.9** If $u = \cos\ x + \sin\ y$ but also $x \equiv [s^2 + t^2]^{1/2}$ and $y \equiv \exp(st)$, then evaluate $\mathrm{d}u$ in terms of $\mathrm{d}s$ and $\mathrm{d}t$.

**Problem 1.5.10** If $z^3 + 4xz^2 = 5xy + 8$, then calculate $\mathrm{d}z$ in terms of $x$, $y$, $\mathrm{d}y$, and $\mathrm{d}y$.

**Indefinite integrals**:

$$\int x^n\,\mathrm{d}x = (n+1)^{-1}x^{n+1} + C \tag{1.5.36}$$

where $C$ is a constant of integration (whose numerical value is set by the "initial conditions"); here integration is seen as "antidifferentiation" (for definite limits $a$ and $b$ this is the fundamental theorem of calculus):

$$\int_{x=a}^{x=b} (\mathrm{d}F/\mathrm{d}x)\mathrm{d}x = F(b) - F(a) \tag{1.5.37}$$

Here are a few known indefinite integrals:

$$\int \sin\ x\ \mathrm{d}x = -\cos\ x + C \tag{1.5.38}$$

$$\int \cos\ x\ \mathrm{d}x = \sin\ x + C \tag{1.5.39}$$

$$\int \tan\ x\ \mathrm{d}x = -\log|\cos\ x| + C \tag{1.5.40}$$

$$\int \mathrm{cotan}\ x\ \mathrm{d}x = -\log|\sin\ x| + C \tag{1.5.41}$$

In some cases, a trigonometric transformation helps:

$$\int \cos\ x^2\,\mathrm{d}x = \{\text{using Eq.}(1.5.9)\} = (1/2)\int(1 + \cos\ 2x)\mathrm{d}x = (x/2) + (1/4)\sin\ 2x + C \tag{1.5.42}$$

**Integration by parts**:

$$\int u\ \mathrm{d}v = uv - \int v\ \mathrm{d}u \tag{1.5.43}$$

Here is an example:

$$\begin{aligned}\int x\cos\ x\ \mathrm{d}x &= \{\text{define } \mathrm{d}v \equiv \cos\ x\ \mathrm{d}x \text{ and } u \equiv x,\ \text{whence } v = \sin\ x \text{ and } \mathrm{d}u = \mathrm{d}x\}\\ &= x\sin\ x - \int \sin\ x\ \mathrm{d}x + C'\\ &= x\sin\ x + \cos\ x + C\end{aligned} \tag{1.5.44}$$

**Definite integral in one dimension**: Riemann defined a definite integral of a function $y = y(x)$ between limits $x = x_1$ and $x = x_w$ as the **area** under the function between those limits:

$$\int_{x=x_1}^{x=x_2} y(x)\,\mathrm{d}x = \sum_i \lim_{\Delta x_i \to 0} y(x + \Delta x_i)\Delta x_i \tag{1.5.45}$$

The integration limits may be numbers or functions. An example is

$$\int_{x=3}^{x=4} x^n \, dx = \left[(n+1)^{-1} x^{n+1}\right]_{x=3}^{x=4} = (n+1)^{-1}\left[4^{n+1} - 3^{n+1}\right] \tag{1.5.46}$$

**Definite integral in two dimensions**:

$$\begin{aligned}\int_{y=2}^{y=5}\int_{x=3}^{x=4} x^m y^n \, dx \, dy &= \left[(n+1)^{-1} y^{n+1}\right]_{y=2}^{y=5}\left[(m+1)^{-1} x^{m+1}\right]_{x=3}^{x=4} \\ &= (n+1)^{-1}(m+1)^{-1}\left[5^{n+1} - 2^{n+1}\right]\left[4^{m+1} - 3^{m+1}\right]\end{aligned} \tag{1.5.47}$$

which is simple. When the integration limits on $x$ involve $y$ explicitly, then the integration over $y$ must be done after the integration over $x$ has been completed:

$$\begin{aligned}\int_{y=2}^{y=5}\int_{x=0}^{x=y} x^m y^n \, dx \, dy &= \int_{y=2}^{y=5} y^n \, dy \int_{x=0}^{x=y} x^m \, dx = \int_{y=2}^{y=5} y^n \, dy\left[(m+1)^{-1} y^{m+1}\right] \\ &= (m+1)^{-1}(n+m+2)^{-1}\left[y^{m+m+2}\right]_{y=2}^{y=5} \\ &= (n+1)^{-1}(n+m+2)^{-1}\left(5^{n+m+2} - 2^{n+m+2}\right)\end{aligned} \tag{1.5.48}$$

**Definite integral in three dimensions**: Here is an example:

$$\begin{aligned}&\int_{z=2}^{z=5} z \, dz \int_{y=1}^{y=3}\int_{x=0}^{x=y} xy^3 \cos(xy) \, dx \, dy \\ &= \{\text{change variables : } t = xy, dt = dx \text{ and use Eq.(1.5.43) for "} u \text{ dv"} = x \cos(xy) \, dx\} \\ &= \int_{z=2}^{z=5} z \, dz \int_{y=1}^{y=3} y^3\left[(x/y)\sin(xy) + \left(1/y^2\right)\cos(xy)\right]_{x=0}^{x=y} dy \\ &= \int_{z=2}^{z=5} z \, dz \int_{y=1}^{y=3}\left[xy^2 \sin(xy) + y\cos(xy)\right]_{x=0}^{x=y} dy \\ &= \int_{z=2}^{z=5} z \, dz \int_{y=1}^{y=3}\left[y^3 \sin\left(y^2\right) + y\cos\left(y^2\right) - y\right] dy \\ &= \left\{\text{replace the argument } y^2 \text{ by } t \equiv y^2, \text{ whence } dt = 2y \, dy \text{ and } dy = dt/(2y) = t^{-1/2} dt/2 \text{ and finally } y = t^{1/2}\right\} \\ &= \int_{z=2}^{z=5} z \, dz \int_{t=1}^{t=\sqrt{3}}\left[t^{3/2}\sin t + t^{1/2}\cos t - t^{1/2}\right](1/2)t^{-1/2} dt \\ &= (1/2)\int_{z=2}^{z=5} z \, dz \int_{t=1}^{t=\sqrt{3}} [t\sin t + \cos t - 1] dt \\ &= \left\{\text{use indefinite integral } \int x \sin x \, dx = \sin x - x\cos x + C\right\} \\ &= (1/2)\int_{z=2}^{z=5} z \, dz[\sin t - t\cos t + \sin t - t]_{1=1}^{t=\sqrt{3}} \\ &= (1/2)\int_{z=2}^{z=5} z \, dz[2\sin t - t\cos t - t]_{t=1}^{t=\sqrt{3}} \\ &= \left[z^2\right]_{z=2}^{z=5}\left[2\sin 3^{1/2} - 3^{1/2}\cos 3^{1/2} - 3^{1/2} - 2\sin 1 - \cos 1 - 1\right] \\ &= 42\sin 3^{1/2} - 21 \times 3^{1/2}\cos 3^{1/2} - 21 \times 3^{1/2} - 42\sin 1 - 21\cos 1 - 21 \\ &= 20.727 + 5.840 - 36.373 - 35.342 - 11.346 - 21 = -77.494\end{aligned} \tag{1.5.49}$$

### 1.5.5 Jacobians

Quite often it may be computationally convenient to change independent variables (e.g. $x, y, z$) to a new set (e.g. $u, v, w$). For example, whereas in the orthogonal Cartesian coordinate system $(x, y, z)$, the volume element is $dV = dx\ dy\ dz$, in the orthogonal spherical polar coordinate system $(r, \theta, \varphi)$ it is $dV = r^2\, dr\ \sin\ \theta\ d\theta\ d\varphi$ (as mentioned in Problem 1.6.28 below), the factor $r^2\ \sin\ \theta$ is the **determinant of the Jacobian** of the transformation $(x, y, z) \rightarrow (r, \theta, \varphi)$. The Cartesian system is named after Descartes.[71]

**Sideline 1.10** Descartes, a late riser, died in Sweden, maybe of pneumonia, because Queen Christina,[72] who had hired him, insisted that he teach her philosophy at dawn. The French broadsheet *La Gazette d'Anvers*, announced Descartes' death by "En Suède un sot vient de mourir qui disait qu'il pouvait vivre aussi longtemps qu'il voulait."

### 1.5.6 Line integral

In certain problems, the integration of a function $f(x, y, z)$ occurs for values of that function restricted on some defined line, or curve (this often used in thermodynamics (Chapter 3) to calculate the work done over a certain defined path). This is indicated by a suffix $C$ under the integral sign and explained by defining the curve. Here is an example: integrate the function $f(x, y, z) = xy + y^2 - xyz$ over a plane parabolic curve defined by $y = x^2 + 3x + 2$ and by $z = 0$, between the points $(x = 2, y = 12, z = 0)$ and $(x = 3, y = 17, z = 0)$:

$$\begin{aligned}
\int_{C\ x=2}^{x=3} (xy + y^2 - 0)dx &= \int_{C\ x=2}^{x=3} \left[x\left(y = x^2 + 3x + 2\right) + \left(y = x^2 + 3x + 2\right)^2\right]dx \\
&= \int_{C\ x=2}^{x=3} \left[x^3 + 3x^2 + 2x + x^4 + 9x^2 + 4 + 6x^3 + 12x + 4x^2\right]dx \\
&= \left[x^4 + 7x^3 + 16x^2 + 14x + 4\right]dx \\
&= \left[(1/5)x^5 + (7/4)x^4 + (16/3)x^3 + 7x^2 + 4x\right]_{x=2}^{x=3} \\
&= [(1/5)(243 - 32) + (7/4)(81 - 16) + (16/3)(27 - 8) + 7(9 - 4) + 4(3 - 2)] \\
&= 42.2 + 113.75 + 101.333 + 35 + 4 \\
&= 296.283
\end{aligned} \tag{1.5.50}$$

### 1.5.7 The Gauss[73]–Green[74]–Stokes[75] Theorem

In analogy to the one-dimensional case of Eq. (1.5.32), for two or three dimensions, we seek two theorems: The first theorem is

$$\int\int_S (\text{integrand})dS = \oint_L (\text{another integrand})dl \tag{1.5.51}$$

where the right-hand side is a line integral; this is Green's theorem if $S$ is a flat surface, but is called Stokes' theorem if $S$ is curved. The second theorem is

$$\int\int\int_V (\text{integrand})dV = \int\int_S (\text{another integrand})dS \tag{1.5.52}$$

71 René Descartes (1596–1650).
72 Queen Christina Alexandra of Sweden (née Kristina Augusta Wasa) (1626–1689).
73 Karl Friedrich Gauss (1777–1855).
74 George Green (1793–1841).
75 Sir George Gabriel Stokes (1819–1903).

which is Gauss' theorem. Together, Eqs. (1.5.50) and (1.5.51) are known as the "Gauss–Green–Stokes theorem." In particular, if the three functions $P(x, y, z)$, $Q(x, y, z)$, and $Q(x, y, z)$ and their three partial derivatives $\partial P(x, y, z)/\partial x$, $\partial Q(x, y, z)/\partial y$, and $\partial R(x, y, z)/\partial z$ are continuous then Gauss's theorem takes the explicit form:

$$\iiint_V (\partial P/\partial x + \partial Q/\partial y + \partial R/\partial z)\mathrm{d}V = \iint_S [P(x,y,z) + Q(x,y,z) + R(x,y,z)]\mathrm{d}S \tag{1.5.53}$$

there are strict requirements on the shape of the surface $S$. Equation (1.5.52) will be rewritten in vector notation in Eq. (1.6.82).

Next, if a planar region $S$ bounded by a closed contour $L$, and if $P(x, y)$ and $Q(x, y)$ and all their partial derivatives are continuous on $L$, then Green's theorem is explicitly:

$$\iint_S (\partial P/\partial x - \partial Q/\partial y)\mathrm{d}S = \oint_L (P\ \mathrm{d}x + Q\ \mathrm{d}y) \tag{1.5.54}$$

Please note the minus sign: if the $xyz$ coordinate system is right-handed, then the contour $L$ must lie to the right of an observer riding on $L$.

### 1.5.8 Techniques of integration

While differentiation is routinely possible, not all integrals yield "analytical" solutions "in closed form," i.e. solutions that use known mathematical functions. There are extensive tabulations of known definite and indefinite integrals [1.22–1.24]. Where integration in closed form is impossible, laborious numerical methods are used. For some integrals contour integration is helpful (Section 1.16)

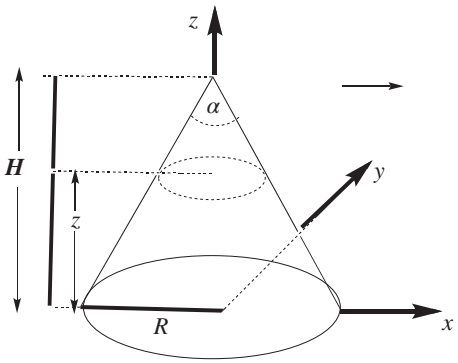
**Problem 1.5.11** Integrate $\int[\cos^2 x]^{-1}\mathrm{d}x$ (indefinite limits).

**Problem 1.5.12** Definite integral: evaluate $\int_{x=\pi/4}^{x=\pi} x\cos x^2\ \mathrm{d}x$.

**Problem 1.5.13** Derive the following definite integral:

$$I_0 = \int_{r=0}^{r=\infty} \exp\left(-ar^2\right) r\,\mathrm{d}r = \frac{1}{2a} \tag{1.5.55}$$

**Problem 1.5.14** Compute the volume of a right circular cone of radius $R$ and height $H$. This cone is widest at $z = 0$: this is an "upside-down ice cream cone." Use cylindrical coordinates $(r, \theta, z)$, where $r$ is the radial direction (range $r = 0$ to $r = R$), $z$ is the height direction (range $z = 0$ to $z = H$), and $\theta$ is the angle (range $\theta = 0$–$2\pi$). Remember that the differential volume element is $\mathrm{d}V = r\ \mathrm{d}r\ \mathrm{d}\theta\ \mathrm{d}z$, but the radius limit starts at $R$ when $z = 0$, but becomes zero at $z = H$, so the radius limit as a function of $z$ is $R\ (1 - z/H)$.




**Problem 1.5.15** For a right circular cone ("upside-down ice cream cone," see illustration) of radius $R$ (at the base), height $H$, opening angle $\alpha$ (which is defined by $R/H$=tan $(\alpha/2)$) prove that $x^2R^{-2} + y^2R^{-2} = (H - z)^2H^{-2}$.

**Problem 1.5.16** Compute the volume of a right circular cone ("upside-down ice cream cone!") of radius $R$ and height $H$ starting with a wide base (circle of radius $R$) in the plane at $z = 0$ and ending in a single point $(x = 0, y = 0, z = H)$. Foolishly

(since Problem 1.5.14 gave such an easy solution) use Cartesian coordinates $(x, y, z)$, where $x$ and $y$ are the directions in the horizontal plane (ranges $x = -R$ to $x = R$, $y = -R$ to $y = R$, but we must define ranges so that a circle is obtained), $z$ is the height direction (the $z$ range is $z = 0$ to $z = H$). The differential volume element is $dV = dx\ dy\ dz$, but the radius limit starts at $R$ when $z = 0$, but becomes zero at $z = H$, so the radius limit, as a function of $z$, is $R\ (1 - z/H)$.

**Problem 1.5.17** Compute the volume of the frustum of a right circular cone of radius $R$ and height $H$ (this is a cone with its top lopped off; the height is not $H$, but $B$, so now the radius as a function of height starts from $r = 0$ at $z = 0$, but does not reach $r = 0$ at $z = H$, but rather "stops" at $r = R(1 - B/H)$).

**Problem 1.5.18** Compute the volume of a pyramid of square base and side $L$ and height $H$. Use rectangular coordinates $(x, y, z)$, where $x$ and $y$ are the directions in the plane (range $x = 0$ to $x = L$, $y = 0$ to $y = L$), $z$ is the height (range $z = 0$ to $z = H$). Remember that the differential volume element is $dV = dx\ dy\ dz$, but the sides vary from $L$ at $z = 0$, become zero at $z = H$, and the side limit as a function of $z$ is $L\ (1 - z/H)$.

**Problem 1.5.19** Compute the volume of a truncated (frustum of $a$) pyramid of square base and side $L$, height $B$, and pyramid vertex height $H$. Use rectangular coordinates $(x, y, z)$, where $x$ and $y$ are the directions in the plane (ranges $x = 0$ to $x = L$ and $y = 0$ to $y = L$), $z$ is the height (range $z = 0$ to $z = H$). Remember that the differential volume element is $dV = dx\ dy\ dz$, but be careful that the sides vary from $L$ at $z = 0$, but become zero at $z = H$, so the side limit is $L(1 - z/H)$.

In some cases, a trick will help with integration. The definite integral

$$I_0 = \int_{r=0}^{r=\infty} \exp\left(-ar^2\right) dr = \frac{1}{2}\sqrt{\frac{\pi}{a}} \tag{1.5.56}$$

resembles Eq. (1.5.55), but is actually an "improper" integral, which must solved in two dimensions by a transformation of the integral to the orthogonal **planar polar coordinate system** $(r, \phi)$ in place of the usual orthogonal **Cartesian system** $(x, y)$:t

$$x = r\cos\phi \quad -\infty \le x \le +\infty; \quad r = +\left(x^2 + y^2\right)^{1/2} \quad 0 \le r \le +\infty \tag{1.5.57}$$

$$y = r\sin\phi \quad -\infty \le y \le +\infty; \quad \phi = \tan^{-1}(y/x) \quad 0 \le \phi \le \pi \tag{1.5.58}$$

the area element of the integrand is

$$dx\ dy = r\ dr\ d\phi \tag{1.5.59}$$

With this, we can integrate Eq. (1.5.56) by considering its square in Cartesian coordinates:

$$\begin{aligned} I_0^2 &= \int_{x=0}^{x=\infty} \exp\left(-ax^2\right) dx \int_{y=0}^{y=\infty} \exp\left(-ay^2\right) dy = \int_{x=0}^{x=\infty} dx \int_{y=0}^{y=\infty} \exp\left(-ax^2\right) \exp\left(-ay^2\right) dy \\ &= \{\text{change to plane polar coordinates}\} = \int_{r=0}^{r=\infty} r\ dr\ \exp\left(-ar^2\right) \int_{\theta=0}^{\theta=\pi/2} d\theta \\ &= (\pi/2)\left[\{-1/(2a)\}\exp\left(-ar^2\right)\right]_{r=0}^{r=\infty} = (\pi/2)\{1/(2a)\} = \pi/(4a) \end{aligned}$$

therefore, the square root of this result is

$$I_0 = \int_{x=0}^{x=\infty} \exp\left(-ax^2\right) dx = (1/2)(\pi/a)^{1/2}$$

The finite integral $(2/\sqrt{\pi})\int_{r=0}^{r=R} \exp(-r^2) dr = \text{erf}(R)$ is the **error function** (Section 1.17).

**Taylor series**:

$$\begin{aligned} f(x) = f(a) &+ (x-a)/1!(d/dx)\, f(x)|_{x=a} + (x-a)^2/2!\left(d^2/dx^2\right) f(x)|_{x=a} \\ &+ (x-a)^3/3!\left(d^3/dx^3\right) f(x)|_{x=a} + \ldots + (x-a)^n/n!\left(d^n/dx^n\right) f(x)|_{x=a} + \ldots \end{aligned} \tag{1.5.60}$$

**Maclaurin series** (=Taylor series for $a = 0$):

$$\exp(x) = e^x = 1 + x + x^2/2 + x^3/6 + x^4/24 + x^5/120 + \ldots + x^n/n! + \ldots \tag{1.5.61}$$

$$\exp(-x) = e^{-x} = 1 - x + x^2/2 - x^3/6 + x^4/24 + x^5/120 + \ldots + (-1)^n x^n/n! + \ldots \tag{1.5.62}$$

$$\cos x = 1 - x^2/2 + x^4/24 - x^6/720 + \ldots + (-1)^n x^{2n}/(2n)! + \ldots \tag{1.5.63}$$

$$\sin x = x - x^3/6 + x^5/120 - x^7/5040 + \ldots + (-1)^n x^{2n+1}/(2n+1)! + \ldots \tag{1.5.64}$$

$$\tan x = x + x^3/3 + (2/15)x^5 + (17/315)x^7 + \ldots \tag{1.5.65}$$

$$\text{cotan } x = 1/x - x/3 - x^2/45 - (2/945)x^5 - x^7/4725 - \ldots \tag{1.5.66}$$

$$\sec x = 1 + x^2/2 + (5/24)x^4 + (61/720)x^6 + (277/8064)x^8 + \ldots \tag{1.5.67}$$

Please note the term $(1/x)$ in Eq. (1.5.66).

**Euler** formula:

$$\exp(ix) = \cos x + i \sin x \tag{1.5.68}$$

where $i \equiv (-1)^{1/2}$; which gives the funny looking but nevertheless true result:

$$\exp(i\pi) = -1 \tag{1.5.69}$$

where $\pi$ is another irrational number:

$$\pi = 3.141\,592\,653\,589\,79\ldots \tag{1.5.70}$$

that has vexed geometers since classical antiquity (the question of "squaring the circle").

**Sideline 1.11** A mnemonic for $\pi$ is to count the letters in the statement: "Yes, I want a drink, of course, after the heavy sessions involving quantum mechanics" [1.26].

### 1.5.9 Differential Forms

Derivatives can also be used in **differential forms**: e.g. instead of writing the derivative

"$d(\sin x)/dx = \cos x$"

as a single expression (where $dx$ in the denominator of the derivative is usually taken as an infinitesimally small quantity), one can also rewrite the derivative in the differential form "$d(\sin x) = \cos x\ dx$".This becomes useful in thermodynamics (see Chapter 3).

**Sums:**

$$\sum_{i=1}^{i=n} a_i \equiv a_1 + a_2 + a_3 + \ldots + a_{n-1} + a_n \tag{1.5.71}$$

The Einstein summation convention is that a sum is understood for any repeated indices over their "full" defined range:

$$a_i b_i \equiv \sum_{i=1}^{i=n} a_i b_i \tag{1.5.72}$$

### 1.5.10 Infinite Series

**Arithmetic series:** If $a_i \equiv a + di$, then

$$\sum_{i=1}^{i=n} a_i = \sum_{i=1}^{i=n} (a + id) = na + [n(n-1)/2]d \tag{1.5.73}$$

**Geometric series:** If $a_i \equiv ad^i$, then

$$\sum_{i=1}^{i=n} a_i = \sum_{i=1}^{i=n} ad^i = a + ad + ad^2 + \ldots + ad^n = a\sum_{i=1}^{i=n} d^i = \frac{a\left(1-d^{n+1}\right)}{1-d} \tag{1.5.74}$$

If $n$ is infinite and $a_i \equiv ad^i$, then:

$$\sum_{i=1}^{i=\infty} ad^i = \frac{a}{1-d} \tag{1.5.75}$$

**Problem 1.5.20** Verify Eq. (1.5.74).

**Problem 1.5.21** Verify Eq. (1.5.75).

**Problem 1.5.22** Prove by differentiating a power series that

$$\sum_{i=1}^{i=\infty} ix^{[i(1-x)]} = x(1-x)^{-2}$$

### 1.5.11 Sum and Integral Killers: The Kronecker[76] Delta and the Dirac[77] Delta Function

The Kronecker delta is defined as

$$\delta_{mn} = 0 \text{ if } m \neq n$$
$$\delta_{mn} = 1 \text{ if } m = n \tag{1.5.76}$$

This Kronecker delta can be called a "sum killer."

$$\sum_{n=0}^{n=\infty} a_n \delta_{mn} = a_m \tag{1.5.77}$$

The Dirac delta function is defined as

$$\delta(x) \equiv 0 \text{ for } x \neq 0 \tag{1.5.78}$$

and

$$\delta(x) \equiv \infty \text{ for } x = 0 \tag{1.5.79}$$

provided that

$$\int_{x=-\infty}^{x=\infty} \delta(x)\,\mathrm{d}x \equiv 1 \tag{1.5.80}$$

The Dirac delta function, very often used in quantum mechanics, is an infinitely tall but infinitely thin function, which is anathema to some mathematicians (it is called an improper function). The Dirac delta function could be called an "integral killer" because

$$\int_{x=-\infty}^{x=\infty} \delta(x-a)f(x)\,\mathrm{d}x = f(a) \tag{1.5.81}$$

The Dirac delta function can also be defined (Problem 1.5.23) as the derivative of the Heaviside[78] step function (which is defined as $H(x) = 0$ for $x < 0$, $H(x) = 0.5$ for $x = 0$, and $H(x) = 1$ for $x > 0$).

76 Leopold Kronecker (1823–1891).
77 Paul Adrien Maurice Dirac (1902–1984).

The Dirac delta function can also be expressed as any of the following three limits:

$$\begin{aligned}\delta(x) &= \lim_{e\to 0}(1/2\pi)\int_{k=-\infty}^{k=\infty} \mathrm{d}k\ \exp(-\mathrm{i}kx - e|k|)\\ &= \lim_{e\to 0}(1/2\pi\mathrm{i})\left[(x-\mathrm{i}e)^{-1} - (x+\mathrm{i}e)^{-1}\right]\\ &= \lim_{e\to 0}(1/\pi)\left[e/\left(x^2+e^2\right)\right]\end{aligned}$$

**Sideline 1.12** Paul Dirac was so taciturn that a rate of human conversation (1 dirac = 1 word per hour) was named humorously after him.

**Problem 1.5.23** Prove $\int f(x)\delta'(x-a)\,\mathrm{d}x = -f'(\mathrm{a})$ (where prime denotes first derivative).

**Problem 1.5.24** Show that the Heaviside function, defined by $\{H(x)\equiv 0$ for $x<0$, $H(x)\equiv 1$ for $x>0$, $H(0)\equiv 1/2\}$ is the integral of the delta function.

**Partial fractions:** It is often convenient or desirable (e.g. in some difficult integrations) to break up a complicated factored polynomial expression in the denominator into **partial fractions** involving new denominators of order no higher than 2. For instance, it can be shown that the fraction on the left can be decomposed into a sum of the simpler fractions on the right:

$$\begin{aligned}(x+3)/\left[(x-1)^2(x-2)(x-3)\left(x^2+2x+2\right)\right] &= A(x-1)^{-1} + B(x-1)^{-2} + C(x-2)^{-1}\\ &+ D(x-3)^{-1} + (Ex+F)\left(x^2+2x+2\right)^{-1}\end{aligned}$$

where the coefficients $A$, $B$, $C$, $D$, and especially $E$ and $F$ are nonzero. These coefficients are found by "brute-force." The two rules for how to set up the partial fractions are (i) if a linear factor $ax+b$ occurs $n$ times in the denominator, then to this factor will correspond a sum of $n$ partial fractions: $A_1(ax+b)^{-1} + A_2(ax+b)^{-2} + \ldots + A_n(ax+b)^{-n}$ with $A_n \neq 0$; (ii) if a quadratic factor $ax^2+bx+c$ occurs $n$ times as factors in the denominator, then to this factor will correspond a sum of $n$ partial fractions: $(A_1x+B_1)(ax^2+bx+c)^{-1} + (A_2x+B_2)(ax^2+bx+c)^{-2} + \ldots + (A_nx+B_n)(ax^2+bx+c)^{-n}$ with $A_nx+B_n \neq 0$.

**Problem 1.5.25** Evaluate the coefficients $A$, $B$, $C$, $D$, $E$, and $F$ in the equation:

$$\begin{aligned}(x+3)/\left[(x-1)^2(x-2)(x-3)\left(x^2+2x+2\right)\right] &= A/(x-1) + B/(x-1)^2 + C/(x-2)\\ &+ D/(x-3) + (Ex+F)/\left(x^2+2x+2\right)\end{aligned}$$

**(Answer**: $A = 2.172\ 625$; $B = -143.2715$; $C = 367.7903$; $D = 3.0$; $E = -372.963$; $F = -1922.4936$)

### 1.5.12 Ordinary Differential Equations (ODEs)

Physics, chemistry, and engineering very often require the solution of equations involving not only functions but also, typically, functions mixed with derivatives of that function: these are the ODEs, which must be integrated. This assumes "continuum physics" that the independent variable (time, position, volume, or speed) is free to assume a continuum of values.

Most ODEs in physics are first order or second order in the derivatives. Their solution, i.e. finding the function which would depend explicitly on a variable in some form, but excluding its derivatives, is known as "solving the ODE": this will involve one or two integrations, as needed, and then satisfying pre-chosen "initial conditions," i.e. the values at some extremum of the problem at hand, so a family of solutions is narrowed down to the "needed one." This can be demonstrated by a simple example:

78 Oliver Heaviside (1850–1925).

**Example 1**
Solve:

$$(\mathrm{d}y/\mathrm{d}x) = 5y - 3 \tag{1.5.82}$$

with initial condition $y(2) = 1$. This is a first-order inhomogeneous ODE with constant coefficients (1, 5, −3). First, separate the variables, i.e. shift everything that depends on $y$ to the left side and everything that depends on $x$ to the right side:

$$\mathrm{d}y/(5y - 3) = \mathrm{d}x \tag{1.5.83}$$

now integrate both sides to get:

$$(1/5)\ln_e(|5y - 3|) = x + C \tag{1.5.84}$$

where $C$ is a constant of integration, and

$$y = \pm (1/5)\exp(5x + 5C) + (3/5) \tag{1.5.85}$$

Which sign is OK? By substitution into the ODE, both signs are OK. Next, using the initial condition $y(2) = 1$ will yield the value of $C$: $1 = y(2) = \pm(1/5)\exp(10 + 5C) + (3/5)$, whence $2 = \pm\exp(10 + 5C) = \pm\exp(10)\exp 5C$, so $\pm\ln_e 2 = 10 + 5C$, or $C = -2 \pm (1/5)\ln_e 2 = -1.8614$ or $C = -2 \pm 0.1386$ so either $C = -1.8613$ or $C = -2.1385$, so the final solution is either $y = 3/5 + (1/5)\exp(-9.3068 + 5x)$ or $y = 3/5 - (1/5)\exp(-10.6925 + 5x)$, explicitly:

$$\text{either } y = -0.6 + 0.000\ 018\ 13\exp(5x) \text{ or } -0.6 - 0.000\ 022\ 72\exp(5x) \tag{1.5.86}$$

**Example 2**
When solving a first-order inhomogeneous ODE

$$(\mathrm{d}y/\mathrm{d}x) + A(x)y(x) = B(x) \tag{1.5.87}$$

a trick is useful: because $(\mathrm{d}/\mathrm{d}x)[y(x)\exp(\int A(x)\,\mathrm{d}x)] = \exp(\int A(x)\,\mathrm{d}x)[A(x)y(x) + (\mathrm{d}y/\mathrm{d}x)]$; therefore, multiplying both sides by $\exp(\int A(x)\,\mathrm{d}x)$ will yield a general solution to Eq. (1.5.87):

$$\begin{aligned} &y(x)\exp\left(\int A(x)\,\mathrm{d}x\right) = \int \mathrm{d}x\, B(x)\exp\left(\int A(x)\,\mathrm{d}x + C\right) \text{ or} \\ &y(x) = C\exp\left[-\int A(x)\,\mathrm{d}x\right] + \left[\exp\left(-\int A(x)\,\mathrm{d}x\right)\right]\int \mathrm{d}x\, A(x)\exp\left[\int \mathrm{d}x\, B(x)\right] \end{aligned} \tag{1.5.88}$$

which looks awful, but becomes quite elegant when properly expanded.

**Example 3**
**Harmonic oscillator**: a very simple homogeneous second-order ODE is

$$\mathrm{d}^2y/\mathrm{d}t^2 - Hy = 0 \tag{1.5.89}$$

With initial condition $y = 0$ at $t = 0$, this differential equation can be integrated twice to yield:

$$y = y_{\max}\sin(\omega t) \tag{1.5.90}$$

**Example 3 (Continued)**

where the angular frequency $\omega$ (radians per second) is defined by

$$\omega \equiv H^{1/2} \tag{1.5.91}$$

and the frequency $\nu$ (in Hz = cycles per second) is

$$\nu \equiv (1/2\pi)\omega = (1/2\pi)H^{1/2} \tag{1.5.92}$$

By setting a different initial condition ($y = y_{max}$ at $t = 0$), the solution to Eq. (1.5.89) can be expressed also by a cosine function $y = y_{max} \cos(\omega t)$, or in general, the solution is the equation for classical simple harmonic motion:

$$y(t) = A \exp\ (i\omega t) + B \exp\ (-i\omega t) \tag{1.5.93}$$

**Problem 1.5.26** Solve $(dy/dx) + y - x^2 = 2$.

**Problem 1.5.27** Solve $x(dy/dx) + (1-x)y(x) = \exp(2x)$.

### 1.5.13 The Lagrange Method of Undetermined Multipliers

To prove important statistical mechanical results in Chapter 4, we need the method of undetermined multipliers, due to Lagrange: it can be enunciated as follows: assume that a function $f(x_1, x_2, \ldots x_n)$ of $n$ variables $x_1, x_2, \ldots x_n$ is subject to two auxiliary conditions:

$$g(x_1, x_2, \ldots, x_n) = 0 \tag{1.5.94}$$

$$h(x_1, x_2, \ldots, x_n) = 0 \tag{1.5.95}$$

We seek extrema (maxima, minima, or saddle points) of $f$, subject to these two conditions. We shall show that there exist two constants, defined as $\alpha$ and $\beta$ (these two are known as the Lagrange multipliers), such that the system of $n + 2$ equations:

$$\partial f(x_1, x_2, \ldots, x_n)/\partial x_i + \alpha \partial g(x_1, x_2, \ldots, x_n)/\partial x_i + \beta \partial h(x_1, x_2, \ldots, x_n)/\partial x_i = 0 \quad (i = 1, 2, \ldots, n) \tag{1.5.95}$$

$$g(x_1, x_2, \ldots, x_n) = 0 \tag{1.5.96}$$

$$h(x_1, x_2, \ldots, x_n) = 0 \tag{1.5.97}$$

when solved, will provide the desired extremum for $f$.

We want to find the conditions for which $df = 0$. Please remember that, while the constraints $g = 0$ and $h = 0$, in general $f \neq 0$.

Define the differential:

$$df \equiv \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right) dx_i \tag{1.5.98}$$

We want to find when $df = 0$. Since $g = 0$ and $h = 0$, therefore also, *a fortiori*:

$$dg = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x_i}\right) dx_i = 0 \tag{1.5.99}$$

and also:

$$dh = \sum_{i=1}^{n} (\partial h/\partial x_i) dx_i = 0 \tag{1.5.100}$$

Rewriting Eq. (1.5.98) now yields:

$$\partial f/\partial x_i = -\alpha\partial g/\partial x_i - \beta\partial h/\partial x_i \quad (i = 1, 2, ..., n) \tag{1.5.101}$$

so that finally:

$$\mathrm{d}f = -\alpha\sum_{i=1}^{n}\left(\frac{\partial g}{\partial x_i}\right)\mathrm{d}x_i - \beta\sum_{i=1}^{n}(\partial h/\partial x_i)\mathrm{d}x_i = -\alpha 0 - \beta 0 = 0 \tag{1.5.102}$$

i.e. we found the condition of Eq. (1.5.98), i.e. that $\mathrm{d}f = 0$, as desired, as conditions for $\alpha$ and $\beta$.

Proving whether this extremum in $f(x_1, x_2, ..., x_n)$, is a maximum, a minimum, or zero, is usually not done analytically (e.g. by further differentiation to make sure that $\mathrm{d}^2f > 0$ for a minimum, $\mathrm{d}^2f > 0$ for a maximum), but instead by recourse to physical arguments. Indeed, the values of the Lagrange multipliers $\alpha$ and $\beta$ are often found by using physical arguments.

### 1.5.14 Partial Differential Equations (PDEs)

Particularly in three dimensions, most physically meaningful differential equations will involve one of more partial derivatives (usually first order or second order in these derivatives), i.e. PDEs. These PDEs will be discussed after the introduction of vector operators which will make everything a bit more elegant. The solution of linear PDE begins with seeking ways of separating variables, so that the overall solution is the product of one-dimensional ODEs for each of the independent variables. The ODEs are then solved one at a time, and combined at the end as a product, satisfying boundary conditions completes the solution of the PDE. This method works for the one-dimensional mechanical wave equation (Section 1.7), the electromagnetic wave equation (Section 1.9), and for the Schrödinger equation (Chapter 2).

The **one-dimensional wave equation** is a linear homogenous PDE with constant coefficients that links the dependent variable ($y$) to the two independent variables $x$ (position) and $t$ (time):

$$\left(\partial^2 y/\partial x^2\right) - A\left(\partial^2 y/\partial t^2\right) = 0 \tag{1.5.103}$$

where $A$ is a constant. "Linear" means that multiplying the solution by a constant multiplies everything by that constant, with no problem. "Homogenous" means that there is no additional constant on the right-hand side of Eq. (1.5.103). It is solved by assuming that $y(x,t)$ can be factored into two terms that depend one on $x$ only, the other on $t$ only:

$$y(x, t) = X(x)T(t) \tag{1.5.104}$$

Substitution into Eq. (1.5.103) yields:

$$T(t)\left(\mathrm{d}^2X/\mathrm{d}x^2\right) = AX(x)\left(\mathrm{d}^2T/\mathrm{d}t^2\right) \tag{1.5.105}$$

Dividing both sides by $y(x, t) = X(x)T(t)$ yields:

$$(1/X(x))\left(\mathrm{d}^2X/\mathrm{d}x^2\right) = (A/T(t))\left(\mathrm{d}^2T/\mathrm{d}t^2\right) = Q \tag{1.5.106}$$

which has placed the $x$-dependence on the left and the $t$-dependence on the right. For Eq. (1.5.106) to hold for all $x$ and $t$, both terms must equal a common constant $Q$ (which we redefine as $-k^2$) and a common constant $Q/A$ (which we redefine as $-k^2$):

$$\left(\mathrm{d}^2X/\mathrm{d}x^2\right) = QX(x) = -k^2X(x) \tag{1.5.107}$$

$$\left(\mathrm{d}^2T/\mathrm{d}t^2\right) = (Q/A)T(t) = -\omega^2T(t) \tag{1.5.108}$$

These two ODEs are already familiar (Eq. (1.5.88)): their **classical harmonic solutions** can be sines, cosines, or complex exponentials, depending on the four initial conditions:

$$X(x) = F\exp(\mathrm{i}kx) + G\exp(-\mathrm{i}kx) \tag{1.5.109}$$

$$T(t) = H\exp(\mathrm{i}\omega t) + K\exp(-\mathrm{i}\omega t) \tag{1.5.110}$$

where $F$, $G$, $H$, and $K$ are constants. So, the overall solution is

$$y(x,t) = [F\exp(\mathrm{i}kx) + G\exp(-\mathrm{i}kx)]\ [H\exp(\mathrm{i}\omega t) + K\exp(-\mathrm{i}\omega t)] \tag{1.5.111}$$

or more simply, if we assume that the wave only travels along $x$ only from left to right:

$$y(x,t) = L\exp(\mathrm{i}kx + \mathrm{i}\omega t) + M\exp(\mathrm{i}kx - \mathrm{i}\omega t) \tag{1.5.112}$$

Finally, $L$ and $M$ are constants (we are getting rid of $A$, and $Q$), $\omega$ is the angular frequency (rad per second) and $k$ is the wavevector ($\mathrm{m}^{-1}$), and $\lambda = k^{-1}$ is the wavelength per cycle.

## 1.6 Mechanics, Vectors, Tensors, and Determinants

This section summarizes further useful concepts: forces and linear momentum; dimensional analysis; unit systems; vectors; derivative operators; tensors; determinants; similarity transformations; the eigenvalue-eigenvector problem; Euler angle rotation; covariant and contravariant vectors; moment of inertia; and angular momentum.

### 1.6.1 Forces and Linear Momentum

**Force** is defined by Newton's second law:

$$F = \partial \boldsymbol{p}/\partial t \tag{1.6.1}$$

where $F$ is the force, $t$ is the time, $\boldsymbol{p}$ is the particle's **linear momentum**, and $m$ is its mass (this equation is relativistically correct). When the momentum is given by the product of mass time velocity $\boldsymbol{v}$:

$$\boldsymbol{p} = m\boldsymbol{v} \tag{1.6.2}$$

(this is not valid at relativistic speeds) and if $\boldsymbol{a} = \mathrm{d}\boldsymbol{v}/\mathrm{d}t$, then:

$$\boldsymbol{F} = m\boldsymbol{a} \tag{1.6.3}$$

### 1.6.2 Dimensional Analysis

Despite the more philosophical discussion in Section 1.2, the **mass** $\boldsymbol{m}$ of any particle can be stated as the proportionality constant between force and acceleration, or as "amount of stuff" with dimension $[M]$. Then force and field have dimensions $[M][L][T]^{-2}$, while energy has units $[M][L]^2[T]^{-2}$.

### 1.6.3 Unit Systems

Five-unit systems should be summarized here:

A) the **SI** (Système International) units use kilograms, meters, seconds, ampères, kelvin, mole ($6.022 \times 10^{23}$ molecules per **gram**-mole, and **not** per kg-mole), and candela for $[M]$, $[L]$, $[T]$, current, absolute temperature, mole, and luminous intensity, respectively. It started from an MKS (m-kg-s) system and included an electrical unit as part of the definition, as first suggested by Giorgi[79] in 1904. There is also a very slight modification of SI, used in nonlinear optics, confusingly dubbed MKS by its users, but called SI′ below in Section 1.9.

B) The older **cgs** units started from the French Academy work of 1793 defining the gram and the meter, and use grams, centimeters, and seconds for $[M]$, $[L]$, and $[T]$, respectively. To define electrical and magnetic quantities, cgs comes in two flavors, **cgs-esu**, or simply **esu** (where statCoulombs are the units for electrical charge) and **cgs-emu**, or simply **emu**, where the Oersted[80] is the unit of magnetic field. There are other variants of the cgs units: Gaussian and Heaviside–Lorentz.

79 Giovanni Giorgi (1871–1950).
80 Hans Christian Oersted (1777–1851).

C) In addition to the SI and cgs systems, we can define a system of Hartree[81] **atomic units** (au). Alas, a slightly different set of Rydberg[82] au also exists, but will not be discussed here. The Hartree atomic units are defined so that (i) the unit of length $[L]$ is $a_0 = 1$ b $= 5.291\,77 \times 10^{-11}$ m = radius of the "Bohr[83] orbit" for hydrogen, (ii) the unit of mass $[M]$ is 1 electron mass $= m_e = 9.109 \times 10^{-31}$ kg, (iii) the unit of [action] is $[M][L]^2[T]^{-1} = (h/2\pi) \equiv \hbar =$ reduced Planck constant of action $= 1.055 \times 10^{-34}$ J s, (iv) the unit of electrical charge is the proton charge $e = 1.602 \times 10^{-19}$ C. This is equivalent to putting $\hbar = 1$, $m_e = 1$, $e = 1$ in all formulas. As a consequence, (v) the unit of time $[T]$ = time for 1 electron to travel 1 bohr radius $= 2.419 \times 10^{-17}$ s, (vi) the unit of energy $[M][L]^2[T]^{-2} = 1$ hartree = twice the ionization energy of the hydrogen atom $= 4.360 \times 10^{-18}$ J.

D) In **Planck units**, of interest to quantum gravity and to early cosmology, $\hbar = 1$, as in the atomic units, but $c$ = speed of light *in vacuo* = 1, and also $G$ = gravitational constant = 1.

E) In **Astronomical units** (unfortunately, and confusingly, also called au) the unit of mass is the solar mass ($1.988\,92 \times 10^{11}$ kg), the unit of length is the mean distance from Earth to Sun ($1.495\,978\,714\,64 \times 10^{11}$ m).

F) As updated in 2018, the kilogram is now defined as "the amount of substance which contains as many elementary entities as there are atoms in 0.012 kg of carbon-12."

For a single particle of mass $m$ and momentum $\boldsymbol{p}$, or velocity $\boldsymbol{v} = (\mathrm{d}\boldsymbol{r}/\mathrm{d}t)$, the **kinetic energy** $T$ (originally dubbed "vis viva," or live energy!) is defined as

$$T \equiv p^2/2m = mv^2/2 \tag{1.6.4}$$

$T$ is the positive definite. The potential energy $U$ introduced above depends on an arbitrary choice of its zero, which depends on "tradition," that is, on the whim of the first or loudest experimenter: $U$ becomes positive or negative, relative to that zero, depending on that tradition.

A concept helpful for solving simple celestial mechanics problems is the **centrifugal acceleration** $a$ of any body moving with a speed $v$ in a circular orbit of radius $r$:

$$a = v^2/r \tag{1.6.5}$$

which, expressed as a vector, is $\boldsymbol{a} = \mathrm{d}^2\boldsymbol{r}/\mathrm{d}t^2 = -v^2\boldsymbol{r}r^{-2}$. Its opposite is the **centripetal acceleration** that will keep a body on its circular path. In general, the acceleration $\boldsymbol{a}$ will have a radial component, the centripetal acceleration $(v^2/r)$ along an unit (inward) vector $\boldsymbol{e}_n$ and a tangential component, along the unit tangent vector $\boldsymbol{e}_t$: $\boldsymbol{a} = \mathrm{d}^2\boldsymbol{r}/\mathrm{d}t^2 = (\mathrm{d}v/\mathrm{d}t)\boldsymbol{e}_t + (v^2/r)\boldsymbol{e}_n$.

**Problem 1.6.1** Given that the average Earth–Moon distance is $R_{em} = 3.844 \times 10^8$ m and that the Moon's revolution around the Earth is 27.3 days (from which its tangential orbital velocity is $v_m = 1.0186 \times 10^3$ m s$^{-1}$), compute the mass of the Earth.

**Problem 1.6.2** Show that the relative gravitational potential energy at a height $h$ above the Earth's surface is

$$U_{rel} = mgh \tag{1.6.6}$$

where $U_{rel} = 0$ at the Earth's surface, by using the Maclaurin series

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + x^4 - \ldots \tag{1.6.7}$$

**Problem 1.6.3** Show that the escape velocity $v_{esc}$ from the Earth's gravitational field is $1.1 \times 10^4$ m s$^{-1}$. Given the necessary escape kinetic energy $(1/2)mv_{esc}^2 = (3/2)k_B T$, where $k_B = 1.3807 \times 10^{-23}$ J K$^{-1}$ atom$^{-1}$ is Boltzmann's constant, which molecules, at an effective temperature of 30 000 K, can leak out from the Earth's atmosphere into space? Is this temperature reasonable?

81 Douglas Rayner Hartree (1897–1958).
82 Johannes Robert Rydberg (1854–1919).
83 Niels Henrik David Bohr (1885–1965).

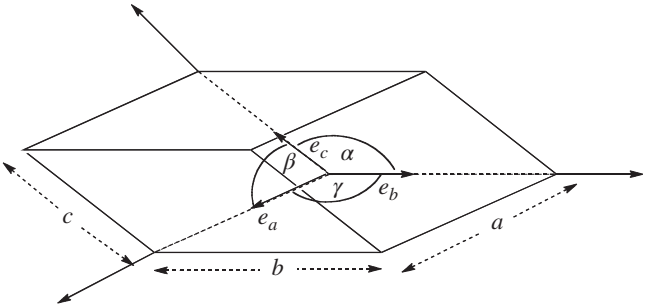



**Figure 1.5** The general unit cell in a triclinic (lowest symmetry) crystal. The unit cell has sides $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, angles $\alpha$, $\beta$, and $\gamma$ , and volume $V$.

**Problem 1.6.4** Show that the gravitational potential energy of an object of mass $m$ at the Earth's surface is only 7% due to the Earth and 93% due to the Sun (the Earth–Sun distance is 149 600 000 km; the mass of the Sun is $1.985 \times 10^{30}$ kg). So why do we not fall off the Earth and tumble toward the Sun?

**Problem 1.6.5** What is the center of gravity in the Earth–Sun trajectory (Sun–Earth distance $=1.496 \times 10^{11}$ m; Earth mass $= 5.977 \times 10^{24}$ kg; Sun mass $= 1.985 \times 10^{30}$ kg)?

**Problem 1.6.6** If a satellite is to reach an orbit 100 km above the surface of the Earth, what tangential velocity must it have as it enters the orbit? How long will it take to make one revolution around the Earth (Earth mass $=5.977 \times 10^{24}$ kg; Earth radius $=6.371 \times 10^{6}$ m)?

### 1.6.4 Vectors

Since forces are represented by vectors, we next review some properties of these **vectors**, with particular applications to crystals. The position vector $\boldsymbol{r}$ is usually given in a Cartesian (orthogonal) system, but crystals are defined as symmetric objects with translational symmetry, with a fundamental or unit cell, which is the basic nonrepeating unit that is not necessarily orthogonal (see Sections 6.1 and 6.10). The unit cell has three axes $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ (measured in nm or in Å); the axes do form a right-handed system, and are, in general, the corners of an oblique parallelepiped (in the lowest symmetry system, the triclinic system, see Figure 1.5); the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ is called $\gamma$; the angle between $\boldsymbol{b}$ and $\boldsymbol{c}$ is called $\alpha$, and the angle between $\boldsymbol{c}$ and $\boldsymbol{a}$ is called $\beta$.

The dot or **scalar inner product** between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is a scalar quantity, defined by

$$\boldsymbol{a} \cdot \boldsymbol{b} \equiv |\boldsymbol{a}||\boldsymbol{b}| \cos \gamma = \boldsymbol{b} \cdot \boldsymbol{a} \tag{1.6.8}$$

where the angle between the $\boldsymbol{a}$ and $\boldsymbol{b}$ axes is $\gamma$; similarly:

$$\boldsymbol{b} \cdot \boldsymbol{c} = |\boldsymbol{b}||\boldsymbol{c}| \cos \alpha = \boldsymbol{c} \cdot \boldsymbol{b} \tag{1.6.9}$$

$$\boldsymbol{c} \cdot \boldsymbol{a} = |\boldsymbol{c}||\boldsymbol{a}| \cos \beta = \boldsymbol{a} \cdot \boldsymbol{c} \tag{1.6.10}$$

The **length**, or norm of a vector, is defined as the positive square root of its inner or dot product with itself:

$$\text{length of } A = + [A \cdot A]^{1/2} \tag{1.6.11}$$

One can also define the $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ system in terms of any arbitrarily oriented Cartesian system:

$$\begin{aligned} \boldsymbol{a} &\equiv a_x \boldsymbol{e}_x + a_y \boldsymbol{e}_y + a_z \boldsymbol{e}_z \\ \boldsymbol{b} &\equiv b_x \boldsymbol{e}_x + b_y \boldsymbol{e}_y + b_z \boldsymbol{e}_z \\ \boldsymbol{c} &\equiv c_x \boldsymbol{e}_x + c_y \boldsymbol{e}_y + c_z \boldsymbol{e}_z \end{aligned} \tag{1.6.12}$$

but in crystallography it is customary to align either the $\boldsymbol{b}$ axis with $\boldsymbol{e}_y$ or the $\boldsymbol{c}$ axis with $\boldsymbol{e}_z$. The inner product can be defined in a space of any dimensions from two to infinity and obeys the associative and commutative laws. Of course, if $\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal, then the dot product $\boldsymbol{a} \cdot \boldsymbol{b}$ is zero; in the Cartesian system used above:

$$\boldsymbol{e}_x \cdot \boldsymbol{e}_y = \boldsymbol{e}_y \cdot \boldsymbol{e}_z = \boldsymbol{e}_z \cdot \boldsymbol{e}_x = 0 \tag{1.6.13}$$

and the length, or norm, of the unit vectors is, of course, unity:

$$\boldsymbol{e}_x \cdot \boldsymbol{e}_x = \boldsymbol{e}_y \cdot \boldsymbol{e}_y = \boldsymbol{e}_z \cdot \boldsymbol{e}_z = 1 \tag{1.6.14}$$

The **vector product** or "cross" product (term coined by Gibbs[84]) is defined only in three-dimensional space: the vector product, or cross product, of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is a vector $\boldsymbol{v}$, whose magnitude is $|\boldsymbol{a}||\boldsymbol{b}|\sin\gamma$, where $\gamma$ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$, and whose direction is perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$, and whose orientation is such that $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{v}$ form a right-handed Cartesian coordinate system:

$$\boldsymbol{v} = \boldsymbol{a} \times \boldsymbol{a} \equiv \boldsymbol{e}_v \, |\boldsymbol{a}||\boldsymbol{b}| \sin\gamma \tag{1.6.15}$$

In a left-handed coordinate system $\boldsymbol{v}$ changes sign: therefore, the vector product is called a **pseudovector**. In a Cartesian coordinate system, by applying the definition of a cross product in this orthogonal system, the unit vectors $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ are related as follows:

$$\boldsymbol{e}_z \times \boldsymbol{e}_y = \boldsymbol{e}_z = -\boldsymbol{e}_y \times \boldsymbol{e}_x, \boldsymbol{e}_y \times \boldsymbol{e}_z = \boldsymbol{e}_x = -\boldsymbol{e}_z \times \boldsymbol{e}_y, \text{ and } \boldsymbol{e}_z \times \boldsymbol{e}_x = \boldsymbol{e}_y = -\boldsymbol{e}_x \boldsymbol{e}_z \tag{1.6.16}$$

while the cross product of any vector with itself vanishes:

$$\boldsymbol{e}_x \times \boldsymbol{e}_x = \boldsymbol{e}_y \times \boldsymbol{e}_y = \boldsymbol{e}_z \times \boldsymbol{e}_z = 0 \tag{1.6.17}$$

**Problem 1.6.7** The sum of two vectors $\boldsymbol{A} + \boldsymbol{B}$ is a new vector $\boldsymbol{C}$, which can be drawn by moving **B** to have the same origin (tail) as $\boldsymbol{A}$; the vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ become sides of a parallelogram. The vector $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ starts from the common origin and has length equal to the longer body diagonal of the parallelogram.

**Problem 1.6.8** The difference of vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ is a new vector $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B}$, which can be drawn by moving $\boldsymbol{B}$ to have the same origin (tail) as $\boldsymbol{A}$; the vector $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B}$ starts from the vertex of $\boldsymbol{B}$ and ends on the vertex of $\boldsymbol{A}$. Equivalently, $\boldsymbol{D}$ is the shorter body diagonal of the parallelogram defined in the previous problem. Show all this in a diagram.

**Problem 1.6.9** Show that the sum of vectors is commutative: $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$.

**Problem 1.6.10** Verify the distributive property of the vector product:

$$\boldsymbol{A} \times (\boldsymbol{B} + \boldsymbol{C} + \boldsymbol{D} + ...) = \boldsymbol{A} \times \boldsymbol{B} + \boldsymbol{B} \times \boldsymbol{C} + \boldsymbol{A} \times \boldsymbol{D} + ... \tag{1.6.18}$$

**Problem 1.6.11** If you multiply a vector by a scalar, the result is a new vector in the same direction, whose length is that scalar times the length of the original vector.

Before continuing, the concept of determinant must be introduced.

### 1.6.5 Determinants

In an $n$-dimensional space, the determinant $\det\,\alpha$ is defined as a single number or function that is evaluated from a rectangular array of numbers (or functions) $\alpha_{ij}$ (where the first [second] suffix indicates the row [the column]):

$$\det\,\alpha \equiv \left|\alpha_{ij}\right| \equiv \begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1,n-1} & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2,n-1} & \alpha_{2n} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \cdots & \alpha_{3,n-1} & \alpha_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-1,1} & \alpha_{n-1,2} & \alpha_{n-1,3} & \cdots & \alpha_{n-1,n-1} & \alpha_{n-1,n} \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{n,n-1} & \alpha_{nn} \end{vmatrix} \tag{1.6.19}$$

84 Josiah Willard Gibbs, Jr. (1839–1903).

by the following recipe:

$$\det\ \boldsymbol{\alpha} = \sum_{i=1}^{i=n}\sum_{j=1}^{j=n}(-1)^h \alpha_{1j_1}\alpha_{2j_2}\alpha_{3j_3}...\alpha_{nj_n} = \sum_{i=1}^{i=n} a_{ik}A^{ik} \quad (k=1,2,...,n) \tag{1.6.20}$$

where the indices $(j_1, j_2, ..., j_n)$ are permutations of the "natural" or perfectly sequential order of the first $n$ integers $(1, 2, 3, ..., n)$ and $h$ is the number of twofold exchanges of any two elements needed to recreate this "natural" order; the **co-factor or minor** $A^{ik}$ of any element $\alpha_{ik}$ is defined as the subdeterminant created by eliminating the $i$th row and the $k$-column of the original determinant, multiplied by $(-1)^{i+k}$. A determinant can be computed as a sum of the minors of any one given row (or column) (choose one):

$$\det\ \boldsymbol{\alpha} = \sum_{i=1}^{i=n} \alpha_{ik}(-1)^{i+k}\left|A^{ik}\right| \quad (k=1,2,...,n) \tag{1.6.21}$$

where the operator $|\ |$ also means "determinant." $\det\ \alpha = 0$ if (i) all elements of any row are zero, or (ii) if all elements of any column are all zero, or (iii) if any two rows have equal elements in the same order, or (iv) if any two columns have equal elements in the same order. The evaluation of determinant with $n \geq 4$ rows and columns are laborious. For a $2\times 2$ determinant, the evaluation is trivial:

$$\det\ \alpha \equiv \left|\alpha_{ij}\right| \equiv \begin{vmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{21} & \alpha_{22}\end{vmatrix} = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} \tag{1.6.22}$$

For $3\times 3$ determinants, Sarrus'[85] rule is simple: repeat and append the *first two* columns at the end and multiply diagonally down three times with $a+1$ prefactor ($a_{11}a_{22}a_{33}$, then $a_{12}a_{23}a_{31}$, then $a_{13}a_{21}a_{32}$), then multiply diagonally up three times with $a-1$ prefactor ($-a_{31}a_{22}a_{13}$, then $-a_{32}a_{23}a_{11}$, then $-a_{33}a_{21}a_{12}$ ). For example,

$$\det\begin{vmatrix}5 & 7 & 9\\ 3 & 2 & 14\\ 12 & 6 & 2\end{vmatrix} = \begin{vmatrix}5 & 7 & 9\\ 3 & 2 & 14\\ 12 & 6 & 2\end{vmatrix}\begin{Bmatrix}5 & 7\\ 3 & 2\\ 12 & 6\end{Bmatrix}$$
$$= (+1)\times 5\times 2\times 2 + (+1)\times 7\times 14\times 12 + (+1)\times 9\times 3\times 6 + (-1)\times 12\times 2\times 9$$
$$+ (-1)\times 6\times 14\times 5 + (-1)\times 2\times 3\times 7$$
$$= 20 + 1176 + 162 - 216 - 420 - 42$$
$$= 680$$

A discussion given below will associate square matrices with determinants.

Now we will put $3\times 3$ determinants to good use. Combining Eqs. (1.6.14) and (1.6.17):

$$\begin{aligned}\boldsymbol{v} &= \boldsymbol{a}\times\boldsymbol{b} = \left(a_x\boldsymbol{e_x} + a_y\boldsymbol{e_y} + a_z\boldsymbol{e_z}\right)\times\left(b_x\boldsymbol{e_x} + b_y\boldsymbol{e_y} + b_z\boldsymbol{e_z}\right) = -\boldsymbol{b}\times\boldsymbol{a}\\ &= \boldsymbol{e_x}\left(a_yb_z - a_zb_y\right) + \boldsymbol{e_y}\left(a_zb_x - a_xb_z\right) + \boldsymbol{e_z}\left(a_xb_y - a_yb_z\right)\end{aligned} \tag{1.6.23}$$

and then using the properties of a $3\times 3$ determinant, one gets:

$$\boldsymbol{v} = \boldsymbol{a}\times\boldsymbol{b} \equiv \begin{vmatrix}\boldsymbol{e_x} & \boldsymbol{e_y} & \boldsymbol{e_z}\\ a_x & a_y & a_z\\ b_x & b_y & b_z\end{vmatrix} \tag{1.6.24}$$

**Problem 1.6.12** Prove:

$$\boldsymbol{A}\times(\boldsymbol{B}\times\boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A}\cdot\boldsymbol{C}) - (\boldsymbol{A}\cdot\boldsymbol{B})\boldsymbol{C} \tag{1.6.25}$$

85 Pierre Frédéric Sarrus (1798–1861).

**Problem 1.6.13** Prove:

$$\boldsymbol{A}\times\boldsymbol{B}\cdot\boldsymbol{C}\times\boldsymbol{D}=(\boldsymbol{A}\cdot\boldsymbol{C})\,(\boldsymbol{B}\cdot\boldsymbol{D})-(\boldsymbol{B}\cdot\boldsymbol{C})\,(\boldsymbol{A}\cdot\boldsymbol{D}) \tag{1.6.26}$$

**Problem 1.6.14** Prove:

$$(\boldsymbol{A}\times\boldsymbol{B})\times(\boldsymbol{C}\times\boldsymbol{D})=\boldsymbol{C}(\boldsymbol{D}\cdot\boldsymbol{A}\times\boldsymbol{B})-\boldsymbol{D}(\boldsymbol{A}\times\boldsymbol{B}\cdot\boldsymbol{C}) \tag{1.6.27}$$

The cross product is anticommutative, i.e. it changes sign when the factors are reversed. Indeed, the cross product is really a "**pseudovector**," or "polar vector," which has all the desirable properties of vectors, plus one undesirable one: a pseudovector is "married" to a right-handed system (in a left-handed system its magnitude is the same, but its sign changes). To differentiate them from their ill-behaved cousins (the pseudovectors), the "normal" vectors are also called sometimes "axial vectors." The anticommutation is also implicit in the properties of a determinant. Geometrically, the magnitude of the vector product of $\boldsymbol{a}$ and $\boldsymbol{b}$ is the area $A$ of the parallelogram with $\boldsymbol{a}$ and $\boldsymbol{b}$ as the sides:

$$A=|\boldsymbol{a}\times\boldsymbol{b}|=|a||b|\ \sin\gamma \tag{1.6.28}$$

We next calculate the volume $V$ of the parallelepiped of Figure 1.5: we need the projection of $\boldsymbol{c}$ onto the direction normal to both $\boldsymbol{a}$ and $\boldsymbol{b}$: this demands the dot product of $\boldsymbol{c}$ times the cross product of $\boldsymbol{a}$ and $\boldsymbol{b}$: indeed (in a cyclic fashion) it can be seen that:

$$V=(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}=(\boldsymbol{b}\times\boldsymbol{c})\cdot\boldsymbol{a}=(\boldsymbol{c}\times\boldsymbol{a})\cdot\boldsymbol{b} \tag{1.6.29}$$

and using the commutative properties of the scalar product and the anticommutative properties of the vector product one sees the cyclic permutation of vectors as giving a positive value for the volume, while any twofold permutation yields a negative volume (left-handed system!):

$$\begin{aligned}V\ &=\boldsymbol{c}\cdot(\boldsymbol{a}\times\boldsymbol{b})=\boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c})=\boldsymbol{b}\cdot(\boldsymbol{c}\times\boldsymbol{a})=-(\boldsymbol{b}\times\boldsymbol{a})\cdot\boldsymbol{c}=-(\boldsymbol{c}\times\boldsymbol{b})\cdot\boldsymbol{a}=-(\boldsymbol{a}\times\boldsymbol{c})\cdot\boldsymbol{b}\\&=-\boldsymbol{c}\cdot(\boldsymbol{b}\times\boldsymbol{a})=-\boldsymbol{a}\cdot(\boldsymbol{c}\times\boldsymbol{b})=-\boldsymbol{b}\cdot(\boldsymbol{a}\times\boldsymbol{c})\end{aligned} \tag{1.6.30}$$

$V$ is also "married" to a right-handed system, so $V$ should be called a **pseudoscalar**. To evaluate $V$, we need the difficult projection of $\boldsymbol{c}$ onto the direction normal to both $\boldsymbol{a}$ and $\boldsymbol{b}$: this is not easy (see Problem 1.6.14). The result is

$$V=abc\left[1-\cos^2\alpha-\cos^2\beta-\cos^2\gamma+2\cos\alpha\ \cos\beta\ \cos\gamma\right]^{1/2} \tag{1.6.31}$$

An easier result to see from Eqs. (1.6.20) and (1.6.26) is

$$V=\boldsymbol{a}\bullet(\boldsymbol{b}\times\boldsymbol{c})=\begin{vmatrix}a_x & a_y & a_z\\ b_x & b_y & b_z\\ c_x & c_y & c_z\end{vmatrix} \tag{1.6.32}$$

A detailed discussion of reciprocal space is given in Section 6.10.

**Problem 1.6.15** Prove Eq. (1.6.32).

### 1.6.6 Matrices

The notion of determinants was introduced earlier. Now we present matrices, with which determinants are frequently associated. A rectangular matrix $\boldsymbol{W}$ with 5 rows and 4 columns is an organized set of $5\times4$ elements (numbers or functions) enclosed by braces:

$$\boldsymbol{W}=\begin{pmatrix}w_{11} & w_{12} & w_{13} & w_{14}\\ w_{21} & w_{22} & w_{23} & w_{24}\\ w_{31} & w_{32} & w_{33} & w_{34}\\ w_{41} & w_{42} & w_{43} & w_{44}\\ w_{51} & w_{52} & w_{53} & w_{54}\end{pmatrix} \tag{1.6.33}$$

The most useful matrices are the $n\times n$ square matrices, which can be associated with determinants.

### 1.6.7 Tensors

While a scalar (a pure number, real or complex) $p$ merely multiplies the length of a vector $\boldsymbol{V}$:

$$\boldsymbol{V}' = p\boldsymbol{V} \tag{1.6.34}$$

(i.e. multiplies all three components of $\boldsymbol{V}$ by $p$), there exist $3 \times 3$ tensors $\boldsymbol{\alpha}$ that can rotate them: they are known as **tensors of rank two** ($3 \times 3$ matrices):

$$\boldsymbol{V}'' \equiv \alpha \boldsymbol{V} \tag{1.6.35}$$

If we are rotating a (column) vector $\boldsymbol{V}$ (or tensor of rank one) in three-dimensional space, then the tensor $\boldsymbol{\alpha}$ of rank two will have, in general, $3 \times 3 = 9$ independent elements:

$$\boldsymbol{\alpha} \equiv \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \tag{1.6.36}$$

where each tensor element $\alpha_{ij}$ is a scalar (i.e. a real or complex number).

If a tensor $\boldsymbol{\beta}$ is **symmetric**, i.e. if $\beta_{ij} = \beta_{ji}$, then this tensor has only six independent elements:

$$\boldsymbol{\beta} \equiv \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{12} & \beta_{22} & \beta_{23} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{pmatrix} \tag{1.6.37}$$

A tensor $\boldsymbol{\gamma}$ is **Hermitian** if its off-diagonal elements are the complex conjugates of the corresponding elements across the main diagonal $\gamma_{ij} = \gamma_{ji}^*$ (this Hermitian condition is written in matrix notation as $\boldsymbol{\gamma} = \boldsymbol{\gamma}^\dagger$); a tensor $\boldsymbol{\delta}$ is **unitary** if the inverse of the matrix is equal to its Hermitian conjugate: $\boldsymbol{\delta} = (\boldsymbol{\delta}^\dagger)^{-1}$.

The trace Tr of a square tensor is the scalar sum of its diagonal elements:

$$\mathrm{Tr}(\boldsymbol{\alpha}) \equiv \sum_i \alpha_{ii} = \{\text{if } \boldsymbol{\alpha} \text{ is a } 3 \times 3 \text{ tensor}\} = \alpha_{11} + \alpha_{22} + \alpha_{33} \tag{1.6.38}$$

Of course, $n \times n$ tensors of rank two can be defined for dimensions $n > 3$: they occur frequently in four-dimensional special and general relativity theories.

### 1.6.8 Similarity Transformations

A matrix $\boldsymbol{A}$ is **singular** if $\det \boldsymbol{A} = 0$; conversely, $\boldsymbol{A}$ is ***nonsingular*** if $\det \boldsymbol{A} \neq 0$.

If the matrix $\boldsymbol{A}$ is symmetric, Hermitian, or unitary, then there is a system of $3 \times 3$ rotation matrices $\boldsymbol{R}$ (and their inverses $\boldsymbol{R}^{-1}$) which will rotate the matrix elements $A_{ij}$ so that the only nonzero elements will appear on the diagonal: this is known as a **similarity transformation** or as a **principal-axis transformation** or **diagonalization**:

$$\boldsymbol{A}_{\text{diag}} = \boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R} \tag{1.6.39}$$

This transformation is crucial when some measured $3 \times 3$ matrix has nine experimental values, and a coordinate system is sought which will highlight the physically significant components of this matrix, which may be as few as three after the appropriate similarity transformation into the "right" coordinate system. Written out in full, the relationship between a matrix and its inverse is

$$\boldsymbol{R}\boldsymbol{R}^{-1} = \boldsymbol{R}^{-1}\boldsymbol{R} = \boldsymbol{E} \tag{1.6.40}$$

where $\boldsymbol{E}$ is the unit matrix:

$$\boldsymbol{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.6.41}$$

### 1.6.9 Eigenvalue-Eigenvector Problem

The problem of finding the rotation matrix that will "diagonalize" some symmetric, Hermitian, or unitary matrix $\boldsymbol{A}$ can be recast as an **eigenvalue-eigenvector** problem: we seek the characteristic solutions to the problem:

$$\boldsymbol{K} \equiv \boldsymbol{A} - \lambda \boldsymbol{E} \tag{1.6.42}$$

where $\lambda$ is one of $n$ scalar (real or complex) "characteristic values" or "eigenvalues," and $\boldsymbol{K}$ is the characteristic matrix of $\boldsymbol{A}$ (a matrix constructed by adjoining all column vectors to each other). In particular, the scalar $n$th degree equation:

$$K(\lambda) \equiv \det \boldsymbol{K} = \det|\boldsymbol{A} - \lambda \boldsymbol{E}| = 0 \tag{1.6.43}$$

has $n$ solutions (the $n$ eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$), so that $K(\lambda)$ can be rewritten as

$$K(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)...(\lambda - \lambda_n) = 0$$

The determinant may factor naturally: this is rare. For $n = 2$, 3, or 4, one can calculate exactly the roots of a quadratic, cubic, or quartic polynomial equation in $\lambda$, respectively; otherwise, one must resort to numerical methods (diagonalize the matrix). For instance, when $n = 7$, then the eigenvalue matrix $\boldsymbol{\Lambda} \equiv \boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R}$ is the diagonal matrix:

$$\boldsymbol{\Lambda} = \begin{vmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_7 \end{vmatrix} \tag{1.6.44}$$

The problem of finding the $n$ eigenvalues ($\lambda_i$, $i = 1, 2, ..., n$) and the $n$ corresponding eigenvectors $\boldsymbol{x}$:

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \tag{1.6.45}$$

is thus equivalent to diagonalizing the matrix $\boldsymbol{A}$. The $n \times n$ **eigenvalue-eigenvector equation** or **secular equation** (so-called because astronomers used it to describe motions of distant planets tracked over the course of several centuries, or *saecula*) is given by

$$(\boldsymbol{A} - \lambda \boldsymbol{E})\boldsymbol{x} = \boldsymbol{O} \tag{1.6.46}$$

where $\boldsymbol{E}$ is the $n \times n$ unit matrix and $\boldsymbol{O}$ is the $n \times n$ null matrix. Its solution $\boldsymbol{x}$ has the eigenvector components:

$$x_i = (-1)^{i+j} \left| \left(\boldsymbol{A} - \lambda_j \boldsymbol{E}\right)_{ij} \right| \quad i = 1, 2, 3; j = 1, 2, ..., n \tag{1.6.47}$$

where the $A_{ij}$ are the $(n-1) \times (n-1)$ "minors" of the original $n \times n$ determinant $|\boldsymbol{A} - \lambda \boldsymbol{E}|$, obtained by eliminating the $i$th row and $j$th column of the original determinant.

The eigenvectors coefficients are redefined by normalization. If the eigenvalues are not degenerate, then the eigenvectors are automatically mutually orthogonal; if the eigenvalues are degenerate, then the eigenvectors can be forced to be orthogonal. The $n \times n$ rotation matrix $\boldsymbol{R}$ is constructed by accosting to each other the $n \times 1$ eigenvectors. The inverse rotation matrix $\boldsymbol{R}^{-1}$ is the transpose of $\boldsymbol{R}$. The transformation of $\boldsymbol{A}$ into $\boldsymbol{\Lambda}$ is known as a similarity transformation:

$$\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R} = \boldsymbol{\Lambda} \tag{1.6.48}$$

It is important to realize that the trace of the matrix $\boldsymbol{A}$ (the sum of the diagonal terms) is invariant under a similarity transformation.

**Problem 1.6.16** Invert the symmetric matrix:

$$\boldsymbol{A} = \begin{pmatrix} 21 & 3 & -6 \\ 3 & 14 & 7 \\ -6 & 7 & 12 \end{pmatrix}.$$

**Problem 1.6.17** Diagonalize the $3 \times 3$ symmetric real matrix:

$$\boldsymbol{A} = \begin{pmatrix} 21 & 3 & -6 \\ 3 & 14 & 7 \\ -6 & 7 & 12 \end{pmatrix}$$

by first solving the determinantal equation:

$$\det \begin{vmatrix} 21-\lambda & 3 & -6 \\ 3 & 14-\lambda & 7 \\ -6 & 7 & 12-\lambda \end{vmatrix} = 0$$

Finally, check that the trace of $\boldsymbol{A}$ is equal to the trace of $\boldsymbol{\Lambda}$.

**Problem 1.6.18** For the matrix:

$$\boldsymbol{A} = \begin{pmatrix} 21 & 3 & -6 \\ 3 & 14 & 7 \\ -6 & 7 & 12 \end{pmatrix}$$

which in Problem 1.6.17 was transformed into the diagonal eigenvalue matrix:

$$\boldsymbol{\Lambda} = \begin{pmatrix} 24.032 & 324\ 397\ 377 & 0 & 0 \\ 0 & 19.474\ 154\ 527\ 374 & 0 & 0 \\ 0 & 0 & 3.493\ 521\ 075\ 249 & \end{pmatrix}$$

find its eigenvectors

**Problem 1.6.19** What is the inverse of the rotation matrix $\boldsymbol{R}$

$$\boldsymbol{R} = \begin{pmatrix} 0.875\,212\,771\,955 & -0.333\,162\,753\,797 & 0.350\,720\,950\,142 \\ -0.072\,043\,610\,084 & -0.806\,709\,369\,55 & -0.586\,540\,460\,086 \\ -0.478\,343\,311\,470 & -0.488\,080\,498\,03 & 0.730\,044\,590\,288 \end{pmatrix}$$

which was computed in Problem 1.6.18?

### 1.6.10 Planar Rotations

One often needs to rotate a coordinate system. In the $xy$ plane, a counterclockwise rotation by a **negative** angle $-\phi$ causes a rotation from old vectors $x$, $y$ to new vectors $x'$, $y'$ **(**Problem 1.6.20):

$$x' = x\cos\phi + y\sin\phi \tag{1.6.49}$$

$$y' = -x\sin\phi + y\cos\phi \tag{1.6.50}$$

which can be represented as a column vector $\boldsymbol{X} = (xy)$ rotated to a new column vector $\boldsymbol{X}' = (x'y')$ by premultiplication by the square rotation matrix $\boldsymbol{A} = (a_{ij})$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.6.51}$$

$$\boldsymbol{X}' = \boldsymbol{A}\boldsymbol{X} \tag{1.6.52}$$

Since the coordinate system is simply rotated, not shrunk or expanded in any way; therefore, the matrix $\boldsymbol{A}$ must be orthonormal:

$$\sum_i a_{ij}a_{jk} = \delta_{ik}(\ j,k = 1,2) \tag{1.6.53}$$

### 1.6.11 Eulerian Rotations

To do a full rotation in three-dimensional space, **Eulerian rotation angles** can be defined, as **three** successive rotations, each in two dimensions. The first rotation (rotation matrix $\boldsymbol{A}$) is a counterclockwise rotation by $\alpha$ degrees about the $z$ axis: it leaves the $z$ axis unchanged ($z' = z$), but rotates the axis $x$ to $x'$ by $\alpha$ degrees, and the axis $y$ to $y'$ by $\alpha$ degrees.

$$\boldsymbol{X}' = \boldsymbol{A}\boldsymbol{X} \tag{1.6.54}$$

The second rotation (rotation matrix $\boldsymbol{B}$) rotates by a counterclockwise rotation by $\beta$ degrees about the axis $x'$: it leaves the $x'$ axis unchanged: $x'' = x'$, but rotates $y'$ to $y''$ by $\beta$ degrees, and $z'$ to $z''$ by $\beta$ degrees:

$$\boldsymbol{X}'' = \boldsymbol{B}\boldsymbol{X}' \tag{1.6.55}$$

The third rotation (rotation matrix $\boldsymbol{C}$) rotates by a counterclockwise rotation by $\gamma$ degrees about the axis $z'$: it leaves the $z''$ axis unchanged: $z''' = z''$, but rotates $x''$ to $x'''$ by $\gamma$ degrees, and $y''$ to $y'''$ by $\gamma$ degrees: $\boldsymbol{X}''' = \boldsymbol{C}\boldsymbol{X}''$. The rotation matrices are

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{pmatrix}, C = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.6.56}$$

The overall Eulerian rotation matrix $\boldsymbol{E} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{C}$ is given by

$$E = \begin{pmatrix} \cos\gamma\cos\alpha - \sin\gamma\sin\alpha\cos\beta & \cos\gamma\sin\alpha + \sin\gamma\cos\alpha\cos\beta & \sin\beta\sin\gamma \\ -\sin\gamma\cos\alpha - \cos\gamma\sin\alpha\cos\beta & -\sin\gamma\sin\alpha + \cos\gamma\cos\alpha\cos\beta & \sin\beta\cos\gamma \\ \sin\alpha\sin\beta & -\cos\alpha\sin\beta & \cos\beta \end{pmatrix} \tag{1.6.57}$$

The inverse matrix $\boldsymbol{E}^{-1}$ is equal to the transpose matrix $\boldsymbol{E}^{\mathrm{T}}$ (where rows and columns are interchanged):

$$E^{-1} = \begin{pmatrix} \cos\gamma\cos\alpha - \sin\gamma\sin\alpha\cos\beta & -\sin\gamma\sin\alpha - \cos\gamma\sin\alpha\cos\beta & \sin\alpha\sin\beta \\ \cos\gamma\sin\alpha + \sin\gamma\cos\alpha\cos\beta & -\sin\gamma\sin\alpha + \cos\gamma\cos\alpha\cos\beta & -\cos\alpha\sin\beta \\ \sin\beta\sin\gamma & \sin\beta\cos\gamma & \cos\beta \end{pmatrix} \tag{1.6.58}$$

**Problem 1.6.20 Rotation of Cartesian coordinate system**. A point $P(x,y)$ in the $(x, y)$ plane is rotated counterclockwise by an angle $\theta$ to a new point $P'(x',y')$, while keeping the distance from $P$ to the origin $O$, the same as the distance from $O$ to $P'$, i.e. length $(OP)$ = length $(OP')$ (Figure 1.6). Express $(x', y')$ in terms of $(x, y)$ and the angle $\phi$, using the angle addition formulas Eqs. (1.5.4) and (1.5.5).

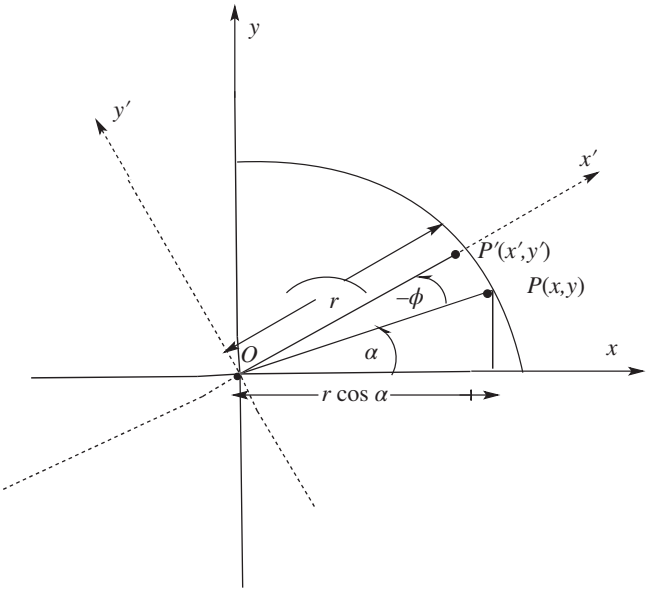



**Figure 1.6** Counterclockwise rotation of planar coordinate system by a negative angle $-\phi$ from $(x, y)$ to $(x', y')$ and of point $P$ to $P'$.

**Problem 1.6.21** Prove Eqs. (1.6.56), (1.6.57), and (1.6.58).

**Problem 1.6.22** If the Eulerian matrix $\boldsymbol{E}$ of Eq. (1.6.56) equals the rotation matrix $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{pmatrix} 0.875\,212\,771\,955 & -0.333\,162\,753\,797 & 0.350\,720\,950\,142 \\ -0.072\,043\,610\,084 & -0.806\,709\,369\,55 & -0.586\,540\,460\,086 \\ -0.478\,343\,311\,470 & -0.488\,080\,498\,03 & 0.730\,044\,590\,288 \end{pmatrix}$$

for particular Eulerian angles $\alpha, \beta$, and $\gamma$, please compute these Eulerian angles $\alpha, \beta$, and $\gamma$. This is very useful, e.g. when the principal axes of a physically measured tensor in a crystal must be oriented relative to laboratory Cartesian axes.

### 1.6.12 Covariant and Contravariant Vectors and Tensors

We define a quantity (vector or tensor) as **covariant** when it behaves like a coordinate system or framework or basis (Cartesian, cylindrical, spherical polar, etc.). In contrast, a quantity (vector or tensor) is called **contravariant** when it behaves like the *contents* of a coordinate system (e.g. position vector, velocity vector, etc.). Covariant vectors are $1 \times n$ row vectors; contravariant vectors are $n \times 1$ column vectors.

When we rotate a **contravariant** $n \times 1$ **column vector** (e.g. for position, velocity, momentum, electric field, etc.) we **pre**-multiply it by an $n \times n$ rotation tensor $\boldsymbol{R}$.

When, instead, we transform the coordinate system in which such vectors are defined, then the coordinate system, or basis, and also, e.g. the $\nabla$ operator, are ***covariant*** $1 \times n$ **row vectors**, that are transformed by the tensor $\boldsymbol{R}^{-1}$ that is the reciprocal of $\boldsymbol{R}$. A "dot product" or inner product $\boldsymbol{a} \cdot \boldsymbol{b}$ must be the multiplication of a row vector $\boldsymbol{a}$ by a column vector $\boldsymbol{b}$, to give a single number (scalar) as the result.

### 1.6.13 Covariant

We first present the transformation laws for **covariant** quantities. We want to transform an "old" set of quantities to a "new" set of quantities, due to a transformation (typically, rotation). Let the old unit cell be represented by the $1 \times 3$ row vector $\boldsymbol{V}_o \equiv (a_o b_o c_o)$, and the new unit cell be represented by the $1 \times 3$ row vector $\boldsymbol{V}_n \equiv (a_n b_n c_n)$. Then there exists a $3 \times 3$ transformation matrix $\boldsymbol{P}$ such that:

$$\boldsymbol{V}_\mathrm{n} \equiv (a_\mathrm{n} b_\mathrm{n} c_\mathrm{n}) = \boldsymbol{V}_\mathrm{o}\boldsymbol{P}(a_\mathrm{o} b_\mathrm{o} c_\mathrm{o}) \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \tag{1.6.59}$$

The inverse transformation from "new" to "old" is given by the matrix $\boldsymbol{Q}$ inverse to $\boldsymbol{P}$: $\boldsymbol{Q} \equiv \boldsymbol{P}^{-1}$, defined so that $\boldsymbol{PQ} = \boldsymbol{I}$:

$$\boldsymbol{V}_\mathrm{o} \equiv (a_\mathrm{o} b_\mathrm{o} c_\mathrm{o}) = \boldsymbol{V}_\mathrm{o}\boldsymbol{Q} = (a_\mathrm{n} b_\mathrm{n} c_\mathrm{n}) \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix} \tag{1.6.60}$$

**Contravariant**: Next, we define the transformation laws for **contravariant** quantities:

1) column (position) vectors $\boldsymbol{U} = u\boldsymbol{a} + v\boldsymbol{b} + w\boldsymbol{c} = [uvw]$ (e.g. atom positions, or "zone axis" $[uvw]$) which are taken from old $\boldsymbol{U}_\mathrm{o}$ to new $\boldsymbol{U}_\mathrm{n}$;
2) the translation vector $\boldsymbol{S}$ that takes us from an old origin $O_\mathrm{o}$ to a new origin $O_\mathrm{n}$.

We collect a column vector representation of these two contravariant quantities:

$$\boldsymbol{U}_\mathrm{o} \equiv \begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix}, S_\mathrm{o} \equiv \begin{pmatrix} x \\ y \\ z \end{pmatrix}, R_\mathrm{o} \equiv \begin{pmatrix} a_0^* \\ b_0^* \\ c_0^* \end{pmatrix} \tag{1.6.61}$$

It turns out that all these contravariant quantities use the inverse transformation $\boldsymbol{Q}$ defined in Eq. (1.6.59) above:

$$\boldsymbol{U}_\mathrm{n} \equiv \begin{pmatrix} u_n \\ v_n \\ w_n \end{pmatrix} = \boldsymbol{Q}\boldsymbol{V}_\mathrm{o} = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix} \begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix} \tag{1.6.62}$$

The inverse transformation of contravariant quantities will use the matrix $\boldsymbol{P}$:

$$\boldsymbol{U}_\mathrm{o} \equiv \begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix} = \boldsymbol{P}\boldsymbol{V}_\mathrm{n} = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \begin{pmatrix} u_n \\ v_n \\ w_n \end{pmatrix} \tag{1.6.63}$$

**Spherical trigonometry** is discussed in Special Topic ST1.6.

### 1.6.14 Moment of Inertia and Angular Momentum

We next discuss momentum, rotation, torque, moment of inertia, and angular momentum. A body with velocity $\boldsymbol{v}$ and mass $m$ is defined to have (linear) momentum $\boldsymbol{p}$:

$$\boldsymbol{p} = m\boldsymbol{v} \tag{1.6.64}$$

For a rigid body with differential of mass d$m$ (kg), the total mass $m$ is given by integration:

$$m = \int \mathrm{d}m \tag{1.6.65}$$

and the **moment of inertia** $I$, or second moment of the mass (units $[M][L]^2$), is given by the rank-two tensor $\boldsymbol{I}$:

$$\boldsymbol{I} = \int \boldsymbol{r}\boldsymbol{r}\ \mathrm{d}m \tag{1.6.66}$$

which for discrete masses is usually written as

$$\boldsymbol{I} = \sum_i \boldsymbol{r}_i \boldsymbol{r}_i m_i \tag{1.6.67}$$

This moment of inertia is essential for the analysis of rotational spectra of molecules. For anisotropic solids or for molecules, the moment of inertia $\boldsymbol{I}$ is a second-rank tensor, with three principal-axis components $I_1$, $I_2$, and $I_3$. This moment of inertia is important when the body rotates with angular frequency $\omega$ radians per second ($\omega$ Hz), or with $\nu$ revolutions per second.

The **angular momentum** of a body of mass $m$ and momentum $\boldsymbol{p}$ is the pseudovector $\boldsymbol{L}$:

$$\boldsymbol{L} \equiv \boldsymbol{r} \times \boldsymbol{p} = m\boldsymbol{r} \times \boldsymbol{v} \tag{1.6.68}$$

with units $[M][L]^2[T]^{-1}$. It is obvious from Eq. (1.6.68) why the nineteenth-century scientists also called angular momentum the (first) "moment of momentum." If this body rotates with constant angular frequency $\omega$ radians per second ($\omega$ Hz), then the tangential velocity $\boldsymbol{v}$ in Eq. (1.6.68) is used to define $\boldsymbol{\omega}$ as a pseudovector oriented along $\boldsymbol{L}$:

$$\boldsymbol{\omega} \equiv \boldsymbol{r} \times \boldsymbol{v} r^{-2} \tag{1.6.69}$$

The kinetic energy $E = (1/2)mv^2$ for linear motion corresponds elegantly to $E = (1/2)I\omega^2$ for angular motion. Using the tangential velocity $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{r}$, the angular momentum $\boldsymbol{L}$ can be rewritten as

$$\boldsymbol{L} = m\boldsymbol{\omega}r^2 = I\boldsymbol{\omega} \tag{1.6.70}$$

If a body with angular momentum $\boldsymbol{L}$ is subjected to a force $\boldsymbol{F}$ perpendicular to it, then it experiences a **torque** or turning force $\boldsymbol{T}$:

$$\boldsymbol{T} \equiv \boldsymbol{r} \times \boldsymbol{F} = \mathrm{d}\boldsymbol{L}/\mathrm{d}t \tag{1.6.71}$$

This torque, or moment of force (units $[M][L]^2[T]^{-2}$), will turn the orientation of the body in space, without changing the magnitude of its angular momentum. All gyroscopes maintain (by electrical means) their angular momentum and orientation $\boldsymbol{L}$ in any gravitational field $\boldsymbol{F}$, pointing into the same direction in inertial space, and are used as "inertial guidance systems." If mass of a gyroscope (Figure 1.7) is large enough, it can be used in large ships and ocean liners as a stabilizer, to decrease the pitch, yaw, and roll of the ship in heavy seas. The concept of torque is also important in describing the changes in orientation of electrons or nuclei in magnetic fields for electron spin resonance and nuclear magnetic resonance.

The circular analog to Newton's second law:

$$\boldsymbol{F} = \mathrm{d}\boldsymbol{p}/\mathrm{d}t = \mathrm{d}(m\boldsymbol{v})/\mathrm{d}t = m\boldsymbol{a} \quad ((1.6.1))$$

can now be written for angular motion:

$$\boldsymbol{T} = \mathrm{d}\boldsymbol{L}/\mathrm{d}t = \mathrm{d}(\boldsymbol{I\omega})/\mathrm{d}t = \boldsymbol{I\alpha} \tag{1.6.72}$$

where $\boldsymbol{\alpha}$ is the angular acceleration:

$$\boldsymbol{\alpha} = \mathrm{d}\boldsymbol{\omega}/\mathrm{d}t \tag{1.6.73}$$

**Problem 1.6.23** A ship with deadweight 50 000 metric tons is stabilized by a gyroscope of mass 5.1 tons, a diameter of 1.8 m, rotating at 1800 Hz. Calculate its moment of inertia and kinetic energy.

**Problem 1.6.24** Show from Eq. (1.6.64) that $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{r}$.

### 1.6.15 Derivative Operators: "Ski Slopes," "Hernias," and "Curls"

In Cartesian space, the "del" (or "nabla" ≡ "Assyrian harp" ≡ "atled" ≡ "backwards delta") operator or function $\nabla$ is given by

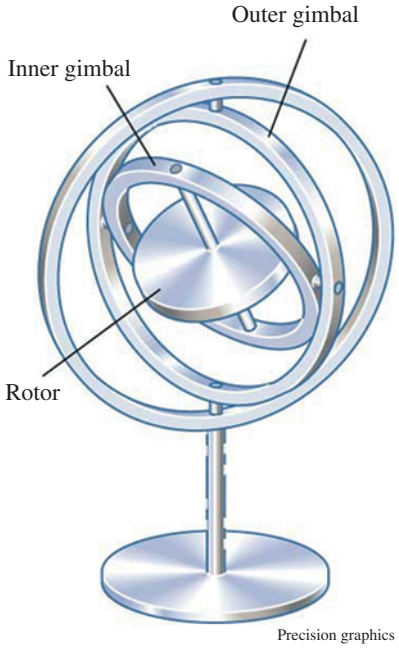



**Figure 1.7** The gyroscope, or universal or Cardano's suspension at the intersection of each set of circles, six mechanical bearings allow complete freedom of rotation about three axes. The rotor's motion is maintained electrically to have constant angular velocity; the rotor will maintain its spin orientation is unchanging in "inertial space," i.e. with respect to the "fixed stars."

$$\nabla \equiv \boldsymbol{e_x}(\partial/\partial x) + \boldsymbol{e_y}(\partial/\partial y) + \boldsymbol{e_z}(\partial/\partial z) \tag{1.6.74}$$

The "del" notation was introduced by Hamilton.[86]

### 1.6.16 Gradient

The **grad** $g \equiv \nabla g$ could also be called the "ski-slope" function: it shows the vector sum of the partial derivatives along three Cartesian axes for a scalar function $g(x,y,z)$ and has the value of a vector:

$$\textbf{grad}\, g(x,y,z) \equiv \nabla g \equiv \boldsymbol{e_x}(\partial g/\partial x) + \boldsymbol{e_y}(\partial g/\partial y) + \boldsymbol{e_z}(\partial g/\partial z) \tag{1.6.75}$$

when applied to a mountain, $\nabla g$ gives the direction of steepest descent for the most ardent and risk-averse skier!

### 1.6.17 Divergence

We could call the div $\boldsymbol{u} = \nabla \boldsymbol{u}$ function the "hernia function": it shows how a vector function $\boldsymbol{u}(x,y,z)$ herniates, i.e. sprouts in the $x$, $y$, and $z$ directions at the point $(x, y, z)$; in Cartesian space, the div or "del dot" operator can be represented by

$$\text{div}\, \boldsymbol{u} \equiv \nabla \cdot \boldsymbol{u} \equiv (\partial u_x/\partial x) + \left(\partial u_y/\partial y\right) + (\partial u_z/\partial z) \tag{1.6.76}$$

### 1.6.18 Curl

The **curl** $\boldsymbol{u}$, or $\nabla \times \boldsymbol{u}$, or **rot** $\boldsymbol{u}$, or "curly function" describes how "curly" a vector field (array) of vectors $\boldsymbol{u}$ is. Verify that in a Cartesian system the pseudovector:

$$\begin{aligned} \textbf{curl}\, \boldsymbol{u} \equiv \textbf{rot}\, \boldsymbol{u} \equiv \nabla \times \boldsymbol{u} \equiv & \begin{vmatrix} \boldsymbol{e_x} & \boldsymbol{e_y} & \boldsymbol{e_z} \\ \partial/\partial x & \partial/\partial x & \partial/\partial z \\ u_x & u_y & u_z \end{vmatrix} \\ = \boldsymbol{e_x}\left(\partial u_z/\partial y - \partial u_y/\partial z\right) & + \boldsymbol{e_y}(\partial u_x/\partial z - \partial u_z/\partial x) + \boldsymbol{e_z}\left(\partial u_y/\partial x - \partial u_x/\partial y\right) \end{aligned} \tag{1.6.77}$$

As defined above $\nabla \times \boldsymbol{u}$ is a pseudovector.

> **Sideline 1.13** Consider the face of a hairy monkey, or an ape-man: if it is taken to be spherical, it can be covered by facial hair everywhere, but the topology of putting hair all over a sphere requires that there be at least two whorls, or points on the sphere where the hair density must be zero, and the curl of the hair is maximized: this is known in topology as the Poincaré[87]–Hopf[88] theorem.

The Laplacian,[89] or del-squared, operator is $\nabla \cdot \nabla = \nabla^2$ (sometimes shown as $\Delta$ to further confuse the poor reader); $\nabla \cdot \nabla$ is given in Cartesian space as

$$\textbf{div}\ \ \textbf{grad} = \nabla \cdot \nabla = \nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2 \tag{1.6.78}$$

**Problem 1.6.25** Verify the following identity, valid for any vector $\boldsymbol{v}$:

$$\nabla \times (\nabla \times \boldsymbol{v}) = \nabla(\nabla \cdot \boldsymbol{v}) - (\nabla \cdot \nabla)\boldsymbol{v} = \nabla(\nabla \cdot \boldsymbol{v}) - \nabla^2 \boldsymbol{v} \tag{1.6.79}$$

86 Sir William Rowan Hamilton (1805–1865).
87 Henri Poincaré (1854–1912).
88 Heinz Hopf (1894–1971).
89 Pierre-Simon, Marquis de Laplace (1749–1827).

or, in a different notation:

$$\text{curl curl}\,\boldsymbol{v} = \text{grad}(\text{div}\,\boldsymbol{v}) - (\text{div grad})\boldsymbol{v} \tag{1.6.80}$$

**Problem 1.6.26** Calculate the volume of a tetrahedron of sides $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ ($|a| = |b| = |c|$).

**Problem 1.6.27** Prove that $\nabla \cdot \nabla \times \boldsymbol{U} = 0$ for any vector $\boldsymbol{U}$.

**Problem 1.6.28** The transformation from spherical polar $(r, \theta, \varphi)$ to Cartesian $(x, y, z)$ coordinates is

$$x = r\sin\theta\cos\varphi$$
$$y = r\sin\theta\sin\varphi$$
$$z = r\cos\theta$$

where the ranges are symmetrical in Cartesian space: $(-\infty \le x < \infty)$, $(-\infty \le y < \infty)$, $(-\infty \le z < \infty)$, but not "symmetrical" in spherical polar coordinates: $(0 \le r \le \infty)$, $(0 \le \text{colatitude } \theta \le \pi)$, $(0 \le \text{longitude } \varphi \le \pi)$, then the reverse transformation is

$$r = +\left(x^2 + y^2 + z^2\right)^{1/2}$$
$$\varphi = \tan^{-1}(y/x)$$
$$\theta = \cos^{-1}\left[z/\left(x^2 + y^2 + z^2\right)^{1/2}\right]$$

The Jacobian of this transformation is

$$dV = dx\,dy\,dz = r^2\sin\theta\;\,dr\,d\theta\,d\varphi$$

Now show that:

$$\begin{aligned}\nabla^2 &= \left(\partial^2/\partial x^2\right) + \left(\partial^2/\partial y^2\right) + \left(\partial^2/\partial z^2\right)\\ &= \left(\partial^2/\partial r^2\right) + (2/r)(\partial/\partial r) + \left(1/r^2\right)\left(\partial^2/\partial\theta^2\right) + \left(\cos\theta/r^2\sin\theta\right)(\partial/\partial\theta) + \left(1/r^2\sin^2\theta\right)\left(\partial^2/\partial\varphi^2\right)\end{aligned}. \tag{1.6.81}$$

(Every student should go through this three-page pain only once in his/her lifetime).

### 1.6.19 Gauss–Green–Stokes Theorem

Returning to a previous discussion, the theorem due to Gauss, Green, Stokes, and also Ostrogradsky,[90] Eq. (1.5.52), can now be rewritten more elegantly in vector notation: it links the volume integral of the divergence of the vector $\boldsymbol{A}$, namely $\textbf{div}\,\boldsymbol{A}$ or $\nabla \cdot \boldsymbol{A}$, with the two-dimensional surface integral of $\boldsymbol{A}$:

$$\oiint_S \boldsymbol{A}\cdot\boldsymbol{n}\,dS = \iiint_V \nabla\cdot\boldsymbol{A}\,dV \tag{1.6.82}$$

which means that the integral over a closed surface $S$ of a vector function $\boldsymbol{A}$ projected onto a unit normal $\boldsymbol{n}$ over that surface is equivalent to the volume integral of $\textbf{div}\,\boldsymbol{A}$ over the volume $V$ enclosed by $S$. The theorem means simply that the algebraic sum of all sources (minus sinks) within a volume equals the net outward flow from its surface.

**Problem 1.6.29** Prove $\nabla^2(1/r) = -4\pi\delta(\boldsymbol{r})$.

90 Mikhaiel Vasilievich Ostrogradsky (1801–1862).

**Example 1**
Consider the vector field $\boldsymbol{A} = 2x\boldsymbol{e_x} + y^2\boldsymbol{e_y} + z^2\boldsymbol{e_z}$ to be integrated over $S$ = the unit sphere ($x^2 + y^2 + z^2 = 1$); the direct computation of the surface integral $\int\int \boldsymbol{A}\cdot\boldsymbol{n}\,dS$ is difficult, but $\nabla\cdot\boldsymbol{A} = 2 + 2y + 2z$, and the volume integral can be split into $\int\int\int 2\,dV + \int\int\int 2x\,dV + \int\int\int 2z\,dV = 2(4/3)\pi + 0 + 0$ (since the $y$ and $z$ integrands change sign from $-y$ to $+y$, $-z$ to $+z$). Hence, $\int\int \boldsymbol{A}\cdot\boldsymbol{n}\,dS = (8/3)\pi$.

### 1.6.20 Combinatorics

Next are a few results from combinatorics, the science of ordering and sorting objects into "boxes" or "containers": these results will be useful for developing the Maxwell–Boltzmann (MB), Fermi–Dirac (FD), and Bose[91]–Einstein (BE) statistics in Chapter 4.

**Problem 1.6.30** Show that the number of ways of ordering $N$ distinguishable objects (=number of permutations of $N$ distinguishable objects) is given by

$$n_1 = N! \tag{1.6.83}$$

**Problem 1.6.31** Prove that the number of ways of placing $N$ distinguishable objects into $R$ distinguishable boxes, so that there are $N_1$ objects in box 1, $N_2$ objects in box 2, etc. is given by

$$n_2 = \frac{N!}{\prod_{i=1}^{i=R} N_i!} \tag{1.6.84}$$

**Problem 1.6.32** Prove that the number of ways of selecting $N$ distinguishable objects from a set of $G$ distinguishable objects (where $G > N$) is given by the binomial coefficient:

$$n_3 = \binom{G}{N} \equiv \frac{G!}{N!(G-N)!} \tag{1.6.85}$$

**Problem 1.6.33** Prove that the number of ways of placing $N$ **indistinguishable** objects into $G$ distinguishable containers is given by

$$n_4 = \frac{(G+N-1)!}{(G-1)!N!} \tag{1.6.86}$$

**Problem 1.6.34** Show that the number of ways of placing $N$ **distinguishable** objects into $G$ distinguishable containers are given by

$$n_5 = \frac{(G+N-1)!}{(G-1)!} \tag{1.6.87}$$

91 Satyendra Nath Bose (1894–1974).

**Problem 1.6.35** Stirling's[92] approximation for large $N$ is given by:

$$N! \approx (2\pi N)^{1/2} N^N \exp(-N) \tag{1.6.88}$$

which can be approximated adequately as

$$\ln N! \approx N \ln N - N \tag{1.6.89}$$

Calculate the percent error between $\ln N!$ and $N \ln N - N$ for $N = 50$, and for $N = 65$.

**Problem 1.6.36** Euler's **gamma function** $\Gamma(x)$ is given by either of two integrals:

$$\Gamma(x) \equiv \int_{t=0}^{t=\infty} t^{x-1} \exp(-t)\, dt = \int_{t=0}^{t=1} [\ln(1/t)]^{x-1} dt \tag{1.6.90}$$

valid for positive real $x$, or for complex $x$ with a positive real part.

(a) First show that:

$$\Gamma(x+1) = x\Gamma(x)$$

(b) Second, show that for integer positive $N$ the gamma function becomes a factorial:

$$\Gamma(N) = (N-1)! \tag{1.6.91}$$

(c) Third, show that:

$$\Gamma(1/2) = \pi^{1/2} \tag{1.6.91}$$

## 1.7 Hooke's Law, One-dimensional Chain, Dispersion Relations, and Stress–Strain Tensors

Hooke[93] suggested in 1660 "ut tensio, sic vis" [1.27], i.e. that a mechanical spring with spring constant $k_H$, if stretched **below** its elastic limit (from its resting length $\boldsymbol{r}_0$ to some distorted length $\boldsymbol{r}$), is subject to a restoring force $k_H(\boldsymbol{r} - \boldsymbol{r}_0)$: this leads to a potential energy:

$$U = (1/2) k_H |\boldsymbol{r} - \boldsymbol{r}_0|^2 \tag{1.7.1}$$

and to a restoring force:

$$\boldsymbol{F} = -k_H(\boldsymbol{r} - \boldsymbol{r}_0) \tag{1.7.2}$$

where $k_H$ is known as the **Hooke's law constant** (SI units: kg $s^{-2}$ or N $m^{-1}$). If a mass $m$ is attached to this spring (Figure 1.8), then, using Newton's second law:

$$\boldsymbol{F} = -k_H(\boldsymbol{r} - \boldsymbol{r}_0) = m d^2(\boldsymbol{r} - \boldsymbol{r}_0)/dt^2 \tag{1.7.3}$$

rewritten in terms of the amount stretched $y \equiv \boldsymbol{r} - \boldsymbol{r}_0$, yields:

$$-k_H y = m d^2 y/dt^2 \tag{1.7.4}$$

This classical harmonic oscillator equation explains the classical analog of quantum vibrational spectroscopy. After setting the initial condition $y = 0$ at $t = 0$, the differential equation can be integrated twice to yield:

$$y = A \sin\left[(k_H/m)^{1/2} t\right] = A \sin(\omega t) \tag{1.7.5}$$

92 James Stirling (1692–1770).
93 Robert Hooke (1635–1703).

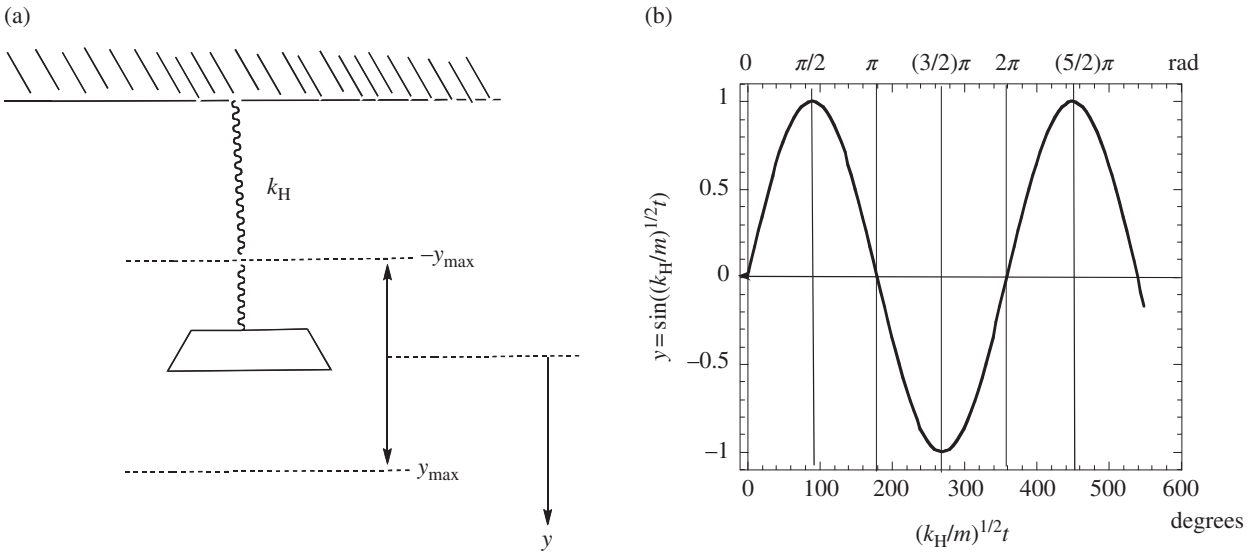



**Figure 1.8** (a) The motion of a mass $m$ suspended from an elastic spring of force constant $k_H$ satisfying Hooke's law. (b) Sinusoidal wave solution $y = \sin((k_H/m)^{1/2}t)$.

where the angular frequency $\omega$ (rad $s^{-1}$) is defined by

$$\omega \equiv (k_H/m)^{1/2} \tag{1.7.6}$$

and the angular frequency $\nu$ (in Hz = cycles per second) is

$$\nu \equiv (1/2\pi)\omega = (1/2\pi)(k_H/m)^{1/2} \tag{1.7.7}$$

The maximum amplitude $A$ depends on the elastic modulus of the spring, i.e. the spring can stretch only so far (say to $y_{max}$), so that it can still recover elastically by a backwards motion, that also obeys Hooke's law, then

$$y = y_{max} \sin\left[(k_H/m)^{1/2}t\right] \tag{1.7.8}$$

At the extrema of motion ($y = \pm y_{max}$), the velocity $v = (dy/dt)$ drops to zero, while the absolute value of the acceleration, as well as $|d^2y/dt^2|$ and the vibrational potential energy $U$ are at their maxima. At $y = 0$, the velocity and the kinetic energy are at maxima, while $U = 0$.

By a change of phase ($y = y_{max}$ at time $t = 0$), the solution to Eq. (1.7.4) can be expressed also by a cosine function of the type $y = A\cos(\omega t)$, or in general, the solution is the equation for classical simple harmonic motion:

$$y(t) = A\exp(i\omega t) + B\exp(-i\omega t) \tag{1.7.9}$$

where $i = (-1)^{1/2}$: this is the equation for **simple harmonic motion or a simple harmonic oscillator**, one of whose constants can be reset by choosing an appropriate initial condition, while the other, as before, depends on the physics, i.e. on the elastic modulus.

**Sideline 1.14** Newton declared (with false modesty?) that his achievements were possible because "he was walking on the shoulders of giants": his competitor, Hooke, was a hunchback! That same remark "*nani gigantium humeris insidentes*," was not original with Newton: it was first attributed to Bernard of Chartres (twelfth century).[94]

**Problem 1.7.1** Find the equation of motion for a pendulum of mass $m$ suspended by a rod of length $l$ from the ceiling, which at any instant of time makes an angle $\theta$ with the vertical. Show that for small angles, where $\sin\theta \approx \theta$, classical simple harmonic motion results. For a grandfather's clock ($l = 1$ m) determine its period.

94 Bernard de Chartres = Bernardus Carnotensis (died after 1124).

**Problem 1.7.2** A mechanical watch has a coil spring ("main spring") that controls the balance wheel, the escapement, and several gears. The spring is rewound periodically. The balance wheel and the main spring function together as a classical harmonic oscillator and provide a constant torque regardless of how well wound the spring is, until the spring is almost completely unwound. If the frequency is set to $\nu = 5$ Hz, compute the period; if the effective mass is 5 g, then what is the Hooke's law constant $k_H$?

We collect next some simple treatments of arrays of atoms or masses in one dimension: these will be quite useful in later analogies with the band structure of solids (Chapter 7).

### 1.7.1 Longitudinal Mechanical Waves Subjected to Hooke's Law

Consider longitudinal waves in a homogeneous line described by $x$, the position of a particular point on that line, and $u$, the longitudinal displacement of that point from its equilibrium position. Consider two points in the medium, at $x$ and $x + dx$. The change in $u$ at these points is $(\partial u/\partial x)_x$ which can be written as

$$(\partial u/\partial x)_{x+dx} \approx (\partial u/\partial x)_x + (\partial^2 u/\partial x^2)dx$$

Let the Hooke's law tension (or force constant) in the medium be $T$; $T$ is the same as $k_H$. Then the net force between $x$ and $x + dx$ is

$$F = T\left[(\partial u/\partial x)_{x+dx} - (\partial u/\partial x)_x\right] = T(\partial^2 u/\partial x^2)dx$$

This force, applied to a material of **mass per unit length** $\mu$, will, by Newton's second law, cause an acceleration $\mu(\partial^2 u/\partial t^2)$ $dx$. Finally, then, $T(\partial^2 u/\partial x^2) = \mu(\partial^2 u/\partial t^2)$ thus the displacement $u$ obeys a **one-dimensional mechanical wave equation**:

$$(\partial^2 u/\partial x^2) = v_o^{-2}(\partial^2 u/\partial t^2) \tag{1.7.10}$$

where the positive definite **phase velocity** is

$$v_o = (T/\mu)^{-1/2} \tag{1.7.11}$$

Equation (1.7.10) is the classical harmonic oscillator problem of Eq. (1.5.103), with solution:

$$u(x,t) = L\exp(ikx + i\omega t) + M\exp(ikx - i\omega t) \qquad ((1.5.112))$$

where $L$ and $M$ are wave amplitudes, and the angular frequency $\omega$ (rad $s^{-1}$), the frequency $\nu$ (Hz, or cycles $s^{-1}$), the wavevector $k$ ($m^{-1}$), and the wavelength $\lambda$ (m $cycle^{-1}$) are related by

$$v_o = \nu\lambda = \omega\lambda/2\pi = \omega/k \tag{1.7.12}$$

The detailed form of the solution depends on the boundary conditions. The wavevector $k$ can serve as the independent variable in the direct (space) domain, while $\omega$ is the independent variable in the time domain.

The **dispersion relation** $\omega(k)$ is the relation between the angular frequency $\omega$ time-domain behavior) and the wavevector $k$ (space-domain behavior). Dispersion means that waves of different angular frequencies $\omega$ can travel at different speeds $v$. In the present simple case, the dispersion relation is the linear relationship:

$$\omega = v_o k \tag{1.7.13}$$

where $v_o$ is the phase velocity of the wave.

**Problem 1.7.3** Construct the same argument, but for transverse waves on a string with tension $T$ and mass per unit length $\mu$: using Figure 1.9, show that $(\partial^2 y/\partial x^2) = (T/\mu)^{1/2}(\partial^2 y/\partial t^2)$.

### 1.7.2 Longitudinal Elastic Waves in a 1D Array of Equal Masses and Springs

Consider next the longitudinal motion of a one-dimensional array of $L$ equal masses (atoms?) $M$ (Figure 1.10). These masses at rest are equidistant, i.e. spaced $\boldsymbol{a}$ (m) apart, and can interact via Hooke's law with force constant $k_H$(N $m^{-1}$), but only with their nearest neighbors. Let $u_n$ be the longitudinal displacement of mass $n$ from its equilibrium position.

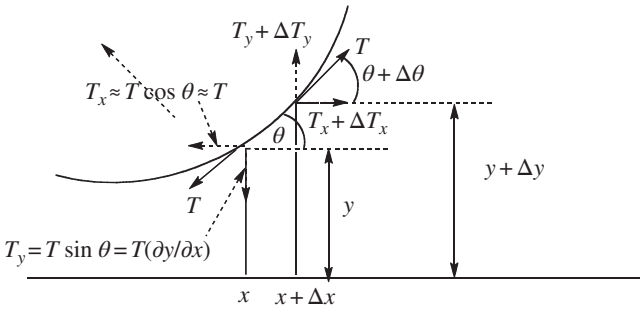



**Figure 1.9** Analysis of the forces for a transverse wave on a string.

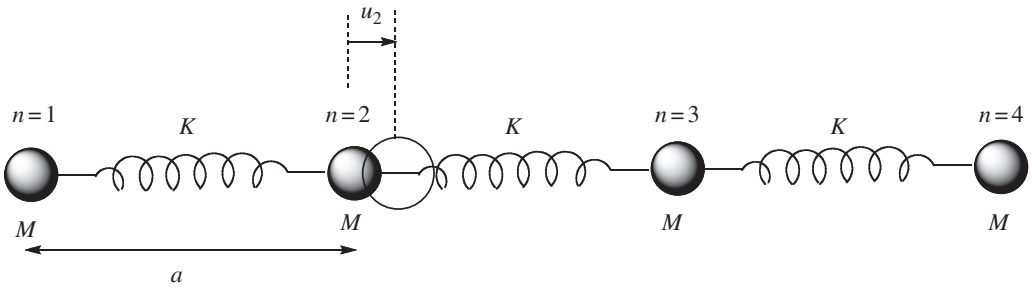



**Figure 1.10** System of four masses $M$ connected by three springs, each with Hooke's law constant $k_H$. The equilibrium distance between adjacent masses is $\boldsymbol{a}$. The second particle has been displaced to the right by $u_2$.

The net Hooke's law force on mass $n$, due to the displacements $u_n$, $u_{n-1}$, and $u_{n+1}$ is

$$F_n = k_H(u_{n+1} - u_n) - k_H(u_n - u_{n-1}) \quad (n = 1, 2, ..., L) \tag{1.7.14}$$

so, the net equation of motion becomes:

$$M\left(\partial^2 u_n/\partial t^2\right) = k_H(u_{n+1} + u_{n-1} - 2u_n) \tag{1.7.15}$$

We assume that the displacement of the last mass ($n = L$) is equal to that of the first mass ($n = 1$): this is the **Born**[95]**–von Kármán**[96] **periodic boundary condition**. We seek traveling or stationary wave solutions of the general form:

$$u_n = A\exp\left[\mathrm{i}(\omega t + kna)\right] \tag{1.7.16}$$

where $(na)$ is the variable used, in place of $x$, to denote the position along the chain. Note that $k$ is not the same as $k_H$. The substitution of Eq. (1.7.16) into Eq. (1.7.15) generates the condition:

$$\begin{aligned} -\omega^2 M &= k_H[\exp(\mathrm{i}ka) + \exp(-\mathrm{i}ka) - 2] \\ &= k_H[2\cos(ka) - 2] \\ &= 2k_H\left[\cos^2(ka/2) - \sin^2(ka/2) - 1\right] \\ &= -4k_H\sin^2(ka/2) \end{aligned} \tag{1.7.17}$$

which means that a solution of the type of Eq. (1.7.14) will exist **if and only if** $\omega$ satisfies the **dispersion relation:**

$$\omega = \pm\left(4k_H/M\right)^{1/2}\sin(ka/2) \tag{1.7.18}$$

95 Max Born (1882–1970).
96 Theodore von Kármán = Szöllöskislaki Kármán Tódor (1881–1963).

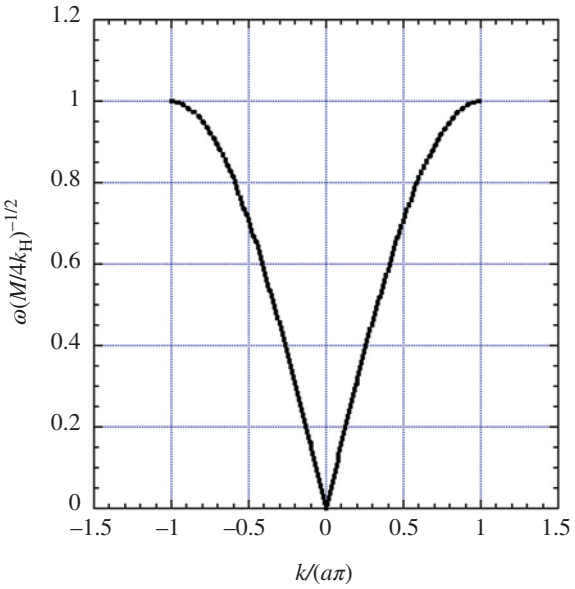



**Figure 1.11** Dispersion relation $\omega = (4k_H/M)^{1/2}\sin(ka/2)$ as a function of $k$, for a one-dimensional chain of masses $M$ linked by springs of Hooke's law constant $k_H$, Eq. (1.7.14). The maxima $\omega_{max} = (4k_H/M)^{1/2}$ are at $k_{max}a = \pm\pi$ rad.

This dispersion relation between the angular frequency $\omega$ and the wavevector $k$ shows a maximum $\omega_{max} = (4k_H/M)^{1/2}$ at the "magical" points $k = \pm\pi/a$. The negative solutions for $\omega$ are thrown away (see Figure 1.11).

The region inside $k = \pm(\pi/a)$ is called the first **Brillouin**[97] **zone** (the **Wigner–Seitz**[98] cell of the reciprocal lattice). Within that region the longitudinal wave can travel: substituting Eq. (1.7.18) into Eq. (1.7.14) yields:

$$u_n = A\exp\left[i\left((4k_H/M)^{1/2}\sin(kat/2) + kna\right)\right] \tag{1.7.19}$$

At the "band edge" or critical value:

$$k_{edge} \equiv \pm\pi/a \tag{1.7.20}$$

the solution of Eq. (1.7.14) is no longer a traveling wave, but is rather a standing wave (independent of $na$, i.e. independent of $x$):

$$u_n = A\exp\left[i(4k_H/M)^{1/2}t\right] \tag{1.7.21}$$

Equation (1.7.20) is very similar to Braggs'[99,100] law for diffraction [$n\lambda = 2d\sin\theta$]; this can be seen by rewriting Eq. (1.7.20) as

$$\lambda_{edge} \equiv 2\pi/k_{edge} = \pm 2a \tag{1.7.22}$$

The values of $k$ larger than $k_{edge}$ are usually "folded" back into the first Brillouin zone, since no new physics arises from the extended-zone scheme (see Chapter 7).

### 1.7.3 One-dimensional Chain with Two Kinds of Atoms: A Band Gap Appears

We discuss next the one-dimensional lattice with two kinds of atoms: atoms of mass $M$ occupy the even-numbered sites, and atoms of smaller mass $m$ occupy the odd-numbered sites. As before, there are $L$ atoms, and the equilibrium distances

97 Léon Brillouin (1889–1969).
98 Frederick Seitz (1911–2008).
99 Sir William Henry Bragg (1852–1942).
100 Sir William Lawrence Bragg (1890–1971).

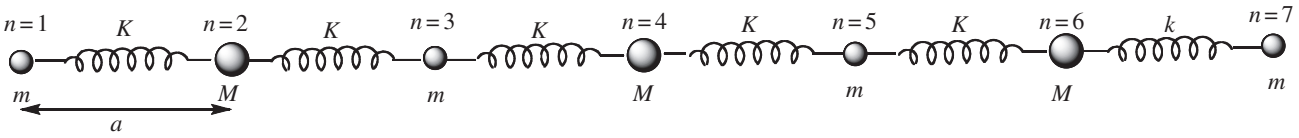



**Figure 1.12** System of four small masses $m$ and three large masses $M$, alternating on a one-dimensional chain, with intermass equilibrium distance $\boldsymbol{a}$. The even labels $2n$ refer to the larger masses, the odd labels to the smaller masses.

between adjacent atoms are equal to $\boldsymbol{a}$. Allow for a Hooke's law force with constant $k_{\mathrm{H}}$ to act between nearest neighbors **only**. Let $u_n$ be again the longitudinal displacement of atom $n$ from equilibrium (Figure 1.12).

We again assume Born–von Kármán periodic boundary conditions for the motion: the $N$th atom has displacement equal to that of the zeroth atom (closed loop). Then the two equations of motion are

$$M\left(\partial^2 u_{2n}/\partial t^2\right) = k_{\mathrm{H}}(u_{2n+1} + u_{2n-1} - 2u_{2n}) \tag{1.7.23}$$

$$m\left(\partial^2 u_{2n+1}/\partial t^2\right) = k_{\mathrm{H}}(u_{2n+2} + u_{2n} - 2u_{2n+1}) \tag{1.7.24}$$

As before, we seek periodic solutions of the type:

$$u_{2n}(t) = A\exp\left[\mathrm{i}\omega t + 2n\mathrm{i}ka\right] \quad \text{even masses} \tag{1.7.25}$$

$$u_{2n+1}(t) = B\exp\left[\mathrm{i}\omega t + (2n+1)\mathrm{i}ka\right] \quad \text{odd masses} \tag{1.7.26}$$

for the even and odd masses, respectively. Again, $k$ is not the same as $k_{\mathrm{H}}$. Thus, $u_{2n+2}(t)$ will resemble $u_{2n}(t)$, while $u_{2n-1}(t)$ will resemble $u_{2n+1}(t)$. When these trial solutions are used in Eqs. (1.7.13) and (1.7.14), and the common factor $\exp[\mathrm{i}\omega t + 2n\mathrm{i}ka]$ is discarded, two coupled homogeneous equations in the two unknowns $A$, $B$ are obtained:

$$0 = A\left[\omega^2 M - 2k_{\mathrm{H}}\right] + B[k_{\mathrm{H}}\exp(\mathrm{i}ka) + k_{\mathrm{H}}\exp(-\mathrm{i}ka)] \tag{1.7.27}$$

$$0 = A[k_{\mathrm{H}}\exp(\mathrm{i}ka) + k_{\mathrm{H}}\exp(-\mathrm{i}ka)] + B\left[\omega^2 m - 2k_{\mathrm{H}}\right] \tag{1.7.28}$$

These have a nontrivial solution if and only if the determinant of their coefficients vanishes:

$$\begin{vmatrix} \omega^2 M - 2k_{\mathrm{H}} + & 2k_{\mathrm{H}}\cos(ka) \\ 2k_{\mathrm{H}}\cos(ka) & \omega^2 m - 2k_{\mathrm{H}} \end{vmatrix} = 0 \tag{1.7.29}$$

This determinantal equation, when solved for $\omega^2$, yields the **dispersion relation**:

$$\frac{\omega^2}{k_H} = \left(\frac{1}{m} + \frac{1}{M}\right) \pm \sqrt{\left(\frac{1}{m} + \frac{1}{M}\right)^2 - \frac{4\sin^2(ka)}{mM}} \tag{1.7.30}$$

When this dispersion relation is plotted as $\omega$ vs. $k$ (Figure 1.13), one finds:

i) a lower-energy, or "acoustical" branch ($A = B$, atoms moving together whether even or odd in combined lower-frequency longitudinal motion);
ii) a higher-frequency, or "optical" branch (even-numbered atoms vibrating together; odd-numbered atoms vibrating together, but separately from the even-numbered atoms);
iii) for intermediate values of $\omega$ and $k$ not on these curves, the value of $k$ would be complex (**forbidden gap frequency**). Thus, one sees for the first time, in a classical problem, an energy gap of forbidden motions, which presages similar, frequent behavior in quantum problems. The arguments given above for longitudinal waves may be also made, *mutatis mutandis*, for transverse waves.

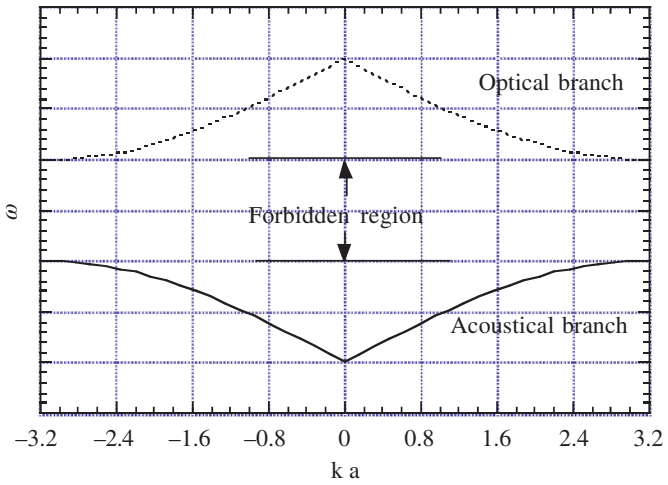



**Figure 1.13** Dispersion relation $\omega = k_H(1/m + 1/M)\{1 \pm [1 - 4\sin^2(ka)(mM)^{-1}(1/m + 1/M)^{-2}]^{1/2}\}$ for a classical chain of $N$ atoms of two kinds of masses, smaller $m$ and larger $M$, Eq. (1.7.30).

### 1.7.4 Stress and Strain Tensors

For a three-dimensional body, discussions of elastic responses in the framework of Hooke's law become more complicated. One defines a $3 \times 3$ **stress tensor** $\boldsymbol{P}$ [1.28], which is the force (units: N) expressed in a Cartesian coordinate system:

$$\boldsymbol{P} = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \tag{1.7.31}$$

In general, the off-diagonal elements of $\boldsymbol{P}$ can be nonzero, but are usually symmetrical, $P_{ij} = P_{ji}$, so that, of the nine terms of $\boldsymbol{P}$, only at most six are unique. If the force $\boldsymbol{P}$ is isotropic or has cubic symmetry, then only one term is unique, $P_{11} = P_{22} = P_{33}$, and the off-diagonal terms vanish. The stress on a body can be positive or tensile (extending the body) or negative and contractile (reducing its size).

As discussed in Section 6.7, crystals, particularly organic crystals, usually exist in lower-symmetry nonorthogonal systems: for them the off-diagonal terms of $\boldsymbol{P}$ become important. The response of a crystal to the stress tensor $\boldsymbol{P}$ is a series of fractional displacements, **small** compared to any dimension of the body; these fractional displacements are called strains and are denoted by the **strain** (or **dilatation**) **tensor** $\boldsymbol{s}$:

$$\boldsymbol{s} \equiv \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \tag{1.7.32}$$

The strain component $s_{12}$ is usually the deformation of the body along axis 1, due to a force along axis 2; the strain tensor $\boldsymbol{s}$ is usually symmetrical, $s_{ij} = s_{ji}$, and thus, of the nine terms of $\boldsymbol{s}$, at most six are unique. Both $\boldsymbol{P}$ and $\boldsymbol{s}$ can be represented as **ellipsoids** of stress and strain, respectively, and can be reduced to a diagonal form, e.g. $\boldsymbol{P}'$, along some preferred orthogonal system of axes, oblique to the laboratory frame or to the frame of the crystal, but characteristic for the solid: the transformation to this diagonal form is a **principal-axis transformation** of the type:

$$\boldsymbol{P}' = \boldsymbol{X}\boldsymbol{P}\boldsymbol{X}^{-1} \tag{1.7.33}$$

$$\boldsymbol{s}' = \boldsymbol{Z}\boldsymbol{s}\boldsymbol{Z}^{-1} \tag{1.7.34}$$

where $\boldsymbol{X}$ is the $3 \times 3$ transformation matrix and $\boldsymbol{X}^{-1}$ (or $\boldsymbol{Z}^{-1}$) is the inverse of $\boldsymbol{X}$ (or $\boldsymbol{Z}$):

$$\boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{Z}\boldsymbol{Z}^{-1} = \boldsymbol{E} \tag{1.7.35}$$

where $\boldsymbol{E}$ is the diagonal $3 \times 3$ unit matrix. In this principal-axis or diagonal system (usually identical for $\boldsymbol{P}$ and $\boldsymbol{s}$), we can talk about the six unique nonzero terms $P'_{11}, P'_{22}, P'_{33}, s'_{11}, s'_{22}$, and $s'_{33}$. The **cubic dilatation**, or change in volume per unit volume, $D$, is the trace, or sum of the diagonal terms of $\boldsymbol{s}'$:

$$D = s'_{11} + s'_{22} + s'_{33} \tag{1.7.36}$$

**Poisson's**[101] **ratio** $\sigma$ is the change in length per unit length, usually the contraction in one direction due to the dilatation in the perpendicular direction (for isotropic elastic bodies $-1.0 < \sigma < +0.5$).

**Young's**[102] **modulus** $\boldsymbol{Y}$, also called the **linear modulus of elasticity**, is the $3 \times 3$ tensor of the stress $\boldsymbol{P}$ divided by the strain $\boldsymbol{s}$:

$$Y_{ij} \equiv P_{ij}/s_{ij} \tag{1.7.37}$$

A body will obey Young's modulus only if it is stretched or compressed within its elastic limit; if this limit is exceeded, structural failure ensues. For a one-dimensional system, or for a cubic crystal, Young's modulus reduces to the Hooke's law constant $k_{\mathrm{H}}$:

$$Y_{11} = Y_{22} = Y_{33} = k_{\mathrm{H}} \tag{1.7.38}$$

The reciprocal of $k_{\mathrm{H}}$ is the scalar compressibility $\kappa$:

$$\kappa = 1/k_{\mathrm{H}} \tag{1.7.39}$$

The volume compressibility tensor $\boldsymbol{\kappa}$ is given by its nine terms ($i, j = 1, 2, 3$):

$$\kappa_{ij} = s_{ij}/P_{ij} \tag{1.7.40}$$

The shear modulus $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu} = \boldsymbol{Y}/2(1 + \boldsymbol{\sigma}) \tag{1.7.41}$$

## 1.8 Lagrange's Function and Hamilton's Function

We now introduce [1.26] the classical expressions for Lagrange's function $L$ and for Hamilton's function $H$. These two functions allow us to solve classical problems, by focusing on the energies of the problem, rather than on forces, and thus present certain conceptual advantages; the mathematical labor is the same: from $F = -m\,\mathrm{d}x^2/\mathrm{d}t^2$ it takes two integrations to obtain $x(t)$; from $F$ to either $L$ and $H$ involves one integration, but then one must do one more integration of $L$ or $H$ with respect to $t$, to obtain $x(t)$. $L$ and $H$ also become important when the Hamiltonian operator is developed in quantum mechanics (Section 2.1).

Let a system of $N$ particles of masses $m_i$ and Cartesian positions $(x_i, y_i, z_i) \equiv (r_{i1}, r_{i2}, r_{i3})$ have kinetic energy $T$, which is a function of only the velocities $(\mathrm{d}r_{ij}/\mathrm{d}t)$ [1.26]:

$$T(\mathrm{d}r_{ij}/\mathrm{d}t) = \frac{1}{2}\sum_{i=1}^{i=N} m_i \sum_{j=1}^{3} (\mathrm{d}r_{ij}/\mathrm{d}t)^2 \tag{1.8.1}$$

This system is called "conservative" or "holonomic," if and only if it satisfies the following two conditions:

i) it is subject to forces that are expressible as derivatives of a potential energy $U$,
ii) this potential energy $U$ does not depend explicitly either on time or on the speed of the particles (frictional systems, for instance, are not holonomic).

101 Siméon Denis Poisson (1781–1840).
102 Thomas Young (1773–1829).

In particular, let $U(r_{ij})$ depend only on the positions of the particles: for instance, for the harmonic oscillator in one dimension, $U = (1/2)k_{\mathrm{H}}x^2$. Then we can define a **Lagrangian function** $\boldsymbol{L}$ as $L \equiv T - U$:

$$L\left(r_{ij}, \mathrm{d}r_{ij}/\mathrm{d}t\right) \equiv T\left(\mathrm{d}r_{ij}/\mathrm{d}t\right) - U\left(r_{ij}\right) \quad (i = 1, 2, ..., N; j = 1, 2, 3) \tag{1.8.2}$$

This Lagrangian is a function of the positions and speeds of the particles. Newton's second-order differential equations of motion:

$$m_i\left(\mathrm{d}^2 r_{ij}/\mathrm{d}t^2\right) = -\partial U/\partial r_{ij} \tag{1.8.3}$$

can be now be replaced formally by Lagrange's second-order differential equations of motion:

$$(\mathrm{d}/\mathrm{d}t)\left(\partial L/\partial\left(\mathrm{d}r_{ij}/\mathrm{d}t\right)\right) = \partial L/\partial r_{ij} \quad (i = 1, 2, ..., N; j = 1, 2, 3) \tag{1.8.4}$$

Rather than deal directly with the speeds of the particles $\mathrm{d}r_{ij}/\mathrm{d}t$, it is useful to rewrite the kinetic energy $T$ in terms of their momenta $\boldsymbol{p}_i = m_i(\mathrm{d}\boldsymbol{r}_i/\mathrm{d}t)$, or their components $p_{ij}$:

$$T\left(p_{ij}\right) = \sum_{i=1}^{i=N} (1/2m_i) \sum_{j=1}^{3} p_{ij}^2 \tag{1.8.5}$$

**Hamilton's function** $\boldsymbol{H}$ is next defined as $H \equiv T + U$:

$$H\left(r_{ij}, p_{ij}\right) \equiv T\left(p_{ij}\right) + U\left(r_{ij}\right) \quad (i = 1, 2, ..., N; j = 1, 2, 3) \tag{1.8.6}$$

Then Hamilton's equations of motion are first-order differential equations:

$$\partial H/\partial p_{ij} = \mathrm{d}r_{ij}/\mathrm{d}t \quad (i = 1, 2, ..., N; j = 1, 2, 3) \tag{1.8.7}$$

$$\partial H/\partial r_{ij} = -\mathrm{d}p_{ij}/\mathrm{d}t \quad (i = 1, 2, ..., N; j = 1, 2, 3) \tag{1.8.8}$$

The mathematics is not easier overall, compared to using Newton's second law directly, but the emphasis is changed.

**Problem 1.8.1** Solve the simple harmonic motion problem by using (i) Lagrange's and (ii) Hamilton's equations of motion.

**PART THREE:** Electricity and Magnetism

## 1.9 Force Two: Electromagnetism

Using two popular systems of units, SI and the older cgs, we give first two fundamental equations, Coulomb's law of 1785 and Ampère's[103] law of 1826, in both systems.

### 1.9.1 Coulomb's Law and Ampère's Law

We restate **Coulomb's law** for electrostatics, Eq. (1.1.2), for the force $\boldsymbol{F}_{12}$ between two electrical charges $q_1$ and $q_2$ separated by a distance $\boldsymbol{r}_{12}$:

$$\boldsymbol{F}_{12} = q_1 q_2 \boldsymbol{r}_{12}/\left(4\pi\varepsilon_0 r_{12}^3\right) \quad (\mathrm{SI}); \quad \boldsymbol{F}_{12} = q_1 q_2 \boldsymbol{r}_{12} r_{12}^{-3} \quad (\text{cgs-esu}) \tag{1.9.1}$$

103 André-Marie Ampère (1775–1836).

**Ampère's law** (or the law of Biot[104] and Savart[105]) for the force between two electrical currents $\boldsymbol{j}_1$ and $\boldsymbol{j}_2$ is given by

$$\boldsymbol{F}_{12} = (\mu_0/4\pi)\int \mathrm{d}v_1 \int \mathrm{d}v_2 \boldsymbol{j}_1 \times \boldsymbol{j}_2 \times \boldsymbol{r}_{12} r_{12}^{-3} \quad \text{(SI)}; \quad \boldsymbol{F}_{12} = \int \mathrm{d}v_1 \int \mathrm{d}v_2 \boldsymbol{j}_1 \times \boldsymbol{j}_2 \times \boldsymbol{r}_{12} r_{12}^{-3} \quad \text{(cgs-emu)} \tag{1.9.2}$$

The older **cgs** definition considers the electrical charge (in cgs-esu) as a secondary quantity, statcoulombs, where 1 statC $\equiv 1$ g$^{1/2}$ cm$^{3/2}$ s$^{-1}$ (i.e. units of $[M]^{1/2}[L]^{3/2}[T]^{-1}$); the advantage of cgs units is that 1 statC$^2$ = 1 dyn. Equation (1.9.2) in cgs-emu form defines an electrical current, the abampere = 1 g$^{1/2}$ cm$^{1/2}$ s$^{-1}$ (i.e. units of $[M]^{1/2}[L]^{1/2}[T]^{-1}$). By convention, 1 abampere $\equiv$ 1 abA = 0.1 A; again, it is nice to see that 1 abA$^2 \equiv$ 1 dyn. The disadvantage of **cgs-esu** and **cgs-emu**, however, is that the electromagnetic unit of current, favored by magneticians, is not the same as the electrostatic unit: in detail, if $\rho_{\text{cgs - esu}}$ is the charge density and $\boldsymbol{u}$ is the velocity with which this charge density moves, and $c$ is the speed of light, then:

$$1.0\boldsymbol{j}_{\text{cgs-emu}} = 1.0\rho_{\text{cgs-esu}}\boldsymbol{u}/c \tag{1.9.3}$$

In **SI** this difficulty is avoided by making the **electrical charge** $[Q]$ **a new, fourth fundamental quantity**, of equal importance as kg for $[M]$, m for $[L]$, or s for $[T]$; SI uses the practical units of Coulomb for electrical charge, A $\equiv$ 1 C s$^{-1}$ for electrical current, H[106] for magnetic inductance, Wb[107] for magnetic flux ($\equiv$1 V s), T[108] $\equiv$W m$^{-2}$ for magnetic induction, C for capacitance, $\Omega$[109] for electrical resistance, S[110] for electrical conductance, etc. To secure this practical advantage, however, SI must introduce two convenient, if physically meaningless constants: $\varepsilon_0$, the "electrical permittivity of vacuum," defined as

$$\varepsilon_0 \equiv 10^7/\left(4\pi c^2\right) = 8.854\,187\,817 \times 10^{-12} \text{ F m}^{-1} \quad \text{(SI)} \tag{1.9.4}$$

where $c$ is the speed of light in vacuum, and $\mu_0$ is the "magnetic permittivity of vacuum," defined as

$$\mu_0 \equiv 4\pi \times 10^{-7} = 1.256\,637\,061\,432 \times 10^{-6} \text{ H m}^{-1} \quad \text{(SI)} \tag{1.9.5}$$

The product of $\varepsilon_0$ and $\mu_0$ is, however, physically meaningful:

$$\varepsilon_0\mu_0 = c^{-2} \quad \text{(SI)} \tag{1.9.6}$$

A minor variant of the SI system, called MKS by experimentalists in nonlinear optics, but called SI′ here, is defined below. Alas, there are also other systems: "rationalized" or "unrationalized," Heaviside–Lorentz, "atomic" (where $m_e = c = \hbar = 1$), etc. The "rationalized" and "unrationalized" versions differ in how they apportion the pesky factor $4\pi$ (surface area of a sphere with unit radius that is involved in surface integrals) between the various electrical and the magnetic variables. The details are given in Special Topic ST1.7.

The **Coulomb force** is mediated by photons, real or virtual. Following Faraday, we introduce the artificial concept of the **electric field** $\boldsymbol{E}$ due to an electric charge distribution $\rho(\boldsymbol{r})$:

$$\boldsymbol{E} = (1/4\pi\varepsilon_0)\int \mathrm{d}v(\boldsymbol{r})\rho(\boldsymbol{r})\boldsymbol{r}r^{-3} \quad \text{(SI)}; \quad \boldsymbol{E} = \int \mathrm{d}v(\boldsymbol{r})\rho(\boldsymbol{r})\boldsymbol{r}r^{-3} \quad \text{(cgs-esu)} \tag{1.9.7}$$

For a single charge $q$ at a distance $\boldsymbol{r}$ from the point of observation, Eq. (1.9.7) reduces to:

$$\boldsymbol{E} = q\boldsymbol{r}/\left(4\pi\varepsilon_0 r^3\right) \quad \text{(SI)}; \quad \boldsymbol{E} = q\boldsymbol{r}r^{-3} \quad \text{(cgs-esu)} \tag{1.9.8}$$

104 Jean-Baptiste Biot (1774–1862).
105 Félix Savart (1791–1841).
106 Joseph Henry (1797–1878).
107 Wilhelm Eduard Weber (1804–1891).
108 Nikola Tesla (1856–1943).
109 Georg Simon Ohm (1789–1854).
110 Ernst Werner von Siemens (1816–1892).

Since magnetic monopoles have never been found, the irreducible magnetic entity must be the magnetic dipole. In **Ampère's law**, we can define the magnetic flux density or magnetic induction $\boldsymbol{B}$, due, or induced, by an electrical current $\boldsymbol{j}$ at a distance $\boldsymbol{r}$, as:

$$\boldsymbol{B} = (\mu_0/4\pi)\int \mathrm{d}v(\boldsymbol{r})\boldsymbol{j}\times\boldsymbol{r}r^{-3} \quad \text{(SI)}; \quad \boldsymbol{B} = (1/c)\int \mathrm{d}v(\boldsymbol{r})\boldsymbol{j}\times\boldsymbol{r}r^{-3} \quad \text{(cgs-emu)} \tag{1.9.9}$$

This result explains electrical motors.

Thus, the fundamental fields are $\boldsymbol{E}$ and $\boldsymbol{B}$. In an older school of thought, the fundamental magnetic field was taken as $\boldsymbol{H}$ because of a symmetry between $\boldsymbol{E}$ and $\boldsymbol{H}$ explained below. We next see how $\boldsymbol{E}$ and $\boldsymbol{B}$ are modified inside matter: this will generate the "constitutive equations."

### 1.9.2 Material Media and Their Reaction to External Fields

In a material medium, a charge distribution can induce some charge separations, or dipoles, which help to minimize the total energy. Similarly, an external magnetic field will induce some magnetic dipoles in the medium to counteract this field. To handle these effects, an **electric polarization** (or electrical dipole moment per unit volume) $\boldsymbol{P}$ and a **magnetization** (or magnetic dipole moment per unit volume) $\boldsymbol{M}$ are defined. If the medium is linear and isotropic, these two new vectors $\boldsymbol{P}$ and $\boldsymbol{M}$ are proportional to $\boldsymbol{E}$ and $\boldsymbol{H}$, respectively

$$\boldsymbol{P} = \varepsilon_0\chi\boldsymbol{E} \quad \text{(SI)}; \quad \boldsymbol{P} = \chi\boldsymbol{E} \quad \text{(cgs-esu)} \tag{1.9.10}$$

$$\boldsymbol{M} = \chi_\mathrm{m}\boldsymbol{H} \quad \text{(SI)}; \quad \boldsymbol{M} = \chi_\mathrm{m}\boldsymbol{H} \quad \text{(cgs-emu)} \tag{1.9.11}$$

where $\chi$ is the $3\times3$ **linear volume electric susceptibility tensor**, and $\chi_\mathrm{m}$ is the **linear volume magnetic susceptibility tensor**. If the medium is isotropic, then both $\chi$ and $\chi_\mathrm{m}$ become scalars; if one divides $\chi$ and $\chi_\mathrm{m}$ by the mass, molar mass, or density of the medium, one gets the corresponding mass, molar, or specific susceptibilities, respectively. If the electric field is very high, there are nonzero higher-order (nonlinear) electric susceptibilities. It is further convenient to define a $3\times3$ linear **dielectric constant (relative electrical permittivity) tensor** $\boldsymbol{\varepsilon}$, and a **dielectric displacement** vector $\boldsymbol{D}$ (as a "net field" due to formal charges and to induced dipoles) in the **first constitutive equation**:

$$\begin{aligned}&\boldsymbol{D} = \varepsilon_0\boldsymbol{\varepsilon}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \varepsilon_0\chi_\mathrm{SI}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P} \quad \text{(SI)}; \quad \boldsymbol{D} = \boldsymbol{\varepsilon}\boldsymbol{E} = \boldsymbol{E} + 4\pi\boldsymbol{P} = \boldsymbol{E} + 4\pi\chi_\mathrm{esu}\boldsymbol{E} \quad \text{(cgs-esu)};\\ &\boldsymbol{D} = \varepsilon_0\boldsymbol{\varepsilon}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \chi_{\mathrm{SI}'}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P} \quad (\mathrm{SI}')\end{aligned} \tag{1.9.12}$$

This introduces SI′, a minor variant of SI, that is conventionally (and confusingly) called MKS, and is used in nonlinear optics.

The magnetic case is similar: using $\boldsymbol{\mu}$, the $3\times3$ linear **relative magnetic permeability** tensor, the magnetic induction $\boldsymbol{B}$ is the appropriately scaled sum of the magnetic field intensity or magnetostatic field $\boldsymbol{H}$ and the magnetization $\boldsymbol{M}$, in the **second constitutive equation**:

$$\boldsymbol{B} = \mu_0\boldsymbol{\mu}\boldsymbol{H} = \mu_0(1+\chi_\mathrm{m})\boldsymbol{H} = \mu_0(\boldsymbol{H}+\boldsymbol{M}) \quad \text{(SI)}; \quad \boldsymbol{B} = \boldsymbol{\mu}\boldsymbol{H} = \boldsymbol{H} + 4\pi\boldsymbol{M} \quad \text{(cgs-emu)} \tag{1.9.13}$$

The magnetic field $\boldsymbol{H}$ is given explicitly by

$$\boldsymbol{H} = \mu_0^{-1}\boldsymbol{B} - \boldsymbol{M} = \mu_0^{-1}\boldsymbol{\mu}^{-1}\boldsymbol{B} \quad \text{(SI)}; \quad \boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M} = \boldsymbol{H}/(1+\chi_\mathrm{m}) \quad \text{(cgs)} \tag{1.9.14}$$

The SI definitions of $\boldsymbol{D} = \boldsymbol{\varepsilon}_0\boldsymbol{\varepsilon}\boldsymbol{E}$ and $\boldsymbol{B} = \mu_0\boldsymbol{\mu}\boldsymbol{H}$ do seem cumbersome but have the advantage that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ have the same magnitudes in SI and in cgs and become $\boldsymbol{\varepsilon} = 1$ or $\boldsymbol{\mu}$ in vacuum. Many texts use the simpler-looking SI definitions $\boldsymbol{D} = \boldsymbol{\varepsilon}\boldsymbol{E}$ and $\boldsymbol{B} = \boldsymbol{\mu}\boldsymbol{H}$.

In material media, we add the empirical **Ohm's law**, a **third constitutive equation**:

$$\boldsymbol{J} = \boldsymbol{\sigma}\times\boldsymbol{E} \quad \text{(SI)}; \quad \boldsymbol{J} = \boldsymbol{\sigma}\times\boldsymbol{E} \quad \text{(cgs-esu)} \tag{1.9.15}$$

where $\boldsymbol{\sigma}$ is the $3\times3$ linear electrical conductivity tensor of the medium (SI: S m$^{-1}$); $\boldsymbol{J}$ is the electrical current; $\boldsymbol{\sigma}$ is the reciprocal of the electrical resistivity tensor $\boldsymbol{\rho}$ (SI: Ω m; esu:abΩ cm).

As will discussed in detail in Section 7.1, Ohm's law, obeyed by most metals, is valid only if the conductivity is limited by scattering processes (electrons or holes scattering off inclusions, lattice defects, or phonons). Semiconductors typically have a nonlinear dependence of the current on the applied electric field because the number of carriers depends on temperature and on reaching or exceeding an Arrhenius-type activation energy ("nonohmic behavior"). In vacuum, or within a single atom or a single molecule, conductivity occurs only by quantum-mechanical tunneling over very short distances (0.1 to about 5 nm): this tunneling is not linear with voltage. A quantum of conductivity can be defined.

### 1.9.3 Maxwell's Equations

Combining all earlier experimental results, in 1865, Maxwell obtained the following four fundamental equations for electromagnetism for media at rest [1.5]:

$$\nabla \cdot \boldsymbol{D} = \rho \quad \text{(SI)}; \quad \nabla \cdot \boldsymbol{D} = 4\pi\rho \quad \text{(cgs)} \tag{1.9.16}$$

$$\nabla \cdot \boldsymbol{B} = 0 \quad \text{(SI)}; \quad \nabla \cdot \boldsymbol{B} = 0 \quad \text{(cgs)} \tag{1.9.17}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \partial \boldsymbol{D}/\partial t \quad \text{(SI)}; \quad \nabla \times \boldsymbol{H} = (4\pi/c)\boldsymbol{J} + (1/c)\partial \boldsymbol{D}/\partial t \quad \text{(cgs)} \tag{1.9.18}$$

$$\nabla \times \boldsymbol{E} = -\partial \boldsymbol{B}/\partial t \quad \text{(SI)}; \quad \nabla \times \boldsymbol{E} = -(1/c)\partial \boldsymbol{B}/\partial t \quad \text{(cgs)} \tag{1.9.19}$$

Of these, the first is Gauss's law, the second could be called Gauss's law for magnetic fields, the third is the generalized form of Ampère's law, and the fourth is Faraday's law of induction. Since electrical charges do exist, while magnetic monopoles have never been found; therefore, the source density is nonzero on the right side of Eq. (1.9.16), but must be zero in Eq. (1.9.17). The integral forms of these four Maxwell equations are

$$\int\int \boldsymbol{D} \cdot \mathrm{d}\boldsymbol{S} = q \quad \text{(SI)}; \quad \int\int \boldsymbol{D} \cdot \mathrm{d}\boldsymbol{S} = 4\pi q \quad \text{(cgs)} \tag{1.9.20}$$

$$\int\int \boldsymbol{B} \cdot d\boldsymbol{S} = 0 \quad \text{(SI)}; \quad \int\int \boldsymbol{B} \cdot d\boldsymbol{S} = 0 \quad \text{(cgs)} \tag{1.9.21}$$

$$\int \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} = \mu_0 J + \varepsilon_0 \mu_0 \mathrm{d}\Phi_\mathrm{E}/\mathrm{d}t \quad \text{(SI)}; \quad \int \boldsymbol{B} \bullet \mathrm{d}\boldsymbol{l} = 4\pi c^{-1} J + c^{-1} \mathrm{d}\Phi_\mathrm{E}/\mathrm{d}t \quad \text{(cgs)} \tag{1.9.22}$$

$$\int \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\mathrm{d}\Phi_\mathrm{B}/\mathrm{d}t \quad \text{(SI)}; \quad \int \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -c^{-1}\mathrm{d}\Phi_\mathrm{B}/\mathrm{d}t \quad \text{(cgs)} \tag{1.9.23}$$

where $\mathrm{d}\boldsymbol{S}$ is the surface element of a closed surface (and the double integrals are over the whole closed surface), $\mathrm{d}\boldsymbol{l}$ is a line element (and the integrals are over a closed loop), $q$ is the total electrical charge enclosed by the surface, $\Phi_\mathrm{E}$ is the electric flux, $\Phi_\mathrm{B}$ is the magnetic flux, and $J$ is the electrical current.

> **Sideline 1.15** Maxwell did not use vector notation, gradients, and curls, but published eight relevant equations in 1865 using quaternions; these later expanded to 20 equations: the brief forms shown above are the culmination of later improvements by Gibbs, Hertz,[111] and Heaviside, and were known variously as the Hertz–Heaviside or Maxwell–Heaviside equations. Since Maxwell unified electricity and magnetism and understood the concept of an electromagnetic wave, the simpler name "Maxwell equations" was finally bestowed on them by Einstein.

To repeat, Eqs. (1.9.20)–(1.9.23) are, respectively, Gauss's law, Gauss's law in magnetism, the generalized form of Ampère's law, and Faraday's law of induction.

**Problem 1.9.1** Prove that $\nabla\cdot$ of Eq. (1.9.19) agrees with Eq. (1.9.17).

111 Heinrich Rudolph Hertz (1857–1894).

The four Maxwell equations (Eqs. (1.9.16)–(1.9.19)) plus two constitutive equations (Eqs. (1.9.12) and (1.9.13)) are six equations in four unknown fields ($\boldsymbol{E}$, $\boldsymbol{H}$, $\boldsymbol{D}$, and $\boldsymbol{B}$) and two unknown tensors ($\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$: thus, they can be solved uniquely.

The **Lorentz force** $\boldsymbol{F}$ on a particle of charge $q$ in an electromagnetic field is given by:

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \quad \text{(SI)}; \quad \boldsymbol{F} = q\left(\boldsymbol{E} + c^{-1}\boldsymbol{v} \times \boldsymbol{B}\right) \quad \text{(cgs)} \tag{1.9.24}$$

The Lorentz force encompasses the Lenz[112] "right-hand rule" between $\boldsymbol{v}$, $\boldsymbol{B}$, and $\boldsymbol{F}$ and also explains how cyclotrons and mass spectrometers work. Practical applications of the Lorentz force are (i) the **cyclotron** (Problem 1.9.2) with its **cyclotron frequency**:

$$\omega = v/r = qB/m \quad \text{(SI)}; \quad \omega = v/r = qB/mc \quad \text{(cgs)} \tag{1.9.25}$$

(ii) other particle accelerators that use bending magnets (such as the synchrotron), (iii) ion cyclotron resonance, and (iv) mass spectrometry.

**Problem 1.9.2 Cyclotron problem**. A particle with mass $m$, charge $q$, and velocity $\boldsymbol{v}$ injected into a magnetic field $\boldsymbol{B}$ normal to its velocity will feel a force $\boldsymbol{F}$ normal to both $\boldsymbol{v}$ and $\boldsymbol{B}$. Using Eq. (1.9.25), show that it will move in a circle of radius $r = mv/qB$. Lawrence[113] and Livingston[114] in the late 1930s accelerated elementary particles to high kinetic energies in an early particle accelerator named the cyclotron.

**Problem 1.9.3** Another application of the Lorentz force is the selection of particles by a mass spectrometer. A dilute plasma of ions, each bearing a charge $q$ and mass $m$, is accelerated by an electric field to a uniform velocity $\boldsymbol{v}$, then drifts into a uniform magnetic field $\boldsymbol{B}$, in which ions of different mass describe circular orbits of different radius $r$. The mass/charge ratio is given by $m/q = rB/v$.

The free energy density $u$ (J m$^{-3}$) in an electromagnetic field is given by

$$u = (1/2)\boldsymbol{E} \cdot \boldsymbol{D} + (1/2)\boldsymbol{B} \cdot \boldsymbol{H} \quad \text{(SI)}; \quad u = (1/8\pi)(\boldsymbol{E} \cdot \boldsymbol{D} + \boldsymbol{B} \cdot \boldsymbol{H}) \quad \text{(cgs)} \tag{1.9.26}$$

The energy flow is best defined by the Poynting[115] pseudovector $\boldsymbol{S}$:

$$\boldsymbol{S} \equiv \boldsymbol{E} \times \boldsymbol{H} \quad \text{(SI)}; \quad \boldsymbol{S} \equiv (c/4\pi)\boldsymbol{E} \times \boldsymbol{H} \quad \text{(cgs)} \tag{1.9.27}$$

For electric fields $\boldsymbol{E}$ and magnetic fields $\boldsymbol{H}$, either in a vacuum or in isotropic media, the wave equation for propagation is the **electromagnetic wave equation** (Problem 1.9.4):

$$\nabla^2\boldsymbol{E} - \varepsilon_0\varepsilon\mu_0\mu\left(\partial^2\boldsymbol{E}/\partial t^2\right) = 0 \quad \text{(SI)}; \quad \nabla^2\boldsymbol{E} - \left(\varepsilon\mu c^{-2}\right)\left(\partial^2\boldsymbol{E}/\partial t^2\right) = 0 \quad \text{(cgs)} \tag{1.9.28}$$

$$\nabla^2\boldsymbol{H} - \varepsilon_0\varepsilon\mu_0\mu\left(\partial^2\boldsymbol{H}/\partial t^2\right) = 0 \quad \text{(SI)}; \quad \nabla^2\boldsymbol{H} - \left(\varepsilon\mu c^{-2}\right)\left(\partial^2\boldsymbol{H}/\partial t^2\right) = 0 \quad \text{(cgs)} \tag{1.9.29}$$

These waves were initially called "Hertzian waves," e.g. by Marconi,[116] the inventor of the radio. The solutions to these wave equations can be shown to be transverse electromagnetic (TEM) sine or cosine waves (Eq (1.5.112), Problem 1.9.5, and Figure 1.14).

In a vacuum ($\mu = 1$ and $\varepsilon = 1$), Maxwell's equations simplify to:

$$\nabla \cdot \boldsymbol{E} = \varepsilon_0\rho \quad \text{(SI)}; \quad \nabla \cdot \boldsymbol{E} = 4\pi\rho \quad \text{(cgs)} \tag{1.9.30}$$

$$\nabla \cdot \boldsymbol{B} = 0 \quad \text{(SI)}; \quad \nabla \cdot \boldsymbol{B} = 0 \quad \text{(cgs)} \tag{1.9.31}$$

$$\nabla \times \boldsymbol{B} = \mu_0\boldsymbol{J} + \mu_0\partial\boldsymbol{D}/\partial t \quad \text{(SI)}; \quad \nabla \times \boldsymbol{B} = 4\pi c^{-1}\boldsymbol{J} + c^{-1}\partial\boldsymbol{D}/\partial t \quad \text{(cgs)} \tag{1.9.32}$$

112 Heinrich Friedrich Emil Lenz (1804–1865).
113 Ernest Orlando Lawrence (1901–1958).
114 Milton Stanley Livingston (1905–1986).
115 John Henry Poynting (1852–1914).
116 Guglielmo Marconi (1874–1937).

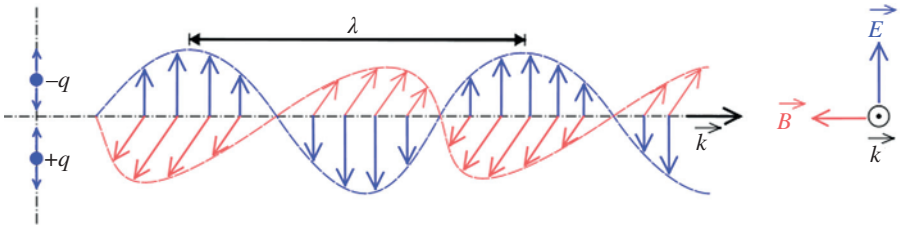



**Figure 1.14** Representation of ***E***, ***B***, and ***k*** vectors for a transverse electromagnetic (TEM) wave *in vacuo*: the ***E*** and ***B*** fields are in phase, and the three mutually orthogonal vectors ***E***, ***B***, and ***k*** form a right-handed set by Lenz's law. In cgs units *in vacuo*, the ***E*** and ***B*** vectors have the same magnitude.

$$\nabla \times \boldsymbol{E} = -\partial \boldsymbol{B}/\partial t \quad \text{(SI)}; \quad \nabla \times \boldsymbol{E} = -c^{-1}\partial \boldsymbol{B}/\partial t \quad \text{(cgs)} \tag{1.9.33}$$

In a conducting medium ($\sigma \neq 0$) with no source of charge ($\rho = 0$), the Maxwell equations yield the **first telegraph equation** (so named from a transmission-line theory for long-range telegraphy developed by Heaviside) [1.29]:

$$\nabla^2 \boldsymbol{E} - \varepsilon_0 \varepsilon \mu_0 \mu \left(\partial^2 \boldsymbol{E}/\partial t^2\right) - \mu_0 \mu \sigma (\partial \boldsymbol{E}/\partial t) = 0 \quad \text{(SI)}; \quad \nabla^2 \boldsymbol{E} - \left(\varepsilon \mu c^{-2}\right)\left(\partial^2 \boldsymbol{E}/\partial t^2\right) - \mu \sigma (\partial \boldsymbol{E}/\partial t) = 0 \quad \text{(cgs)} \tag{1.9.34}$$

The third term is a damping term, which allows for the possibility that a wave is absorbed by the medium: this is called the **evanescent wave**: the quantity $\sigma$ is also called the optical conductivity (at zero frequency it becomes the electrical conductivity).

This evanescent wave is exploited in **near-field scanning optical microscopy**. If waves propagate along $x$, so that $\partial/\partial y = 0$, $\partial/\partial z = 0$, then $E_x = H_x = 0$. Next, assume $E_y(x, t) = f(x)\exp(\mathrm{i}\omega t) = 0$ and $E_z(x, t) = g(x)\exp(\mathrm{i}\omega t) = 0$, i.e. assume **plane-polarized light** with the ***E*** vector in the $xy$ plane: then the differential equation to be solved is more simply:

$$\mathrm{d}^2 f/\mathrm{d}x^2 + \left(\varepsilon\mu\omega^2 - \mathrm{i}\sigma\mu\omega\right) f = 0$$

The space-dependent solution is

$$f(x) = A \exp\left[\pm \mathrm{i}\left(\varepsilon\mu\omega^2 - \mathrm{i}\sigma\mu\omega\right)^{1/2} x\right]$$

One can define a **complex index of refraction** $\aleph$ by its real part $n$ and its imaginary part $-\mathrm{i}k n$:

$$\left(\varepsilon_0 \varepsilon \mu_0 \mu \omega^2 - \mathrm{i}\sigma\mu_0\mu\omega\right)^{1/2} \equiv \omega n(1 - \mathrm{i}k) \equiv \omega\aleph \quad \text{(SI)}; \quad c^{-1}\left(\varepsilon\mu\omega^2 - \mathrm{i}4\pi\sigma\mu\omega\right)^{1/2} \equiv \omega n(1 - \mathrm{i}k) \equiv \omega\aleph \quad \text{(cgs)} \tag{1.9.35}$$

A **second telegraph equation** [1.29] can be written for the magnetic induction:

$$\nabla^2 \boldsymbol{B} - \varepsilon_0 \varepsilon \mu_0 \mu \left(\partial^2 \boldsymbol{B}/\partial t^2\right) - \mu_0 \mu \sigma (\partial \boldsymbol{B}/\partial t) = 0 \quad \text{(SI)}; \quad \nabla^2 \boldsymbol{B} - \left(\varepsilon \mu c^{-2}\right)\left(\partial^2 \boldsymbol{B}/\partial t^2\right) - \mu \sigma (\partial \boldsymbol{B}/\partial t) = 0 \quad \text{(cgs)} \tag{1.9.36}$$

The two telegraph equations are derived in Problem 1.9.11.

Next, the three assumptions $\sigma = 0$, $\rho = 0$, and $\boldsymbol{E}(\boldsymbol{r},t) = \boldsymbol{E}(\boldsymbol{r})\exp(\mathrm{i}\omega t)$ will lead from Eq. (1.9.34) to the real form of the **Helmholtz**[117] **equation**:

$$\nabla^2 \boldsymbol{E} + \varepsilon_0 \varepsilon \mu_0 \mu \omega^2 \boldsymbol{E} = 0 \quad \text{(SI)}; \quad \nabla^2 \boldsymbol{E} + \varepsilon\mu\omega^2 c^{-2} \boldsymbol{E} = 0 \quad \text{(cgs)} \tag{1.9.37}$$

with a real wavevector $k \equiv (\varepsilon_0 \varepsilon \mu_0 \mu)^{1/2}(\omega/c)$.

The Helmholtz equation resembles the spatial part of the **classical wave equation** for matter waves (waves in ocean, sound waves, vibrations of a string, electromagnetic waves in vacuum, etc.) of amplitude $F = F(\boldsymbol{r},t)$:

$$\nabla^2 F(\boldsymbol{r}, t) + v^{-2}\left[\partial^2 F(\boldsymbol{r}, t)/\partial t^2\right] = 0 \tag{1.9.38}$$

117 Heinrich Ludwig Ferdinand von Helmholtz (1821–1894).

The solution to this classical wave equation may be factored into space and time factors:

$$F(\boldsymbol{r},t) = f(\boldsymbol{r})g(t) = f(\boldsymbol{r})[A\exp(\mathrm{i}t/v) + B\exp(-\mathrm{i}t/v)] \tag{1.9.39}$$

where $\nabla^2 f(\boldsymbol{r}) = -v^{-2}f(\boldsymbol{r})$ and $\mathrm{d}^2g(t)/\mathrm{d}t^2 = -v^{-2}g(t)$, $v$ is the (assumed constant) wave speed, and $A$ and $B$ are constants whose values are determined from the boundary conditions.

Finally, Maxwell's four equations (1.9.30)–(1.9.33), the three constitutive equations (1.9.12), (1.9.13), and (1.9.15), and the three assumptions $\rho = 0$, $\boldsymbol{E}(\boldsymbol{r},t) = \boldsymbol{E}(\boldsymbol{r})\exp(\mathrm{i}\omega t)$, and $\boldsymbol{H}(\boldsymbol{r},t) = \boldsymbol{H}(\boldsymbol{r})\exp(\mathrm{i}\omega t)$, when considered together, yield (Problem 1.9.11) the **complex form of the Helmholtz equation**:

$$\nabla^2\boldsymbol{E} + \omega^2\mu_0\mu(\varepsilon_0\varepsilon - \mathrm{i}\sigma/\omega)\boldsymbol{E} = 0 \quad (\mathrm{SI}); \quad \nabla^2\boldsymbol{E} + \omega^2\mu(\varepsilon - \mathrm{i}\sigma/\omega)\boldsymbol{E} = 0 \quad (\mathrm{cgs}) \tag{1.9.40}$$

which, rewritten as $\nabla^2\boldsymbol{E} + \kappa^2\boldsymbol{E} = 0$, defines a complex wavevector $\kappa$:

$$\kappa \equiv \omega[\mu_0\mu(\varepsilon_0\varepsilon - \mathrm{i}\sigma/\omega)]^{1/2} \quad (\mathrm{SI}); \quad \kappa \equiv \omega[\mu(\varepsilon - \mathrm{i}\sigma/\omega)]^{1/2} \quad (\mathrm{cgs}) \tag{1.9.41}$$

If in Eq. (1.9.35) $\sigma = 0$ and $\kappa = 0$, then the real part of the refractive index is given by

$$n^2 = \mu_0\mu\varepsilon_0\varepsilon \quad (\mathrm{SI}); \quad n^2 = \mu\varepsilon \quad (\mathrm{cgs}) \tag{1.9.42}$$

**Problem 1.9.4** Derive the wave equation for the $\boldsymbol{E}$ field, Eq. (1.9.28), as follows [1.30]:

i) Consider a region with no net charges ($\rho = 0$) and no external sources of electromotive force, where $\mu$ and $\varepsilon$ are scalars that are isotropic (not dependent on coordinates) and also independent of time;
ii) Operate with $\nabla\times$, i.e. take the curl of both sides of Eq. (1.9.33), but replace $\boldsymbol{B}$ by $\mu\mu_0\boldsymbol{H}$;
iii) Substitute Ampère's law, Eq. (1.9.18) into the previous result, but replace $\boldsymbol{D}$ by $\varepsilon\varepsilon_0\boldsymbol{E}$;
iv) Use the vector identity, Eq. (1.6.25) $\nabla\times\nabla\times\boldsymbol{v} = \nabla(\nabla\bullet\boldsymbol{v}) - (\nabla^2)\boldsymbol{v}$, then remember from Gauss's law, Eq. (1.9.16) that for $\rho = 0$, $\nabla\cdot\boldsymbol{D} = \varepsilon\varepsilon_0\nabla\cdot\boldsymbol{E} = 0$;
v) Use Ohm's law, Eq. (1.9.15), to get rid of $\boldsymbol{J}$ and obtain the general wave equation, valid in isotropic media or in free space;
vi) Show that in an insulator ($\sigma = 0$), the third term drops out;
vii) In conductors, the term $\partial^2\boldsymbol{E}/\partial t^2$ is usually vanishingly small and can be neglected.

**Problem 1.9.5** Show that, both in a dielectric insulator and in a vacuum, a plane-wave electromagnetic field solution propagating along $x$, whose amplitude depends only on the coordinate $x$ and on the time $t$, can have no component along $x$, i.e. show that it must be a transverse electric wave [1.29].

**Problem 1.9.6** For the situation of Problem 1.9.4, find the plane-wave solution.

**Problem 1.9.7** For the plane-polarized wave propagating along $x$, as described in Problem 1.9.6, find the magnetic field.

**Problem 1.9.8** A photon of wavelength $\lambda = 500$ nm propagating in vacuum ($\varepsilon = 1$) has energy $W = hc/\lambda$. If it is equivalent to a wave packet constrained in a volume of 1 nm$^3$, then estimate its electric field (ignore its magnetic field).

**Problem 1.9.9** If the electric field $E$ of an electromagnetic wave in vacuum is $3\times10^4$ V m$^{-1}$, please estimate its magnetic field $H$.

**Problem 1.9.10** Let an electron of mass $m_e = 9.11\times10^{-31}$ kg and charge $e = -1.602\times10^{-19}$ C rotate in the first elliptical (quasi-circular) Bohr orbit (radius = $a_0 = 0.0529$ nm) around a proton ($M_p = 1.673\times10^{-27}$ kg) at a speed of $2.22\times10^6$ m s$^{-1}$. Compute (i) the gravitational energy, (ii) the electrical energy, and (iii) the magnetic energy of a system of two such electrons in coplanar circular orbits around protons 2 nm apart.

**Problem 1.9.11** Derive the two "**telegraph equations**," Eqs. (1.9.34) and (1.9.25).

**Problem 1.9.12** From Eq. (1.9.34) derive the Helmholtz equation, Eq. (1.9.37) [1.29].

**Problem 1.9.13** Start from Maxwell's equations (1.9.30)–(1.9.33), use the three constitutive equations (1.9.12), (1.9.13), and (1.9.15), then assume $\rho = 0$ to get:

$$\nabla \cdot \boldsymbol{E} = 0$$
$$\nabla \cdot \boldsymbol{H} = 0$$
$$\nabla \times \boldsymbol{E} + \mu_0 \mu \partial \boldsymbol{H}/\partial t = 0$$
$$\nabla \times \boldsymbol{H} - \sigma \times \boldsymbol{E} - \varepsilon\varepsilon_0 \partial \boldsymbol{E}/\partial t = 0$$

For plane waves, $\boldsymbol{E}(\boldsymbol{r},t) = \boldsymbol{E}(\boldsymbol{r})\exp(\mathrm{i}\omega t)$ and $\boldsymbol{H}(\boldsymbol{r},t) = \boldsymbol{H}(\boldsymbol{r})\exp(\mathrm{i}\omega t)$:

$$\nabla \times \boldsymbol{E} + \mathrm{i}\omega\mu_0\mu\boldsymbol{H} = 0$$
$$\nabla \times \boldsymbol{H} - \sigma\boldsymbol{E} - \mathrm{i}\omega\varepsilon_0\varepsilon\boldsymbol{E} = 0$$

Now derive the complex form of the Helmholtz equation, Eq. (1.9.40) [1.29].

**Problem 1.9.14** An electromagnetic wave of wavelength $\lambda = 589.3$ nm (and therefore angular frequency $\omega = 3.20 \times 10^{15}$ Hz), penetrates into bulk Cu as $\exp(-\omega n k z/c)$. Estimate the **skin depth** $d \equiv [2\mu^{-1}\sigma^{-1}\omega^{-1}]^{1/2}$ at which the wave is attenuated to $(1/e)$ of its value at the surface. Use $n^2k^2 = (c^2/2)\left\{\left[\varepsilon_0^2\varepsilon^2\mu_0^2\mu^2 + (\mu\mu_0\sigma/\omega)^2\right]^{1/2} - \varepsilon_0\varepsilon\mu_0\mu\right\}$, assuming $\mu = 1$, $\varepsilon = 1$, $\sigma = 5.8 \times 10^7$ S m$^{-1}$, $\omega = 3.20 \times 10^{15}$ Hz, and $n = 0.62$. (Since $n < 1$, the phase velocity of the light wave in Cu will exceed the speed of light, but of course, its group velocity will not (and cannot) do so) [1.29].

**Problem 1.9.15** The original version of Ampère's law had been simply:

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} \quad \text{(SI)} \tag{1.9.44}$$

But Maxwell realized that calculating the divergence $(\nabla\cdot)$ of Eq. (1.9.34) would yield:

$$\nabla \bullet \mu_0 \boldsymbol{J} = 0 \quad \text{(SI)} \tag{1.9.44}$$

which would disagree with a **continuity equation** that must hold whenever the charge density varied with time:

$$\nabla \bullet \mu_0 \boldsymbol{J} = -\partial\rho/\partial t \quad \text{(SI)} \tag{1.9.45}$$

How did Maxwell resolve this dilemma?

### 1.9.4 Scalar and Vector Potentials

It is often useful to define an electrical (scalar) potential $\phi(\boldsymbol{r})$ and a magnetic (vector) potential $\boldsymbol{A}(\boldsymbol{r})$ such that:

$$\boldsymbol{E}(\boldsymbol{r}) = -\nabla\phi(\boldsymbol{r}) - \partial\boldsymbol{A}/\partial t \quad \text{(SI)}; \quad \boldsymbol{E}(\boldsymbol{r}) = -\nabla\phi(\boldsymbol{r}) - (1/c)\partial\boldsymbol{A}/\partial t \quad \text{(cgs)} \tag{1.9.46}$$
$$\boldsymbol{B}(\boldsymbol{r}) = \nabla \times \boldsymbol{A}(\boldsymbol{r}) \quad \text{(SI)}; \quad \boldsymbol{B}(\boldsymbol{r}) = \nabla \times \boldsymbol{A}(\boldsymbol{r}) \quad \text{(cgs)} \tag{1.9.47}$$

One can use these two potentials instead of the fields upon which they are based.
These potentials are invariant under the "gauge transformations":

$$\phi \rightarrow \phi + (\partial\psi/\partial t)$$
$$\boldsymbol{A} \rightarrow \boldsymbol{A} + \nabla\psi$$

To remove this arbitrariness, one can force these two potentials to satisfy the Lorentz condition:

$$\nabla \cdot \boldsymbol{A} + \varepsilon\varepsilon_0\mu\mu_0(\partial\phi/\partial t) + \mu\mu_0\sigma\phi = 0 \quad \text{(SI)}; \quad \nabla \cdot \boldsymbol{A} + (\varepsilon\mu/c)(\partial\phi/\partial t) + \mu\sigma\phi = 0 \quad \text{(cgs)} \tag{1.9.48}$$

Using Eq. (1.9.48) one can obtain two **inhomogenous wave equations**:

$$\nabla^2\phi - \varepsilon\varepsilon_0\mu\mu_0\left(\partial^2\phi/\partial t^2\right) - \mu\mu_0\sigma(\partial\phi/\partial t) = -\rho/\varepsilon_0\varepsilon \quad \text{(SI)}$$
$$\nabla^2\phi - \varepsilon\mu\left(\partial^2\phi/\partial t^2\right) - \mu\sigma(\partial\phi/\partial t) = -(4\pi/\varepsilon)\rho \quad \text{(cgs)} \tag{1.9.49}$$

$$\nabla^2\boldsymbol{A} - \varepsilon\varepsilon_0\mu\mu_0\left(\partial^2\boldsymbol{A}/\partial t^2\right) - \mu\mu_0\sigma(\partial\boldsymbol{A}/\partial t) = -\mu_0\mu\boldsymbol{J} \quad \text{(SI)}$$
$$\nabla^2\boldsymbol{A} - \varepsilon\mu\left(\partial^2\boldsymbol{A}/\partial t^2\right) - \mu\sigma(\partial\boldsymbol{A}/\partial t) = -(4\pi\mu/c)\boldsymbol{J} \quad \text{(cgs)} \tag{1.9.50}$$

which mathematically look identical, but involve the charge density $\rho$ and the current density $\boldsymbol{J}$ as sources, respectively. These equations are related to Eqs. (1.9.34) and (1.9.36).

If the scalar potential $\phi$ does not depend on time, then Eq. (1.9.49) becomes **Poisson's equation:**

$$\nabla^2\phi = -\rho/\varepsilon_0\varepsilon \quad \text{(SI)}; \quad \nabla^2\phi = -4\pi\rho/\varepsilon \quad \text{(cgs)} \tag{1.9.51}$$

and if there is no source of charge, then Eq. (1.9.51) becomes **Laplace's equation**:

$$\nabla^2\phi = 0 \quad \text{(SI)}; \quad \nabla^2\phi = 0 \quad \text{(cgs)} \tag{1.9.52}$$

### 1.9.5 Polarization of Electromagnetic Waves

The **polarization direction** of an electromagnetic wave is usually taken to be along its $\boldsymbol{E}$ vector and is measured relative to some laboratory axes. Please do not confuse this **polarization direction** (=a direction in space) with the **electric polarization** $\boldsymbol{P}$ (=electrical dipole moment per unit volume) of Eq. (1.9.10).

Light emitted by individual atoms or molecules is **polarized** (but we may not know in which direction). For a single photon in vacuum, the $\boldsymbol{E}$ and $\boldsymbol{B}$ vectors are perpendicular to each other and to the propagation direction $\boldsymbol{k}$, but are randomly oriented relative to laboratory coordinates: for the experimenter this photon is **unpolarized**. Also, a macroscopic light source emitting many photons (Avogadro's number's worth) at random times will also have random polarizations: the beam is unpolarized or **natural light**. If the photon interacts with matter (other photons, or a gas, solid, or liquid), it can be partially or completely polarized (Section 1.12). Depending on the emission mode, a laser beam can be polarized (see Section 9.10).

A **polarizer** or **polarization filter** (usually an oriented crystal) can suppress most of the light polarized in one direction, relative to the optical axes of the crystal, and transmit most of the component polarized in the perpendicular direction: if the polarization is complete, it will be **plane-polarized (or linearly polarized)** light. If this polarization is incomplete, the light may consist of two mutually perpendicular plane-polarized components; depending on the amplitude of these two waves and their relative phase, the combined electric vector traces out an ellipse; the wave is **elliptically polarized**. Elliptical and plane polarization can be interconverted by **birefringent crystal filters**.

**Circular polarization** of TEM waves has the tip of $\boldsymbol{E}$ (or $\boldsymbol{H}$ if you like) at a fixed point in space describing a **circle** as time progresses and the magnitude of $\boldsymbol{E}$ stays constant. **Right** or **left** polarization describes how $\boldsymbol{E}$ rotates. Also, there are two conventions: (i) for physics and chemistry, right-handed or positive polarization, or right-threaded screw motion or right-handed helix as viewed from the target point toward the source point (Figure 1.15); (ii) for engineering: left-threaded screw or helix. Circular polarization is a limiting case of **elliptical polarization**. **Circular dichroism (CD)** is the differential absorption of left- and right-handed circularly polarized light, due to molecules (or acentric crystals) that have **optical handedness** or **optical activity** (this happens for most molecules of biological interest: the $\alpha$ helix, $\beta$ sheet, and random coil regions of proteins and the double helix of nucleic acids have recognizable CD spectral signatures).

The other special case is **linear polarization**: two circularly polarized waves of opposite helicity add up to a linearly polarized wave.

As shown **later in this book**, "optically active" molecules which have asymmetric chemical bonding (differences in chemical groups) around at least one C or other "active centers" will rotate plane-polarized light in a quantified way (optical rotation). Since natural evolution has favored one kind of rotation (e.g. dextro-rotatory activity) over the opposite (levo-rotatory activity), therefore, the study of relative optical activity coupled with structural analysis can monitor the biological activity of such molecules. There are also solid "polarizers," which are polymers, or crystals that can rotate linear polarization in either absorption or reflection.

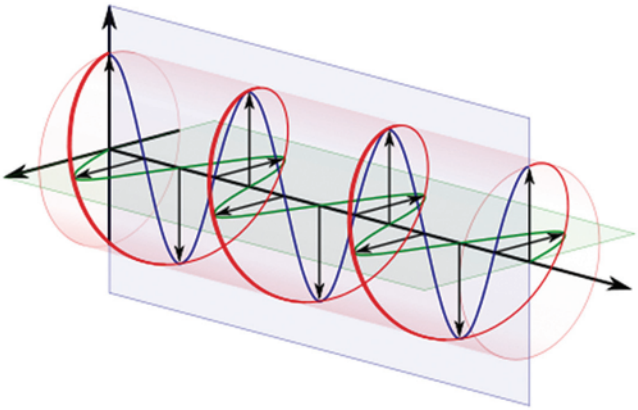


**Figure 1.15** Right-handed circular polarization of a TEM wave.

If a linearly polarized (plane-polarized) light of intensity $I_0$ is incident onto a polarizer, then the intensity of the transmitted light $I$ will depend upon the angle $\theta$ between the direction of the light polarization and the orientation of maximum polarization of the crystal: $I = I_0 \cos^2 \theta$.

### 1.9.6 Multipoles

One can define [1.30] the free energy $G$ (see Section 4.6) of a localized charge distribution $\rho(\boldsymbol{r})$, placed in an external potential $\phi(\boldsymbol{r})$ (which has no extra charge distribution associated with it):

$$G \equiv \int\int\int \rho(\boldsymbol{r})\phi(\boldsymbol{r})\, \mathrm{d}v(\boldsymbol{r}) \tag{1.9.53}$$

If $\phi(\boldsymbol{r})$ varies slowly enough in the region where $\rho(\boldsymbol{r})$ is significant, then one can use a Maclaurin expansion of $\phi(\boldsymbol{r})$ around some suitably chosen origin $\boldsymbol{r} = 0$:

$$\phi(\boldsymbol{r}) = \phi(0) + \boldsymbol{r} \bullet \nabla\phi(0) + (1/2)\sum_i \sum_j x_i x_j \left(\partial^2 \phi(0)/\partial x_i \partial x_j\right) + \ldots \tag{1.9.54}$$

where the summations extend over the three Cartesian components ($x_i$, $i = 1, 2, 3$) of $\boldsymbol{r}$. By using Eq. (1.9.47) in the form $\boldsymbol{E}(\boldsymbol{r}) = -\nabla\phi(\boldsymbol{r})$, and, using Eq. (1.9.30) in the simplified form $\nabla \cdot \boldsymbol{E} = 0$ (to justify adding a mathematically convenient term $-(1/6)\boldsymbol{r}^2 \nabla \cdot \boldsymbol{E}(0)$, since the trace of a quadrupole moment cannot be measured experimentally), one obtains:

$$\phi(\boldsymbol{r}) = \phi(0) - \boldsymbol{r} \cdot \boldsymbol{E}(0) - (1/6)\sum_i \sum_i \left(3x_i x_j - r^2 \delta_{ij}\right)\left(\partial E_j(0)/\partial x_i\right) + \ldots \tag{1.9.55}$$

where $\delta_{ij}$ is the Kronecker delta:

$$\delta_{ij} = 0 \text{ if } i \neq j, \quad = 0 \text{ if } i = j \tag{((1.6.55))}$$

Thus, finally, the free energy contributions to $G$ show explicitly that the electrostatic charge $q$ interacts with the potential, the electric dipole moment vector $\boldsymbol{m}$ (Figure 1.16) interacts with the external electric field $\boldsymbol{E}$, the traceless electric quadrupole moment $Q_{ij}$ interacts with the external field gradient, and so on:

$$G = q\phi(0) - \boldsymbol{m} \cdot \boldsymbol{E}(0) - (1/6)\sum_i \sum_i Q_{ij}\left(\partial E_j(0)/\partial x_i\right) + \ldots \tag{1.9.56}$$

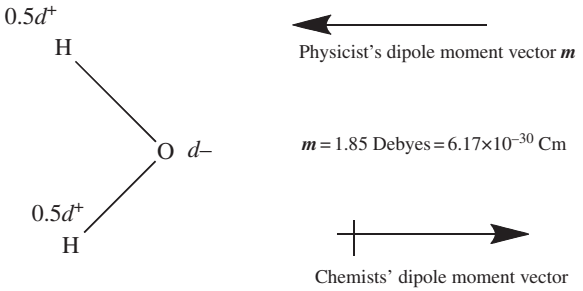



**Figure 1.16** The dipole moment $\boldsymbol{m}$ of water, $H_2O$ in the gas phase, with partial positive charges ($0.5d$) conceptually localized on the hydrogen atoms and a partial negative charge ($-d$) localized on the oxygen atom. The chemists and physicists disagree about the direction of the static electric dipole moment for this or any other molecule: the physicists run the positive dipole moment vector direction from the locus of partial negative charge (tail) to the locus of positive charge (head); unfortunately, the chemists follow Pauling's (Linus Carl Pauling (1901–1994)) convention in *The Nature of the Chemical Bond,* and do the reverse!

where

$$q \equiv \int \rho(\boldsymbol{r})\, \mathrm{d}v(\boldsymbol{r}) \tag{1.9.57}$$

$$\boldsymbol{m} \equiv \int \rho(\boldsymbol{r})\boldsymbol{r}\, \mathrm{d}v(\boldsymbol{r}) \tag{1.9.58}$$

$$Q_{ij} \equiv \int \rho(\boldsymbol{r})\left(3x_i x_j - r^2\delta_{ij}\right)\mathrm{d}v(\boldsymbol{r}) \tag{1.9.59}$$

### 1.9.7 Electric Dipole Moments

In many cases, the (di)electric polarization $\boldsymbol{P}$ is proportional to the electric field strength $\boldsymbol{E}$. The relation between the electric displacement $\boldsymbol{D}$ and the electric field strength $\boldsymbol{E}$ is given by

$$\begin{aligned}&\boldsymbol{D} = \varepsilon_0\boldsymbol{\varepsilon}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \varepsilon_0\boldsymbol{\chi}^{(1)}_{\mathrm{SI}}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P} \quad (\mathrm{SI}); \quad \boldsymbol{D} = \boldsymbol{\varepsilon}\boldsymbol{E} = \boldsymbol{E} + 4\pi\boldsymbol{\chi}^{(1)}_{\mathrm{esu}}\boldsymbol{E} = \boldsymbol{E} + 4\pi\boldsymbol{P} \quad (\text{cgs-esu});\\ &\boldsymbol{D} = \varepsilon_0\boldsymbol{\varepsilon}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \boldsymbol{\chi}^{(1)}_{\mathrm{SI}'}\boldsymbol{E} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P} \quad (\mathrm{SI}')\end{aligned} \tag{(1.9.12)}$$

If $\varepsilon$ and $\chi$ are scalars, then the (di)electric polarization $\boldsymbol{P}$ is given by

$$\begin{aligned}&\boldsymbol{P} = \varepsilon_0(\varepsilon - 1)\boldsymbol{E} = \varepsilon_0\chi^{(1)}_{\mathrm{SI}}\boldsymbol{E} \quad (\mathrm{SI}); \quad \boldsymbol{P} = [(\varepsilon - 1)/4\pi]\boldsymbol{E} = \chi^{(1)}_{\mathrm{esu}}\boldsymbol{E} \quad (\text{cgs-esu});\\ &\boldsymbol{P} = \varepsilon_0(\varepsilon - 1)\boldsymbol{E} = \chi^{(1)}_{\mathrm{SI}'}\boldsymbol{E} \quad (\mathrm{SI}')\end{aligned} \tag{1.9.60}$$

This polarization has two components: the induced polarization $\boldsymbol{P}_\alpha$ (due to movement of the centers of charge, or to the static electric dipole polarizability $\boldsymbol{\alpha}$ of molecules), and the dipole polarization $\boldsymbol{P}_\mu$ (due to the orientation of the permanent dipoles $\boldsymbol{m}$ in the applied electric field $\boldsymbol{E}$):

$$\boldsymbol{P} = \boldsymbol{P}_\alpha + \boldsymbol{P}_\mu \tag{1.9.61}$$

This second effect will be computed first. A molecule of permanent electric dipole moment $\boldsymbol{\mu}$ (C m), when put into an external electric field $\boldsymbol{E}$ (V $\mathrm{m}^{-1}$), assumes an angle $\theta$ with respect to $\boldsymbol{E}$; its energy $\Delta U$ (J) is then:

$$\Delta U = -\boldsymbol{m}\cdot\boldsymbol{E} = -mE\cos\theta \tag{1.9.62}$$

### 1.9.8 Electric Multipole Moments, Polarizabilities and Hyperpolarizabilities

In intense electric fields $\boldsymbol{E}$, the electric dipole moment vector $\boldsymbol{m}$ is no longer a constant, but acquires field-dependent higher-order contributions:

$$\boldsymbol{m}_i(\boldsymbol{E}) = \boldsymbol{m}_i^0 + \sum_k \alpha_{ik}E_k + \sum_k\sum_l \beta_{ikl}E_kE_l + \sum_k\sum_l\sum_m \gamma_{iklm}E_kE_lE_m + \ldots \tag{1.9.63}$$

where $\boldsymbol{\alpha}$ is the rank-two **electric polarizability tensor** with 9 elements, $\boldsymbol{\beta}$ is the rank-three first hyperpolarizability tensor with $3 \times 3 \times 3 = 27$ elements, and $\boldsymbol{\gamma}$ is the rank-four second **hyperpolarizability tensor**, with $3 \times 3 \times 3 \times 3 = 81$ elements. The next tensor $\boldsymbol{\delta}$ has not yet been measured. The indices $k$, $l$, and $m$ range over values 1, 2, 3, corresponding to $x$, $y$, and $z$ coordinates. In abbreviated notation, Eq. (1.9.65) becomes:

$$\boldsymbol{m} = \boldsymbol{m}_0 + \boldsymbol{\alpha E} + \boldsymbol{\beta EE} + \boldsymbol{\gamma EEE} + \ldots \tag{1.9.64}$$

The macroscopic version of this expansion in powers of $\boldsymbol{E}$ involves the electric susceptibility $\chi$:

$$\chi_{ij}(\boldsymbol{E}) = \chi^0_{ij} + \sum_k \chi^{(1)}_{ijk} E_k + \sum_k \sum_l \chi^{(2)}_{ijkl} E_k E_l + \sum_k \sum_l \sum_m \chi^{(3)}_{ijklm} E_k E_l E_m + \ldots \tag{1.9.65}$$

where $\chi$ is the $3 \times 3$ electrical susceptibility tensor, and $\chi^{(n)}$ is the $n$th order contribution to it. One talks about tensor elements $\chi^{(2)}_{1213}$or, equivalently, about $\chi^{(2)}_{xyxz}$. Since the susceptibility $\chi$ is now a tensor, the electric (or dielectric) polarization (or polarization density) $\boldsymbol{P}$ becomes in a formal extension of Eq. (1.9.10):

$$\boldsymbol{P} = \varepsilon_0 \boldsymbol{\chi E} \quad \text{(SI)}; \quad \boldsymbol{P} = \boldsymbol{\chi E} \quad \text{(cgs-esu)}; \quad \boldsymbol{P} = \boldsymbol{\chi E} \quad (\text{SI}') \tag{1.9.66}$$

The polarizabilities are calculated from the electrical field dependence of either the total molecular energy or the dipole moment; the $\chi^{(2)}_{ijkl}$in crystals obey the laws of crystal symmetry and are measured using powerful laser sources. Nonlinear optics are discussed further in Special Topic ST1.8.

**PART FOUR:** Quantum Numbers, Relativity, Optics, and More Mathematics

## 1.10 The Size of Fundamental Particles and the Physical Meaning of Quantum Numbers

Here we discuss a few fundamental issues.

**How big is an electron**? (i) The "classical radius" of an electron (or Lorentz radius or Thomson scattering length), $r_{\text{cl}} = e^2 m_{\text{e}}^{-1} c^{-2} = 2.892 \times 10^{-15}$ m, originates from equating the Coulomb potential to the Einstein rest-mass energy (Problem 1.10.1). (ii) Another measure of the electron size is its Compton[118] wavelength: $\lambda_{\text{C}} = h/m_{\text{e}}c = 2.42 \times 10^{-12}$ m, i.e. the length below which particle creation and annihilation (and therefore quantum field theory) come into play: if a photon of energy in excess of $m_{\text{e}}c^2$ is used to "find" the electron, then this photon may itself become a new electron, thus mooting the question of where the electron had been! More precisely, as discussed in Section 2.1, the position uncertainty $\Delta x$ and the momentum uncertainty $\Delta p_x$ are linked by the Heisenberg uncertainty principle: $\Delta x \Delta p_x \geq \hbar/2$; if we use for $\Delta p_x$ the relativistic momentum of the electron, $p = m_{\text{e}}c$, then the uncertainty in its position is given by $\Delta x \geq (\hbar/2)/m_{\text{e}}c = \lambda_{\text{C}}/4\pi = 1.92 \times 10^{-13}$ m. (iii) Electron–electron scattering provides an estimated collision distance between two electrons of about $10^{-16}$ m (much smaller than that of an atom, $10^{-10}$ m, or even of a nucleus, $10^{-14}$ m). (iv) Of course, electrons seem to be "point particles" with negligible radius: is this correct?

Scattering experiments, often using electrons as "bullets," yield a scattering cross section for hadrons, and from this cross section one can obtain an estimated size for the hadrons. The original Rutherford[119] experiment used a beam of $({}_2\text{He}^4)^{2+}$ nuclei (α particles) from a Ra source, and arrived at a surprisingly small nuclear size, relative to the size of atom.

**How big is a photon**? If one considers the wave nature of light, one may think that the photon size is roughly equal to its wavelength (say 500 nm): however, when the photon is absorbed by an atom, it "disappears" within a body of radius 0.5 nm: this is a manifestation of the intricacies of the wave-particle duality, which are discussed in Section 2.39.

**Problem 1.10.1** Evaluate the classical radius of the electron $r_{\text{cl}}$, by assuming that the rest mass of the electron $m_{\text{e}}$ is totally due to its electrostatic potential $e^2/4\pi\varepsilon_0 r_0$.

118 Arthur Holly Compton (1892–1962).
119 Ernest Rutherford, first Baron Rutherford of Nelson (1871–1937).

**Problem 1.10.2** Evaluate the speed of the orbital motion of an electron that has orbital angular momentum $m_e v r_0 = h$, if the mass is concentrated at the electron radius $r_0$ estimated above either from electron–electron scattering ($10^{-16}$ m) or from its "classical radius" ($2.892 \times 10^{-15}$ m) or from the Compton wavelength $\lambda_C (2.426 \times 10^{-12}$ m).

**What do quantum numbers mean**? As we shall see in Section 2.5, the three spatial quantum numbers ($n$, $l$, $m_l$) for the H atom identify the allowed eigenstates for the solution of the Schrödinger equation, with certain characteristic energies and spatial features (e.g. the angular momentum quantum number $l$ describes how much angular momentum the atom has). But then, what is the meaning of the "electron spin" quantum number?

Electron spin can be visualized as the "helicity" of the particle. Schrödinger suggested a zig-zag picture or "Zitterbewegung" or "trembling motion" for the electron: from the solutions to the relativistically correct Dirac equation in space-time, the electron seems to oscillate ("zig-zag") between two states: a massless "zig" particle with right-handed helicity and massless "zag" particle with left-handed helicity: these two particles are mutual sources for each other, and are coupled by a coupling constant that bears the rest mass of the electron. In normal three-dimensional space, the velocity of "zig" and "zag" (at the speed of light) continually reverses, but the direction of spin remains constant.

Since photons are absorbed to change the spin projection of an electron from $-1/2$ to $+1/2$, therefore, the spin of the photon must be 1. The graviton has been postulated to carry a spin of 2 because of the symmetry of the equations in Einstein's general theory of relativity (gravity comes from a rank-two stress-energy tensor).

Other quantum numbers (charm, strangeness, etc.) were invented *ad hoc* to preserve, at least for fermions, the hypothesis that every fundamental particle must have a unique set of quantum numbers. These newer quantum numbers may have more abstruse physical significance.

## 1.11 Special Relativity

There was a young lady named Bright,
whose speed was much faster than light;
she started one day
on her relative way
and came back on the previous night {Edward Teller[120]}.

There was a sprinter in action,
who lost his race by a fraction:
he thought he had won,
except he shrank by a
Lorentz-FitzGerald[121] contraction [? Edward Teller].

Special relativity was introduced by Einstein in 1905 [1.8], to explain the null result of the 1887 Michelson–Morley experiment, that the speed of light was independent of the seasons, i.e. it did not add or subtract vectorially from the speed of the "ether wind." Einstein postulated that the speed of light *in vacuo*, or the speed of information and energy transfer, is a **universal constant** $c$, independent of frame of reference. Phase velocities in excess of $c$ are allowed, but group velocities for waves (which transmit energy or information) are limited by $c$ *in vacuo*. In material media, the speed is $nc$, where $n$ is the index of refraction. **Tachyons** (particles moving faster than light) are just gleams in some theoreticians' eyes. The following discussion is adapted from [1.31].

**Sideline 1.16** Alas, in 1894 Michelson stated "The more important fundamental laws and facts of physical science have all been discovered, and these are now so firmly established that the possibility of their ever being supplanted in consequence of new discoveries is exceedingly remote [...]. Many other instances might be cited, but these will suffice to justify the statement that 'our future discoveries must be looked for in the sixth place of decimals'." But Michelson lived long enough to regret that pronouncement, and gave valid summaries of special and general relativity.

120 Edward Teller (1908–2003).
121 George Francis FitzGerald (1851–1901).

If a particle moves at speed $V$ in the $x$-direction, then the usual "Galilean"[122] transformation of coordinates from a stationary system $S$ to a system $S'$ moving with the particle is

$$x' = x - Vt \tag{1.11.1}$$

$$y' = y \tag{1.11.2}$$

$$z' = z \tag{1.11.3}$$

$$t' = t \tag{1.11.4}$$

This says that the time clocks are the same in both systems.

This is incompatible with Maxwell's equations, as shown below by using Gauss's law, Eq. (1.9.16) and the Lorentz force, Eq. (1.9.24). Assume that the two systems $S$ and $S'$ move at velocities $\boldsymbol{v}$ and $\boldsymbol{v}'$, and relative velocity $\boldsymbol{V} = \boldsymbol{v}' - \boldsymbol{v}$. If we use the Galileian transformation, and assume that the charge $q$ and the electric displacement $\boldsymbol{D}$ is the same in the two systems:

$$\nabla' \cdot \boldsymbol{D} = \rho' = q'/\Delta x' \Delta y' \Delta z' = q/\Delta x \Delta y \Delta z = \rho = \nabla \cdot \boldsymbol{D} \quad \text{(SI)} \tag{1.11.5}$$

the Lorentz force must also be the same in both systems: $\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) = \boldsymbol{F}' = q(\boldsymbol{E}' + \boldsymbol{v}' \times \boldsymbol{B}')$ then:

$$\boldsymbol{E} = \boldsymbol{E}' - \boldsymbol{V} \times \boldsymbol{B}' \tag{1.11.6}$$

For simplicity, assume zero electrical polarization: $\boldsymbol{P}' = \boldsymbol{P} = 0$: then $\boldsymbol{D} = \boldsymbol{D}' - \varepsilon_0 \boldsymbol{V} \times \boldsymbol{B}'$. Finally:

$$\nabla' \cdot \boldsymbol{D}' = \rho' - \varepsilon_0 \nabla' \cdot \boldsymbol{V} \times \boldsymbol{B}' \tag{1.11.7}$$

which brings out a "second term," which depends on $\boldsymbol{V}$. All attempts (including the 1887 Michelson–Morley experiment) to measure the speed of light at various times during the year, hoping to measure the speed $\boldsymbol{V}$ of the "luminiferous ether" (which was thought necessary to propagate electromagnetic waves) failed: the speed of light was the same, no matter where the Earth was with respect to the cosmos. This forced a shift away from the Galileian transformation at relativistic speeds.

When this speed $V$ becomes large enough to approach the speed of light ("relativistic"), then the **Lorentz transformation** steps in:

$$x' = \left(1 - V^2 c^{-2}\right)^{-1/2} (x - Vt) = \gamma(x - \beta ct) \tag{1.11.8}$$

$$y' = y \tag{1.11.9}$$

$$z' = z \tag{1.11.10}$$

$$t' = (1 - V^2 c^{-2})^{-1/2} (t - Vx/c^2) = \gamma(t - \beta x/c) \tag{1.11.11}$$

where :

$$\gamma \equiv \left(1 - V^2 c^{-2}\right)^{-1/2} \tag{1.11.12}$$

$$\beta \equiv V/c \tag{1.11.13}$$

Equation (1.11.8) is called the Lorentz–Fitz Gerald **contraction of space**; Einstein's new idea is Eq. (1.11.11), the **dilatation of time**: a clock advances more slowly in a system moving at a high speed $V$. When $V \ll c$, $\gamma \approx 1$, $\beta \approx 0$, and the Lorentz transformation reduces to the Galilean transformation.

The Lorentz transformation has the following cute property. If two events are measured in coordinate system $S$ as separated by $\Delta x$, $\Delta y$, and $\Delta z$, and time $\Delta t$, and they are measured also in coordinate system $S'$ as being separated by different amounts of space $\Delta x'$, $\Delta y'$, $\Delta z'$, and time $\Delta t'$, then the Lorentz invariance requires:

$$(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2(\Delta t)^2 = (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2 - c^2(\Delta t')^2 \tag{1.11.14}$$

122 Galileo Galilei (1564–1642).

When time is not involved, this invariance is the well-known and quite ordinary rotation of a coordinate system in $x$, $y$, $z$ space:

$$(\Delta r)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2 \tag{1.11.15}$$

which causes no dilatation or contraction of three-space. The Lorentz invariance is best analyzed in **four-space**, by introducing a $1 \times 4$ column vector $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{pmatrix} x1 \\ x2 \\ x3 \\ x4 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \\ \mathrm{i}ct \end{pmatrix} \tag{1.11.16}$$

(here $\mathrm{i} = (-1)^{1/2}$). There are other possible definitions, with and without an explicit i, with and without $+$. For the Lorentz transformation given above, the transformation matrix is

$$\boldsymbol{C} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & \mathrm{i}\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\mathrm{i}\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \tag{1.11.17}$$

This matrix equation relating the four-vector $\boldsymbol{X}$ to the four-vector $\boldsymbol{X}'$ is

$$\boldsymbol{X}' = \boldsymbol{C}\boldsymbol{X} \tag{1.11.18}$$

or, using the Einstein summation convention:

$$x'_\mu = c_{\mu\nu} x_\nu \quad (\mu = 1, 2, 3, 4) \tag{1.11.19}$$

which means, when written out in full:

$$x'_\mu = \sum_{\nu=1}^{\nu=4} c_{\mu\nu} x_\nu \quad (\mu = 1, 2, 3, 4) \tag{1.11.20}$$

It can be shown that $\det \boldsymbol{C} = 1$ (Problem 1.11.1). The norm of $\boldsymbol{C}$, or trace of $\boldsymbol{C}$, or sum of its diagonal terms, is $2 + 2\gamma$. Since $\det \boldsymbol{C} = 1$, therefore, we can consider the Lorentz transformation matrix $\boldsymbol{X}$ like the four-dimensional analog of the Eulerian rotation in three-space. We now seek quantities that are "covariant with the Lorentz transformation," i.e. are "relativistically correct." We next define, in this new four-space a few essential quantities:

The **proper time** $\Delta\tau$ is defined by

$$(\Delta\tau)^2 \equiv -c^{-2} - \Delta x_\mu \Delta x_\mu = (\Delta t)^2 - c^{-2}\left[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\right] \tag{1.11.21}$$

If $(\Delta\tau)^2 > 0$, then $\Delta\tau$ is real and represents a "time-like" interval; if $(\Delta\tau)^2 < 0$, then $\Delta\tau$ is imaginary and represents a "space-like" interval in proper time. We can also write:

$$(\Delta\tau)^2 = (\Delta t)^2 - c^{-2}(\Delta r)^2 = (\Delta t)^2\left[1 - V^2 c^{-2}\right] = \gamma^{-2}(\Delta t)^2 \tag{1.11.22}$$

A **velocity four-vector** is obtained by differentiating the position four-vector:

$$U_\mu \equiv \mathrm{d}x_\mu/\mathrm{d}\tau \quad (\mu = 1, 2, 3, 4) \tag{1.11.23}$$

whence we can obtain, for the Lorentz transformation, Eq. (1.11.16), using $(\mathrm{d}x/\mathrm{d}\tau) = (\mathrm{d}x/\mathrm{d}t)(\mathrm{d}t/\mathrm{d}\tau)$, and so forth:

$$U_1 = \gamma(\mathrm{d}x/\mathrm{d}t), U_2 = \gamma(\mathrm{d}y/\mathrm{d}t), U_3 = \gamma(\mathrm{d}z/\mathrm{d}t), U_4 = \mathrm{i}c\gamma \tag{1.11.24}$$

The **four-vector linear momentum** is defined by:

$$P_\mu \equiv \mu_0 U_\mu \quad (\mu = 1, 2, 3, 4) \tag{1.11.25}$$

where $m_0$ is the rest mass (more about $m_0$ below). Therefore, the momentum components are:

$$P_1 = \gamma m_0(\mathrm{d}x/\mathrm{d}t), P_2 = \gamma m_0(\mathrm{d}y/\mathrm{d}t), P_3 = \gamma m_0(\mathrm{d}z/\mathrm{d}t), P_4 = \mathrm{i}cm_0\gamma \tag{1.11.26}$$

The **four-vector force** then is

$$F_\mu = \mathrm{d}\big(m_0\big(\mathrm{d}x_\mu/\mathrm{d}\tau\big)\big)/\mathrm{d}\tau \quad (\mu = 1, 2, 3, 4) \tag{1.11.27}$$

The **relativistic mass** $m$ is defined as:

$$m \equiv \gamma m_0 \tag{1.11.28}$$

The **total energy** $W$ of a particle is defined by:

$$W \equiv cP_4/\mathrm{i} = m_0\gamma c^2 = m_0c^2 + T = m_0c^2\left[1 + (1/2)V^2c^{-2} + (3/8)V^4c^{-4} + \ldots\right] \tag{1.11.29}$$

Therefore, the kinetic energy $T$ is equal to $(1/2)m_0V^2$ only for speeds $V \ll c$. Rest mass and kinetic energy $T$ may be interconverted: indeed the first term on the expanded right-hand side of Eq. (1.11.29), namely $m_0c^2$, corresponds to Eq. (1.2.5), the interconversion between rest mass and energy. The law of conservation of momentum becomes:

$$\sum_i P_\mu(i) = \sum_j P_\mu(j) \quad (\mu = 1, 2, 3, 4) \tag{1.11.30}$$

for the particles $i = 1, 2, \ldots$ present before the collision or interaction (on left) and for the particles $j = 1, 2, \ldots$ present after the event (on the right). In particular, the conservation of $P_4$ implies from Eq. (1.11.29) that rest mass and kinetic energy may be interconverted. Hence also, if $p^2 = m_0^2\left[(\mathrm{d}x/\mathrm{d}t)^2 + (\mathrm{d}y/\mathrm{d}t)^2 + (\mathrm{d}z/\mathrm{d}t)^2\right]$ is the square of the ordinary momentum, then:

$$W^2 = p^2c^2 + m_0^2c^4 \tag{1.11.31}$$

As discussed in Chapter 2, the Dirac equation is a covariant version of the nonrelativistic quantum-mechanical Schrödinger equation:

$$H\psi = [c\boldsymbol{\alpha}(\hbar/\mathrm{i})\nabla - |e|\boldsymbol{A}] + \left[\beta m_0c^2 + e\phi\right]\psi = -(\hbar/\mathrm{i})(\partial\psi/\partial t) \tag{1.11.32}$$

where, as above, $-|e|$ is the charge on the electron, $c$ is the speed of light, $\boldsymbol{A}$ is the magnetic vector potential, $\phi$ is the scalar electric potential, but now $\boldsymbol{\alpha}$ is a traceless $4 \times 4$ matrix with the following Cartesian components:

$$\alpha_x \equiv \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \alpha_y \equiv \begin{pmatrix} 0 & 0 & 0 & -\mathrm{i} \\ 0 & 0 & \mathrm{i} & 0 \\ 0 & -\mathrm{i} & 0 & 0 \\ \mathrm{i} & 0 & 0 & 0 \end{pmatrix}, \alpha_z \equiv \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{1.11.33}$$

and $\beta$ is a scalar traceless $4 \times 4$ matrix with components:

$$\beta \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{1.11.34}$$

**Problem 1.11.1** For Eq. (1.11.17), show that $\det \boldsymbol{C} = 1$.

**Problem 1.11.2** For Eq. (1.11.17), find $\boldsymbol{C}^{-1}$, the inverse matrix to $\boldsymbol{C}$.

**Problem 1.11.3** Verify Eq. (1.11.24).

## 1.12 Feynman Diagrams

As a mnemonic for calculations in quantum electrodynamics, Feynman[123] found it convenient to identify the most important interactions between particles, antiparticles, and photons (in QCD, the antiparticles are shown in the collision as going "backwards in time" [Stückelberg[124] convention]). One can discuss all electron–electron interactions in terms of the second electron being in the electric field caused the first one, and vice versa, or in terms of an exchange of virtual photons between the two electrons: this is usually represented by a Feynman diagram (Figure 1.17).

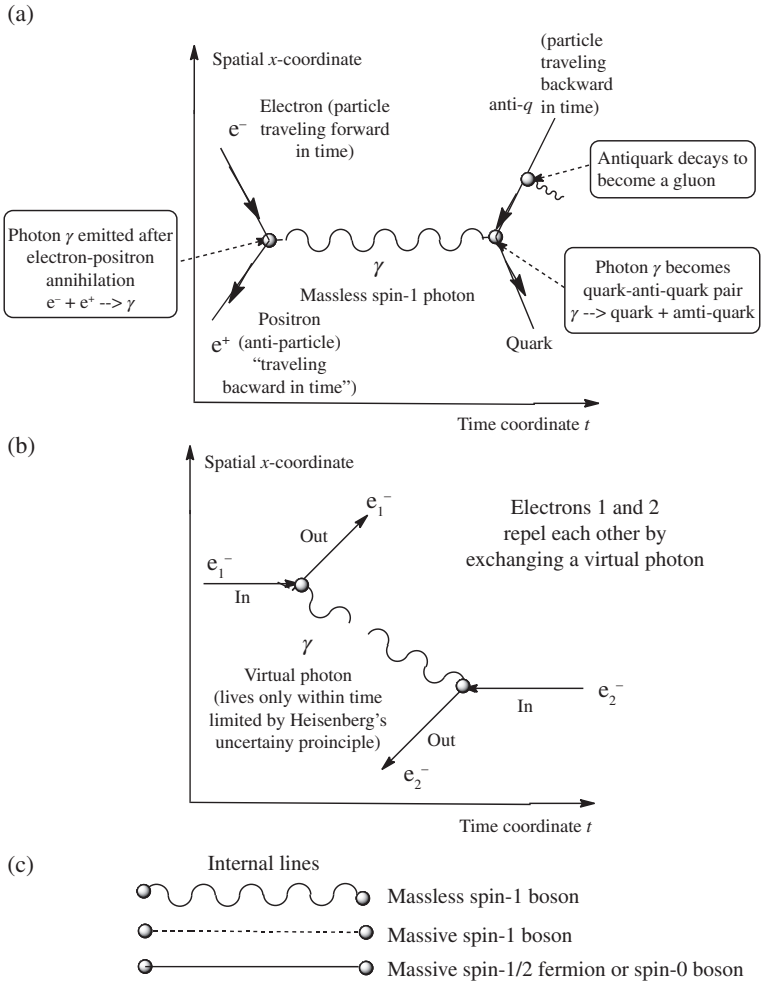



**Figure 1.17** Two simple Feynman diagrams (a) for a particle–antiparticle collision (b) for an electron–electron repulsion. (c) Particles are shown as solid lines with a time's arrow (particles going forward in time, antiparticles going backward in time), while spin-1 bosons (photons) are shown as a wavy line, and a massive spin-1 boson is shown as a dotted line. In (a), a gluon is shown as a very squiggly line.

**Problem 1.12.1** Draw the Feynman diagram for the spontaneous "β" decay (mediated by a massive vector boson W−) of a neutron ($_0n^1$, with quark contents $uu\overline{d}$) (Table 1.2) into a proton ($_1p^1$, with quark contents $uu\overline{d}$), electron (e), and electron antineutrino ($\overline{\nu}_e$) (Eq. (1.3.1). Two of the three component quarks (first and third) are mere "spectators" in the reaction. The interaction is also called Fermi coupling (Figure 1.18).

**Sideline 1.17** Feynman was a great scientist, teacher, writer, and public speaker, but was also an inveterate punster. Here is an original Feynman story. In 1963, a seminar with the approximate title "Hartree–Fock Calculations on the

123 Richard Phillips Feynman (1918–1988).
124 Ernst Carl Gerlach Stückelberg (1905–1984).

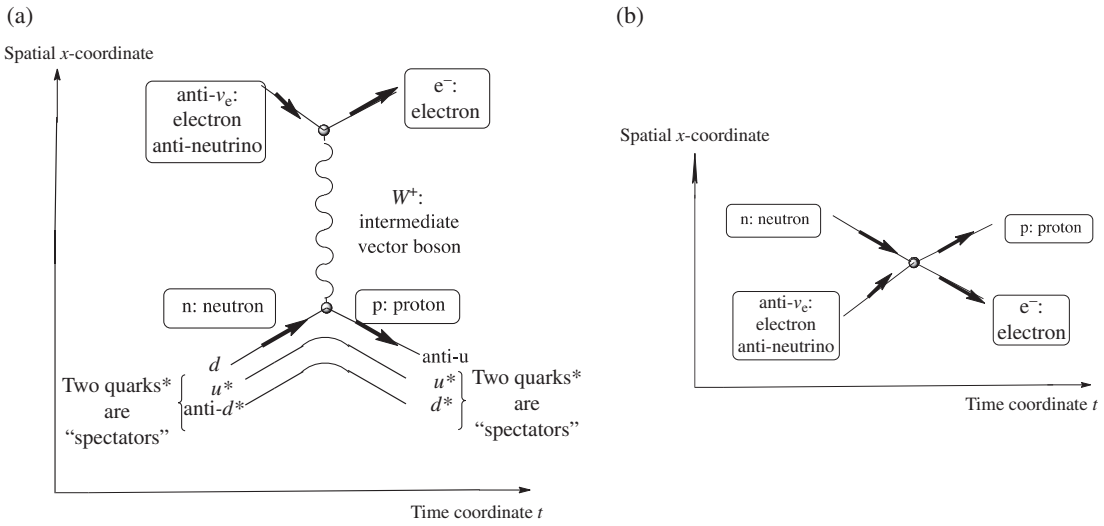



**Figure 1.18** Feynman diagrams for the spontaneous "β" decay of a neutron ${}_0n^1 \rightarrow {}_1p^1 + e^- + \bar{\nu}_e$ (a) including the intermediate vector boson $W^+$ and (b) omitting the $W^+$ boson.

$1s^2$ Wavefunction Amplitudes in Lanthanum" was presented by A.J. Freeman[125] at a solid state physics seminar series at Caltech. Feynman came in late, as the seminar speaker was droning on, and sat down in the fifth row, next to this author (RMM). In a very loud sotto voce, Feynman asked RMM "What kind of obscene theory is this Hartree–Fock?" Stuck as straight man, RMM answered that he did not know; the fellow graduate students in the fifth row, all fellow chemists from the McConnell[126] group), broke out in snickers. Feynman went on to say that, when he had been at Los Alamos during World War II, and shared an office with two Brits, one day a secretary came to the open door and said "I have a letter here for a Dr. Fuchs. Do you know him?" (This would be Klaus Fuchs,[127] who was later unmasked as a spy for the Soviet Union.) Feynman said that he did not react fast enough, but one of the Brits answered: "No, we do not know him, but the word surely sounds familiar." Very loud snickers from the McConnell group. At the end of seminar, the seminar coordinator, Rudolf Mößbauer[128] thanked the seminar speaker for the talk, and the audience "for very audible comments." In 1988, RMM called Prof. Freeman by telephone: was he the same A.J. Freeman of 25 years before? Indeed, he was: RMM repeated to him the above story. But Freeman gave the "other side" of the story: he had done calculations that Mößbauer did not believe, so he had been invited to Caltech to defend himself. At a first general physics colloquium, Feynman asked many detailed questions, and declared Freeman's calculation to be correct. As a result, Freeman was "on cloud nine," but was happy to be told what had gone on in the fifth row during his second seminar at Caltech: Feynman was just having some fun.

## 1.13 General Relativity, Gravity as a Curvature of Space-Time and the Size of the Universe

In his general theory of relativity, Einstein starts with a generalization of Lorentz invariance, Eq. (1.11.14) in non-Cartesian coordinate systems by introducing tensor notation and the Einstein summation convention for the space-time line element d$s$:

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2 \equiv g_{\mu\nu}dx^\mu dx^\nu \equiv \sum_{\mu=1}^{\mu=4}\sum_{\nu=1}^{\nu=4} g_{\mu\nu}dx^\mu dx^\nu \tag{1.13.1}$$

125 Arthur J. Freeman (1930–2016).
126 Harden Marsden McConnell (1927–2014).
127 Emil Julius Klaus Fuchs (1911–1988).
128 Rudolf Ludwig Mößbauer (1929–2011).

which for a Cartesian system has four nonzero covariant terms $g_{11} = g_{22} = g_{33} = 1$, $g_{44} = -c^2$ but $g_{\mu\nu} = 0$ for $\mu \neq \nu$. In non-Cartesian spaces, the expression is more complicated. The use of Eq. (1.13.1) is called the **principle of covariance**. Note that if you change coordinate systems from $(x, y, z, ct)$ to $(x', y', z', ct')$ then

$$g'_{\mu\nu} = (\partial x^{\alpha}/\partial x'^{\mu})(\partial x^{\beta}/\partial x'^{\nu})g_{\alpha\beta} \tag{1.13.2}$$

For the next big step, the **principle of equivalence**, we start from Poisson's equation for gravitation in three-space:

$$(\partial^2 V/\partial x^2) + (\partial^2 V/\partial y^2) + (\partial^2 V/\partial z^2) = 4\pi G\rho \tag{1.13.3}$$

where $V$ is the gravitational potential, $G$ is the universal gravitational constant, and $\rho$ is the density of matter. The gravitational potential $V$ is replaced by a momentum-energy contravariant tensor $T_{\mu\nu}$:

$$T^{\mu\nu} = \begin{pmatrix} T^{11} & T^{12} & T^{13} & T^{14} \\ T^{21} & T^{22} & T^{23} & T^{24} \\ T^{31} & T^{32} & T^{33} & T^{34} \\ T^{41} & T^{42} & T^{43} & T^{44} \end{pmatrix} = \begin{pmatrix} P_{xx} & P_{xy} & P_{xz} & cg_x \\ P_{yx} & P_{yy} & P_{yz} & cg_y \\ P_{zx} & P_{zy} & P_{zz} & cg_z \\ cg_x & cg_y & cg_z & c^2\rho \end{pmatrix} \tag{1.13.4}$$

Einstein introduced a contracted tensor $R_{\mu\nu}$:

$$R_{\mu\nu} \equiv \{\mu\sigma, \alpha\}\{\alpha\nu, \sigma\} - \{\mu\sigma, \alpha\}\{\alpha\sigma, \sigma\} + (\partial/\partial x^{\alpha})\{\mu\sigma, \sigma\} - (\partial/\partial x^{\alpha})\{\mu\nu, \sigma\} \tag{1.13.5}$$

where the three-index (nontensor) symbols of the type $\{\mu\nu, \sigma\}$ are defined as

$$\{\mu\nu, \sigma\} \equiv (1/2)g^{\sigma\lambda}\left[\left(\partial g_{\mu\lambda}/\partial x^{\nu}\right) + (\partial g_{\nu\lambda}/\partial x^{\mu}) - \left(\partial g_{\mu\nu}/\partial x^{\lambda}\right)\right]$$

Then the Einstein field equation for general relativity is, finally:

$$R_{\mu\nu} - (1/2)Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \left(8\pi G c^{-4}\right)T_{\mu\nu} \tag{1.13.6}$$

where $R$ is the scalar curvature, $c$ is the speed of light in vacuum, $\Lambda$ is the *ad hoc* important "cosmological constant" (at present $\Lambda$ is thought to be small but positive, since the universe is expanding). Each of the $4 \times 4$ tensors in Eq. (1.13.6) has 16 components, of which 10 can be independent of each other. The value of the numerical factor in Eq. (1.13.6) is

$$\left(8\pi G c^{-4}\right) = 2.073 \times 10^{-48}\ \mathrm{cm}^{-1}\ \mathrm{g}^{-1}\ \mathrm{s}^2 = 2.073 \times 10^{-53}\ \mathrm{m}^{-1}\ \mathrm{kg}^{-1}\ \mathrm{s}^2 \tag{1.13.7}$$

The experimental evidence for Eq. (1.13.6) consists of (i) the advance of the perihelion of the planet Mercury, (ii) the bending of starlight when it passes close by the Sun, (iii) a (small) gravitational shift in stellar spectral lines, and (iv) a mediating spin-2 boson called the graviton, first detected in 2015.

The Schwarzschild spherically symmetric solution to Eq. (1.13.6) can be obtained by first transforming the Cartesian coordinate system $(x, y, z, t)$ of Eq. (1.13.5) to spherical polar coordinates $(r, \theta, \phi, -t)$:

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - c^2\mathrm{d}t^2 = g_{11}\mathrm{d}r^2 + g_{22}\mathrm{d}\theta^2 + g_{33}\mathrm{d}\phi^2 + g_{44}\mathrm{d}t^2 \tag{1.13.8}$$

After exploiting the assumed spherical symmetry and some labor, the metric becomes:

$$\mathrm{d}s^2 = \left(1 - 2GM/c^2 r\right)\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right) - c^2\left(1 - 2GM/c^2 r\right)\mathrm{d}t^2 \tag{1.13.9}$$

This yields a singularity in space and time when the Schwarzschild radius $r_{SS}$ is reached:

$$r_{SS} = 2GM/c^2 \tag{1.13.10}$$

This Schwarzschild singularity (discussed in Problem 1.4.3) around a massive stationary black hole, constitutes an **event horizon**: all photons trying to leave the black hole are sucked back in by the intense gravity, while all photons approaching it from afar are also sucked in.



**Figure 1.19** False-color image of the Event Horizon EHT2017 around a massive black hole in Galaxy M87. *Source:* The Event Horizon Collaboration [1.31].

After decades-long theoretical speculation of what an event horizon would look like, 8 Earth-based radio telescopes around the world were used as interferometers to see the first event horizon (dubbed EHT2017), observed at 200 GHz (infrared) around a rotating massive black hole of mass $6.5 \times 10^6$ solar masses and distant $4 \times 10^{10}$ km from Earth in galaxy M87 in the Virgo cluster close to the center of the Milky Way [1.32] (Figure 1.19).

> **Sideline 1.18** The key algorithm for analyzing the data collected by the Event Horizon project was designed by K.L. Bouman,[129] a graduate student in computer science.

A small side-trip will present the classical Doppler[130] effect, which explains why the frequency $f_0$ emitted by a source (train whistle or light photon) has a different frequency $f$ measured at a receiver if there is relative motion. For example, the source emits at the frequency $f_0$ either a photon at speed of light $c = 3 \times 10^8$ m s$^{-1}$ or a whistle sound that travels in air at the speed $c = 3.43 \times 10^2$ m s$^{-1}$. If the source and receiver travel at speed $v_s$ and the receiver travels at a speed $v_r$ (both relative to a stationary object and both collinear with $c$), then the Doppler-effect frequency $f$ detected at the receiver will be given by

$$f/f_0 = (c \pm v_r)/(c \pm v_s) \tag{1.13.11}$$

The plus sign is used in the numerator if the receiver moves at speed $v_r$ toward the source; the plus sign is used in the denominator if the source moves with speed $v_s$ away from the receiver. So, the frequency $f$ will be smaller than $f_0$ if either source or receiver move away from each other. Since the speeds of light and sound are both typically much larger than $v_r$ or $v_s$, an approximate formula is used:

$$f/f_0 \approx v_{wr}/v_{ws} \approx v_r/c \tag{1.13.12}$$

where $v_{ws}$ and $v_{wr}$ are the speeds for the wave relative to the source or the receiver, respectively.

Typically, $v_{wr} \approx v_r$ and $v_{ws} \approx c$ are assumed. Next, we return to discussing general relativity.

The emission spectrum of excited hydrogen atoms from far-away galaxies speeding away from ours observational platform on Earth at a speed $v_{Doppler}$ is red-shifted by the Doppler effect; this shift, coupled with an estimated distance $d$ of the source from Earth, yields the very important Hubble[131] parameter $H$:

$$H = v_{Doppler}/d \tag{1.13.13}$$

that also allows guesses for the age of the universe and its size (Problem 1.13.1).

The question of whether the universe is unstable or stable was addressed by several scientists: Friedmann,[132] Lemaitre,[133] and Hubble. A density $\rho_{critical}$ for matter in the universe that is critical for its long-term stability is given by the Friedmann equation:

$$\rho_{critical} = 3H^2/8\pi G \tag{1.13.14}$$

This $\rho_{critical}$ can be calculated in terms of the number of hydrogen atoms in a cubic meter (Problem 1.13.2). Finally, Doppler shifts have also been measured in the microwave emission of electron spin-flips of H atoms in distant galaxies (Problem 1.13.3).

129 Kathleen Louise Bouman (1989–).
130 Christian Doppler (1803–1853).
131 Edwin Powell Hubble (1889–1953).
132 Alexander Alexandrovich Friedmann (1888–1925).
133 The Reverend Georges Henri Joseph Édouard Lemaître (1894–1966).

**Problem 1.13.1** The emission spectrum from excited-state hydrogen atoms located in a source in space at an estimated distance $d_{\text{source}}$ from Earth and moving away at a speed $v_{\text{Doppler}}$ relative to Earth is red-shifted (decreased in energy) by the Doppler effect. If $d_{\text{source}}$ is known or estimated, then $v_{\text{Doppler}} = Hd_{\text{source}}$, where $H$ is the Hubble parameter. The present estimate for the Hubble parameter is $H = 67.8$ km s$^{-1}$ megaparsec$^{-1}$. If 1 parsec $= 3.0860 \times 10^{16}$ m, compute $H$ in units of 1/seconds, and discuss the significance of $1/H$ and of $c/H$, where $c$ is the speed of light in vacuum.

**Problem 1.13.2** The density of the universe can be estimated from Eq. (1.13.14) , where $H$ is the Hubble parameter (not strictly a constant) $H = 67.8$ km s$^{-1}$/megaparsec $= 2.198 \times 10^{-18}$ s$^{-1}$ that estimates the rate of expansion in the universe, and $G$ is the universal gravitational constant $G = 6.674 \times 10^{-11}$ m$^3$ kg$^{-1}$ s$^{-2}$. The critical density is thought to be that density at which the universe will halt its expansion, albeit only after infinite time. Compute this critical density.

**Problem 1.13.3** The hyperfine transition in the H spectrum at 21 cm = 420 MHz is due the flip of the electron spin ($S = 1/2$, $m_S = +1/2$ to $m_S = -1/2$) from parallel to antiparallel, relative to the proton spin ($I = 1/2$, $m_I = +1/2$). This transition has been called "the song of hydrogen." This transition can be Doppler-shifted to lower energies (200–9 MHz) if the H cloud moves away from Earth. Compute the relative Doppler velocities $V$ for these two frequencies $f_0$ and $f$.

## 1.14 Elements of Optics

We first discuss Young's double-slit experiment (1803). Start with a source of light $S$ that creates a beam of monochromatic light of wavelength $\lambda$ (either a modern monochromatic laser or a multichromatic source followed by a wavelength-range-limiting diffraction-grating or prism) (Figure 1.20). This light impinges on a single slit, then goes through a double slit with slit-slit separation $h$, followed (in a modern version) by a photographic plate or fluorescent screen at a distance $D$ from the double slit (pinholes would be acceptable, but the analysis is simpler assuming slits of finite width and infinite height). The light from source $S$ travels a path $r_1$ from the upper slit to the screen, and a different path $r_2$ from the lower slit to the same point on the screen $P$. When two waves with vector electric fields $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ interfere with each other, then the intensity $I$ of their vector sum $\boldsymbol{E}_1 + \boldsymbol{E}_2$ can be written as

$$I = \langle E^2 \rangle = \langle (\boldsymbol{E}_1 + \boldsymbol{E}_2) \cdot (\boldsymbol{E}_1 + \boldsymbol{E}_2) \rangle = \langle \boldsymbol{E}_1^2 \rangle + \langle \boldsymbol{E}_2^2 \rangle + 2\langle \boldsymbol{E}_1 \cdot \boldsymbol{E}_2 \rangle \tag{1.14.1}$$

This can be rewritten in terms of the individual intensities $I_1$ and $I_2$:

$$I = I_1 + I_2 + 2(I_1 I_2)^{1/2} \cos \delta \tag{1.14.2}$$

The intensity at $P$ is modulated by this **phase difference** $\delta$ between the two paths:

$$\delta = \phi_2 - \phi_1 + k(r_2 - r_1) \equiv \Delta\phi + (2\pi/\lambda)(r_2 - r_1) \tag{1.14.3}$$

We seek convenient measurables. The angles are shown geometrically as

$$\sin \theta_1 = (r - r_1)/(h/2) \tag{1.14.4}$$

$$\sin \theta_2 = (r_2 - r)/(h/2) \tag{1.14.5}$$

If the observation point is at a distance $x$ such that $x < D$, so that angles $\theta_1$ and $\theta_2$ are small, then $\sin \theta_1 \approx \tan \theta_1$ and $\sin \theta_2 \approx \tan \theta_2$, so from equivalent triangles:

$$\sin \theta_1 \approx \tan \theta_1 = x/D \tag{1.14.6}$$

$$\sin \theta_2 \approx \tan \theta_2 = (x + h/2)/D. \tag{1.14.7}$$

If also $h < x < D$, then, ignoring a term in $h^2$, we get:

$$r_2 - r \approx xh/2D \tag{1.14.8}$$

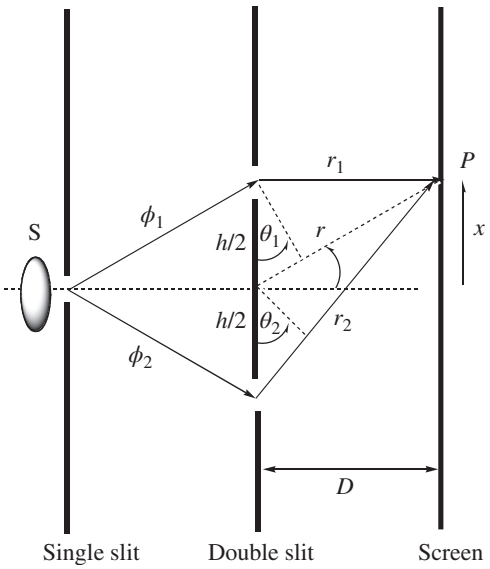



**Figure 1.20** Young double-slit experiment.

similarly:

$$r - r_1 \approx xh/2D$$

Substituting the sum of these last two results into Eq. (1.14.3) yields:

$$\delta = \Delta\phi - 2\pi xh/\lambda D \tag{1.14.9}$$

**Problem 1.14.1** Prove Eq. (1.14.8).

Whenever the argument $\delta$ of the cosine function changes by $2\pi$, the detected output on the screen at $P$ goes from light to light, or from dark to dark. Thus, the condition for light maxima in the Young double-slit experiment is

$$(2\pi x_1 h/\lambda D - \Delta\phi) - (2\pi x_2 h/\lambda D - \Delta\phi) = 2\pi \tag{1.14.10}$$

which simplifies to

$$(x_1 - x_2) = \lambda D/h \tag{1.14.11}$$

The above discussion implicitly obeys **Huygens'**[134] **principle**, that each point on a spherical wavefront can be regarded as the source of a secondary wavelet (another spherical wave), as well as **Fermat's**[135] **principle of least time**.

The **index of refraction** $\boldsymbol{n}$ for a given medium at any frequency is defined by the speed of light in that medium, $v$, divided by the speed of light in vacuum, $c$:

$$n \equiv c/v \tag{1.14.12}$$

Here are some values, measured at the yellow Na D line (wavelength $\lambda = 589$ nm): for vacuum, $n = 1$ by definition; for air $n = 1.000\,294$; for $CO_2$ $n = 1.000\,449$; for glass $n = 1.33$; for water $n = 1.333$; for most other materials, $1 < n < 2$.

If there is absorption within the medium, then the refractive index becomes complex, as discussed earlier, Eq. (1.9.35) can be rewritten as:

$$\aleph \equiv n - ik \tag{1.14.13}$$

where $n$ and $k$ are real but $\aleph$ is complex. If the medium is anisotropic, then $n$ becomes a $3 \times 3$ tensor $\boldsymbol{n}$ with three principal-axis diagonal values $n_a$, $n_b$, and $n_c$ localized within the crystal or oriented polymer.

When a monochromatic parallel beam of light propagating in a medium of one refractive index arrives at a second medium of different refractive index (real or complex scalar, or tensor), several things can happen: (i) reflection back into the first medium, (ii) refraction, (iii) anisotropic refraction, or (iv) absorption into the second medium. To be specific, let an electromagnetic wave travel in the direction $\boldsymbol{u}$ (Figure 1.21). Define the interface (dividing) plane between medium 1 (with scalar real index of refraction $n_1$) and medium 2 (with scalar real index of refraction $n_2$) as the $xy$ plane, and the direction normal to the dividing plane as the $z$-axis. Call the $xz$ plane the plane of incidence (this plane, normal to the dividing $xy$ plane, contains the vector $\boldsymbol{u}$). Without loss of generality, let the $\boldsymbol{u}$ vector make an angle $\theta_i$ with the $z$-axis:

$$u = \boldsymbol{e_x} \sin\theta_i + \boldsymbol{e_z} \cos\theta_i \tag{1.14.14}$$

The reflected beam $\boldsymbol{u}'$ makes the angle of reflection $\theta_r$ with the negative $z$-axis and the angle $\phi_r$ with the $x$-axis:

$$u' = \boldsymbol{e_x} \sin\theta_r \cos\phi_r + \boldsymbol{e_y} \sin\theta_r \sin\phi_r - \boldsymbol{e_z} \cos\theta_r \tag{1.14.15}$$

and the refracted (transmitted) beam $\boldsymbol{u}''$ makes the angle of refraction $\theta_t$ with the $z$-axis and the angle $\phi_t$ with the $x$-axis:

$$u'' = \boldsymbol{e_x} \sin\theta_t \cos\phi_t + \boldsymbol{e_y} \sin\theta_t \sin\phi_t + \boldsymbol{e_z} \cos\theta_t \tag{1.14.16}$$

The electric field vectors are then, respectively:

$$\boldsymbol{E} = \boldsymbol{E}_0 \exp[i\omega_i t - (x \sin\theta_i + z\cos\theta_i)n_1/c] \tag{1.14.17}$$

$$\boldsymbol{E}' = \boldsymbol{E}'_0 \exp[i\omega_r t - (x \sin\theta_r \cos\phi_r + y \sin\theta_r \sin\phi_r - z\cos\theta_r)n_1/c + i\delta_r] \tag{1.14.18}$$

134 Christiaan Huygens (1629–1695).
135 Pierre de Fermat (1601–1665).

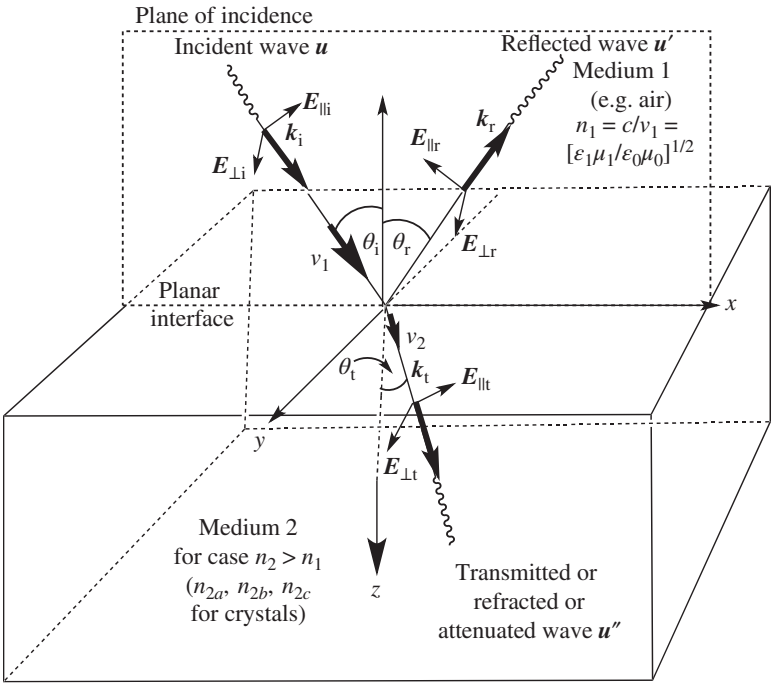



**Figure 1.21** Incident wave $\boldsymbol{u}$, reflected wave $\boldsymbol{u}'$, and refracted wave $\boldsymbol{u}''$. The index of refraction, or refractive index, of medium 1 (top) is $n_1$; the index of refraction of medium 2 (bottom) is $n_2$ (shown for case $n_2 > n_1$). The incident plane contains the incident and the reflected beam wave vectors. The parallel (||) and perpendicular (⊥) polarization components of the electric field are defined relative to the plane of incidence.

$$\boldsymbol{E}'' = \boldsymbol{E}_0'' \exp\left[\mathrm{i}\omega_t t - (x \sin\theta_t \cos\phi_t + y \sin\theta_t \sin\phi_t + z\cos\theta_t) n_2/c + \mathrm{i}\delta_t\right] \tag{1.14.19}$$

where, for generality, changes in the angular frequency $\omega_r$, $\omega_t$ and phase shifts $\delta_r$ and $\delta_t$ have been allowed. The magnetic field vectors are:

$$\boldsymbol{H} = (\varepsilon_1/\mu_1)^{1/2}\boldsymbol{u} \times \boldsymbol{E} \tag{1.14.20}$$

$$\boldsymbol{H}' = (\varepsilon_1/\mu_1)^{1/2}\boldsymbol{u}' \times \boldsymbol{E}' \tag{1.14.21}$$

$$\boldsymbol{H}'' = (\varepsilon_2/\mu_2)^{1/2}\boldsymbol{u}'' \times \boldsymbol{E}''' \tag{1.14.22}$$

At the interface the boundary conditions on the field amplitudes must be:

$$E_x + E_x' = E_x'', E_y + E_y' = E_y'' \text{ and } H_x + H_x' = H_x'', \; H_y + H_y' = H_y'' \tag{1.14.23}$$

which can be satisfied if and only if the following five conditions are met simultaneously:

i) $\omega_i = \omega_r = \omega_t$ : no change in frequency,
ii) $\delta_r$ and $\delta_t$= either 0 or $\pi$ radians ($\delta_r = \pi$ for reflection, $\delta_t = 0$ for refraction),
iii) $\phi' = \phi'' = 0$: both reflected and refracted rays must lie in the plane of incidence,
iv) $\sin\theta_i = \sin\theta_r$: angle of incidence = angle of reflection,
v) $n_1 \sin\theta_i = n_2 \sin\theta_t$: this is Snell's[136] law.

Thus, an electromagnetic wave reflected at an interface between media of different refractive indices undergoes a phase shift of $\pi$ radians; when it is refracted, it suffers no phase shift. The fifth condition above is **Snell's law** of refraction:

$$n_{12} \equiv n_2/n_1 = \sin\theta_i/\sin\theta_t \tag{1.14.24}$$

136 Willebrord Snell van Royen (1591–1626).

When these conditions are met, one can evaluate the transmission ($t$) and reflection ($r$) coefficients [1.29].

For perpendicular ($N$, $\perp$, $\sigma$, or senkrecht) polarization ($\boldsymbol{E}$ vector perpendicular to plane of incidence but parallel to the dividing plane) the transmission coefficient $t_\perp$ and the reflection coefficient $r_\perp$ are given by:

$$t_\perp = 2/(1 + \mu_1 \tan\ \theta_i/\mu_2 \tan\ \theta_t) \approx 2\cos\theta_i \sin\theta_t/\sin(\theta_I + \theta_t) \tag{1.14.25}$$

$$r_\perp = (1 - \mu_1 \tan\ \theta_i/\mu_2 \tan\ \theta_t)/(1 + \mu_1 \tan\ \theta_i/\mu_2 \tan\ \theta_t) \approx -\sin(\theta_i - \theta_t)/\sin(\theta_i + \theta_t) \tag{1.14.26}$$

For parallel ($P$, $\pi$, || or waagerecht) polarization ($\boldsymbol{E}$ vector in the plane of incidence), the transmission $t_\parallel$ and reflection coefficients $r_\parallel$ are:

$$\begin{aligned} t_\parallel &= 2\cos\theta_i \sin\theta_t/(\cos\theta_i \sin\theta_t + (\mu_1/\mu_2)\sin\theta_i \cos\theta_i) \\ &\approx 2\cos\theta_i \sin\theta_t/\sin(\theta_i + \theta_t)\cos(\theta_i - \theta_t) \end{aligned} \tag{1.14.27}$$

$$\begin{aligned} r_\parallel &= (\mu_1 \sin 2\theta_i/\mu_2 - \sin\theta_t)/(1 + \mu_1 \tan\ \theta_i/\mu_2 \tan\ \theta_t) \\ &\approx -\tan(\theta_i - \theta_t)/\tan(\theta_i + \theta_t) \end{aligned} \tag{1.14.28}$$

Equivalently, **Fresnel's**[137] **formulas** [1.29] give the parallel (||) and perpendicular ($\perp$) field amplitude components of the reflected ($E'$) and refracted ($E''$) light beams:

$$E'_\parallel = E_\parallel r_\parallel = E_\parallel \tan(\theta_i - \theta_t)/\tan(\theta_i + \theta_t) \tag{1.14.29}$$

$$E'_\perp = E_\perp r_\perp = -E_\perp \sin(\theta_i - \theta_t)/\sin(\theta_i + \theta_t) \tag{1.14.30}$$

$$E''_\parallel = E_\parallel t_\parallel = E_\parallel 2\sin\theta_i \cos\theta_t/[\sin(\theta_i - \theta_t)\sin(\theta_i + \theta_t)] \tag{1.14.31}$$

$$E''_\perp = E_\perp t_\perp = E_\perp 2\cos\theta_i \sin\theta_t/\sin(\theta_i + \theta_t) \tag{1.14.32}$$

**Problem 1.14.2** Derive Fresnel's formulas, Eqs. (1.14.29)–(1.14.32).

To obtain energy ratios, we use the Poynting vectors, which involve the squares of the electric field amplitudes, and obtain the reflectivity $R$ and the transmittivity $T$ (where $R + T \equiv 1$):

$$R = |E''|^2/|E|^2 \tag{1.14.33}$$

and

$$T = n_2 \cos\theta_t |E''|^2/n_1 \cos\theta_i |E|^2 \tag{1.14.34}$$

If light goes from a low-index medium to a higher-index medium ($n_2 > n_1$), then the reflectivity $R$ becomes large, as Snell's law "fails" beyond relatively larger incident angles $\theta_i$; if the light goes from a higher-index medium to a lower-index medium ($n_1 > n_2$), then the reflectivity becomes 1.0 beyond relatively smaller incident angles [1.29]. This **total internal reflection** can be understood from Snell's law (1.14.24), where the angles $\theta_i$ and $\theta_t$ are defined only between 0 and $\pi/2$ radians (Figures 1.22 and 1.23): as $\theta_i$ increases to a critical value $\theta_c$, $\theta_t$ reaches $\pi/2$ (this $\theta_c$ is the **critical angle of incidence**), Snell's law fails, there is **no** refracted beam, and all light is reflected back into the same medium (total internal reflection back into medium 1).

A different phenomenon is **Brewster's**[138] **angle** for maximum parallel polarization. As the angle of incidence $\theta_i$ is changed systematically, the relative intensities of the reflected and refracted rays change. At a critical angle $\theta_B$, called **Brewster's angle**, the reflected beam for parallel polarization has **zero** intensity ($T = 0$), and all the energy of the parallel-polarized incident wave goes into the refracted wave. If $\theta_B = \theta_i = \pi/2 - \theta_t$ then $\sin\ \theta_t = \cos\ \theta_i$ and Snell's law, Eq. (1.14.24), reduces to Brewster's law:

$$n_{12} = n_2/n_1 = \tan\ \theta_B \tag{1.14.35}$$

137 Augustin-Jean Fresnel (1788–1827).
138 Sir David Brewster (1781–1868).

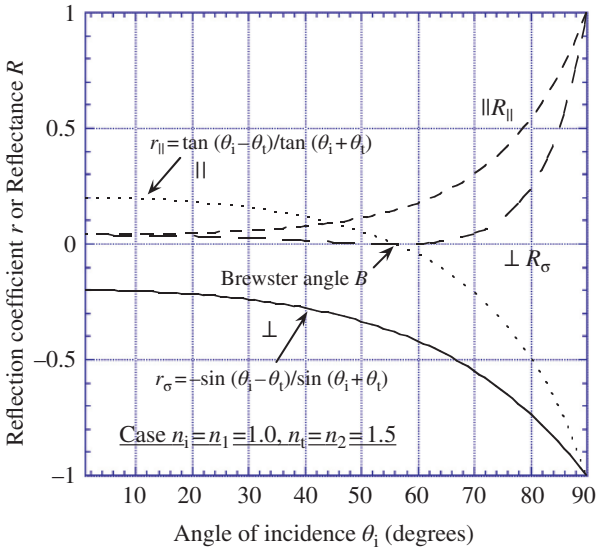



**Figure 1.22** Reflectivity $R$ and reflection coefficient $r$ for $n_i = n_1 = 1.0 < n_t = n_2 = 1.5$.

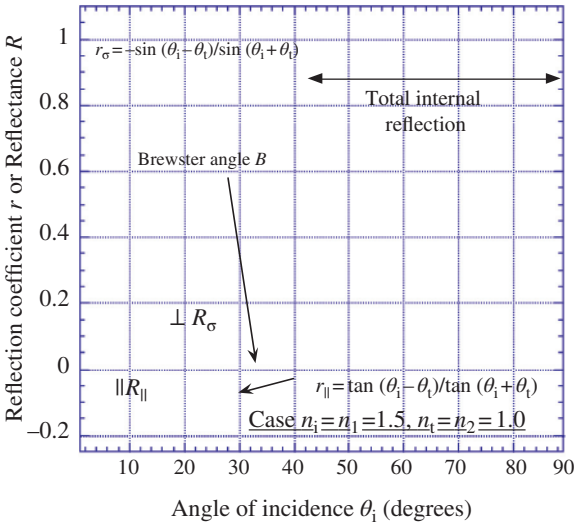



**Figure 1.23** Reflectivity $R$ and reflection coefficient $r$ for $n_i = n_1 = 1.5 > n_t = n_2 = 1.0$.

This reflected beam intensity goes to zero only for parallel polarization, as can be seen in Eq. (1.14.29) when $\tan(\vartheta_i + \vartheta_t)$ becomes infinite, or $r_P$ goes to zero. When Eq. (1.14.29) holds, the unpolarized incident light $\boldsymbol{u}$ will yield a **plane-polarized refracted** ray $\boldsymbol{u}''$ at Brewster's angle $\theta_B$ (the refracted ray $\boldsymbol{u}''$, at 90° to the reflected ray, will be partially polarized, but very weak). The light intensity refracted into medium 2 has the maximum relative intensity at Brewster's angle, but this phenomenon can be seen for a few degrees around the Brewster angle.

**Problem 1.14.3** Show that the ratio $R$ of the reflected intensity $I'$ to the intensity $I$ at normal incidence ($\theta = 0$) is [1.29]

$$R = I'/I = E'^2/E^2 = (n_{12} - 1)^2/(n_{12} + 1)^2 \tag{1.14.36}$$

**Problem 1.14.4** Consider a complex index of refraction $N$ for a metal [1.29]:

$$N \equiv n(1 - ik) \tag{1.14.37}$$

(there are many different sign conventions). In metals, for instance, $n$ and $k$ become functions of the dielectric constant $\varepsilon$, the magnetic permeability $\mu$, the electrical conductivity $\sigma$, and the light frequency $\nu$:

$$n^2 - n^2k^2 = \varepsilon\mu \tag{1.14.38}$$

$$n^2k = \mu\sigma/\nu \tag{1.14.39}$$

Show that Snell's law must now be modified to:

$$n_{12}(1 - ik_{12}) = \sin\theta / \sin\theta''$$

### 1.14.1 Jones Vector

A TEM wave propagating along $z$ has an electric field vector $\boldsymbol{E}$ in the $xy$ plane, and its **polarization direction** can be described by a normalized **Jones**[139] **vector** (1941):

$$|\psi\rangle \equiv \begin{pmatrix} E_x(t)/\sqrt{E_x(t)^2 + E_y(t)^2} \\ E_y(t)/\sqrt{E_x(t)^2 + E_y(t)^2} \end{pmatrix} \tag{1.14.40}$$

Thus, if light is linearly polarized along $x$, along $y$, and at 45° from both $x$ and $y$, the Jones vectors are, respectively:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \tag{1.14.41}$$

### 1.14.2 Anisotropic Indices of Refraction

Crystals can have anisotropic indices of refraction $\boldsymbol{n}$, with differing optical properties. The $3 \times 3$ index of refraction tensor $\boldsymbol{n}$ will typically have at most three principal axis values. Values for liquids, glasses, and relevant crystal structures are given in Table 1.15. Refractive indices vary considerably with wavelength, particularly in the visible region of the electromagnetic spectrum (see discussion of prisms below): that is why the frequency or wavelength of the radiation used to measure $\boldsymbol{n}$ must be specified.

**Isotropic (cubic or rhombohedral) minerals**: halite (NaCl); fluorite ($CaF_2$); garnets ($X_3Y_2(SiO_4)_3$, where X = Mg, Mn, Fe(II), Ca, and Y = Al, Fe(III), Cr; periclase (MgO)) have a single index.

**Anisotropic minerals** (**uniaxial** or **biaxial**) differ from isotropic minerals because they exhibit **birefringence**: (i) the velocity of light varies depending on the direction through the mineral; (ii) they show **double refraction**; (iii) the index of refraction $\boldsymbol{n}$ is a $3 \times 3$ tensor, with two or three diagonal values $n_a$, $n_b$, and $n_c$. When light enters an anisotropic mineral, it is split into **two rays of different velocity**, which vibrate at right angles to each other. In anisotropic minerals, there are one or two directions through the mineral, along which light behaves **as if** the mineral were isotropic: this (these) direction(s) is(are) referred to as the **optic axis (axes)**. Birefringence can also arise in certain rare magnetic materials with anisotropic magnetic permeabilities. Many colored anisotropic materials also display a change of color with orientation: this is **pleochroism** or **dichroism**.

Hexagonal and tetragonal minerals (e.g. calcite $CaCO_3$, quartz $SiO_2$, $MgF_2$, tourmaline, BN) have **one** optic axis and are **optically uniaxial**. Orthorhombic, monoclinic, and triclinic minerals (e.g. sulfur, mica, turquoise, selenite) have **two** optic axes and are **optically biaxial**.

For instance, a light beam traveling through calcite ($CaCO_3$), a uniaxial anisotropic mineral, is split into two rays that vibrate at right angles to each other because they cross lattice atoms or ions with different efficiencies: (i) the **ordinary (or slow) ray**, labeled $\omega$, which does obey Snell's law, Eq. (1.14.24) with refractive index $n_\omega = n_a = 1.658$ and a speed of $1.81 \times 10^8$ m s$^{-1}$ ($=3.0 \times 10^8/1.658$) and (ii) the **extraordinary (or fast) ray**, labeled $\varepsilon$, which does **not** follow Snells'

139 Robert Clark Jones (1916–2004).

**Table 1.15** Index of refraction components $n_a$, $n_b$, and $n_c$ (measured at Na D line, $\lambda$ = 589.29 nm).

| Name | Formula | Type | Optic axis | $n_a$ | $n_b$ | $n_c$ |
|---|---|---|---|---|---|---|
| Water at 20 °C | $H_2O$ | Isotropic liquid | — | 1.333 0 | 1.333 0 | 1.333 0 |
| Benzene at 20 °C | $C_6H_6$ | Isotropic liquid | — | 1.501 | 1.501 | 1.501 |
| Ethanol at 20 °C | $C_2H_5OH$ | Isotropic liquid | — | 1.361 | 1.361 | 1.361 |
| Silicone oil | $(SiO_2)_n$ | Isotropic liquid | — | 1.520 45 | 1.520 45 | 1.520 45 |
| Diamond | C | Isotropic | — | 2.419 | 2.419 | 2.419 |
| Sr titanate | $SrTiO_3$ | Isotropic | — | 2.41 | 2.41 | 2.41 |
| Fused silica | $SiO_2$ | Isotropic (dis.) | — | 1.458 46 | 1.458 46 | 1.458 46 |
| Pyrex® glass | $BSi_xO_y$ | Isotropic (dis.) | — | 1.470 | 1.470 | 1.470 |
| Halite | NaCl | Isotropic | — | 1.516 | 1.516 | 1.516 |
| Water ice | $H_2O$ | Uniaxial | | 1.309 | 1.313 | 1.313 |
| Sellaite | $MgF_2$ | Uniaxial | | 1.378 | 1.390 | 1.390 |
| Quartz | $SiO_2$ | Uniaxial | | 1.544 24 | 1.553 35 | 1.553 35 |
| Wurtzite | ZnS | Uniaxial | | 2.356 | 2.378 | 2.378 |
| Rutile | $TiO_2$ | Uniaxial | | 2.616 | 2.903 | 2.903 |
| Cinnabar | HgS | Uniaxial | | 2.854 | 3.201 | 3.201 |
| Calcite | $CaCO_3$ | Uniaxial | 111 | 1.658 | 1.486 | 1.486 |
| Tourmaline | Complex | Uniaxial | | 1.669 | 1.638 | 1.638 |
| Sapphire | $\alpha$ - $Al_2O_3$ | Uniaxial | | 1.768 1 | 1.759 9 | 1.759 9 |
| Tridymite | $SiO_2$ | Biaxial | | 1.469 | 1.47 | 1.473 |
| Mica (muscovite) | Complex | Biaxial | | 1.560 1 | 1.593 6 | 1.597 7 |
| Turquoise | Complex | Biaxial | | 1.61 | 1.62 | 1.66 |
| Topaz | $Al_2SiO_4(F, OH)_2$ | Biaxial | | 1.619 | 1.62 | 1.627 |
| Sulfur | $S_8$ | Biaxial | | 1.95 | 2.043 | 2.240 |
| Borax | $Na_2B_4O_7 \cdot 10H_2O$ | Biaxial | | 1.447 | 1.47 | 1.472 |
| Lanthanite | Complex | Biaxial | | 1.52 | 1.587 | 1.613 |
| Stibnite | $Sb_2S_3$ | Biaxial | | 3.194 | 4.303 | 4.46 |

Complex = chemical formula too complex for this table; dis.=disordered

law, with $n_\varepsilon = n_b = n_c = 1.486$, and a speed of $2.02 \times 10^8$ m s$^{-1}$(=$3.0 \times 10^8/1.486$). For this extraordinary ray, the electric displacement $\boldsymbol{D}$ and the electric field $\boldsymbol{E}$ are no longer parallel. The difference $\Delta \equiv n_\varepsilon - n_\omega$ is called the **optical retardation**.

The direction of the principal axes of the index of refraction tensor $\boldsymbol{n}$ can be described by the **indicatrix**. For isotropic crystals, the indicatrix is a **sphere**. For **positive uniaxial crystals,** it is a **prolate spheroid** ($n_\varepsilon > n_\omega$); for **negative uniaxial crystals,** it is an **oblate spheroid** ($n_\omega > n_\varepsilon$). For orientations away from the principal axis orientations, the extraordinary ray will have a refractive index $n_{\varepsilon'}$ intermediate between $n_\omega$ and $n_\varepsilon$.

**Ellipsometry** measures the orientation of polarized light undergoing oblique reflection from a sample surface. Linearly polarized light, when reflected from a surface, will become elliptically polarized because of presence of the thin layer of the boundary surface between two media. Dependence between optical constants of a layer and parameters of elliptically polarized light can be found on basis of the Fresnel formulas described above (see Special Topic ST1.10).

**X-ray diffraction** will be discussed in Section 7.3.

### 1.14.3 Mirrors

For reflection from a plane **mirror**, an observer situated at a distance $u$ from it sees a "virtual image" of himself/herself at a distance

$$v = -u \tag{1.14.42}$$

"behind" the mirror; the observer sees in this virtual image a mirror reflection ("left-hand" appears as "right-hand," but "up" remains "up" and "down" is seen as "down"). In the discussion below, we assume that the angle of incidence $\theta_i$ is equal to the angle of reflection $\theta_r$:

$$\theta_i = \theta_r \tag{1.14.43}$$

The relative intensities of the primary and reflected rays are not considered here.

For a **concave mirror** of radius of curvature $R$ (Figure 1.24) a point object at a distance $u$ from the mirror forms a real image at a distance $v$ from the mirror, where

$$1/u + 1/v \approx 2/R \tag{1.14.44}$$

("Real image" means that the image is on the same side of the mirror as the object.) This relationship, known as the "thin-lens formula," is approximate; it holds exactly only when $u = v = R$, i.e. when a point source and its image are both at the center of curvature $C$. Also, if the object is moved to $I$, then the image will form at $O$, i.e. the points $O$ and $I$ are "conjugate," and $u$ and $v$ are called conjugate distances. Finally, if $u$ becomes infinite, $u = \infty$, then $v$ assumes the special value $f$, called the **principal focal length**: all parallel rays coming in from infinity will focus onto the same point $F$. Conversely, a point source at $F$ will create a parallel bundle of rays going out to infinity. In Eq. (1.14.44), setting $u = \infty$ and $v = f$ yields $f = R/2$. Therefore, we can rephrase Eq. (1.14.44) as

$$1/u + 1/v \approx 1/f \tag{1.14.45}$$

Similarly, for the plane mirror, $R = \infty$, Eq. (1.14.44) will reduce to Eq. (1.14.42).

**Problem 1.14.5** Prove Eq. (1.14.44).

When the mirror is convex instead of concave, then the image is "virtual," i.e. it forms on the opposite side as the object. The same Eq. (1.14.44) holds, if we assume $u > 0$ always, but $v < 0$, $R < 0$, and $f < 0$ for the "virtual" case, and $v > 0$, $R > 0$, and $f > 0$ for the "real" case.

The reciprocal of the focal length $f$ (expressed in $m$) is the diopter, which measures the "strength" of a mirror or lens. The images produced by mirrors or lenses can be constructed (Figure 1.25).

The transverse linear magnification $m$, that is, the ratio of the length of transverse length $h' = II'$ of the image to the transverse length of the object $h = OO'$, is obtained from Figure 1.19a: from the congruent triangles $OO'C$ and $CII'$ one sees that $|h'|/|h| = II'/OO' = CI/CO = (R - v)/(u - R)$. However, it is wise to leave $h$ as a positive quantity for a right-side-up object and negative for an inverted (upside down) image. So, we define $m \equiv -h'/h = -(R - v)/(u - R)$, using Eq. (1.14.44) and $R = 2uv/(u + v)$, one gets:

$$\text{transverse linear magnification (spherical mirror)}\ m \equiv -h'/h = -v/u \tag{1.14.46}$$

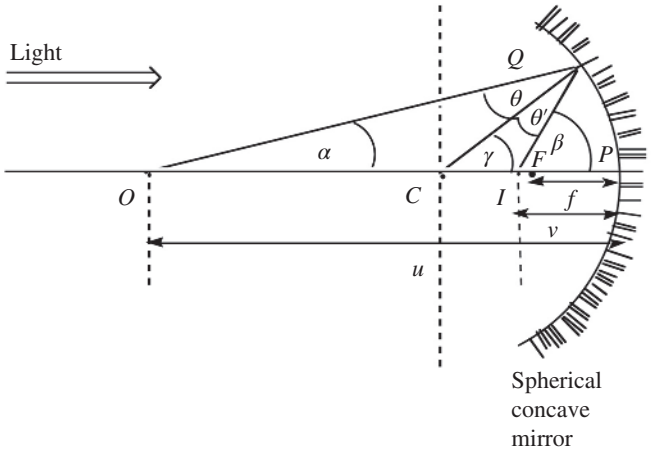



**Figure 1.24** A spherical concave mirror of radius $R = \leftrightarrow CP$ reflects an object at $O$ (mirror-object distance $\leftrightarrow OP \equiv u$) into its image at $I$ (mirror-image distance $\leftrightarrow IP \equiv v$). If object moves to infinity ($u = \infty$), then all rays from that object will converge at the principal focus point $F$: $f = \leftrightarrow FP = (1/2)\ CP = R/2$, i.e. the principal focal length $f$ is half of the radius of curvature.

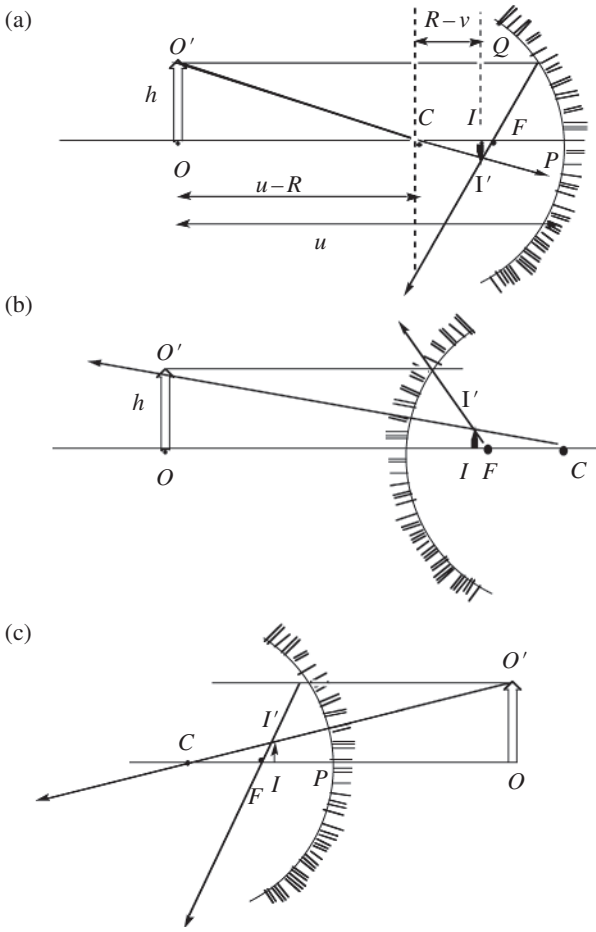



**Figure 1.25** Graphical location of object $OO'$ and image $II'$. (a) Concave mirror, $f > 0$: real object $OO'$ and real image $II'$: draw parallel line from $O'$ to mirror surface (point $Q$). From $Q$ draw an arrow through the focus $F$. From $O'$ draw arrow through center $C$. The two arrows meet at the image point $I'$. (b) Convex mirror, $f < 0$: real object and virtual image. (c) Concave mirror, $f > 0$: virtual object and real image.

**Problem 1.14.6** Complete the proof of Eq. (1.14.46).

For the **longitudinal magnification**, i.e. the magnification along the optical axis of a spherical mirror, one calculates the differential of Eq. (1.14.46): from $1/u + 1/v = 1/f =$ constant, the differential form is $-u^{-2}\,\mathrm{d}u - v^{-2}\,\mathrm{d}v = 0$, whence

$$\text{longitudinal linear magnification (spherical mirror)}\ \mathrm{d}v/\mathrm{d}u = -v^2/u^2 \tag{1.14.47}$$

Other conic sections exist: paraboloidal mirrors, ellipsoidal mirrors, and hyperboloidal mirrors. In paraboloidal mirrors, all rays (from infinity or not) converge at the same focus. In ellipsoidal mirrors, all rays emanating at focus 1 converge at focus 2.

For refraction at a single spherical interface of radius $R$ that separates two media of refractive indices $n_1$ and $n_2$, the **lens equation** is

$$n_1/u + n_2/v = (n_2 - n_1)/R \tag{1.14.48}$$

and the first and second principal focal lengths or foci are defined as $f_2$ (by setting $u = \infty$), and as $f_1$ (by setting $v = \infty$):

$$f_2 = Rn_2/(n_2 - n_1) \tag{1.14.49}$$

$$f_1 = Rn_1/(n_2 - n_1) \tag{1.14.50}$$

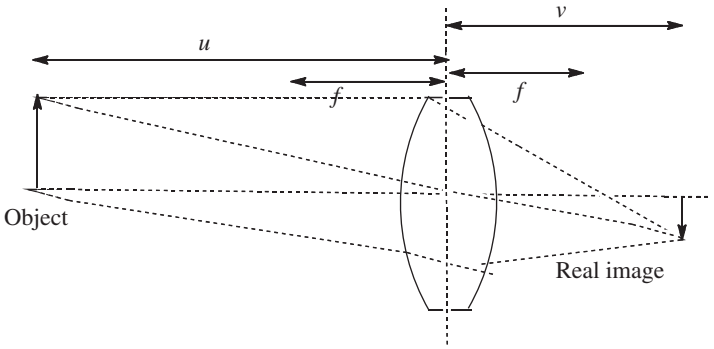



**Figure 1.26** Real image formation for a biconvex lens (positive, converging) with refractive index $n > 1$. As long as $u > f$, a real image forms, upside down, on the opposite side of the lens from the object.

For two thin lenses with spherical segments (Figure 1.26), the two effects are additive, and yield:

$$1/u + 1/v = (n-1)(1/R_1 - 1/R_2) = 1/f \tag{1.14.51}$$

while the transverse linear magnification becomes:

$$m = -v/u$$

For compound lenses, as in a single-lens reflex camera, the focal lengths add as follows:

$$1/f = 1/f_1 + 1/f_2 + \ldots \tag{1.14.52}$$

The **lensmaker's equation** is an elaboration of Eq. (1.14.51):

$$1/f = (n-1)[1/R_1 - 1/R_2 + (n-1)d/(nR_1R_2)] \tag{1.14.53}$$

where $f$ is the **focal length** of the lens, $n$ is the refractive index of the lens material, $R_1$ is the radius of curvature of the lens surface closest to the light source ($R_1 > 0$ if lens surface is convex), $R_2$ is the radius of curvature of the lens surface farthest from the light source ($R_2 < 0$ if the lens surface is concave), and $d$ is the thickness of the lens (distance along the lens axis between the two surface vertices). As above, if $f$ is given in meters, then $1/f$ is in "**diopters**." If $d \ll (R_1, R_2)$ ("thin-lens approximation") then

$$1/f \approx (n-1)[1/R_1 - 1/R_2] \tag{1.14.54}$$

### 1.14.4 Prisms and Gratings

For a prism of refractive index $n$ and refracting angle $\alpha$, there is an angle of incidence $\theta_i$ for which the deviation angle $\delta$ is a minimum: this can be used, e.g. to determine the refractive index $n$ of a liquid placed inside the hollowed prism (Figure 1.27).

We next discuss how to wavelength-select visible radiation. There have been two traditional kinds of optical elements: **prisms** and **gratings** (Figure 1.28). **Prisms exploit the variation of the refractive index with wavelength.** In particular, the **Bunsen**[140] **prism** is a 60° prism, made of fused silica (normal glass absorbs too much light below 350 nm); a single natural quartz crystal is acentric and will thus introduce polarization, so the **Cornu**[141] **biprism** consists of two quartz crystals, 30° each, cut from right-handed quartz and left-handed quartz and glued together. Another design is the **Littrow**[142] **prism**, which uses a 30° cut and a reflective metal mirror on the back.

140 Robert Wilhelm Eberhard Bunsen (1811–1899).
141 Marie Alfred Cornu (1841–1902).
142 Otto von Littrow (1843–1864).

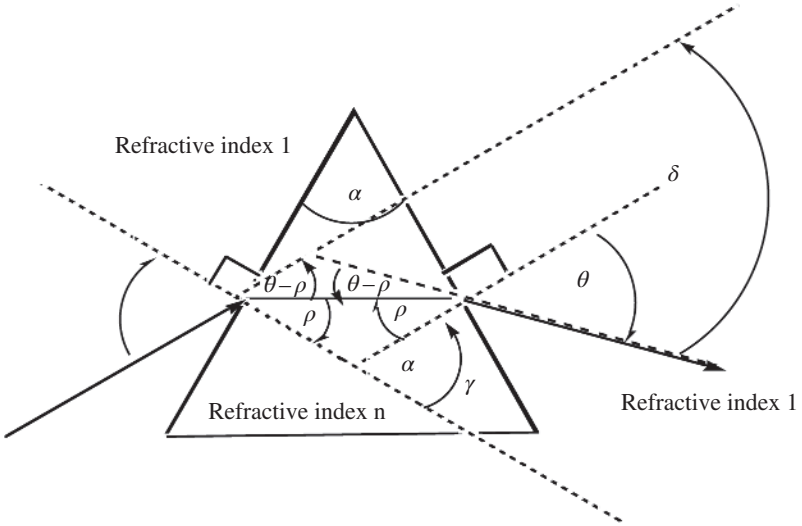



**Figure 1.27** Prism of measured refractive angle $\alpha$ and unknown refractive index $n$ bends light through a measured angle $\delta$, whose minimum is found by scanning the incident angle $\theta$. External angles of triangles yield $\delta = 2(\theta - \rho)$ and $\alpha = 2\rho$, and Snell's law gives $n = \sin\ \theta/\sin\ \rho = \sin[(\alpha + \delta)/2]/\sin(\alpha/2)$.

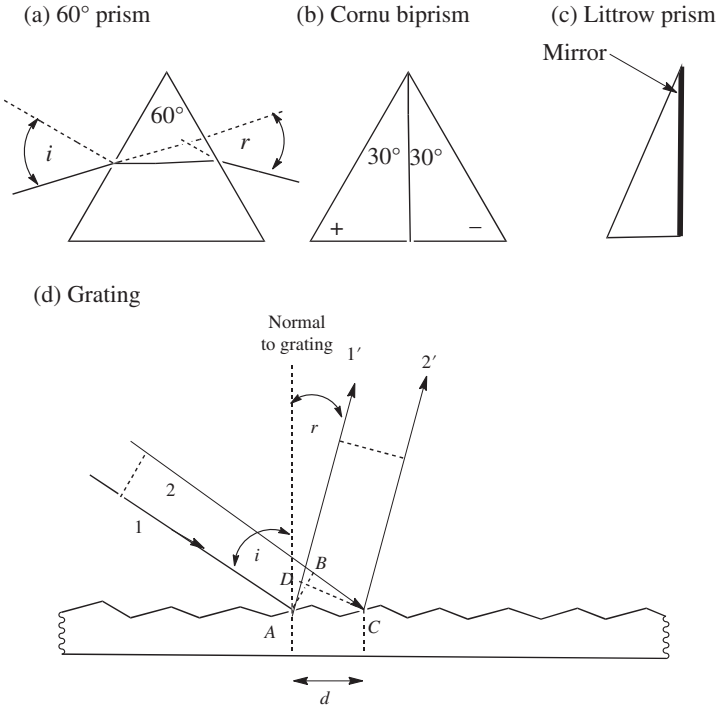



**Figure 1.28** Prisms and grating.

Primary **gratings** are quartz substrates with a set of a few thousand very finely cut grooves, all exactly parallel with a distance $d$ between them: they are labor intensive and expensive to manufacture. **Replica gratings** are much cheaper: a soft polymer block (e.g. poly-methylmethacrylate) is pressed into a primary grating, cured, and dried and then covered with a thin layer of Al or Ni. The **grating equation** resembles Braggs' law (discussed in Section 7.3): a planar wavefront of wavelength $\lambda$ with angle of incidence $i$ is refracted to a smaller angle $r$, but has the wavelets in phase: this requires that the path length difference $AB$ plus $CD$ must be an integer ($n$) times the wavelength $\lambda$:

$$n\lambda = AB + CD = d(\sin i + \sin r) \tag{1.14.55}$$

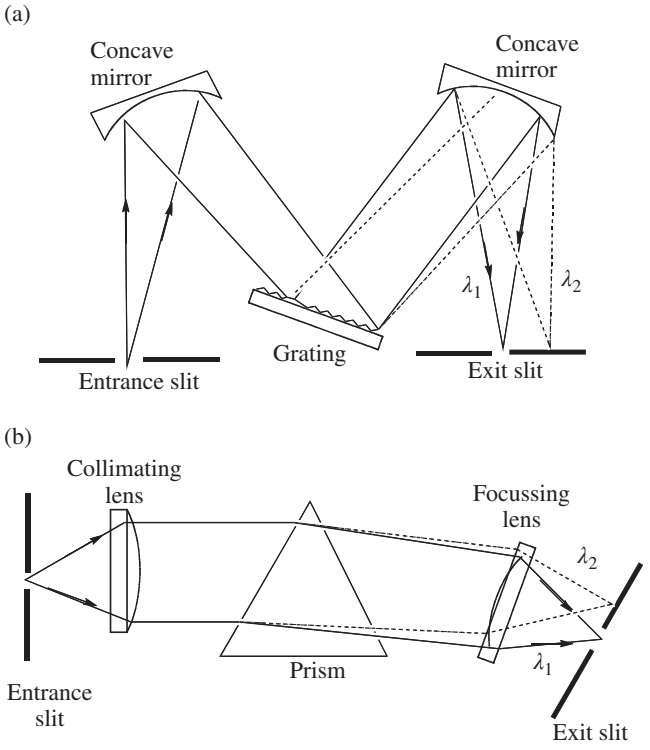



**Figure 1.29** Monochromators: (a) Czerny (Marianus Czerny (1896–1985))–Turner (Arthur Francis Turner (1906–1996)) grating monochromator; (b) Bunsen prism monochromator.

Prisms and gratings are usually rotated mechanically inside a sealed unit called a monochromator Figure 1.29). The light-gathering power of a monochromator is given, as in photographic lens systems, by its *f*-number; its ability to resolve small differences in wavelength $\lambda$ is defined by its angular dispersion,

$$\mathrm{d}r/\mathrm{d}\lambda = nf/d\cos r \tag{1.14.56}$$

or by its reciprocal dispersion $1/D$:

$$1/D \equiv (\mathrm{d}\lambda/\mathrm{d}r) = d\cos r/nf \tag{1.14.57}$$

where $f$ is the focal length, and $d$, $r$, and $n$ are defined in Eq. (1.14.55).

## 1.15 Transforms

Various mathematical transforms and their reverse, or "back" transforms are often used in succession to "clean up" and filter out noise effects from experimental data, by first calculating the full transform, then eliminating the higher-order terms which are ascribed to noise, and finally computing the back transform. They are also very useful for solving differential equations.

Furthermore, transforms are often used only one way, taking data from coordinate space to momentum space, or from coordinate space to time space, etc. In common to all these techniques is the transformation integral:

$$F(k) = \int_{x=a}^{x=b} f(x)K(k,x)\,\mathrm{d}x \tag{1.15.1}$$

where $K(k,x)$ is the "kernel." The limits of integration $a$, $b$ differ from transform to transform. A short list of transforms is

1) Fourier[143] transform [$K(k, x) = \exp(\mathrm{i}kx)$] [1.33]
2) Laplace transform [$K(k, x) = \exp(-kx)$]
3) Hadamard[144] transform
4) Wavelet transform
5) Abel transform [$K(k, x) = 2x(x^2 - k^2)^{-1/2}$]
6) Hankel[145] or Fourier–Bessel transform ($K(k, x) = J_\mathrm{m}(kx)$, where $J_\mathrm{m}(kx)$, is the $m$th-order Bessel function)
7) Hartley[146] transform
8) Hilbert transform[147] [$K(k, z) = \mathrm{PV}(k - z)^{-1}$], where PV = principal value
9) Linear canonical transform
10) Mellin[148] transform [$K(k, z) = k^{z-1}$]
11) Radon[149] transform
12) Stieltjes[150] transform
13) "Sumudu"[151] or "smooth" transform [$K(p, z) = p^{-1}\exp(-x/p)$].

We shall discuss only the first four here.

The **Fourier transform** of a one-dimensional function $f(x)$, is $F(k)$, defined by:

$$F(k) = \int_{x=-\infty}^{x=\infty} f(x)\exp(\mathrm{i}kx)\,\mathrm{d}x \tag{1.15.2}$$

and the corresponding **reverse** or **back Fourier transform** is

$$f(x) = \frac{1}{2\pi}\int_{k=-\infty}^{k=\infty} F(k)\exp(-\mathrm{i}kx)\,\mathrm{d}k \tag{1.15.3}$$

The factor of $2\pi$ is sometimes evenly assigned as a factor $(2\pi)^{1/2}$ for both the forward and inverse transforms. Remember $\exp(\mathrm{i}kx) = \cos(kx) + \mathrm{i}\sin(kx)$. There are real (cosine) and complex (sine) versions of the transform. If the function is three-dimensional, the transform becomes:

$$F(\boldsymbol{k}) = \int_{r=-\infty}^{r=\infty} f(\boldsymbol{r})\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = \int_{z=-\infty}^{z=\infty}\int_{y=-\infty}^{y=\infty}\int_{x=-\infty}^{x=\infty} f(\boldsymbol{r})\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{1.15.4}$$

And the inverse transform is

$$f(\boldsymbol{r}) = \frac{1}{2\pi}\int_{k=-\infty}^{k=\infty} F(\boldsymbol{k})\exp(-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})\,\mathrm{d}k \tag{1.15.5}$$

The Fourier transforms of aperiodic functions $f(x)$ are boring.

If, however, the functions $f(x)$ or $f(\boldsymbol{r})$ are periodic with period $L$ or $\boldsymbol{L}$, then the transforms and inverse transforms become **sums** instead of integrals. If the three Dirichlet[152] conditions are satisfied: (i) periodicity: $f(x) = f(x + L)$; (ii) $f(x)$

143 Jean-Baptiste Fourier (1768–1830).
144 Jacques Solomon Hadamard (1866–1963).
145 Hermann Hankel (1839–1873).
146 Ralph Vinton Lyon Hartley (1888–1970).
147 David Hilbert (1862–1943).
148 Hjalmar Mellin (1854–1933).
149 Johann Karl August von Radon (1887–1956).
150 Thomas Joannes Stieltjes (1856–1894).
151 "Sumudu" may be a pseudonym, just as "Nicolas Bourbaki" was a pseudonym for a group of several French mathematicians.
152 Johann Peter Gustav Lejeune Dirichlet (1805–1859).

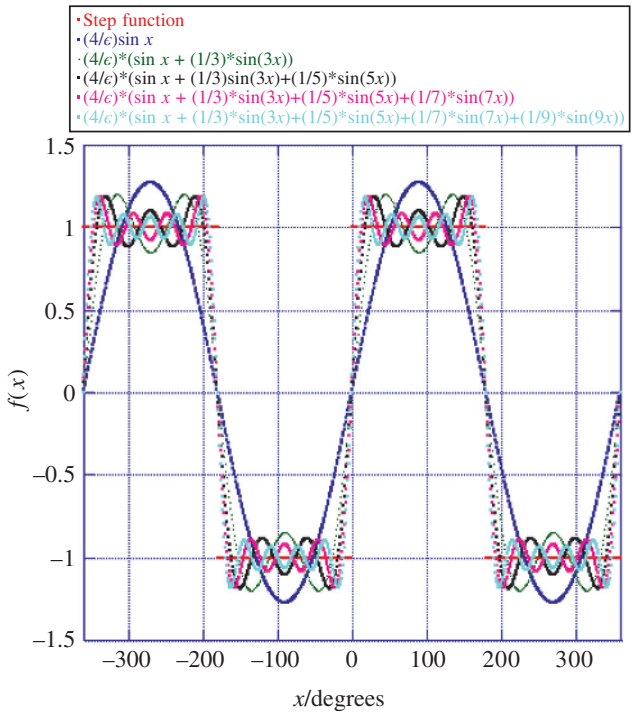



**Figure 1.30** The odd periodic step function $\{f(x) \equiv 1$ for $0 < x < \pi, f(x) \equiv -1$ for $\pi < x < 2\pi\}$ has a Fourier sine series $f(x) = (4/\pi)[\sin x + (1/3) \sin 3x + (1/5) \sin 5x + \ldots]$. The illustration shows how including more and more successive terms in the Fourier series improves the fit to the step function.

is continuous between $0 \leq x \leq L$ except at a finite number of points; and (iii) if $f(x)$ has a finite number of maxima or minima in the period $L$, then the Fourier series for $f(x)$ is given by two infinite series:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{n=\infty} a_n \cos(2\pi nx/L) + \mathrm{i} \sum_{n=1}^{n=\infty} b_n \sin(2\pi nx/L) \tag{1.15.6}$$

where the coefficients are

$$a_n \equiv \frac{2}{L} \int_{x=0}^{x=L} f(x) \cos(2\pi nx/L) \mathrm{d}x \tag{1.15.7}$$

$$b_n \equiv \frac{2}{L} \int_{x=0}^{x=L} f(x) \sin(2\pi nx/L) \mathrm{d}x \tag{1.15.8}$$

Note that if the function is even: $f(x) = f(-x)$, then only the coefficients $a_n$ in the cosine series are nonzero; if the function is odd: $f(x) = f(-x)$, then only the coefficients $b_n$ in the sine series are nonzero. Figure 1.30 shows how including more terms of a Fourier expansion for the odd periodic step function ($f(x) = 1$ for $0 < x < \pi$ radians, $f(x) = -1$ for $\pi < x < 2\pi$, etc.) approaches a reasonable representation of the original step function.

If the sums in Eq. (1.15.6) are truncated at $n = N - 1$, then one has a discrete Fourier transform. The discrete Fourier transform is then:

$$X(k) = \sum_{n=0}^{n=N-1} x_n \exp(2\pi \mathrm{i} kn/N) \quad \{k = 0, 1, 2, \ldots, N-1\} \tag{1.15.9}$$

and the inverse discrete Fourier transform is

$$x_n = \frac{1}{N}\sum_{k=0}^{k=N-1} X(k)\exp(-2\pi ikn/N) \quad \{n = 0, 1, 2, ..., N-1\} \tag{1.15.10}$$

There are trigonometric additive and multiplicative relations for both cos(*nx*) or sin(*nx*) in terms of (ultimately) sin *x* and cos *x*.

$$\sin(nx) = \sin[(n-1)x + x] = \sin[(n-1)x]\cos x + \cos[(n-1)x]\sin x \tag{1.15.11}$$

$$\cos(nx) = \cos[(n-1)x + x] = \cos[(n-1)x]\cos x - \sin[(n-1)x]\sin x \tag{1.15.12}$$

Once a relatively more compute-intensive calculation of sin *x* and cos *x* is finished, the trigonometric expansions with multiplications and sums are relatively fast: this computational advantage was used in the **Beevers**[153]**–Lipson**[154] method of computing Fourier maps in early X-ray structure determinations.

A digitally efficient, if tolerably imprecise, **fast Fourier transform** (FFT), due to Cooley[155] and Tukey,[156] has been implemented in digital computer programs for Fourier transform spectroscopy.

Another very useful result is

$$\frac{1}{2\pi}\int_{k=-\infty}^{k=\infty} \exp(ikx)\,dk = \delta(x) \tag{1.15.13}$$

Very significant feature: the Fourier transform of a Gaussian is another Gaussian.

### 1.15.1 Convolution Theorem

The convolution (German: Faltung, i.e. folding) of a function $f(x)$ times a function with a different origin $g(x)$ is the very useful function $C(y)$ defined by:

$$C(y) \equiv f*g \equiv \int_{x=0}^{x=y} f(x)g(x+y)\,dx \tag{1.15.14}$$

If $C(y) = 0$ for $y < 0$, then the above upper integration limit $y$ can safely go to infinity:

$$C(y) = \int_{x=0}^{x=\infty} f(x)g(x+y)\,dx \tag{1.15.15}$$

The convolution theorem states (i) that the Fourier transform of $C$ is the product of the Fourier transform of $f$ and of the complex conjugate of the Fourier transform of $g$.

There is also another kind of convolution:

$$D(y) \equiv f**g = \int_{x=0}^{x=y} f(x)g(x-y)\,dx \tag{1.15.16}$$

with similar properties. Thus, if the Fourier transform of $f$ is $F$, and if the Fourier transform of $g$ is $G$, then the Fourier transform of $f*g$ (Eq. (1.15.14) is $FG*$ (where $*$ denotes the complex conjugate); the Fourier transform of $f**g$ (Eq. (1.15.16) is $FG$. A similar theorem is valid for Laplace transforms.

**Problem 1.15.1** Show that:

$$\int_{x=0}^{x=L} \exp(-2\pi imx/L)\exp(2\pi inx/L)dx = \delta_{mn}L.$$

153 Cecil Arnold Beevers (1908–2001).
154 Henry Lipson (1910–1991).
155 James William Cooley (1926–2016).
156 John Wilder Tukey (1915–2000).

**Problem 1.15.2** Prove Eq. (1.15.13).

**Problem 1.15.3** Prove **Parseval's**[157] **theorem**: if $F(k) = \int_{x=-\infty}^{x=\infty} f(x) \exp(\mathrm{i}kx)\, \mathrm{d}x$ is the Fourier transform of $f(x)$, and if $f(x)$ is the inverse transform of $F(k)$: $f(x) = (1/2\pi)\int_{k=-\infty}^{k=\infty} F(k) \exp(-\mathrm{i}kx)\, \mathrm{d}k$, then the Fourier transform of $|f(x)|^2$ is $|F(k)|^2$.

**Problem 1.15.4** Show that the three-dimensional Fourier transform of the normalized Gaussian function $f(\boldsymbol{r}) = 2^{3/4}\pi^{-3/4}a^{-3/2}\exp(-r^2/a^2)$ is another Gaussian function, namely $F(\boldsymbol{k}) = (2\pi a^2)^{3/4}\exp(-k^2a^2/4)$.

### 1.15.2 Laplace Transforms

When Eqs. (1.15.3) and (1.15.4) are modified by using a real transform kernel $\exp(-kx)$ instead of $\exp(\mathrm{i}kx)$, then we have the Laplace transform:

$$F(k) = \int_{x=0}^{x=\infty} f(x) \exp(-kx)\, \mathrm{d}x = L[f(x), k] = L[f(x)] \tag{1.15.17}$$

and the corresponding inverse transform is

$$f(x) = \frac{1}{2\pi\mathrm{i}} \int_{k=c-\mathrm{i}\infty}^{k=c+\mathrm{i}\infty} F(k) \exp(kx)\, \mathrm{d}k \tag{1.15.18}$$

This transform is abbreviated as $L[f(x), k]$ or as $L[f(x)]$. The Laplace transform is linear: the transform of a sum of functions is equal to the sum of transforms of each function.

The Laplace transform of the first derivative of function $f'(x) \equiv \mathrm{d}f(x)/\mathrm{d}x$ is given by

$$L[f'(x)] = kL[f(x)] - f(0) \tag{1.15.19}$$

Similarly, for the second derivative:

$$L[f''(x)] = k^2L[f(x)] - f'(0) - kf(0) \tag{1.15.20}$$

$$L\left[\mathrm{d}^2f(x)/\mathrm{d}x^2\right] = kL[\mathrm{d}f(x)/dx] - \mathrm{d}f(0)/\mathrm{d}x \tag{1.15.21}$$

The Laplace transform of the integral of a function is given by

$$L\left[\int_{x=0}^{x=x} f(x)dx\right] = k^{-1}L[f(x)] \tag{1.15.22}$$

The inverse Laplace transform of $L(k)$ is given by

$$L^{-1}[F(k)] = f(z) \tag{1.15.23}$$

Formally, this inverse transform is given by

$$L^{-1}[F(k)] = \frac{1}{2\pi\mathrm{i}} \int_{k=c-\mathrm{i}\infty}^{k=c+\mathrm{i}\infty} F(k) \exp(kx)\, \mathrm{d}k \tag{1.15.24}$$

Some transforms are listed in Table 1.16. Laplace transforms are useful, *inter alia*, in complicated chemical kinetics problems.

**Problem 1.15.5** Solve by Laplace transform methods the classical linear harmonic oscillator differential equation $m\mathrm{d}^2y/\mathrm{d}t^2 = -k_{\mathrm{H}}y(t)$, where $k_{\mathrm{H}}$ is the Hooke's law force constant, with the initial condition $\mathrm{d}y/\mathrm{d}t = 0$ at $t = 0$. Note: Use $p$ for the Laplace transform variable, to avoid confusing it with the Hooke's law force constant $k_{\mathrm{H}}$!

157 Marc-Antoine Parseval (1755–1836).

**Table 1.16** A few Laplace transforms $F(k) \equiv \int_{x=0}^{x=\infty} f(x)\,\mathrm{d}x \exp(-kx) \equiv L[f(x)] \equiv L[(f(x), k)$.

| $f(x)$ | $F(k)$ |
|---|---|
| $a$ | $ak^{-1}$ |
| $f(x)$ | $\int_{x=0}^{x=\infty} \exp(-kx) f(x)\,\mathrm{d}x$ |
| $\mathrm{d}f(x)/\mathrm{d}x$ | $kF(k) - f(0^+)$ |
| $\mathrm{d}^n f(x)/\mathrm{d}x^n$ | $k^n F(k) - \sum_{i=1}^{i=n} k^{i-1} \left[d^{n-i} f(x)/\mathrm{d}x^{n-1}\right]_{x=0+}$ |
| $x$ | $k^{-2}$ |
| $1/x$ | 1 |
| $x^n$ | $n!\,k^{-n-1}$ |
| $x^{-n}$ | $[(n-1)!]^{-1} k^{n-1}$ |
| $1/(x-a)$ | $\exp(ak)$ |
| $\delta(x-a)$ $\{a>0\}$ | $\exp(-ka)$ |
| $\exp(-ax)$ $\{a$ is complex$\}$ | $(k+a)^{-1}$ |
| $x\exp(-ax)$ $\{a$ is complex$\}$ | $(k+a)^{-2}$ |
| $a^{-1}[1-\exp(-ax)]$ | $k^{-1}(k+a)^{-1}$ |
| $(1-ax)\exp(-ax)$ | $k(k+a)^{-2}$ |
| $(a-b)^{-1}[a\exp(-ax) - b\exp(-ax)]$ | $k(k+a)^{-1}(k+b)^{-1}$ |
| $(b-a)^{-1}[\exp(-ax) - \exp(-bx)]$ | $(k+a)^{-1}(k+b)^{-1}$ |
| $(b-a)^{-1}(c-a)^{-1}\exp(-ax) + (a-b)^{-1}(c-b)^{-1}\exp(-bx) + (a-c)^{-1}(b-c)^{-1}\exp(-cx)$ | $(k+a)^{-1}(k+b)^{-1}(k+c)^{-1}$ |
| $a^{-2}[1-\exp(-ax) - ax\exp(-ax)]$ | $k^{-1}(k+a)^{-2}$ |
| $(a-b)^{-1}[a\exp(-ax) - b\exp(-bx)]$ | $k(k+a)^{-1}(k+b)^{-1}$ |
| $(c-b)^{-1}[(a-b)\exp(-bx) - (a-c)\exp(-cx)]$ | $(k+a)(k+b)^{-1}(k+c)^{-1}$ |
| $\sin(ax)$ | $a(k^2+a^2)^{-1}$ |
| $\cos(ax)$ | $k(k^2+a^2)^{-1}$ |
| $\sinh(ax)$ | $a(k^2-a^2)^{-1}$ |
| $\cosh(ax)$ | $k(k^2-a^2)^{-1}$ |
| $[(n-1)!]^{-1}x^{(n-1)}$ | $k^{-n}$ |
| $(\pi x)^{-1/2}$ | $k^{-1/2}$ |
| $2(x/\pi)^{1/2}$ | $k^{-3/2}$ |
| $(y/2)(\pi s x^3)^{-1/2}\exp(-y^2/4sx)$ | $\exp[-(k/s)^{1/2}y]$ |
| $(s/\pi x)^{1/2}\exp(-y^2/4sx)$ | $(s/k)^{1/2}\exp[-(k/s)^{1/2}y]$ |
| $\mathrm{erfc}[(y/2)(sx)^{-1/2}]$ | $k^{-1}\exp[-(k/s)^{1/2}y]$ |
| $\exp(a^2x)\,\mathrm{erfc}(ax^{1/2})$ | $k^{-1/2}(k^{1/2}+a)^{-1}$ |

*Sources:* Erdélyi et al. [1.24]; Steinfeld et al. [1.33]; Bard and Faulkner [1.34].

The **Hadamard transform** is also called the Walsh[158]–Hadamard or Hadamard–Rademacher[159] $2^m \times 2^m$ matrix, consisting of elements that are either 1 or −1, which can be defined recursively;

$$H_m \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} H_{m-1} & H_{m-1} \\ H_{m-1} & -H_{m-1} \end{pmatrix} \quad \text{for } m > 0 \tag{1.15.25}$$

158 Joseph Leonard Walsh (1895–1973).
159 Hans Adolph Rademacher (1892–1969).

and

$$H_0 \equiv 1 \tag{1.15.26}$$

so that:

$$H_1 \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{1.15.27}$$

$$H_2 \equiv \frac{1}{\sqrt{8}}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \tag{1.15.28}$$

This transform is used in quantum computation, in signal processing, and in JPEG-4 compression of visual data.

### 1.15.3 Wavelet Transform

In the continuous wavelet transform, a function $f(x)$ is decomposed into a set of (unspecified) orthonormal and square-integrable basis functions $\psi(s, \tau, x)$:

$$F(s,\tau) = \int f(x)\psi*(s,\tau,x)\,\mathrm{d}x \tag{1.15.29}$$

and the corresponding inverse wavelet transform is

$$f(x) = \int\int F(s,\tau)\psi(s,\tau,x)\,\mathrm{d}\tau\,\mathrm{d}s \tag{1.15.30}$$

This transform uses the *unspecified* single mother wavelet $\psi(s, \tau, x)$ to generate other wavelets by scaling and translation:

$$\psi(s,\tau,x) = s^{-1/2}\psi((x-\tau)/s) \tag{1.15.31}$$

The advantage of this transform is that its kernel $\psi(s, \tau, x)$ is left unspecified. The discrete wavelet transform was invented by Haar,[160] used by petroleum geologists to extract meaningful data from noisy seismograms, and later utilized in JPEG2000 pixel compression.

## 1.16 Contour Integration and Kramers–Kronig Relations

For a complex function $f(z) = f_1(z) + \mathrm{i}f_2(z)$ of the complex variable $z$, which is analytic in the upper half plane of $z$ and decays faster than $|z|^{-1}$, the two Kramers[161]–Kronig[162] relations are:

$$f_1(z) = \frac{1}{\pi}P\int_{t=-\infty}^{t=\infty}\frac{f_2(t)}{t-z}\mathrm{d}t \tag{1.16.1}$$

and

$$f_2(z) = -\frac{1}{\pi}P\int_{t=-\infty}^{t=\infty}\frac{f_1(t)}{t-z}\mathrm{d}t \tag{1.16.2}$$

160 Alfréd Haar (1885–1933).
161 Hendrik Anthony Kramers (1894–1952).
162 Ralph Kronig (1904–1995).

where $P$ denotes the Cauchy[163] principal value. Thus, the real and imaginary parts of $f(z)$ are interrelated. This result can be obtained from the **Cauchy integral formula**: if a function $f$ is analytic everywhere within and on a closed contour $C$, and if $z$ is any point interior to $C$, then

$$f(z) = \frac{1}{2\pi i}\oint_C \frac{f(t)}{t-z}\mathrm{d}t \tag{1.16.3}$$

where the integral is taken in the positive sense (counterclockwise) around $C$. This powerful result has been used to evaluate many a difficult integral.

To derive the Kramers–Kronig relation, the contour $C$, shown in Figure 1.31, follows the real ("$x$") axis, except for a hump over the "**pole**" at $x = t$ (defined as the point where $[z-t]^{-1}$ becomes infinite) and a semicircle in the upper half plane at infinity. The integral is then split into three parts. The length of the segment at infinity increases proportionally to |z|, but its integral component vanishes, if and only if f(z) vanishes faster than $|z|^{-1}$. What is left is the segment along the real axis and the half-circle around the pole:

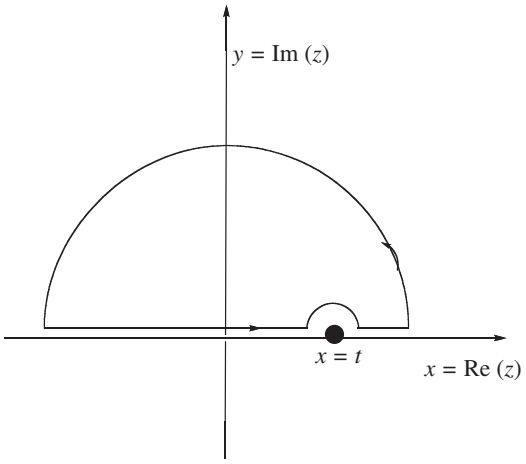



**Figure 1.31** Contour for Kramers–Kronig relations, including the pole at $x = t$.

$$\oint_C \frac{f(t)}{t-z}\mathrm{d}t = -\mathrm{i}\pi f(t) + P\int_{z=-\infty}^{z=\infty} \frac{f_2(z)}{z-t}\mathrm{d}z = 0 \tag{1.16.4}$$

Rearranging yields a compact form of the **Kramers–Kronig relations**:

$$f(z) = \frac{1}{\mathrm{i}\pi}P\int_{z=-\infty}^{z=\infty} \frac{f(z)}{z-t}\mathrm{d}z = 0 \tag{1.16.5}$$

The spectral response function $\chi(\omega)$, a function of the angular frequency $\omega$, is the sum of an in-phase real part $\chi_1(\omega)$, which is an even function of $\omega$, and an out-of-phase (dissipative) imaginary part $\chi_2(\omega)$, which is an odd function of $\omega$:

$$\chi(\omega) = \chi_1(\omega) + \mathrm{i}\chi_2(\omega) \tag{1.16.6}$$

Two Kramers–Kronig relations show $\chi_1(\omega)$ as the integral over the complex part $\chi_2(\omega')$:

$$\chi_1(\omega) = \frac{2}{\pi}P\int_{\omega'=0}^{\omega'=\infty} \frac{\omega'\chi_2(\omega')}{\omega'^2-\omega^2}\mathrm{d}\omega' \tag{1.16.7}$$

and $\chi_2(\omega)$ as the integral over the real part $\chi_1(\omega')$:

$$\chi_2(\omega) = -\frac{2\omega}{\pi}P\int_{\omega'=0}^{\omega'=\infty} \frac{\chi_1(\omega')}{\omega'^2-\omega^2}d\omega' \tag{1.16.8}$$

This seems trivial, but is the very important result that, if one measures $\chi_1(\omega)$, one can calculate $\chi_2(\omega)$ and conversely; put differently, the Kramers–Kronig relations show that the absorptive and dispersive properties of a medium are not independent of each other. An experimental difficulty is that one must truncate the integrations at some maximum measured frequency $\omega$: this may lead to considerable error.

A similar result is obtained for the complex dielectric constant $\varepsilon$, which also consists of a real and even function of $\omega$, $\varepsilon_1(\omega)$, and an odd and imaginary part $\varepsilon_2(\omega)$ (often written as $\kappa$):

$$\varepsilon(\omega) = \varepsilon_1(\omega) + \mathrm{i}\varepsilon_2(\omega) = \varepsilon_1(\omega) + \mathrm{i}\kappa(\omega) \tag{1.16.9}$$

These relationships have been used extensively in nonlinear optics.

163 Augustin-Louis, Baron Cauchy (1789–1857).




**Figure 1.32** Major error! On 22 October 1895 a locomotive and its train overshoot Monparnasse train station, Paris, France: one news-stand lady killed instantly, a few passengers mildly injured. *Source:* Paul Popper/Popperfoto/Getty Images.

## 1.17 Error Analysis

### 1.17.1 Errors

An error associated with a measurement, called the "uncertainty," is usually the smallest reading that can be read or estimated (by interpolation) from an instrument, or it is the "resolution" of that instrument (the smallest interval of the value measured available on that instrument). Assume that you make $N$ measurements of a quantity $x$, and get results (data) $x_1, x_2, ..., x_N$. Of course, you cannot know which datum of these $N$ data is the "true" value. But you can evaluate the mean, or average, trivially:

$$\langle x \rangle \equiv (1/N) \sum_{i=1}^{i=N} x_i \equiv (x_1 + x_2 + x_3 + ... + x_N)/N \tag{1.17.1}$$

There is no guarantee that the mean, $\langle x \rangle$, is the true value, either: the true value is in the hands of God. But you can guess that, barring systematic errors (misreading of instruments, neglect of major factors [Figure 1.32] affecting the measurement), a large number of repetitive measurements will yield a mean or average that may approach the "true" value (as $N$ tends to infinity, $\langle x \rangle$ may tend to the true value). The deviation of any datum from the mean, or its "residual," is

$$d_i \equiv x_i - \langle x \rangle \tag{1.17.2}$$

of course, $d_i$ can be positive or negative.

The average deviation, $\langle d \rangle$ is defined as

$$\langle d \rangle \equiv (1/N) \sum_{i=1}^{i=N} |x_i - \langle x \rangle| \equiv [|x_1 - \langle x \rangle| + [|x_2 - \langle x \rangle| + ...[ + |x_N - \langle x \rangle|]]/N \tag{1.17.3}$$

Here | denotes the absolute sign: $|x| = +x$ if $x > 0$, $|x| = -x$ if $x < 0$. Before we proceed, we must admit that, if $\langle x \rangle$ is known and used, then of the $N + 1$ values $\{x_1, x_2, ..., x_N$, plus $\langle x \rangle\}$, only $N$ are mutually independent. Therefore, the denominator $N$ in Eq. (1.16.3) should be replaced by $N - 1$. Accordingly, the variance $\sigma^2$ is defined by

$$\sigma^2 \equiv [1/(N-1)] \sum_{i=1}^{i=N} |x_i - \langle x \rangle|^2 \tag{1.17.4}$$

its square root is the **estimated standard deviation,** or **sample standard deviation,** $\sigma$, which is universally quoted as a good estimate of the probable error $\varepsilon$ in the measurement:

$$\sigma = (N-1)^{-1/2} \left[ \sum_{i=1}^{i=N} |x_i - \langle x \rangle|^2 \right]^{1/2} \tag{1.17.5}$$

### 1.17.2 Propagation of Errors

**Rule 1:** When adding or subtracting measurements with their associated errors, the *absolute* errors are summed (never subtracted). For instance, in "$17.0 \pm 0.5\,^{\circ}$C," the "$0.5\,^{\circ}$C" is the absolute error.

**Rule 2:** When multiplying or dividing measurements with associated errors, the percentage errors (or the fractional errors) are summed.

Assume that in your experiment you have measured four variables $x$, $y$, $z$, and $w$ and have determined errors $\Delta x$, $\Delta y$, $\Delta z$, and $\Delta w$ in them. Assume further that a formula $r = f(x, y, z, w)$ must be used to determine a final result $r$. What is the expected error in $r$?

The approximate expression for the differential $dr$ is given by

$$\mathrm{d}r = (\partial f/\partial x)_{y,z,w}\mathrm{d}x + (\partial f/\partial y)_{z,w,x}\mathrm{d}y + (\partial f/\partial z)_{w,x,y}\mathrm{d}z + (\partial f/\partial w)_{x,y,z}\mathrm{d}w \tag{1.17.6}$$

which, in practice, is used as follows:

$$\Delta r = (\partial f/\partial x)_{y,z,w}\Delta x + (\partial f/\partial y)_{z,w,x}\Delta y + (\partial f/\partial z)_{w,x,y}\Delta z + (\partial f/\partial w)_{x,y,z}\Delta w \tag{1.17.7}$$

### 1.17.3 Gaussian Normal Error Probability Function

This standard deviation $\sigma$ occurs in the Gaussian or normal error probability function $P_{\text{Gaussian}}(x)$:

$$P_{\text{Gaussian}}(x) = (2\pi)^{-1/2}\sigma^{-1}\exp\left(-x^2/2\sigma^2\right) \tag{1.17.8}$$

The infinite integral of this function equals 1. That is what "normal" or "normalized" means, but in general parlance, all distributions are defined to be normalized. Thus, the name "normal" for the Gaussian distribution is probably a testament to its prevalence.

$$\int_{x=-\infty}^{x=+\infty} P_{\text{Gaussian}}(x)\,\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}}\int_{x=-\infty}^{x=+\infty}\exp\left(-x^2/2\sigma^2\right)\mathrm{d}x = 1 \tag{1.17.9}$$

As already mentioned in Section 1.5, its finite one-sided integral is known as the **error function**:

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_{t=0}^{t=x}\exp\left(-t^2\right)\mathrm{d}t = \frac{\sqrt{2}}{\sigma\sqrt{\pi}}\int_{t=0}^{t=x}\exp\left(-t^2/2\sigma^2\right)\mathrm{d}t \tag{1.17.10}$$

The Gaussian distribution function is most valid when the number of data, $N$, is very large. When there are fewer data, in a smaller sample, a different distribution function is better: it was designed by Gosset,[164] who, required by his employer (Guinness Brewery, 1908) to publish anonymously, chose "Student" as his *nom de plume*, and therefore known as the "**Student-t**" distribution:

$$T_{\text{student}}(t) \equiv \Gamma(N/2)(N-1)^{-1/2}[\Gamma((N-1)/2)]^{-1}\left\{1 + (N-1)^{-1}t^2\right\}^{-N/2} \tag{1.17.11}$$

where $\Gamma(x)$ is the gamma function. The Student-$t$ distribution function is normalized:

$$\int_{t=-\infty}^{t=+\infty} T_{\text{student}}(t)\,\mathrm{d}t = 1 \tag{1.17.12}$$

When $N$ is large, $T_{\text{student}}(t)$ does approach $P_{\text{Gaussian}}(x)$.

If we wish to find the range of values $x$ over which the integral of the Student-$t$ distribution function has a value $T_{0.95}$, of say, 0.95 (i.e. 95%), this is the integral:

$$T_{0.95} = 0.95 = \int_{x=-t}^{x=t} T_{\text{Student}}(x)\,\mathrm{d}x \tag{1.17.13}$$

This has been tabulated extensively (see Table 1.17; but watch out for subtly different conventions, i.e. different limits may be set on the integral in Eq. (1.17.13)).

It is often said that (i) "within a confidence limit of 95%, the mean of a large number of measurements is within $2\sigma$ of the true value" and that (ii) "within a confidence limit of 99%, the mean is within $3\sigma$ of the true value." These two estimates can be found in Table 1.17 (i) under entry $P = 0.95$, for $f = \infty$, $t = 2.04$; (ii) under $P = 0.99$ for $f = \infty$, $t = 2.75 \approx 3$.

164 William Sealy Gosset (1876–1937).

**Table 1.17** For $f$ = number of data = degrees of freedom, values of $t$ for the Student $t$-distribution, where the heading $P$ is the probability ($0 \leq P \leq 1$) that the mean of the population for an infinite number of data does not differ from the finite sample mean by a factor of more than $t$.

| | P | | | | | | |
|---|---|---|---|---|---|---|---|
| $f$ | 0.50 | 0.80 | 0.90 | 0.95 | 0.98 | 0.99 | 0.999 |
| 1 | 1.00 | 3.08 | 6.31 | 12.7 | 31.8 | 63.7 | 637.0 |
| 2 | 0.816 | 1.89 | 2.92 | 4.30 | 6.96 | 9.92 | 31.6 |
| 3 | 0.765 | 1.64 | 2.35 | 3.18 | 4.54 | 5.84 | 12.9 |
| 4 | 0.741 | 1.53 | 2.13 | 2.78 | 3.75 | 4.60 | 8.61 |
| 5 | 0.727 | 1.48 | 2.02 | 2.57 | 3.36 | 4.03 | 6.87 |
| 6 | 0.718 | 1.44 | 1.94 | 2.45 | 3.14 | 3.71 | 5.96 |
| 7 | 0.711 | 1.41 | 1.89 | 2.36 | 3.00 | 3.50 | 5.41 |
| 8 | 0.706 | 1.40 | 1.86 | 2.31 | 2.90 | 3.36 | 5.04 |
| 9 | 0.703 | 1.38 | 1.83 | 2.26 | 2.82 | 3.25 | 4.78 |
| 10 | 0.700 | 1.37 | 1.81 | 2.23 | 2.76 | 3.17 | 4.59 |
| 15 | 0.691 | 1.34 | 1.75 | 2.13 | 2.60 | 2.85 | 4.07 |
| 20 | 0.687 | 1.33 | 1.72 | 2.09 | 2.53 | 2.85 | 3.85 |
| 30 | 0.683 | 1.31 | 1.70 | 2.04 | 2.46 | 2.75 | 3.65 |
| ∞ | 0.683 | 1.31 | 1.70 | 2.04 | 2.46 | 2.75 | 3.65 |

### 1.17.4 Binomial Distribution

The probability of $m$ "successes" and $(n–m)$ "failures" is given by

$$P_{\text{binomial}}(m) = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \tag{1.17.14}$$

It can be shown that for large $n$, the binomial distribution becomes the Gaussian distribution (Problem 1.15.2)

### 1.17.5 Poisson Distribution

Another limit of the binomial distribution will yield the Poisson distribution: by letting both $n \to \infty$ and $p \to 0$, but keeping the product $np$ constant, e.g. $nm = a$ (where $a$ is a constant), we will see that $[n\,!\,/(n-m)!] \to nm$ and $(1-p)^{n-m} \to \exp(-a)$, whence

$$P_{\text{Poisson}}(m) = a^m \exp(-a)/m! \tag{1.17.15}$$

Of course, $\sum_{m=0}^{m=\infty} P_{\text{Poisson}}(m) = 1$. The Poisson distribution is applicable for the case of a large number of experiments, but each with a small probability of success $a$.

**Problem 1.17.1** Prove that for large $n$, the binomial distribution becomes the Gaussian distribution.

### 1.17.6 Least Squares or Linear Regression Analysis

A traditional method of treating $N$ measured data ($x_i$, $y_i$, $i = 1, 2, \ldots, N$), if the theoretical equation governing them is known $y_{\text{theo}}(x)$; the simplest case is the linear equation (that chemists really love!):

$$y_{\text{theo}}(x) = mx + b \tag{1.17.16}$$

is to fit the data to the equation, obtaining the best slope $m$ and the best intercept $b$, so that that the sum of the squares of the deviations from the average $\langle x\rangle = (1/N)\sum_{i=1}^{i=N} x_i$ be minimized. The general procedure is as follows: define and compute the four sums:

$$\begin{aligned}\sum x &\equiv \sum_{i=1}^{i=N} x_i\\ \sum y &\equiv \sum_{i=1}^{i=N} y_i\\ \sum x^2 &\equiv \sum_{i=1}^{i=N} x_i^2\\ \sum xy &\equiv \sum_{i=1}^{i=N} x_i y_i\end{aligned} \tag{1.17.17}$$

(note that the sum $\sum_{i=1}^{i=N} y_i^2$ is not computed and that the problem is asymmetrical: it assumes that the *independent variable* $x$ *is free from error* and that the error is all in the measurement of the dependent variable $y$). Then the **"best" slope** and **"best" intercept** are given by

$$\begin{aligned}m_{\text{best}} &= \left[N\sum xy - \sum x\sum y\right] \Big/ \left[N\sum x^2 - \left(\sum x\right)^2\right]\\ b_{\text{best}} &= \left[\sum x^2 \sum y - \sum x \sum xy\right] \Big/ \left[N\sum x^2 - \left(\sum x\right)^2\right]\end{aligned} \tag{1.17.18}$$

when this is done, and the sum $\sum_{i=1}^{i=N} y_i^2$ will have been minimized. This is the **method of least squares**, introduced in 1805 by Legendre, later used in a statistical study of the late 1800s that showed that the fraction of tall people in a sample human population tended to return or "regress" to a lower average height: the name "regression analysis" stuck.

Another measure of interest is the **sample correlation coefficient** or **Pearson's *r*** or **Pearson's**[165] **product-moment formula** or **the linear correlation of *x* and *y***:

$$r = \left[N\sum xy - \sum x\sum y\right]\left[N\sum x^2 - \left(\sum x\right)^2\right]^{-1/2}\left[N\sum y^2 - \left(\sum y\right)^2\right]^{-1/2} \tag{1.17.19}$$

where $\sum y^2 \equiv \sum_{i=1}^{i=N} y_i^2$. Note that $-1 \le r \le 1$: **if $r = 1$, then all points lie exactly on one line**; if $|r| = 0$, then no straight line is any better than any other.

If the function $f(x)$ is not linear in $x$, then **nonlinear regression** is followed: the transcendental equation is re-cast in a linear form by suitable (if often approximate) transformations (e.g. computing its logarithm), and one proceeds as in the linear case, with a suitable back-transformation at the end.

## 1.18 Statistics

We introduce a few results from **combinatorics**, the science of ordering and sorting objects into "boxes" or "containers": these results will be useful for developing the Maxwell–Boltzmann, Fermi–Dirac, and Bose[92]–Einstein statistics in Chapter 5.

**Problem 1.18.1** Show that the number of ways of ordering $N$ distinguishable objects (=number of permutations of $N$ distinguishable objects) is given by:

$$n_1 - N! \tag{1.18.1}$$

165 Karl Pearson (1857–1936).

**Problem 1.18.2** Prove that the number of ways of placing $N$ distinguishable objects into $R$ distinguishable boxes, so that there are $N_1$ objects in box 1, $N_2$ objects in box 2, etc. is given by:

$$n_2 = \frac{N!}{\prod_{i=1}^{i=R} N_i!} \tag{1.18.2}$$

**Problem 1.18.3** Prove that the number of ways of selecting $N$ distinguishable objects from a set of $G$ distinguishable objects (where $G > N$) is given by the binomial coefficient:

$$n_3 = \binom{G}{N} \equiv \frac{G!}{N!(G-N)!} \tag{1.18.3}$$

**Problem 1.18.4** Prove that the number of ways of placing $N$ indistinguishable objects into $G$ distinguishable containers is given by:

$$n_4 = \frac{(G+N-1)!}{(G-1)!N!} \tag{1.18.4}$$

**Problem 1.18.5**
Show that the number of ways of placing $N$ distinguishable objects into $G$ distinguishable containers is given by:

$$n_5 = \frac{(G+N-1)!}{(G-1)!} \tag{1.18.5}$$

**Problem 1.18.6** Stirling's[93] approximation for large $N$ is given by:

$$N! \approx (2\pi N)^{1/2} N^{\mathrm{N}} \exp(-N) \tag{1.18.6}$$

which can be approximated adequately as:

$$\ln N! \approx N \ln N - N \tag{1.18.7}$$

Calculate the percent error between ln $N$! and $N$ln $N$ – $N$ for $N = 50$, and for $N = 65$.

**Problem 1.18.7**
The **gamma function** $\Gamma(x)$, invented by Leonhard Euler, is given by either of two integrals:

$$\Gamma(x) \equiv \int_{t=0}^{t=\infty} t^{x-1} \exp(-t)\,\mathrm{d}t = \int_{t=0}^{t=1} \left[\ln(1/t)^{x-1}\mathrm{d}t\right] \tag{1.18.8}$$

valid for positive real $x$, or for complex $x$ with a positive real part.
(a) First show that:

$$\Gamma(x+1) = x\Gamma(x)$$

(b) Second, show that for integer positive $N$ the gamma function becomes a factorial:

$$\Gamma(N) = (N-1)! \tag{1.18.9}$$

(c) Third, show that:

$$\Gamma(1/2) = \pi^{1/2} \tag{1.18.10}$$

## 1.19 General References

Mathematical methods: [1.36–1.41] (some of these are "golden oldies"). Integrals, series, and other tables: [1.19–1.25]. Of course, much labor can now be saved by using the commercial computer program package Mathematica.

### Special Topic ST1.1 Lifetimes from Natural Linewidths

As discussed in Section 1.4 above, the **lifetimes** $\tau$, or **half-lives** $t^{1/2}$ (where $\tau \equiv 1.442\,695 t^{1/2}$) can measured directly when $\tau \geq 1.0 \times 10^{-12}$ s or so. Shorter half-lives are inferred from the measured "natural" or Breit–Wigner or Lorentzian linewidth $\Delta E$ of their energy $E$ (see Figure 1.33) by using the Heisenberg uncertainty principle condition (discussed further in Section 2.1)

$$\Delta E \Delta \tau \geq (h/4\pi \equiv \hbar/2) \tag{ST1.1.1}$$

where $h$ is Planck's constant of action (one can argue whether $\tau$ is a lifetime or a half-life in the Heisenberg sense). In practice, one estimates the half-life $\tau$ from the width of the resonance $\Delta E_{12}$ (=width at half-maximum height; see Figure 1.33)

$$\tau \equiv \hbar/2\Delta E_{12} \tag{ST1.1.2}$$

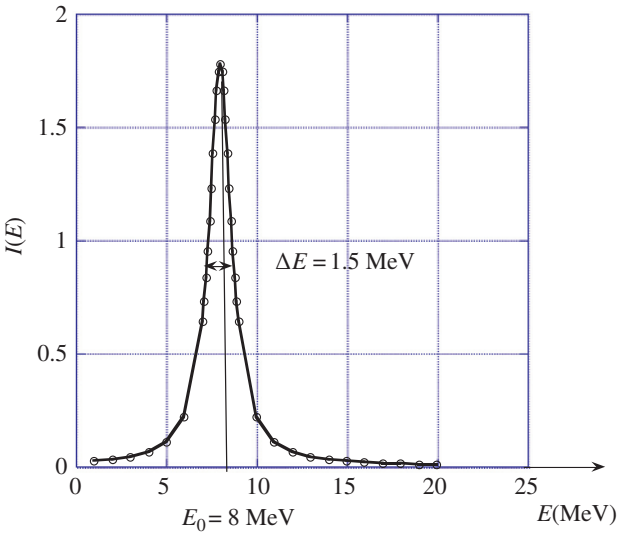



**Figure 1.33** Natural or Breit–Wigner or Lorentzian linewidth $I(E) = [(E - E_0)^2 + \Delta E^2/4]^{-1}$ for $\Delta E = 1.5$ MeV at half-maximum centered around $E_0 = 8.0$ MeV.

### Special Topic ST1.2 Nuclear Shell Model

The nuclear shell model, developed by Göppert-Mayer and Jensen, posits populating nuclear states as if the nucleons occupied the lowest possible quantum states for a three-dimensional harmonic oscillator, but with an energy correction due to a **nuclear spin–orbit interaction**: the nuclear "spin" quantum numbers *I* and "orbital" quantum numbers *M*, couple strongly as $\boldsymbol{I} \cdot \boldsymbol{M}$; this nuclear spin–orbit interaction (invented in analogy to the electron spin–orbit interaction) is, however, due to an unknown potential function; nevertheless, this model does account nicely for magic number stability and nuclear excited states. There is an acrostic "**spuds if pug dish of pig**" that serves as a mnemonic for the ordering 1s, 1p, 1d, 2s, 1f, 2p, 1g, 2d, 3s, 1h, 2f, 3p, 1i, 2g (before nuclear spin–orbit splittings). Note also that Table X1 (please refer www.wiley/go/metzger/physchemtoolbox2) shows only the ground state of nuclei.

The force that accounts for nuclear stabilities must, of course, be the strong force, or rather its residual component that works outside of quark confinement.

Excited states of nuclei have been studied and tabulated extensively, but the details exceed the objectives of the present volume.

### Special Topic ST1.3 Factoring a Polynomial With Integer Coefficients by Long Division

We start with a polynomial equation of order *n* with integer coefficients

$$a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1 x + a_0 = 0 \quad (n = 2, 3, 4, \ldots) \tag{ST1.3.1}$$

All linear factors with rational coefficients can be found by using the rational root test: all possible linear factors must be of the form $b_1 x - b_0$, where $b_1$ is an **integer factor** of the leading term $a_n$, while $b_0$ is an **integer factor** of the last term $a_0$. This rational root test, also known as rational root theorem, or as the rational zero test, or as the *p/q* theorem, is a consequence of Gauss' lemma.

**Problem ST1.3.1** Find the quotient and the remainder of the **dividend** $x^3 - 2x^2 - 4$ by the **quotient** $x - 3$ (procedure is similar to that of "long division").

### Special Topic ST1.4 Other Solutions of the Cubic Equation

A solution to Eq. (1.5.3) can be obtained by using Viète's substitution:

$$\begin{aligned} &\lambda = u - p/3u - a/3, \ \text{where :} \\ &p \equiv b - a^2/3 \\ &q \equiv (2/27)a^3 - (1/3)ab + c \\ &u \equiv \left\{ -q/2 \pm \left[ (1/4)q^2 + p^3/27 \right]^{1/2} \right\}^{1/3} \end{aligned} \tag{ST1.4.1}$$

Two distinct roots are possible, for the two alternatives for ±. The three cube roots of 1: (i) $\exp(2\pi i/3) = \cos(120^\circ) + i\sin(120^\circ) = -0.500\,000 + i0.866\,025 = -(1/2) + i(3^{1/2}/2)$, (ii) $\exp(4\pi i/3) = \cos(240^\circ) + i\sin(240^\circ) = -0.500\,000 - i0.866\,025 = -(1/2) - i(3^{1/2}/2)$, and (iii) $\exp(6\pi i/3) = 1$ provide three roots; this times two, is six roots, which do reduce to only three.

**Special Topic ST1.4 (Continued)**

To get the three correct roots $\lambda_1$, $\lambda_2$, and $\lambda_3$, it is essential that $u$ (and not $\lambda$) be premultiplied by factors of exp(6πi/3), exp(2πi/3), and exp(4πi/3), or else wrong results will be obtained. The discriminant $\Delta \equiv 18abc - 4a^3c + a^2b^2 - 4b^3 - 27c^2$ determines the nature of the three roots: if $\Delta > 0$, then there are three distinct real roots; if $\Delta = 0$, then all three roots are real (but some are degenerate); if $\Delta < 0$, then there are one real root and two complex and mutually conjugate roots.

Equivalently, one can define:

$$A \equiv \left\{ -q/2 + \left[ (1/4)q^2 + p^3/27 \right]^{1/2} \right\}^{1/3}$$
$$B \equiv \left\{ -q/2 - \left[ (1/4)q^2 + p^3/27 \right]^{1/2} \right\}^{1/3} \qquad \text{(ST1.4.2)}$$

The three solutions are then:

$$\lambda_1 = A + B$$
$$\lambda_2 = -(1/2)(A+B) + \mathrm{i}\left(3^{1/2}/2\right)(A-B)$$
$$\lambda_3 = -(1/2)(A+B) - \mathrm{i}\left(3^{1/2}/2\right)(A-B) \qquad \text{(ST1.4.3)}$$

Consider $\delta \equiv (1/4)q^2 + p^3/27$; within overall multiplicative factors, this $\delta$ is equivalent to, but opposite in sign to, the discriminant $\Delta \equiv 18abc - 4a^3c + a^2b^2 - 4b^3 - 27c^2$ defined above.

If $\delta > 0$, there will be 1 real root and 2 conjugate imaginary roots;
if $\delta = 0$, there will be 3 real roots, of which at least 2 are equal;
if $\delta < 0$, there will be 3 real and unequal roots; if $\delta < 0$, then define:

cos $\theta \equiv (-q/2)/(-p^3/27)^{1/2}$ and the 3 unequal real roots (for $\delta < 0$ and $\Delta > 0$) are

$$\lambda_1 = -a/3 + 2(-p/3)^{1/2}\cos(\theta/3)$$
$$\lambda_2 = -a/3 + 2(-p/3)^{1/2}\cos(\theta/3 + 120^\circ)$$
$$\lambda_3 = -a/3 + 2(-p/3)^{1/2}\cos(\theta/3 + 240^\circ) \qquad \text{(ST1.4.4)}$$

**Special Topic ST1.5 Solving a Quartic Equation**

The quartic equation [1.41]:

$$a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0 \qquad \text{(ST1.5.1)}$$

was first solved by Ferrari in 1540, and published by Cardano in 1545; Euler and Viète later provided deep insights.

### Discriminant Analysis

For Eq. (ST1.5.1) with five real nonzero coefficients $a$, $b$, $c$, $d$, and $e$, the roots are mainly determined by the sign of the **discriminant** $\Delta$:

$$\begin{aligned}\Delta \equiv\ & 256a_4^3a_0^3 - 192a_4^2a_3a_1a_0^2 - 128a_4^2a_2^2a_0^2 + 144a_4^2a_2a_1^2a_0 - 27a_4^2a_1^4 \\ & + 144a_4a_3^2a_2a_0^2 - 6a_4a_3^2a_1^2a_0 - 80a_4a_3a_2^2a_1a_0 + 18a_4a_3a_2a_1^3 + 16a_4a_2^4a_0 \\ & - 4a_4a_2^3a_1^2 - 27a_3^4a_0^2 + 18a_3^3a_2a_1a_0 - 4a_3^3a_1^3 - 4a_3^3a_2^3a_0 + a_3^2a_2^2a_1^2\end{aligned} \qquad \text{(ST1.5.2)}$$

and also, by the signs of four other polynomials $P$, $R$, $\Delta_0$, and $D$ defined next. The first is

$$P \equiv 8a_4a_2 - 3a_3^2 \qquad \text{(ST1.5.3)}$$

**Special Topic ST1.5 (Continued)**

where $P/8a_4^2$ is the second-degree coefficient of the associated depressed quartic shown below,

$$R \equiv a_3^3 + 8a_1a_4^2 - 4a_4a_3a_2 \tag{ST1.5.4}$$

where $R/8a_4^3$ is the first-degree coefficient of that associated depressed quartic, and:

$$\Delta_0 \equiv a_2^2 - 3a_3a_1 + 12a_4a_0 \tag{ST1.5.5}$$

which vanishes ($\Delta_0 = 0$) if the quartic has a triple root; and finally,

$$D \equiv 64a_4^3a_0 - 16a_4^2a_2^2 + 16a_4a_3^2a_2 - 16a_4^2a_3a_1 - 3a_3^4 \tag{ST1.5.6}$$

equals 0 ($D = 0$) if the quartic has two double roots. The resultant discriminant analysis shows that the possible cases for the roots are as follows:

**(1)** If < 0, then Eq. (ST1.5.1) has two distinct real roots and two complex conjugate roots.
**(2)** If > 0, then either the four roots of Eq. (ST1.5.1) are all real or none of them is.
**(2a)** If P < 0 and D < 0, then all four roots are real and distinct.
**(2b)** If P > 0 or D > 0, then there are two pairs of nonreal complex conjugate roots.
**(3)** If $\Delta_0 = 0$, then (and only then) the quartic has multiple roots. Here are the different cases that can occur: If P < 0 and D < 0 and $\Delta_0 = 0$, there are 1one real double root plus two real simple roots.

**(3a)** If D > 0 or (P > 0 and (D $\neq$ 0 or R $\neq$ 0)), there are one real double root and 2two complex conjugate roots.
**(3b)** If $\Delta_0 = 0$ and D $\neq$ 0, there are a triple root and a simple root, all real.
**(3c)** If D = 0, then:
**(3da)** If P < 0, there are two real double roots.
**(3db)** If P > 0 and R = 0, there are two complex conjugate double roots.
**(3dc)** If $\Delta_0 = 0$, all four roots are equal to $-b/4a$.

Cases not covered in the list above cannot occur. (End of discriminant analysis.)

Instead of the five subscripted coefficients ($a_4$, $a_3$, $a_2$, $a_1$, and $a_0$) in Eq. (ST1.5.1), one can make do with four unsubscripted constants ($a$, $b$, $c$, and $d$) in a slightly simpler quartic equation:

$$x^4 + ax^3 + bx^2 + cx + d = 0 \tag{ST1.5.7}$$

Viète found interesting results for the four roots $x_1$, $x_2$, $x_3$, and $x_4$ of Eq. (ST1.5.7):

$$x_1 + x_2 + x_3 + x_4 = -a \tag{ST1.5.8}$$

$$x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 = +b \tag{ST1.5.9}$$

$$x_1x_2x_3 + x_2x_3x_4 + x_1x_2x_4 + x_1x_3x_4 = -c \tag{ST1.5.10}$$

$$x_1x_2x_3x_4 = +d \tag{ST1.5.11}$$

For Eq. (ST1.5.7) several websites give the following, **possibly** valid, set of four roots ($x_1$–$x_4$) combined for convenience into a single complicated formula for the general root $x_{1,\,2,\,3,\,4}$. For convenience, three parentheses carry four signs each: {**–, –, +, +**}, {**–, +, –, +**}, and {**–, –, +, +**} are used, to allow choosing the relevant sign as follows: the first root $x_1$ goes with the first signs of the three indicated parentheses, e.g. the first parenthesis shown {**–, –, +, +**} indicates that for root $x_1$ you choose the first sign (**–**); for root $x_2$ you chose the second sign (**–**); and so forth. For all three parentheses, the signs shown explicitly just before the three parentheses are the signs for $x_1$. Also, given how easily errors can creep in when using this complicated formula, alphabetic capital subscripts are added in bold to each parenthesis opened or closed (except for trivially simple quotients like ($a/b$)): e.g. $_{\mathbf{A}}$ is used for the outermost parenthesis opened, $_{\mathbf{B}}$ is used for the next parenthesis opened, etc., and $_{\mathbf{-A}}$ is shown when that first parenthesis is closed, $_{\mathbf{-B}}$ for when the next-outermost parenthesis is closed, etc.

**Special Topic ST1.5 (Continued)**

$$
\begin{aligned}
x_{1,2,3,4} = &-(a/4)-\{-,-,+,+\}(1/2)\Big\{_{\mathbf{A}}\left(a^2/4\right)-(2b/3)+2^{1/3}3^{-1}\left(_{\mathbf{B}}b^2-3ac+12d\right)\\
&-\mathbf{B}\Big[_{\mathbf{B}}2b^3-9abc+27c^2+27a^2d-72bd+\Big[_{\mathbf{C}^{-4}}\left(_{\mathbf{D}}b^2-3ac+12d\right)_{-\mathbf{D}^3}\\
&+\left(_{\mathbf{D}}2b^3-9abc+27c^2+27a^2d-72bd\right)_{-\mathbf{D}^2}\Big]_{-\mathbf{C}^{1/2}}\Big]_{-\mathbf{B}^{-1/3}}\\
&+\Big[_{\mathbf{B}}\Big(_{\mathbf{C}}2b^3-9abc+27c^2+27a^2d-72bd+\Big[_{\mathbf{D}^{-4}}\left(_{\mathbf{E}}b^2-3ac+12d\right)_{-\mathbf{E}^3}\\
&+\left(_{\mathbf{E}}2b^3-9abc+27c^2+27a^2d-72bd\right)_{-\mathbf{E}^2}\Big]_{-\mathbf{D}^{1/2}}\Big)_{-\mathbf{C}}/54\Big]_{-\mathbf{B}^{1/3}}\Big\}_{-\mathbf{A}^{1/2}}\\
&-\{-,+,-,+\}(1/2)\Big\{_{\mathbf{A}}\left(a^2/2\right)-(4b/3)-2^{1/3}3^{-1}\left(_{\mathbf{B}}b^2-3ac+12d\right)\\
&-\mathbf{B}\Big[_{\mathbf{B}}2b^3-9abc+27c^2+27a^2d-72bd+\Big[_{\mathbf{C}^{-4}}\left(_{\mathbf{D}}b^2-3ac+12d\right)_{-\mathbf{D}^3}\\
&+\left(_{\mathbf{D}}2b^3-9abc+27c^2+27a^2d-72bd\right)_{-\mathbf{D}^2}\Big]_{-\mathbf{C}^{1/2}}\Big]_{-\mathbf{B}^{-1/3}}-\Big[_{\mathbf{B}}\Big(_{\mathbf{C}}2b^3-9abc\\
&+27c^2+27a^2d-72bd\Big[_{\mathbf{D}^{-4}}\left(_{\mathbf{E}}b^2-3ac+12d\right)_{-\mathbf{E}^3}+\left(_{\mathbf{E}}2b^3-9abc+27c^2-27a^2d-72bd\right)_{-\mathbf{E}^2}\Big]\\
&_{-\mathbf{D}^{1/2}}\Big)_{-\mathbf{C}}/54\Big]_{-\mathbf{B}^{1/3}}-\{-,-,+,+\}(1/4)\left(_{\mathbf{B}}-a^3+4ab-8c\right)\\
&-\mathbf{B}\Big\{_{\mathbf{B}}\left(a^2/4\right)-(2b/3)+2^{1/3}3^{-1}\left(_{\mathbf{C}}b^2-3ac+12d\right)_{-\mathbf{C}}\left(_{\mathbf{C}}2b^3-9abc+27c^2+27a^2d-72bd\right.\\
&+\Big[_{\mathbf{D}}\Big\{_{\mathbf{E}^{-4}}\left(b^2-3ac+12d\right)_{-\mathbf{E}^3}+\left(_{\mathbf{E}}2b^3-9abc+27c^2+27a^2d-72bd\right)_{-\mathbf{E}^2}\Big\}_{-\mathbf{D}^{1/2}}\Big]_{-\mathbf{C}^{-1/3}}\\
&+\Big[_{\mathbf{C}}\left(_{\mathbf{D}}2b^3-9abc+27c^2+27a^2d-72bd+\Big[_{\mathbf{E}^{-4}}\left(_{\mathbf{F}}b^2-3ac+12d\right)_{-\mathbf{F}^3}\right.\\
&\left.\left.+\left(_{\mathbf{F}}2b^3-9abc+27c^2+27a^2d-72bd\right)_{-\mathbf{F}^2}\Big]_{-\mathbf{E}^{1/2}}\right)_{-\mathbf{D}}/54\Big]_{-\mathbf{C}^{1/3}}\Big)_{-\mathbf{B}^{-1/2}}\Big\}_{-\mathbf{A}^{1/2}}
\end{aligned}
\tag{ST1.5.12}
$$

Thus, shorn of the paraphernalia, the third root $x_3$ is specifically:

$$
\begin{aligned}
x_3 = &-(a/4)+(1/2)\left\{\left(a^2/4\right)-(2b/3)\right.\\
&+2^{1/3}3^{-1}\left(b^2-3ac+12d\right)\Big[2b^3-9abc+27c^2+27a^2d-72bd\\
&+\left[-4\left(b^2-3ac+12d\right)^3+\left(2b^3-9abc+27c^2+27a^2d-72bd\right)^2\right]^{1/2}\Big]^{-1/3}\\
&+\Big[\left(2b^3-9abc+27c^2+27a^2d-72bd\right.\\
&\left.\left.+\left[-4\left(b^2-3ac+12d\right)^3+\left(2b^3-9abc+27c^2+27a^2d-72bd\right)^2\right]^{1/2}\right)/54\Big]^{1/3}\Big\}^{1/2}\\
&-(1/2)\left\{\left(a^2/2\right)-(4b/3)-2^{1/3}3^{-1}\left(b^2-3ac+12d\right)\left(2b^3-9abc+27c^2+27a^2d-72bd\right.\right.\\
&+\left[-4\left(b^2-3ac+12d\right)^3+\left(2b^3-9abc+27c^2+27a^2d-72bd\right)^2\right]^{1/2}\Big)^{-1/3}\\
&-\Big[\left(2b^3-9abc+27c^2+27a^2d-72bd\right.\\
&\left.+\left[-4\left(b^2-3ac+12d\right)^3+\left(2b^3-9abc+27c^2-27a^2d-72bd\right)^2\right]^{1/2}\right)/54\Big]^{1/3}\\
&+(1/4)\left(-a^3+4ab-8c\right)\left\{\left(a^2/4\right)-(2b/3)\right.\\
&+2^{1/3}3^{-1}\left(b^2-3ac+12d\right)\left(2b^3-9abc+27c^2+27a^2d-72bd\right.\\
&\left.+\left[-4\left(b^2-3ac+12d\right)^3+\left(2b^3-9abc+27c^2+27a^2d-72bd\right)^2\right]^{1/2}\right)^{-1/3}\\
&+\Big[\left(2b^3-9abc+27c^2+27a^2d-72bd\right.\\
&\left.\left.\left.+\left[-4\left(b^2-3ac+12d\right)^3+\left(2b^3-9abc+27c^2+27a^2d-72bd\right)^2\right]^{1/2}\right)/54\Big]^{-1/2}\right\}^{1/2}
\end{aligned}
\tag{ST1.5.13}
$$

## Special Topic ST1.5 (Continued)

By defining three auxiliary quantities:

$$\phi \equiv a_2^2 - 3a_3a_1 + 12a_0 \equiv b^2 - 3ac + 12d \tag{ST1.5.14}$$

$$\xi \equiv 2a_2^3 - 9a_3a_2a_1 + 27a_1^2 + 27a_3^2a_0 - 72a_2a_0 \equiv 2b^3 - 9abc + 27c^2 + 27a^2d - 72bd \tag{ST1.5.15}$$

$$\omega \equiv -a_3^3 + 4a_3a_2 - 8a_1 \equiv -a^3 + 4ab - 8c \tag{ST1.5.16}$$

and inserting them into Eq. (ST1.5.12), we get a simpler-looking combined root:

$$\begin{aligned}
x_{1,2,3,4} = &-(a/4) - \{-,-,+,+\}(1/2)\Big\{\mathbf{A}(a^2/4) - (2b/3) + 2^{1/3}3^{-1}\phi\Big[\mathbf{B}\xi + \big[\mathbf{C} - 4\phi^3 + \xi^2\big]_{-\mathbf{C}^{1/2}}\Big]_{-\mathbf{B}^{-1/3}} \\
&+ \Big[\mathbf{B}\big(\mathbf{C}\xi + \big[\mathbf{D} - 4\phi^3 + \xi^2\big]_{-\mathbf{D}^{1/2}}\big)_{-\mathbf{C}}/54\Big]_{-\mathbf{B}^{1/3}}\Big\}_{-\mathbf{A}^{1/2}} \\
&- \{-,+,-,+\}(1/2)\Big\{\mathbf{A}(a^2/2) - (4b/3) + 2^{1/3}3^{-1}\phi\Big(\mathbf{B}\xi + \big[\mathbf{C} - 4\phi^3 + \xi^2\big]_{-\mathbf{C}^{1/2}}\Big)_{-\mathbf{B}^{-1/3}} \\
&- \Big[\mathbf{B}\big(\mathbf{C}\xi + \big[\mathbf{D} - 4\phi^3 + \xi^2\big]_{-\mathbf{D}^{1/2}}\big)_{-\mathbf{C}}/54\Big]_{-\mathbf{B}^{1/3}} \\
&- \{-,-,+,+\}(1/4)\omega\Big[\mathbf{B}(a^2/4) - (2b/3) + 2^{1/3}3^{-1}\phi\big(\mathbf{C}\xi + \big[\mathbf{D} - 4\phi^3 + \xi^2\big]_{-\mathbf{D}^{1/2}}\big)_{-\mathbf{C}^{-1/3}} \\
&+ \Big[\mathbf{C}\big(\mathbf{D}\xi + \big[\mathbf{E} - 4\phi^3 + \xi^2\big]_{-\mathbf{E}^{1/2}}\big)_{-\mathbf{D}}/54\Big]_{-\mathbf{C}^{1/3}}\Big\}_{-\mathbf{B}^{-1/2}}\Big\}_{-\mathbf{A}^{1/2}}
\end{aligned} \tag{ST1.5.17}$$

$$\begin{aligned}
x_{1,2,3,4} = &-(a/4) - \{-,-,+,+\}(1/2)\Big\{(a^2/4) - (2b/3) + 2^{1/3}3^{-1}\phi\Big[\xi + \big[-4\phi^3 + \xi^2\big]^{-1/2}\Big]^{-1/3} \\
&+ \Big[\Big(\xi + \big[-4\phi^3 + \xi^2\big]^{-1/2}\Big)/54\Big]^{-1/3}\Big\}^{-1/2} \\
&- \{-,+,-,+\}(1/2)\Big\{(a^2/2) - (4b/3) + 2^{1/3}3^{-1}\phi\Big(\xi + \big[-4\phi^3 + \xi^2\big]^{-1/2}\Big)^{-1/3} \\
&- \Big[\Big(\xi + \big[-4\phi^3 + \xi^2\big]^{-1/2}\Big)/54\Big]^{-1/3} \\
&- \{-,-,+,+\}(1/4)\omega\Big[(a^2/4) - (2b/3) + 2^{1/3}3^{-1}\phi\Big(\xi + \big[-4\phi^3 + \xi^2\big]^{-1/2}\Big)^{-1/3} \\
&+ \Big[\Big(\xi + \big[-4\phi^3 + \xi^2\big]^{-1/2}\Big)/54\Big]^{-1/3}\Big\}^{-1/2}\Big\}^{-1/2}
\end{aligned} \tag{ST1.5.18}$$

**Problem ST1.5.1** Check the validity of Eq. (ST1.5.12) by finding the four roots of the equation $x^4 + 15x^3 + 8x^2 - 3x + 2 = 0$. (a) First perform discriminant analysis, then (b) use a website equation solver, then finally (c) use Eq. (ST1.5.17) or Eq. (ST1.5.18).

**Problem ST1.5.2** Solve: $x^4 - 10x^3 + 21x^2 + 40x - 100 = 0$:

## Special Topic ST1.6 Spherical Trigonometry

Spherical trigonometry is essential for terrestrial and celestial navigation. Its brief presentation here should help scientific travelers estimate their travel distance and appreciate how the angles we all learned in plane trigonometry are quite different on a spherical surface. In a **spherical triangle** (a triangle localized on the surface of a sphere), angles *A*, *B*, *C* as well as sides *a*, *b*, *c* are measured in radians (Figure 1.34) [1.43].

**Special Topic ST1.6 (Continued)**

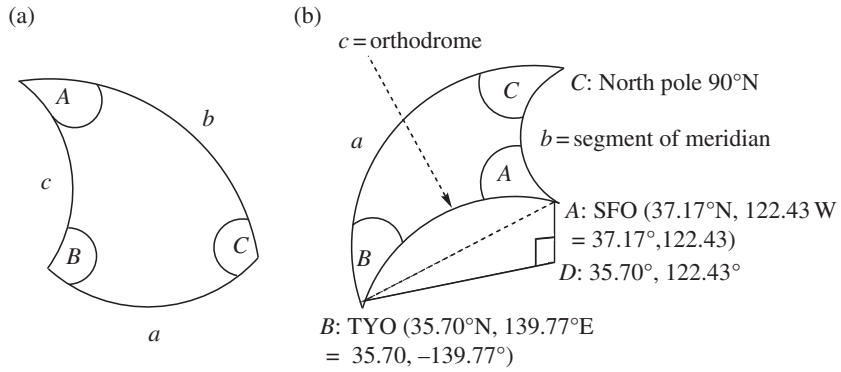



**Figure 1.34** (a) Spherical triangle with sides *a*, *b*, *c* and opposite angles *A*, *B*, and *C*. (b) Computation of orthodrome between San Francisco (SFO) and Tokyo (TYO), and plane triangle *ABD*. *Source:* Nielsen and Vanlonkuyyzen [1.43].

Here are 10 properties of spherical triangles:

1) Any angle *A*, *B*, or *C* must be less than 180°;
2) $a + b + c < 360°$;
3) any side is less than the sum of the other two sides;
4) $180° < A + B + C < 540°$;
5) the sine law is

$$\sin A/\sin a = \sin B/\sin b = \sin C/\sin c \tag{ST1.6.1}$$

6) the cosine law for the sides is

$$\begin{aligned} \cos a &= \cos b \cos c + \sin b \sin c \cos A \\ \cos b &= \cos c \cos a + \sin c \sin a \cos B \\ \cos c &= \cos a \cos b + \sin a \sin b \cos C \end{aligned} \tag{ST1.6.2}$$

7) the cosine law for the angles is

$$\begin{aligned} \cos A &= -\cos B \cos C + \sin B \sin C \cos a \\ \cos B &= -\cos C \cos A + \sin C \sin A \cos b \\ \cos C &= -\cos A \cos B + \sin A \sin B \cos c \end{aligned} \tag{ST1.6.3}$$

8) the half-angle formulas are

$$\begin{aligned} \sin(A/2) &= [\sin(s-b)\sin(s-c)/\sin b \sin c]^{1/2} \\ \sin(B/2) &= [\sin(s-c)\sin(s-a)/\sin c \sin a]^{1/2} \\ \sin(C/2) &= [\sin(s-a)\sin(s-b)/\sin a \sin b]^{1/2} \\ &\text{where } s \equiv (1/2)(a+b+c) \end{aligned} \tag{ST1.6.4}$$

and:

$$\begin{aligned} \tan(A/2) &= p/\sin(s-a) \\ \tan(B/2) &= p/\sin(s-b) \\ \tan(C/2) &= p/\sin(s-c) \\ &\text{where } p \equiv [\sin(s-a)\sin(s-b)\sin(s-c)/\sin s]^{1/2} \end{aligned} \tag{ST1.6.5}$$

**Special Topic ST1.6 (Continued)**

9) the four analogies of Napier[166]:

$$\begin{aligned}
&\sin[(A-B)/2]/\sin[(A+B)/2] = \tan[(a-b)/2]/\tan(c/2)\\
&\cos[(A-B)/2]/\cos[(A+B)/2] = \tan[(a+b)/2]/\tan(c/2)\\
&\sin[(a-b)/2]/\sin[(a+b)/2] = \tan[(A-B)/2]/\mathrm{cotan}(C/2)\\
&\cos[(a-b)/2]/\cos[(a+b)/2] = \tan[(A+B)/2]/\mathrm{cotan}(C/2)
\end{aligned} \tag{ST1.6.6}$$

and cyclically:

$$\begin{aligned}
&\sin[(B-C)/2]/\sin[(B+C)/2] = \tan[(b-c)/2]/\tan(a/2)\\
&\cos[(B-C)/2]/\cos[(B+C)/2] = \tan[(b+c)/2]/\tan(a/2)\\
&\sin[(b-c)/2]/\sin[(b+c)/2] = \tan[(B-C)/2]/\mathrm{cotan}(A/2)\\
&\cos[(b-c)/2]/\cos[(b+c)/2] = \tan[(B+C)/2]/\mathrm{cotan}(A/2)\\
&\sin[(C-A)/2]/\sin[(C+A)/2] = \tan[(c-a)/2]/\tan(b/2)\\
&\cos[(C-A)/2]/\cos[(C+A)/2] = \tan[(c+a)/2]/\tan(b/2)\\
&\sin[(c-a)/2]/\sin[(c+a)/2] = \tan[(C-A)/2]/\mathrm{cotan}(B/2)\\
&\cos[(c-a)/2]/\cos[(c+a)/2] = \tan[(C+A)/2]/\mathrm{cotan}(B/2)
\end{aligned} \tag{ST1.6.7}$$

10) Gauss's formula:

$$\sin[(A-B)/2] = \sin[(a-b)/2]\cos(c/2)/\sin(c/2) \tag{ST1.6.8}$$

Note that one, two, or three right angles may coexist in the same right spherical triangle!

The **orthodrome** between points *A* and *B* on the surface of a geode or Earth is the shortest distance between *A* and *B* on this surface; the orthodrome is a segment of a great circle (e.g. a meridian) passing through both *A* and *B*.

A **loxodrome** (from the Greek λοξος = slanted and δρομος = course) is a path on the Earth's surface that is followed when a compass is kept pointing in the same direction: it is a straight line on a Mercator[167] projection of the globe, precisely because such a projection is designed to have the property that all paths along the Earth's surface that preserve the same directional bearing appear as straight lines. Nunes[168] thought that the loxodrome was the shortest distance between two points on a sphere: he was wrong. Many centuries ago, it was difficult for a ship's navigator to follow a great circle because this required constant changes of compass heading. The solution was to follow a loxodrome, also known as a rhumb line, by navigating along a constant direction. In middle latitudes, at least, this did not lengthen the journey unduly. If a loxodrome is continued indefinitely around a sphere, it will produce a spherical spiral, or a logarithmic spiral on a polar projection. The distance between two points, measured along a loxodrome, is simply the absolute value of the secant of the bearing (azimuth) times the north–south distance (except for circles of latitude). The airports of Helsinki, Finland and Anchorage, AK are almost on the same parallel, so one can fly due W from Helsinki and get comfortably close to Anchorage, but it surely is not the shortest path: the loxodrome distance would be 9493 km, while the orthodrome, the shortest distance, is 6520 km. The loxodrome spirals from one pole to the other, with an angle setting equal to the "compass setting." Close to the poles, the loxodromes closely resemble logarithmic spirals. The total length of the loxodrome from the N pole to the S pole is, assuming a perfect sphere, the length of the meridian divided by the cosine of the bearing away from true North. On a sphere which has coordinates $\phi$ (**latitude**), $\lambda$ (**longitude**), and $\alpha$ (**azimuth**), the equation of a loxodrome is

$$gd^{-1}(\phi) = \cosh^{-1}(\sec\phi) = \ln[\sec\phi(1+\sin\phi)] \tag{ST1.6.9}$$

$$\lambda = \tan\alpha\, gd^{-1}(\phi) + \lambda_0 \tag{ST1.6.10}$$

166 John Napier, Eighth Laird of Merchiston (1550–1617).
167 Gerardus Mercator = Gerhard de Cremer (1512–1594).
168 Pedro Nunes (1502–1578).

**Special Topic ST1.6 (Continued)**

where $arcgd(\phi) \equiv gd^{-1}(\phi)$ is the inverse Gudermannian[169] function, and $\lambda_0$ is the longitude where the loxodrome passes the equator. Here the Gudermannian function is

$$g\,dz \equiv 2\tan^{-1}(\exp(z)) - \pi/2 \tag{ST1.6.11}$$

while the inverse Gudermannian function is given by

$$gd^{-1}(z) \equiv \ln[\tan(\pi/4 + z/2)] = \int_{z=z_1}^{z=z_2} \sec z\,dz = \ln(\sec z + \tan z) \tag{ST1.6.12}$$

Another way of defining a loxodrome is

$$\tan\alpha = (\lambda_2 - \lambda_1)/\left[\cosh^{-1}(\sec\phi_2) - \cosh^{-1}(\sec\phi_1)\right] = \frac{(\lambda_2 - \lambda_1)}{\int_{\phi=\phi_1}^{\phi=\phi_2} \sec\varphi\,d\phi} \tag{ST1.6.13}$$

The **celestial sphere** is a fixed sphere of infinite radius, concentric with the center of the Earth. The celestial North and South poles ($P_N$ and $P_S$) are an extension of the Earth's North and South poles to infinity. The celestial equator is the great circle whose poles are $P_N$ and $P_S$. The (local) zenith point $Z$ is the point vertically above an observer at some arbitrary point on the Earth's surface: the point at the antipodes is the nadir point $Z'$ (Figure 1.35). The (local) celestial horizon corresponds to what we call conventionally the horizon, but is strictly the great circle whose poles are $Z$ and $Z'$. Consider a point $B$ within view of the observer (e.g. a star): the great circle that passes through $P_N$, $P_S$, and $B$ is called the hour circle. The hour circle of $Z$ is called the (local) observer's meridian. The astronomical triangle is $\Delta P_N Z B$ (or $\Delta P_N Z' B$ in the southern hemisphere), with arc $\overset{\frown}{PNZ} = 90^\circ - \lambda$ = co-latitude of observer or codeclination of zenith), arc $\overset{\frown}{P_N B} = 90^\circ - d$ = codeclination of star, arc $\overset{\frown}{BZ} = 90^\circ - h$ = coaltitude of star, angle $\angle ZP_N B = t$ = hour angle of star, angle $\angle P_N ZB = z$ = azimuth of star, and finally angle $\angle ZBP_N = \phi$ = position angle of star.

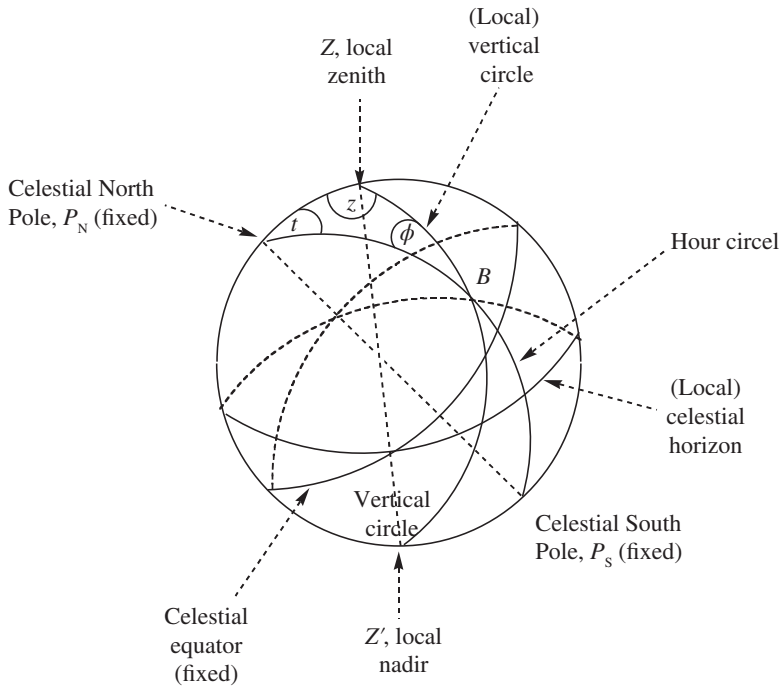



**Figure 1.35** The celestial sphere, with local zenith $Z$, and a point $B$ (a star). Note the spherical triangle $BZP_N$ with interior angles $\phi$, $z$, and $t$.

169 Christoph Gudermann (1798–1852).

## Special Topic ST1.6 (Continued)

**Problem ST1.6.1** Given two sides ($a = 135.8233^\circ$, $c = 60.0817^\circ$) and the included angle ($B = 142.2100^\circ$) of a spherical triangle, compute $b$, $A$, and $C$ [1.42].

**Problem ST1.6.2** Given two angles ($A = 57.9480^\circ$, $B = 137.3425^\circ$) and the included side ($c = 94.8017^\circ$) of a spherical triangle, compute $a$, $b$, and $C$ [1.42].

**Problem ST1.6.3** Compute the orthodrome, or shortest distance on the Earth's surface, between San Francisco, CA (SFO, latitude $37^\circ\,40'\,\text{N} = 37.17^\circ\,\text{N}$; longitude $122^\circ\,26'\,\text{W} = 122.43^\circ\,\text{W}$) and Tokyo, Japan (TYO, latitude $35^\circ\,42'\,\text{N} = 35.70^\circ\,\text{N}$, longitude $139^\circ\,46'\,\text{E} = 139.77^\circ\,\text{E}$), assuming that the Earth is a sphere of circumference 40 million meters. Compute also the initial compass heading as you leave San Francisco in the direction of Tokyo (this compass heading changes continuously as you travel the orthodrome).

## Special Topic ST1.7 Summary of Several Unit Systems for Electromagnetism

A convenient way of comparing many unit systems used in electromagnetism is to rewrite Maxwell's equations, by introducing six constants ($k_1$–$k_6$) plus the dimensionless Sommerfeld fine-structure constant $\alpha$, and remembering that for SI $\varepsilon_0 \equiv 10^7/4\pi c^2$ and $\mu_0 \equiv 4\pi \times 10^{-7}$:

$$\nabla \cdot \boldsymbol{D} \equiv 4\pi k_1 \rho \tag{ST1.7.1}$$

$$\nabla \cdot \boldsymbol{B} \equiv 0 \tag{ST1.7.2}$$

$$\nabla \times \boldsymbol{H} \equiv 4\pi k_2 k_3 \boldsymbol{J} + (k_2 k_3/k_1)(\partial \boldsymbol{D}/\partial t) \tag{ST1.7.3}$$

$$\nabla \times \boldsymbol{E} + (1/k_3)(\partial \boldsymbol{B}/\partial t) \equiv 0 \tag{ST1.7.4}$$

$$\boldsymbol{D} = k_4 \boldsymbol{E} + k_5 \boldsymbol{P} \tag{ST17.5}$$

$$\boldsymbol{H} = (1/k_6)\boldsymbol{B} - k_5 \boldsymbol{M} \tag{ST1.7.6}$$

$$\alpha \equiv 1/137.035\,999\,68 = e^2/\hbar c \ \ (\text{in cgs-esu}) = e^2/4\pi\varepsilon_0 \hbar c \ \ (\text{in SI}) \tag{ST1.7.7}$$

The values of the six constants $k_1$–$k_6$ are listed in Table ST1.7.1. The resultant equations are given in Table ST1.7.2. There is some ambiguity in defining magnetic fields in the "Hartree" atomic units: we follow here the Gaussian convention, in which a plane electromagnetic field has electric and magnetic fields of equal magnitudes in vacuum. In the alternate Hartree–Lorentz convention (not used here), the magnetic field is derived from the Lorentz force, and the magnetic induction $\boldsymbol{B}$ will contain $\alpha$ implicitly.

**Table ST1.7.1** Constants $k_1$ through $k_6$ needed to write Eqs. (ST1.7.1)–(ST1.7.7) in several unit systems.

| Unit system | [L] | [M] | [T] | [Q] | $k_1$ | $k_2$ | $k_3$ | $k_4$ | $k_5$ | $k_6$ |
|---|---|---|---|---|---|---|---|---|---|---|
| cgs-esu (unrat.) | cm | g | s | — | 1 | 1 | 1 | 1 | $4\pi$ | $c^{-2}$ |
| cgs-emu (unrat.) | cm | g | s | — | 1 | 1 | 1 | $c^{-2}$ | $4\pi$ | 1 |
| cgs-Gaussian (unrat.) | cm | g | s | — | 1 | $c^{-2}$ | $c$ | 1 | $4\pi$ | 1 |
| cgs-Heaviside-Lorentz (rat.) | cm | g | s | — | $1/4\pi$ | $(4\pi c^2)^{-1}$ | $c$ | 1 | 1 | 1 |
| au (Hartree) (unrat.) | $a_0$ | $m_e$ | $t_0$ | | 1 | 1 | $\alpha$ | 1 | $4\pi$ | $c^{-2}$ |
| SI (rat.) and SI′ (rat.) | m | kg | s | C | $1/4\pi$ | $1/4\pi$ | 1 | $10^7/4\pi c^2$ | 1 | $4\pi \times 10^{-7}$ |

Here, $c$ = speed of light in cgs or SI units, $a_0$ is the bohr radius (see Section 2.1), and $t_0 = \left(2\varepsilon_0^2 h^3/\pi m_e e^4\right)$. The label "rat." ≡ rationalized means that a factor of $4\pi$ appears explicitly in some of the Maxwell equations (ST1.7.1)–(ST1.7.4), but not in the constitutive equations (ST1.7.5) and (ST1.7.6). For "unrat." $4\pi$ is absent in Maxwell's equations, but reappears in the constitutive equations.

## Special Topic ST1.7 (Continued)

**Table ST1.7.2** Maxwell's and constitutive equations, and Lorentz force in various unit systems.

| Unit system | Maxwell equations | Constitutive equations | Lorentz force |
|---|---|---|---|
| cgs-esu (unrationalized) | $\nabla\cdot\boldsymbol{D}=4\pi\rho,\ \nabla\cdot\boldsymbol{B}=0,\ \nabla\times\boldsymbol{H}=4\pi\boldsymbol{J}+\partial\boldsymbol{D}/\partial t,\ \nabla\times\boldsymbol{E}+\partial\boldsymbol{B}/\partial t=0$ | $\boldsymbol{D}=\boldsymbol{E}+4\pi\boldsymbol{P}\ \boldsymbol{H}=c^2\boldsymbol{B}-4\pi\boldsymbol{M}$ | $\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}$ |
| cgs-emu (unrationalized) | $\nabla\cdot\boldsymbol{D}=4\pi\rho,\ \nabla\cdot\boldsymbol{B}=0,\ \nabla\times\boldsymbol{H}=4\pi\boldsymbol{J}+\partial\boldsymbol{D}/\partial t,\ \nabla\times\boldsymbol{E}+\partial\boldsymbol{B}/\partial t=0$ | $\boldsymbol{D}=c^{-2}\boldsymbol{E}+4\pi\boldsymbol{P}\ \boldsymbol{H}=\boldsymbol{B}-4\pi\boldsymbol{M}$ | $\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}$ |
| cgs-Gaussian (unrationalized) | $\nabla\cdot\boldsymbol{D}=4\pi\rho,\ \nabla\cdot\boldsymbol{B}=0,\ c\nabla\times\boldsymbol{H}=4\pi\boldsymbol{J}+\partial\boldsymbol{D}/\partial t,\ c\nabla\times\boldsymbol{E}+\partial\boldsymbol{B}/\partial t=0$ | $\boldsymbol{D}=\boldsymbol{E}+4\pi\boldsymbol{P}\ \boldsymbol{H}=\boldsymbol{B}-4\pi\boldsymbol{M}$ | $\boldsymbol{E}+c^{-1}\boldsymbol{v}\times\boldsymbol{B}$ |
| cgs-Heaviside–Lorentz (rationalized) | $\nabla\cdot\boldsymbol{D}=\rho,\ \nabla\cdot\boldsymbol{B}=0,\ c\nabla\times\boldsymbol{H}=\boldsymbol{J}+\partial\boldsymbol{D}/\partial t,\ c\nabla\times\boldsymbol{E}+\partial\boldsymbol{B}/\partial t=0$ | $\boldsymbol{D}=\boldsymbol{E}+\boldsymbol{P}\ \boldsymbol{H}=\boldsymbol{B}-\boldsymbol{M}$ | $\boldsymbol{E}+c^{-1}\boldsymbol{v}\times\boldsymbol{B}$ |
| au (Hartree) (unrationalized) | $\nabla\cdot\boldsymbol{D}=4\pi\rho,\ \nabla\cdot\boldsymbol{B}=0,\ \alpha^{-1}\nabla\times\boldsymbol{H}=4\pi\boldsymbol{J}+\partial\boldsymbol{D}/\partial t,\ \nabla\times\boldsymbol{E}+\alpha\partial\boldsymbol{B}/\partial t=0$ | $\boldsymbol{D}=\boldsymbol{E}+4\pi\boldsymbol{P}\ \boldsymbol{H}=c^2\boldsymbol{B}-4\pi\boldsymbol{M}$ | $\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}$ |
| SI (and SI′) (rationalized) | $\nabla\cdot\boldsymbol{D}=\rho,\ \nabla\cdot\boldsymbol{B}=0,\ \nabla\times\boldsymbol{H}=\boldsymbol{J}+\partial\boldsymbol{D}/\partial t,\ \nabla\times\boldsymbol{E}+\partial\boldsymbol{B}/\partial t=0$ | $\boldsymbol{D}=\varepsilon_0\boldsymbol{E}+\boldsymbol{P}\ \boldsymbol{H}=\mu_0^{-1}\boldsymbol{B}-\boldsymbol{M}$ | $\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}$ |

For au, the unit of energy is the Hartree, not the Rydberg, and the Gaussian convention is used: an electromagnetic wave in vacuum has $\boldsymbol{E}$ and $\boldsymbol{H}$ components of equal magnitude.

## Special Topic ST1.8 Nonlinear Optics

When materials have **nonlinear optical characteristics**, then Eq. (1.9.62) must be modified by writing $\boldsymbol{P}$ as a power series in the electric field $\boldsymbol{E}$:

$$\begin{aligned}\boldsymbol{P}&=\chi^{(0)}_{\text{esu}}+\chi^{(1)}_{\text{esu}}\boldsymbol{E}+\chi^{(2)}_{\text{esu}}\boldsymbol{EE}+\chi^{(3)}_{\text{esu}}\boldsymbol{EEE}+\ldots\quad(\text{esu})\\ \boldsymbol{P}&=\varepsilon_0\left[\chi^{(0)}_{\text{SI}}+\chi^{(1)}_{\text{SI}}\boldsymbol{E}+\chi^{(2)}_{\text{SI}}\boldsymbol{EE}+\chi^{(3)}_{\text{SI}}\boldsymbol{EEE}+\ldots\right]\quad(\text{SI})\\ \boldsymbol{P}&=\chi^{(0)}_{\text{SI}'}+\chi^{(1)}_{\text{SI}'}\boldsymbol{E}+\chi^{(2)}_{\text{SI}'}\boldsymbol{EE}+\chi^{(3)}_{\text{SI}'}\boldsymbol{EEE}+\ldots\quad(\text{SI}')\end{aligned}\tag{ST1.8.1}$$

where the units are (i) dimensionless for $\chi^{(1)}_{\text{esu}}$,$\text{cm}^2\ \text{s}\ \text{C}^{-1}$ for $\chi^{(2)}_{\text{esu}}$, and $\text{cm}^4\ \text{s}\ \text{C}^{-2}$ for $\chi^{(3)}_{\text{esu}}$; (ii) dimensionless for $\chi^{(1)}_{\text{SI}}$, $\text{m}\ \text{V}^{-1}$ (in practice, a few $\text{pm}\ \text{V}^{-1}$) for $\chi^{(2)}_{\text{SI}}$, and$\text{m}^2\ \text{V}^{-2}$ for $\chi^{(3)}_{\text{SI}}$; (iii) $\text{C}^2\ \text{J}^{-1}\ \text{m}^{-1}$ for $\chi^{(1)}_{\text{SI}'}$, $\text{C}^3\ \text{J}^{-2}$ for $\chi^{(2)}_{\text{SI}'}$, and $\text{C}^4\ \text{m}\ \text{J}^{-3}$ for $\chi^{(3)}_{\text{SI}'}$.$\chi^{(1)}$ is a rank-two tensor with 9 terms; $\chi^{(2)}$ is a rank-three tensor with 27 terms; $\chi^{(3)}$ is a rank-four tensor, with 81 terms.

If the molecule or the solid has a center of inversion symmetry, then the permanent dipole moment $\boldsymbol{m}_0$ vanishes, as do the even-rank tensors $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and the even-rank tensors $\chi^{(2)}$, $\chi^{(4)}$, $\chi^{(6)}$. All matter, with or without a center of inversion symmetry, has nonzero values for the odd-rank molecular tensors $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and all odd-rank tensors $\chi^{(1)}$, $\chi^{(3)}$, $\chi^{(5)}$, etc. If the crystal has symmetry, then the number of unique tensor components is vastly reduced. The components have values that depend seriously on the frequency of the electromagnetic radiation used to probe them. A practical application of nonlinear optics is frequency doubling of the high-powered near-infrared Nd-YAG laser light from its most intense wavelength 1.06 μm to the frequency-doubled wavelength 530 nm, typically using expensive lithium niobate crystals.

To confuse theoreticians, the experimentalists use half-sized $\boldsymbol{d}$ tensor components:

$$2d_{ij}\equiv\chi^{(2)}_{ij}\tag{ST1.8.2}$$

Finally, one can define an electric displacement vector $\boldsymbol{D}(\omega)$, and three tensors, for the **optical dielectric constant** $\boldsymbol{\varepsilon}(\omega)$, the effective **optical susceptibility** $\chi_{\text{eff}}(\omega)$, and the **optical index of refraction** $\boldsymbol{n}(\omega)$:

$$\boldsymbol{D}(\omega)=\varepsilon_0\boldsymbol{E}(\omega)+\boldsymbol{P}(\omega)=\varepsilon_0\boldsymbol{\varepsilon}(\omega)\boldsymbol{E}(\omega)\quad(\text{SI});\quad\boldsymbol{D}(\omega)=\boldsymbol{E}(\omega)+4\pi\boldsymbol{P}(\omega)=\boldsymbol{\varepsilon}(\omega)\boldsymbol{E}(\omega)\quad(\text{cgs})\tag{ST1.8.3}$$

**Special Topic ST1.8 (Continued)**

$$\boldsymbol{P}(\omega) = \varepsilon_0 \chi_{eff}(\omega)\boldsymbol{E}(\omega) \quad \text{(SI)}; \quad \boldsymbol{P}(\omega) = \chi_{eff}(\omega)\boldsymbol{E}(\omega) \quad \text{(cgs)} \tag{ST1.8.4}$$

$$\boldsymbol{n}(\omega)\boldsymbol{n}(\omega) = \boldsymbol{\varepsilon}(\omega) = 1 + 4\pi\ \chi_{\text{eff}}(\omega) \quad \text{(SI)}; \quad \boldsymbol{n}(\omega)\boldsymbol{n}(\omega) = \boldsymbol{\varepsilon}(\omega) = 1 + 4\pi\ \chi_{\text{eff}}(\omega) \quad \text{(cgs)} \tag{ST1.8.5}$$

including tensor effects:

$$\chi_{\text{eff}} = \chi^{(1)} + \chi^{(2)} :\ \boldsymbol{E} + \chi^{(3)} :\ \boldsymbol{EE} + \chi^{(4)} :\ \boldsymbol{EEE} + \ldots \tag{ST1.8.6}$$

If a sample is subjected to the sum of two electrical fields, a direct current (DC), i.e. frequency-independent component $E_{\text{DC}}$ and a frequency-dependent "optical field" $E(\omega)$:

$$E = E_{\text{DC}} + E(\omega) = E_{\text{DC}} + E_0 \cos\ (\omega t - kz) \tag{ST1.8.7}$$

then, ignoring tensor effects, the polarization becomes:

$$P = \chi^{(1)}[E_{\text{DC}} + E_0 \cos\ (\omega t - kz)] + \chi^{(2)}[E_{\text{DC}} + E_0 \cos\ (\omega t - kz)]^2 + \chi^{(3)}[E_{\text{DC}} + E_0 \cos\ (\omega t - kz)]^3 + \ldots \tag{ST1.8.8}$$

Keeping only the expansion terms in $\omega$, and using the trigonometric identity $\cos(3\omega) = 4\cos^3\ \omega - 3\cos\ \omega$, we get:

$$\begin{aligned} P(\omega) &= \chi^{(1)}E_0 \cos\ (\omega t - kz) + 2\chi^{(2)}E_{\text{DC}}E_0 \cos\ (\omega t - kz) + 3\chi^{(3)}E_{\text{DC}}^2 E_0 \cos\ (\omega t - kz) \\ &\quad + (3/4)\chi^{(3)}E_0^3 \cos\ (\omega t - kz) + \ldots \approx \chi_{\text{eff}}E_0 \cos\ (\omega t - kz) \end{aligned} \tag{ST1.8.9}$$

The nonlinear effective refractive index $n(\omega)$ can be defined by

$$n(\omega)^2 \equiv 1 + 4\pi\chi_{\text{eff}} \tag{ST1.8.10}$$

The above then gives:

$$n(\omega)^2 \approx 1 + 4\pi\left[\chi^{(1)} + 2\chi^{(2)}E_{\text{DC}} + 3\chi^{(3)}E_{\text{DC}}^2 + (3/4)\chi^{(3)}E_0^2 + \ldots\right] \tag{ST1.8.11}$$

If the linear refractive index is $n_0(\omega)$, and if one also defines:

$$n_0(\omega)^2 \equiv 1 + 4\pi\chi^{(1)} \tag{ST1.8.12}$$

then one gets:

$$n(\omega)^2 \approx n_0(\omega)^2\left[1 + \left(8\pi\chi^{(2)}E_{\text{DC}}\right)n_0(\omega)^{-2} + \left(12\pi\chi^{(3)}E_{\text{DC}}^2 + 3\pi\chi^{(3)}E_0^2\right)n_0(\omega)^{-2}\right] \tag{ST1.8.13}$$

which, after taking square roots and using the series expansion $[1 + x]^{1/2} = 1 + x/2$, valid for small $x$, yields:

$$n(\omega) \approx n_0(\omega) + 4\pi\chi^{(2)}E_{\text{DC}}n_0(\omega)^{-1} + 6\pi\chi^{(3)}E_{\text{DC}}^2 n_0(\omega)^{-1} + 1.5\pi\chi^{(3)}E_0^2 n_0(\omega)^{-1} \tag{ST1.8.14}$$

It is useful to define the light intensity $I(\omega)$ as

$$I(\omega) \equiv cn_0(\omega)E_0^2/8\pi \quad (cgs) \tag{ST1.8.15}$$

where $c$ is again the speed of light, so that the nonlinear index of refraction, or refractive index, at the frequency $\omega$ can be rewritten as

$$n(\omega) \approx n_0(\omega) + n_1 E_{\text{DC}} + n_2(0)E_{\text{DC}}^2 + n_2(\omega)I(\omega) \quad \text{(cgs)} \tag{ST1.8.16}$$

where the **linear electrooptic (or Pockels)**[170] **effect** is given by

$$n_1 \equiv 4\pi\chi^{(2)}/n_0(\omega) \quad \text{(cgs)} \tag{ST1.8.17}$$

170 Friedrich Carl Alwin Pockels (1865–1913).

**Special Topic ST1.8 (Continued)**

while the **quadratic electrooptic effect** is given by

$$n_2(0) \equiv 6\pi\chi^{(3)}/n_0(\omega) \quad \text{(cgs)} \tag{ST1.8.18}$$

and the **optical Kerr**[171] **effect** is given by

$$n_2(\omega) = 12\pi^2\chi^{(3)}/cn_0(\omega)^2 \quad (cgs) \tag{ST1.8.19}$$

Note that both $n_1$ and $n_2(0)$ depend on the DC electric field, but not on the frequency-dependent light intensity $I(\omega)$, while $n_2(\omega)$ depends on the intensity of the AC electric field at the frequency $\omega$, i.e. on the intensity of the input light $I(\omega)$.

Similar theoretical expressions can be written for the magnetic susceptibility, but the nonlinear effects are much milder.

**Special Topic ST1.9 Conic Sections, Ellipses, and Kepler's[172] Laws of Planetary Motion**

### The ellipsoid

The general equation of a 3D ellipsoid with semiaxes $a$, $b$, and $c$ is

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1 \tag{ST1.9.1}$$

If $a = b > c$, then the ellipsoid becomes an oblate spheroid; if $a = b < c$, then the ellipsoid becomes a prolate spheroid; if $a = b = c$, then the ellipsoid becomes a sphere.

For a 2D ellipse the general equation is

$$x^2/a^2 + y^2/b^2 = 1 \tag{ST1.9.2}$$

where $a$ and $b$ are the major and minor semiaxes (Figure 1.36).

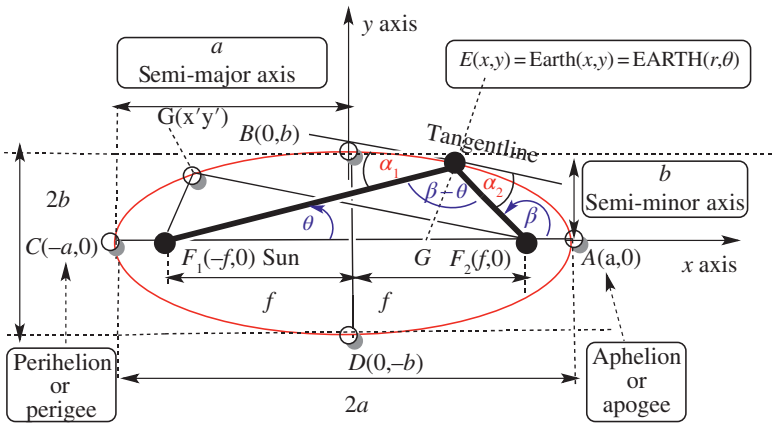



**Figure 1.36** Ellipse with semimajor axes $a$ and $b$, vertices $A(a,0)$, $B(0,b)$, $C(-a,0)$, and $D(0,-b)$. By the definition of an ellipse, the length of the vectors going from one focus to any point on the ellipse and from there to the second focus is constant: $F_1E + EF_2 = F_1G + GF_2 = F_1A + AF_2$. (Here $E(x,y)$ and $G(x',y')$ were chosen arbitrarily.)

171 John Kerr (1824–1907).
172 Johannes Kepler (1571–1630).

**Special Topic ST1.9 (Continued)**

The eccentricity $e$ is defined as

$$e \equiv \left[1 - b^2/a^2\right]^{1/2} \quad \text{(ellipse)} \tag{ST1.9.3}$$

For all ellipses $0 < e < 1$; when the ellipse becomes a circle $a \to b$ and $e \to 0$.

This ellipse is a planar curve such that the sum of the distances from the general point $(x, y)$ on the ellipse to the two foci $F_1(-f, 0)$ and $F_2(f, 0)$ inside the ellipse is constant: these distances are called generator lines. If $f$ is the distance from either focus to the center of the ellipse, it can be shown (Problem ST1.9.1) that (for $a > b$):

$$f^2 = a^2 - b^2 \tag{ST1.9.4}$$

It can also be shown (Problem ST1.9.2) that

$$f = ea \tag{ST1.9.5}$$

The area of an ellipse is given by **(**Problem ST1.9.3):

$$A = \pi ab \tag{ST1.9.6}$$

One can rewrite Eq. (ST1.9.2) using polar coordinates $(r, q)$ with origin at (say) the left-hand focus: after some labor, the result is

$$r = \left(b^2/a\right)/(1 \pm e\cos\ \theta) \tag{ST1.9.7}$$

where there is some ambiguity in sign, depending on how $\theta$ is defined (Problem ST1.9.4).

Ellipses, parabolas, and hyperbolas are called conic sections (Figure 1.37).

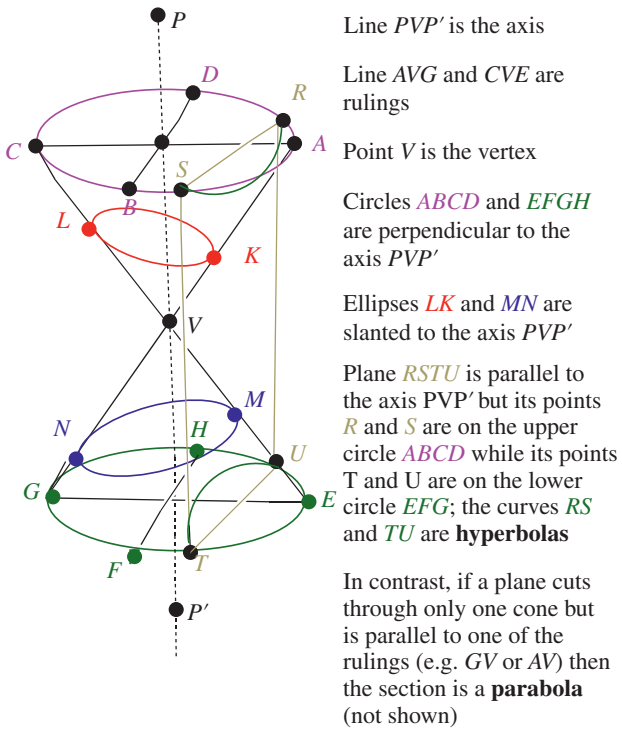



**Figure 1.37** Conic sections, showing circles, ellipses, and hyperbolas.

**Special Topic ST1.9 (Continued)**

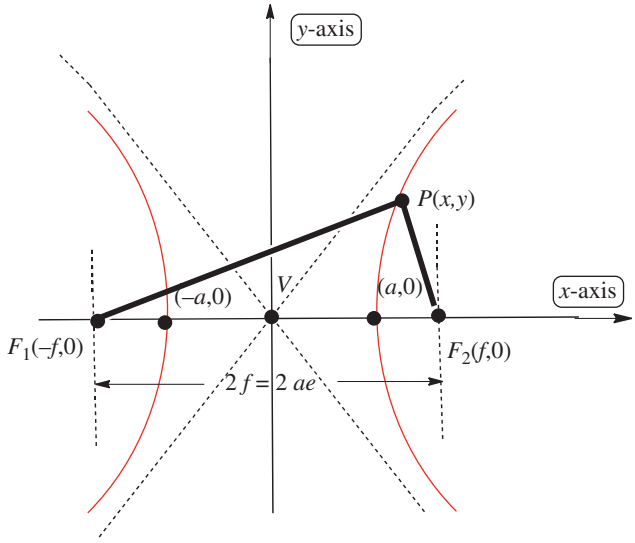



**Figure 1.38** Hyperbola with semimajor axes $a$ and $b$, and vertices $A(a,0)$, $B(0,b)$, and foci $F_1(-f, 0)$ and $F_2(f, 0)$. By the definition of the hyperbola, the total length of the two vectors going from one focus to any point $P(x,y)$ on the hyperbola and from there to the second focus is constant: $F_1P + PF_2 = 2a$. The focal length is $f = (1/2)(F_1V + VF_2)$.

The equation for a hyperbola (Figure 1.38) is

$$x^2/a^2 - y^2/b^2 = 1 \qquad \text{(ST1.9.8)}$$

It is defined as the curve whose points are equidistant from the two foci. The hyperbola has two branches, symmetric with respect to the vertex $V$. The eccentricity $e$ of a hyperbola is

$$e = \left[1 + b^2/a^2\right]^{1/2} \quad \text{(hyperbola)} \qquad \text{(ST1.9.9)}$$

The motion of planets and artificial satellites can be explained by ellipses. It is presumed that eight planets in our solar system (Mercury, Venus, Mars, Earth, Jupiter, Saturn, Uranus, and Neptune) were formed from ejecta in one whoosh from the Sun about $4.6 \times 10^9$ years ago (hence their orbits are almost coplanar), while Pluto is a quasi-planet from outer space that was captured by the Sun's gravitational field (its orbit is quite tilted compared to the other planets). The age of the solar system is estimated from the decay rate of certain long-lived potassium and uranium isotopes ($_{19}K^{40}$, half-life $1.248 \times 10^9$ years, measured in rocks with its locally generated and trapped daughter $_{18}Ar^{40}$ and $_{92}U^{238}$ half-life $4.47 \times 10^9$ years) found on Earth, which presumably were formed at the same time as the Earth formed.

Historically, the motions of the planets and the Sun, as observed from the Earth, were explained by a stationary Earth and several epicycles each per planet or Sun (circles running inside several other circles like a gimbal "cardanic" suspension). Instead, Copernicus[173] found a simpler model of all planets and also the Earth revolving around an almost stationary Sun, in his posthumous work "De Revolutionibus Orbium Coelestium" (1453). After favorable publicity and experimental support of the "Copernican Revolution" by Galileo, Kepler used astronomical data collected by Ulugbek[174] and by Brahe.[175] With these data, Kepler showed in three empirical laws that the motion of all planets (including the Earth) around the Sun is (Kepler law 1) an ellipse with the Sun at one of the two foci (Kepler law 2), the partial areas of

173 Nicolaus Copernicus = Mikolaj Kopernick (1473–1543).
174 Mizra Muhammad Taraghay bin Shakrukh = Ulugbek = Ulugh Beg (1394–1449)
175 Tycho Brahe (1546–1601).

**Special Topic ST1.9 (Continued)**

the ellipse swept by the planet's motion are proportional to the time elapsed in orbit (Kepler law 3) the square of the orbital period of a planet is proportional to the cube of the elliptical semimajor axis $a$. The dramatic impact of Copernicus' ideas, which directly contradicted medieval interpretations of the Jewish Bible (which had implied a stationary Earth), and led in the 1700s to a new use of the term "revolution" for political upsets of any old established order.

Kepler's three laws are explained by Isaac Newton's universal law of gravitation (Eq. (1.1.1)) that pulls the planet toward the Sun, plus an added and essential tangential velocity (due to how the planet was kicked out of the Sun), that prevents the planet from falling into the Sun. The derivations are worked out in Problems ST1.9.6, ST1.9.7, and ST1.9.8.

**Problem ST1.9.1** Prove $f^2 = a^2 - b^2$, Eq. (ST1.9.4).

**Problem ST1.9.2** Prove $f = ea$, Eq. (ST1.9.5).

**Problem ST1.9.3** Show that the area of an ellipse is $f = ea$, Eq. (ST1.9.6).

**Problem ST1.9.4** Prove $r = (b^2/a)/(1 \pm e\cos\theta)$, Eq. (ST1.9.7): there is some ambiguity in sign! {This is a laborious proof}.

**Problem ST1.9.5 Whisper gallery effect: the focusing of sound at the foci of an ellipse**. There is an unusual but important property of a 3D ellipsoid or a 2D ellipse: a sound emitted at one focus is reflected with maximum intensity at the second focus. This is demonstrated in "whisper galleries," e.g. in St. Paul's Cathedral in London, the astronomical observatory on the campus of the University of North Carolina at Chapel Hill, and at Grand Central Station, New York (in front of the Oyster Bar), or at the Natural History Museum, New York, or at the Temple of Heaven in Beijing, or the Gol Gumbaz mausoleum in Bijapur, India, or the Cathedral of Agrigento, Sicily (where secret confessions whispered to a priest in a small confessional were overheard distinctly at a point behind the main altar!), or the Mormon Tabernacle in Salt Lake City, or the gallery in the US Capitol rotunda in Washington, DC, etc. The sound is reflected at point $E(x,y)$ (off an imaginary line that is tangent to the ellipse) with angles $\alpha_1$ and $\alpha_2$ (see Figure 1.36). Prove this effect.

**Problem ST1.9.6 Kepler's first law:** Using Newton's second law and Newton's law of universal gravitation, Eq. (1.1.1), prove that the orbit of any planet of mass $m_P$ around the Sun of mass $M_S$ is an ellipse with the Sun at one of the two foci. In the derivation, (i) assume a constant angular momentum

$$\boldsymbol{L} = m_P \boldsymbol{v} \times \boldsymbol{r} = m_P(\mathrm{d}\boldsymbol{r}/\mathrm{d}t) \times \boldsymbol{r} \qquad \text{(ST1.9.9)}$$

because the pseudo-vector $\boldsymbol{L}$ is normal to both $\boldsymbol{r}$ and also to $(\mathrm{d}^2\boldsymbol{r}/\mathrm{d}t^2)$; (ii) use this constant $\boldsymbol{L}$ to rewrite the centripetal acceleration $m_P(\mathrm{d}\boldsymbol{r}/dt)^2\boldsymbol{r}^{-1}$.

**Problem ST1.9.7 Kepler's second law:** That the partial areas of the ellipse swept by the planet's motion are proportional to the time elapsed in orbit.

**Problem ST1.9.8 Kepler's third law:** Prove that the square of the orbital period $T_P$ of a planet is proportional to the cube of the elliptical semimajor axis $a$. Indeed, empirically in the following table (within the significant digits used), the ratio $a^3 T_P^{-2}$ is almost constant.

| Name (symbol) | Mass (kg) | Radius (Mm) | Revolution period $T_P$ (days) | Rotation period (days) | Semimajor axis $a$ (Gm) | Eccentricity ($e$) | Ratio $a^3 T_P^{-2}$ | Inclination degrees (°) |
|---|---|---|---|---|---|---|---|---|
| Mercury (☿) | $3.3 \times 10^{23}$ | 2.44 | 87.9 | 58.7 | 57.9 | 0.206 | 25.1 | 3.38 |
| Venus (♀) | $4.9 \times 10^{24}$ | 6.05 | 225 | 243 | 108 | 0.007 | 24.9 | 3.86 |
| Earth (⊕) | $6.0 \times 10^{24}$ | 6.38 | 365.25 | 0.997 | 149 | 0.017 | 24.8 | 7.155 |
| Mars (♂) | $6.4 \times 10^{23}$ | 3.39 | 687 | 1.026 | 228 | 0.093 | 25.1 | 5.65 |
| Jupiter (♃) | $1.9 \times 10^{27}$ | 71.9 | 4333 | 0.413 | 779 | 0.048 | 25.2 | 6.09 |
| Saturn (♄) | $5.7 \times 10^{26}$ | 60.4 | 10 800 | 0.483 | 1 430 | 0.056 | 25.1 | 5.51 |
| Uranus (♅) | $8.8 \times 10^{25}$ | 23.5 | 30 700 | 0.451 | 2 870 | 0.047 | 25.1 | 6.48 |
| Neptune (♆) | $1.0 \times 10^{26}$ | 24.6 | 60 200 | 0.625 | 4 500 | 0.008 | 25.1 | 6.43 |
| Pluto (♇) | — | — | 90 700 | — | 5 900 | 0.249 | 25.0 | 17.88 |

### Special Topic ST1.9 (Continued)

Note also from the last column that the relatively large inclination of the orbit of Pluto (17.88°), relative to the Sun's equator, is why in 2006 the International Astronomical Union has downgraded Pluto from "true planet" (formed, like the others, c. five billion years ago from presumably simultaneous ejecta from the Sun) to "dwarf planet."

**Problem ST1.9.9 Artificial Earth satellite in low orbit.** Assume that a rocket launched from the surface of the Earth (radius = $6.380 \times 10^6$ m, mass $M_E = 6.0 \times 10^{24}$ kg) reaches an altitude of 370 km, i.e. $6.750 \times 10^6$ m above the center of the Earth before its engine is shut off. The artificial satellite assumes an elliptical path around the Earth, where the center of the Earth is at one of the foci. Using Eq. (ST7), estimate the time $T_P$ for a full orbit.

**Problem ST1.9.11 Artificial Earth satellite in geostationary orbit.** Assume that a rocket launched from the surface of the Earth (radius = $6.380 \times 10^6$ m, mass $M_E = 6.0 \times 10^{24}$ kg) reaches an altitude $x$ large enough for the satellite to have an orbit time of exactly $T_P$ = 23 hours and 56 minutes = 86 160 seconds, i.e. a geostationary orbit, such that the satellite is always at the zenith for a local observer. Using Eq. (ST1.8.17), estimate $x$.

### Special Topic ST1.10 Ellipsometry in Detail

In a beam of light (TEM radiation), the three vectors of electric field ***E***, magnetic field ***H***, and light propagation direction ***k*** are mutually perpendicular, forming a right-handed set (***E***, ***H***, ***k***) by Lenz' law. The anisotropy or **polarization** is usually specified by the direction of the vector ***E*** relative to some fixed laboratory axis system. Light **emitted** by individual atoms or molecules is **always polarized**, while a macroscopic light beam emitted by randomly oriented atoms or molecules will consist of a huge number (Avogadro's number's worth) of such emitters, and the direction of ***E*** relative to laboratory coordinates will become random: such light is called **unpolarized or natural light**.

A **polarizer** or **polarization filter** (usually an oriented crystal) can suppress most of the light polarized in one direction, relative to the optical axes of the crystal, and transmit most of the component polarized in the perpendicular direction: if the polarization is complete, it will be **plane-polarized (or linearly polarized)** light. In general, the polarized light will consist of two mutually perpendicular plane-polarized components; depending on the amplitude of these two waves and their relative phase, the combined electric vector traces out an ellipse; the wave is **elliptically polarized**. Elliptical and plane polarization can be interconverted by **birefringent crystal filters**. In a **circularly polarized** wave ***E*** moves like in a coil. The equipment for ellipsometry is presented in Section 10.25.

If a linearly polarized (plane-polarized) light of intensity $I_0$ is incident onto a polarizer, then the intensity of the transmitted light $I$ will depend upon the angle $\theta$ between the direction of the light polarization and the orientation of maximum polarization of the crystal: $I = I_0 \cos^2 \theta$. **Ellipsometry** measures the orientation of polarized light undergoing oblique reflection from a sample surface. Linearly polarized light, when reflected from a surface, will become elliptically polarized because of presence of the thin layer of the boundary surface between two media. Dependence between optical constants of a layer and parameters of elliptically polarized light can be found on basis of the Fresnel formulas described above.

Using any of several unit systems (see Table 1.11) and Cartesian coordinates, Maxwell's equations for $\boldsymbol{J} = 0$ and $\rho = 0$ can be combined into a 6 × 6 matrix form:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -\partial/\partial z & \partial/\partial y \\ 0 & 0 & 0 & \partial/\partial z & 0 & -\partial/\partial x \\ 0 & 0 & 0 & -\partial/\partial y & \partial/\partial x & 0 \\ 0 & \partial/\partial z & -\partial/\partial y & 0 & & 0 \\ -\partial/\partial z & 0 & \partial/\partial x & 0 & 0 & 0 \\ \partial/\partial y & -\partial/\partial x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \\ H_x \\ H_y \\ H_z \end{pmatrix} = Q(\partial/\partial t) \begin{pmatrix} D_x \\ D_y \\ D_z \\ B_x \\ B_y \\ B_z \end{pmatrix} \qquad \text{(ST1.10.1)}$$

**Special Topic ST1.10 (Continued)**

where $Q \equiv 1$ for SI and cgs units, but $Q \equiv (1/c)$ for Gaussian units. To abbreviate Eq. (ST1.10.1), let $\boldsymbol{G}$ be the 6 × 1 column vector involving the components of electric field $\boldsymbol{E}$ and of magnetic field $\boldsymbol{H}$ on the left-hand side; let $\boldsymbol{C}$ be the 6 × 1 column vector involving the electric displacement $\boldsymbol{D}$ and the magnetic induction $\boldsymbol{B}$ on the right-hand side; let the 6 × 6 traceless matrix on the left-hand side be called $\boldsymbol{O}$, then Eq. (ST1.10.1) becomes:

$$\boldsymbol{OG} = \boldsymbol{Q}(\partial/\partial t)\boldsymbol{C} \quad \text{(SI)}; \quad \boldsymbol{OG} = \boldsymbol{Q}(\partial/\partial t)\boldsymbol{C} \quad \text{(cgs)} \tag{ST1.10.2}$$

The constitutive equations are:

$$\boldsymbol{D} \equiv \varepsilon_0 \boldsymbol{\varepsilon E} \quad \text{(SI)}; \quad \boldsymbol{D} \equiv \boldsymbol{\varepsilon E} \quad \text{(cgs-esu)} \tag{(1.9.12)}$$

$$\boldsymbol{B} \equiv \mu_0 \boldsymbol{\mu H} \quad \text{(SI)}; \quad \boldsymbol{B} \equiv \boldsymbol{\mu H} \quad \text{(cgs-emu)} \tag{(1.9.13)}$$

$$\boldsymbol{D} \equiv \boldsymbol{\rho H} \quad \text{(SI)}; \quad \boldsymbol{D} \equiv \boldsymbol{\rho H} \quad \text{(cgs)} \tag{ST1.10.3}$$

$$\boldsymbol{B} \equiv \boldsymbol{\rho}' \boldsymbol{E} \quad \text{(SI)}; \quad \boldsymbol{B} \equiv \boldsymbol{\rho}' \boldsymbol{E} \quad \text{(cgs)} \tag{ST1.10.4}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\varepsilon}$, $\boldsymbol{\rho}$, and $\boldsymbol{\rho}'$ are **rank-two** tensors with complex matrix elements that depend on the electronic and magnetic properties of the material: $\boldsymbol{\varepsilon}$ is the dielectric tensor or relative electric permittivity tensor, $\boldsymbol{\mu}$ is the magnetic permeability tensor, $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$ are optical rotation tensors, needed to describe optical activity, Faraday effect, etc. In Eqs. (ST1.10.3) and (ST1.10.4), it is assumed that the medium has no nonlinear optical effects ($\boldsymbol{D}$ is taken to be linearly dependent on $\boldsymbol{E}$) and no nonlinear magnetic effects ($\boldsymbol{B}$ is assumed to be linearly dependent on $\boldsymbol{H}$). Equations (ST1.10.1) and (ST1.10.2) do not directly include the linear response of the material to $\boldsymbol{E}$ or $\boldsymbol{H}$; that linear response is given by Eqs. (1.9.12), (1.9.13), (ST1.10.3), and (ST1.10.4), which can be recast into a new and formally convenient form: the column vectors $\boldsymbol{G}$ and $\boldsymbol{C}$ are next linked by a second 6 × 6 matrix $\boldsymbol{M}$, called the **optical matrix**:

$$\boldsymbol{MG} = \boldsymbol{C} \tag{ST1.10.5}$$

$\boldsymbol{M}$ contains all the information about the anisotropic optical properties of the medium that supports the electromagnetic fields; some of its 36 tensor components are the target of the ellipsometric measurements of anisotropic crystals. This optical matrix $\boldsymbol{M}$ is thus defined by

$$\boldsymbol{M} \equiv \begin{pmatrix} \boldsymbol{\varepsilon} & \boldsymbol{\rho} \\ \boldsymbol{\rho}' & \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} & \rho_{11} & \rho_{12} & \rho_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} & \rho_{21} & \rho_{22} & \rho_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} & \rho_{31} & \rho_{32} & \rho_{33} \\ \rho'_{11} & \rho'_{12} & \rho'_{13} & \mu_{11} & \mu_{12} & \mu_{13} \\ \rho'_{21} & \rho'_{22} & \rho'_{23} & \mu_{21} & \mu_{22} & \mu_{23} \\ \rho'_{31} & \rho'_{32} & \rho'_{33} & \mu_{31} & \mu_{32} & \mu_{33} \end{pmatrix} \tag{ST1.10.6}$$

where $\boldsymbol{\varepsilon} \equiv [\varepsilon_{ij} = M_{ij}(i, j = 1, \ldots, 3)]$ is the complex dielectric (or relative electric permittivity) tensor, $\boldsymbol{\mu} \equiv [\mu_{ij} = M_{i+3, j+3},\ i, j = 1, \ldots, 3]$ is the permeability tensor, and $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$ are the optical rotation tensors $\boldsymbol{\rho} \equiv [\rho_{ij} = M_{i, j+3},\ i, j = 1, \ldots, 3]$, and $\boldsymbol{\rho}' \equiv [\rho_{ij} = M_{i+3, j},\ i, j = 1, \ldots, 3]$.

The 3 × 3 complex refractive index tensor $N \equiv n - \mathrm{i}k$ is related to the dielectric tensor $\boldsymbol{\varepsilon}$ and the magnetic tensor $\boldsymbol{\mu}$ by the **Maxwell relation**:

$$\varepsilon_0 \mu_0 \boldsymbol{\varepsilon} \cdot \boldsymbol{\mu} = (\boldsymbol{n} - \mathrm{i}\boldsymbol{k})^2 \quad \text{(SI)}; \quad \boldsymbol{\varepsilon} \cdot \boldsymbol{\mu} = (\boldsymbol{n} - \mathrm{i}\boldsymbol{k})^2 \quad \text{(cgs)} \tag{ST1.10.7}$$

where $\mathrm{i} = (-1)^{1/2}$, $\boldsymbol{n}$ is the ratio of the phase velocity of an electromagnetic wave in vacuum to its phase velocity in the anisotropic crystal (Snell's law), while $\boldsymbol{k}$ is related to the absorption of light energy during the propagation of the light wave within the crystal. Of course, for nonmagnetic media $\mu = 1$.

The experimental aim is to derive from ellipsometric data the tensor $\boldsymbol{\varepsilon}$

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{pmatrix} \tag{ST1.10.8}$$

**Special Topic ST1.10 (Continued)**

which has three complex principal-axis values $\varepsilon_1$, $\varepsilon_2$, $\varepsilon_3$ in a Cartesian principal-axis system 1, 2, 3 that is oriented in some fashion relative to the known crystal unit cell axes of the anisotropic crystalline material. From these complex values, assuming $\mu = 1$, one obtains the principal-axis values of the real and imaginary parts of the refractive index tensor:

$$\begin{aligned} \varepsilon_1 &= (n_1 - \mathrm{i}k_1)^2 \\ \varepsilon_2 &= (n_2 - \mathrm{i}k_2)^2 \\ \varepsilon_3 &= (n_3 - \mathrm{i}k_3)^2 \end{aligned} \tag{ST1.10.9}$$

Thus, in principle, one can obtain the complex index of refraction from the upper left-hand quarter of the optical tensor $\boldsymbol{M} = [M_{ij}, (i, j = 1, \ldots, 3)]$. We now explain how these optical constants can be derived from ellipsometry.

By combining Eqs. (ST1.10.1) and (ST1.10.7), the full spatial wave equation to be solved becomes:

$$\boldsymbol{OG} = Q(\partial/\partial t)\boldsymbol{MG} \tag{ST1.10.10}$$

The geometry of the ellipsometric measurement is shown in Figure 1.13. The anisotropic material medium under study has coordinate $z > 0$, while the isotropic ambient medium (air) has $z < 0$; the electromagnetic field components of interest are the tangential components $E_x$, $E_y$, $H_x$, and $H_y$; the plane of incidence is the $xz$ plane.

Let the time dependence of the electromagnetic fields be the phase factor $\exp(\mathrm{i}\omega t)$, where $\omega$ is the angular frequency of the light. By the symmetry of Figure 1.20, there is no variation of any field component along the $y$-direction: hence, in Eq. (ST1.10.1):

$$(\partial/\partial y) = 0 \tag{ST1.10.11}$$

can be set. If $\xi$ denotes the $x$ component of the wave vector of the incident wave of frequency $\omega$, then all fields should vary in the $x$-direction as $\exp(\mathrm{i}\omega\xi/c)$; thus, another condition for Eq. (ST1.8.1) is

$$(\partial/\partial x) = -\mathrm{i}\omega\xi/c \tag{ST1.10.12}$$

These two conditions reduce Eq. (ST1.9.1) to:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -\partial/\partial z & 0 \\ 0 & 0 & 0 & \partial/\partial z & 0 & \mathrm{i}\omega\xi/c \\ 0 & 0 & 0 & 0 & -\mathrm{i}\omega\xi/c & 0 \\ 0 & \partial/\partial z & -\partial/\partial y & 0 & 0 & 0 \\ -\partial/\partial z & 0 & -\mathrm{i}\omega\xi/c & 0 & 0 & 0 \\ \partial/\partial y & \mathrm{i}\omega\xi/c & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \\ H_x \\ H_y \\ H_z/\xi \end{pmatrix} = (\mathrm{i}\omega/c)M \begin{pmatrix} E_x \\ E_y \\ E_z \\ H_x \\ H_y \\ H_z/\xi \end{pmatrix} \tag{ST1.10.13}$$

The above matrix is equivalent to two linear homogeneous algebraic equations (the third and sixth equations) and four linear differential equations (the first, second, fourth, and fifth equations); the third equation is

$$-\xi H_y = M_{31}E_x + M_{32}E_y + M_{33}E_z + M_{34}H_x + M_{35}H_y + M_{36}H_z \tag{ST1.10.14}$$

and the sixth equation is

$$\xi E_y = M_{61}E_x + M_{62}E_y + M_{63}E_z + M_{64}H_x + M_{65}H_y + M_{66}H_z \tag{ST1.10.15}$$

We need the tangential components of $\boldsymbol{E}$ and $\boldsymbol{H}$ (namely $E_x$, $E_y$, $H_x$, and $H_y$) as explicit variables, so we want to get rid of $H_z$ and $E_z$; this is done by solving Eqs. (ST1.10.14) and (ST1.10.15) explicitly and simultaneously for $E_z$ and $H_z$ in terms of $E_x$, $E_y$, $H_x$, and $H_y$. These expressions for $E_z$ and $H_z$ are then substituted into the remaining four differential equations, to

**Special Topic ST1.10 (Continued)**

produce four linear homogeneous first-order differential equations in the four tangential field variables $E_x$, $E_y$, $H_x$, $H_z$. For convenience, we define a 4 × 1 generalized field vector $\boldsymbol{\psi}$:

$$\boldsymbol{\psi} \equiv \begin{pmatrix} E_x \\ H_y \\ E_y \\ -H_x \end{pmatrix} \tag{ST1.10.16}$$

The 4 × 4 matrix involving $\boldsymbol{\psi}$ can be abbreviated to:

$$(\partial y/\partial z) = -\mathrm{i}(\omega/c)\boldsymbol{\Delta}\boldsymbol{\psi} = -\mathrm{i}(\omega/c)\begin{pmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} & \Delta_{14} \\ \Delta_{21} & \Delta_{22} & \Delta_{23} & \Delta_{24} \\ \Delta_{31} & \Delta_{32} & \Delta_{33} & \Delta_{34} \\ \Delta_{41} & \Delta_{42} & \Delta_{43} & \Delta_{44} \end{pmatrix}\boldsymbol{\psi} \tag{ST1.10.17}$$

where $\boldsymbol{\Delta}$ is a new 4 × 4 matrix, called the differential propagation matrix, that depends on $\boldsymbol{M}$ and $\xi$. The relations between the elements of the 4 × 4 matrix $\boldsymbol{\Delta}$ and the elements of the 6 × 6 matrix $\boldsymbol{M}$ are:

$$\begin{aligned}
\Delta_{11} &= M_{51} + (M_{53} + \xi)a_{31} + M_{56}a_{61} \\
\Delta_{12} &= M_{55} + (M_{53} + \xi)a_{35} + M_{56}a_{65} \\
\Delta_{13} &= M_{52} + (M_{53} + \xi)a_{32} + M_{56}a_{62} \\
-\Delta_{14} &= M_{54} + (M_{53} + \xi)a_{34} + M_{56}a_{64} \\
\Delta_{21} &= M_{11} + M_{13}a_{31} + M_{16}a_{61} \\
\Delta_{22} &= M_{15} + M_{13}a_{35} + M_{16}a_{65} \\
\Delta_{23} &= M_{12} + M_{13}a_{32} + M_{16}a_{62} \\
-\Delta_{24} &= M_{14} + M_{13}a_{34} + M_{16}a_{64} \\
-\Delta_{31} &= M_{41} + M_{43}a_{31} + M_{46}a_{61} \\
-\Delta_{32} &= M_{45} + M_{43}a_{35} + M_{46}a_{65} \\
-\Delta_{33} &= M_{42} + M_{43}a_{32} + M_{46}a_{62} \\
\Delta_{34} &= M_{44} + M_{43}a_{34} + M_{46}a_{64} \\
\Delta_{41} &= M_{21} + M_{23}a_{31} + (M_{26} - \xi)a_{61} \\
\Delta_{42} &= M_{25} + M_{23}a_{35} + (M_{26} - \xi)a_{65} \\
\Delta_{43} &= M_{22} + M_{23}a_{32} + (M_{26} - \xi)a_{62} \\
-\Delta_{44} &= M_{24} + M_{23}a_{34} + (M_{26} - \xi)a_{64}
\end{aligned} \tag{ST1.10.18}$$

with the auxiliary definitions:

$$\begin{aligned}
a_{31} &\equiv (M_{61}M_{36} - M_{31}M_{66})/d \\
a_{32} &\equiv [(M_{62} - \xi)M_{36} - M_{32}M_{66}]/d \\
a_{34} &\equiv (M_{64}M_{36} - M_{34}M_{66})/d \\
a_{35} &\equiv [M_{65}M_{36} - (M_{36} + \xi)M_{66}]/d \\
a_{61} &\equiv (M_{63}M_{31} - M_{33}M_{61})/d \\
a_{62} &\equiv [M_{63}M_{32} - M_{33}(M_{62} - \xi)]/d \\
a_{64} &\equiv (M_{63}M_{34} - M_{33}M_{64})/d \\
a_{65} &\equiv [M_{63}(M_{62} + \xi) - M_{33}M_{65}]/d
\end{aligned} \tag{ST1.10.19}$$

and further:

$$d \equiv M_{33}M_{66} - M_{36}M_{63} \tag{ST1.10.20}$$

**Special Topic ST1.10 (Continued)**

$$\xi \equiv (\omega/c) n_{\mathrm{m}} \sin \theta_{\mathrm{i}} \tag{ST1.10.21}$$

where $n_{\mathrm{m}}$ is the real refractive index of the isotropic ambient medium (air). Formally, if one has the experimental values of the dielectric tensor $\boldsymbol{\varepsilon}$, the magnetic permeability tensor $\boldsymbol{\mu}$, and the optical rotation tensors $\boldsymbol{\rho}, \boldsymbol{\rho}'$ for the substrate, one can construct first the optical matrix $\boldsymbol{M}$, then the differential propagation matrix $\boldsymbol{\Delta}$ and $\xi$ which, to repeat, is the $x$-component of the wave vector of the incident wave. Once $\boldsymbol{\Delta}$ is known, the law of propagation (wave equation) for the generalized field vector $\boldsymbol{\psi}$ (the components of $\boldsymbol{E}$ and $\boldsymbol{H}$ parallel to the $x$- and $y$-axes) is specified by Eq. (ST1.10.18). Experimentally, one travels this path backward.

Consider the relationship between $\boldsymbol{\Delta}$ and the dielectric tensor $\boldsymbol{\varepsilon}$. In ellipsometry, there is reflection and transmission by the surface ($z = 0$) of a semi-infinite anisotropic substrate (biaxial crystal) into an isotropic ambient (air, for $z < 0$). Suppose that this semi-infinite anisotropic medium (the crystal) is homogeneous, and that its optical matrix $\boldsymbol{M}$ is independent of $z$ (if $\boldsymbol{\Delta}$ does depend on $z$, i.e. on how far into the crystal one goes, then the problem becomes much more difficult). If the optical matrix $\boldsymbol{M}$ of the substrate is independent of $z$, then so is the differential propagation matrix $\boldsymbol{\Delta}$; if $\boldsymbol{\Delta}$ is independent of $z$, and has a value $\xi$, to be found below, the solution of Eq. (ST1.10.17) is given by

$$\boldsymbol{\psi}(z) = \boldsymbol{\psi}(0) \exp(-\mathrm{i}\omega\zeta z/c) \tag{ST1.10.22}$$

Here $\boldsymbol{\psi}(0)$ is the value of the generalized field vector of the incident plane wave at $z = 0$ (the crystal surface). Substituting Eq. (ST1.10.22) into the original differential equation, Eq. (ST1.10.17), one gets:

$$-\mathrm{i}(\omega\zeta/c)\boldsymbol{\psi}(z) = -\mathrm{i}(\omega/c)\boldsymbol{\Delta}\boldsymbol{\psi}(z) \tag{ST1.10.23}$$

which has unique solutions if and only if the following determinant vanishes:

$$\det |\, \boldsymbol{\Delta} - \zeta \boldsymbol{I} \,| = 0 \tag{ST1.10.24}$$

where $\boldsymbol{I}$ is the $4 \times 4$ identity matrix. This yields four eigenvalues $\zeta_1$, $\zeta_2$, $\zeta_3$, and $\zeta_4$ and four plane-wave solutions $\{\boldsymbol{\psi}(z) = \boldsymbol{\psi}(0) \exp(-\mathrm{i}\omega\zeta_m z/c),\ m = 1, \ldots, 4\}$ for $E_x$, $H_y$, $E_y$, and $-H_x$. The four eigenvalues correspond to four values of the $z$-component of the propagation vectors of the electromagnetic waves. Two of these are associated with waves propagating in the $+z$ direction (into the biaxial crystal), and two waves will propagate in the negative $z$-direction (out of the crystal into the air).

Consider the two reflected plane-waves propagating in the $+z$ direction, generated by the incident waves (which traveled in the $-z$ direction); let $\zeta_1$ and $\zeta_2$ be the corresponding two eigenvalues of Eq. (ST1.10.24) with a positive real part; let $c_1\boldsymbol{\psi}_1(0)$ and $c_2\boldsymbol{\psi}_2(0)$ be the associated eigenvectors, which are known up to the constant amplitude factors $c_1$ and $c_2$. These coefficients $c_1$ and $c_2$ will be determined by matching tangential electric and magnetic field components, in the ambient (air) and in the substrate (crystal), at their common interface $z = 0$. In terms of the generalized field vectors, the boundary conditions assume the form:

$$\boldsymbol{\psi}_{\mathrm{i}}(0^-) + \boldsymbol{\psi}_{\mathrm{r}}(0^-) = c_1\boldsymbol{\psi}_1(0^+) + c_2\boldsymbol{\psi}_2(0^+) \tag{ST1.10.25}$$

where the subscripts $i$ and $r$ indicate the incident and reflected wave components of the total field $\boldsymbol{\psi}$ in the ambient (air): $0^+$ and $0^-$ represent the ambient and substrate sides of the $z = 0$ interface. Equation (ST1.10.25) shows, on the left-hand side, the generalized field of the incident and reflected waves on the air side of the interface, while the right-hand side shows the waves generated inside the biaxial crystal. (Equation (ST1.10.25) is a conservation of energy condition, but since $\boldsymbol{\psi}_{\mathrm{r}}(0^-)$ has propagation vector opposite to $\boldsymbol{\psi}_{\mathrm{i}}(0^-)$, there is a positive sign in front of $\boldsymbol{\psi}_{\mathrm{r}}(0^-)$.)

We now need a relationship between $\boldsymbol{H}$ and $\boldsymbol{E}$, involving the dielectric constant $n$. Consider the plane-wave case, i.e. the case in which $\boldsymbol{E}$ depends only on one of the three coordinates, say $x$; that is, $E_x$, $E_y$, and $E_z$ can depend only on $x$, but not on $y$ or $z$. For this plane wave Eqs. (1.9.30) and (1.9.28) can be shown to yield $E_x = 0$ and $H_x = 0$. The wave equation, Eq. (1.10.10) now reduces to:

$$\partial^2 E_y/\partial x^2 = \varepsilon\mu c^{-2}\partial^2 E_y/\partial t^2 \tag{ST1.10.26}$$

A similar equation is written for $E_z$. The general solution to Eq. (ST1.10.26) is

$$E_y = A \exp\left[\mathrm{i}\omega\left(t - \varepsilon^{1/2}\mu^{1/2} x\, c^{-1}\right)\right] \tag{ST1.10.27}$$

**Special Topic ST1.10 (Continued)**

This solution is inserted into Eqs. (ST1.10.28) and (ST1.10.26):

$$-(\mu/c)(\partial H_z/\partial t) = \partial E_y/\partial x = -\mathrm{i}\omega\varepsilon^{1/2}\mu^{1/2}c^{-1}A\exp\left[\mathrm{i}\omega\left(t-\varepsilon^{1/2}\mu^{1/2}x/c\right)\right] \tag{ST1.10.28}$$

After integrating with respect to $t$,

$$H_z = \varepsilon^{1/2}\mu^{-1/2}\mathrm{i}_y \quad (\mathrm{cgs}) \tag{ST1.10.29}$$

For nonmagnetic media, $\mu = 1$, so the Maxwell relation, Eq. (ST1.10.7) yields finally:

$$H_z = nE_y \quad (\mathrm{cgs}) \tag{ST1.10.30}$$

Using this in Equation (ST1.10.25), we want to know the generalized field vectors $\boldsymbol{\psi}_\mathrm{i}$, $\boldsymbol{\psi}_\mathrm{r}$ of the incident and reflected waves. We can write down $\boldsymbol{\psi}_\mathrm{i}$ and $\boldsymbol{\psi}_\mathrm{r}$ in terms of components parallel (||) and perpendicular (s) to the plane of incidence. In a nonmagnetic ($\mu = 1$) optically isotropic medium, Eq. (ST1.10.30) shows that the magnetic field components are simply related to their associated orthogonal electric field components through the index of refraction $n$:

$$H_\parallel/E_\perp = H_\perp/E_\parallel = n \tag{ST1.10.31}$$

Consider the plane-polarized incident beam, with amplitude $E_{\parallel\mathrm{i}}$ in the plane of incidence (with $y$ and $z$ components, see Figure 1.14) and $E_{\mathrm{si}}$ in the direction normal to the plane of incidence (the $x$-axis in Figure 1.14), traversing an isotropic medium (air) of refractive index $n_\mathrm{m}$, and making an angle of incidence $\theta_\mathrm{i}$ with the surface. Since the beam is polarized in the plane of incidence, $E_y = 0$; with Eq. (ST1.10.31) this yields also $H_x = 0$. From the definition of the generalized field vector, Eqs. (ST1.10.16) and (ST1.10.31) we can get the $s$ and $p$ components of $\psi_\mathrm{i}$:

$$\psi_{\parallel\mathrm{i}} = \psi_{\parallel\mathrm{i}}\begin{pmatrix}0\\0\\1\\n_\mathrm{m}\cos\ \vartheta_\mathrm{i}\end{pmatrix};\quad \psi_{\perp\mathrm{i}} = E_{\perp\mathrm{i}}\begin{pmatrix}0\\0\\1\\n_\mathrm{m}\cos\ \vartheta_\mathrm{i}\end{pmatrix} \tag{ST1.10.32}$$

and

$$\psi_\mathrm{i} \equiv \psi_{\parallel\mathrm{i}} + \psi_{\perp\mathrm{i}} \tag{ST1.10.33}$$

Also:

$$\psi_{\parallel\mathrm{r}} = E_{\parallel\mathrm{r}}\begin{pmatrix}-\cos\ \vartheta_\mathrm{i}\\n_\mathrm{m}\\0\\0\end{pmatrix};\quad \psi_{\perp\mathrm{r}} = E_{\perp\mathrm{r}}\begin{pmatrix}0\\0\\1\\-n_\mathrm{m}\cos\ \vartheta_\mathrm{i}\end{pmatrix} \tag{ST1.10.34}$$

and

$$\boldsymbol{\psi}_\mathrm{r} \equiv \boldsymbol{\psi}_{\parallel\mathrm{r}} + \boldsymbol{\psi}_{\perp\mathrm{r}} \tag{ST1.10.35}$$

where i and r denote the incident and reflected beams, respectively, assuming $\theta_\mathrm{i} = \theta_\mathrm{r}$. After substituting into Eq. (ST1.10.26), the $\boldsymbol{\psi}_\mathrm{i}$ and $\boldsymbol{\psi}_\mathrm{r}$ given by Eqs. (ST1.10.32)–(ST1.10.35), and denoting by $n_\mathrm{i}$ the index of refraction of the ambient (=air), one gets:

$$\begin{aligned}
&(E_{\parallel\mathrm{i}} - E_{\parallel\mathrm{r}})\cos\ \theta_\mathrm{i} = c_1\psi_{11} + c_2\psi_{12}\\
&(E_{\parallel\mathrm{i}} + E_{\parallel\mathrm{r}})n_\mathrm{m} = c_1\psi_{21} + c_2\psi_{22}\\
&(E_{\perp\mathrm{i}} + E_{\perp\mathrm{r}}) = c_1\psi_{31} + c_2\psi_{32}\\
&(E_{\perp\mathrm{i}} - E_{\perp\mathrm{r}})n_\mathrm{m}\cos\ \theta_\mathrm{i} = c_1\psi_{41} + c_2\psi_{42}
\end{aligned} \tag{ST1.10.36}$$

**Special Topic ST1.10 (Continued)**

where $\psi_{ij}$ is the $j$th component ($i = 1, 2$; $j = 1, \ldots, 4$) of the column eigenvectors $\boldsymbol{\psi}_1(0^+)$ and $\boldsymbol{\psi}_2(0^+)$. Define the 1 × 2 column vector $\boldsymbol{c}$, whose two elements determine the relative amplitudes ($c_1$ and $c_2$) of the two refracted (transmitted) waves propagating into the crystal. Define the total electric field vectors

$$\boldsymbol{E}_\mathrm{i} \equiv \boldsymbol{E}_{\mathrm{i}\parallel} + \boldsymbol{E}_{\mathrm{i}\perp} \tag{ST1.10.37}$$

$$\boldsymbol{E}_\mathrm{r} \equiv \boldsymbol{E}_{\mathrm{r}\parallel} + \boldsymbol{E}_{\mathrm{r}\perp} \tag{ST1.10.38}$$

define the 2 × 2 complex-amplitude reflection matrix $\boldsymbol{R}$ by

$$\boldsymbol{E}_\mathrm{r} \equiv \boldsymbol{R}\boldsymbol{E}_\mathrm{i} \tag{ST2.10.39}$$

or, in detail:

$$\begin{pmatrix} E_{\parallel\mathrm{r}} \\ E_{\perp\mathrm{r}} \end{pmatrix} = \begin{pmatrix} R_{\parallel\parallel} & R_{\parallel\perp} \\ R_{\perp\parallel} & R_{\perp\perp} \end{pmatrix} \begin{pmatrix} E_{\parallel\mathrm{i}} \\ E_{\perp\mathrm{i}} \end{pmatrix} = \begin{pmatrix} R_{\parallel\parallel}E_{\parallel\mathrm{i}} + R_{\parallel\perp}E_{\perp\mathrm{i}} \\ R_{\perp\parallel}E_{\parallel\mathrm{i}} + R_{\perp\perp}E_{\perp\mathrm{i}} \end{pmatrix} \tag{ST1.10.40}$$

Then the four equations in Eq. (ST1.10.40) become:

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{S}_1(\boldsymbol{E}_\mathrm{i} - \boldsymbol{E}_\mathrm{r}) \\ \boldsymbol{c} &= \boldsymbol{S}_2(\boldsymbol{E}_\mathrm{i} - \boldsymbol{E}_\mathrm{r}) \end{aligned} \tag{ST1.10.41}$$

where we define further:

$$\boldsymbol{S}_1 \equiv \cos\ \theta_\mathrm{i} \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{41}/n_\mathrm{m} & \psi_{42}/n_\mathrm{m} \end{pmatrix}^{-1} \text{ and } \boldsymbol{S}_2 \equiv \begin{pmatrix} \psi_{21}/n_\mathrm{m} & \psi_{22}/n_\mathrm{m} \\ \psi_{31} & \psi_{32} \end{pmatrix}^{-1} \tag{ST1.10.42}$$

Elimination of $\boldsymbol{c}$ finally leads to an expression of the reflection matrix $\boldsymbol{R}$ in terms of the generalized field eigenvalues and the index of refraction of the medium (air):

$$\boldsymbol{R} = (\boldsymbol{S}_1 + \boldsymbol{S}_2)^{-1}(\boldsymbol{S}_2 - \boldsymbol{S}_1) \tag{ST1.10.43}$$

So far, however, one still needs an expression for the reflection matrix that shows how to extract from it the tensor elements for the refractive index tensor of the biaxial medium. We seek the reflection matrix $\boldsymbol{R}$ for the semi-infinite anisotropic biaxial medium. Using Eqs. (ST1.10.8) and (ST1.10.21), we can relate the 4 × 4 differential propagation matrix $\boldsymbol{\Delta}$ to the dielectric tensor $\boldsymbol{\varepsilon}$ from Eqs. (ST1.10.21) and (ST1.10.24). Then it can be shown that:

$$\boldsymbol{\Delta} = \begin{pmatrix} 0 & 1 - \xi^2/\varepsilon_3 & 0 & 0 \\ \varepsilon_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \varepsilon_2 - \xi^2 & 0 \end{pmatrix} \tag{ST1.10.44}$$

The eigenvalues of $\boldsymbol{\Delta}$ are given by

$$\zeta_{1,3} = \pm \left\{\varepsilon_1\left(1 - \xi^2/\varepsilon_3\right)\right\}^{1/2} \tag{ST1.10.45}$$

$$\zeta_{2,4} = \pm \left\{\varepsilon_2 - \xi^2\right\}^{1/2} \tag{ST1.10.46}$$

The eigenvectors are then:

$$\psi_1 = \begin{pmatrix} \sqrt{1 - \xi^2/\varepsilon_3} \\ \sqrt{\varepsilon_1} \\ 0 \\ 0 \end{pmatrix}, \psi_3 = \begin{pmatrix} \sqrt{1 - \xi^2/\varepsilon_3} \\ -\sqrt{\varepsilon_1} \\ 0 \\ 0 \end{pmatrix} \tag{ST1.10.47}$$

**Special Topic ST1.10 (Continued)**

$$\psi_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \sqrt{\varepsilon_2 - \xi^2} \end{pmatrix}, \psi_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -\sqrt{\varepsilon_2 - \xi^2} \end{pmatrix} \tag{ST1.10.48}$$

where positive eigenvalues are associated with $\psi_1$ and $\psi_2$. We need next to obtain the reflection matrix $\boldsymbol{R}$ for this case. It has been shown that the reflection matrix reduces to:

$$\begin{aligned} R_{|||} &= R_{11} = \left[\varepsilon_1^{1/2}\cos\ \theta_\mathrm{i} - n_\mathrm{m}\left(1-\xi^2\varepsilon_3^{-1}\right)^{1/2}\right] / \left[\varepsilon_1^{1/2}\cos\ \theta_\mathrm{i} + n_\mathrm{m}\left(1-\xi^2\varepsilon_3^{-1}\right)^{1/2}\right] \\ R_{\|\perp} &= R_{12} = 0 \\ R_{\perp\perp} &= R_{22} = \left[n_\mathrm{m}\cos\ \theta_\mathrm{i} - \left(\varepsilon_2 - \xi^2\right)^{1/2}\right] / \left[n_\mathrm{m}\cos\ \theta_\mathrm{i} + \left(\varepsilon_2 - \xi^2\right)^{1/2}\right] \\ R_{\perp\|} &= R_{21} = 0 \end{aligned} \tag{ST1.10.49}$$

In the monoclinic case, in which the dielectric tensor is not in the principal axis system:

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & 0 & \varepsilon_{xz} \\ 0 & \varepsilon_{yy} & 0 \\ \varepsilon_{xz} & 0 & \varepsilon_{zz} \end{pmatrix} \tag{ST1.10.50}$$

the reflection matrix values are

$$\begin{aligned} R_{|||} = R_{11} = &\left[\left(\varepsilon_{xx} - \varepsilon_{xz}^2\varepsilon_{zz}^{-1}\right)^{1/2}\cos\ \theta_\mathrm{i} - n_\mathrm{m}\left(1-\xi^2\varepsilon_{zz}^{-1}\right)^{1/2}\right] \\ &\left[\left(\varepsilon_{xx} - \varepsilon_{xz}^2\varepsilon_{zz}^{-1}\right)^{1/2}\cos\ \theta_\mathrm{i} + n_\mathrm{m}\left(1-\xi^2\varepsilon_{zz}^{-1}\right)^{1/2}\right]^{-1} \\ R_{\|\perp} = R_{12} = 0 \\ R_{\perp\perp} = R_{22} = &\left[n_\mathrm{m}\cos\ \theta_\mathrm{i} - \left(\varepsilon_{yy} - \xi^2\right)^{1/2}\right]\left[n_\mathrm{m}\cos\ \theta_\mathrm{i} + \left(\varepsilon_{yy} - \xi^2\right)^{1/2}\right]^{-1} \\ R_{\perp\|} = R_{21} = 0 \end{aligned} \tag{ST1.10.51}$$

We next calculate the null setting of an ellipsometer from the reflection matrix in an anisotropic sample. The Jones vector for the reflected light is given by Eq. (ST1.10.44); for an anisotropic sample the off-diagonal elements of reflection matrix $\boldsymbol{R}$ are nonzero. We finally seek the fundamental equation of ellipsometry. In the isotropic case (where $n''$ and $n$ are both scalar), the traceless 2 × 2 reflection matrix $\boldsymbol{R}$ is given by

$$\begin{pmatrix} E'_\| \\ E'_\perp \end{pmatrix} = \begin{pmatrix} R_\| & 0 \\ 0 & R_\perp \end{pmatrix}\begin{pmatrix} E_\| \\ E_\infty \end{pmatrix} \tag{ST1.10.52}$$

The complex **reflectance ratio** $\rho$ is defined by

$$\rho \equiv \left(R_\| / R_\perp\right) = \left[E'_\| / E_\|\right] / \left[E'_\perp / E_\perp\right] \tag{ST1.10.53}$$

using $\boldsymbol{E}(t) = \boldsymbol{E}\exp[\mathrm{i}(\omega t + \phi)] = \boldsymbol{E}\exp[\mathrm{i}\omega t]\exp[\mathrm{i}\phi]$ for all polarizations we get:

$$\rho \equiv \left(R_\| / R_\perp\right) = \left[E'_\| E_\perp / E'_\perp E_\|\right]\exp\left\{\mathrm{i}\left[\left(\phi'_\| - \phi'_\perp\right) - \left(\phi_\| - \phi_\perp\right)\right]\right\} \tag{ST1.10.54}$$

this can be simplified by defining four new angles, $\psi$, $\psi'$, $\delta$, and $\delta'$, where:

$$\tan\psi \equiv \left[E_\| / E_\perp\right],\ \tan\psi' \equiv \left[E'_\| / E'_\perp\right], \delta \equiv \phi_\| - \phi_\perp,\ \text{ and }\ \delta' \equiv \phi'_\| - \phi'_\perp \tag{ST1.10.55}$$

then

$$\rho = \left[\tan\psi' / \tan\psi\right]\exp\left[\mathrm{i}(\delta' - \delta)\right] \tag{ST1.10.56}$$

**Special Topic ST1.10 (Continued)**

and with two more trivial definitions:

$$\tan\psi \equiv [\tan\psi'/\tan\psi] \text{ and } \Delta \equiv [\delta' - \delta] \tag{ST1.10.57}$$

we finally get the fundamental equation of ellipsometry:

$$\rho = \tan\psi \exp(\mathrm{i}\Delta) \tag{ST1.10.58}$$

In the ellipsometer, let $P$, $A$, and $Q$ be the angular settings of the **polarizer, analyzer,** and **compensato**r, respectively: these angles are measured counterclockwise for an observer looking into the light beam and are zero when the transmission axes of the polarizer and analyzer and the fast axis of the compensator are in the plane of incidence. If the compensator is a perfect quarter-wave plate and if the electric field intensities are normalized, then the Jones vector representing the polarization of the light before it strikes the surface of the sample is given by

$$\begin{aligned}(E_{\|\mathrm{i}}) &= (\cos\ Q\cos\ (P-Q) + \mathrm{i}\sin\ Q\sin\ (P-Q))\\ (E_{\perp\mathrm{i}}) &= (\sin\ Q\sin\ (P-Q) - \mathrm{i}\cos\ Q\sin\ (P-Q))\end{aligned} \tag{ST1.10.59}$$

If the reflected beam can be extinguished by the analyzer, then the phases of $E_{\|\mathrm{r}}$ and $E_{\perp\mathrm{r}}$ must be equal, and the imaginary part of the ratio $E_{\|\mathrm{r}}/E_{\perp\mathrm{r}}$ must be zero. Substituting this condition into Eq. (ST1.10.59) yields:

$$F\sin\ (2P-2Q) + G\cos\ (2P-2Q) + H = 0 \tag{ST1.10.60}$$

where $F$, $G$, and $H$ are defined by

$$F \equiv \mathrm{Re}\left(R_{11}R_{22}^{*} - R_{12}R_{21}^{*}\right) \tag{ST1.10.61}$$

$$G \equiv \left[\mathrm{Im}\left(R_{11}R_{22}^{*} + R_{12}R_{21}^{*}\right)\right]\sin 2Q + \left[\mathrm{Im}\left(R_{11}R_{21}^{*} - R_{12}R_{22}^{*}\right)\right]\cos 2Q \tag{ST1.10.62}$$

$$H \equiv -\mathrm{Im}\left(R_{11}R_{21}^{*} + R_{12}R_{22}^{*}\right) \tag{ST1.10.63}$$

The two solutions of Eq. (ST1.10.60) are

$$P_1 = (1/2)\sin^{-1}\left[H/\left(F^2+G^2\right)^{1/2}\right] - \tan^{-1}(G/F) + Q \tag{ST1.10.64}$$

$$P_2 = (1/2)\pi - \sin^{-1}\left[H/\left(F^2+G^2\right)^{1/2}\right] - \tan^{-1}(G/F) + Q \tag{ST1.10.65}$$

where $H^2 < F^2 + G^2$ must hold. Equation (ST1.10.60) gives the relation between the polarizer and compensator that will provide a null setting. Equations (ST1.10.64) and (ST1.10.65) give a solution for $P$ for a fixed compensator or a fixed polarizer. The plane-polarized reflected beam is extinguished when the setting of analyzer angle $A$ satisfies the following equation:

$$\tan A = -\left(R_{11}E_{\|\mathrm{i}} + R_{12}E_{\perp\mathrm{i}}\right)/\left(R_{21}E_{\|\mathrm{i}} + R_{22}E_{\perp\mathrm{i}}\right) \tag{ST1.10.66}$$

Two real values of $A$ are obtained, one for each of the two values of $P$ from Eq. (ST1.10.66). The two sets of null settings for $P$ and $A$ are related to the different zone settings, but sign conventions vary from instrument to instrument.

In brief, the steps needed to calculate the instrumental settings $P$, $A$ from given values of the permittivity $\boldsymbol{\varepsilon}$, permeability $\boldsymbol{\mu}$, and optical rotation $\boldsymbol{\rho}$, $\boldsymbol{\rho}'$ tensors are

1) Set up $\boldsymbol{\varepsilon}$, $\boldsymbol{\mu}$, $\boldsymbol{\rho}$, and $\boldsymbol{\rho}'$ and compute the elements of $\boldsymbol{M}$.
2) Compute the elements of the differential propagation matrix $\boldsymbol{\Delta}$.
3) Determine the eigenvalues $\zeta_{\mathrm{i}}$ and the eigenvectors $\psi_{\mathrm{i}}$ of $\boldsymbol{\Delta}$.
4) Find the two eigenvectors $\psi_{\mathrm{i}}$ associated with each of the eigenvalues $\zeta_{\mathrm{i}}$.
5) Compute the refractive index values from the relationship between $\psi_{\mathrm{i}}$, $\zeta_{\mathrm{i}}$, and $\boldsymbol{\Delta}$.
6) Compute the reflection matrix $\boldsymbol{R}$.
7) Compute the polarizer and analyzer angles $P$ and $A$ from the reflection matrix $\boldsymbol{R}$.

### Special Topic ST1.10 (Continued)

**The scheme above might be called the "forward" calculation.**

The practical method to compute "backward," i.e. to calculate the dielectric constant tensor values, or the complex index of refraction, from a set of the observed polarizer and analyzer angles is not presented here. Instead, for a biaxial crystal, the technique indicated below is

1) Use a single wavelength of light.
2) Collect $P, A$ values at 15° intervals (which will repeat after a 180° rotation) in four zones (thus, 12 × 4 unique data), and at two angular settings of the ellipsometer (two different angles of incidence): thus, there are 96 unique data.
3) Test initial values of the complex index of refraction ($n_x - ik_x, n_y - ik_y, n_z - ik_x$), and of three Eulerian rotation angles for the crystal face. Thus, there are 9 parameters to be determined, for 96 data points: the system is mathematically overdetermined.
4) Once a reasonable starting set of values for the 9 parameters is found, refine these parameters using the Simplex algorithm.

### Special Topic ST1.11 Standard Model, the Eightfold Way: Are Octonions Helpful?

Between 1950 and 2000 AD, many new elementary particles and short-lived resonances were found, creating hope that their discovery or new theoretical concepts would produce a final explanation for "subatomic physics." Their masses $m$, charges $Q$, and spin $S$ were measured. In reactions observed between particles and resonances, certain "conservation theorems" and ensuing new "quantum numbers" were invented: "Parity" ($P$) to describe handedness; "isotopic spin or isospin" $T$ (to show how close in mass the proton and neutron are) and from it, the "z-component of isospin" $T_z$, "strangeness" $S$, "charm" Ch, "beauty or bottomness" By, and finally "truth or topness" Tr were invented as new and presumed useful quantum numbers. Alas, in the steady decline of elegant-sounding Greek or Latin neologisms, common English words were assigned new significance: there is no color in "color," no flavor in "flavor," no charm in "charm" no location value in "up," "down," "top," "bottom," or moral value in "charm." "Gauge" should be really called "phase." One wonders which of these quantum numbers will "survive" in a future full theory.

The "Eightfold Way" of Murray Gell-Mann borrows a label from the "Eightfold Way to Buddhism" and organizes both the spin-1/2 baryons (Figure 1.39) into an octet, and the spin-0 mesons (Figure 1.40) into an octet; the Eightfold Way was also applied to the spin-3/2 baryons (Figure 1.41) that form a decuplet (Figure 1.42).

For instance, strangeness St is conserved in the reaction:

$$\pi^-(\mathrm{St}=0) + p^+(\mathrm{St}=0) \rightarrow K^0(\mathrm{St}=1) + \Lambda^0(\mathrm{St}=-1) \qquad (\mathrm{ST1.11.1})$$

The photon-induced introduction of a neutron excited state $\mathrm{n}^{0*}$, and resultant charge separation and interconversion of quarks is exhibited in:

$$\gamma + \mathrm{n}^0\left(ud\overline{d}\right) \rightarrow \mathrm{n}^{0*}\left(ud\overline{d}\ u\overline{u}\right) \rightarrow p^+(u\overline{u}d) + \pi^-\left(u\overline{d}\right) \qquad (\mathrm{ST1.11.2})$$

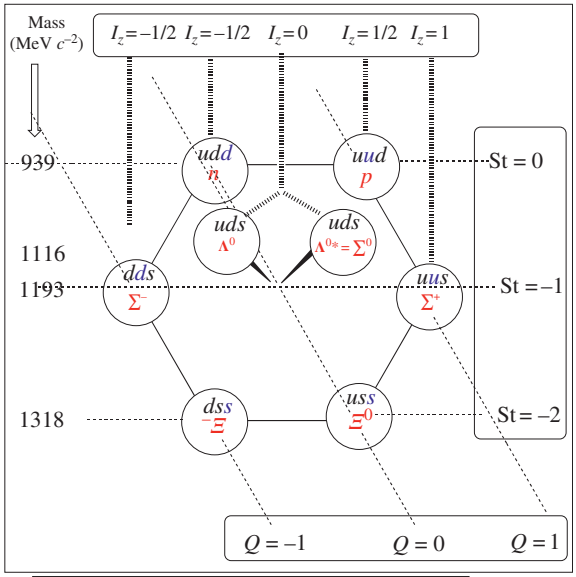
We repeat an earlier result:

$$Q = T_z + (\mathrm{Ba} + \mathrm{St} + \mathrm{Ch} + \mathrm{By} + \mathrm{Tr})/2 \qquad ((1.4.1))$$

The Gell–Mann–Okubo and Coleman–Glashow equations predict some approximate and relative rest mass ($m$) relationships:

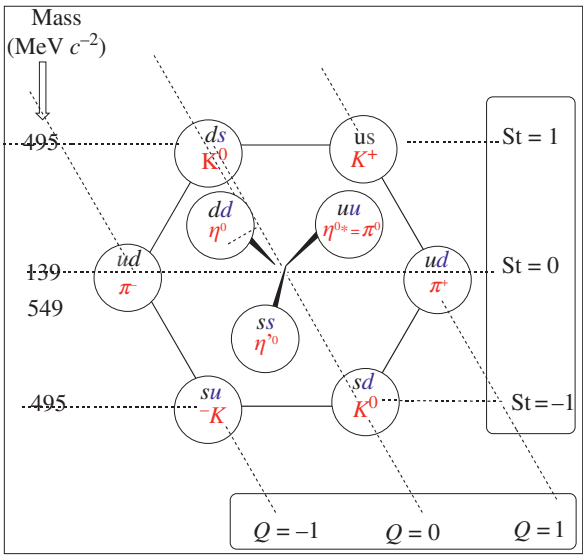
$$m(n) + m\left(\Xi^0\right) \text{ vs. } (3/2)m\left(\Lambda^0\right) + m\left(\Sigma^0\right) \qquad (\mathrm{ST1.11.3})$$

**Special Topic ST1.11 (Continued)**




(1) Here St = strangeness; $Q$ = charge;
(2) A repeated second quark ($d$, $s$, or $u$)is an anti-quark, and is shown in blue
(3) the $\Lambda^0$ or its excited state $\Sigma^0$ are "above" the hexagonal plane

**Figure 1.39** The hexagon arrangement in the "Eightfold Way" for eight spin-1/2 baryons (the "baryon octet").




(1) Here St = strangeness; $Q$ = charge;
(2) A repeated second quark ($d$, $s$, or $u$)is an anti-quark, and is shown in blue
(3) the $\pi^0$ or its excited state $\eta^0$ or $\eta'^0$ are "above" the hexagonal plane

**Figure 1.40** Hexagonal arrangement in the "Eightfold Way" for eight spin-0 mesons (the "meson octet"): $\eta'$ is an excited $\eta$.

**Special Topic ST1.11 (Continued)**

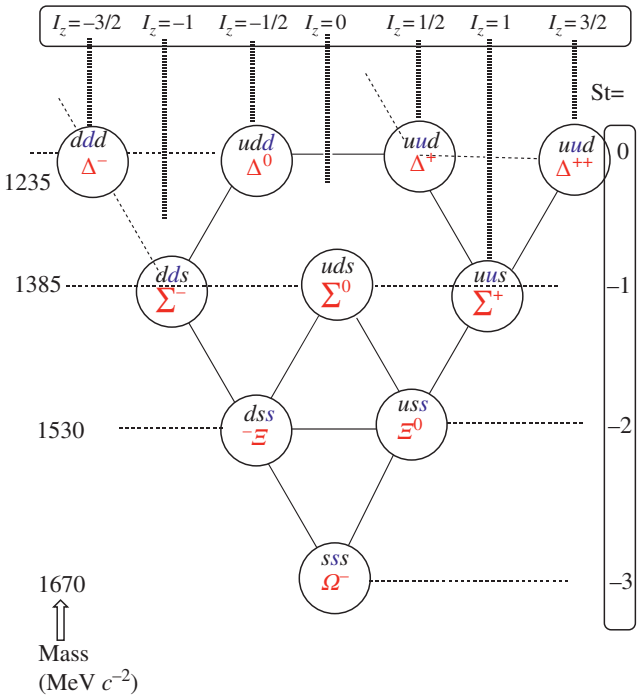



**Figure 1.41** The "baryon decuplet," including the Ω– particle (predicted in 1961, and found in 1964).

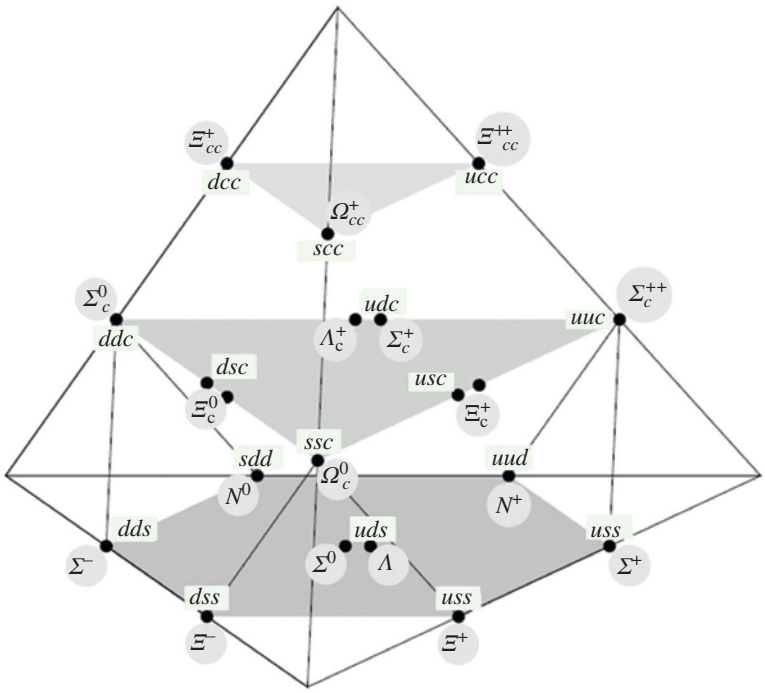



**Figure 1.42** A baryon supermultiplet using a basis of four quarks (*u*, *d*, *s*, *c*) and half-integer spin.

**Special Topic ST1.11 (Continued)**

where the left-hand mass sum is 1839 + 1192 = 3031 MeV, while the right-hand side mass sum is (3/2)1116 + 1315 = 2989 MeV (not too bad), and:

$$m(\Xi^-) - m(\Xi^0) \text{ vs. } m(\Sigma^-) - m(\Sigma^+) + m(p) - m(n) \tag{ST1.11.4}$$

where the left-hand mass sum is 1321 − 1315 = −6 MeV, while the right-hand side mass sum is 1197 − 1189 + 1838 − 1839 = 7 MeV: not bad at all.

The finite group SU(2) (here "S" is for special, "U" is for unitary) can be used for both spin *S* and for isospin *T*. For electron spin *S*, the three generators are multiples of the traceless Pauli spin matrices:

$$J_1 = (\hbar/2)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, J_2 = (\hbar/2)\begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, J_3 = (\hbar/2)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{ST1.11.5}$$

and the two spin eigenfunctions are

$$\alpha = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{ST1.11.6}$$

For isospin *T* the three generators are simply:

$$t_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, t_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, t_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{ST1.11.7}$$

and the isospin eigenfunctions are the proton and the neutron:

$$p = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, n = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{ST1.11.8}$$

For the group SU(3) with rank = 2 and dimension = 8 there are 8 traceless "generators" or Gell-Mann matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_2 = \begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{ST1.11.9}$$

$$\lambda_5 = \begin{pmatrix} 0 & 0 & -\mathrm{i} \\ 0 & 0 & 0 \\ \mathrm{i} & 0 & 0 \end{pmatrix}, \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix}, \lambda_8 = 3^{-1/2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \tag{ST1.11.10}$$

where $\lambda_1$–$\lambda_3$ are the augmented three-dimensional versions of the well-known two-dimensional Pauli spinors (first 2 rows and 2 columns). These 8 generators ($\lambda_1$–$\lambda_8$) were loosely associated by Gell-Mann with the 8 particles mentioned in Figures 1.39 and 1.40.

For instance, one can write for the three quark colors *R*, *G*, and *B* the column vectors:

$$R = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, G = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{ST1.11.11}$$

which are simultaneous eigenfunctions of $\lambda_3$ and $\lambda_8$. This can be seen in a diagram (Figure 1.43).

For future growth, the Standard Model needs some new mathematics that can describe the short-range potentials that bind elementary particles and nuclei together. In the meantime, one can associate the strong force with the Lie group $SU(3)_C$, the weak force by the Lie group $SU(2)_L$, and electromagnetism by the Lie group $U(1)_Y$. Is a fusion of these possible? Can octonions be of help?

**Special Topic ST1.11 (Continued)**

An up-to-date picture (Figure 1.44) shows how the Higgs boson, when it finds its new ground state and "breaks the symmetry" (lower part of Figure 1.44), it somehow confers to the particles individual rest masses as needed. Parity and handedness (see quantum number *P*) must also be included in the theory.

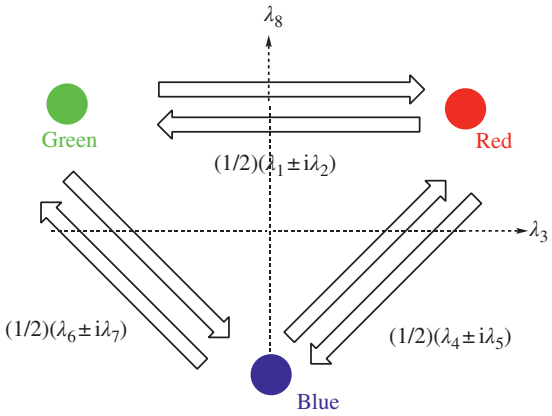



**Figure 1.43** Interconversion of the colors (R, B, G) using the generators of SU(3).

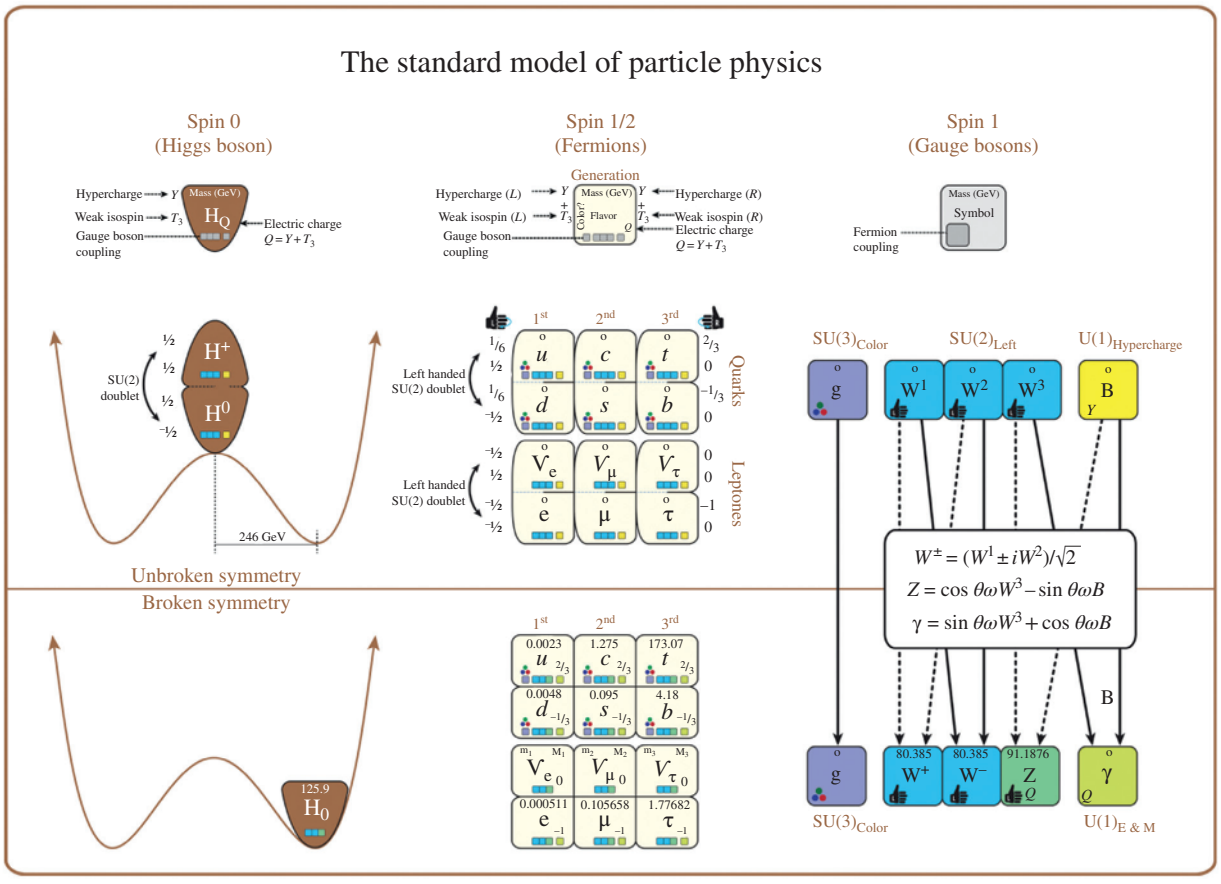



**Figure 1.44** The Present Understanding (drawn by Latham Boyle). *Source:* Drawn by Latham Boyle/Wikimedia Commons, CC BY-SA 4.0. https://commons.wikimedia.org/wiki/File:Standard_Model_Of_Particle_Physics--Most_Complete_Diagram.png.

## References


**[1.1]** Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica*. London: Pepys.

**[1.2]** B. P. Abbott, et al. (LIGO Scientific Collaboration + VIRGO Collaboration) "Observation of gravitational waves from a binary black hole merger", *Phys. Rev. Lett.* **116**: #061102 (2016).

**[1.3]** Serway, R.A. (1996). *Physics for Scientists and Engineers*, 4e. Philadelphia, PA: Saunders.

**[1.4]** de Coulomb, C.A. (1785). [Trois] Mémoires sur l'Électricité et le Magnétisme. *Hist. Acad. Roy. Sci.* 569–577, 578–611, and 612–638.

**[1.5]** Maxwell, J.C. (1865). A dynamical theory of the electromagnetic field. *Phil. Trans. Roy. Soc. London* **155**: 459–512.

**[1.6]** Fermi, E. (1934). Attempt at a theory of β-rays. *Nuovo Cimento* **11**: 1.

**[1.7]** Mohr, P.J., Taylor, B.N., and Newell, D.B. (2012). *The 2010 CODATA Recommended Values of the Fundamental Physical Constants, Rev. Mod. Phys*. **84**: 1527–1605.

**[1.8]** Einstein, A. (1905). Zur Elektrodynamik bewegter Körper. *Ann. Physik* **17**: 891–921.

**[1.9]** Michelson, A.A. and Morley, E. (1887). On the relative motion of the Earth and the luminiferous ether. *Am. J. Sci.* **34**: 333–345.

**[1.10]** Einstein, A. (1915). Zur allgemeinen Relativitätstheorie. *Sitzungsber. K. Preuss. Akad. Wissenschaften zu Berlin* (4 November 1915), pp. 778–786 and "Nachtrag", pp. 799–801.

**[1.11]** Wheeler, J.A. (1963). *Geometrodynamics*. New York, NY: Academic Press.

**[1.12]** Libby, W.F., Anderson, E.C., and Arnold, J.R. (1949). Age determination by radiocarbon content: World-wide assay of natural radiocarbon. *Science* **109**: 227–228.

**[1.13]** Nuclear Binding energy curve. Wikipedia (2016). en.wikipedia.org/wiki/File:Binding_energy_curve__common.isotopes.svg.

**[1.14]** (a) G. Friedlander, M. Perlman, J R. Stehn, & E. F. Clancy, *Chart of the Nuclides* (Knolls Atomic Power Laboratory, operated by the General Electric Co., 5, Apr 1956) (operated by Bechtel Marine Propulsion, 17e, 2010, www.nuclidechart.com);
(b) National Nuclear Data Center, Brookhaven National Laboratory, Upton, NY 11973-5000, USA;
(c) Korea Atomic Energy Research Institute (2000). Daejeon, South Korea;
(d) Universal Nuclear Chart (www.nucleonica.net).

**[1.15]** (a) D. Mendeleev, "Experience on the System of the Elements" [in Russian] *Zhurnal Russkogo Khimicheskogo Obshchestva* **1**(2–3): 35 (1869);
(b) D. Mendeleev, "Experience on the System of the Elements" [in German], *Z. Prakt. Chem.* **106**: 251 (1869);
(c) D. Mendeleev, "On the Relationship of the Properties of the Elements to their Atomic Weights [in Russian], *Zhurnal Russkogo Khimicheskogo Obshchestva* **1**: 60–77 (1869). [Abstracted in *Z. Chem.* **12**: 405–406 (1869)];
(d) D. Mendeleev, "Natural System of the Elements and its Application to Prediction of Properties of Yet Undiscovered Elements" [in Russian], *Zhurnal Russkogo Khimicheskogo Obshchestva* 3: 25–56 (1871);
(e) D. Mendelejeff, "Zur Frage über das System der Elemente", *Ber. Deutsch. Chem. Gesell.* **4**: 348–352 (1871).

**[1.16]** Wolfram, S. (2003). *The Mathematica Book*, 5e. Champaign, IL, USA: Wolfram Research.

**[1.17]** Wolfram, S. (ed.) (2015). *Mathematica*, vol. **11**. Champaign, IL, USA: Wolfram Research.

**[1.18]** Abel, N.H. (1824). *Mémoire sur les équations algébriques, où on démontre l'impossibilité de la résolution de l'équation générale du cinquième degré*. Norway: Christiania.

**[1.19]** Abramowitz, M. and Stegun, I.A. (ed.) (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Applied Mathematics Series, vol. **55**. Washington, DC: U. S. Government Printing Office.

**[1.20]** Zhang, S. and Jin, J. (1996). *Computation of Special Functions*. New York, NY: Wiley-Interscience.

**[1.21]** Jahnke, E. and Emde, F. (1945). *Tables of Functions with Formulae and Curves*, 4e. New York, NY: Dover.

**[1.22]** Jolley, L.B.W. (1961). *Summation of Series*, 2e. New York, NY: Dover.

**[1.23]** Dwight, H.B. (1961). *Tables of Integrals and Other Mathematical Data*, 4e. New York, NY: Macmillan.

**[1.24]** Gradshteyn, I.S. and Ryzhik, I.M. (1965). *Table of Integrals, Series, and Products*, 4e. New York, NY: Academic Press.

**[1.25]** Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. (1954). *Tables of Integral Transforms*, vol. 1 and 2. New York, NY: McGraw-Hill.

**[1.26]** Davidson, N. (1962). *Statistical Mechanics*. New York, NY: McGraw-Hill.

**[1.27]** Hooke, R. (1678). *Ut Tensio Sic Vis*. In: *A Description of Helioscopes and Some Other Instruments*. London: J. Martyn.

**[1.28]** Joos, G. (1950). *Theoretical Physics*, 3e. New York, NY: Hafner.

**[1.29]** Guenther, R.D. (1990). *Modern Optics*. New York, NY: Wiley.

**[1.30]** Jackson, J.D. (1999). *Classical Electrodynamics*, 3e. New York, NY: Wiley.
**[1.31]** Leighton, R.B. (1959). *Principles of Modern Physics*. New York, NY: McGraw-Hill.
**[1.32]** The Event Horizon Collaboration (2019). First M87 Event telescope results. V. Physical origin of the asymmetric ring. *Astrophys. J. Lett.* **875**: L5.
**[1.33]** Churchill, R.V. (1963). *Fourier Series and Boundary-Value Problems*, 2e. New York, NY: McGraw-Hill.
**[1.34]** Steinfeld, J.I., Francisco, J.S., and Hase, W.L. (1998). *Chemical Kinetics and Dynamics*, 2e. Upper Saddle River, NJ: Prentice-Hall.
**[1.35]** Bard, A.J. and Faulkner, L.R. (2001). *Electrochemical Methods – Fundamentals and Applications*, 2e. New York, NY: J. Wiley.
**[1.36]** Courant, R. (1959). *Differential and Integral Calculus*, vol. **II**, 17–18. New York: Interscience.
**[1.37]** Perrin, C.L. (1970). *Mathematics for Chemists*. New York, NY: Wiley-Interscience.
**[1.38]** Mathews, J. and Walker, R.L. (1970). *Mathematical Methods of Physics*, 2e. Menlo Park, CA: W. A. Benjamin.
**[1.39]** Margenau, H. and Murphy, G.M. (1961). *The Mathematics of Physics and Chemistry*, 2e. Princeton, NJ: D. Van Nostrand.
**[1.40]** Margenau, H. and Murphy, G.M. (1964). *The Mathematics of Physics and Chemistry*, vol. **2**. New York, NY: D. Van Nostrand-Reinhold.
**[1.41]** Churchill, R.V. (1960). *Complex Variables and Applications*. New York, NY: McGraw-Hill.
**[1.42]** Knaust H. (1999) www.sosmath.com.algebra/fac12/fac12.html (accessed 25 February 2018): copyright 1999-2018 MathMedics, LLC. P.O.Box 12395, El Paso, TX 77913, USA.
**[1.43]** Nielsen, K.L. and Vanlonkuyyzen, J.H. (1946). *Plane and Spherical Trigonometry*. New York, NY: Barnes and Noble.
**[1.44]** Cooper, P.W. (1966). Through the Earth in forty minutes. *Am. J. Phys.* **34**: 68–70.
**[1.45]** Cooper, P.W. (1966). Further commentary. *Am. J. Phys.* **34**: 703.

## End-of-Chapter Problems

**Problem E1-1** **The golden ratio** $\varphi$ (=golden mean = golden section = divine section = sectio aurea = sectio divina): two quantities $a$ and $b$ are in the **golden ratio** $\varphi$ if their ratio is the same as the ratio of their sum to the larger of the two quantities (assuming $a > b > 0$):

$$(a+b)/a = a/b \equiv \varphi \equiv \left(1+\sqrt{5}\right)/2 = 1.618\,033\,988\,7$$

$\varphi$, known since Egyptian times, was studied by Euclid. Prove that $\varphi \equiv \left(1+\sqrt{5}\right)/2 = 1.618\,033\,988\,7$.

**Problem E1-2** **Integer numbers: need for zero.** The concept of "zero" was developed independently in Babylon, India, and the Mayan empire. The word "zero" is a Venetian contraction of "zefiro" from the pre-Islamic word sfir ("emptiness") which rendered the concept of emptiness or zero (sunya in Sanskrit). Show how zero is needed for subtraction. Show that the set of all positive and negative integers and zero is "closed" under the operations of additions and subtraction.

**Problem E1-3** **Rational numbers.** Show that postulating the operation of division alongside the operation of multiplication forces into existence the rational numbers, which can be represented by $p/q$, where both $p$ and $q$ are integers. The set of integer plus rational numbers is "closed" under the four operations of addition, subtraction, multiplication, and division.

**Problem E1-4** **Irrational numbers.** Starting with a set of rational numbers (i.e. integers and fractions) ordered, let us, say, between 1 and 2, a Dedekind[176] cut divides this set into a Lower set L and an upper set U: all numbers in L are smaller than all numbers in U; neither L nor U is empty. Then show that an irrational number, e.g. $\sqrt{2} = 2^{1/2} = 1.414\ \ 213\ \ 562.$ lies between these two sets, i.e. cannot be represented by any rational fraction $p/q$ in either set. Prove this by squaring.

176 Julius Wilhelm Richard Dedekind (1831–1916).

**Problem E1-5 Complex numbers.** When dealing with square roots, the system of complex numbers is born. Indeed, the square root of negative numbers is not real. $\sqrt{(-1)}$ causes problems if a real number representation is attempted: Squaring $\sqrt{(-1)}$ would mean $\left[\sqrt{(-1)}\right]^2 \equiv \sqrt{(-1)}\sqrt{(-1)} = -1$, but this is not possible. Show that inventing $\mathrm{i} \equiv \sqrt{(-1)}$, where i is "pure imaginary" resolves the difficulty. Show how to represent complex numbers in a 2D Cartesian coordinate system: $z \equiv x + \mathrm{i}y$ (where $x$ and $y$ are real) and in a plane polar coordinate system.

**Problem E1-6 Quaternions.** In 1843, Hamilton extended the notion of complex numbers $z = x + \mathrm{i}y$ defined in a 2D space to quaternions in a 4D space, with numbers

$$a + bi + cj + ck$$

where $a$, $b$, $c$, and $d$ are real numbers, while **i**, **j**, **k** are the fundamental quaternion units, with definitions:

$$i^2 = j^2 = k^2 = \mathbf{ijk} = -1$$

(The multiplication of quaternions is not commutative.)

**Sideline 1.19** Hamilton incised the above in the stone of Brougham Bridge over the River Liffey in Dublin (but his graffiti are no longer visible today). Maxwell's equations for electromagnetism were first written in quaternions, but by the 1880s, quaternions were displaced by the vector analysis introduced by Gibbs. Quaternions have had a recent revival.

Show by right multiplication that

$$\mathrm{ij} = \mathrm{k}, \mathrm{jk} = \mathrm{i}, \mathrm{ki} = \mathrm{j}, \mathrm{ji} = -\mathrm{k}, \mathrm{kj} = -\mathrm{i}, \mathrm{ik} = -\mathrm{j}$$

and then construct the group multiplication table.

**Problem E1-7 Straight line.** Equations for a straight line in the $xy$ plane: The two-point form of a straight line going through points $(x_1, y_1)$ and $(x_2, y_2)$ is given by

$$(y - y_1) = [(y_2 - y_1)/(x_2 - x_1)](x - x_1) \qquad (\mathrm{E1} - 7.1)$$

**(a)** Its slope-intercept form is

$$y = mx + b \qquad (\mathrm{E1} - 7.2)$$

express $m$ and $b$ in terms of $(x_1, y_1)$ and $(x_2, y_2)$.

**(b)** Its intercepts form is given by

$$x/p + y/q = 1 \qquad (\mathrm{E1} - 7.3)$$

Express $p$ and $q$ in terms of $(x_1, y_1)$ and $(x_2, y_2)$.

**Problem E1-8 Surveying.** A person standing 70 m from the base of a tree, uses a sextant to measure the angle between the ground and the top of the tree as 20°. Estimate the height h of the tree to the nearest tenth of a meter.

**Problem E1-9 Trigonometric equation**. Solve the equation $\sec^2 x - 2\tan x = 4$ for $x$.

**Problem E1-10 Trigonometric equation**. Simplify $(\sin^4 x - \cos^4 x)/(\sin^2 x + \cos^2 x)$.

**Problem E1-11 Trigonometric equation**. Simplify: $\sec x \cos^2 x/(1 + \sec x)$.

**Problem E1-12 Trigonometric equation**. Solve for $x$ in the equation $2\sin^2 x - \sin x - 1 = 0$ in the interval $[0, 2\pi]$.

**Problem E1-13** **Trigonometric equation**. Solve $\sec^2 x - 2\tan x = 4$.

**Problem E1-14** **Trigonometric equation**. Solve $\cos^3 x - \tan^3 3x = 2$ (cannot be done completely: it is too messy).

**Problem E1-15** **Trigonometric equation**. Solve $\text{cotan}^2 x - \cos x = 0$.

**Problem E1-16** **Trigonometric equation**. Verify $2\tan x/\tan 2x = 1 - \tan^2 x$.

**Problem E1-17** **Trigonometric equation**. Evaluate: $5\sin 2x = 3$.

**Problem E1-18** **Pythagoras' theorem**. Prove Pythagoras' theorem: in a right triangle, the sum of the squares of the two sides $a$ and $b$ equals the square of the hypotenuse $c$: $a^2 + b^2 = -c^2$. This is shown in Figure 1.45.

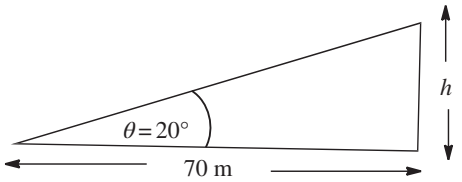



**Figure 1.45** Use of sextant.

**Sideline 1.20 Fermat's last theorem**. The equation $a^n + b^n = c^n$ can be solved for $a$, $b$, and $c$ integers when $n = 1$ (trivial) or $n = 2$ (Pythagoras). For instance, we can see that $3^2 + 4^2 = 5^2$ or that $6^2 + 8^2 = 10^2$ or that $12^2 + 16^2 = 20^2$. However, $a^n + b^n = c^n$ cannot be solved for $a$, $b$, and $c$ integers when $n > 2$: this was "Fermat's last theorem" (1637), which was finally proved in 1995 (358 years later!).

**Problem E1-19** **Derivative.** Evaluate $(d^3/dx^3)\cos^{-1} x$.

**Problem E1-20** **Derivative.** Evaluate $(d^2/dx^2)\tan x$.

**Problem E1-21** **Derivative.** Given $\sinh x \equiv (1/2)[\exp(x) - \exp(-x)]$, evaluate $(d^2/dx^2)\sinh x$.

**Problem E1-22** **Derivative.** In the equation $x + a(xy)^{1/2} - y^2 = b$, where $a$ and $b$ are constants, find $(dy/dx)$ in terms of $x$ and $y$, without solving for $y$.

**Problem E1-23** **Derivative.** Evaluate $(d/dz)\ln_e(z^2 - 4)$.

**Problem E1-24** **Derivative.** Evaluate $(d/dz)\ln_e[(1 - \cos x)/(1 + \cos x)]^{1/3}$.

**Problem E1-25** **Partial derivative.** For an equation $f(x, y, z) \equiv x^2 + 2yz - 2y^2 - z^2 = 11$, find $(\partial f/\partial z)_x$ the partial derivative of $f$ with respect to $z$, with $x$ held constant.

**ProblemE1-26** **Partial derivative.** For $f \equiv y^3 \sin x + \exp(y)\cos x$, find $(\partial f/\partial x)$ and $(\partial f/\partial y)$.

**Problem E1-27** **Partial derivative.** For $f \equiv 4r^2 \sin 2\theta$, find $(\partial f/\partial r)$ at point $(r = -2, \theta = \pi/4)$.

**Problem E1-28** **Partial derivative.** For $f = [x^2 - y^2]^{1/2}$, find all second partial derivatives with respect to $x$ and $y$.

**Problem E1-29** **Partial derivative.** Prove that $\nabla \cdot (\boldsymbol{E} \times \boldsymbol{F}) = \boldsymbol{F} \cdot \nabla \times \boldsymbol{E} - \boldsymbol{E} \cdot \nabla \times \boldsymbol{F}$.

**Problem E1-30** **Indefinite integral: complete the square.** Evaluate $\int dx[x^2 + 6x + 13]^{-1}$.

**Problem E1-31** **Indefinite integral: partial fractions.** Evaluate $\int dx\,(5x + 16)/(x - 4)(2x + 1)$ by partial fractions.

**Problem E1-32** **Indefinite integral.** Evaluate $\int \sec^2 x \tan x (4 - \tan^2 x)^{-3/2}\,dx$.

**Problem E1-33** **Definite integral.** Evaluate $\int_{x=1}^{x=3} dx\,(4 - x)^{1/2}(2 + x)^{-1/2}$.

**Problem E1-34** **Elliptic integral.** While extensive tables of complicated integrals exist, not all expressions can be integrated in "closed form," i.e. in terms of known algebraic, trigonometric, logarithmic, and transcendental functions. An interesting class is the elliptic integrals of the type $\int dx[(1 - x^2)(1 - k^2x^2)]^{1/2}$ for which a series of elliptic functions was developed. What can be done instead?

**Problem E1-35** **Definite integral.** Evaluate $\int_{\theta=0}^{\theta=\pi/4} \sin^3 \theta \cos \theta\, d\theta$.

**Problem E1-36** **Definite integral.** Evaluate $\int_{\theta=0}^{\theta=2\pi} (1 + \cos \theta)^{1/2}\,d\theta$.

**Problem E1-37** **Double integral: volume.** Compute the volume $V$ of a solid under the plane $x + 3y + 4z = 12$ and above a triangle in the $xy$-plane whose sides are defined by the equations $x = 2$, $y = 0$, and $y = x$. The volume is calculated using the general formula $V = \int\int_A f(x, y)\,dA$, with the function $f(x, y) = z = (12 - x - 3y)/4$. The area $A$ is the area above the specified triangle for $z = 0$ but under the specified plane for $z > 0$.

**Problem E1-38** **Double integral: volume.** Find the volume $V$ of the region bounded by the circular cylinder $x^2 + y^2 = 4$, the $xy$-plane and the plane $x - y + 2z = 4$. (Hint: To integrate a function involving $(4 - x^2)$, change variables to $x = 2 \sin \theta$.)

**Problem E1-39** **Triple integral: volume.** Compute the volume of a sphere of radius $R$.

**Problem E1-40** **Triple integral: volume.** Compute the volume of a right circular cylinder of radius $R$ and height $H$.

**Problem E1-41** **Triple integral: volume.** Volume of a doughnut (in mathematics lingo, it is a torus). Call $r$ the inner radius of the doughnut and $R$ as the distance from the center of the doughnut hole to the center of the ring. Compute the volume of the doughnut.

**Problem E1-42** **Differential equation.** Find the general solution of the first-order differential form: $(3x^2 - y)dx + (3y^2 - x)dy = 0$.

**Problem E1-43** **Differential equation.** Solve $dy/dx = xy/(x - 1)$.

**Problem E1-44 Differential equation.** Solve $x\,\mathrm{d}^2y/\mathrm{d}x^2 = \mathrm{d}y/\mathrm{d}x$.

**Problem E1-45 Differential equation.** For the differential equation: $(\mathrm{d}^2y/\mathrm{d}x^2) + p(\mathrm{d}y/\mathrm{d}x) + \theta = 0$, where $p$ and $q$ are constants, consider an auxiliary equation: $m^2 + pm + q = 0$.

(a) If this auxiliary equation has real and distinct roots, call them $m_1$ and $m_2$, then verify that the general solution to the differential equation is $y = A\exp(m_1x) + B\exp(m_2x)$.

(b) If the auxiliary equation has imaginary roots $\alpha \pm \mathrm{i}\beta$, where $\mathrm{i} \equiv (-1)^{1/2}$, then verify that the general solution to the differential equation is $y = \exp(\alpha x)[A\sin(\beta x) + B\cos(\beta x)$.

(c) If the auxiliary equation has two equal roots $m$, then verify that the general solution to the differential equation is $y = (A + Bx)\exp(mx)$. Throughout, $A$ and $B$ are constants of integration set by the initial conditions.

**Problem E1-46 PDE: the heat equation.** Heat transfer in one dimension along a rod of length $l$ can be analyzed in terms of the change in temperature $\theta = \theta(t,x)$ as a function of both time $t$ and distance $x$. The operative PDE is

$$-\beta^2\left(\partial^2\theta/\partial x^2\right) + \left(\partial\theta/\partial t\right) = 0$$

where $\beta^2$ is a constant known as the thermal diffusivity, with SI units of $\mathrm{m}^2\ \mathrm{s}^{-1}$. It is related by the equation $\beta^2 = \kappa/\rho C_\mathrm{P}$ to the thermal conductivity $\boldsymbol{\kappa}$, the sample density $\rho$, and the constant-pressure specific heat capacity $C_\mathrm{P}$. Solve the PDE by assuming that the solution $\theta(t,x)$ is a product of two functions of only one variable each: $\theta(x,t) \equiv X(x)T(t)$. For a complete discussion, we should set up both initial conditions for $\theta(t = 0, x)$ and boundaries for the $x$-dependence.

**Problem E1-47 PDE.** Solve the one-dimensional wave equation: $-(\partial^2W/\partial x^2) - q(\partial^2W/\partial t^2) = 0$ by separation of variables: $W(x,t) = X(x)T(t)$.

**Problem E1-48 Circumference of Earth.** Eratosthenes,[177] author of the now lost *Geographika* and acknowledged founder of geography, worked in Ptolemaic Egypt. He measured **correctly** the circumference of the Earth by noting that, at the winter solstice (21 December) at high noon in the town of Syene, the Sun would be vertically overhead, so that the shadow of a man standing over a deep pit would vanish completely within the pit, while in Alexandria the Sun's rays were not vertical. Agricultural surveys had shown that the distance from Syene to Alexandria was 5000 Attic stadia ($5000 \times 185\ \mathrm{m} = 925$ km). Eratosthenes used a gnomon (a stick planted vertically in the ground) at high noon in Alexandria and found that the angle of the Sun from the local zenith was not zero, but one-fiftieth of a circle away from the zenith, i.e. at $7^\circ12'$. Eratosthenes made three assumptions: (i) the Sun's rays are roughly parallel everywhere as they strike the Earth, (ii) Alexandria was directly North of Syene, and (iii) the Earth is a sphere, and then calculated that the circumference of the Earth was 46 250 km (in modern units). The correct modern value is 40 075 km. Repeat Eratosthenes's calculation. Toscanelli[178] had it wrong: this almost doomed Columbus'[179] first voyage.

**Problem E1-49 The catenary: part 1.** The equation for the catenary (Figure 1.46) was derived by Leibniz, Johann Bernoulli,[180] and Huygens; it is a hyperbolic cosine curve seen in (i) loose chains suspended at both ends, (ii) electrical wires between poles, (iii) the Gateway Arch at St. Louis, MO (inverted catenary), (iv) some arches in large buildings (inverted catenaries). The equation for the catenary curve can be derived as follows (Figure 1.47).

Consider a section of cable $PA$ between points $P$ and $A$. At point $A$, there is a **horizontal tension** $H$ pulling the cable sideways to the left, due to the intrinsic materials strength of the cable. At point $P$, there is a **tangential tension** $T$ pulling the cable along its length. The section $PA$ is also pulled down by **gravity** by a downward force $\lambda gs$, where $s =$ length of cable between $A$ and $P$, $g =$ acceleration due to gravity ($9.80\ \mathrm{m\ s}^{-2}$), and $\lambda(\mathrm{kg\ m}^{-1}) =$ the uniform linear mass density of the cable between points $P$ and $A$. At equilibrium, the forces must balance in the $x$- and $y$-directions, respectively:

177 Eratosthenes (276–194 BC).
178 Paolo dal Pozzo Toscanelli (1397–1482).
179 Christopher Columbus= Cristoforo Colombo = Cristobál Colón (1451–1506).
180 Johann Bernoulli (1667–1748).

**Figure 1.46** Pythagoras' theorem proved geometrically.

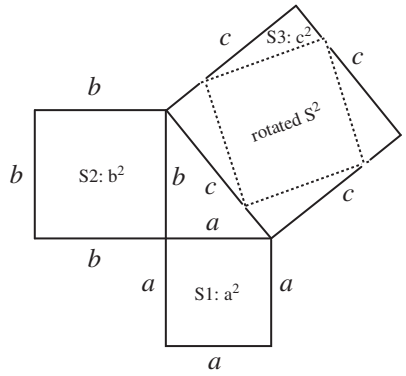
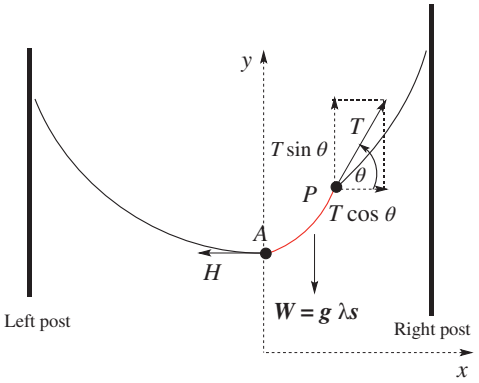



**Figure 1.47** Catenary problem.




$$T\cos\theta = H \quad \text{and} \quad T\sin\theta = W = \lambda g s \tag{E1 – 49.1}$$

Show that the ratio of these two equations yields the Whewell[181] equation:

$$\tan\ \theta = (\lambda g s/H) \tag{E1 – 49.2}$$

By definition at $P$, we see $\tan\ \theta \equiv (\mathrm{d}y/\mathrm{d}x)$, whence:

$$(\mathrm{d}y/\mathrm{d}x) = (\lambda g s/H). \tag{E1 – 49.3}$$

The SI units are $[\lambda] = \mathrm{kg\ m^{-1}}$; $[s] = \mathrm{m}$; $[g] = \mathrm{m\ s^{-2}}$; $[H] = N = \mathrm{kg\ m\ s^{-2}}$.

**Problem E1-50** **The catenary:** part 2. The independent variable s in Eq. (E1-49.3) is inconvenient. Since, to first order, $\mathrm{d}s^2 \approx \mathrm{d}x^2 + \mathrm{d}y^2$, therefore, differentiating both sides of Eq. (E1-49.3) with respect to x yields a second-order ODE in y:

$$\begin{aligned}
\left(\mathrm{d}^2y/\mathrm{d}x^2\right) &= (\lambda g/H)(\mathrm{d}s/\mathrm{d}x) = (\lambda g/H)\left[(\mathrm{d}s/\mathrm{d}x)^2\right]^{1/2} \\
&= (\lambda g/H)\left[\mathrm{d}s^2/\mathrm{d}x^2\right]^{1/2} = (\lambda g/H)\left[\left(\mathrm{d}x^2 + \mathrm{d}y^2\right)/\mathrm{d}x^2\right]^{1/2} \\
&= (\lambda g/H)\left[1 + \mathrm{d}y^2/\mathrm{d}x^2\right]^{1/2} = (\lambda g/H)\left[1 + (\mathrm{d}y/\mathrm{d}x)^2\right]^{1/2} \\
\left(\mathrm{d}^2y/\mathrm{d}x^2\right) &= (\lambda g/H)\left[1 + (\mathrm{d}y/\mathrm{d}x)^2\right]^{1/2}
\end{aligned} \tag{E1 – 50.1}$$

181 William Whewell (1794–1868).

**Problem E1-51** **The catenary: part 3**. Equation (E1-50.1) can be simplified by introducing $z \equiv (dy/dx)$, whence $(dz/dx) = (d^2y/dx^2)$, so that the equation to be solved becomes:

$$(dz/dx) = (\lambda g/H)\left[1 + z^2\right]^{1/2} \qquad (E1-51.1)$$

Solve this first-order differential equation in $z$ by separating the variables and prove:

$$y = (H/\lambda g)\cosh(\lambda g x/H) + \text{constant} \qquad (E1-51.2)$$

**Problem E1-52** **The catenary: part 4**. Note that (i) Eq. (E1-51.2) depends on $H$ but not on $T$; (ii) cosh(x) is a symmetric function of $x$; (iii) the only minimum of cosh($x$) is at $x = 0$, where cosh(0) = 1. To set the constant in Eq. (E1-51.2), use the boundary condition ($x = 0$, $y = 1$), so that the equation for the catenary finally becomes:

$$y = [(H/\lambda g)[\cosh(\lambda g x/H) - 1] + 1 \qquad (E1-52.1)$$

If instead the minimum of the catenary is at $[x = 0, y = y_0]$, then we have:

$$y = [(H/\lambda g)[\cosh(\lambda g x/H) - 1] + y_0 \qquad (E1-52.2)$$

If the maxima occur at $\pm x_1$, then $y_1 = [(H/\lambda g)[\cosh(\lambda g x_1/H) - 1] + y_0$ and the sag becomes:

$$\Delta y \equiv y_1 - y_0 = [(H/\lambda g)[\cosh(\lambda g x_1/H) - 1] \qquad (E1-52.3)$$

**Problem E1-53** **The catenary: part 5**. Let a chain span a distance $L$ between posts of equal height. Prove that the sag $d$ at the midpoint is

$$d = \lambda g L^2/8H \qquad (E1-53.1)$$

**Problem E1-54** **The catenary: part 6**. Prove that at any point x the radius of curvature $R \equiv [1 + (dy/dx)^2]^{1/2}|d^2y/dx^2|^{-1}$ of the catenary $y = a\cosh(x/a)$ is given by $R = y^2/a$.

**Problem E1-55** **The catenary: part 7**. Compute the linear mass density (kg m$^{-1}$) for a copper cable of diameter 10 cm, length $L = 100$ m, and volumetric density 8.93 g cm$^{-3}$.

**Problem E1-56** **The catenary: part 8**. An Al wire of radius 5 mm has cross-sectional area $= \pi 52\,\text{mm}^2 = 78.5\,\text{mm}^2 = 7.85 \times 10^{-5}$ m$^2$. The volumetric density of Al is 2.70 g cm$^{-3}$ = 2700 kg m$^{-3}$ and for an Al cable of radius 5 mm and cross-sectional area $7.85 \times 10^{-5}$ m$^2$ the linear density of this Al wire is $\lambda = 2700 \times 7.85 \times 10^{-5} = 0.212$ kg m$^{-1}$. The specific tensile strength of Al is 100 N mm$^{-2}$, whence $H = 100 \times 78.5 = 7850$ N. With these data, using Eq. (E1-53.1), compute the sag d in that wire is if its suspended between posts 100 m apart.

**Problem E1-57** **The catenary: part 9**. Power cable for a train line. Use a cylindrical Cu cable, radius 0.5 cm, volumetric density = 8.93 g cm$^{-3}$ = $8.93 \times 10^3$ kg m$^{-3}$, whence $\lambda$ = linear mass density = $(8.93 \times 10^3)[\pi \times (5 \times 10^{-3})^2] = 0.701$ kg m$^{-1}$, with $L = 100$ m between poles and $g = 9.80$ m s$^{-2}$ and a sag $d = 1$ m, what would $H$ have to be?

**Problem E1-58** **The catenary: part 10**. For Cu, the yield strength is 70 MPa; the ultimate tensile strength is 220 MPa (Wikipedia); for Cu wire the tensile strength is $H \approx 200$ to 250 N mm$^{-2}$ if annealed; for Al the tensile strength is $H \approx 100$ N mm$^{-2}$ (Wikipedia). Are we doing OK?

**Problem E1-59** **The catenary: part 11**. How much does the 0.5 cm radius Cu cable discussed in Problem E1-57 sag half-way between the power poles (i.e. at the point $x = 50$ m, which is half-way between the two poles spaced $L = 100$ m apart)?

**Problem E1-60 The catenary: Part 12**. It can also be shown that that the catenary $y = a\cosh(x/a)$ is the solution of the variational problem of finding the curve $y(x)$ which (i) minimizes the definite integral

$$\int_{x=a}^{x=b} y\left[1 + (\mathrm{d}y/\mathrm{d}x)^2\right]^{1/2} \mathrm{d}x$$

while (ii) keeping constant the definite integral

$$\int_{x=a}^{x=b} \left[1 + (\mathrm{d}y/\mathrm{d}x)^2\right]^{1/2} \mathrm{d}x$$

(which functions as a constraint equation).

**Problem E1-61** The equation for the cable of a suspension bridge (such as the Golden Gate Bridge in San Francisco, CA, finished in 1937), where the cable supports a very heavy roadway, is **not** a catenary, but a **parabola**: indeed $T_0 A\ \mathrm{d}y = \mu g x\ \mathrm{d}x$ integrates immediately to the parabola: $y = (\mu g/2AT_0)x^2 + \text{constant}$. Since there are two cables, the equation becomes $y = (\mu g/4AT_0)x^2$. The span between the two suspension towers of the Golden Gate Bridge is 1280 m, while the distance between the land abutments in San Francisco and Marin County is 2788 m. The sag of the roadway at the midpoint is only 3.30 m (but that is not relevant in the present problem). The yield strength of the two main galvanized steel cables is $T_0 = 182\,000\ \text{psi} = 1.254 \times 10^9\ \text{N m}^{-2}$. The roadway mass from land to land for the present steel roadway and no cars is $1.54 \times 10^8$ kg; therefore, the fractional mass between the two suspension towers is $1.54 \times 10^8 \times (1280/2788) = 7.07 \times 10^7$ kg. To this we should add the weight of cars filling up the roadway: 2 ton cars of length 120 in. each = 3 m filling up the roadway: 6 lanes $\times (1280/3)(2000) = 5.12 \times 10^6$ kg, whence the total mass between the two suspension towers must be $7.07 \times 10^7 + 5.12 \times 10^6 = 7.5 \times 10^7$ kg, and the mass per unit length is $\mu = 7.5 \times 10^7\ \text{kg}/1280 = 5.86 \times 10^4\ \text{kg m}^{-1}$. The galvanized steel cable diameter is 0.92 m, whence the cross-sectional area of the cable is $A = \pi(0.92/2)^2 = 0.665\ \text{m}^2$. The height of the suspension towers above the mean high-water mark is 227 m; the height of the roadway above the mean high-water mark is 67 m. Check how well the parabola equation coefficients fit these dimensions.

**Problem E1-62 Gravity train or "brachistochrone" problem** (first opined by Hooke, then inspired by Voyage to the Center of the Earth (1864) by Verne[182]; the mathematics were calculated by Cooper[183] [1.43, 1.44]. Suppose a straight hole is dug between any two points on the surface of the Earth, e.g. Washington and Los Angeles, or Paris and Beijing. Suppose a frictionless train is dropped at the initial point, accelerated by gravity, this idealized train will arrive at the final point in exactly 42 minutes and 12 seconds (if the travel is purely frictionless!) independent of the distance between the points on the surface.

Let $x(t)$ be the coordinate of the train along the tunnel, with $x = 0$ at the middle of the trajectory, and let $r$ be the distance from the train to the center of the Earth. Let the angle $\alpha$ between the vertical and the direction to the middle of the tunnel, then $x = \pm r\cos\ \alpha$.

The gravity force on the train of mass m is given by $F = -GmM(r)r^{-2}$, where $M(r)$ is the mass of the Earth beneath the train (the mass above it does not matter!) and $G$ is the universal gravitational constant, $G = 6.673\,84 \times 10^{-11}\ \text{m}^3\ \text{kg}^{-1}\ \text{s}^{-2}$. If the assumed uniform and isotropic density of the Earth is $\rho = 5.52\ \text{g cm}^{-3} = 5.52 \times 10^3\ \text{kg m}^{-3}$, then $M(r) = (4\pi/3)\rho r^3$, whence: $F = -(4\pi/3)Gm\rho r$. The component of the force along the tunnel is $F\cos\alpha$. By Newton's second law, the equation of motion becomes: $m(\mathrm{d}^2x/\mathrm{d}t^2) = F\cos\ \alpha = -(4\pi/3)Gm\rho r\cos\ \alpha = -(4\pi/3)Gm\rho x$, which becomes the equation of motion for a classical harmonic oscillator: $\mathrm{d}^2x/\mathrm{d}t^2 = -k^2x$, with solution: $x = A\sin(kt) + B\cos(kt)$, where: $k = [(4\pi/3)G\rho]^{1/2}$. The constants $A$ and $B$ are determined by the initial conditions. The half-period (the time for a single one-way trip) is

$$\pi/k = \pi\left[(4/3)\pi G\rho\right]^{-1/2} = 2\pi^{1/2}3^{1/2}G^{-1/2}\rho^{-1/2}$$

Compute this half-period.

182 Jules Verne (1828–1905).
183 Paul W. Cooper (b. c. 1935–).

**Problem E1-63** Archimedes'[184] principle (in memory of Waldman,[185] a former policeman and amateur scientist even at age 97!). A solid body of density $\rho_1$ is immersed in water whose density is $\rho_2$: if $\rho_1 \leq \rho_2$, then the body will float; if $\rho_1 > \rho_2$, then the body will sink. Archimedes took a bath, floated, and raised the water level (and/or displaced some of it on the floor) by an amount equal to his volume (this is Archimedes' principle: both water and Archimedes have roughly the same density $\rho = 1\ \text{g cm}^{-3} = 10^3\ \text{kg m}^{-3}$); Archimedes ran through the streets of Syracuse, Sicily yelling "eureka" (ευρηκα) ("I have found it"). A hollow ship of mass $M$ and density $D$ (for its wood or metal body) will have a volume $V$ which can be full of air, or partially full of cargo, and which provides buoyancy. The load line or Plimsoll[186] line is a large stripe or marking painted or scribed on the side of a ship (ever since ancient times) to prevent overloading the ship with cargo: if the Plimsoll line (a long horizontal line bisecting a circle, similar to the present-day symbol of the London Underground), is beneath the water level, the ship is in immediate danger of sinking!

(a) Assume an empty ship shaped as an ugly but simple parallelepiped of length 100 m, width (or beam) 10 m, and height 30 m (and therefore total volume $V = 3 \times 10^4\ \text{m}^3$), with a total metal mass (=displacement) of $M = 20\,000$ mt, i.e. $2 \times 10^7$ kg. By how many meters ($h$) will this empty ship be immersed in the water? The acceleration due to gravity is $g = 9.80\ \text{m s}^{-2}$, the density of water is $\rho_{\text{warer}} = 10^3\ \text{kg m}^{-3}$, and the density of air is $\rho_{\text{air}} = 1\ \text{kg m}^{-3}$. (b) Assume next that the boat will be loaded with $C = 5 \times 10^6$ kg of cargo. By how many meters ($h$) will the loaded boat sink into the water? (The vertical distance from sea level to the bottom of the ship is called the draft.)

**Problem E1-64** Determinants. Prove that if $G = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$ $H = \begin{vmatrix} a & e \\ b & f \end{vmatrix}$ then $G + xH = \begin{vmatrix} a & c + xe \\ b & d + xf \end{vmatrix}$.

**Problem E1-65** In octonions, the next step up from quaternions, there are eight unit vectors ($\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$, $\boldsymbol{e}_4$, $\boldsymbol{e}_5$, $\boldsymbol{e}_6$, $\boldsymbol{e}_7$); the group multiplication rules are as follows:

$$e_i e_j = e_j \text{ when } i = 0$$

$$e_i e_j = e_j \text{ when } j = 0$$

$$e_i e_j = -d_{ij} e_0 + e_{ijk} e_k \text{ otherwise}$$

where $\delta_{ij}$ is the Kronecker delta, and $e_{ijk} = +1$ when $ijk = 123$, 145, 176, 246, 257, 347, and 365. Construct the group multiplication table.

**Problem E1-66** Compute the height h above the Earth for a geostationary orbit on an artificial satellite that is equally attracted by the Earth (mass = $5.9736 \times 10^{24}$ kg) and the Moon (mass = $7.3467 \times 10^{22}$ kg) if the Earth–Moon mean distance is 384 400 km.

184 Archimedes (c. 287–212 BC).
185 Walter Sylvester Waldman (1916–2015).
186 Samuel Plimsoll (1824–1898).