

14 SOME APPLICATIONS OF LASERS

In the early years of their development, lasers were regarded by skeptics as “a solution looking for a problem.” More and more “problems” were found, and lasers have become an important part of the science and technology of our time, with applications ranging from medical to military. Lasers have been used in distance and velocity measurements, holography, printers, bar coding, CD players, surgery, and many other areas for many years now, and such “everyday” applications will not be touched upon here. We will instead consider, in addition to some aspects of the medical applications of lasers, just a few examples of the importance of lasers in research and emerging technology. The special role of diode and fiber lasers in telecommunications is the main subject of the next chapter.

14.1 LIDAR

Light detection and ranging (lidar) dates back to the 1930s, but because of lasers it has become one of the primary tools in atmospheric and environmental research. There are several types of lidar, all involving a transmitter of laser radiation and a receiver for the detection and analysis of backscattered light (Fig. 14.1). Before describing some lidar techniques and the information they provide, we will derive an equation for the number of photons counted at the receiver. Our analysis will apply directly to the most common type of lidar system, that in which the transmitter and receiver are located at essentially the same place; the laser beam in this case is typically sent through the receiver telescope. This is called the “monostatic,” as opposed to “bistatic,” configuration.

Consider first the case in which the backscattered light is resonance fluorescence, that is, spontaneous emission from atoms (or molecules) excited by laser radiation. Suppose that a laser beam propagates vertically from ground level to an altitude z and excites atoms within a column between altitudes z and $z + \Delta z$ ($\Delta z \ll z$). The radiated energy in time Δt from these atoms is

$$\Delta E_{\text{atoms}}(z, t) = h\nu A_{21} \Delta t \int_z^{z+\Delta z} dz' N(z') \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy p(x, y, z', t), \quad (14.1.1)$$

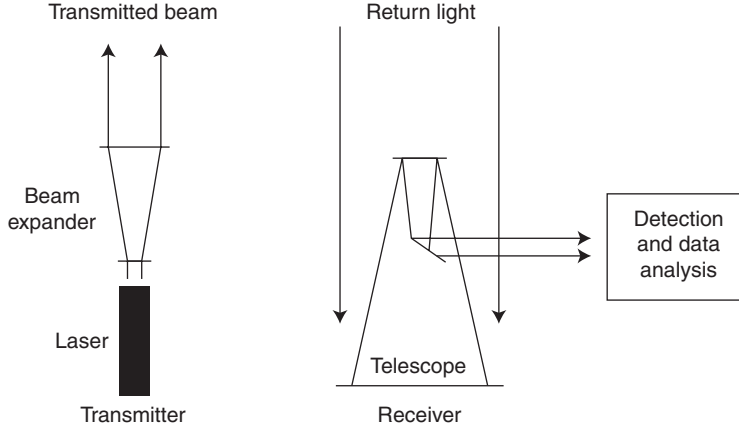


Figure 14.1 Basic elements of a standard lidar system. A beam expander is usually used at the transmitter in order to reduce the divergence of the laser beam before it propagates into the atmosphere. The receiver includes a wavelength filter, a photodetector, and computers and electronics for data acquisition and analysis.

where $N(z)$ is the number density of atoms at altitude z , ν is the transition frequency (assumed to be equal to the laser frequency), A_{21} is the spontaneous emission rate, and $p(x, y, z, t)$ is the excited-state probability at time t for an atom at a point x, y, z . If the excited atoms radiate isotropically, the power incident on a receiver of area A_r is simply

$$\begin{aligned} \text{Pwr}(z, t) &= T_0(z) \frac{A_r}{4\pi z^2} \Delta E_{\text{atoms}}(z, t) \\ &= T_0(z) \frac{A_r}{4\pi z^2} h\nu A_{21} \int_z^{z+\Delta z} dz' N(z') \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy p(x, y, z', t), \end{aligned} \quad (14.1.2)$$

where $T_0(z)$ is the atmospheric transmission coefficient at the transition frequency for propagation to ground from altitude z (or from ground to z), and is assumed not to vary significantly between z and $z + \Delta z$, and we continue to write Pwr for power, to distinguish it easily from P 's denoting pumping rate and pressure and probability, etc.

We will assume that the excited-state probability $p(x, y, z, t)$ changes predominantly by spontaneous emission; this is a reasonable assumption at high altitudes, where molecular densities are sufficiently small that collisional deexcitation is negligible. Suppose that the laser radiation is in the form of a pulse of duration τ_p long compared to the radiative lifetime $1/A_{21}$. Then, for times t much longer than the radiative lifetime (Problem 14.1),

$$p(x, y, z, t) \cong \frac{\sigma(z, \nu)}{A_{21} h\nu} I(x, y, z, t), \quad (14.1.3)$$

where $\sigma(z, \nu)$ is the absorption cross section at z at the laser frequency ν , and so

$$\text{Pwr}(z, t) = T_0(z) \frac{A_r}{4\pi z^2} \int_z^{z+\Delta z} dz' N(z') \sigma(z', \nu) \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy I(x, y, z', t) \quad (14.1.4)$$

in this approximation.

Now consider the number of photons counted in a time interval from $2z/c$ to $2z/c + \tau_d$ by a detector at the receiver. $2z/c$ is the time it takes for the laser radiation to propagate to altitude z , plus the time for photons radiated at z to propagate back to ground.¹ The number of photons counted from the “bin” $[z, z + \Delta z]$ in this time interval is

$$\begin{aligned}\mathcal{N}(z) &= \eta \times \frac{1}{h\nu} \times \int_{2z/c}^{2z/c + \tau_d} dt \text{Pwr}(z, t) \\ &= \eta T_0(z) \frac{A_r}{4\pi z^2} \frac{1}{h\nu} \int_z^{z+\Delta z} dz' N(z') \sigma(z', \nu) \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{2z/c}^{2z/c + \tau_d} dt I(x, y, z', t),\end{aligned}\quad (14.1.5)$$

where $\eta (\leq 1)$ is the photon counting efficiency at the receiver. If τ_d is much larger than the pulse duration, the integral of $I(x, y, z', t)$ over x, y , and t in (14.1.5) is just the pulse energy $E(z')$ at z' . Since $E(z') = T_0(z')E_p \cong T_0(z)E_p$, where E_p is the energy of the pulse leaving the transmitter, we write (14.1.5) as

$$\mathcal{N}(z) = \eta T_0^2(z) \frac{A_r}{4\pi z^2} \frac{E_p}{h\nu} \int_z^{z+\Delta z} dz' N(z') \sigma(z', \nu). \quad (14.1.6)$$

The column length Δz from which photons can reach the detector during the time interval $[2z/c, 2z/c + \tau_d]$ is determined by $2\Delta z/c = \tau_d$, or $\Delta z = c\tau_d/2$. We have assumed that this *range bin length* is made small enough that $N(z')$ and $\sigma(z', \nu)$ do not vary significantly within it. Then we obtain from (14.1.6) the *lidar equation*

$$\mathcal{N}(z) = \eta T_0^2(z) \frac{A_r}{4\pi z^2} \frac{E_p}{h\nu} N(z) \sigma(z, \nu) \frac{c\tau_d}{2}, \quad (14.1.7)$$

for the number of photon counts at the receiver from a range bin of length $\Delta z = c\tau_d/2$ at altitude z . There is clearly a trade-off between the range bin length and the strength of the photon counting signal: The smaller the range bin, the greater the “range resolution” but the smaller the number of photons backscattered from it.

Suppose some atmospheric constituent is probed as a function of altitude and the detection electronics is designed to “reset to zero” at the end of each counting interval, after which it begins counting photons from the next range bin. This *range gating* continues for some number of range bins, and photon counts per range bin are accumulated with multiple laser pulses fired at some repetition rate R_{rep} over some “integration time” τ . The number of photons accumulated from a range bin of length Δz over the integration time τ is obtained by multiplying (14.1.7) by $R_{\text{rep}}\tau$. Then the lidar equation for the number of photon counts from a range bin of length Δz at altitude z takes the form

$$\mathcal{N}(z) = \eta T_0^2(z) \frac{(\text{Pwr})_L \tau A_r}{h\nu z^2} N(z) \sigma^B(z, \nu) \Delta z, \quad (14.1.8)$$

where $(\text{Pwr})_L = E_p R_{\text{rep}}$ is the laser power. We have assumed that the scattering (or resonance fluorescence) is isotropic, i.e., that it is the same over all 4π steradians about the scatterer; in this case the cross section for scattering in the backward direction

¹Whether we use here the speed of light in vacuum (c) or the phase velocity or, more appropriately, the group velocity, is inconsequential because the atmosphere is so weakly dispersive.

is $\sigma(z, \nu)/4\pi$, where $\sigma(z, \nu)$ is the total (over all angles) scattering cross section. But in general the scattering (or resonance fluorescence) is not isotropic, and to indicate this we have written $\sigma^B(z, \nu)$ in (14.1.8) instead of $\sigma(z, \nu)/4\pi$, where $\sigma^B(z, \nu)$ is the “differential” cross section for backscattering, and reduces to $\sigma(z, \nu)/4\pi$ if the scattering happens to be isotropic.

Various expressions relating to (14.1.7) and (14.1.8) are found in the lidar literature. For example, the average power at the receiver implied by (14.1.7) is

$$\begin{aligned} \text{Pwr}(z) &= \frac{h\nu\mathcal{N}(z)}{\eta\tau_d} = T_0^2(z) \frac{A_r}{z^2} E_p N(z) \sigma^B(z, \nu) \frac{c}{2} \\ &= (\text{Pwr})_0 T_0^2(z) \frac{A_r}{z^2} N(z) \sigma^B(z, \nu) \frac{c\tau_p}{2}, \end{aligned} \quad (14.1.9)$$

where $(\text{Pwr})_0$ is the single-pulse average power defined as E_p/τ_p . Similarly, if we take the time τ_d to be some arbitrary number of single pulse durations we infer from (14.1.7) that the photon count obtained from a single laser pulse is

$$\mathcal{N}(z) = \eta T_0^2(z) \frac{A_r E_p}{z^2 h\nu} N(z) \sigma^B(z, \nu) \frac{c\tau_p}{2}. \quad (14.1.10)$$

What is called “the lidar equation” in the literature may refer to any of expressions (14.1.7)–(14.1.10) or to some other variant of (14.1.7).

In practice the number of “background counts” $\mathcal{N}_B$, which includes detector dark counts (Section 12.7) and counts from any background light sources (e.g., the sun), must be added to the right-hand side of whatever form of the lidar equation is used.

If laser pulses propagate from ground at a zenith angle ψ , where $\psi = 0$ defines vertical propagation, then the distance of atoms at altitude z from the receiver is $R = z/\cos \psi$, and this distance replaces z in the lidar equation. Similarly the lidar equation is often written with $T_0^2(R)$ expressed as an integral over the path to R :

$$T_0^2(R) = \exp \left[-2 \int_0^R a(R') dR' \right], \quad (14.1.11)$$

where $a(R')$ is the atmospheric attenuation coefficient at propagation distance R' at the wavelength of interest. For instance, $a(R')$ due to Rayleigh scattering will obviously decrease with altitude as the density of scatterers decreases.

It has been assumed that the backscattered photons come from resonance fluorescence, but the lidar equation is in fact more broadly applicable, as its form suggests. If the backscattered photons are due to scattering by air molecules (predominantly N_2 and O_2), the cross section $\sigma^B(z, \nu)$ in the lidar equation is the Rayleigh backscattering cross section, which is

$$\sigma_R^B(z, \nu) = \frac{\pi^2 [n^2(z, \lambda) - 1]^2}{N^2(z) \lambda^4} \quad \begin{array}{l} \text{(Rayleigh backscattering cross section} \\ \text{for wavelength } \lambda = c/\nu), \end{array} \quad (14.1.12)$$

as shown below. The laser radiation is assumed to be narrowband and at the same frequency as the scattered radiation. Since Rayleigh scattering does not involve a change

in molecular energy levels, it is often called *elastic scattering* in the lidar literature. This terminology applies also to the more general case of *Mie scattering* in which the scatterers are not necessarily small compared to the wavelength. The lidar equation is based on the approximation that multiple scattering is negligible; this is the approximation that each molecule scatters only the laser radiation and not the radiation scattered in its direction by another molecule.

We have implicitly assumed in deriving the lidar equation that the field of view of the receiving optics is larger than the laser beam divergence. It should also be noted that, although we have referred to photon counts at the detector, as is appropriate when return signals are measured with photomultiplier tubes or avalanche photodiodes in the Geiger mode, strong return signals are usually recorded in the analog mode and then converted to digital form (Section 12.7).

- The electric field from an electric dipole oscillating along a direction $\hat{\mathbf{e}}$ is proportional to $(\hat{\mathbf{e}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \hat{\mathbf{e}}$ in the radiation zone, where $\hat{\mathbf{r}}$ is the unit vector pointing from the dipole to the point of observation [Eq. (8.9.13)]. The scattered power in the direction $\hat{\mathbf{r}}$ is therefore proportional to

$$[(\hat{\mathbf{e}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \hat{\mathbf{e}}]^2 = 1 - (\hat{\mathbf{e}} \cdot \hat{\mathbf{r}})^2. \quad (14.1.13)$$

For a dipole induced by an incident field, $\hat{\mathbf{e}}$ is the polarization unit vector of the incident field. Let θ be the scattering angle, that is, the angle between the propagation direction $\hat{\mathbf{z}}$ of the incident field and the direction $\hat{\mathbf{r}}$ in which the scattered field is observed. We write

$$\hat{\mathbf{r}} = \hat{\mathbf{x}} \sin \theta \cos \phi + \hat{\mathbf{y}} \sin \theta \sin \phi + \hat{\mathbf{z}} \cos \theta, \quad (14.1.14)$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are orthogonal unit vectors in the plane perpendicular to $\hat{\mathbf{z}}$ and ϕ is the angle between $\hat{\mathbf{x}}$ and the projection of $\hat{\mathbf{r}}$ onto the xy plane. Then, for incident radiation polarized along $\hat{\mathbf{x}}$ or $\hat{\mathbf{y}}$, respectively, we have

$$1 - (\hat{\mathbf{e}} \cdot \hat{\mathbf{r}})^2 = 1 - \sin^2 \theta \cos^2 \phi \quad (x \text{ polarization}), \quad (14.1.15a)$$

$$= 1 - \sin^2 \theta \sin^2 \phi \quad (y \text{ polarization}). \quad (14.1.15b)$$

In the case of unpolarized incident radiation the angular dependence of the scattered power can be obtained simply by taking the average of (14.1.15a) and (14.1.15b), since the x and y components of the incident field have the same (average) intensity:

$$\frac{1}{2}(1 - \sin^2 \theta \cos^2 \phi) + \frac{1}{2}(1 - \sin^2 \theta \sin^2 \phi) = 1 - \frac{1}{2} \sin^2 \theta = \frac{1}{2}(1 + \cos^2 \theta). \quad (14.1.16)$$

The same result is obtained for incident light that is circularly polarized because the x and y field components again have equal intensities. The integral of (14.1.16) over all solid angles Ω is

$$\int_{4\pi} d\Omega \frac{1}{2}(1 + \cos^2 \theta) = \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta \frac{1}{2}(1 + \cos^2 \theta) = \frac{8\pi}{3}, \quad (14.1.17)$$

and so we define the *differential scattering cross section* $d\sigma/d\Omega$ such that $\int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta (d\sigma/d\Omega) = \sigma_R$, where σ_R is the *total* cross section for Rayleigh scattering defined by (8.9.5):

$$\frac{d\sigma}{d\Omega} = \sigma_R \times \frac{3}{8\pi} \frac{1}{2}(1 + \cos^2 \theta) = \frac{\pi^2 [n^2(z, \lambda) - 1]^2}{\lambda^4 N^2(z)} \frac{1}{2}(1 + \cos^2 \theta), \quad (14.1.18)$$

for $n(\lambda) \cong 1$. Similarly, for incident light linearly polarized in the x direction,

$$\frac{d\sigma}{d\Omega} = \frac{\pi^2[n^2(z, \lambda) - 1]^2}{\lambda^4 N^2(z)} (1 - \sin^2 \theta \cos^2 \phi), \quad (14.1.19)$$

and for y -polarized light $\cos^2 \phi$ is replaced by $\sin^2 \phi$ in this formula. Therefore, the differential cross section for backscattering ($\theta = \pi$) is

$$\sigma_R^B(z, \nu) \equiv \frac{d\sigma}{d\Omega}(\theta = \pi) = \frac{\pi^2[n^2(z, \lambda) - 1]^2}{N^2(z)\lambda^4} \quad (14.1.20)$$

for *any* polarization of incident light.

Using (14.1.20) and the formula (8.10.5) for the refractive index of air, we obtain for the quantity $N(z)\sigma_R^B(z, \nu)$ in the lidar equation for Rayleigh scattering the approximation (Problem 14.2)

$$N(z)\sigma_R^B(z, \nu) \cong \frac{3.3 \times 10^4 P(z)}{\lambda^4} \frac{P(z)}{T(z)} \text{ m}^{-1}, \quad (14.1.21)$$

where $P(z)$ and $T(z)$ are, respectively, the atmospheric pressure (millibars) and temperature (Kelvin) at altitude z and λ is the wavelength in nanometers.²

Rayleigh lidars typically probe the atmosphere at altitudes above about 30 km, where scattering from aerosols is negligible.³ They have been used to infer atmospheric temperature distributions at such altitudes. To get a rough idea of the number of photon counts predicted by the lidar equation, consider a fairly typical sort of Rayleigh lidar system in which 532-nm, 1-J laser pulses at a repetition rate of 30 Hz propagate vertically to 30 km, and backscattered photons are returned from a 150-m range bin ($\tau_d = 1 \mu\text{s}$) to a 50-cm-diameter receiver aperture over an integration time of 1 min. At $z = 30$ km the atmospheric pressure and temperature are approximately 12 mbar and 250K, respectively; then, from (14.1.8) and (14.1.21) (Problem 14.2),

$$\mathcal{N}(z) \sim 3.2 \times 10^6 \times \eta T_0^2(z) \quad (14.1.22)$$

from a 150-m range bin at 30 km. A reasonable estimate of $T_0^2(z)$ at 30 km is about 0.6 for visible wavelengths far from any absorption resonances of air; the attenuation in this case is due predominantly to Rayleigh scattering from air molecules and to scattering from aerosols. Assuming $\eta \sim 0.5$ for the detection efficiency, we obtain $\mathcal{N}(z) \sim 10^6$ photon counts. Note that the power $h\nu\mathcal{N}(z)/(\eta\tau)$ backscattered to the receiver during the integration time in this example is about 10^{-14} W, compared to an average laser power of (30 Hz)(1 J) = 30 W. •

The contribution of Rayleigh scattering by air molecules to the backscattered signal can be calculated, since the backscattering cross section and the attenuation coefficient for Rayleigh scattering are known as a function of pressure and temperature. The spectral width of the backscattered light depends on the velocity distribution of the air molecules. Since any aerosol particles present are much more massive and therefore have much smaller thermal velocities than air molecules, the spectral width of backscattered light from them is much narrower than that for Rayleigh scattering by air molecules. Narrow-bandwidth filters can be used to distinguish the contributions of air and aerosol

²Since $\sigma_R^B(z, \nu)$ is a *differential* cross section, $N(z)\sigma_R^B(z, \nu)$, strictly speaking, has dimensions of $(\text{length})^{-1}$ (steradian)⁻¹ rather than $(\text{length})^{-1}$. This distinction does not affect computations of backscattered photon numbers.

³In the lidar literature Rayleigh scattering refers specifically to scattering by molecules rather than by aerosol particles, which may or may not be small compared to the wavelength.

particles to the elastic backscatter, and therefore to determine distributions of aerosol concentrations. Similarly, whereas the absolute Doppler shift due to line-of-sight wind is the same for light scattered by aerosols and by air molecules, the *relative* Doppler shift for aerosols is much larger; in other words, the ratio of the Doppler shift resulting from being carried along by wind to the average thermal velocity is larger for aerosols. The large relative frequency shift allows the Doppler shift to be accurately measured as a separation between two narrow “spikes” in the spectrum of the backscattered light. This allows accurate profiling of atmospheric winds and is the basis for the atmospheric laser Doppler instrument (ALADIN) planned for operation aboard the European Space Agency’s ADM-Aeolus “wind watch” satellite. This lidar system will operate at the 355-nm wavelength of frequency-tripled Nd : YAG laser radiation. It is expected to provide 100 wind profiles per hour with a range resolution of 1 km for altitudes up to 30 km.

Elastic-backscatter lidars can detect the presence of large, nonspherical particles (e.g., ice crystals or soot). For spherical particles the backscattered light has the same polarization as the laser radiation. Nonspherical particles, however, result in a “depolarization” of backscattered light; analysis of this depolarization can therefore determine whether such particles are present. Polarization-sensitive elastic-backscatter lidar has been useful, for instance, in studies of the distribution of ice crystals in cirrus clouds and of aerosols in dust layers.

Resonance fluorescence lidar studies of sodium and other mesospheric metallic species (e.g., potassium and iron) believed to result from meteoric ablation have also been important in atmospheric research. Sodium is found in a layer about 30 km wide centered at about 95 km.⁴ It has been especially useful in resonance fluorescence lidar because of its strong D line transitions and its *relatively* large density (only a few thousand atoms per cubic centimeter!) in the mesosphere.

To be specific, let us consider, as in Section 3.13, the $3S_{1/2}(F=2) \leftrightarrow 3P_{3/2} D_2$ transition. We can estimate as follows the number of photon counts $\mathcal{N}$ that can be obtained from the full width of the sodium layer with laser radiation at this transition frequency. Assuming n_b range bins of length Δz [$\sim 30 \text{ km}/n_b$] in the lidar equation (14.1.8), we write

$$\begin{aligned} \mathcal{N} &\sim \sum_{m=1}^{n_b} \mathcal{N}(z + m \Delta z) \sim \eta T_0^2 \frac{(\text{Pwr})_L \tau A_r}{h\nu} \frac{1}{z_{\text{eff}}^2} \sigma_{\text{Na}}^B \sum_{m=1}^{n_b} N(z + m \Delta z) \Delta z \\ &= \eta T_0^2 \frac{(\text{Pwr})_L \tau A_r}{h\nu} \frac{1}{z_{\text{eff}}^2} \sigma_{\text{Na}}^B C_s, \end{aligned} \quad (14.1.23)$$

where C_s (m^{-2}) is the sodium *column density*. We have replaced $T_0(z)$ and z within the sodium layer by effective values T_0 and z_{eff} , and σ_{Na}^B is the backscattering cross section for sodium resonance fluorescence, which for our rough estimates is assumed to be isotropic; thus we take

$$\sigma_{\text{Na}}^B = \frac{1}{4\pi} (9.3 \times 10^{-16} \text{ m}^2), \quad (14.1.24)$$

⁴The sodium layer is seen as a faint, thin yellow arc in a photograph taken by an astronaut during the September, 1992 flight of the space shuttle *Endeavour*. This photograph is reproduced in W. J. Wild and R. Q. Fugate, *Sky and Telescope* **87**, 10 (1994).

where $9.3 \times 10^{-16} \text{ m}^2$ is the total (absorption) cross section calculated in Section 3.13 [Eq. (3.13.10)] for $T = 200\text{K}$, which is approximately the temperature in the mesosphere. Then, from (14.1.23) with $z_{\text{eff}} = 95 \text{ km}$, we obtain the number of photons per unit area and per unit time arriving at the receiver during the integration time τ :

$$\frac{\mathcal{N}}{\eta A_r \tau} \sim 2.4 \times 10^{-8} T_0^2 (\text{Pwr})_L (W) C_s (\text{m}^{-2}) \text{ photons/m}^2/\text{s}. \quad (14.1.25)$$

The sodium column density C_s varies with location, time of year, and even time of day, typical values being $3 - 6 \times 10^{13} \text{ m}^{-2}$, corresponding to peak sodium densities of around a few thousand atoms per cubic centimeter. Using $C_s = 4 \times 10^{13} \text{ m}^{-2}$ and $T_0^2 \sim 0.5$, we estimate that $\mathcal{N}/(\eta A_r \tau) \sim 5 \times 10^5 T_0^2$ photons/m²/s per watt of laser power arrive at the detector. This is much smaller than the photon flux from low-altitude Rayleigh backscattering, which can be eliminated from the measured sodium resonance fluorescence signal by range gating [for pulses shorter than about $(95 \text{ km})/c \sim 300 \mu\text{s}$] such that photon counting at the receiver is begun only after a sufficiently long time following the launch of a laser pulse.

We have made assumptions about column density and temperature in order to estimate photon returns, but, of course, one application of lidar is to *determine* such quantities from measured photon counts. The temperature profile of the mesosphere, for example, has been measured by sodium resonance fluorescence lidar based on the temperature dependence of the absorption cross section shown in Fig. 3.20. The maximum and minimum backscattering cross sections occur at frequencies we denote by ν_{max} and ν_{min} , respectively, and the ratio of these cross sections depends on the temperature. The ratio of range-gated photon return signals taken at the two laser frequencies ν_{max} and ν_{min} , for example, can therefore be compared with the theoretical ratio (Fig. 3.20) to infer the temperature as a function of altitude in the mesosphere. Figure 14.2 shows results of

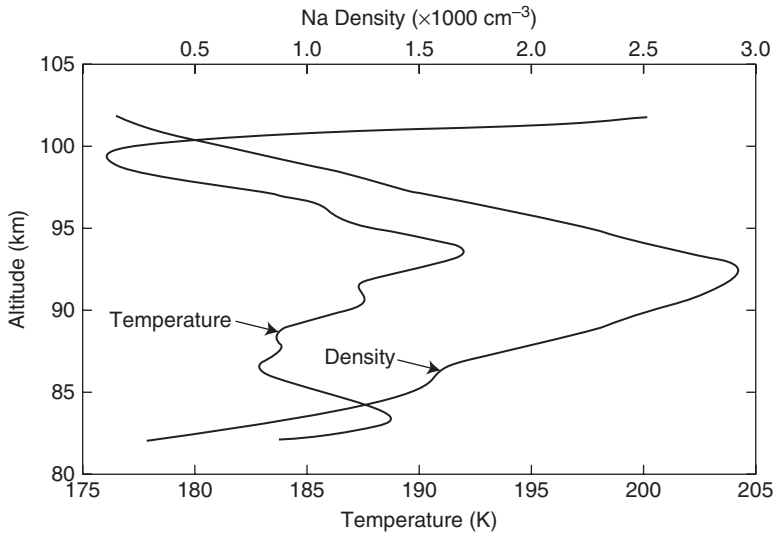


Figure 14.2 Temperature and sodium density in the mesosphere over Ft. Collins, Colorado, obtained from sodium resonance fluorescence lidar. [From R. E. Bills, C. S. Gardner, and C.-Y. She, *Optical Engineering* **30**, 13 (1991).]

such measurements for the temperature as well as the sodium density. The relatively high-resolution ($\Delta z = 1$ km) data were obtained with a cw dye laser tuned alternately between two frequencies.

- The sodium density can be obtained by normalizing the sodium photocounts to the Rayleigh photocounts from aerosol-free altitudes ($\gtrsim 35$ km) at which the Rayleigh scattering is due almost entirely to air molecules; this has the advantage of effectively eliminating uncertainties in the atmospheric transmission coefficient $T_0(z)$. Thus, from Eq. (14.1.8) it follows that the sodium density $N_{\text{Na}}(z)$ at altitude z can be expressed as (Problem 14.3)

$$N_{\text{Na}}(z) = \left[\frac{N_R(z_R) \sigma_R^B(z_R, \nu)}{\sigma_{\text{Na}}^B(z, \nu)} \right] \frac{z^2 \mathcal{N}_{\text{Na}}(z)}{z_R^2 \mathcal{N}_R(z_R)}, \quad (14.1.26)$$

Here z_R and z are the altitudes for which the Rayleigh and sodium photon counts $\mathcal{N}_R(z_R)$ and $\mathcal{N}_{\text{Na}}(z)$, respectively, are taken, and $N_R(z_R)$ is the density of molecular (Rayleigh) scatterers at z_R . The factor in brackets on the right-hand side is known theoretically as a function of the laser frequency [cf. (14.1.21)], z^2/z_R^2 is determined by the altitudes probed, and the ratio of sodium and Rayleigh photocounts is measured to infer $N_{\text{Na}}(z)$.

In deriving the lidar equation, we used the proportionality (14.1.3) between the excited-state population and the intensity; this ignores the possibility of saturation or coherent excitation effects that depend on the temporal variations of the electric field. In most lidar systems saturation or coherent excitation effects are of no concern, and the forms of the lidar equation we have written are quite adequate. It should be remembered, however, that backscattering cross sections generally depend on the polarization of the laser. In narrowband lidars, moreover, it is often necessary to account for the laser linewidth via a convolution of the frequency-dependent cross section with the laser spectrum. •

In *differential absorption lidar* (DIAL) the wavelength dependence of the transmission coefficient T_0 is used to determine the densities of absorbing species. Consider the lidar equation for the power at the receiver [cf. (14.1.9)] when the laser is at wavelength λ :

$$\text{Pwr}(R, \lambda) = (\text{Pwr})_0(\lambda) \frac{A_r}{R^2} N(R) \sigma^B(R, \lambda) \Delta R e^{-2 \int_0^R dR' a(R', \lambda)}, \quad (14.1.27)$$

where we have used (14.1.11) and have written R instead of z to allow for arbitrary zenith angles; ΔR is the range bin length, $(\text{Pwr})_0(\lambda)$ is the laser power at $R = 0$ at wavelength λ , and $a(R', \lambda)$ is the atmospheric attenuation coefficient at wavelength λ and range R' . The basic idea of DIAL is to measure photon return signals at two wavelengths λ_{on} and λ_{off} , where λ_{on} is a wavelength that is “on” resonance with an absorption line of the atmospheric molecule of interest and λ_{off} is “off” resonance. In other words, the molecule of interest absorbs at wavelength λ_{on} but not at λ_{off} . From (14.1.27) we obtain the ratio

$$\frac{\text{Pwr}(R, \lambda_{\text{on}})}{\text{Pwr}(R, \lambda_{\text{off}})} = \exp \left(-2 \int_0^R dR' [a(R', \lambda_{\text{on}}) - a(R', \lambda_{\text{off}})] \right). \quad (14.1.28)$$

We are assuming that the laser powers as well as the range bin lengths are the same at the two wavelengths, and that the two backscattering cross sections due, for instance, to Rayleigh scattering, are also approximately the same.

The attenuation at the two wavelengths is attributable to scattering from air molecules and aerosols as well as to absorption by any molecules with absorption lines near these wavelengths. Suppose, however, that only a particular molecular species contributes significantly to the integral over R of the *difference* $a(R, \lambda_{\text{on}}) - a(R, \lambda_{\text{off}})$; this, of course, is the species probed by DIAL with (slightly) different laser wavelengths λ_{on} and λ_{off} . Then we can ignore all contributions to (14.1.28) except that from this particular molecular species. Since

$$a(R', \lambda_{\text{on}}) - a(R', \lambda_{\text{off}}) = N_a(R')[\sigma_a(R', \lambda_{\text{on}}) - \sigma_a(R', \lambda_{\text{off}})], \quad (14.1.29)$$

where N_a and σ_a are the number density and absorption cross section, respectively, of this particular molecule, it follows from (14.1.28) that

$$\frac{\text{Pwr}(R, \lambda_{\text{on}})}{\text{Pwr}(R, \lambda_{\text{off}})} = \exp \left[-2 \int_0^R dR' N_a(R') \Delta\sigma_a(R') \right], \quad (14.1.30)$$

where we define the “differential absorption cross section”

$$\Delta\sigma_a(R') = \sigma_a(R', \lambda_{\text{on}}) - \sigma_a(R', \lambda_{\text{off}}). \quad (14.1.31)$$

Differentiation of (14.1.30) yields

$$N_a(R) = \frac{-1}{2\Delta\sigma_a(R)} \frac{d}{dR} \left[\ln \frac{\text{Pwr}(R, \lambda_{\text{on}})}{\text{Pwr}(R, \lambda_{\text{off}})} \right], \quad (14.1.32)$$

or, since the return signals are, of course, recorded in discrete range bins ΔR ,

$$N_a(R) \cong \frac{1}{2\Delta\sigma_a(R) \Delta R} \ln \left[\frac{\text{Pwr}(R + \Delta R, \lambda_{\text{off}})}{\text{Pwr}(R, \lambda_{\text{off}})} \frac{\text{Pwr}(R, \lambda_{\text{on}})}{\text{Pwr}(R + \Delta R, \lambda_{\text{on}})} \right]. \quad (14.1.33)$$

Sometimes DIAL is said to be (approximately) “self-calibrating”: by measuring the return signals at the two wavelengths λ_{on} and λ_{off} , the number density $N_a(R)$ in different range bins is determined independently of the area of the receiving aperture, the photon-counting efficiency, the absolute laser power, the density and cross sections of the back-scatterers, and the atmospheric transmission coefficient $T_0(R)$. It is, of course, essential that the differential absorption cross section $\Delta\sigma_a(R)$ of the molecule of interest be known from theory or from laboratory experiments. DIAL has played a very important role in the determination of absolute concentrations of ozone, water vapor, carbon dioxide, and other environmentally important absorbers. It has also been used for temperature profiling of the atmosphere based on the known concentration of O_2 and the temperature dependence of its differential absorption cross section.

- Lidar signals are often very weak, and it is sometimes essential to detect them in such a way as to discriminate against background noise. This is done by *heterodyne detection*, as opposed to the “direct detection” we have presumed. In heterodyne detection the return light is superposed with light from a “local oscillator,” a cw laser at frequency ω_{LO} . Let $E_S \cos(\omega_S + \phi_S)t$ and $E_{\text{LO}} \cos(\omega_{\text{LO}}t + \phi_{\text{LO}})$ denote the electric fields of the signal and local oscillator, respectively. The electric current i_D of a photodetector will be proportional to the total intensity $\epsilon_0 c [E_S \cos(\omega_S t + \phi_S) + E_{\text{LO}} \cos(\omega_{\text{LO}}t + \phi_{\text{LO}})]^2$, or

$$i_D \propto \frac{1}{2} E_S^2 + \frac{1}{2} E_{\text{LO}}^2 + E_S E_{\text{LO}} \cos[(\omega_S - \omega_{\text{LO}})t + (\phi_S - \phi_{\text{LO}})] \quad (14.1.34)$$

when terms at frequencies $2\omega_S$, $2\omega_{LO}$, and $\omega_S + \omega_{LO}$ are dropped under the assumption that these frequencies are so large that the detector can only respond to their time average, which is zero. In addition to a dc component, the current at the detector oscillates at the difference (beat) frequency $\omega_S - \omega_{LO}$, typically at radio frequencies, and with an amplitude proportional to $\sqrt{(\text{Pwr})_S(\text{Pwr})_{LO}}$, where $(\text{Pwr})_S$ and $(\text{Pwr})_{LO}$ are the powers of the signal and local oscillator. The principal advantage of heterodyning is the greater signal-to-noise ratio achievable with a sufficiently strong local oscillator field.

Heterodyne detection is “coherent” in that it responds to the signal field rather than directly to the intensity; it records information about the phase (and polarization) as well as the power of the signal. Lidars employing heterodyne detection are referred to as coherent DIAL, coherent Doppler, and so forth. Heterodyne detection requires a highly stable pulsed laser and local oscillator as well as sufficiently fast detectors; while offering better noise performance than direct detection, it is more sensitive to phase perturbations and misalignments. Because of the speckle effect due to atmospheric turbulence (Sections 8.11 and 13.8), there are shot-to-shot fluctuations in the return signal from each range bin, and a time averaging must be performed to smooth out these fluctuations. The heterodyne signal decreases for ranges R greater than the coherence length of the source (Section 13.11). ●

Other lidar techniques are based on Raman scattering, an *inelastic* scattering process involving a change in the vibrational-rotational state of a molecule. The scattered radiation is shifted in frequency from the incident laser radiation by the change in the vibrational-rotational energy. Since the energy levels are distributed according to a Boltzmann distribution at the ambient temperature, Raman lidars have been used for temperature profiling of the atmosphere. Unlike resonance fluorescence lidar or DIAL, Raman lidars do not require specific laser wavelengths that match an absorption line of an atmospheric constituent; Raman cross sections are approximately proportional to λ^{-4} , as in Rayleigh scattering, so that the laser wavelengths used (often between about 320 and 550 nm) are typically smaller than in other lidars. The cross sections tend to be small, rendering Raman lidars most useful for probing constituents with relatively high concentrations. One of their principal applications is to measurements of water vapor concentrations in the troposphere, the lowest and densest layer of Earth’s atmosphere containing nearly all its water vapor.⁵

The implementations and results of lidar are far too diverse to describe in any detail here. Mobile lidar systems housed in trucks and airplanes are used to monitor concentrations in air of many constituents including ozone, carbon dioxide, methane, water vapor, industrial emissions, and pollutants. Airborne lasers in the visible and ultraviolet (e.g., frequency-doubled Nd:YAG laser radiation at 532 nm) are used in fluorescence lidars in which chlorophyll, for example, absorbs at 532 nm and fluoresces at 685 and 740 nm; the strength of the radiation at these wavelengths provides information about forest ecosystems and environmental variations. Fluorescence submarine lidars have been developed for the detection of accidental or illegal chemical discharges in seawater. All such lidars employ some variant of the basic lidar equation and software for the “inversion” of the lidar equation to retrieve the desired information from measured return signals.

The first spaceborne lidar—the Lidar-in-Space Technology Experiment (LITE)—was launched in September, 1994 as the primary payload aboard the U.S. space shuttle

⁵The troposphere extends from ground to ~ 8 km at the poles and to ~ 18 km at the equator. It is where weather “happens.”

Discovery. The transmitter was a Nd:YAG laser with frequency doublers and triplers generating 532 and 355 nm wavelengths at a 10-Hz pulse repetition rate. The receiving telescope had a 1-m diameter, and photomultipliers (for 532 and 355 nm) and an avalanche photodiode (for 1064 nm) were used for the detection of return signals; the vertical range resolution was about 15 m. This experimental system, the data from which were validated by six lidar-equipped aircraft as well as more than 90 ground-based systems in 20 countries, demonstrated that return signals in space could be obtained from ground or close to ground, and it succeeded in identifying storm systems, dust layers, and complex cloud structures. Earth-orbiting lidar systems under development will take comprehensive data from clouds and aerosols in order to obtain information needed for climate modeling.

Lasers have long been used for distance and velocity measurements. Range finders employ photodetection and timing electronics to measure the time T it takes for a light pulse to reach an object and reflect back to the transmitter: The distance to the object is $d = cT/2$. The advantages of laser pulses over radio-frequency systems (radar) are their very short durations, directionality, and their intensities that result in much stronger echo signals. On July 20, 1969, Apollo 11 astronauts placed an 18-inch-square reflector on the moon, and return pulses from a ruby laser at the Lick Observatory in California were first detected on August 1. Distances between points on Earth and the moon were determined to an accuracy of a few inches. The U.S. Laser Geodynamics Satellite (LAGEOS) launched in 1976 is a sphere of mass 407 kg and diameter 60 cm covered with 426 retro-reflectors for satellite laser ranging, similar in principle to lunar laser ranging. Laser tracking of the position of LAGEOS in its orbital altitude ~ 5900 km revealed small variations in the length of the Earth day. A second LAGEOS satellite was launched in 1992 to provide more data relevant to seismic activity. Both satellites remain in service as of this writing.

Laser ranging can also be done interferometrically. In a Michelson interferometer (Section 13.10), for example, the fringe maxima and minima are interchanged when the arm separation is changed by $\lambda/2$: The magnitude of this change in the arm separation can be determined in terms of the wavelength λ by counting the number of fringe shifts as the change occurs. This technique is used routinely in length calibrations and machining applications.

Lasers are also used to measure velocities based on the Doppler effect. Laser velocimeters, some employing heterodyne detection, allow accurate measurements of velocities of aircraft and fluid flows, for example. Velocities can also be measured by *time of flight*: The time T for a laser pulse to propagate to a target and back implies the distance $d = cT/2$ to the object, and the slope of the plot of distance vs. time is the velocity. This is how police lidar guns work. They typically employ GaAs diode lasers at an eye-safe wavelength (904 nm) and average power (~ 50 μ W). Since ~ 100 pulses are used to compute the velocity, a typical pulse repetition rate of 300 Hz implies that only ~ 0.3 s is needed for a velocity determination—not much time for a speeding driver to react and slow down.

14.2 ADAPTIVE OPTICS FOR ASTRONOMY

In Chapter 8 we introduced concepts such as the refractive index structure constant C_n^2 , the coherence diameter r_0 , and the seeing angle θ that are especially important in the

theory of the propagation of light in the turbulent atmosphere. We showed that even under good seeing conditions a large ground-based telescope, while collecting more light and allowing the observation of far fainter objects, will have no greater image resolution than a much smaller telescope of comparable optical quality. Were it possible to eliminate the effects of atmospheric turbulence, the image-resolving capability of ground-based telescopes could exceed that of a space telescope (Section 8.11). Substantial enhancement of the imaging performance of ground-based telescopes is in fact possible with *adaptive optics*.

The degradation of image quality caused by atmospheric turbulence arises primarily from phase distortions across an incoming wavefront; recall the discussion in Section 8.11, where we showed that relatively small phase variations can substantially degrade image quality. The basic idea of adaptive optics is to measure the phase distortions, that is, the variations in local time lags across a wavefront, and to correct for them by advancing the phase at points where the atmosphere has retarded it and retarding it at points where the atmosphere has advanced it. The result, ideally, is a “corrected” wavefront in which transverse phase variations have been removed and the image of an object in a telescope is free of the blurring effect of the phase-distorting atmosphere. The advancement and retardation of the local phases are done with a deformable mirror surface, or “rubber mirror,” as indicated in Fig. 14.3. The phase distortions across an incident wavefront—or more precisely the transverse phase *gradients*—are measured by a wavefront sensor and used by a wavefront reconstructor to compute approximations to the actual phase gradients. These are converted to voltages that drive an array of mechanical actuators (or “pistons”) that adjust the shape of the deformable mirror surface so that the phase variations across an incoming wavefront are (ideally) absent in the reflected wavefront.

The most common type of wavefront sensor in adaptive optics is the *Shack–Hartmann sensor*, an array of lenslets or “subapertures” (Fig. 14.4). Each lenslet produces a spot on a detector array in the focal plane, and the displacement of the spot from its local null position is a measure of the local phase gradient—and therefore the local ray propagation direction—of the incoming wavefront at that subaperture. (The null positions of the focal spots can be defined using an undistorted incoming wavefront

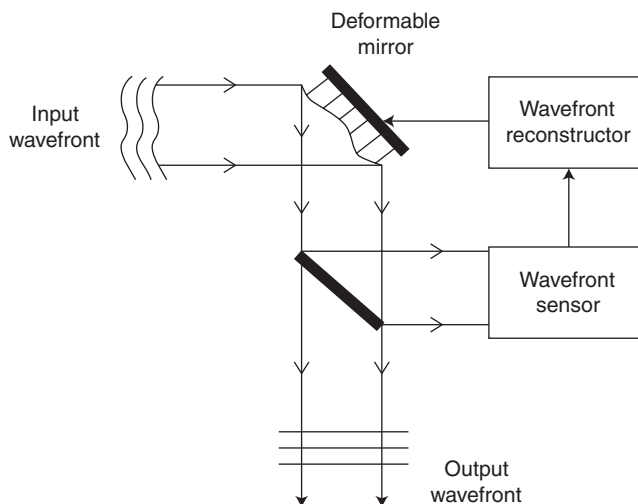


Figure 14.3 Schematic of the principal components of an adaptive optical system.

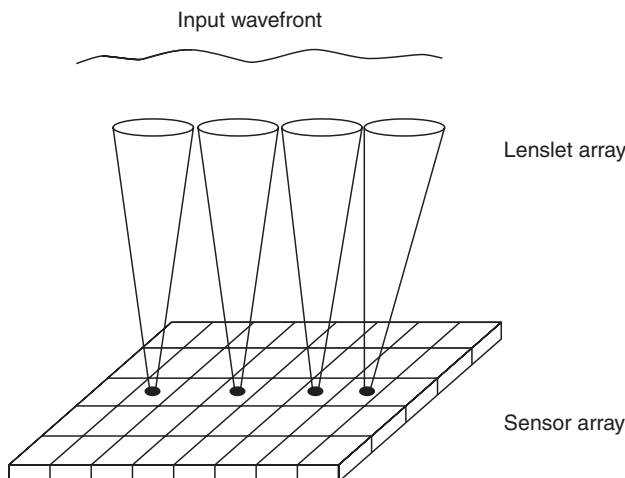


Figure 14.4 Shack–Hartmann sensor consisting of an array of small lenslets (subapertures), each of which focuses a small portion of an incoming wavefront onto a spot on a detector array in the focal plane.

from a laser.) The Shack–Hartmann sensor thus transforms the local phase gradients of the incoming wavefront into a matrix of focal-spot displacements. The spots have a diffracted-limited form if the lenslet diameters are smaller than the atmospheric coherent diameter r_0 (Section 8.11). The detection array used to determine the phase gradient associated with each subaperture is usually a “quad cell,” which consists of four detectors (e.g., avalanche photodiodes) forming the quadrants of a square: The different photon counts in the four quadrants provide the measure of the local phase gradient.

The performance of the adaptive optical system depends, among other things, on how accurately the local phase gradients can be measured. Measurement errors arise from uncertainties in the centroid positions of the focal spots in the Shack–Hartmann sensor. These uncertainties are due to photon number fluctuations of the incoming wavefront and to detector readout noise: Both types of noise diminish as the average number of incoming photons per subaperture increases.

The wavefront “reconstruction” process involves algorithms and software for estimating the actual phase profile from the measured matrix of phase gradients. This reconstruction provides the information used to adjust the deformable mirror and retard or advance the phase in such a way as to obtain the flat, undistorted output wavefront indicated in Fig. 14.3.

Deformable mirrors are either segmented or continuous. The segmented type consists of closely spaced planar mirror segments, each of which can be independently positioned, whereas the continuous type uses a single reflecting sheet (Fig. 14.5). Continuous deformable mirrors require more sophisticated control algorithms but avoid alignment and edge diffraction complications that arise with segmented mirrors.

When employed at a telescope the adaptive optical system of Fig. 14.3 operates as follows. Light that has been phase distorted by atmospheric turbulence enters the telescope and reflects off the deformable mirror. Part of the reflected light is incident on the wavefront sensor. The sensor measures the phase “errors” (deviations of focal spots from null positions), which would vanish if the deformable mirror were such that the reflected

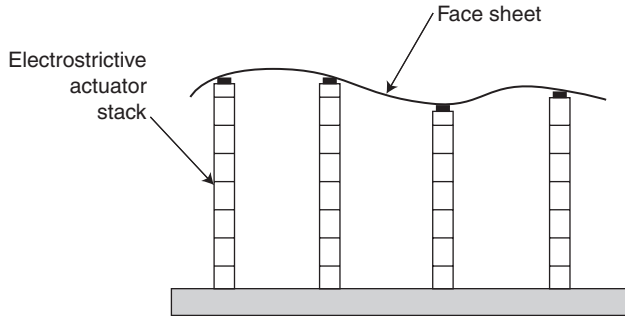


Figure 14.5 Continuous deformable mirror whose shape is controlled by voltages applied to a stack of electrostatic actuators (pistons). Deflections $\sim \pm 1\mu\text{m}$ are typically obtained with drive voltages $\sim \pm 100\text{ V}$.

wavefront had no phase distortions. The wavefront reconstructor then converts the sensor output into voltages that drive the “pistons” of the deformable mirror to change its shape in such a way as to reduce the phase errors measured by the sensor. In other words, the adaptive optical system functions as a servo-control loop—a very complicated one because of the large number of elements (pistons) that must be simultaneously controlled. Moreover, it must correct for phase errors on a time scale short compared to the $\sim 10\text{ ms}$ in which phase distortions of incoming light change because of atmospheric fluctuations (Section 8.11). Deformable mirrors typically have much faster response times, so the rapidity of atmospheric fluctuations is not problematic.

Adaptive optical systems also employ a fast steering or “tip-tilt” mirror to remove the overall phase slope or tilt (Section 8.11) of the wavefront; tilt causes the image wander or “dancing” that is most responsible for the blurring of images obtained with long exposure times (Section 8.12). A tip-tilt sensor controls the steering mirror to stabilize the image for correction by the deformable mirror of “higher-order” phase distortions. Tilt typically accounts for more than 80% of the total power in wavefront distortion, but it can be followed and removed by the steering mirror with the light from stars as dim as magnitude 17–19; stars of the required magnitude cover about 60% of the sky.⁶

- A highly simplified model shows how wavefront sensing errors decrease with increasing light intensity. Consider instead of a quad cell a two-segment detector geometry that divides the image plane in the neighborhood of a focal spot from a subaperture into the regions $x < 0$ and $x > 0$. Let N_L and N_R be the number of photons counted in the regions $x < 0$ and $x > 0$, respectively, in a given time interval. Since the direction of ray propagation at the subaperture is perpendicular to the local wavefront, a phase tilt implies $N_R \neq N_L$. The approximately linear phase difference across the small subaperture results in a measured phase gradient

$$\Phi \propto \frac{N_R - N_L}{N_R + N_L}, \quad (14.2.1)$$

⁶It should be noted that wavefront tilt cannot be compensated using (single-wavelength) artificial laser guide stars because of “reciprocity.” The downward-propagating light from the guide star follows the same path as the upward-propagating laser beam, and so the guide star image will not appear to wander in the focal plane of the telescope.

and fluctuations ΔN_R and ΔN_L in N_R and N_L will cause a fluctuation

$$\begin{aligned}\Delta\Phi &\propto \frac{(N_R + N_L)(\Delta N_R - \Delta N_L) - (N_R - N_L)(\Delta N_R + \Delta N_L)}{(N_R + N_L)^2} \\ &= \frac{N_L \Delta N_R - N_R \Delta N_L}{(N_R + N_L)^2}\end{aligned}\quad (14.2.2)$$

in the measured phase. Assuming $\langle \Delta N_R \rangle = \langle \Delta N_L \rangle = \langle \Delta N_R \Delta N_L \rangle = 0$, $\langle N_R \rangle = \langle N_L \rangle \equiv N/2$, and Poisson statistics for N_R and N_L [$\langle \Delta N_R^2 \rangle = \langle N_R \rangle$, $\langle \Delta N_L^2 \rangle = \langle N_L \rangle$], we obtain a mean-square fluctuation

$$\langle \Delta\Phi^2 \rangle \propto 1/N \quad (14.2.3)$$

for the measured phase, where N is the average photon count, or, more precisely, the average number of photoelectrons collected per subaperture. Wavefront sensing therefore increases in accuracy with increasing light intensity. •

Astronomical sources are too faint for the wavefront sensor to make accurate enough phase measurements for correction of higher-order distortions. Adaptive optical systems for astronomy therefore use the brighter light from a reference “guide star.” Ideally the light from the guide star is distorted in the same way by atmospheric turbulence as the light from an astronomical object of interest, so that removal of the phase distortions of the light from the guide star will result in an improved image of the object. In order for the light from the guide star and the object to be phase distorted in the same way by atmospheric turbulence, the distortions should derive from the same “patch” of sky. The angle θ subtended by this patch at the receiver should be no larger than roughly the coherence diameter r_0 divided by the effective height (~ 10 km) of the atmosphere, in this case, $\theta \lesssim 20 \text{ cm}/10 \text{ km} = 4 \text{ arcsec} = 19 \mu\text{rad}$ for $r_0 = 20 \text{ cm}$ in the visible. This maximum allowable angular separation between the object and the guide star is called the *isoplanatic angle*. Unfortunately, there is a dearth of sufficiently bright stars that can be used as guide stars—it is estimated, for instance, that only about 0.1% of the sky has stars sufficiently bright to serve as guide stars in the visible.

However, lasers allow *artificial* guide stars to be created anywhere in the sky. For example, the Rayleigh-backscattered light from a laser beam propagating up into the atmosphere can serve as a guide star. The resonance fluorescence from laser-excited mesospheric sodium atoms provides another type of artificial guide star. Mesospheric sodium has a major advantage over Rayleigh scattering as a guide star: It is at an altitude (~ 95 km) much higher than can be probed with sufficient photon returns by Rayleigh scattering. The light from it therefore suffers phase distortions attributable to essentially the full height of the atmosphere, and so an adaptive optical system employing it as a guide star will “correct” the wavefront from an astronomical object at “infinity.” Figure 14.6 shows how an adaptive optical system and an artificial guide star can be employed at a ground-based telescope. Experiments have demonstrated the feasibility of adaptive optics using artificial as well as natural guide stars (Figs. 14.7 and 14.8).⁷ We will confine the remainder of our discussion mainly to the sodium guide star.

⁷For a comprehensive review of optical technology for improving the resolution of large ground-based telescopes see, for instance, M. C. Roggemann, B. M. Welsh, and R. Q. Fugate, *Reviews of Modern Physics* **69**, 437 (1997).

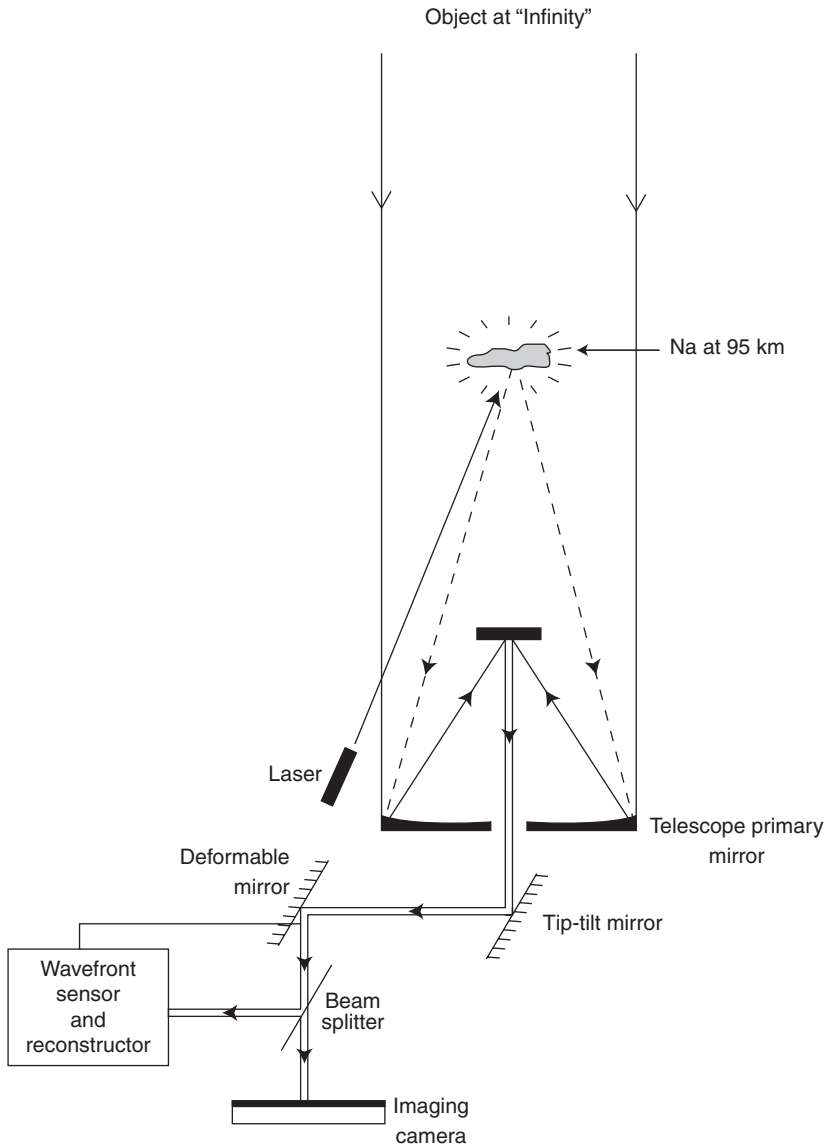


Figure 14.6 Schematic showing how adaptive optics with an artificial guide star can be used to improve the resolution of a telescope.

It is estimated that photon returns from the sodium layer of $\sim 10^6$ photons/m²/s are required for adaptive optics as described above to result in a Strehl ratio of $\exp(-1.0) = 0.37$, a nominal value associated with generally adequate imaging.⁸ Calculations of photon returns from the sodium guide star are the same as for the sodium resonance

⁸Estimates of required photon numbers for desired Strehl ratios, and comparisons of theoretical and experimental photon returns for various laser pulse formats, are discussed, for instance, in P. W. Milonni, R. Q. Fugate, and J. M. Telle, *Journal of the Optical Society of America A* **15**, 217 (1998).



Figure 14.7 The 3.5-m telescope facility at the Starfire Optical Range in New Mexico with a 589-nm beam propagating to the mesosphere to produce a sodium guide star. (Photo courtesy of J. M. Telle.)

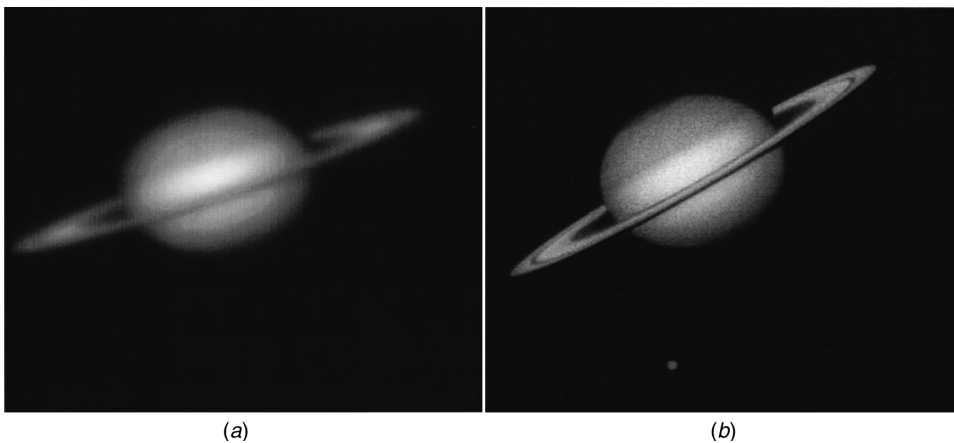


Figure 14.8 One of the early demonstrations of adaptive optics in astronomy (July, 1994). Images of Saturn and one of its moons (Titan) obtained (a) without and (b) with adaptive optics using a Rayleigh guide star at 656.4 nm with the 1.5-m telescope at the Starfire Optical Range in New Mexico. (Photo courtesy of R. Q. Fugate.)

fluorescence lidar and are generally consistent with observations. For example, formula (14.1.25) predicts a return photon flux of 1.6×10^6 photons/m²/s per watt of cw laser power at 589 nm, compared with a reported 1.9×10^6 photons/m²/s per watt when $T_0^2 C_s = 6.7 \times 10^{13} \text{ m}^{-2}$.⁹ The reported data were actually the average of photon returns obtained with linearly and circularly polarized excitation, which gave 1.5×10^6 photons/m²/s/W and 2.2×10^6 photons/m²/s/W, respectively. For reasons discussed in the following section, circular polarization generally results in larger photon return fluxes. With a 30-W, circularly polarized transmitted beam, for example, the photon return was approximately 70×10^6 photons/m²/s, whereas formula (14.1.25), which is applicable for laser intensities sufficiently small that the polarization effects discussed in the following section are insignificant, predicts a return of 52×10^6 photons/m²/s for 30 W of transmitted power.

- The intensity of light from an artificial guide star can be characterized by the “apparent magnitude” m used to quantify the visually perceived brightness of stars. For reasons grounded more in tradition than logic, the apparent magnitude is defined as

$$m = -[19 + 2.5 \log I], \quad (14.2.4)$$

where I is the intensity in W/m². The smaller m , the brighter the object. For the sun, the brightest object in the sky, $I \cong 1.4 \text{ kW/m}^2$ at Earth’s surface, and so $m \cong -26.9$. For the full moon, $m = -12.6$, while for Sirius (the brightest star aside from the sun), $m = -1.5$. The faintest stars observable with the naked eye have apparent magnitudes $m \cong 6$, whereas the Hubble space telescope can detect objects as faint as magnitude 30 in the visible. A sodium guide star yielding 70×10^6 photons/m²/s at ground level has an apparent magnitude of about 7.6. •

Quasi-monochromatic irradiation of mesospheric sodium excites only a narrow velocity group within the approximately 3-GHz-wide Doppler absorption profile (Fig. 3.20). By appropriate phase modulation the spectrum of laser radiation can be made to cover essentially the entire Doppler absorption profile. Various pulse-train excitation schemes have been employed, including trains of subnanosecond pulses producing the coherent excitation effects described in Chapter 9.

Saturation of the sodium transitions occurs at sufficiently high laser powers. If, for example, the D₂ line is excited by phase-modulated light whose spectrum is relatively flat over the absorption cross section, then formula (4.11.2) with the D₂ radiative rate $A_{21} = 1/(16 \text{ ns})$ and an “average” cross section $\bar{\sigma} \sim 2 \times 10^{-16} \text{ m}^2$ at 200K (Fig. 3.20) implies $I^{\text{sat}} \sim 5 \text{ W/cm}^2$ for the saturation intensity. Suppose, for example, that such a phase-modulated pulse with a duration long compared to the radiative lifetime saturates the D₂ transition such that the excited-state probability for an atom with coordinates (x, y) transverse to the propagation direction has the quasi-steady-state value

$$\bar{p}(x, y) = \frac{\frac{1}{2}I(x, y)/I^{\text{sat}}}{1 + I(x, y)/I^{\text{sat}}} \quad (14.2.5)$$

⁹J. Telle, J. Drummond, C. Denman, P. Hillman, G. Moore, S. Novotny, and R. Fugate, *Proceedings of the SPIE (International Society for Optical Engineering)* **6215**, 62150K-1 (2006).

during the pulse. Suppose also that the average intensity at the mesosphere may be approximated by the Gaussian form¹⁰

$$I(x, y) = I_0 e^{-2(x^2+y^2)/w^2} = I_0 e^{-2r^2/w^2} \quad (14.2.6)$$

(w is typically ~ 1 m). Then the rate at which photons are backscattered from such a pulse to a receiving aperture of area A_r at ground is

$$\begin{aligned} \mathcal{R} &= \frac{T_0 C_s}{4\pi z^2} A_r \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \frac{A_{21} \times \frac{1}{2} I(x, y) / I^{\text{sat}}}{1 + I(x, y) / I^{\text{sat}}} \\ &= \frac{T_0 C_s}{4\pi z^2} A_r (2\pi) \int_0^{\infty} dr r \frac{A_{21} \times \frac{1}{2} I_0 e^{-2r^2/w^2} / I^{\text{sat}}}{1 + I_0 e^{-2r^2/w^2} / I^{\text{sat}}} \\ &= \frac{T_0 C_s}{4\pi z^2} A_r \frac{\pi w^2}{2} \ln \left(1 + \frac{I_0}{I^{\text{sat}}} \right), \end{aligned} \quad (14.2.7)$$

aside from a possible factor accounting for anisotropy of the backscatter. Such an approximate functional dependence of $\mathcal{R}$ on the peak intensity I_0 provides a good fit to observed photon returns from “long,” phase-modulated laser pulses.⁸

Work on sodium guide star includes research into methods of generating high-power radiation at 589 nm. Chains of dye lasers have been used to generate trains of pulses of sufficient intensity to strongly saturate the mesospheric sodium D₂ line, but these systems are large and complicated and the overall efficiency is low. Sum-frequency (three-wave mixing) generation using Nd:YAG laser transitions at 1.064 and 1.319 μm ($1/1064 + 1/1319 = 1/589$) appears especially promising for laser guide stars,⁹ and fiber lasers capable of yielding sufficiently high powers at 589 nm are also an attractive possibility.

In Section 8.11 we introduced the seeing angle as a measure of angular resolution, and estimated that under good seeing conditions (characterized by a coherence diameter $r_0 = 10$ cm in the visible) the angular resolution of a 10-m telescope is about 1 arcsec compared to its theoretical diffraction-limited value of 0.01 arcsec. In other words, compensation for atmospheric turbulence could ideally result in an improvement in angular resolution by a factor of 100! Aside from far more detailed imaging, such an improvement in angular resolution would have important benefits for ground-based astronomical spectroscopy: It would allow the use of smaller spectroscopic slits and therefore a reduction of background radiation in the spectroscopy of very faint objects.

- Adaptive optics was first proposed in 1953 by the astronomer H. W. Babcock, who envisioned a deformable mirror based on an oil film whose thickness over a mirror surface could be controlled electrostatically. Similar ideas were advanced by V. P. Linnik. These concepts were impossible to implement without the computers and image-monitoring devices that came much later.

The concept of an artificial guide star, like much of adaptive optics, originated in classified military research on satellite surveillance and other applications. One type of guide star or “beacon” used in satellite surveillance is a “glint” of sunlight reflected by a satellite. The use

¹⁰The laser radiation propagating to the mesosphere is of course subject to speckle and other effects of turbulence discussed in Chapter 8.

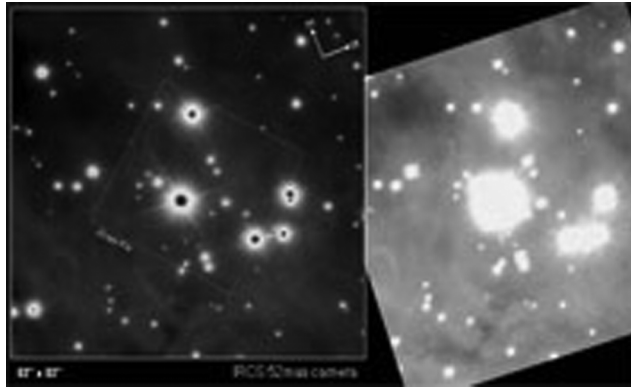


Figure 14.9 One of the early demonstrations of the use of the sodium guide star for adaptive optics on a large telescope (October, 2006). Images of the Trapezium region of the Orion Nebula with (left) and without (right) adaptive optics. (© Subaru Telescope, National Astronomical Observatory of Japan.)

of Rayleigh-scattered light as a beacon for adaptive optics was first suggested in the late 1970s. In 1982 mesospheric sodium was proposed as a beacon for adaptive optics in a classified report;¹¹ in 1985 it was suggested as a guide star for adaptive optics on telescopes,¹² and the first experimental studies in that direction were reported in 1987.¹³ In 1991 most of the military research on adaptive optics and laser guide stars in the United States was declassified. •

Adaptive optics with a sodium guide star is not a perfect solution to the problem of imaging through atmospheric turbulence. For one thing, *focus anisoplanatism*—the fact that the light from the guide star is a spherical wave rather than the plane wave from an astronomical object at “infinity”—implies that the guide star light does not “sample” exactly the same part of the turbulent atmosphere as the light from an object at infinity. And as already mentioned, a natural guide star is still required in order to compensate for the overall phase tilt caused by the atmosphere. But the results thus far have been impressive, and development of adaptive optical telescope systems with the sodium guide star is proceeding rapidly. Figure 14.9 shows an example of the image improvement obtained with the 8.2-m Subaru telescope at Mauna Kea, Hawaii, one of a growing number of very large (6–10 m) ground-based telescopes employing adaptive optics with sodium guide stars. The adaptive optics system used a 188-subaperture Shack–Hartmann wavefront sensor, about 2 m wide, and a 13-cm-diameter deformable mirror with 188 actuators. The 589-nm beam was obtained by sum-frequency generation with 1064- and 1319-nm Nd : YAG laser radiation; the 589-nm radiation was sent by a photonic crystal fiber from the room housing the lasers and the nonlinear optics to the launch telescope. The diffraction-limited angular resolution at the 2.2- μm imaging wavelength is $\sim 2.2 \times 10^{-6} / (8.2) = 0.27 \mu\text{rad} = 0.06 \text{ arcsec}$, compared to a 0.6-arcsec resolution without any adaptive optics. With adaptive optics and the sodium guide star the angular resolution was nearly diffraction-limited.

¹¹W. Happer, G. J. MacDonald, C. E. Max, and F. J. Dyson, *Journal of the Optical Society of America B* **11**, 263 (1994).

¹²R. Foy and A. Labeyrie, *Astronomy and Astrophysics* **152**, L29 (1985).

¹³L. A. Thompson and C. S. Gardner, *Nature* **328**, 229 (1987).

14.3 OPTICAL PUMPING AND SPIN-POLARIZED ATOMS

Aside from occasional allusions to the possibility of using polarized light to preferentially populate particular magnetic substates (cf. Sections 3.7 and 4.11), we have presumed that degenerate states differing only in the magnetic quantum number m are equally populated. In the calculation of the absorption cross section for the sodium D_2 line in Section 3.13, for example, the three magnetic substates of the $3S_{1/2}(F=1)$ level and the five magnetic substates of the $3S_{1/2}(F=2)$ level were assigned equal populations, as must be the case for degenerate states in thermal equilibrium. The process by which a departure from a thermal distribution occurs due to irradiation with light, as in the creation of population inversion in dye or solid-state lasers, for example, is called *optical pumping*. The term often refers specifically to the redistribution of hyperfine levels and magnetic substates by a resonant atom–field interaction, and it is in this sense that we will use it here. Optical pumping as such was well understood by the late 1940s but, as with so many other aspects of spectroscopy, it came into very widespread use primarily because of the quasi-monochromaticity, directionality, tunability, and intensity of light made possible by the laser.

Let us briefly review the physical significance of the magnetic substates. An atom has a magnetic dipole moment proportional to its angular momentum $\mathbf{F}$, and for a level with total angular momentum quantum number F there are $2F + 1$ degenerate substates corresponding to “magnetic” quantum numbers $m = -F, -F + 1, \dots, -1, 0, 1, \dots, F - 1, F$. (Recall the simplified discussion of the hydrogen atom in Section 2.2, or Fig. 3.19 for the hyperfine structure of the sodium D_2 line.) An atom in a state with magnetic quantum number m will be found in a measurement to have a component of angular momentum $m\hbar$ along *any* chosen “quantization axis,” which is usually called the z axis. We are denoting the total angular momentum by $\mathbf{F}$, the standard notation for the sum of the electron orbital angular momentum $\mathbf{L}$, the electron spin angular momentum $\mathbf{S}$, and the nuclear spin angular momentum $\mathbf{I}$. Atoms with an odd isotope number (the sum of the number of protons and the number of neutrons in the nucleus) have a net nuclear spin, hyperfine structure, and a total angular momentum $\mathbf{F} = \mathbf{L} + \mathbf{S} + \mathbf{I}$. For atoms with even isotope numbers the net nuclear spin $\mathbf{I} = 0$ and the total angular momentum is $\mathbf{L} + \mathbf{S}$; in this case $\mathbf{F}$ in the discussion to follow, unless otherwise noted, is actually $\mathbf{L} + \mathbf{S}$ and is conventionally denoted by $\mathbf{J}$.

In the presence of a weak magnetic field $\mathbf{B} = B_z \hat{z}$ the degeneracy of the different magnetic substates is removed by the Zeeman effect: The state F, m is shifted in energy by an amount proportional to $\mathbf{F} \cdot \mathbf{B} = mB_z$:

$$\Delta E_{F,m} = \mu_B g_F m B_z, \quad (14.3.1)$$

where $\mu_B = eh/2m_e = 9.274 \times 10^{-24}$ joule/tesla is the *Bohr magneton* (e and m_e are the electron charge and mass, respectively) and g_F is a so-called Landé g factor. For example, the Zeeman shift of a substate m of a hyperfine level with total angular momentum quantum number F (Section 3.13) is given approximately by (14.3.1) with

$$\begin{aligned} g_F &\cong g_J \frac{F(F+1) + J(J+1) - I(I+1)}{2F(F+1)}, \\ g_J &\cong 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \end{aligned} \quad (14.3.2)$$

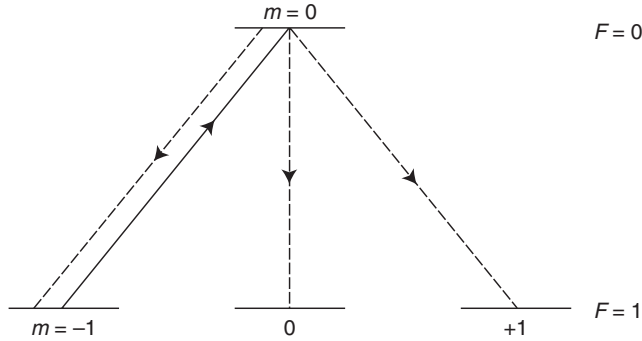


Figure 14.10 Optical pumping with σ_+ light of an atomic transition with lower- and upper-level angular momentum quantum numbers 1 and 0, respectively. The σ_+ light induces a transition indicated by the solid arrow, whereas spontaneous emission can occur on the transitions shown by the dashed arrows.

For sodium, $I = \frac{3}{2}$, $S = \frac{1}{2}$, $g_{J=3/2}(3P_{3/2}) = \frac{4}{3}$, and $g_{J=1/2}(3S_{1/2}) = 2$. Then, for example, $g_{F=1}(3S_{1/2}) = -\frac{1}{2}$ and $g_{F=1}(3P_{3/2}) = \frac{2}{3}$.

Now the fact that light carries intrinsic angular momentum implies that there can be a change Δm in the magnetic quantum number when light is absorbed or emitted. Circularly polarized photons carry angular momentum ± 1 (in units of $\hbar$) directed along the propagation direction $\hat{z}$, and the *selection rule* for allowed transitions is $\Delta m = \pm 1$ when the quantization axis is taken to be the $\hat{z}$ direction. For linearly polarized photons the selection rule is $\Delta m = 0$ with the quantization axis along the polarization direction, as discussed below. Based on these selection rules, it is easy to see how polarized light can be used to preferentially populate a particular magnetic substate, or in other words to “align” (or “spin-polarize”) an atom so that along a particular direction its magnetic dipole moment has only one possible value. We will illustrate this with a few examples.

The solid arrow in Fig. 14.10 shows the $\Delta m = +1$ transition allowed when light with σ_+ circular polarization is resonant with an atomic transition having a lower level with three degenerate magnetic substates ($m = -1, 0, 1$) and an upper level with a single magnetic substate ($m = 0$).¹⁴ The only allowed absorptive transition that can be induced by σ_+ light is between the lower state with $m = -1$ and the upper state with $m = 0$.¹⁵ Spontaneous emission from the upper states, however, is constrained only by the selection rule $\Delta m = 0, \pm 1$, since we cannot associate any special direction or quantization axis with it. The upper state $m = 0$ can therefore decay spontaneously into any of the three lower states in Fig. 14.10. When a spontaneous transition occurs from the upper state to either the $m = 0$ or $m = 1$ lower state, the atom remains in that lower state

¹⁴ σ_+ and σ_- circular polarizations are defined with respect to the *transitions* they can produce rather than with respect to left- or right-hand circular polarization of the light itself. Thus, for example, σ_+ light propagating in the $+z$ direction (the direction of the quantization axis) is said in optics to be left-hand circularly polarized, as is σ_- light propagating in the $-z$ direction. $\sigma_{\pm}$ photons have angular momentum $\pm \hbar$ along their direction of propagation and therefore can induce $\Delta m = \pm 1$ absorptive transitions. It is best in this context to avoid referring to “left” or “right” circular polarizations.

¹⁵Stimulated emission from the upper $m = 0$ state to the lower $m = -1$ state is also allowed, but it does not change the fact that the atom ends up eventually with zero population in the $m = -1$ state.

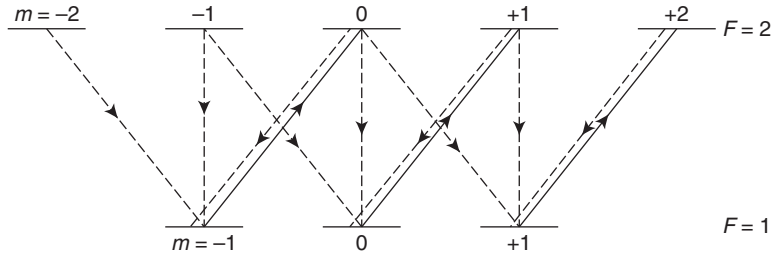


Figure 14.11 Optical pumping with σ_+ light of an atomic transition with lower- and upper-level angular momentum quantum numbers 1 and 2, respectively. σ_+ light induces transitions shown by solid arrows, whereas spontaneous emission occurs on transitions indicated by dashed arrows.

because there are no allowed transitions from it that can be induced by the applied σ_+ light. It is obvious then that, after a sufficient number of absorption and emission transitions, the atom will end up with zero probability of being in the $m = -1$ state, and therefore will cease to absorb any σ_+ light. In fact it is easy to see that any transition with a lower level having a greater (or equal) number of magnetic substates than the upper level will eventually cease to absorb σ_+ light; the same conclusion holds for applied σ_- light.

The same sort of thing happens when resonant linearly polarized light (labeled by π instead of σ) is incident on the atom with the upper and lower levels of Fig. 14.10. Because of the $\Delta m = 0$ selection rule for induced transitions in this case, while again $\Delta m = 0$, ± 1 for spontaneous emission transitions, the atom after a few induced and spontaneous transitions will have zero probability of being in the $m = 0$ lower state and will not absorb the applied π light. The conclusion that π light will cease to absorb is also reached (of course!) if we choose to take the direction of light propagation rather than the direction of polarization as the quantization axis and regard the π -polarized light as a superposition of σ_+ and σ_- light [cf. (14.3.4)]. In terms of this quantization axis, with the magnetic quantum number denoted m' , absorptive transitions occur only from the $m' = \pm 1$ lower states, and after a spontaneous transition from the upper state to the $m' = 0$ lower state the atom can no longer absorb.¹⁶

Consider next an example in which the lower level of a transition has a smaller number of magnetic substates than the upper level. Figure 14.11 shows allowed induced and spontaneous transitions when resonant σ_+ light is incident on an atom with lower and upper levels having three and five degenerate magnetic substates, respectively. It is clear that in this case the atom will eventually have zero probability of being in any state other than the lower $m = 1$ state or the upper $m = 2$ state. After it is fully “pumped” by σ_+ light it can only make spontaneous and stimulated transitions between these two states—it is a *two-state atom*. The circularly polarized light has “spin-polarized” the atom in the sense that its component of angular momentum along the direction $\hat{z}$ of field propagation (the quantization axis) is either $\hbar$ (when it is in the lower, $m = 1$ state) or $2\hbar$ (when it is in the upper, $m = 2$ state); if the σ_+ light is shut

¹⁶The m' states can be expressed as a linear combination of the m states, and vice versa. In general magnetic substates corresponding to two different quantization axes making an angle θ are related by rotation matrices $d_{mm'}^{(F)}(\theta)$. See, for instance, A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, Princeton University Press, Princeton, NJ, 1996.

off, the atom will stay in the lower $m = 1$ state after making a spontaneous transition to that state. With σ_- light, similarly, a two-state atom is created with lower and upper states $m = -1$ and $m = -2$, respectively.

- We will derive the *general* selection rule $\Delta m = 0, \pm 1$ for electric dipole transitions using the example of a one-electron, spinless atom and the fact that a transition from a state 1 of energy E_1 to a state 2 of energy $E_2 > E_1$ is governed by the matrix element [see Eq. (9.3.16) and Problem 14.6]

$$\mathbf{x}_{21} \cdot \hat{\mathbf{e}} = \int \phi_2^*(\mathbf{x})(\mathbf{x} \cdot \hat{\mathbf{e}})\phi_1(\mathbf{x}) d^3x. \quad (14.3.3)$$

The polarization unit vector $\hat{\mathbf{e}}$ will in general have components in all three directions $\hat{x}$, $\hat{y}$, and $\hat{z}$ defined by some Cartesian coordinate system. Consider polarization in the xy plane. It is convenient in this case to combine the x and y components and define σ_+ and σ_- circularly polarized light with complex unit polarization vectors (Problem 3.4)

$$\hat{\mathbf{e}}_{\pm} = \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y}), \quad (14.3.4)$$

where the orthogonal unit vectors $\hat{x}$ and $\hat{y}$ are perpendicular to the light propagation direction $\hat{z}$. To evaluate (14.3.3) we express the hydrogen wave functions in terms of spherical coordinates (r, θ, ϕ) defined with respect to the same (x, y, z) coordinate system, i.e., $x = r \sin \theta \cos \phi$, $y = r \sin \theta \sin \phi$, and $z = r \cos \theta$. Now all we need to know to derive the selection rule for the magnetic quantum number is that, for a state of our hydrogen atom with principal, orbital angular momentum, and magnetic quantum numbers n, ℓ, m , respectively, the dependence of the wave function $\phi_{n\ell m}(r, \theta, \phi)$ on the azimuthal angle ϕ is described entirely by a factor $\exp(im\phi)$. Thus, for an electric dipole transition from a state with quantum numbers n, ℓ, m to a state with quantum numbers n', ℓ', m' the matrix element (14.3.3) is proportional to

$$\int_0^{2\pi} e^{-im'\phi} e^{\pm i\phi} e^{im\phi} d\phi = \int_0^{2\pi} e^{-i(m'-m\mp 1)\phi} d\phi, \quad (14.3.5)$$

since $\mathbf{x} \cdot \hat{\mathbf{e}}_{\pm} = (x\hat{x} + y\hat{y} + z\hat{z}) \cdot \hat{\mathbf{e}}_{\pm} = (1/\sqrt{2})(x \pm iy) = (1/\sqrt{2})(r \sin \theta \cos \phi \pm ir \sin \theta \sin \phi) = (1/\sqrt{2})r \sin \theta \exp(\pm i\phi)$. m and m' have only integer values, and consequently the integral (14.3.5) is zero unless the exponent in the integrand is zero, i.e., unless the selection rule $m' = m + 1$ (for σ_+ light) or $m' = m - 1$ (for σ_- light) is satisfied.

The matrix element for the stimulated emission transition from the excited state 2 with magnetic quantum number m' to the lower energy state 1 with magnetic quantum number m is just the complex conjugate of (14.3.3) (Problem 14.6). In this case the same selection rule applies ($\Delta m = \pm 1$) but the magnetic quantum number *decreases* when the transition is induced by σ_+ light and *increases* when the transition is induced by σ_- light.

For light polarization parallel to the z axis we have $\mathbf{x} \cdot \hat{\mathbf{e}} = \mathbf{x} \cdot \hat{z} = z = r \cos \theta$, and the matrix element (14.3.3) is proportional to

$$\int_0^{2\pi} e^{-im'\phi} e^{im\phi} d\phi = \int_0^{2\pi} e^{-i(m'-m)\phi} d\phi, \quad (14.3.6)$$

which vanishes unless $m = m'$. In other words, for linearly polarized light we have the selection rule $\Delta m = 0$ for allowed transitions. In this case the quantization axis (z) has been chosen to be along the direction of field polarization, whereas for circular polarization we took it to be the direction of field propagation.

Selection rules for the total angular momentum quantum number are derived in textbooks on quantum mechanics or atomic physics: $\Delta F = 0, \pm 1$, but $F = 0 \leftrightarrow F' = 0$ transitions are forbidden, that is, their electric dipole transition matrix elements are zero. If F is the total angular momentum quantum number in the case of nonzero nuclear spin ($\mathbf{F} = \mathbf{J} + \mathbf{I}$, the sum of electronic and nuclear angular momenta), we also have the selection rules $\Delta J = 0, \pm 1$ ($J = 0 \leftrightarrow J' = 0$ forbidden) and $\Delta m_J = 0, \pm 1$.

There are also selection rules for magnetic dipole and electric quadrupole transitions, which have much smaller transition rates than electric dipole transitions (typically about 10^5 and 10^8 times smaller, respectively). The selection rules for magnetic dipole transitions are $\Delta F = 0, \pm 1$ and $\Delta m = 0, \pm 1$, whereas for electric quadrupole transitions $\Delta F = 0, \pm 1, \pm 2$ and $\Delta m = 0, \pm 1, \pm 2$. In all cases $F = 0 \leftrightarrow F' = 0$ transitions are forbidden. •

Let us return to the example of the sodium D_2 line. In Section 3.13 we calculated the absorption cross section shown in Fig. 3.20 under the assumption that the eight magnetic substates of the two $3S_{1/2}$ hyperfine levels were equally populated, as is approximately the case in thermal equilibrium at temperatures for which $k_B T$ is large compared to the 1.77 GHz $3S_{1/2}(F = 1) \leftrightarrow 3S_{1/2}(F = 2)$ hyperfine splitting. Suppose we irradiate the sodium atom with 589-nm σ_+ light that can induce transitions out of both the $3S_{1/2}$ hyperfine levels. From Fig. 3.19 and the selection rules $\Delta F = 0, \pm 1$ and $\Delta m = 1$ it can be seen that, for irradiation times long compared to the 16-ns radiative lifetime of the excited states, the sodium atom will be a two-state atom: only the $3S_{1/2}(F = 2, m = 2)$ and $3P_{3/2}(F = 3, m = 3)$ states will have nonzero occupation probabilities. The optical pumping to this two-state system occurs in ~ 20 excitation and decay transitions, after which the absorption cross section will have only a single peak instead of the two appearing in Fig. 3.20. As discussed below, this results in stronger absorption than in the case of the unpolarized D_2 line. If the σ_+ light is shut off after the two-state atom has been realized, the sodium atom after spontaneous emission from the $3P_{3/2}(F = 3, m = 3)$ state will be in the single, spin-polarized state $3S_{1/2}(F = 2, m = 2)$. In this state the electron with spin z -component m_S and the nucleus with spin z -component m_I have their spins aligned: $m = m_F = m_L + m_S + m_I = 0 + \frac{1}{2} + \frac{3}{2} = 2$.

Suppose instead that we irradiate the sodium atoms with narrowband, linearly polarized (π) radiation that can only induce transitions from the $3S_{1/2}(F = 2)$ level. Because the atoms excited to the $3P_{3/2}(F = 1, 2, 3)$ levels can undergo spontaneous transitions to $3S_{1/2}(F = 1, m = \pm 1)$, the sodium atoms in this case will become *transparent* to the π light after several excitation and decay transitions.

Absorption can increase or decrease in strength, therefore, when the incident light causes different, degenerate magnetic substates to have different populations. Such optical pumping occurs in the absence of collisions and magnetic fields that tend to redistribute the m -state populations. In collisions of sodium atoms, for example, the valence electrons can be exchanged; these *spin-exchange* collisions can quickly “thermalize” the populations of the magnetic substates. Spin-relaxation collisions can have large collision cross sections ($\sim 10^{-14}$ cm² for alkali atoms) and can prevent or rapidly destroy spin polarization. In a dilute vapor with relatively infrequent atom–atom collisions, atom–wall collisions can likewise prevent spin polarization.

Spin relaxation rates are greatly reduced in experiments with atomic beams, where collisions are effectively avoided. They are also substantially reduced by the use of “buffer” gases, especially inert gases, that cause very little spin relaxation in collisions with polarized atoms. Buffer gases at pressures of typically a few Torr are used to slow down the diffusion of polarized atoms to the cell walls and to maintain their polarization

for times on the order of seconds or longer in the case of alkali atoms. Coating the cell walls with paraffins and other materials that do not cause significant spin relaxation is another frequently used method for preserving spin polarization.

Magnetic fields, including Earth's magnetic field, can also induce transitions between different magnetic substates F, m and F, m' . Such *static* $\mathbf{B}$ fields cannot induce transitions between states of different energy, and, in particular, cannot cause transitions between states of different total angular momentum quantum number F . The frequency with which a static $\mathbf{B}$ field induces oscillations between different m states is on the order of $\mu_B B / 2\pi\hbar$, as shown below. For Earth's magnetic field, for example, $B \approx 0.5 \text{ G}$ (gauss) $= 0.5 \times 10^{-4} \text{ T}$ (tesla), and so the time scale for spin depolarization due to the geomagnetic field is typically on the order of a microsecond. Optical pumping experiments often employ a magnetic field applied along the direction of spin polarization; since the energy of the magnetic dipole moment $\mathbf{M}$ of the atom in a magnetic field $\mathbf{B}$ is $-\mathbf{M} \cdot \mathbf{B}$, the field serves to maintain the spin polarization in the presence of any stray, weak magnetic fields $\mathbf{B}_s$ for which $\mathbf{M} \times \mathbf{B}_s \neq 0$.

- For an example of how a magnetic field can affect optically pumped atoms, consider the experiment sketched in Fig. 14.12. Light propagating in the y direction and linearly polarized along x is resonant with an atomic transition that we assume for simplicity has a lower level with $F = 0$ and an upper level with $F = 1$. There is also a static $\mathbf{B}$ field along the z direction, which we will take to be the quantization axis.

For the incident electric field we write

$$\begin{aligned} \mathbf{E}(y, t) &= \hat{x}E_0 e^{-i(\omega t - ky)} = \frac{1}{2}[(\hat{x} + i\hat{y}) + (\hat{x} - i\hat{y})]E_0 e^{-i(\omega t - ky)} \\ &= \frac{1}{\sqrt{2}}[\hat{\mathbf{e}}_+ + \hat{\mathbf{e}}_-]E_0 e^{-i(\omega t - ky)}, \end{aligned} \quad (14.3.7)$$

where as always it is implicit that we are to take the real part. The $\hat{\mathbf{e}}_+$ and $\hat{\mathbf{e}}_-$ components result in nonvanishing probability amplitudes for the $F = 1, m = 1$ and $F = 1, m = -1$ atomic states, respectively. We denote the stationary-state wave functions (eigenfunctions) of these two states by $\Phi_1(\mathbf{x})$ and $\Phi_{-1}(\mathbf{x})$, respectively, and similarly let $\Phi_0(\mathbf{x})$ denote the wave function for the

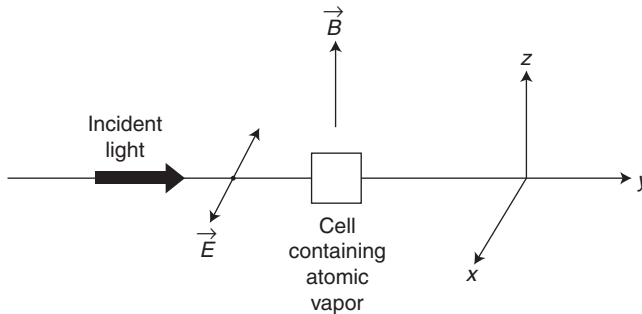


Figure 14.12 Light polarized along x propagates in the y direction and excites atoms in a dilute vapor. In the absence of a magnetic field the atoms do not radiate any light in the x direction. With a magnetic field $\mathbf{B}$ in the z direction, however, resonance fluorescence polarized along y is observed in the x direction. This is the Hanle effect.

$F = 0, m = 0$ lower state of the (one-electron) atom. Then for the wave function at time t for an atom in the field (14.3.7) we write

$$\psi(\mathbf{x}, t) = a_0(t)\Phi_0(\mathbf{x}) + a_1(t)\Phi_1(\mathbf{x}) + a_{-1}(t)\Phi_{-1}(\mathbf{x}). \quad (14.3.8)$$

Let us calculate the expectation value $\mathbf{p}(t)$ of the electric dipole moment of an atom described by this wave function:

$$\mathbf{p}(t) = e \int d^3x \psi^*(\mathbf{x}, t) \mathbf{x} \psi(\mathbf{x}, t) = e\mathbf{x}_{0,1}a_0^*(t)a_1(t) + e\mathbf{x}_{0,-1}a_0^*(t)a_{-1}(t) + \text{c.c.}, \quad (14.3.9)$$

since the only nonvanishing dipole matrix elements are between the state $F = 0, m = 0$ and the states $F = 1, m = \pm 1$. In terms of spherical coordinates r, θ, ϕ , $\Phi_0(\mathbf{x}) = \Phi_0(r, \theta)$ and $\Phi_{\pm 1}(\mathbf{x}) = \Phi(r, \theta)e^{\pm i\phi}$, so that for the x, y , and z components of $\mathbf{x}_{0,\pm 1}$ we obtain

$$\begin{aligned} x_{0,\pm 1} &= \int d^3x \Phi_0^*(r, \theta) x \Phi(r, \theta) e^{\pm i\phi} \\ &= \int_0^\infty dr r \int_0^\pi d\theta \sin \theta \Phi_0^*(r, \theta) r \sin \theta \Phi(r, \theta) \int_0^{2\pi} d\phi \cos \phi e^{\pm i\phi} \\ &\equiv \mu \int_0^{2\pi} d\phi \cos \phi e^{\pm i\phi} = \pi\mu, \end{aligned} \quad (14.3.10a)$$

$$y_{0,\pm 1} = \mu \int_0^{2\pi} d\phi \sin \phi e^{\pm i\phi} = \pm i\pi\mu, \quad (14.3.10b)$$

$$z_{0,\pm 1} = \int_0^\infty dr r \int_0^\pi d\theta \sin \theta \Phi_0^*(r, \theta) r \cos \theta \Phi(r, \theta) \int_0^{2\pi} d\phi e^{\pm i\phi} = 0, \quad (14.3.10c)$$

and therefore $\mathbf{x}_{0,\pm 1} = x_{0,\pm 1}\hat{x} + y_{0,\pm 1}\hat{y} + z_{0,\pm 1}\hat{z} = \pi\mu(\hat{x} \pm i\hat{y})$ and

$$\mathbf{p}(t) = \pi e\mu[a_0^*(t)a_1(t)(\hat{x} + i\hat{y}) + a_0^*(t)a_{-1}(t)(\hat{x} - i\hat{y})] + \text{c.c.} \quad (14.3.11)$$

We will assume now that the light incident on an atom in Fig. 14.12 is in the form of a short pulse that excites the atom at time $t = t_0$, after which the state probability amplitudes evolve according to the time-dependent Schrödinger equation [cf. (9.2.2)]

$$\dot{a}_0 = 0, \quad (14.3.12a)$$

$$\dot{a}_1 = -i(\omega_0 + \omega_L)a_1, \quad (14.3.12b)$$

$$\dot{a}_{-1} = -i(\omega_0 - \omega_L)a_{-1}, \quad (14.3.12c)$$

where $\hbar\omega_L = \mu_B g_1 B$ and $-\hbar\omega_L$ are the Zeeman shifts of the $F = 1, m = 1$ and $F = 1, m = -1$ states, respectively [Eq. (14.3.1)] and ω_0 is the $F = 0 \leftrightarrow 1$ transition frequency of the unperturbed atom. Then

$$a_0(t) = c_0, \quad (14.3.13a)$$

$$a_1(t) = c_1 e^{-i(\omega_0 + \omega_L)(t-t_0)}, \quad (14.3.13b)$$

$$a_{-1}(t) = c_{-1} e^{-i(\omega_0 - \omega_L)(t-t_0)}, \quad (14.3.13c)$$

where c_0, c_1 , and c_{-1} are the probability amplitudes at $t = t_0$. Since the linearly polarized pulse is a superposition of $\hat{\mathbf{e}}_+$ and $\hat{\mathbf{e}}_-$ components of equal amplitude, we can assume that

$c_0^*c_1 = c_0^*c_{-1} \equiv \rho$; we also assume that the phases of the probability amplitudes are chosen to make ρ real. Then it follows from (14.3.11) and some simple algebra that

$$\mathbf{p}(t) = 4\pi e\mu\rho \cos \omega_0(t - t_0) \{ \hat{x} \cos[\omega_L(t - t_0)] + \hat{y} \sin[\omega_L(t - t_0)] \}, \quad (14.3.14)$$

where $\omega_L = \mu_B g_1 B / \hbar$ is the *Larmor frequency*.

This result suggests—correctly—that when $\mathbf{B} = 0$ ($\omega_L = 0$) the resonance fluorescence from the atoms in the experiment of Fig. 14.12 is always linearly polarized along the same axis (x) as the polarization of the incident field: The atomic dipole oscillates along the x axis and therefore does not radiate in the $\pm x$ directions. When $\mathbf{B} \neq 0$, however, the dipole has a component along y , produces resonance fluorescence polarized along y , and therefore *does* radiate in the $\pm x$ directions. This effect of a magnetic field on resonance fluorescence is called the *Hanle effect*.

The Hanle effect is easily explained with the classical electron oscillator model (Chapter 3) when we include the force $e\mathbf{v} \times \mathbf{B}$ in the equation of motion for the electron displacement $\mathbf{x}$. In our example the electric field polarized along x causes the electron to oscillate with a component of velocity $\mathbf{v}$ along x , so that a magnetic field along z results in a force $e\mathbf{v} \times \mathbf{B}$ along y . The oscillating electric dipole moment of the atom therefore acquires a y component that results in y -polarized radiation in the $\pm x$ directions.

We have not accounted for the damping of the expectation value $\mathbf{p}(t)$ of the atomic dipole moment as a result of the spontaneous emission from the excited states. To do so requires only that we replace the density matrix element ρ in (14.3.14) by $\rho \exp(-\gamma t/2)$, where γ is the spontaneous emission rate, which is the same for the degenerate magnetic substates of a given energy level:

$$\mathbf{p}(t) = 4\pi e\mu\rho \cos \omega_0(t - t_0) e^{-\gamma(t-t_0)/2} \{ \hat{x} \cos[\omega_L(t - t_0)] + \hat{y} \sin[\omega_L(t - t_0)] \}. \quad (14.3.15)$$

If a polarizer (or “analyzer”) on the x axis is oriented to fully transmit radiation polarized along a direction $\hat{a}$ making an angle φ with respect to $\hat{z}$, the time-averaged intensity $I(t)$ of the resonance fluorescence measured by a detector behind the polarizer will be proportional to the time average of $[\mathbf{p}(t) \cdot \hat{a}]^2$, or (Problem 14.7)

$$I(t) \propto e^{-\gamma(t-t_0)} \cos^2\{\omega_L(t - t_0) - \varphi\}. \quad (14.3.16)$$

This is the basis of a useful technique for measuring excited-state lifetimes ($1/\gamma$). In fact “the Hanle effect has been developed into one of the most reliable methods for measuring the lifetimes of excited levels of atoms and molecules.”¹⁷ This is explored further in Problem 14.7.

Two points about our simplified approach to the Hanle effect are noteworthy. First, $\rho \neq 0$ means that the Hanle effect involves off-diagonal coherence of the atomic density matrix (Chapter 9); this explains why it attracted interest during the development of quantum theory in the 1920s. Second, as suggested by the classical oscillator model, expressions like (14.3.15) for the Hanle effect, with relatively small modifications, describe more general cases than the simplest one we have considered of a transition between levels with angular momentum quantum numbers 0 and 1. •

We have considered specifically some examples involving polarized light, but in fact a redistribution of magnetic substates occurs also for *unpolarized* light, provided it is anisotropic (e.g., unidirectional). As long as the light is not isotropic the spherical symmetry of an unperturbed atom is “broken” and different magnetic substates defined with respect to some quantization axis interact differently with different field polarization

¹⁷A. Corney, *Atomic and Laser Spectroscopy*, Oxford University Press, Oxford, 2006, p. 478. Chapter 15 of this book is an extensive treatment of the Hanle effect.

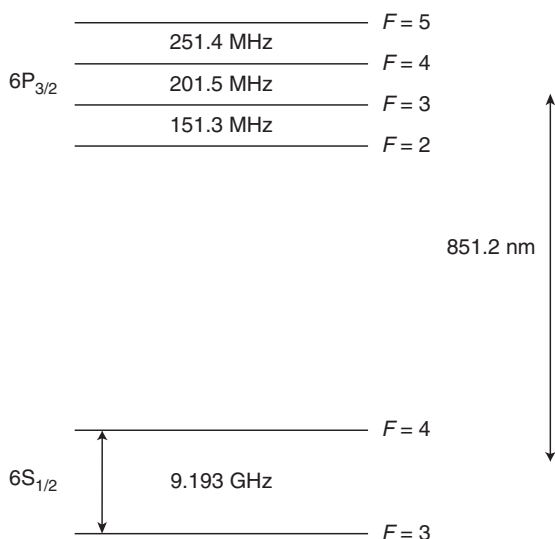


Figure 14.13 Simplified energy-level diagram of ^{133}Cs , showing the hyperfine *clock transition* at 9.193 GHz. The nuclear and electronic angular momenta ($I = \frac{7}{2}$ and $J = \frac{1}{2}$, respectively) result in a splitting of the $6S_{1/2}$ ground level into two hyperfine levels with $F = 3$ ($= \frac{7}{2} - \frac{1}{2}$) and $F = 4$ ($= \frac{7}{2} + \frac{1}{2}$).

components, regardless of whether the field has a definite polarization. Polarized emission lines from the solar corona, for example, are observed and explained as a consequence of the interplay of the sun's magnetic field and optical pumping with the (unpolarized) radiation from the photosphere.¹⁸

Optical pumping has been a very useful tool in basic atomic physics, especially in experimental studies of hyperfine structure and nuclear magnetic moments, atomic collisions, and other aspects of spectroscopy. In the remainder of this section we briefly discuss an application of optical pumping in atomic frequency standards. The following section describes another important application.

- The idea that ground- and excited-state populations of atomic energy levels can deviate substantially from a thermal distribution as a result of resonant atom–field interactions, and in particular that atoms can become spin polarized, is attributed to Alfred Kastler, who was awarded the 1966 Nobel Prize in Physics for his research on optical pumping. Kastler's work included the prediction and observation of the polarization of the “twilight glow” resulting from the excitation of the mesospheric sodium D lines by sunlight. The earliest observations of spin polarization of atomic ground levels by optical pumping were reported by J. Brossel, Kastler, and J. Winter in 1952 and by W. B. Hawkins and R. H. Dicke in 1953. •

Since 1967 the second in the International System of Units (SI) has been defined as “the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom” (Fig. 14.13). Time and frequency standards have evolved over many years and are an

¹⁸J. Trujillo Bueno, E. Landi Degl’Innocenti, M. Collados, L. Merenda, and R. Manso Sainz, *Nature* **415**, 403 (2002).

integral part of such applications as satellite communications and especially the U.S. Global Positioning System (GPS), which would not exist without them; every GPS satellite is equipped with an atomic clock. They are also required for basic scientific experiments aimed at determining whether various fundamental “constants” actually vary in time. The details of the design and implementation of these systems are complex, but the basic operating principles are relatively simple. In the most widely used type of atomic clock the frequency of radiation near an atomic absorption line is varied and controlled to lock it to the peak absorption frequency ν_0 . Then a number $\mathcal{N}$ of oscillation periods $1/\nu_0$ of the radiation is counted to determine a “standard” time interval; thus the time taken to count $\mathcal{N} = 9,192,631,770$ cycles of radiation at the $6S_{1/2}(F=3) \rightarrow 6S_{1/2}(F=4)$ transition of cesium is, by definition, a second. Until recently it has been necessary in atomic clocks to employ microwave frequencies, which are small enough to allow accurate electronic counting of cycles (Section 14.7).

Cesium has been used in atomic clocks since the 1950s, and hundreds of commercial time and frequency standards based on its “clock transition” are currently in operation. Like all atomic transition frequencies, the cesium clock frequency is fundamentally the same everywhere aside from generally calculable shifts due to electric, magnetic, and gravitational fields. Cesium, while hardly the only atom used for atomic clocks, has a vapor pressure that allows relatively intense atomic beams to be produced, a large mass resulting in small thermal velocities and Doppler shifts and, in common with all alkali atoms, only two lower hyperfine levels (Fig. 14.13). Other advantages of cesium include its relatively large clock transition frequency, so that a large Q factor $\nu_0/\Delta\nu_0$ is obtained for a given linewidth $\Delta\nu_0$, and the fact that this transition is only weakly affected by small electric fields that may be present.

In a type of atomic clock that served as the primary frequency standard from the late 1960s until about 1990, a beam of cesium atoms, all in the (approximately equally populated) $F=3$ and $F=4$ hyperfine levels, pass through a strong and spatially inhomogeneous “Stern–Gerlach” magnetic field in region A that deflects atoms in different hyperfine states (and therefore with different magnetic moments) by different amounts (Fig. 14.14) and in one of two directions determined by the “polarity” (the sign of the magnetic quantum number m) with respect to the field.¹⁹ A second inhomogeneous magnetic field in region B is designed to deflect atoms further, such that, absent anything else, no atoms would be “focused” onto a hot-wire atom detector; such a detector causes ionization of atoms incident upon it and thereby an electric signal proportional to the number of incident atoms. However, if an atom undergoes a transition between the $F=3$ and $F=4$ levels in a region C between A and B, the deflection caused by magnet B reverses that caused by A, so that *those atoms that have undergone a transition* are focused onto the atom detector. In other words, magnet A selects atoms in certain magnetic substates, while magnet B is used for the detection of atoms that have made a transition. Transitions are effected in region C by a 3.26-cm microwave field that is resonant with the $F=3, m=0 \leftrightarrow F=4, m=0$ transition in the presence of a weak, (nearly) uniform magnetic field (the “C field”) that Zeeman splits the different magnetic substates; the $m=0$ substates are chosen because they have no (first-order) Zeeman shifts [Eq. (14.3.1)]. Region C is surrounded by a high-permeability material in order to shield it from Earth’s magnetic field. The C field, perpendicular to the atomic

¹⁹Because the atoms have a distribution of velocities, it is generally not possible by this method to put only atoms in one particular magnetic substate in region C of Fig. 14.14.

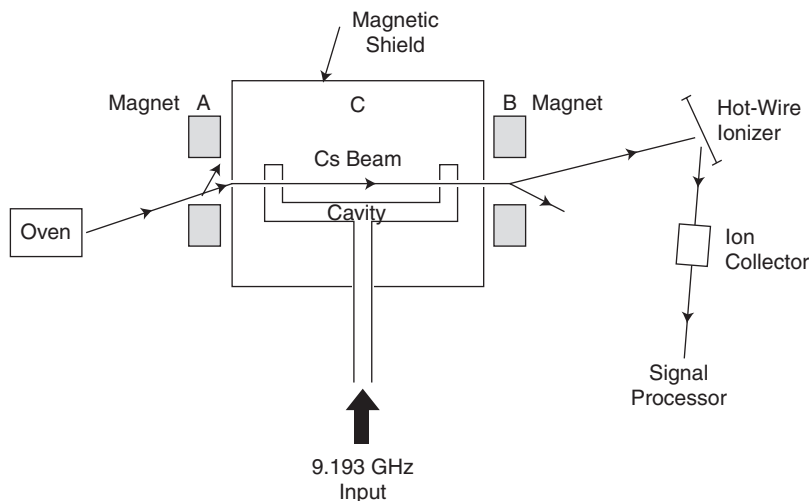


Figure 14.14 Schematic of a cesium atomic clock. The atomic beam is split by a state-selecting magnetic field A into two beams. Atoms selected to enter region C are irradiated with microwave radiation resonant with the clock transition indicated in Fig. 14.13. Those that have undergone the clock transition are focused by the magnetic field B onto an atom detector, the response of which is maximized when the microwave frequency is matched to the clock transition. Feedback circuitry keeps the microwave frequency locked to this value.

beam, acts to Zeeman shift $m \neq 0$ states in the presence of residual stray magnetic fields away from the $F = 3, m = 0 \leftrightarrow F = 4, m = 0$ resonance, which has only a very small, second-order ($\propto B^2$) Zeeman shift.

The narrow linewidth of the clock transition results in a sharp resonance frequency ν_0 , the frequency at which the applied microwave field produces the largest signal from the atom detector. A feedback loop keeps the field locked to this frequency, and the field cycles provide the periodic “clicks” used to keep time. The locked microwave frequency is downconverted electronically and used in a servo loop to lock a quartz oscillator at a frequency of 5 MHz; this is used to generate a signal consisting of one pulse per second, the “output” of the atomic clock.

In this scheme most of the atoms in the atomic beam—about $\frac{15}{16}$ of them if only the $6S_{1/2}(F = 4, m = 0)$ state, say, were selected—are rejected, resulting in a smaller signal-to-noise ratio than would otherwise be possible. State *preparation* by optical pumping, however, has increased the accuracy of atomic frequency standards. For example, all the atoms entering region C can be prepared in the $6S_{1/2}(F = 3)$ level by irradiating them with laser radiation tuned to the $6S_{1/2}(F = 4) \leftrightarrow 6P_{3/2}(F = 3)$ transition, the upper state of which has a radiative lifetime of about 30 ns. Spontaneous transitions occur to both $6S_{1/2}(F = 4)$ and $6S_{1/2}(F = 3)$, and transitions out of the latter level do not occur because they are too detuned from the laser radiation; therefore all the atoms are pumped into that level. A second laser can be used to optically pump all the atoms into the $6S_{1/2}(F = 3, m = 0)$ state, thus preparing them for the $6S_{1/2}(F = 3, m = 0) \leftrightarrow 6S_{1/2}(F = 4, m = 0)$ clock transition in region C.

Optical pumping can also be used for state *detection*. For example, a laser tuned to the $6S_{1/2}(F = 4) \leftrightarrow 6P_{3/2}(F = 5)$ transition results in allowed spontaneous transitions from $6P_{3/2}(F = 5)$ to $6S_{1/2}(F = 4)$, so that the detection of photons radiated in these

transitions indicates that atoms in region C have undergone the clock transition. In other words, in optical pumping schemes the state selection magnets A and B are replaced by lasers, the atom detector is replaced by a photodetector, and signal-to-noise ratios are increased because atoms are *prepared* in desired states rather than just selected from an ensemble of atoms distributed over many states. Optical pumping is employed in the cesium frequency standard NIST-7 of the U.S. National Institute of Standards and Technology (NIST). This served as the primary frequency standard in the United States from 1993 to 1999 but has been replaced in this role by the NIST-F1 standard employing laser cooling (Section 14.4).

The accuracy $\Delta\nu_0$ with which the cesium resonance frequency can be determined is limited by the transit time t_d over which the atoms interact with the microwave field; the fractional width $\Delta\nu_0/\nu_0$ is inversely proportional to the “interrogation time” t_d , i.e., $\Delta\nu_0/\nu_0 \propto 1/\nu_0 t_d$ (Section 9.11). It is therefore advantageous to make t_d as large as possible. Cesium, in addition to the fact that it has the largest vapor pressure and the largest ground-level hyperfine splitting of any alkali, has the advantage of a large mass and therefore small thermal velocities (typically ~ 250 m/s). Relatively long interrogation times (~ 0.004 s for a path of length $L \sim 1$ m) are therefore obtained with cesium. However, it is impractical to produce magnetic fields that are sufficiently homogeneous over such lengths; small field inhomogeneities give rise to line-broadening effects. For this reason atomic frequency standards, which require extremely narrow resonance lines, employ a U-shaped microwave cavity (Fig. 14.14) that takes advantage of the Ramsey method of separated oscillatory fields. As discussed in Section 9.11, this avoids line broadening due to field inhomogeneities while allowing large transit times and the very sharp resonance of the central Ramsey fringe.

Inaccuracies $\Delta\nu_0/\nu_0$ as small as about 5×10^{-15} have characterized cesium frequency standards of the type just described. This corresponds to a clock accuracy of about 1 s in 6 million years. The atomic fountain clocks described in the following section are more accurate by about an order of magnitude. These measurements of the cesium clock frequency are probably the most accurate measurements ever made of any physical quantity.

The cesium atomic clock is *passive*, as opposed to *active* frequency standards based on masers. The hydrogen maser frequency standard operating on the 21-cm hyperfine transition $1S_{1/2}(F=1, m=0) \rightarrow 1S_{1/2}(F=0, m=0)$ of atomic hydrogen, for example, operates on the basis of a quartz oscillator locked to the frequency of the hydrogen maser radiation.

The technical details involved in the operation of atomic clocks and their application to the GPS, for example, are complex and too far removed from laser physics to delve into here. The interested reader can easily find more information on the websites of companies that sell atomic clocks and of national laboratories, such as NIST, that maintain “primary” time and frequency standards.

In Section 3.7 we wrote rate equations for the occupation probabilities of degenerate upper states and degenerate lower states of transitions between two energy levels. We briefly discussed conditions under which the degenerate states of equal energy could be assumed to be equally populated, including the case where the atom is irradiated with isotropic rather than unidirectional radiation, or where collisions act to equalize the degenerate-state populations, or where the light intensity is too small to produce significant optical pumping. Under such conditions degeneracy results simply in factors like the ratio g_2/g_1 appearing in formulas for the small-signal gain or absorption

coefficient. This is the case for most naturally occurring optical phenomena and may therefore be called the case of “natural excitation.” In this respect a remark from a classic work on atomic spectroscopy is revealing: “If the excitation occurs in some definitely non-isotropic way, as by absorption from a unidirectional beam of light . . . , large departures from natural excitation may be produced. The study of such effects raises a whole complex of problems somewhat removed from the main body of spectroscopy.”²⁰ Whether *laser* spectroscopy is considered to be “somewhat detached” from the main body of spectroscopy is only a matter of viewpoint, of course. However, the reader should be aware that formulas for absorption, emission, and dispersion found in pre-laser reference material may not be directly applicable to atoms prepared by laser radiation, which is definitely not isotropic!

- It is often necessary in detailed computations to know how electric dipole transition matrix elements depend on the quantum numbers F, m and F', m' of the two states of an atomic transition. Formulas expressing this dependence may be found in the book by Condon and Shortley²⁰ and other monographs.²¹ Here we present a few pertinent formulas.

The transition electric dipole moment between states with angular momentum quantum numbers F, m and F', m' is nonvanishing only if $q = m' - m$ is 0 or ± 1 , as discussed earlier. This dipole matrix element is denoted $\langle F'm'|d_q|Fm\rangle$ in the “bra-ket” notation of quantum theory. According to the *Wigner–Eckart theorem*, it has the form

$$\langle F'm'|d_q|Fm\rangle = (-1)^{F'-m'} \begin{pmatrix} F' & 1 & F \\ -m' & q & m \end{pmatrix} \langle F' || d || F \rangle, \quad (14.3.17)$$

where $\langle F' || d || F \rangle$ is the *reduced matrix element* and the quantity in large parentheses is the *3j symbol*, numerical values for which can be found in books²¹ or on the Web.

The reduced matrix element is independent of m and m' . In the case of hyperfine transitions it depends not only on F and F' but also on the electron angular momentum quantum numbers J and J' of the two states and on the nuclear spin angular momentum quantum number I :

$$\langle F' || d || F \rangle = (-1)^F \sqrt{(2F+1)(2F'+1)} \begin{Bmatrix} J' & I & F' \\ F & 1 & J \end{Bmatrix} \langle J' || d || J \rangle. \quad (14.3.18)$$

Here $\langle J' || d || J \rangle$ is a further “reduced” matrix element, and the quantity in curly brackets is the *6j symbol* that, like the *3j symbol*, is tabulated in various places.²¹ The numerical value of $\langle J' || d || J \rangle$ follows from the formula

$$\frac{1}{\tau_{\text{rad}}} = \frac{4\bar{\omega}^3}{3\hbar c^3} \frac{1}{2J'+1} |\langle J' || d || J \rangle|^2 \quad (14.3.19)$$

for the radiative lifetime τ_{rad} of the upper state with quantum number J' ; the value of τ_{rad} is usually provided by experiment. Here, owing to the smallness of the hyperfine splittings, $\bar{\omega}$ is simply

²⁰E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra*, Cambridge University Press, London, 1959, p. 97.

²¹See, for instance, R. D. Cowan, *The Theory of Atomic Structure and Spectra*, University of California Press, Berkeley, CA, 1981. Different conventions, which are of no physical consequence when followed consistently, are used in the definitions of the *3j* and *6j* coefficients. For example, the reduced matrix element $\langle J' || d || J \rangle$ is sometimes defined such that a factor $2J+1$ appears on the right-hand side of Eq. (14.3.19). We follow the conventions of Cowan, which is consistent with Condon and Shortley. We also write the electric dipole operator, which we have usually denoted by μ , as d here in order to conform to a conventional notation used in defining reduced matrix elements and related quantities.

an average transition frequency, for example, $\bar{\omega} = 2\pi c/\lambda$ with $\lambda = 589.0$ nm in the case of the sodium D_2 transitions. The spontaneous emission rate for the particular transition $F', m' \rightarrow F, m$ is

$$A(F', m' \rightarrow F, m) = \frac{4\bar{\omega}^3}{3\hbar c^3} |\langle F' m' | d_q | F m \rangle|^2, \quad (14.3.20)$$

and the total rate of spontaneous emission from the state F', m' may be shown from the properties of the $3j$ and $6j$ symbols to be just $1/\tau_{\text{rad}}$:

$$\sum_{F, m} A(F', m' \rightarrow F, m) = \frac{1}{\tau_{\text{rad}}}. \quad (14.3.21)$$

So all the hyperfine states of the excited energy level have the same radiative lifetime, as implied by (14.3.19).

The Rabi frequency for the $F, m \leftrightarrow F', m'$ transition in the case of a linearly polarized field $\mathbf{E}(t) = E_0 \hat{x} \cos(\omega t + \phi)$ is

$$\chi_{F', m'; F, m} = \langle F' m' | d_{q=0} | F m \rangle \frac{E_0}{\hbar}, \quad (14.3.22)$$

whereas for a circularly polarized field $\mathbf{E}(t) = (1/\sqrt{2})E_0[\hat{x} \cos(\omega t + \phi) \pm \hat{y} \sin(\omega t + \phi)]$,

$$\chi_{F', m'; F, m} = \langle F' m' | d_{q=\pm 1} | F m \rangle \frac{E_0}{\hbar}. \quad (14.3.23)$$

•

14.4 LASER COOLING

Atoms recoil when they emit or absorb light, as required by conservation of linear momentum and the fact that photons carry linear momentum. Einstein (1909) inferred from his analysis of thermal radiation that atoms recoil not only when they absorb radiation but also when they undergo spontaneous or stimulated emission: They must do so if their average kinetic energy as they absorb and emit thermal radiation at temperature T is to be equal to the value $\frac{3}{2}k_B T$ required by the equipartition theorem of statistical mechanics. (k_B is Boltzmann's constant.) As discussed in Chapter 12, the recoil of an atom in spontaneous emission in particular provides strong evidence for the validity of the photon concept. In fact, the Doppler effect can be understood simply as a consequence of the fact that a photon of frequency ν absorbed (or emitted) by a moving atom carries linear momentum $\hbar\nu/c$ (Problem 14.9).

The recoil forces exerted on atoms by resonant laser radiation are used to slow atoms to very small velocities and thereby to cool gases to extremely low temperatures. In this section we discuss some of the basic physics of laser cooling.

Since a photon carries a linear momentum $\hbar\nu/c = \hbar\omega/c = \hbar k$, the force (= rate of change of linear momentum) on an atom is $F = \hbar k R_{\text{abs}}$, where R_{abs} is the rate at which photons are absorbed. In the case of a two-state atom,

$$R_{\text{abs}} = \left(\frac{dP_2}{dt} \right)_{\text{abs}}. \quad (14.4.1)$$

$(dP_2/dt)_{\text{abs}}$, the rate at which the upper-state probability P_2 changes due to absorption, may be obtained from the optical Bloch equations (9.5.1). Since $P_2 = \rho_{22} = \frac{1}{2}(\rho_{22} - \rho_{11}) + \frac{1}{2}(\rho_{22} + \rho_{11}) = \frac{1}{2}(w + 1)$,

$$\left(\frac{dP_2}{dt}\right)_{\text{abs}} = \frac{1}{2}\left(\frac{dw}{dt}\right)_{\text{abs}} = -\frac{1}{2}\chi v, \quad (14.4.2)$$

which in steady state is obtained as usual by equating the derivatives in (9.5.1) to zero. The resulting expression for the absorption rate in a monochromatic field has a familiar form:

$$R_{\text{abs}} = \left(\frac{dP_2}{dt}\right)_{\text{abs}} = \frac{\frac{1}{2}\chi^2/\beta}{1 + \Delta^2/\beta^2 + \chi^2/\beta A_{21}} = \frac{\frac{1}{2}A_{21}I/I^{\text{sat}}}{1 + \Delta^2/\beta^2 + I/I^{\text{sat}}}, \quad (14.4.3)$$

where we have used the fact that the intensity I is proportional to χ^2 to write R_{abs} in terms of the line-center saturation intensity I^{sat} . The force on the atom is therefore

$$F = \frac{\frac{1}{2}A_{21}I/I^{\text{sat}}}{1 + \Delta^2/\beta^2 + I/I^{\text{sat}}} \hbar k. \quad (14.4.4)$$

In the case under consideration of radiative broadening, $\beta = A_{21}/2$ (Section 9.4).

Consider as an example a sodium atom that has been optically pumped by circularly polarized light into the state $3S_{1/2}(F = 2, m = 2)$. For $I = I^{\text{sat}} = 6.3 \text{ mW/cm}^2$ the acceleration is found from (14.4.4) to be

$$a = 4.6 \times 10^5 \text{ m/s}^2 \quad (14.4.5)$$

for laser radiation at resonance ($\Delta = 0$) with the atom (Problem 14.8). This is 5×10^4 times the acceleration due to gravity and indicates that resonant laser radiation of very modest intensity can significantly affect how atoms move about. It should be noted, however, that for this simplified calculation we have ignored the Doppler shift $kv = \omega v/c$, which contributes to an atom-field detuning.

The force (14.4.4) will obviously slow down an atom moving oppositely to the propagation direction of laser radiation, but eventually it will make the atom turn around and speed up. Suppose, however, that there are *two* laser beams, propagating in opposite directions. Assuming that both lasers have the same intensity and the same detuning Δ from resonance with a stationary atom, the total force they exert on an atom with velocity v is

$$F = \frac{\frac{1}{2}A_{21}I/I^{\text{sat}}}{1 + (\Delta + kv)^2/\beta^2 + I/I^{\text{sat}}} \hbar k - \frac{\frac{1}{2}A_{21}I/I^{\text{sat}}}{1 + (\Delta - kv)^2/\beta^2 + I/I^{\text{sat}}} \hbar k. \quad (14.4.6)$$

The first term is the force exerted by the laser beam with propagation direction parallel to that of the atom's velocity v , so that it exerts a force in the same direction and is resonant with the atomic transition when $\omega = \omega_0 + kv$. (Recall that $\Delta = \omega_0 - \omega$, the detuning of the field frequency ω from the atomic transition frequency ω_0 .) The second term is the force exerted by the laser with propagation direction opposed to the atom's velocity, so

that it is at resonance when $\omega = \omega_0 - kv$. For velocities such that the Doppler shift kv is small in magnitude compared to β and $|\Delta|$, a binomial expansion of (14.4.6), together with the fact that $\beta = A_{21}/2$, gives

$$F \cong -\kappa v, \quad \kappa = \frac{4(I/I^{\text{sat}})\hbar k^2 \Delta/\beta}{(1 + \Delta^2/\beta^2 + I/I^{\text{sat}})^2} \quad (14.4.7)$$

for the total force. This force acts to “damp” the atom’s velocity if $\kappa > 0$, i.e., if $\Delta > 0$, which means that the laser frequency ω is smaller than the atomic transition frequency ω_0 . This has a simple interpretation. The laser beam propagating in the direction opposite to the atom velocity and exerting a retarding force on it is seen by the atom to be Doppler shifted closer to resonance, since its frequency in the laboratory frame is below the atomic resonance frequency; the counterpropagating beam exerting a force acting to increase the atom’s velocity, however, is seen by the atom to be Doppler shifted further away from resonance. The net effect, called *Doppler cooling*, is therefore to slow down the atom.

The net force (14.4.6) is zero if $\Delta = 0$ and acts to increase rather than decrease an atom’s velocity if the lasers are tuned *above* the atomic resonance ($\Delta < 0$). Figure 14.15 plots F vs. v for $\Delta = \beta$ and $I = 0.1I^{\text{sat}}$. It is seen from this figure and Eq. (14.4.6) that the force is always in a direction opposite to v for $\Delta > 0$, but is very small for atomic velocities much greater in magnitude than $\Delta/k = \lambda\Delta/2\pi$. For our sodium example with $\Delta = \beta$, $\lambda\Delta/2\pi = \lambda A_{21}/4\pi = 3$ m/s: An atom’s velocity must already be relatively small in order to slow it further by Doppler cooling. $|kv| < \beta$ defines a “velocity capture range,” i.e., the velocities for which atoms are significantly slowed by Doppler cooling. The cooling of “warmer” atoms is discussed below.

The force (14.4.7) implies that an atom in the field of two counterpropagating laser beams tuned below resonance will eventually come to rest. But our analysis thus far

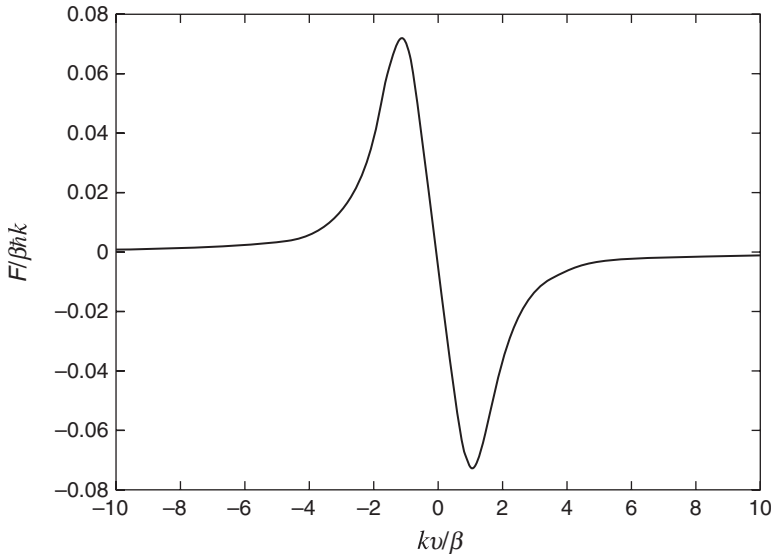


Figure 14.15 The force (14.4.6) vs. kv/β for $\Delta = \beta$ and $I = 0.1I^{\text{sat}}$.

has ignored the fact that the absorption process itself, while producing no net recoil when $v \cong 0$ [Eq. (14.4.7)], causes the mean-square momentum of the atom to increase: The atom can recoil with momentum $\pm \hbar k$, depending from which laser it absorbs a photon. Thus, for small v , the rate at which the mean-square momentum of an atom increases due to absorption is $2R_{\text{abs}}$, where R_{abs} is given by (14.4.3) (for $v \cong 0$) and the factor 2 comes from the fact that we have two laser beams of equal intensity. In addition, the atom can be driven from the ground state back to the ground state by resonance fluorescence and in so doing recoil with momentum $\hbar k$. This recoil, again, has no preferred direction and is therefore zero *on average*, but, like absorption, it causes the mean-square momentum of the atom to grow at the rate $2R_{\text{abs}}(\hbar k)^2$. The increase of the average kinetic energy E of an atom of mass M due to absorption followed by emission is therefore

$$\left(\frac{dE}{dt}\right)_{\text{heating}} = \frac{d\langle p^2 \rangle}{dt} \frac{1}{2M} = \frac{4R_{\text{abs}}(\hbar k)^2}{2M}, \quad (14.4.8)$$

and this prevents an atom from coming to a complete stop.

The cooling force (14.4.6) causes the average kinetic energy to decrease at the rate

$$\left(\frac{dE}{dt}\right)_{\text{cooling}} = \langle Fv \rangle = -\kappa \langle v^2 \rangle = -\frac{2\kappa}{M} \left(\frac{1}{2} M \langle v^2 \rangle \right) = -\frac{2\kappa}{M} E. \quad (14.4.9)$$

Setting the sum of (14.4.8) and (14.4.9) to zero, we obtain the equilibrium kinetic energy $E = (R_{\text{abs}}/\kappa)(\hbar k)^2$, or, from (14.4.3) and (14.4.7),

$$E \approx \frac{\hbar \beta^2}{4\Delta} \left(1 + \frac{\Delta^2}{\beta^2} \right). \quad (14.4.10)$$

We are assuming for simplicity that $I/I^{\text{sat}} \ll 1$, that is, that we are in the linear absorption regime. This is consistent with our neglect of stimulated emission in deriving (14.4.8).

According to the equipartition theorem of statistical mechanics, in thermal equilibrium at temperature T the average kinetic energy $E = (\frac{1}{2})k_B T$ for motion along one axis, that axis here being defined by the two counterpropagating laser beams. Then (14.4.10) implies the equilibrium temperature

$$T \approx \frac{\hbar \beta^2}{2\Delta k_B} \left(1 + \frac{\Delta^2}{\beta^2} \right) \quad (14.4.11)$$

for a gas in the field of two counterpropagating laser beams tuned below resonance. $dT/d\Delta = 0$ when $\Delta = \beta$, and for this value of the detuning the temperature is minimized:

$$T \approx \frac{\hbar \beta}{k_B}. \quad (14.4.12)$$

The smallest possible temperature achievable by Doppler cooling is therefore estimated to be

$$T_{\min}^D = \frac{\hbar\beta}{k_B} = \frac{\hbar A_{21}}{2k_B}. \quad (14.4.13)$$

For the sodium D₂ line, for example, $T_{\min}^D \approx 240 \mu\text{K}$, corresponding to an rms velocity $\approx 0.3 \text{ m/s}$. Note also that the time scale for Doppler cooling in this case, which according to Eq. (14.4.9) is $\sim M/2\kappa$, is about $16 \mu\text{s}$ for $I/I^{\text{sat}} = 0.1$. In other words, extremely low temperatures can be achieved in very short times by Doppler cooling with very modest laser intensities.

Two counterpropagating laser beams slow the atoms' velocities only along one axis. Doppler cooling in all three dimensions of space is realized using three pairs of counterpropagating beams, each with detuning $\Delta > 0$, along three orthogonal axes (x , y , and z). An atom moving in any direction then experiences “viscosity” similar to that of a particle in a fluid, while the random “kicks” (recoils) it gets in absorption and emission are analogous to the thermal fluctuations resulting from collisions of a particle with the molecules of a fluid. Because of these analogies, a vapor undergoing three-dimensional Doppler cooling has come to be referred to as *optical molasses*.

Optical molasses was first observed in 1985. The temperature of Doppler-cooled sodium vapor was inferred from the time taken for atoms to leave the confinement region after the lasers were all turned off, and the result was consistent with the expected $T_{\min}^D \approx 240 \mu\text{K}$. Further experimentation during the next few years at Bell Laboratories and NIST, however, resulted in temperatures $\approx 40 \mu\text{K}$, well below the theoretical minimum for Doppler cooling.

The explanation of this sub-Doppler cooling invokes effects not included in the derivation of (14.4.13), beginning with the fact that the atoms undergo optical pumping in a field having a *spatially varying polarization*. We follow here a simplified model that brings out the essential features,²² assuming two counterpropagating plane waves with orthogonal linear polarizations. At a point z along the axis of propagation we write the total electric field as (the real part of)

$$\mathbf{E}(z, t) = \frac{1}{\sqrt{2}} E_0 (\hat{x} e^{ikz} + \hat{y} e^{-ikz}) e^{-i\omega t}, \quad (14.4.14)$$

or

$$\mathbf{E}(z, t) = \frac{1}{\sqrt{2}} E_0 (\hat{x} + \hat{y} e^{-2ikz}) e^{-i(\omega t - kz)} = \hat{\mathbf{e}}(z) E_0 e^{-i(\omega t - kz)}, \quad (14.4.15)$$

where

$$\hat{\mathbf{e}}(z) = \frac{1}{\sqrt{2}} (\hat{x} + \hat{y} e^{-2ikz}) \quad (14.4.16)$$

²²J. Dalibard and C. Cohen-Tannoudji, *Journal of the Optical Society of America B* **6**, 2023 (1989).

is the complex unit polarization vector, defined with respect to propagation in the positive z direction. Now

$$\begin{aligned}
 \hat{\mathbf{e}}(z) &\rightarrow \frac{1}{\sqrt{2}}(\hat{x} + \hat{y}) && \text{for } kz = \frac{2\pi z}{\lambda} = 0, \\
 &\rightarrow \frac{1}{\sqrt{2}}(\hat{x} - i\hat{y}) = \hat{\mathbf{e}}_- && \text{for } kz = \frac{\pi}{4}, \\
 &\rightarrow \frac{1}{\sqrt{2}}(\hat{x} - \hat{y}) && \text{for } kz = \frac{\pi}{2}, \\
 &\rightarrow \frac{1}{\sqrt{2}}(\hat{x} + i\hat{y}) = \hat{\mathbf{e}}_+ && \text{for } kz = \frac{3\pi}{4}, \\
 &\rightarrow \frac{1}{\sqrt{2}}(\hat{x} + \hat{y}) && \text{for } kz = \pi,
 \end{aligned} \tag{14.4.17}$$

and so on: As z changes in steps of $\lambda/8$, the field polarization changes from linear (along $\hat{x} + \hat{y}$) to σ_- circular to linear (along $\hat{x} - \hat{y}$) to σ_+ circular to linear (along $\hat{x} + \hat{y}$) . . . , and so forth.²³ Because the transition matrix elements between magnetic substates are different for these different field polarizations, the populations of different magnetic substates of an atom will vary as it moves along z . This occurs whenever the two counter-propagating fields have different polarizations and the total field therefore has a polarization gradient.

There is another effect that was neglected in our discussion of Doppler cooling: Energy levels of atoms are shifted in an electric field. This is an electric analog of the Zeeman shifts in a magnetic field. As is the case with Zeeman shifts, these *Stark shifts* (or “light shifts”) are generally different for different magnetic substates. Light shifts of ground magnetic states play a key role in sub-Doppler cooling; in the simplified model followed here it is assumed that the ground level has an orbital angular momentum quantum number $J = \frac{1}{2}$ and therefore only two magnetic substates, $m = \pm\frac{1}{2}$. For the excited level it is assumed that $J = \frac{3}{2}$ and therefore $m = -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}$ (Fig. 14.16a). We will assume laser intensities small enough that excited-state populations are negligible and focus attention on the $m = \pm\frac{1}{2}$ ground states. As shown below, the light shifts of these two states at any point z in the field (14.4.14) are

$$\Delta E(m = -\tfrac{1}{2}) = -2V - V \sin 2kz, \tag{14.4.18a}$$

$$\Delta E(m = +\tfrac{1}{2}) = -2V + V \sin 2kz. \tag{14.4.18b}$$

For our purposes it will not be necessary to have a numerical value for V , which is positive for the case of interest in which the laser frequency ω is less than the atomic transition frequency ω_0 .

²³Note that any phase difference between the two counterpropagating waves can be discarded simply by defining the $z = 0$ origin appropriately.

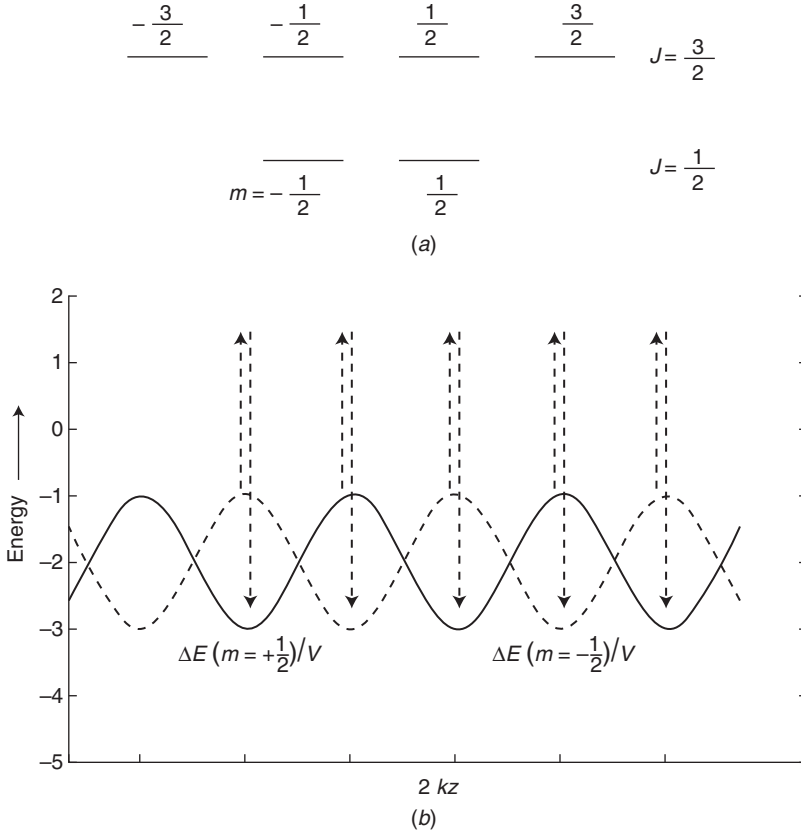


Figure 14.16 (a) Magnetic substates of a $J = \frac{1}{2} \leftrightarrow J = \frac{3}{2}$ transition. (b) Light shifts (14.4.18) of the $m = \pm \frac{1}{2}$ states of the $J = \frac{1}{2}$ ground level. The dashed arrows indicate transitions between the light-shifted $J = \frac{1}{2}$, $m = \pm \frac{1}{2}$ ground states and the $J = \frac{3}{2}$ level at points for which $\sin 2kz = 1$.

Optical pumping by the field (14.4.14) results in transfer of population between the ground $m = \pm \frac{1}{2}$ states. Writing (14.4.16) in the form

$$\begin{aligned} \hat{\mathbf{e}}(z) &= \frac{1}{\sqrt{2}} \left[\frac{1}{\sqrt{2}} (\hat{\mathbf{e}}_+ + \hat{\mathbf{e}}_-) - \frac{i}{\sqrt{2}} (\hat{\mathbf{e}}_+ - \hat{\mathbf{e}}_-) e^{-2ikz} \right] \\ &= \frac{1}{2} \hat{\mathbf{e}}_+ (1 - ie^{-2ikz}) + \frac{1}{2} \hat{\mathbf{e}}_- (1 + ie^{-2ikz}), \end{aligned} \quad (14.4.19)$$

where $\hat{\mathbf{e}}_{\pm}$ are the complex unit polarization vectors for $\sigma_{\pm}$ light [Eq. (14.3.4)], we see that the intensity of σ_+ light at any point z is proportional to $(\frac{1}{4})|1 - i \exp(-2ikz)|^2 = (\frac{1}{2})(1 - \sin 2kz)$, while the intensity of σ_- light is proportional to $(\frac{1}{4})|1 + i \exp(-2ikz)|^2 = (\frac{1}{2})(1 + \sin 2kz)$:

$$I(\sigma_{\pm}) = \frac{1}{2} I_0 (1 \mp \sin 2kz), \quad (14.4.20)$$

where I_0 is the maximum intensity. The intensities $I(\sigma_{\pm})$ determine the time scale τ_p for population transfer by optical pumping between the $m = \pm \frac{1}{2}$ ground states. For σ_+ light,

for instance, absorptive transitions are allowed from the $m = -\frac{1}{2}$ ground state to the excited $m = +\frac{1}{2}$ state and from the $m = +\frac{1}{2}$ ground state to the $m = +\frac{3}{2}$ excited state. As discussed in the preceding section, all the population in steady state will be in the $m = +\frac{1}{2}$ ground state if the intensity is small enough that saturation of the $m = +\frac{1}{2} \leftrightarrow m = +\frac{3}{2}$ transition is negligible, even if all the population initially resides in the $m = -\frac{1}{2}$ ground state. In other words, σ_+ light transfers population from the $m = -\frac{1}{2}$ ground state to the $m = +\frac{1}{2}$ ground state. Similarly, σ_- light transfers population from the $m = +\frac{1}{2}$ ground state to the $m = -\frac{1}{2}$ ground state. Depending on z , either transfer rate can be dominant since the relative strengths of the σ_+ and σ_- components vary with z according to (14.4.20).

We can now understand, qualitatively, how the polarization gradient and optical pumping cause atoms to lose kinetic energy. Consider an atom at a point z such that the field polarization is σ_- . At such a point the atom will be in the state $J = \frac{1}{2}$, $m = -\frac{1}{2}$ in steady state as a result of optical pumping. Now suppose the atom is moving to the right with a velocity v such that $v\tau_p \approx \lambda/4$, so that after the time τ_p it will be at a point where the field polarization is σ_+ (Fig. 14.16). In other words, the atom will have “climbed the potential hill” from a trough to a peak in the light-shifted energy. At the peak the light polarization is σ_+ and the population transfer rate from $m = -\frac{1}{2}$ to $m = +\frac{1}{2}$ is greatest. The population transfer is not instantaneous, and if we imagine that the atom stays in the $m = -\frac{1}{2}$ state as it climbs the potential hill, and is then optically pumped at the top of the hill into the $m = +\frac{1}{2}$ state at the bottom, we can see from Fig. 14.16*b* that it emits a larger photon frequency than it absorbs and therefore that its kinetic energy decreases. After the absorption and emission the atom again starts climbing the potential hill. In reality, of course, the absorption, emission, and optical pumping are statistical processes. But it should be clear that, because the population transfers between the $m = \pm\frac{1}{2}$ states are not instantaneous but require a finite time $\sim \tau_p$, there is a net tendency for kinetic energy to be converted to potential energy and therefore for the atoms to be slowed. Regardless of whether an atom is moving to the right or to the left, it finds itself continually climbing a potential hill and losing kinetic energy in this so-called *Sisyphus cooling*.²⁴

- The light shifts can be derived by first considering the work involved in inducing an electric dipole moment $\mathbf{p} = \alpha\mathbf{E}$. Since the potential energy of an electric dipole moment $\mathbf{p}$ in an incremental electric field $d\mathbf{E}$ is $-\mathbf{p} \cdot d\mathbf{E}$, this work is

$$W = - \int_0^{\mathbf{E}} \mathbf{p} \cdot d\mathbf{E} = -\alpha \int_0^{\mathbf{E}} \mathbf{E} \cdot d\mathbf{E} = -\frac{1}{2} \alpha \mathbf{E}^2. \quad (14.4.21)$$

The factor $\frac{1}{2}$ appears because the dipole is *induced* by the field. In the case of a monochromatic field $\mathbf{E}_0 \cos \omega t$, the cycle-averaged energy

$$\Delta E_i = -\frac{1}{4} \alpha_i(\omega) \mathbf{E}_0^2 \quad (14.4.22)$$

²⁴“The gods had condemned Sisyphus to ceaselessly rolling a rock to the top of a mountain, whence the stone would fall back of its own weight.”—Albert Camus, *The Myth of Sisyphus and Other Essays*, Vintage Books, New York, 1955, p. 119.

is the “light shift” of an atom in a state with polarizability $\alpha_i(\omega)$. The polarizability is in general complex, and $\alpha_i(\omega)$ here is implicitly assumed to be its real part.

The nearly resonant atom–field interaction is characterized by a detuning $\Delta = \omega_0 - \omega$, in which case the polarizability of the lower state of the transition can be inferred from Eq. (9.6.30) and the formula $n(\omega) = 1 + N\alpha(\omega)/2\epsilon_0$ for the refractive index $n(\omega)$ of a dilute gas of N atoms per unit volume [cf. Section 3.15]:

$$\alpha_i(\omega) = \frac{|\mu_{ji}|^2}{\hbar} \frac{\Delta}{\Delta^2 + \beta^2} \quad (14.4.23)$$

for the lower state.²⁵ In the more general case in which the state i can make allowed transitions to more than one other state, this generalizes to

$$\alpha_i(\omega) = \frac{\Delta}{\hbar(\Delta^2 + \beta^2)} \sum_j |\mu_{ji}|^2, \quad (14.4.24)$$

and the light shift (14.4.22) becomes

$$\Delta E_i = -\frac{\Delta}{4\hbar(\Delta^2 + \beta^2)} \mathbf{E}_0^2 \sum_j |\mu_{ji}|^2 \equiv -CI \sum_j |\mu_{ji}|^2, \quad (14.4.25)$$

where I is the intensity and C is positive for $\Delta > 0$ (i.e., for the field tuned below resonance). It is assumed that the states j all have the same energy, so that a single detuning Δ appears in this equation, and that all the transitions $i \leftrightarrow j$ are characterized by the same homogeneous linewidth. This simplification applies in fact to the example of interest here.

Consider now the two ground states of Fig. 14.16. For the $J = \frac{1}{2}$, $m = -\frac{1}{2}$ state,

$$\Delta E(m = -\frac{1}{2}) = -CI(\sigma_+) \left| \langle \frac{3}{2} \frac{1}{2} | d_1 | \frac{1}{2} - \frac{1}{2} \rangle \right|^2 - CI(\sigma_-) \left| \langle \frac{3}{2} - \frac{3}{2} | d_{-1} | \frac{1}{2} - \frac{1}{2} \rangle \right|^2, \quad (14.4.26)$$

since it has an allowed transition to $J = \frac{3}{2}$, $m = \frac{1}{2}$ for σ_+ light and an allowed transition to $J = \frac{3}{2}$, $m = -\frac{3}{2}$ for σ_- light. We use here the notation of Eq. (14.3.17) for the electric dipole matrix elements. Since we are ignoring any hyperfine structure for our model atom, the matrix elements $\langle J'm' | d_q | Jm \rangle$ (with $q = m' - m$) are given simply by (14.3.17) with $F' = J'$ and $F = J$. Then

$$\left| \langle \frac{3}{2} \frac{1}{2} | d_1 | \frac{1}{2} - \frac{1}{2} \rangle \right|^2 = \begin{pmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{pmatrix}^2 \left| \langle \frac{3}{2} | d | \frac{1}{2} \rangle \right|^2 = \frac{1}{12} \left| \langle \frac{3}{2} | d | \frac{1}{2} \rangle \right|^2 \quad (14.4.27)$$

and

$$\left| \langle \frac{3}{2} - \frac{3}{2} | d_{-1} | \frac{1}{2} - \frac{1}{2} \rangle \right|^2 = \begin{pmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ \frac{3}{2} & -1 & -\frac{1}{2} \end{pmatrix}^2 \left| \langle \frac{3}{2} | d | \frac{1}{2} \rangle \right|^2 = \frac{1}{4} \left| \langle \frac{3}{2} | d | \frac{1}{2} \rangle \right|^2. \quad (14.4.28)$$

Therefore

$$\Delta E(m = -\frac{1}{2}) = -C \left[\frac{1}{12} I(\sigma_+) + \frac{1}{4} I(\sigma_-) \right] \left| \langle \frac{3}{2} | d | \frac{1}{2} \rangle \right|^2, \quad (14.4.29)$$

²⁵Note from (9.6.30) that the polarizability of the upper state j is $\alpha_j(\omega) = -\alpha_i(\omega)$. Here we are concerned only with ground-state light shifts and therefore ground-state polarizabilities.

and, from Eq. (14.4.20),

$$\Delta E(m = -\frac{1}{2}) = -\frac{1}{12}CI_0 \left| \langle \frac{3}{2} \frac{3}{2} | d | \frac{1}{2} \frac{1}{2} \rangle \right|^2 (2 + \sin 2kz) \equiv -V(2 + \sin 2kz) \quad (14.4.30)$$

for the light shift of an atom in state $J = \frac{1}{2}$, $m = -\frac{1}{2}$ at the point z in the field (14.4.14). Similarly,

$$\begin{aligned} \Delta E(m = +\frac{1}{2}) &= -CI(\sigma_+) \left| \langle \frac{3}{2} \frac{3}{2} | d_1 | \frac{1}{2} \frac{1}{2} \rangle \right|^2 - CI(\sigma_-) \left| \langle \frac{3}{2} -\frac{1}{2} | d_{-1} | \frac{1}{2} \frac{1}{2} \rangle \right|^2 \\ &= -C \left[\frac{1}{4}I(\sigma_+) + \frac{1}{12}I(\sigma_-) \right] \left| \langle \frac{3}{2} \frac{3}{2} | d | \frac{1}{2} \frac{1}{2} \rangle \right|^2 \\ &= -V(2 - \sin 2kz). \end{aligned} \quad (14.4.31)$$

•

Since V and the optical pumping time τ_p are directly proportional and inversely proportional, respectively, to the intensity, the rate $1/\tau_{\text{cool}} = E^{-1}(dE/dt)_{\text{cool}} \propto V\tau_p$ for Sisyphus cooling is independent of intensity, in contrast to the rate (14.4.9) for Doppler cooling, which is proportional to intensity. As in the case of Doppler cooling, the heating rate is proportional to the intensity. Since the equilibrium temperature is proportional to the heating rate times τ_{cool} , it is proportional to intensity for Sisyphus cooling. The velocity capture range is also intensity dependent, again in contrast to Doppler cooling: Our simplified discussion shows that the optimal situation for Sisyphus cooling occurs when $v\tau_p \sim \lambda$, that is, when an atom with velocity v moves a distance $\sim \lambda$ during an optical pumping time τ_p , so that $kv \sim 1/\tau_p$, implying a velocity capture range inversely proportional to τ_p and therefore directly proportional to the laser intensity.

At low laser intensities the cooling in optical molasses is observed to be less effective in the presence of magnetic fields, which cause transitions among magnetic substates and thereby weakens the Sisyphus effect. If the laser intensity is sufficiently large, however, the cooling becomes less sensitive to magnetic fields because the light shifts and optical pumping rates become larger. Detailed analyses of Sisyphus cooling are found to be consistent with such experimental observations as well as with the extremely low temperatures—as small as a few microkelvins—that have been realized in optical molasses.

The temperatures obtained by Sisyphus cooling can be lowered by lowering the ground-state light shift, for example, by decreasing the laser intensity or by increasing the detuning. But there are limits even to Sisyphus cooling. The smallest average kinetic energy for atoms absorbing and emitting photons is that associated with the recoil of a nearly stationary atom when it absorbs or emits a *single photon*. The mean-square recoil momentum, $(\hbar\omega_0/c)^2$, implies a minimum average kinetic energy $(\hbar\omega_0/c)^2/2M = \hbar^2/2M\lambda^2$ for an atom with mass M and transition wavelength λ , and therefore a temperature

$$T_{\text{recoil}} = \frac{\hbar^2}{Mk_B\lambda^2}. \quad (14.4.32)$$

For $M = 23$ amu and $\lambda = 589$ nm (sodium), $T_{\text{recoil}} = 2.4$ μK ; for $M = 133$ amu and $\lambda = 852$ nm (cesium), $T_{\text{recoil}} = 0.2$ μK . Temperatures several times larger than T_{recoil} have been obtained by Sisyphus cooling. By employing quantum interference effects similar to those used to inhibit absorption in electromagnetically induced transparency (Section 9.10), it has been possible to cool atoms to temperatures well below the recoil limit.

- The force on an electric dipole moment $\mathbf{p}$ in an electromagnetic field is given by the “Lorentz force,”

$$\mathbf{F} = (\mathbf{p} \cdot \nabla) \mathbf{E} + \frac{\partial \mathbf{p}}{\partial t} \times \mathbf{B}. \quad (14.4.33)$$

To obtain the force on an electric dipole in the electromagnetic field

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2} \mathbf{E}_0(\mathbf{r}) e^{-i\omega t} + \frac{1}{2} \mathbf{E}_0^*(\mathbf{r}) e^{i\omega t}, \quad \mathbf{B}(\mathbf{r}, t) = \frac{1}{2} \mathbf{B}_0(\mathbf{r}) e^{-i\omega t} + \frac{1}{2} \mathbf{B}_0^*(\mathbf{r}) e^{i\omega t}, \quad (14.4.34)$$

we write

$$\mathbf{p}(\mathbf{r}, t) = \frac{1}{2} \mathbf{p}_0(\mathbf{r}) e^{-i\omega t} + \frac{1}{2} \mathbf{p}_0^*(\mathbf{r}) e^{i\omega t}, \quad (14.4.35)$$

with $\mathbf{p}_0(\mathbf{r}) = \alpha(\omega) \mathbf{E}_0(\mathbf{r})$ and $\alpha(\omega) = \alpha_R(\omega) + i\alpha_I(\omega)$ the complex polarizability. It follows from the Maxwell equation $\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t$ that $\mathbf{B}_0 = -(i/\omega) \nabla \times \mathbf{E}_0$, and from these expressions that the cycle-averaged z component of the force (14.4.33) is

$$F_z(\mathbf{r}) = \frac{1}{4} \alpha_R(\omega) \frac{\partial}{\partial z} |\mathbf{E}_0(\mathbf{r})|^2 - \frac{1}{2} \alpha_I(\omega) \text{Im} \left[E_{0x}(\mathbf{r}) \frac{\partial E_{0x}^*}{\partial z} + E_{0y}(\mathbf{r}) \frac{\partial E_{0y}^*}{\partial z} + E_{0z}(\mathbf{r}) \frac{\partial E_{0z}^*}{\partial z} \right]. \quad (14.4.36)$$

The x and y components of the force are obtained by replacing $\partial/\partial z$ with $\partial/\partial x$ and $\partial/\partial y$, respectively. The formula (14.4.36) also gives the dipole force on a small dielectric sphere (Section 14.5).²⁶

Consider as an example a two-state atom, for which the complex, near-resonance polarizability can be inferred from (9.6.18), for instance:

$$\alpha(\omega) = \frac{|\mu|^2}{\hbar} \frac{1}{\Delta - i\beta} \quad (\Delta = \omega_0 - \omega). \quad (14.4.37)$$

The first term on the right-hand side of (14.4.36) is therefore the z component of

$$\mathbf{F}_{\text{dipole}} = -\nabla[U(\mathbf{r})], \quad (14.4.38)$$

where

$$U(\mathbf{r}) = -\frac{|\mu|^2}{4\hbar} \frac{\Delta}{\Delta^2 + \beta^2} |E_0|^2, \quad (14.4.39)$$

which is just the light shift (14.4.22). In other words, the “dipole force” $\mathbf{F}_{\text{dipole}}$ is the force resulting from a gradient of the light-shift energy; this is the basis for the *optical lattices* discussed in the following section.

To interpret the other part of the force in (14.4.36), let us assume for simplicity the plane wave

$$\mathbf{E}_0(\mathbf{r}) = \hat{x} E_0 e^{ikz}. \quad (14.4.40)$$

²⁶See, for instance, P. C. Chaumet and M. Nieto-Vesperinas, *Optics Letters* **25**, 1065 (2000).

In this case the part of the total force that is proportional to the imaginary part of the polarizability in (14.4.36) is

$$\mathbf{F}_{\text{recoil}} = \frac{1}{2} \frac{|\mu|^2}{\hbar} \frac{\beta}{\Delta^2 + \beta^2} E_0^2 k \hat{z} = \hbar k \hat{z} \frac{\frac{1}{2} \chi^2 \beta}{\Delta^2 + \beta^2} = \hbar k \hat{z} R_{\text{abs}}, \quad (14.4.41)$$

for a two-state atom, R_{abs} being just the absorption rate [cf. Eq. (14.4.3)] when the intensity is well below I^{sat} , as assumed here. $\mathbf{F}_{\text{recoil}}$ is just the recoil force (14.4.4) exerted on the atom.

The dipole force arises from spatial variations in the Stark shift $U(\mathbf{r})$ of an atom, whereas the recoil force is a direct consequence of absorption (and, more generally, emission). The recoil force has its maximum value at resonance ($\Delta = 0$) and decreases as $1/\Delta^2$ when the field is far off resonance, whereas the dipole force vanishes at resonance and decreases more slowly, as $1/|\Delta|$, far off resonance.²⁷ The recoil force saturates with intensity according to the formula (14.4.3). The potential energy (14.4.39) determining the dipole force via (14.4.38) at large intensities takes the form (Problem 14.10)

$$U(\mathbf{r}) = -\frac{1}{2} \hbar \Delta \ln \left[1 + \frac{I}{I_v^{\text{sat}}} \frac{\beta^2}{\Delta^2 + \beta^2} \right]. \quad (14.4.42)$$

We have assumed in our discussion of laser cooling that only the recoil force affects the motion of an atom. This assumption is justified if the field is well described as a plane wave ($\nabla E_0 = 0$). In general, however, the dipole force must be taken into account and, as discussed in the following section, it can significantly affect the motion of atoms in optical fields.

We mentioned at the beginning of this section that Einstein invoked recoil in absorption and emission in his treatment of thermal radiation. Assuming an isotropic and unpolarized field with spectral density $\rho(\omega)$, Einstein obtained a cooling rate of the form (14.4.9) with “friction” coefficient

$$\kappa = \left(\frac{\hbar \omega}{c^2} \right) (P_1 - P_2) B \left[\rho(\omega) - \frac{\omega}{3} \frac{d\rho}{d\omega} \right]. \quad (14.4.43)$$

P_1 and P_2 are the lower- and upper-state probabilities of the two-state transition of frequency ω and B is the Einstein coefficient for absorption in a broadband field (Section 3.6). The retarding force $F = -\kappa v$ arises from Doppler shifts and “aberration” in the broadband field. The cooling rate in thermal equilibrium is balanced by the recoil heating rate, which Einstein calculated to be

$$\left(\frac{dE}{dt} \right)_{\text{heating}} = \frac{1}{3} \left(\frac{\hbar \omega}{c} \right)^2 P_1 B \rho(\omega). \quad (14.4.44)$$

From this result, together with (14.4.9), (14.4.43), and $E = (\frac{1}{2}) k_B T$, therefore, the sum of the heating and cooling rates is zero when $\rho(\omega)$ satisfies

$$\rho - \frac{\omega}{3} \frac{d\rho}{d\omega} = \frac{\hbar \omega}{3 k_B T} \frac{P_1}{P_1 - P_2} \rho \quad (14.4.45)$$

with $P_2/P_1 = \exp(-\hbar \omega / k_B T)$ in thermal equilibrium. The solution for $\rho(\omega)$ with “initial condition” $\rho(0) = 0$ is exactly the Planck spectrum. •

²⁷In terminology that seems to generate more confusion than insight, the recoil and dipole forces are sometimes attributed to “real” and “virtual” transitions, respectively.

The lowest temperatures obtained by laser cooling are realized with trapped atoms (Section 14.5). In order for atoms to be captured by a trap, their velocities must be small, much smaller than the typical mean velocity of atoms emerging from an oven, which is ~ 800 m/s for sodium atoms at a vapor temperature $T = 600$ K. Before atoms can be trapped and cooled to temperatures in the microkelvin range and below, therefore, they must first be cooled to temperatures well below their oven vapor temperature. This is done with a counterpropagating laser beam with frequency ω tuned below the atomic resonance frequency ω_0 , so that an atom with an initial velocity v experiences the largest retarding force when $\omega_0 = \omega + kv$, that is, when it is at resonance with the Doppler-shifted field frequency. The atom moves increasingly out of resonance, however, as its velocity decreases, and the degree of slowing it experiences is substantially reduced unless ω or ω_0 can be varied to maintain resonance. The field frequency can be changed by chirping, but the most commonly employed method of maintaining resonance is to change the atomic resonance frequency with a magnetic field that varies along the atomic beam. In this *Zeeman slowing* the magnetic field along the atomic beam is varied by varying the winding of a solenoid, as indicated in Fig. 14.17.

- To get an idea of the sort of numbers involved in Zeeman slowing, let us assume that atoms start out with a velocity $v_i = 800$ m/s and are brought to a complete stop ($v_f = 0$) after a distance z . Using $a = 4.6 \times 10^5$ m/s² [Eq. (14.4.5)] for the deceleration of sodium atoms in a field of intensity $I = I^{\text{sat}} = 6.3$ mW/cm², we calculate that the time taken for the atoms to be stopped is $t = (v_i - v_f)/a = 1.7$ ms, and the distance over which this occurs is $z = \frac{1}{2}at^2 = 70$ cm.

The Doppler shift at a distance z from the trap (Fig. 14.17) is $kv(z) = \omega v(z)/c$. To estimate the magnetic field required to compensate for this Doppler shift, we assume that the Landé g factors of the upper and lower states of the atomic transition differ by a factor ~ 1 —a reasonable approximation—so that the Zeeman shift of the transition frequency is $\Delta\omega_0 \sim \mu_B B_z/\hbar$ for σ_+ light ($\Delta m = +1$). The magnetic field required to keep an atom in resonance with the field as it slows down is therefore

$$B_z(z) = \frac{\hbar\omega}{\mu_B c} v(z) \sim \frac{\hbar\omega_0}{\mu_B c} at = \frac{\hbar\omega_0}{\mu_B c} \sqrt{2az}. \quad (14.4.46)$$

For the sodium transition at 589 nm we obtain, assuming again the acceleration $a = 4.6 \times 10^5$ m/s²,

$$B_z(z) \sim 120\sqrt{z} \text{ gauss}, \quad (14.4.47)$$

where z (in centimeters) is the distance from the trap and we have used the fact that a tesla is 10^4 gauss. •

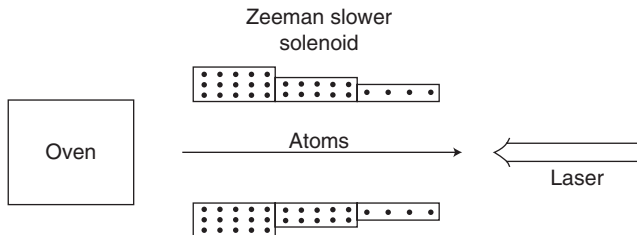


Figure 14.17 A Zeeman slower. Atoms are slowed by the force of a counterpropagating laser beam and are kept near resonance with the Doppler-shifted laser field by the Zeeman shift of their transition frequency in a magnetic field that varies along the direction of the atomic beam. The laser and magnetic field parameters are chosen such that the atoms are slowed enough to allow them to be captured in a trap.

Laser cooling is used to slow cesium atoms and thereby to obtain the extremely small transit-time line-broadening characteristic of atomic fountain clocks that, as noted earlier (Section 14.3), are about 10 times more accurate than atomic beam clocks. Some features of a fountain clock are indicated in Fig. 14.18. Cesium atoms in a trap (Section 14.5) are cooled to microkelvin temperatures at the intersection of six laser beams, as discussed above, and then “launched” upward. The launching can be done by the force on the atoms of a single laser beam, but this results in heating of the atoms as they absorb and emit photons and suffer recoil kicks. The atoms are instead launched by up-shifting the frequency of the upward-propagating laser beam and down-shifting the frequency of the downward-propagating laser beam to produce a *moving standing wave*. If the up-shifted and down-shifted frequencies are, respectively, $\nu + \Delta\nu$ and $\nu - \Delta\nu$, this standing wave moves upward with a velocity $v = \lambda \Delta\nu$ (Problem 14.11), and the atoms move along with the standing wave without any average recoil velocity: The Doppler effect causes the atoms to see the same frequency ν for both fields, resulting in optical molasses in the moving frame. The atoms continue moving upward at a few meters/second after the lasers are turned off (Problem 14.11), and, after being optically pumped into one of the states of the clock transition, they enter the microwave cavity, which serves as the first field for the Ramsey method of separated oscillatory fields. The atoms pass through the cavity, reach their apogee, and then fall by gravity and experience the second Ramsey field as they pass through the microwave cavity a second time. For a total atom path length of ~ 30 cm the Ramsey interrogation time $T \sim 1$ s and therefore the Ramsey fringes are extremely sharp (Section 9.11). The detection of atoms that have made the clock transition is done by irradiating the atoms that fall through the microwave cavity with laser radiation and counting resonance fluorescence photons, and the peak of the atomic resonance is determined by the frequency of the microwave field that produces the largest fluorescence signal. The essential difference from the older atomic beam clock, then, is the narrower resonance of the cesium clock transition resulting from the long interrogation time made possible by laser cooling.

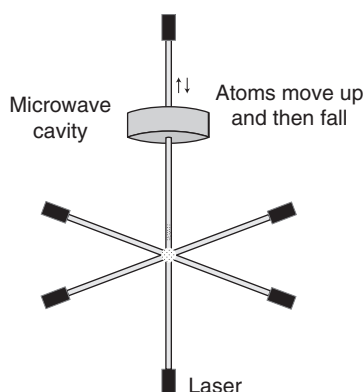


Figure 14.18 An atomic fountain in which atoms make two passes through a microwave cavity, corresponding in effect to the Ramsey method of separated oscillatory fields. The long Ramsey interrogation time (Section 9.11) is the primary reason for the high degree of accuracy of atomic fountain clocks.

14.5 TRAPPING ATOMS WITH LASERS AND MAGNETIC FIELDS

Laser cooling as such does not *trap* atoms since atoms can diffuse out of the laser fields. For trapping we require forces acting to spatially confine atoms. The most ubiquitous atom trap is the *magneto-optical trap* (MOT) employing both lasers and magnetic fields. Before discussing the operating principles of a MOT we will consider briefly the trapping of atoms by a static magnetic field, specifically the *quadrupole* magnetic field

$$\mathbf{B} = \mathcal{B}[x\hat{x} + y\hat{y} - 2z\hat{z}], \quad (14.5.1)$$

where $\mathcal{B}$ is the magnetic field gradient along the x and y axes; the gradient of $\mathbf{B}$ along the z axis is $-2\mathcal{B}$, as required by the Maxwell equation $\nabla \cdot \mathbf{B} = 0$. The field (14.5.1) is produced by a pair of Helmholtz coils (Fig. 14.19) when the coil separation is 1.25 times the coil radius.

The Zeeman-shifted energy of an atom in a state i in a magnetic field of magnitude B has the form [Eq. (14.3.1)]

$$E_i(B) = E_i(B = 0) - \mu_i B, \quad (14.5.2)$$

where $\mu_i (\propto m\mu_B)$ is the magnetic dipole moment of state i along the direction of the magnetic field. A state with $\mu_i > 0$ therefore has its lowest Zeeman-shifted energy at points in space where the magnetic field is strongest, and an atom in such a state will experience a force $-\nabla(-\mu_i B)$ acting to move it to a point where the magnetic field is higher. A state with $\mu_i > 0$ is said to be a *high-field seeker*. Similarly a state with $\mu_i < 0$ is a *low-field seeker*.²⁸ The potential energy function $-\mu_i B \sim \mu_B B = 9.274 \times 10^{-24} B$ J, or $0.67 B$ K when expressed as temperature in degrees Kelvin (K). Thus, for the magnetic fields typical of laboratory experiments ($B \ll 1$ tesla), the “depth” of a magnetic trap is much less than 1K; atoms must be cooled to such temperatures before they can be magnetically trapped. Trapping is accomplished much more effectively by combining magnetic fields with lasers, as we now discuss.

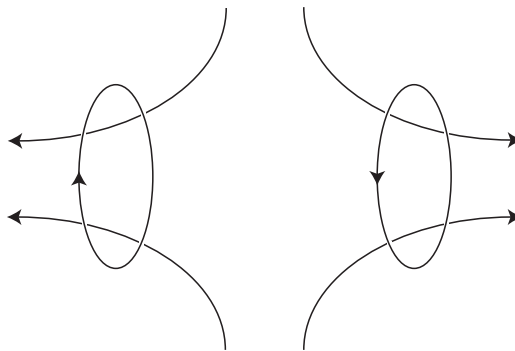


Figure 14.19 Two Helmholtz coils with opposing electric currents produce a quadrupole magnetic field.

²⁸In regions free of electric currents there are no local maxima of $|\mathbf{B}|$, and consequently only low-field seekers can be trapped by magnetic fields alone. This is discussed by W. H. Wing, *Progress in Quantum Electronics* **8**, 181 (1984).

Equation (14.5.2) implies that atomic transition frequencies in a magnetic field such as (14.5.1) are position-dependent. The force exerted by a laser field on an atom therefore depends on the position of the atom, and it is this position-dependent force that confines atoms in a magneto-optical trap. Unlike a purely magnetic trap, a MOT acts not only to trap atoms but also to cool them with counterpropagating laser beams as discussed in the preceding section. To understand this in more detail we consider a model atom with a $J = 0$ ground level and a $J = 1$ upper level. The atom is assumed to be on the z axis, where the magnetic field (14.5.1) is $\mathbf{B} = -2\mathcal{B}z\hat{z}$. The frequencies of the $\Delta m = 0$ and $\Delta m = \pm 1$ transitions are

$$\begin{aligned}\omega_0(J=0, m=0 \rightarrow J=1, m=0) &= \omega_0, \\ \omega_0(J=0, m=0 \rightarrow J=1, m=+1) &= \omega_0 - \frac{\mu_B g B_z}{\hbar} = \omega_0 + \frac{2\mu_B g \mathcal{B}z}{\hbar} = \omega_0 + \Delta\omega_0, \\ \omega_0(J=0, m=0 \rightarrow J=1, m=-1) &= \omega_0 + \frac{\mu_B g B_z}{\hbar} = \omega_0 - \frac{2\mu_B g \mathcal{B}z}{\hbar} = \omega_0 - \Delta\omega_0.\end{aligned}\tag{14.5.3}$$

where $\Delta\omega_0 = 2\mu_B g \mathcal{B}z/\hbar$ and ω_0 is the transition frequency in the absence of the magnetic field. The recoil force exerted on the atom by $\sigma_{\pm}$ light of frequency ω is given by Eq. (14.4.4) with $\Delta = \omega_0 - \omega \pm \Delta\omega_0$:

$$F(\sigma_{\pm}) = \frac{\frac{1}{2}A_{21}I/I^{\text{sat}}}{1 + (1 \pm \Delta\omega_0/\beta)^2} \hbar k \cong \frac{\frac{1}{4}A_{21}I/I^{\text{sat}}}{1 \pm \Delta\omega_0/\beta} \hbar k \tag{14.5.4}$$

for $\Delta\omega_0 \ll \beta$ and $I \ll I^{\text{sat}}$. To simplify the algebra we have set $\omega_0 - \omega = \beta = A_{21}/2$, and we are ignoring for the moment any motion of the atom. Under these assumptions the average recoil force when the atom is irradiated with σ_+ light propagating in the $+z$ direction and σ_- light propagating in the $-z$ direction (Fig. 14.20) is

$$\begin{aligned}F(\sigma_+) - F(\sigma_-) &\cong \left[\frac{\frac{1}{4}A_{21}I}{I^{\text{sat}}} \right] \left[\frac{1}{1 + \Delta\omega_0/\beta} - \frac{1}{1 - \Delta\omega_0/\beta} \right] \hbar k \cong - \left[\frac{\frac{1}{2}A_{21}I}{I^{\text{sat}}} \right] \hbar k \frac{\Delta\omega_0}{\beta} \\ &= - \left[\frac{2I}{I^{\text{sat}}} \frac{\mu_B g \mathcal{B}}{\hbar} \right] \hbar k z \equiv -k_s z\end{aligned}\tag{14.5.5}$$

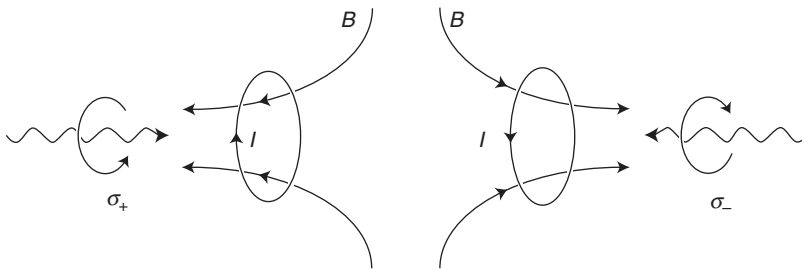


Figure 14.20 One-dimensional model of a magneto-optical trap (MOT). After W. D. Phillips in *Fundamental Systems in Quantum Optics*, eds. J. Dalibard, J.-M. Raimond, and J. Zinn-Justin (North-Holland, Amsterdam, 1992), p. 165.

when the two fields have the same intensity I . The combined effects of the magnetic field and the counterpropagating laser fields produce a restoring force with “spring constant” k_s . The approximation (14.5.5) is valid when the magnetic field and the distance from the center of the trap ($z = 0$) are sufficiently small.

In addition to the trapping effect in a MOT there is cooling by the counterpropagating laser beams: The MOT produces a trapped optical molasses. Let us continue with our one-dimensional model and include the effects of both the trapping and Doppler cooling forces on the motion of an atom of mass M . The equation of motion for the coordinate z of the atom is $M d^2z/dt^2 = -k_s z - av$, or

$$\frac{d^2z}{dt^2} + \gamma_t \frac{dz}{dt} + \omega_t^2 z = 0. \quad (14.5.6)$$

Here the oscillation frequency ω_t for an atom in the MOT is given by

$$\omega_t^2 = \frac{k_s}{M} = \frac{2I\mu_B g \mathcal{B} k}{MI^{\text{sat}}} \quad (14.5.7)$$

and the damping constant $\gamma_t = \kappa/M$, where κ is defined by (14.4.7). As in (14.5.4) we simplify some algebra by taking $\Delta = \omega_0 - \omega = \beta$. Then, if I/I^{sat} is small,

$$\gamma_t \cong \frac{\hbar k^2}{MI^{\text{sat}}}. \quad (14.5.8)$$

Equation (14.5.6) is the familiar equation for a damped harmonic oscillator and expresses the fact that atoms in a MOT are both trapped and cooled. For $M = 23$ amu, $\lambda = 589$ nm, and $g \sim 1$, we estimate $\omega_t/2\pi \sim 2$ kHz and $\gamma_t \sim 10^5 \text{ s}^{-1}$ for $I/I^{\text{sat}} = 0.3$ and $\mathcal{B} = 10$ G/cm, as is fairly typical of MOTs. The atom oscillations are typically “overdamped” ($\gamma_t > \omega_t$), as this example suggests. In an actual MOT the oscillation frequencies are different along the x , y , and z axes but have similar magnitudes.

The force (14.5.5) exerted on atoms by the counterpropagating laser beams in a MOT is typically considerably larger than the magnetic force $\nabla(\mu_B g \mathcal{B}) = \mu_B g \mathcal{B}$ for atoms at a distance of a wavelength or so from the center of the trap (Problem 14.12). Typically, $10^{18} - 10^{21}$ atoms/m³ are trapped, and the size of the trapped “cloud” can be inferred from the temperature T : According to the equipartition theorem, the average potential energy $\frac{1}{2}k_s z^2$ and the average kinetic energy $\frac{1}{2}Mv^2$ are equal to $\frac{1}{2}k_B T$ in thermal equilibrium, so that the cloud size is estimated to be

$$z_{\text{cloud}} \sim \sqrt{k_B T / k_s}, \quad (14.5.9)$$

which is typically $\sim 10^{-2}$ cm for temperatures $T \sim 200$ μK obtained by Doppler cooling (Problem 14.12). The temperature T can be inferred by turning off the trap and determining the velocities of escaping atoms by their time of flight from the trap.

These results of our highly simplified, one-dimensional model for a MOT are in reasonable accord with those of three-dimensional numerical analyses that account for hyperfine structure and all the magnetic substates of real atoms. When such analyses include polarization gradients and other effects, they become quite complicated, and in fact the physics of atom cooling and trapping is a subject to which entire books are

devoted.²⁹ Magnetic traps have been developed over many years, especially in connection with the confinement of hot plasmas in nuclear fusion research, and there is an extensive literature on the subject.

Techniques for the cooling and trapping of *ions* differ in several respects from those we have described for neutral atoms. In this case the electric charge of the particles allows them to be efficiently cooled and trapped with static electric and magnetic fields. In a *Penning trap*, for example, ions are confined in the x and y directions by a magnetic field along z while a quadrupole electric field confines them in the z direction. Cooling can be done by applying a laser beam along some other direction; the laser field cools ions moving toward it and pushes them back to the center of the trap where they exchange energy in collisions with hotter ions, resulting in a cooling of the ion “cloud.” Ion cooling with a single laser beam is done similarly in a *Paul trap*, which confines ions with oscillating electric fields. The repulsive forces between ions make it difficult or impossible to realize Bose–Einstein condensation and, since we highlight that topic in the following section, we will not consider ion traps any further.

- We have described the motion of atoms in a MOT using elementary classical mechanics rather than quantum mechanics. Classical theory is a good approximation provided the atoms are not too cold and therefore their de Broglie wavelengths are not too large. We discuss this point a bit more in the following section.

According to the formula (2.3.8) for a particle in a harmonic-oscillator potential, the motion of an atom in a MOT is characterized by the quantized energies $E_n = \hbar\omega_t(n + \frac{1}{2})$, $n = 0, 1, 2, 3, \dots$, and so we can expect classical theory to be accurate if $\hbar\omega_t/k_B T$ is small. But at sufficiently low temperatures quantum effects become important, and in particular the zero-point energy $\frac{1}{2}\hbar\omega_t$ puts a lower limit on the temperature to which atoms can be cooled in a MOT:

$$T_{\min} = \frac{\hbar\omega_t}{k_B}. \quad (14.5.10)$$

This limiting temperature is very small (Problem 14.12), but it has been closely approached in experiments in which a single beryllium ion is confined in a Paul trap with $\omega_t \sim 100$ MHz. •

There is a great deal more to be said about laser cooling and trapping of atoms than is warranted here.²⁹ In the remainder of this section we will touch on two applications of the dipole force (14.4.38).

The light shifts (14.4.18) we calculated in connection with Sisyphus cooling imply that the model atoms in their $m = \pm \frac{1}{2}$ ground states experience periodic potentials in the counterpropagating laser fields; the period of these potentials is half the wavelength of the presumed plane wave. The dipole force given by Eq. (14.4.38),

$$\mathbf{F}_{\text{dipole}} = -\nabla[U(\mathbf{r})] = \nabla\left[\frac{1}{4}\alpha_R(\omega)\mathbf{E}_0^2(\mathbf{r})\right] \quad (14.5.11)$$

when we take $\mathbf{E}_0$ to be real, implies that an atom finds itself in a periodic potential $U(\mathbf{r})$ whenever the light intensity is periodic, for example, when the atom is in a standing-wave field. The intersection of two plane waves at an angle θ results in a periodic

²⁹See, for instance, H. J. Metcalf and P. van der Straten, *Laser Cooling and Trapping*, Springer, New York, 1999. A tutorial on the design and operation of an “inexpensive” MOT is given by C. Wieman, G. Flowers, and S. Gilbert, *American Journal of Physics* **63**, 317 (1995).

potential with period $\lambda/[2 \sin(\theta/2)]$ for a polarizable particle [cf. Eq. (11.14.31)], and, more generally, pairs of interfering fields can produce a great variety of two- and three-dimensional potentials with different periodicities along different directions. An atom in such an “egg-crate” potential, called an *optical lattice*, behaves very much like an electron in a crystal, where the allowed energies are restricted to certain bands depending on the crystal structure (Chapter 2). The light shifts can be comparable to or greater than the average kinetic energy of atoms loaded into the lattice from a MOT. By varying the laser intensity or detuning, the atoms can be tightly or weakly bound by the potential wells of the “egg crate”; they can be confined—“held in midair” by laser beams—in potential wells comparable in their spatial extent to an optical wavelength. If the field frequency is such that $\alpha_R(\omega) > 0$, the dipole force moves atoms toward maxima of the electric field; if $\alpha_R(\omega) < 0$, it moves atoms toward field minima. Either way, the atoms can be trapped by the optical lattice, and the trapping potentials are increased or decreased as the field intensity is increased or decreased. Thus it is possible with optical lattices to obtain a “phase-space density” of atoms large enough to realize “all-optical” Bose–Einstein condensation (Section 14.6).

Absorption and recoil in optical lattices can be minimized by detuning the lasers far from atomic resonances. The peak values of the light-shift potentials are usually expressed in units of the atomic recoil energy (Problem 14.14). Another attractive feature of optical lattices is that the interaction strengths of atoms at different lattice sites can be controlled by designing the lattice appropriately, and relaxation processes can be effectively eliminated. Fundamental quantum phenomena predicted for electrons in crystal lattices but difficult to observe because of electron–electron and electron–photon interactions can be probed very “cleanly” with atoms stored in optical lattices. Transferring atoms from a Bose–Einstein condensate in a magnetic trap to an optical lattice has led to some remarkable experimental studies, one of which is described in the following section.

We have restricted our discussion of laser trapping thus far to atoms, but lasers are also used to trap many other particles including, for example, DNA molecules, biological cells, and small dielectric particles. The physical basis for this trapping is again the dipole force (14.5.11). Consider a nonabsorbing dielectric sphere with radius a smaller than the wavelength of incident light. With n the refractive index of the dielectric material, and n_b the refractive index of the medium in which the sphere is placed, the polarizability α_R of the sphere is given by electromagnetic theory as

$$\alpha_R = 4\pi\epsilon_0 n_b^2 \left(\frac{n^2 - n_b^2}{n^2 + 2n_b^2} \right) a^3. \quad (14.5.12)$$

(α_R , n , and n_b are all evaluated at the frequency of the incident light.) It follows that the dipole force (14.5.11) is in the direction of increasing electric field if $n > n_b$, and in the direction of decreasing electric field if $n < n_b$. A sphere immersed in water and irradiated with a Gaussian beam at an optical wavelength will experience a radial force toward the axis of the beam if its refractive index n is greater than about 1.33. However, there is also a force of radiation pressure in the direction of propagation of the beam (Problem 14.15):

$$\mathbf{F}_{\text{rad}} = \frac{n_b^2}{4\pi\epsilon_0} \frac{\omega^4}{3c^3} \alpha_R^2(\omega) \mathbf{E}_0^2 = \frac{8\pi}{3c} n_b^5 \left(\frac{\omega}{c} \right)^4 \left(\frac{n^2 - n_b^2}{n^2 + 2n_b^2} \right)^2 a^6 I. \quad (14.5.13)$$

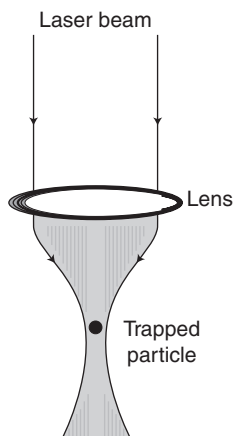


Figure 14.21 “Optical tweezer” used in microscopy. A particle is trapped where the field intensity is greatest.

In order to trap a particle with $n > n_b$ both radially and axially with a focused Gaussian beam, for example, the axial gradient of the field in the direction of the focal spot must be large enough that the dipole force (14.5.11) acting to push the particle toward the focus exceeds the force $\mathbf{F}_{\text{rad}}$ acting to push it away. This condition is not difficult to satisfy. Figure 14.21 shows a commonly used setup for trapping a dielectric particle with a single laser beam, an example of an *optical tweezer*. (The terminology comes from the use of laser beams to noninvasively “take hold” of a particle and move it from one place to another.) Diode-pumped Nd:YAG lasers are commonly used in biological applications of laser tweezers; the 1064-nm wavelength is not absorbed by water and does not damage biological samples. The frequency-doubled Nd:YAG radiation at 532 nm, by contrast, is strongly absorbed by many samples and can serve as an “optical scissors” used in conjunction with an optical tweezer. Optical tweezer technology and its applications are advancing rapidly and, as with other applications of lasers, we cannot go into it in any depth without straying into topics quite distinct from laser physics as such.³⁰

14.6 BOSE–EINSTEIN CONDENSATION

The classical mechanics of an object with linear momentum p is valid if the de Broglie wavelength $\lambda_{\text{deB}} = h/p$ is sufficiently small, just as ray optics well describes some aspects of the propagation of light when the wavelength λ [equal to the photon de Broglie wavelength $h/(h\nu/c)$] is small compared to apertures and other things affecting the propagation. For an ideal gas with N atoms per unit volume at temperature T , for example, the wave nature of the atoms is not expected to play any role unless their de Broglie wavelengths are comparable to the interatomic spacing $\sim N^{-1/3}$. Using

³⁰A large list of papers on optical tweezers, with succinct commentaries, has been compiled by M. J. Lang and S. M. Block, *American Journal of Physics* **71**, 201 (2003).

$p^2/2M = \frac{3}{2}k_B T$ to calculate λ_{deB} for atoms of mass M , we estimate therefore that the temperature must be less than

$$T_c \sim \frac{1}{3} \frac{h^2}{Mk_B} N^{2/3} \quad (14.6.1)$$

if the wave nature of the atoms is to be important. A gas below this critical temperature might be expected to behave as a single “matter wave” rather than as a collection of independently moving atoms, much as laser light is better described as a wave rather than as a collection of independently propagating photons. The possibility that particles of *matter* can behave *collectively* as a wave exhibiting interference and diffraction effects follows from a prediction in 1924–1925 by Einstein, who showed that, at sufficiently low temperatures, *noninteracting* particles can “condense” into a single quantum state of zero velocity. This is called *Bose–Einstein condensation* (BEC).

The critical temperature T_c for BEC is generally extremely small. Consider, for example, a gas of ^{87}Rb atoms at a density $N = 2.5 \times 10^{18} \text{ m}^{-3}$, about 10^{-7} times the density of air at standard temperature and pressure (STP). Using (14.6.1) we calculate $T_c \sim 130 \text{ nK}$, much smaller even than the sort of temperatures quoted in our discussions of Doppler and Sisyphus cooling. Nevertheless, as discussed below, BEC has in fact been realized in atomic clouds, including ^{87}Rb at the density $2.5 \times 10^{18} \text{ atoms/m}^3$.

- Einstein deduced the condensation effect along the lines of the following simplified argument. Consider the number of possible quantum states $d\mathcal{N}_E$ of a particle with energy in the interval $[E, E+dE]$ in an ideal gas of volume V . This is analogous to the number (3.12.12) of field modes; in the case here of particles of mass M , all occupying the same internal quantum state (e.g., the same electronic state of an atom),

$$d\mathcal{N}_E = \frac{1}{(2\pi)^3} V d^3k = \frac{V}{(2\pi)^3} 4\pi k^2 dk. \quad (14.6.2)$$

This is derived in the same fashion as (3.12.12), except that (i) no factor of 2 associated with photon polarization appears, and (ii) k here is related to the energy E by $E = p^2/2M = \hbar^2 k^2/2M$, i.e., by relating a particle’s momentum p to its de Broglie wavelength ($p = 2\pi\hbar/\lambda_{\text{deB}} = \hbar k$). It follows that $d\mathcal{N}_E = \rho(E) dE$, where the “density of states” is

$$\rho(E) = \frac{VM^{3/2}}{\pi^2 \hbar^3 \sqrt{2}} E^{1/2}. \quad (14.6.3)$$

If the atoms are bosons, the number of particles $\mathcal{N}$ at temperature T is

$$\mathcal{N} = \mathcal{N}_0(T) + \int_0^\infty \frac{\rho(E) dE}{z(T)e^{E/k_B T} - 1}. \quad (14.6.4)$$

The first and second terms on the right-hand side are the numbers of particles with energy $E = 0$ and $E > 0$, respectively. Because of $z(T)$, whose functional form we will not require, the denominator in the second term differs from the $(e^{E/k_B T} - 1)^{-1}$ familiar from the case of thermal photons. $z(T)$ appears because, unlike the number of photons in thermal equilibrium, the total number of particles (atoms) of an ideal gas in thermal equilibrium is a conserved quantity; $z(T)$ is *determined* by this number. Since $[z(T)e^{E/k_B T} - 1]^{-1}$ represents the average number of particles in a particular state of energy $E > 0$, and as such must be positive, we must have $z(T) \geq 1$. It then

follows from (14.6.3) that the integrand of the second term in (14.6.4) vanishes when $E = 0$: as stated, the integral is the density of particles with energy E *greater than* 0.

The Bose–Einstein “condensation” of particles into the state with $E = 0$ occurs only at very low temperatures. It can be seen from (14.6.4) that, as the temperature decreases, $z(T)$ must also decrease in order to keep the total number of particles fixed. As $T \rightarrow 0$, $z(T) \rightarrow 1$, its smallest possible value and the value that maximizes the number of particles with $E > 0$. Assuming $z(T_c) \cong 1$, we define the critical temperature T_c for BEC as the temperature at which $E > 0$ for *all* the particles:

$$\mathcal{N} = \int_0^\infty \frac{\rho(E) dE}{e^{E/k_B T_c} - 1} = \frac{VM^{3/2}}{\pi^2 \hbar^3 \sqrt{2}} \int_0^\infty \frac{\sqrt{E} dE}{e^{E/k_B T} - 1} = 2.612V \left(\frac{2\pi M k_B T_c}{h^2} \right)^{3/2}, \quad (14.6.5)$$

or

$$T_c = 0.084 \frac{h^2}{M k_B} N^{2/3} \quad (N = \mathcal{N}/V). \quad (14.6.6)$$

Below this temperature *some* of the particles will be in the condensed phase [$\mathcal{N}_0(T) > 0$]. T_c as defined here is about $\frac{1}{4}$ the cruder estimate (14.6.1). At temperatures $T < T_c$,

$$\begin{aligned} \mathcal{N} &= \mathcal{N}_0(T) + \int_0^\infty \frac{\rho(E) dE}{e^{E/k_B T} - 1} = \mathcal{N}_0(T) + 2.612 \left(\frac{2\pi M k_B T}{h^2} \right)^{3/2} \\ &= \mathcal{N}_0(T) + 2.612 \left(\frac{2\pi M k_B T_c}{h^2} \right)^{3/2} \left(\frac{T}{T_c} \right)^{3/2} = \mathcal{N}_0(T) + \mathcal{N} \left(\frac{T}{T_c} \right)^{3/2}. \end{aligned} \quad (14.6.7)$$

The fraction of particles in the condensed state is therefore

$$\frac{\mathcal{N}_0(T)}{\mathcal{N}} = 1 - \left(\frac{T}{T_c} \right)^{3/2}. \quad (14.6.8)$$

The length

$$\lambda_T = \left(\frac{h^2}{2\pi M k_B T} \right)^{1/2} \quad (14.6.9)$$

is called the *thermal de Broglie wavelength*, and $N\lambda_T^3$ is called the *phase-space density*. From (14.6.7) it follows that the condition $T < T_c$ for BEC can be expressed in terms of the phase-space density:

$$N\lambda_T^3 > 2.612. \quad (14.6.10)$$

Usually $z(T)$ is written as $e^{-\mu/k_B T}$, where μ , the *chemical potential*, is a thermodynamic quantity defined as the change in the energy of a system when a particle is added while the volume and entropy are kept constant. In the case of a BEC at a fixed temperature, the number of particles with energy $E > 0$ is constant [cf. Eq. (14.6.4)], so that any added particles must become part of the zero-velocity condensate. In other words, the chemical potential is nearly zero and therefore $z(T) \cong 1$ for a BEC, as we have assumed in our derivation of T_c . In Einstein’s work a parameter related to $z(T)$ appeared via a “Lagrange multiplier” used in imposing the constraint that the particle number is constant. Einstein suggested that condensation might be observed with a gas of electrons—this was before it was understood that particles are either bosons or fermions, that no two fermions can occupy the same quantum state, and therefore that the condensation effect cannot occur with (unpaired) electrons or any other fermions. •

Before laser cooling and trapping, experimental studies of Bose–Einstein condensates were limited primarily to liquid ^4He . At the required temperatures most substances solidify, and in those that do not the atoms interact too strongly to approximate an ideal gas. Liquid ^4He , the superfluidity of which has traditionally been associated with Bose–Einstein condensation, is the major exception. The small mass and therefore large zero-point energy of helium atoms prevents solidification at atmospheric pressures, and below a critical temperature of about 3K, some fraction of the atoms go into a superfluid state in which the liquid has zero viscosity (Problem 14.16). At lower temperatures nearly all the atoms are in the superfluid component. While many studies strongly support the conclusion that superfluid ^4He is a Bose–Einstein condensate, the density is sufficiently large that interatomic interactions are significant, as evidenced by the mere fact that low-temperature ^4He is a liquid. One of the most remarkable features of Einstein’s prediction, after all, is that condensation can occur in the absence of any particle interactions.

In dilute atomic clouds at temperatures below that required for Bose–Einstein condensation, the average distance between atoms is fairly large, typically $\sim 10^2$ nm. At such separations, interactions resulting in collisions that are deleterious to BEC are weak if the atoms are spin polarized, as discussed below. However, these weak interactions actually turn out to be essential for meeting the major experimental challenge, which has been to realize the exceedingly small critical temperatures that, because of the low densities of dilute gases, are much smaller even than those obtained by Sisyphus cooling, as noted earlier. If we assume a density of 10^{18} atoms/ m^3 in an atom trap, the critical temperature predicted by Eq. (14.6.6) is about 20 nK for ^{87}Rb , for example, whereas the recoil limit is about 400 nK. Actually, as also discussed below, critical temperatures for trapped atoms can be considerably larger than that given by (14.6.6), but they are still much smaller than temperatures reached by laser cooling.

The temperatures required for BEC in atomic clouds are realized by *evaporative cooling*. In this technique the six cooling laser fields (Section 14.4) are turned off and the low-field-seeking atoms are magnetically trapped; the lasers are turned off in order to avoid photon scattering processes that limit the achievable densities and temperatures. The magnetic trapping field is then slowly reduced, so that the (Zeeman energy) depth of the magnetic trap is lowered, allowing atoms with sufficient kinetic energy to escape the trap while cooler atoms remain. An rf magnetic field can also be applied to induce transitions between magnetic substates such that atoms with the largest kinetic energies are changed from low-field seekers to (untrapped) high-field seekers. The effects of the reduced trapping field and the rf field are then to allow hotter atoms to be removed from the cloud while cooler ones remain, analogous to the cooling of a cup of coffee by evaporation; note that atom interactions in the form of elastic collisions are essential for evaporative cooling, that is, for keeping the cloud thermalized as the evaporation proceeds. Evaporative cooling from a laser-cooled temperature of ~ 10 μK to ~ 200 nK, sufficient for the formation of a spin-polarized condensate of about 2000 ^{87}Rb atoms, was first achieved in 1995; a zero-velocity BEC component of the cloud was verified by a time-of-flight measurement of the atoms’ velocity distribution when the cloud was allowed to freely expand. Following those experiments, BEC has been demonstrated by similar methods in different gases. The basic technique has usually been to use Zeeman slowing to cool an atomic beam from an oven, capture and laser-cool the slowed atoms in a MOT, turn off the lasers and magnetically trap

the atoms, and then evaporatively cool the atoms (typically a few million of them) to below the critical BEC temperature.³¹

A serious problem that had to be overcome in the first experiments concerned the trapping in the static magnetic quadrupole field after the lasers were turned off. An atom moving in such a field experiences a time-dependent field that can cause it to make a transition from a low-field-seeking magnetic substate to an untrapped, high-field-seeking state if this time variation is sufficiently rapid, that is, if the magnetic field the atom sees has a Fourier frequency component near the frequency for a transition between two Zeeman-shifted magnetic substates. Near the center of the trap the magnetic field and therefore the Zeeman splittings are small enough for even a slow atom to experience a time-dependent field that causes its magnetic moment to flip into an untrapped state: The quadrupole trap has in effect a “hole” near its center that prevents atoms from being trapped for very long. In one of the original experiments this problem was solved by applying an oscillatory magnetic field such that the total (time-averaged) magnetic field did not vanish anywhere in the trap and therefore there was no “hole.” In another a “blue-tuned” laser at a frequency giving a negative atomic polarizability was applied such that atoms were repelled from the hole. Experiments that followed have employed different magnetic field configurations to circumvent the problem.

- To derive the critical temperature T_c for atoms in a magnetic trap with restoring forces along the x , y , and z axes and atom oscillation frequencies ω_x , ω_y , and ω_z along these axes, we first recall that the quantized energy levels for the atomic motion in the trap are

$$E = (n_x + \frac{1}{2})\hbar\omega_x + (n_y + \frac{1}{2})\hbar\omega_y + (n_z + \frac{1}{2})\hbar\omega_z \quad (n_x, n_y, n_z = 0, 1, 2, \dots). \quad (14.6.11)$$

Consider the number $\mathcal{N}_E$ of possible states with energies less than E . For cases of interest in the trapping of neutral atoms, the energies may be assumed to be much greater than the zero-point energies $\frac{1}{2}\hbar\omega_x$, $\frac{1}{2}\hbar\omega_y$, and $\frac{1}{2}\hbar\omega_z$. Then for the calculation of $\mathcal{N}_E$ the integers n_x , n_y , and n_z may be replaced by continuous variables $E_x/\hbar\omega_x$, $E_y/\hbar\omega_y$, and $E_z/\hbar\omega_z$:

$$\mathcal{N}_E = \frac{1}{\hbar^3 \omega_x \omega_y \omega_z} \int_0^E dE_x \int_0^{E-E_x} dE_y \int_0^{E-E_x-E_y} dE_z = \frac{E^3}{6\hbar^3 \omega_x \omega_y \omega_z}. \quad (14.6.12)$$

The number of states with energy between E and $E + dE$ is therefore

$$d\mathcal{N}_E = \frac{E^2 dE}{2\hbar^3 \omega_x \omega_y \omega_z} = \rho(E) dE, \quad (14.6.13)$$

which defines the density of states $\rho(E)$. Proceeding now as in (14.6.5)–(14.6.8), using this density of states instead of the density of states (14.6.3) for untrapped particles, we obtain the critical temperature

$$T_c = 0.94 \frac{\hbar(\omega_x \omega_y \omega_z)^{1/2}}{k_B} \mathcal{N}^{1/3} \quad (14.6.14)$$

³¹For references to the original work and more detailed discussions see, for instance, C. J. Pethick and H. Smith, *Bose–Einstein Condensation in Dilute Gases*, Cambridge University Press, Cambridge, 2004.

and the condensate fraction

$$\frac{\mathcal{N}_0(T)}{\mathcal{N}} = 1 - \left(\frac{T}{T_c}\right)^3. \quad (14.6.15)$$

Unlike the critical temperature (14.6.6) for free particles, (14.6.14) depends on the number of particles rather than the density. In the original ^{87}Rb experiments the critical temperature was about 170 nK.

Another important consideration in the physics of cold atomic gases is spin polarization. BECs in magnetic traps normally consist of spin-polarized atoms, e.g., alkali atoms in which the valence electrons have “spin up.” The atom interaction involving spin-up (or spin-down) electrons differs from the interaction when the spins are opposite because the Pauli principle allows two electrons in different spin states to occupy the same atomic orbital, as occurs in covalent bonding. If the two electrons have the same spin, however, the Pauli principle forbids them from sharing the same orbital, and the atom–atom interaction in this case is consequently weaker. Of particular interest for BEC are the so-called *doubly polarized* and *maximally stretched* states of the electronic ground state; spin-depolarizing collisions of atoms in these states have very small cross sections and the atoms are low-field seekers. In the doubly polarized state the spin components of the electron (m_J) and the nucleus (m_I) are aligned and have their largest allowed values, for example, $m_I = \frac{3}{2}$ and $m_J = \frac{1}{2}$ for Na or ^{87}Rb (nuclear spin $I = \frac{3}{2}$); the quantization axis is in the direction of the magnetic field. Thus, for Na or ^{87}Rb , both of which have $F = 1$ and $F = 2$ hyperfine ground levels, the doubly polarized state has the quantum numbers $F = 2, m_F = 2$. In the maximally stretched state, $F = I - \frac{1}{2}$ and $m_F = -(I - \frac{1}{2})$, for example, $F = 1, m_F = -1$ for Na or ^{87}Rb . Collisions of atoms in these states are primarily the “good” collisions necessary to keep the gas thermalized during evaporative cooling, as opposed to the “bad” collisions that act to quench the low-field-seeking trapped states. •

The technique of evaporative cooling was developed about a decade prior to its playing an essential role in the first observations of BEC in dilute gases. It was conceived for the cooling of magnetically trapped hydrogen, which was considered at the time the most promising candidate for BEC of an atomic gas; this consideration stemmed from the fact that the interactions of spin-polarized hydrogen atoms are very weak, preventing the formation of hydrogen molecules. The longest-wavelength transition from the ground state of hydrogen is the Lyman- α line at 122 nm, too short for laser cooling with available lasers. This short wavelength and the small mass of the H atom also imply too large a recoil temperature ($T_{\text{recoil}} = 1.3$ mK) for BEC (Problem 14.16). BEC in magnetically trapped hydrogen was achieved in 1998 by cryogenic cooling followed by evaporative cooling. The BEC transition was observed at $T_c = 50$ μK at a density of 1.8×10^{20} atoms/ m^3 . The presence of a condensate was inferred from a large shift in the absorption frequency of the two-photon $1S \rightarrow 2S$ transition (Problem 14.9).

Atoms in a Bose–Einstein condensate can be “loaded” into an optical lattice from a magnetic trap by applying off-resonance laser beams to produce the desired lattice spacings and light-shift potential depths. With the magnetic trapping fields turned off, the atoms are left trapped solely by the laser fields forming the lattice. The potential depth is usually sufficiently large that, according to classical mechanics, the atoms cannot cross the potential energy barrier between sites and move about the lattice. However, according to quantum theory there is a nonzero probability that they can “tunnel” across the barriers, and in the case of a Bose–Einstein condensate this can result in an atomic matter wave whose phase is approximately constant while the number of atoms at

each lattice site fluctuates. This is analogous to a coherent light wave in which the phase but not the number of photons is fixed (Chapter 12): The smaller the fluctuations in the phase, the greater the fluctuations in the photon number, and vice versa. In the case of a fixed number of atoms at each site of an optical lattice, the phase of the matter wave is not well defined and, by analogy to a similar situation for electrons in a crystal, the atoms are said to form a *Mott insulator*. In this case, no interference effects involving a coherent matter wave are observed when the gas is released from the lattice. If, on the other hand, the phase of the matter wave is well defined while the number of atoms at each site fluctuates, matter–wave interference effects can be observed. By changing the potential depth of an optical lattice it has been possible to observe the *Mott transition* between a Mott insulator and the (“superfluid”) case of coherent matter waves exhibiting interference. It has also been demonstrated that evaporative cooling leading to a Bose–Einstein condensate can be done with atoms trapped in an optical lattice, without any magnetic trapping field. In this case the evaporative cooling results from lowering the power of the laser beams defining the lattice.

Two-slit interference and other effects of coherent matter waves have been observed in experiments with optical lattices and other cold-atom systems, and the new field of *atom optics* is based on the wave properties of cold atoms. Atomic interferometers employing atom–wave analogs of mirrors, gratings, and other optical elements will very likely be used in applications such as the detection of rotations and gravitational gradients.³²

Bose–Einstein condensates have been produced by trapping atoms with magnetic fields from planar wire structures fabricated lithographically on a substrate. The small scale ($\sim 0.1 - 10 \mu\text{m}$) of these structures allows strong magnetic field gradients to be produced with small currents ($< 1 \text{ A}$) and low power dissipation, and the wire patterns can be designed to form, for instance, magnetic traps similar to the Helmholtz coil quadrupole trap described earlier. The tight trapping results in rapid cooling and the formation of a BEC in a second or less. The motion and positions of atoms trapped a few microns above the surface of these structures are more controllable than in the case of MOTs in which the trapping fields are produced by lasers and coils outside the atomic cloud. For example, it has been demonstrated that a BEC cloud can be moved about in a prescribed fashion on such an *atom chip*. Integrated atom chips including tiny lasers and readout electronics might eventually be the basis of miniaturized sensors, atomic clocks, and other devices.

- Atoms in which the numbers of electrons, protons, and neutrons are even (odd) integers are bosons (fermions). Neutral atoms (number of electrons equal to the number of protons) are therefore bosons or fermions depending on whether the number of neutrons is even or odd, respectively. Thus, ^7Li (atomic number $Z = 3$, neutron number $N = 4$) is a boson, whereas ^6Li ($Z = 3$, $N = 3$) is a fermion. It is a remarkable consequence of quantum statistics that a Bose–Einstein condensate can be made with ^7Li but not with ^6Li , which differs from ^7Li only in having one less neutron in the nucleus.

While fermionic atoms cannot form BECs, they can be cooled and trapped with lasers and magnetic fields. Techniques have been developed to control with applied fields the interatomic forces in cold fermion gases in such a way that pairs of atoms behave compositely as bosons, somewhat similar to the way coupled pairs of electrons behave as bosons in a superconductor. •

³²See, for instance, M. A. Kasevich, *Science* **298**, 1363 (2002) for an overview of matter–wave experiments and applications of atom interferometry.

14.7 APPLICATIONS OF ULTRASHORT PULSES

Ultrashort laser pulses make possible an astounding degree of time resolution, and the extremely high intensities that often characterize them are of interest for basic research as well as for a growing number of applications. In this section brief introductions are given to three new fields created by ultrashort laser pulses.

Time Resolution of Atomic and Molecular Processes

In the 1870s the photographer Eadweard Muybridge carried out experiments on the Palo Alto, California, farm of Leland Stanford to answer the question whether all four hooves of a horse are off the ground at any point in its trot. Using a series of cameras whose shutters were triggered by strings placed along a track, and later a periodic mechanical triggering, Muybridge eventually achieved millisecond time resolution—probably the record at the time—and found that the answer to the question was Yes.

Great progress in the resolution of short-duration events followed advances in electronics and other areas. But before femtosecond lasers³³ there was no way of following in real time such things as the vibrations of atoms in molecules or the detailed time evolution of a chemical reaction. Chemical reactions at room temperatures typically involve atomic or molecular velocities $\sim 10^3$ m/s and distances $\sim 10^{-10}$ m, implying a time scale $\sim (10^{-10} \text{ m}) / (10^3 \text{ m/s}) = 100$ fs. Observation of the evolution of a reaction therefore requires a time resolution of about $1/10 \times 100 \text{ fs} = 10$ fs, about 1000 times shorter than that possible with the fastest available electronics.

However, resolution on femtosecond time scales is still possible. We saw an example of this in Section 11.13, where we described how the durations of femtosecond laser pulses can be inferred from measurements of a field autocorrelation function. The “event” of a laser pulse is recorded via an interference of the pulse with itself: The pulse is used to measure its own time evolution. But how can femtosecond pulses be used to time-resolve processes occurring in and among atoms and molecules? To illustrate how this is done, we will focus on the vibrations of a diatomic molecule.

Recall that the potential energy function $V(R)$ for the bonding of the atoms of a diatomic molecule is well approximated near its minimum by a parabola, that is, by a harmonic-oscillator potential [cf. Fig. 2.6]. The vibrational energy levels are therefore approximately those of a harmonic oscillator [Eq. (2.3.8)]; anharmonic effects are accounted for by formula (2.3.12), where the frequency ω_e characterizes the molecule and the vibrational mode. To the extent that the vibrations are harmonic, we can picture a diatomic molecule as two atoms whose separation R oscillates sinusoidally in time with a period $1/\omega_e$ that is typically $\sim 10^{-13}$ s.

Figure 14.22 shows potential energy functions $V_0(R)$, $V_1(R)$, and $V_2(R)$ for three electronic levels 0, 1, and 2 of a hypothetical diatomic molecule. A field of frequency ν_{pump} can induce transitions between vibrational states of electronic levels 0 and 1, and a field of frequency ν_{probe} can induce transitions between vibrational states of electronic levels 1 and 2, as indicated. In accordance with the Franck–Condon principle (Section 11.11), transitions are indicated by vertical lines, i.e., they occur without a change in the

³³“Femtosecond” here refers to pulse durations between roughly 1 fs and 0.1 ps = 100 fs.

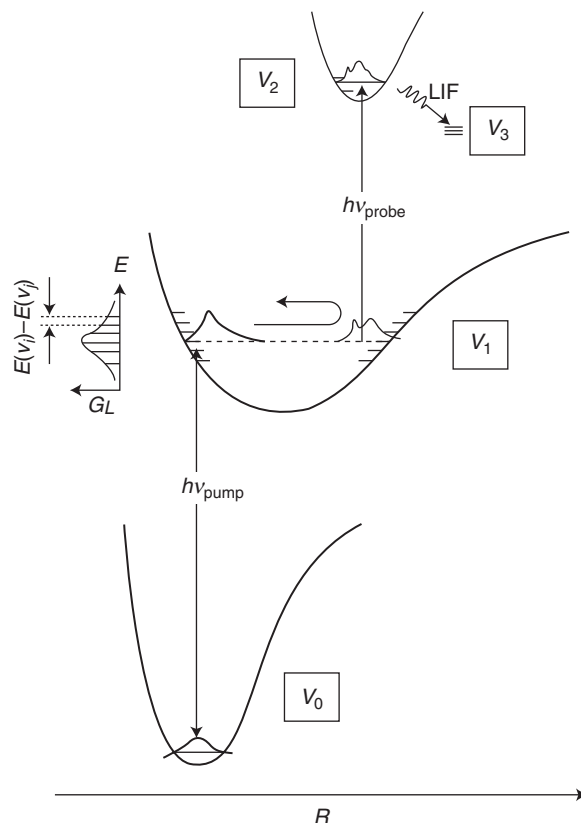


Figure 14.22 Time resolution of molecular vibrations using femtosecond pump and probe pulses. The pump pulse excites a molecule to an electronic level 1 with potential energy function $V_1(R)$, and after some time delay τ the probe pulse is applied to excite the molecule to a level 2 that decays by fluorescence to level 3. During the time τ the interatomic separation R changes, so that the strength of the fluorescence signal depends on τ . The variation with τ of this signal therefore reflects the variation with τ of the distance between the atoms. [From M. Gruebele and A. H. Zewail, *Journal of Chemical Physics* **98**, 883 (1993).]

interatomic separation R . Excitation to the electronic level 2 by absorption of radiation at ν_{pump} followed by absorption of radiation at ν_{probe} can be monitored by detection of the fluorescence accompanying transitions from 2 to some other level 3 (“laser-induced fluorescence,” indicated by LIF in Fig. 14.22).

Suppose a laser pulse at frequency ν_{probe} is applied after some time τ following a pulse at frequency ν_{pump} . During this time the interatomic separation R of a vibrating molecule in electronic level 1 will have changed. If the durations of the pump and probe pulses are short compared to the vibrational period, the variation with τ of the LIF from level 2 signifies the variation of R with τ . Oscillations of the interatomic separation R are reflected in oscillations in the observed LIF signal. If, as estimated above, the period of the molecular vibrations in electronic level 1 is 100 fs, these vibrations can be observed in “real time” using pump and probe pulses of duration less than 100 fs. This is the basic idea behind time-resolved studies of atomic and

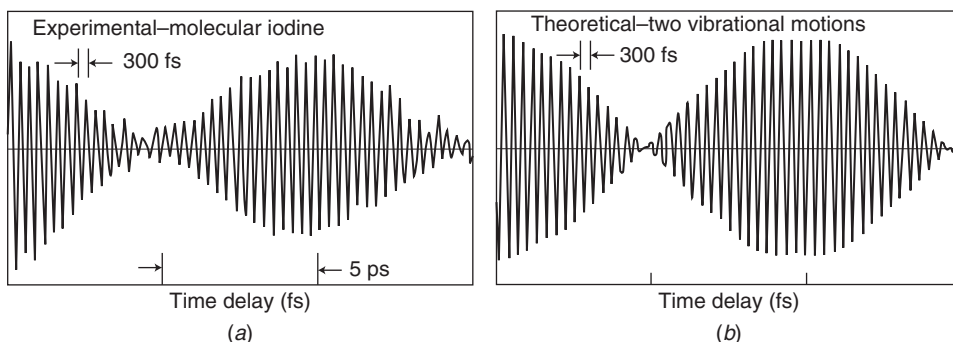


Figure 14.23 (a) Femtosecond pump-probe data showing vibrations of molecular iodine with an oscillation period of 300 fs. The much longer oscillation period of 10 ps is due to the anharmonicity of the vibrations. (b) Calculated signal with the two principal oscillation periods seen in (a). [From M. Dantus, R. M. Bowman, and A. H. Zewail, *Nature* **343**, 737 (1990).]

molecular processes with femtosecond lasers: Pump pulses set the “zero” of time from which temporal variations are determined with time-delayed probe pulses. The time delay τ between the pump and probe pulses can be varied by changing the difference ΔL in their path lengths: $\tau = \Delta L/c$. A path difference of a micron, for example, corresponds to a time delay of 3.33 fs.

Figure 14.23 shows the first reported time resolution of molecular vibrations. The molecule in the experiments was room temperature I_2 , for which the vibrational period of the electronic level “2” is known from spectroscopic studies to be ≈ 300 fs. The pump and probe wavelengths were 620 and 310 nm, respectively. The vibrational period is clearly seen in the data. The amplitude of the LIF signal is modulated at a longer period (10 ps) as a consequence of anharmonicity. On a much longer time scale a period ~ 600 ps is observed as a consequence of the *rotation* of the I_2 molecule.

In our simplified discussion we have imagined that the interatomic separation R follows a classical trajectory, when in fact it must be described statistically using a quantum-mechanical wave function $\Psi(R, t)$: The probability that R is between R and $R + dR$ ($dR \ll R$) at time t is $|\psi(R, t)|^2 dR$. The lowest energy ($v = 0$) vibrational stationary state (or eigenstate), for example, is described by a time-independent wave function whose squared modulus in the harmonic-oscillator approximation is a Gaussian function of R . (Recall the remarks near the end of Section 11.10.) This is indicated for the electronic level with potential curve $V_0(R)$ in Fig. 14.22. The bandwidth of the pump pulse results in nonvanishing occupation probabilities for a range of vibrational eigenstates of the electronic level 1, as also indicated in Fig. 14.22. To simplify matters, let us suppose that only two eigenstates, with vibrational quantum numbers v and $v + 1$ and wave functions $\phi_v(R)$ and $\phi_{v+1}(R)$, have significant excitation probabilities, so that the wave function describing the vibrational state of the molecule in electronic level 1 at the time $t = 0$ immediately after irradiation by the pump pulse is

$$\psi(R, 0) = a_v(0)\phi_v(R) + a_{v+1}(0)\phi_{v+1}(R). \quad (14.7.1)$$

The probability amplitudes satisfy the time-dependent Schrödinger equations

$$i\hbar\dot{a}_v(t) = E_v a_v(t) \quad \text{and} \quad i\hbar\dot{a}_{v+1}(t) = E_{v+1} a_{v+1}(t) \quad (14.7.2)$$

in the absence of any perturbation of the molecule [cf. Eq. (3.A.7)], so that $a_v(t) = a_v(0)\exp(-iE_v t/\hbar)$, $a_{v+1}(t) = a_{v+1}(0)\exp(-iE_{v+1} t/\hbar)$, and the vibrational wave function is

$$\psi(R, t) = a_v(0)e^{-iE_v t/\hbar}\phi_v(R) + a_{v+1}(0)e^{-iE_{v+1} t/\hbar}\phi_{v+1}(R) \quad (14.7.3)$$

at a time t after irradiation of the molecule by the pump pulse and before irradiation by the probe pulse. The probability distribution for the interatomic separation R is therefore

$$\begin{aligned} |\psi(R, t)|^2 &= |a_v(0)|^2 |\phi_v(R)|^2 + |a_{v+1}(0)|^2 |\phi_{v+1}(R)|^2 \\ &\quad + 2\text{Re}\left[a_{v+1}^*(0)a_v(0)\phi_{v+1}^*(R)\phi_v(R)e^{i(E_{v+1}-E_v)t/\hbar}\right]. \end{aligned} \quad (14.7.4)$$

This oscillates in time at the frequency $(E_{v+1} - E_v)/\hbar$ which, in the harmonic-oscillator approximation, is just $c\omega_e$ [Eq. (2.3.12)]. This oscillatory behavior of the probability distribution for R is indicated for the potential curve $V_1(R)$ in Fig. 14.22.

Now consider the probability amplitude for the transition to a vibrational eigenstate v' of electronic level 2 when the probe pulse is applied at a time τ after the pump pulse. Assuming it is small enough for lowest-order perturbation theory to be accurate, this amplitude is proportional to

$$\begin{aligned} \int_{-\infty}^{\infty} \phi_{v'}^*(R) D \psi(R, \tau) dR &= a_v(0)e^{-iE_v \tau/\hbar} \int_{-\infty}^{\infty} \phi_{v'}^*(R) D \phi_v(R) dR \\ &\quad + a_{v+1}(0)e^{-iE_{v+1} \tau/\hbar} \int_{-\infty}^{\infty} \phi_{v'}^*(R) D \phi_{v+1}(R) dR \\ &\equiv D_{v',v} a_v(0)e^{-iE_v \tau/\hbar} + D_{v',v+1} a_{v+1}(0)e^{-iE_{v+1} \tau/\hbar}, \end{aligned} \quad (14.7.5)$$

where $\phi_{v'}(R)$ is the wave function for the vibrational eigenstate v' for the potential energy curve $V_2(R)$ and D is an electric dipole moment operator that depends on the polarization of the probe field. The probability amplitude will of course also depend on the electric field of the probe; this dependence can be ignored for our purpose here, which is simply to note that the LIF signal intensity of interest, which is proportional to $|\int_{-\infty}^{\infty} \phi_{v'}^*(R) D \psi(R, \tau) dR|^2$ for a pump-probe delay τ , will, like (14.7.4), oscillate at the frequency $(E_{v+1} - E_v)/\hbar \cong c\omega_e$. Because the vibrational states are approximately equally spaced, this conclusion holds also if more than two vibrational states of the electronic level 1 are populated by the pump pulse. The (anharmonic) deviations from equal spacings result in additional frequency components in the LIF signal.

Quantum mechanics therefore predicts the same sort of LIF signal intensity expected from a classical trajectory picture of molecular vibrations: The signal oscillates with the pump-probe delay τ at principally the vibrational frequency $c\omega_e$, with additional frequency components due to anharmonicity. The wave function (14.7.3) depends on

the probability amplitudes [$a_v(t)$ and $a_{v+1}(t)$] of the two states described by $\phi_v(R)$ and $\phi_{v+1}(R)$ and not just their occupation probabilities [$|a_v(t)|^2$ and $|a_{v+1}(t)|^2$]. In other words, it is a coherent superposition of two states, just as the wave function (9.3.3) is a coherent superposition of the states of an atom modeled as a two-state system. In general, the broad bandwidth of a femtosecond pulse, together with the approximately equal spacing of vibrational levels, results in a so-called *coherent wave packet*:

$$\psi(R, t) = \sum_{v=0}^{\infty} a_v(t) \phi_v(R). \quad (14.7.6)$$

The number of vibrational levels with significant probability amplitudes $a_v(t)$ is typically ~ 5 – 10 in studies of the type considered here. Of course, a more realistic theory requires that rotational energy levels be included. Since rotational periods are typically $\sim 10^3$ times larger than vibrational periods, an effect of rotations is to introduce periodicities on a much longer time scale than is shown in Fig. 14.23.

Fourier analysis of time series such as the data plotted in Fig. 14.23 can be used to extract spectroscopic information about molecular vibrations and rotations, i.e., to determine vibrational and rotational constants. This *femtosecond wave packet spectroscopy* can be performed on molecules in cells as well as in beams, since collision times are so much longer than the pulse durations and delay times (Problem 14.17).

Femtosecond lasers have made it possible to follow the movements of atoms in chemical processes, creating the field of *femtochemistry*. For example, the breaking of the chemical bond between two atoms of a diatomic molecule has been time-resolved by the femtosecond pump-probe technique. A pump pulse creates a coherent wave packet that oscillates within the potential energy well characterizing the bond. If ν_{pump} is sufficiently large, the outer wings of the wave packet can escape the well, that is, the molecule can dissociate. If the time-delayed probe pulse is at the frequency for which there is absorption when the molecules are closest together, for instance, an LIF signal oscillates as the molecule vibrates, but the oscillation is damped because of dissociation. Fluorescence from a dissociation product excited by a laser at the appropriate wavelength can also track the dissociation in real time.

The femtosecond pump-probe method has been employed in studies of subpicosecond processes in liquids and solids as well as in gases. It has spawned another field, *femtobiology*, based on time-resolved studies of phenomena such as vision and photosynthesis at the molecular level.

With the advent of attosecond pulses (see below) it has become possible to observe phenomena on a time scale characterizing the motion of atomic electrons.³⁴ Because electrons have a much smaller mass than atoms, their movements can be expected, generally speaking, to be faster and therefore to be significant on much shorter time scales; the relevant time scale for an atomic electron can be roughly estimated using the Bohr model, according to which the orbital period of the electron in the ground state of hydrogen is about 150 attoseconds (as). With pump-probe techniques it is even possible to follow the oscillations in time of the electric field of a femtosecond optical pulse. The details of such measurements are complicated, but the basic concept can

³⁴See, for instance, R. Kienberger, E. Goulielmakis, M. Uiberacker, A. Baltuska, V. Yakovlev, F. Bammer, A. Scrinzi, Th. Westerwalbesloh, U. Kleineberg, U. Heinzmann, M. Drescher, and F. Krausz, *Nature* **427**, 817 (2004).

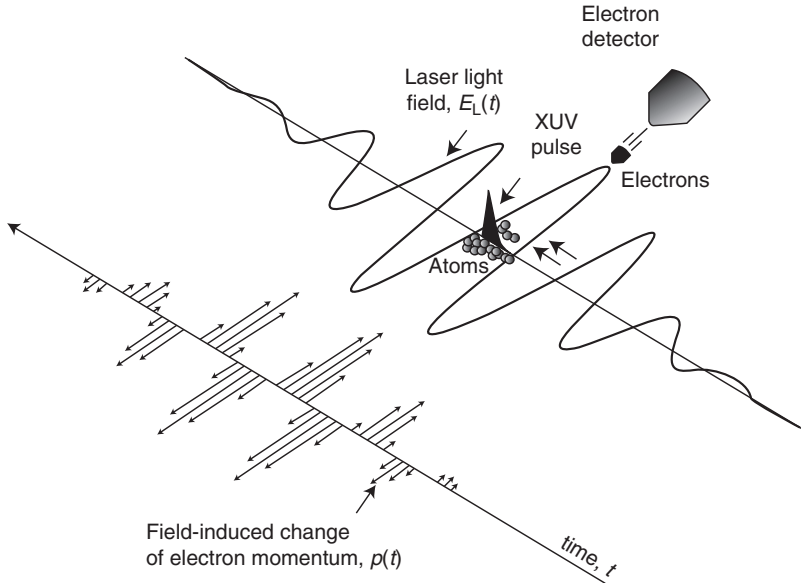


Figure 14.24 Attosecond burst of radiation (XUV pulse) photoionizes an atom, and a time-delayed femtosecond pulse then imparts a momentum impulse to the freed electron, affecting its time of flight to a detector. By repeating the TOF measurements with identical pulses and different time delays, the oscillations of the electric field of the femtosecond pulse can be determined. [From E. Goulielmakis, M. Uiberacker, R. Kienberger, A. Baltuska, V. Yakovlev, A. Scrinzi, Th. Westerwalbesloh, U. Kleineberg, U. Heinzmann, M. Drescher, and F. Krausz, *Science* **305**, 1267 (2004).]

be simply illustrated as in Fig. 14.24. The arrows in the lower left part of the figure indicate the electric field $E(t)$ of a few-cycle, linearly polarized femtosecond pulse whose oscillations are to be measured. This field is applied at a time t after the time $t = 0$ at which an extreme-ultraviolet (XUV) attosecond “burst” of radiation photoionizes an atom to produce a free electron. The field $E(t)$ does not cause any significant ionization of the atom, but it does impart to the electron a momentum impulse

$$\Delta p(t) = e \int_t^{\infty} E(t') dt'. \quad (14.7.7)$$

Depending on the sign of this impulse along the direction pointing to the electron detector in Fig. 14.24, the time of flight of the electron is either increased or decreased; in other words, time-of-flight (TOF) measurements on the electrons provide information about the *amplitude and phase* of $E(t)$, and repeated TOF measurements with identical pulses and different time delays then allow the temporal variation of $E(t)$ to be mapped out (Fig. 14.25). Thus, for two different times differing by a time δt much smaller than the period of the femtosecond field oscillations,

$$\Delta p\left(t - \frac{1}{2}\delta t\right) - \Delta p\left(t + \frac{1}{2}\delta t\right) = e \int_{t-\frac{1}{2}\delta t}^{t+\frac{1}{2}\delta t} E(t') dt' \cong e\delta t E(t), \quad (14.7.8)$$

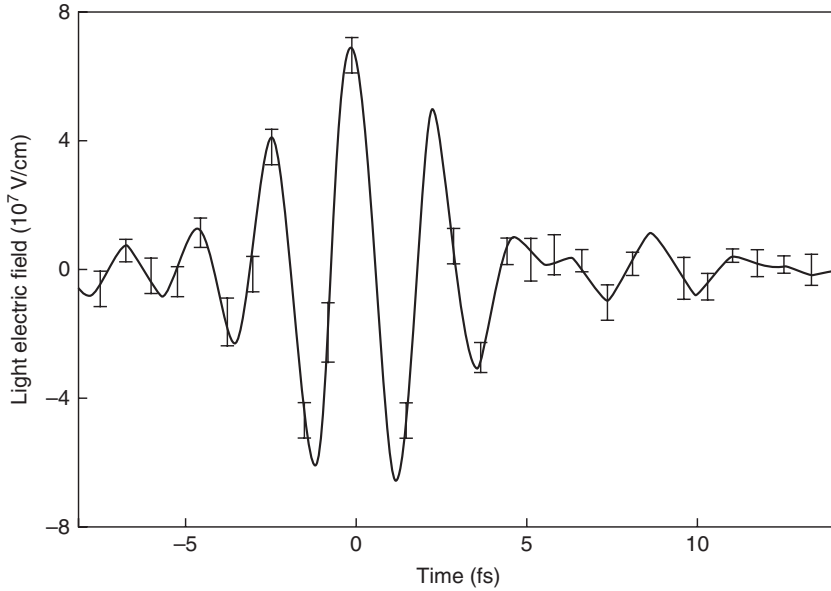


Figure 14.25 Electric field oscillations of a few-cycle femtosecond pulse, determined from electron time-of-flight data. [From E. Goulielmakis, M. Uiberacker, A. Baltuska, V. Yakovlev, F. Bammer, A. Scrinzi, Th. Westerwalbesloh, U. Kleineberg, U. Heinzmann, M. Drescher, and F. Krausz, *Science* **305**, 1267 (2004).]

or

$$E(t) \cong \frac{\Delta p(t - \frac{1}{2}\delta t) - \Delta p(t + \frac{1}{2}\delta t)}{e\delta t}, \quad (14.7.9)$$

and $E(t)$ can be determined by TOF measurements in which t is varied over the duration of the femtosecond pulse.

- Among the highly nontrivial details ignored in our discussion is the need for precise timing between the attosecond and femtosecond pulses. For the data shown in Fig. 14.25, the ionizing pulses were 250-as bursts of 13.4-nm (93-eV) XUV radiation produced via harmonic generation with 0.5-mJ, 750-nm, ~ 5 -fs Ti : sapphire laser pulses incident on a neon target (see below). The attosecond and femtosecond pulses then propagated along the same direction to a second neon target where the photoionization occurs. This second neon target, a < 50 - μm jet, was at the focus of a specially manufactured spherical mirror, such that the focused femtosecond pulse incident on the jet had a diameter > 60 μm . Because of their much longer wavelengths, the femtosecond pulses diverged more in their propagation to the mirror, spreading over a diameter on the mirror of about 25 mm compared to a much smaller spot for the attosecond pulses. Piezoelectric adjustment of the movable central part of the mirror with nanometer precision allowed the time delay between the femtosecond and attosecond pulses incident on the second neon target to be controlled with attosecond precision [$(1 \text{ nm})/c = 3.3 \text{ as}$]. •

High-Harmonic Generation

In Chapter 10 we explained various nonlinear optical phenomena by expanding the polarization $\mathbf{P}$ in powers of the electric field strength. As remarked at the end of that

chapter, this perturbative approach cannot be usefully applied when, for instance, the radiation is so intense that many harmonics of the incident radiation are generated. In this regime of “extreme nonlinear optics” the field so strongly affects an atom that it is anything but a small perturbation.

The electric fields resulting from mode locking and chirped pulse amplification can approach or exceed the fields acting on atomic electrons (Section 11.13). For the hydrogen atom, for example, the electric field at the electron is about 5×10^9 V/cm; a laser with this electric field strength would have an intensity of about 3×10^{16} W/cm², which is smaller than the intensities obtainable by mode-locking and CPA techniques (Problem 11.9). Similarly the energies of photons that can be created by harmonic generation can exceed the binding energies of atomic electrons, for example, the 13.6-eV ionization energy of the hydrogen atom. For this reason photons created in high-harmonic generation are often characterized in terms of energies in electron volts rather than their associated wavelengths or frequencies: near-IR to optical wavelengths correspond to $\sim 1\text{--}2$ eV ($\sim 1200\text{--}600$ nm), XUV wavelengths to $\sim 10\text{--}100$ eV ($\sim 120\text{--}12$ nm), and soft X rays are characterized by photon energies up to ~ 1 keV (~ 1 nm). For a Ti:sapphire laser at a wavelength of 800 nm, by comparison, the photon energy is about 1.5 eV.

It is not surprising that an extremely intense field can generate high-harmonic radiation, given that the response of matter to high-intensity radiation is nonlinear (Section 10.1 and Problem 10.1). The electric field from an atom is proportional to $d^2\mathbf{p}/dt^2$ in the radiation zone, where $\mathbf{p}(t)$ is the induced electric dipole moment. Integrating the equations determining $\mathbf{p}(t)$ in the model of a two-state atom in a field of frequency ω , for example, and then computing the Fourier transform $S(\Omega)$ of the field radiated by $\mathbf{p}(t)$, one finds for sufficiently high intensities that $S(\Omega)$ has peaks at harmonics $\Omega = N\omega$ for a large set of odd integers N . By performing such an exercise, one can also demonstrate the inefficacy of an approach based on a perturbative expansion of $\mathbf{p}(t)$ in powers of the electric field.

High-harmonic generation (HHG) has most often been achieved by focusing laser pulses onto a small cell or jet of atoms with a high ionization potential, i.e., an inert gas. Figure 14.26 shows an HHG spectrum measured in one of the earliest such experiments. The basic features of this harmonic distribution have been amply confirmed in many later experiments: There is a steep decline in the intensity of the first few harmonics, followed by a plateau of roughly comparable intensities and then a sharp cutoff beyond which no higher harmonics are observed. Note also that only odd harmonics appear, as expected from inversion symmetry (Chapter 10).

The development of Ti:sapphire systems yielding extremely intense pulses has greatly extended the range of HHG photon energies, so much so that compact femtosecond laser sources are now used to generate coherent (laserlike) XUV and soft X-ray radiation for various applications. The photon energies are limited by the cutoff feature of HHG spectra, e.g., $N = 33$ in Fig. 14.26. Experimental results have been consistent with the following “universal” cutoff, i.e., the largest photon energy achievable by HHG in a gas of atoms with ionization potential I_p :

$$E_{\max} = I_p + 3.17U_p, \quad (14.7.10)$$

where U_p is the *ponderomotive energy*, defined as $e^2 E_0^2 / 4m\omega^2$, where E_0 is the electric field strength of the laser radiation at the fundamental (angular) frequency ω and e and m

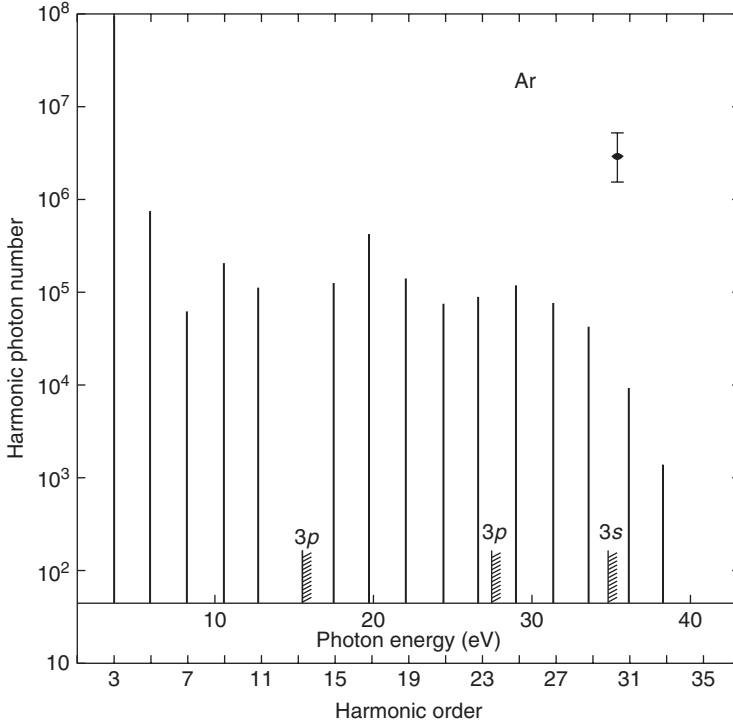


Figure 14.26 Distribution of harmonics in an early experiment on high-harmonic generation. The harmonics were generated when $3 \times 10^{13} \text{ W/cm}^2$ mode-locked Nd:YAG pulses were focused onto a 15-Torr, $\sim 1\text{-mm}$ argon jet. $3s$, $3p$, and $3p^+$ indicate ionization energies for levels of Ar and Ar^+ . [From X. F. Li, A. L’Huillier, M. Ferray, L. A. Lompré, and G. Mainfray, *Physical Review A* **39**, 5751 (1989).]

are the electron charge and mass, respectively. U_p is the cycle-averaged “quiver energy” of an electron in an electric field $\mathbf{E}_0 \cos \omega t$:

$$m\ddot{\mathbf{x}} = e\mathbf{E}_0 \cos \omega t, \quad v(t) = \dot{\mathbf{x}}(t) = \frac{e\mathbf{E}_0}{m\omega} \sin \omega t, \quad \frac{1}{2}mv^2(t) = \frac{e^2E_0^2}{2m\omega^2} \sin^2 \omega t, \\ U_p \equiv \frac{e^2E_0^2}{4m\omega^2} = 0.93 \times 10^{-13} I \lambda^2 \text{ eV}, \quad (14.7.11)$$

where I (W/cm^2) is the intensity and λ (μm) is the wavelength of the presumed monochromatic field. The relation (14.7.10) has a surprisingly simple explanation, as follows.³⁵

At the high intensities needed for HHG we can expect ionization. It is not necessary for our purposes to delve into the quantum theory of the ionization process; in fact we will describe the freed electron using classical mechanics, assuming that at the instant t_i at which ionization occurs, the electron is momentarily at rest at the position $x = 0$

³⁵K. C. Kulander, K. J. Schafer, and J. L. Krause, in *Super-Intense Laser-Atom Physics*, eds. B. Piraux, A. L’Huillier, and K. Rzażewski (Plenum, New York, 1993), p. 106; P. B. Corkum, *Physical Review Letters* **71**, 1994 (1993).

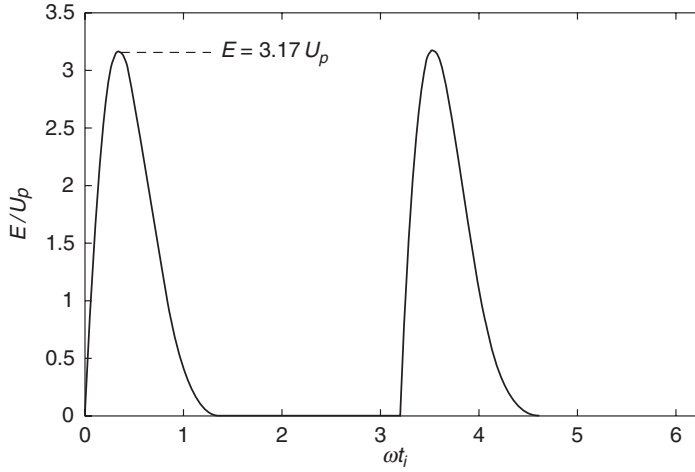


Figure 14.27 Numerical results from Eqs. (14.7.12) for the kinetic energy of an electron at the moment it returns to the parent ion after ionization at time t_i . The largest kinetic energy the electron can have when it hits the ion is computed to be $3.17U_p$.

of the ion in a field $E_0 \cos \omega t$ linearly polarized along the x direction. Ignoring any effect of the parent ion, we describe the electron's motion at times $t > t_i$ by the classical Newton equation, $m\ddot{x} = eE_0 \cos \omega t$, which gives

$$v(t, t_i) = \dot{x}(t, t_i) = \frac{eE_0}{m\omega}(\sin \omega t - \sin \omega t_i), \quad (14.7.12a)$$

$$x(t, t_i) = -\frac{eE_0}{m\omega^2}[\cos \omega t - \cos \omega t_i + \omega(t - t_i) \sin \omega t_i] \quad (14.7.12b)$$

for the chosen initial conditions. Equation (14.7.12b) can be solved for the time $t_f (> t_i)$ at which an electron freed from the atom at time t_i and “quivering” in the field can collide with the ion, that is, the time t_f for which $x(t_f, t_i) = 0$. The velocity at which the electron strikes the ion in this classical model is $v(t_f, t_i)$, and the kinetic energy is $E(t_f, t_i) = \frac{1}{2}mv^2(t_f, t_i)$. Thus, for any particular ionization time t_i we can calculate the time t_f at which the freed electron can “recollide” with the ion, and from that the electron's kinetic energy $E(t_f, t_i)$ at the moment of impact (Problem 14.19). Figure 14.27 shows numerical results for $E(t_f, t_i)$ obtained in this way for values of ωt_i between 0 and 2π . For ωt_i such that $E(t_f, t_i) = 0$ in the figure the freed electron never returns to the ion.³⁶

The largest possible value of $E(t_f, t_i)$ is $3.17U_p$, which is found to occur at $\omega t_i \cong 0.30 \text{ rad} = 17^\circ$; the corresponding value of t_f is given by $\omega t_f \cong 4.45 \text{ rad}$. Thus, $t_f - t_i = (4.45 - 0.30)/\omega = 1.8 \text{ fs}$ for an 800-nm Ti:sapphire pulse, compared with the 2.7-fs period of a monochromatic wave at 800 nm. In reality, of course, the electric field is not periodic, and during a few-cycle pulse, for instance, the electron can return to the ion a few times at most. The main point for our purposes is that *in an initial recollision the electron's kinetic energy cannot exceed $3.17U_p$* . For a gas of atoms in which these ionizations and recollisions occur, the kinetic energies vary over the distribution

³⁶We are ignoring any motion of the electron transverse to the direction of field polarization. Such motion will reduce the number of electrons that actually recollide with their parent ions.

of ionization times, but the kinetic energy with which any electron can hit its parent ion cannot exceed $E_{\max} = 3.17U_p$.

We have avoided any analysis of the ionization process, and likewise we will not analyze in any detail what happens when the electron collides with the ion. It can scatter off the ion and in so doing absorb energy from the field, or it can lose energy by releasing another electron from the ion by collisional ionization. It can also lose energy by radiation since together with the ion it forms a dipole driven by the laser field; based on the remarks above, any emitted photons must be at odd harmonics of the field. The classical model, of course, can only be taken so far, but it suggests that the largest possible energy that the electron could lose whenever it recollides with the ion is $E_{\max} + I_p$, this being the largest possible energy lost when the electron is recaptured by the ion and falls into the ground atomic level of energy $-I_p$ from which it was freed in the first place. Equating this largest possible energy loss to the largest possible energy of a radiated photon, we arrive at the HHG cutoff relation (14.7.10). Electrons recolliding with their ions with kinetic energies smaller than $E_{\max}$ result in radiation at lower harmonics.

Further support for this model of high-harmonic generation may be found in its predictions regarding the laser polarization, which we have assumed to be linear. If we write the Newton equation of motion for an ionized electron in a circularly polarized field, we find that the electron never returns to the ion, in which case there should be no HHG. In fact it is found experimentally that high-harmonic intensities decrease very rapidly with increasing polarization ellipticity of the laser field, which is 0 for linear polarization and 1 for circular.

- The model just described is called the *three-step model*: (1) an atom is ionized by the field, (2) the freed electron is driven back into the ion by the field, and (3) an electron returning to the ion emits a high-harmonic photon. The ionization is assumed to occur by a tunneling process that is best understood in the limiting case of a *static* electric field $\mathcal{E}$. In this case an electron in the hydrogen atom, for example, has a total potential energy $-e^2/r + e\mathcal{E}z$ if the applied static field is along the z direction. Thus, the applied field effectively lowers the “Coulomb barrier” confining the electron, allowing it to escape (ionize) by quantum mechanical tunneling; the stronger the applied field, the more the barrier is lowered and the greater is the rate of tunneling.

Although the applied field in the three-step model for HHG is certainly not static, the assumption that ionization occurs by tunneling appears to be a good approximation if the field frequency is not too large. This condition may be expressed in terms of the *Keldysh parameter* defined as

$$\gamma = \sqrt{\frac{I_p}{2U_p}} = \frac{\omega}{\omega_t}, \quad (14.7.13)$$

where $\omega_t = eE_0/\sqrt{2mI_p}$. γ delineates approximately between ionization by tunneling and by a multiphoton ionization process, tunneling dominating if $\gamma < 1$. The assumption of tunneling ionization in the three-step model is implicit in the initial condition that the electron velocity is zero at time t_i , since in the case of tunneling ionization, as opposed to multiphoton ionization, the initial kinetic energy of the electron can be assumed to be small. In the case of tunneling, furthermore, the ionization typically occurs in a short time compared to the field period, and the amplitude of the electron quiver is larger than an atomic dimension (Problem 14.19). For the ground state of the hydrogen atom, for example, the rate of tunneling ionization in the static-field approximation is given by quantum theory as

$$R_{\text{ion}}(t) = 4\omega_{\text{at}} \frac{E_{\text{at}}}{E(t)} \exp\left[-\frac{2}{3} \frac{E_{\text{at}}}{E(t)}\right], \quad (14.7.14)$$

where $\omega_{\text{at}} = (1/4\pi\epsilon_0)^2 me^4/\hbar^3 = 4.1 \times 10^{16} \text{ s}^{-1}$ and $E_{\text{at}} = (1/4\pi\epsilon_0)e/a_0^2 = 5 \times 10^9 \text{ V/cm}$ is the electric field on the electron ($a_0 = \text{Bohr radius}$). Note the *exponential sensitivity* of the tunneling rate on the electric field strength $E(t)$. For a multiphoton ionization process involving the absorption of q photons, by contrast, the ionization rate is approximately proportional to E_0^{2q} .

For a linearly polarized field $E(t) = E_0 \cos \omega t$ the cycle-averaged tunneling ionization rate is

$$\bar{R}_{\text{ion}} = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} R_{\text{ion}}(t) d(\omega t) = \frac{4\omega_{\text{at}}}{\pi} \frac{1}{F} \int_{-\pi/2}^{\pi/2} \frac{1}{\cos x} e^{-2/(3F \cos x)} dx, \quad (14.7.15)$$

where $F \equiv E_0/E_{\text{at}} (>0)$. Since the integrand is strongly peaked around $x=0$, we make the approximation $1/\cos x \cong 1/(1-x^2/2) \cong 1+x^2/2$ in the exponential and replace $\cos x$ by 1 in the prefactor:

$$\begin{aligned} \bar{R}_{\text{ion}} &\approx \frac{4\omega_{\text{at}}}{\pi} \frac{1}{F} \int_{-\pi/2}^{\pi/2} e^{-(2/3F)(1+x^2/2)} dx = \frac{4\omega_{\text{at}}}{\pi} \frac{2}{F} e^{-2/3F} \int_0^{\pi/2} e^{-x^2/3F} dx \\ &= \frac{4\omega_{\text{at}}}{\pi} \frac{2}{F} e^{-2/3F} \sqrt{3F} \int_0^{\pi/(2\sqrt{3F})} e^{-u^2} du \cong \frac{4\omega_{\text{at}}}{\pi} \frac{2}{F} e^{-2/3F} \sqrt{3F} \int_0^{\infty} e^{-u^2} du \\ &= 4\omega_{\text{at}} \sqrt{\frac{3}{\pi F}} e^{-2/3F} = \sqrt{\frac{3F}{\pi}} R_{\text{ion}}^{\text{static}}, \end{aligned} \quad (14.7.16)$$

where $R_{\text{ion}}^{\text{static}} = (4\omega_{\text{at}}/F) \exp(-2/3F)$ is the ionization rate in a static field of strength E_0 . We have assumed that $F \ll 1$, as is the case in most situations of interest, allowing us to replace the upper integration limit by ∞ in the second line.

For circularly polarized light the magnitude of the electric field is independent of time, and the tunneling ionization rate is the same as that for a static field.

The calculation of the HHG spectrum and cutoff in the three-step model is based on the expression [cf. Eq. (3.A.13)]

$$\mathbf{p}(t) = \int d^3x \psi^*(\mathbf{x}, t) e \mathbf{x} \psi(\mathbf{x}, t) \quad (14.7.17)$$

for the expectation value of the electric dipole moment when an atomic electron is described by a wave function $\psi(\mathbf{x}, t)$. The electron in a strong field can either remain in the initial atomic ground state with wave function $\psi_g(\mathbf{x}, t)$, or it can be in a continuum state with wave function $\psi_c(\mathbf{x}, t)$ if ionization occurs. If it is assumed that the occupation probabilities of excited bound states are negligible, $\psi(\mathbf{x}, t)$ in Eq. (14.7.17) takes the form $\psi(\mathbf{x}, t) = a_g \psi_g(\mathbf{x}, t) + a_c \psi_c(\mathbf{x}, t)$, where $|a_g|^2$ and $|a_c|^2$ are, respectively, the probabilities that the electron is in the ground state or a continuum state. Assuming furthermore that the ionization probability is small, or in other words that the ground state is negligibly depleted, the dipole moment (14.7.17) is determined approximately by the real part of $\int d^3x \psi_g^*(\mathbf{x}, t) e \mathbf{x} \psi_c(\mathbf{x}, t)$, and this expression can be evaluated “semiclassically” using results of the classical model of electron–ion recollisions.³⁵ •

According to the three-step model, electrons ionized by the laser field at different times recollide after different times with their parent ions, and their different kinetic energies upon recollision result in a distribution of radiated photon energies over the odd harmonics of the laser. The HHG radiation is assumed to be sufficiently weak that it does not produce any significant ionization. Each electron that does recollide with its parent ion does so by first being pulled away from the ion by the field and

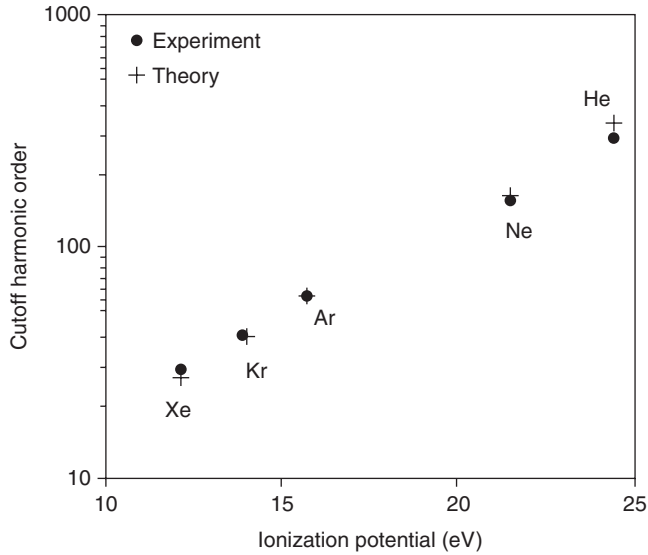


Figure 14.28 Experiment vs. theory for the cutoff HHG photon energies in inert gases. Note the logarithmic scale. [From Z. Chang, A. Rundquist, H. Wang, M. M. Murnane, and H. C. Kapteyn, *Physical Review Letters* **79**, 2967 (1997).]

then being forced back into the ion, the process occurring over a half-cycle of the field. After each half-cycle of the laser field, therefore, there is a burst of radiation whose spectrum consists of all the harmonics up to the cutoff photon energy, and whose duration is much shorter than a field cycle.

The cutoff relation (14.7.10) is an approximate, theoretical upper limit to the HHG photon energy, given the ionization potential and the laser intensity and frequency. In practice, the observed cutoff energy is generally smaller than that predicted by (14.7.10) for several reasons.³⁷ One is that at sufficiently high intensities the ground state is fully depleted and all the atoms are ionized by the leading edge of the laser pulse; the intensity I appearing in (14.7.11) is then replaced by a smaller “saturation” intensity I_s that can be estimated from quantum theoretical calculations of ionization rates. Figure 14.28 shows a comparison of experimentally observed HHG photon cutoff energies to the theoretical cutoff given by (14.7.10) with calculated values of I_s used for I in the expression for the ponderomotive energy. In these experiments 800-nm, 26-fs pulses from a Ti:sapphire laser system were focused onto gas jets from 1-mm-diameter nozzles, such that the gas pressures were ~ 8 Torr in the $\sim 100\text{-}\mu\text{m}$ focal region. The peak intensity of the 20-mJ laser pulses was $\sim 6 \times 10^{15} \text{ W/cm}^2$.

The three-step model applies to a single atom, and as such does not account for propagation effects such as phase mismatching. Furthermore, the intensity used to estimate the cutoff energy might well be higher than that in the medium under conditions of significant ionization, resulting in a smaller than calculated cutoff energy: The refractive index of a plasma, given approximately by Eq. (3.14.13), decreases with electron density, which is largest where the field intensity is greatest and causes the most ionization.

³⁷See, for instance, K. Miyazaki and H. Takada, *Physical Review A* **52**, 3007 (1995).

Therefore, a beam with a Gaussian transverse spatial profile, for example, will experience a smaller refractive index near the beam axis and can undergo self-defocusing as it ionizes atoms in its path. The good agreement between the experimental results shown in Fig. 14.28 with the predictions based on the single-atom cutoff relation (14.7.10) was attributed in part to the low gas density; at higher densities, where propagation effects are more significant, the cutoff energies were smaller, as expected.

Thus far, we have considered HHG mainly in terms of the frequencies generated, and have not addressed the question of the temporal nature of the HHG pulses aside from concluding from the three-step model that their duration should be extremely short. Their short wavelength, broad spectral width, and low intensities make them unamenable to the standard autocorrelation techniques described in Section 11.13, and studies of their temporal characteristics became possible only relatively recently. It is now well established that HHG produces attosecond pulse trains, and, as discussed below, can even produce the *single* attosecond pulses desired for time-resolved studies of subfemtosecond phenomena [cf. Figs. 14.23 and 14.24].

The very broad frequency spectrum associated with high-harmonic generation immediately suggests the possibility of producing attosecond pulses. Recall from our discussion of mode locking in Section 6.8 that if N phase-locked fields with equal amplitudes and with angular frequencies separated by Δ are superposed, the resulting field is a train of pulses separated in time by $T = 2\pi/\Delta$, each pulse having a duration $\tau = T/N$. In HHG, where only odd harmonics of the fundamental (laser) frequency ω appear, we have frequencies separated by 2ω , suggesting that the generation of N harmonics might result in a pulse train with $T = \pi/\omega$ and $\tau = \pi/N\omega$. For example, 100 harmonics of 880-nm radiation might form a train of 15-as pulses separated by 1.5 fs. However, the equal spacing of the high harmonics is by no means sufficient to make an attosecond pulse: The harmonics must also be phase-locked.

Quantum mechanical calculations reveal that the harmonics generated by a single atom are not phase-locked. It is found that, in the language of the semiclassical three-step model, the main contribution to each harmonic comes from two electron trajectories that correspond to different ionization and recollision times but that recollide with the ion with the same kinetic energy. Based on their differences in ionization and recollision times, these two types of trajectory are referred to as “long” and “short.” The two types of trajectory give rise to single-atom HHG emission in the form of a pulse train in which there are mainly *two* pulses per half-cycle ($T = \pi/\omega$) of the laser field. However, nonlinear propagation effects in a gas act in such a way that the harmonics are generated by either the long or the short trajectories, depending on the focusing geometry, *and the harmonics are phase-locked*. The HHG emission is then in the form of a train of extremely short pulses, *one* per half-cycle.³⁸ Experimental results are consistent with these predictions. It has been found, for instance, that the 5 odd harmonics 11–19 generated by 40-fs Ti:sapphire laser pulses ($T = \pi/\omega = 1.35$ fs) in argon are phase-locked and can combine to produce trains of 250-as pulses separated by 1.35 fs.³⁹

Pump-probe time resolution of subfemtosecond phenomena, as discussed in the preceding subsection, generally requires *single* attosecond pulses rather than pulse trains. *Pulse pickers* that select a single pulse (or a sequence of pulses to produce a train with

³⁸P. Antoine, A. L’Huillier, and M. Lewenstein, *Physical Review Letters* **77**, 1234 (1996).

³⁹P. M. Paul, E. S. Toma, P. Breger, G. Mullot, F. Audebert, P. Balcou, H. G. Muller, and P. Agostini, *Science* **292**, 1689 (2001).

a smaller repetition rate) from a mode-locked train of picosecond or femtosecond pulses typically employ a Pockels cell between two crossed polarizers, much like a light modulator (Fig. 5.19). The repetition rate of attosecond pulses in HHG, however, is far too fast for any electronics that can switch a Pockels cell. Single attosecond pulses have been produced by a *polarization gating* technique that relies on the strong dependence of HHG on the laser ellipticity, as noted earlier; the idea, basically, is to form a laser pulse with a time-dependent ellipticity, such that HHG can occur only within a narrow time window. Here we will briefly describe single attosecond pulse generation in the particular case of a few-cycle laser pulse. Based on the three-step model in which electrons emit short bursts of radiation only every half-cycle of the field, a few-cycle laser pulse can be expected to produce at most only a few pulses of high-harmonic radiation.

For the electric field of a linearly polarized laser pulse used to generate high harmonics we write

$$E(t) = \mathcal{E}(t) \cos(\omega t + \phi). \quad (14.7.18)$$

Usually, we can make the assumption that the envelope $\mathcal{E}(t)$ is slowly varying in time compared to the carrier wave $\cos(\omega t + \phi)$. Since $\omega \approx 10^{15} \text{ s}^{-1}$ at optical frequencies, this is a valid approximation for pulses in which $\mathcal{E}(t)$ varies negligibly on a femtosecond time scale, for example, for picosecond pulses; for such pulses the constant phase ϕ in (14.7.18) is unimportant. We used this approximation in Chapter 9, for instance, to formulate the theory of resonant pulse propagation based on the Maxwell–Bloch equations. For pulses lasting only a few cycles of the carrier wave, this approximation of a slowly varying envelope breaks down. Moreover, the phase ϕ for such pulses plays an important role in HHG, which depends sensitively on how the electric field varies over the duration of the laser pulse. In particular, different half-cycles of the field have different electric field amplitudes (Fig. 14.29) and will therefore lead to different peak (cutoff) photon energies.

Consider the example of a Gaussian pulse envelope:

$$E(t) = \mathcal{E}_0 e^{-t^2/\tau^2} \cos(\omega t + \phi), \quad (14.7.19)$$

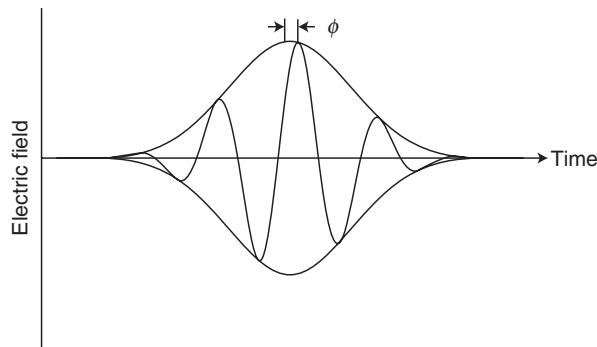


Figure 14.29 Carrier-envelope phase ϕ of a few-cycle pulse gives the phase difference between the peak of the pulse envelope and the nearest peak of the carrier wave.

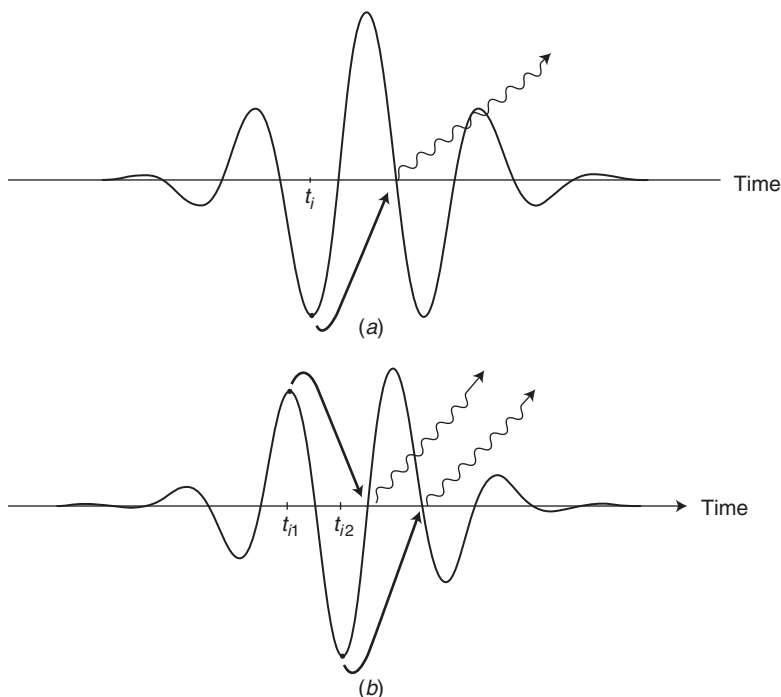


Figure 14.30 Dependence of peak-energy high-harmonic generation on the carrier-envelope phase of a few-cycle pulse. The peak pulse intensity is large enough to give rise to tunneling ionization and therefore bursts of high-harmonic radiation at times near that at which the peak of the pulse envelope occurs. In (a) the CEP is zero. Atoms undergo tunneling ionization at a time t_i , and an electron freed at this time, as indicated by the solid arrow, recollides with the parent ion half a field cycle later, emitting a burst of radiation according to the three-step model. In (b) the CEP is nonzero, and tunneling ionization of atoms is indicated at the two times t_{i1} and t_{i2} . The radiation that occurs half a field cycle after ionization now results in two bursts of radiation. In each case the electron “trajectories” shown are those for which the electron recolliding with the ion experiences the strongest field half-cycle, resulting in bursts of radiation at the peak photon energy.

for a pulse duration amounting to just a few cycles (Fig. 14.29). The pulse envelope peaks at $t = 0$, which coincides with the peak modulus of the electric field only if $\phi = 0$, where ϕ , the *carrier-envelope phase* (CEP), specifies the difference in time from the peak of the envelope to the nearest peak of the carrier. As illustrated in Fig. 14.30, a few-cycle laser pulse with zero carrier-envelope phase results in a single burst of radiation at the peak photon energy, whereas a nonvanishing CEP can produce two. Techniques have been developed to measure and control the carrier-envelope phase of few-cycle laser pulses and thereby to reliably produce single attosecond pulses by high-harmonic generation.⁴⁰

As discussed in Chapter 10, harmonic generation is generally very inefficient without phase matching. Gaseous media are nearly always used for high-harmonic generation

⁴⁰See C. A. Haworth, L. E. Chipperfield, J. S. Robinson, P. L. Knight, J. P. Marangos, and J. W. G. Tisch, *Nature Physics* **3**, 52 (2007) and references therein.

in order to minimize absorption, and the ionization that accompanies the process results in large phase mismatches between the fundamental and the co-propagating harmonics due to the plasma dispersion. Since gases are not birefringent, they do not permit angle phase matching based on birefringence. It has been demonstrated, however, that HHG conversion efficiencies can be substantially increased by a quasi-phase-matching in which counterpropagating Ti:sapphire pulses in a waveguide containing an inert gas modulate the field in such a way that phase matching is effectively realized periodically along the waveguide.⁴¹ Enhancement of HHG conversion efficiencies will have important consequences for attophysics and in a broad range of applications requiring coherent XUV and soft X-ray sources.

Frequency Combs and Optical Frequency Metrology

Throughout its history optical spectroscopy has been based on measurements of wavelengths rather than frequencies. Wavefront distortions result in different path lengths for different parts of a wave, limiting the most precise measurements to relative accuracies ($\Delta\lambda/\lambda$) of about 10^{-10} . Measurements of *frequencies* could be far more accurate since, as noted in Section 6.8, atomic clocks allow measurements of time to relative accuracies on the order of 10^{-15} . While frequency counters allow measurements of microwave frequencies as high as a few tens of gigahertz, no available electronics can count the $\sim 10^{15}$ cycles per second of *optical* waves. This problem was circumvented in the late 1990s with the use of *frequency combs*, allowing measurements of absolute optical frequencies (*optical frequency metrology*). By an “absolute” frequency (or an absolute measurement of frequency) we mean one that involves the second as *defined* by the cesium clock transition (Section 14.3); absolute frequencies are measured by referencing them to the oscillations of this transition. Since optical frequencies are $\sim 10^5$ times larger than that of the clock transition, measurements of absolute optical frequencies require that a very large frequency gap must somehow be bridged from the microwave to the optical.

Recall that the output of a mode-locked laser consists of a continuous, periodic train of pulses (Figs. 6.9 and 6.10). The spectrum of the pulse train is a *comb* of frequencies, such that the (angular-frequency) spacing ω_r between the “teeth” of the comb is equal to $2\pi/T$, where T is the spacing in time between pulses [Fig. 6.8 and Eq. (6.8.5b)]. ω_r depends on the length of the laser cavity and is in the radio-frequency range, typically 10^7 – 10^9 s^{−1}. The frequency combs used in optical frequency metrology are obtained from mode-locked lasers; for reasons discussed below, the comb frequencies are

$$\omega_n = n\omega_r + \omega_0, \quad (14.7.20)$$

where ω_0 , like ω_r , is very much smaller than an optical frequency; n is a positive integer, typically $\sim 10^6$ (Section 6.8). This formula, as well as an expression for the “offset” frequency ω_0 , will be derived below. We will also describe how ω_0 can be measured.

A frequency comb acts as a “ruler” for the measurement of optical frequencies. To determine the absolute frequency ω_L of a single-mode laser, for example, a measurement

⁴¹X. Zhang, A. L. Lytle, T. Popmintchev, X. B. Zhou, H. C. Kapteyn, M. M. Murnane, and O. Cohen, *Nature Physics* **3**, 270 (2007).

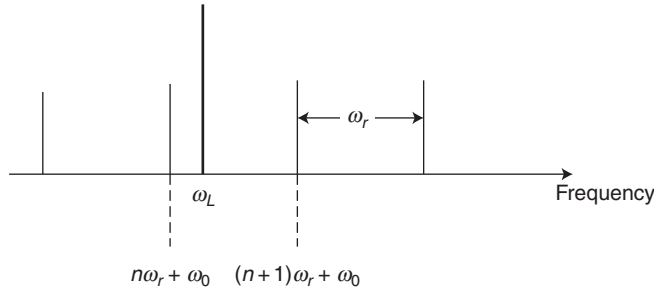


Figure 14.31 Frequency comb spectrum and position of optical frequency ω_L with respect to the closest comb frequency $n\omega_r + \omega_0$. Heterodyning results in beat notes determined by $\omega_L - n\omega_r - \omega_0$.

can be made of the signal detected by a photodetector when the single-mode laser and the mode-locked laser fields are superposed (heterodyned). As can be seen from Fig. 14.31, the rf signal will peak not only at the comb frequencies but also at beat frequencies determined by the difference $\omega_L - n\omega_r - \omega_0$ between the laser frequency and the nearest comb frequency (Problem 14.21). Since ω_r and ω_0 are known, n can be determined if ω_L is known beforehand (e.g., by having measured the wavelength with a wave-meter) to within $\pm \omega_r/4$; this yields ω_L . If such information about ω_L is not known beforehand, n can be determined by varying the comb spacing ω_r of the mode-locked laser. The frequency comb “ruler” can be applied in other ways. For example, the single-mode laser radiation can be frequency doubled, and the difference frequencies $\omega_1 = \omega_L - n\omega_r - \omega_0$ and $\omega_2 = 2\omega_L - 2n\omega_r - \omega_0$ between ω_L and ω_{2L} and the n th and $2n$ th comb lines can be measured by heterodyning. Again n can be determined if ω_L is known beforehand to within $\pm \omega_r/4$, or by “dithering” ω_r if it is not. Then ω_L is given in terms of the known quantities ω_r , ω_1 , ω_2 and n by the relation $\omega_2 - \omega_1 = \omega_L - n\omega_r$. The measured frequencies are “absolute” when they are referenced to a frequency standard, for example, when cycles per second are counted with a cesium atomic clock. It is the fact that n is very large ($\sim 10^6$) that allows the “bridging of the gap” between rf and optical frequencies.

Measurements of absolute optical transition frequencies have been made by locking the frequency of a laser to a transition (Section 5.13) and then measuring the absolute frequency of the laser by the frequency comb technique. Such methods have been used to measure the 1S–2S transition of the hydrogen atom, for example, to a few parts in 10^{14} . These ultra-high-precision techniques are valuable not only for spectroscopy and for applications including atomic clocks and navigation, but also for basic research in quantum electrodynamics and in determining whether the fundamental constants of nature might actually change over time.

To explain the form (14.7.20) of the frequency comb of a mode-locked laser, consider the electric field $E(t)$ from the laser incident on a photodiode; $E(t)$ is a sum of pulses separated by a time $T = 1/\omega_r$. Assuming the pulses are identical and described by an electric field $\mathcal{E}(t) \exp(-i\omega_c t)$, where $\mathcal{E}(t)$ is a pulse envelope and ω_c a carrier frequency, we write

$$E(t) = \sum_m \mathcal{E}(t - mT) e^{-i\omega_c(t - mT)}, \quad (14.7.21)$$

where the summation is over a large number of integers m . The Fourier transform of this field is

$$\begin{aligned}
 \tilde{E}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} E(t) e^{i\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \sum_m \mathcal{E}(t - mT) e^{-i\omega_c(t-mT)} e^{i\omega t} \\
 &= \frac{1}{2\pi} \sum_m \int_{-\infty}^{\infty} dt \mathcal{E}(t) e^{-i\omega_c t} e^{i\omega(t+mT)} \\
 &= \frac{1}{2\pi} \sum_m e^{im\omega T} \int_{-\infty}^{\infty} dt \mathcal{E}(t) e^{i(\omega - \omega_c)t} \\
 &= \tilde{\mathcal{E}}(\omega - \omega_c) \sum_m e^{im\omega T}, \tag{14.7.22}
 \end{aligned}$$

where $\tilde{\mathcal{E}}(\omega)$ is the Fourier transform of the envelope of a single pulse. We have already encountered the sum in the last line in Section 6.7. For the large number of summands appropriate for a mode-locked laser, it consists of a series of peaks separated by $2\pi/T$, i.e., a series of peaks at $\omega_n = 2\pi n/T$, where n is an integer and T is the pulse spacing. So (14.7.22) merely tells us what we already know from Chapter 6 about mode-locked pulse trains.

The assumption of identical pulses in a mode-locked train, however, is not valid when the pulses are very short and therefore have a large bandwidth causing significant dispersion in the laser medium and optical elements. In this case the pulse envelope will propagate within the laser cavity at a group velocity v_g while the carrier wave at frequency ω_c propagates at the phase velocity v_p (Sections 8.3 and 8.4). This results in a carrier-envelope phase difference that we denote by $\Delta\phi$. Now based on the “bouncing-ball” picture of the intracavity field of a mode-locked laser (Section 6.8), we infer that each successive pulse from the laser has an additional carrier-envelope phase difference $\Delta\phi$ over that of the preceding pulse, so aside from some overall phase we can assign a carrier-envelope phase $m\Delta\phi$ to the m th pulse of the mode-locked train. We therefore replace (14.7.21) by

$$E(t) = \sum_m \mathcal{E}(t - mT) e^{-i\omega_c(t-mT)} e^{im\Delta\phi} \tag{14.7.23}$$

and (14.7.22) by

$$\begin{aligned}
 \tilde{E}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} E(t) e^{i\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \sum_m \mathcal{E}(t - mT) e^{-i\omega_c(t-mT)} e^{im\Delta\phi} e^{i\omega t} \\
 &= \frac{1}{2\pi} \sum_m e^{im(\omega T + \Delta\phi)} \int_{-\infty}^{\infty} dt \mathcal{E}(t) e^{i(\omega - \omega_c)t} \\
 &= \tilde{\mathcal{E}}(\omega - \omega_c) \sum_m e^{im(\omega T + \Delta\phi)}. \tag{14.7.24}
 \end{aligned}$$

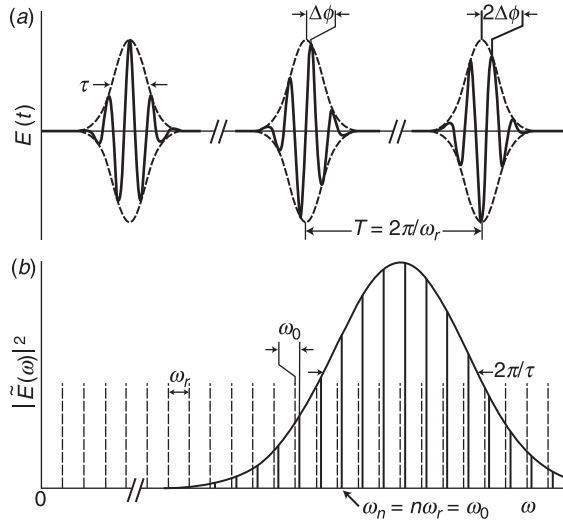


Figure 14.32 A mode-locked pulse train in (a) the time domain and (b) the frequency domain. Pulse-to-pulse carrier-envelope phase shifts $\Delta\phi$ in the time domain translate in the frequency domain to an offset ω_0 from integral multiples of the pulse repetition frequency ω_r . [After S. T. Cundiff and J. Ye, *Reviews of Modern Physics* **75**, 325 (2003).]

The only difference from (14.7.22) is that the peaks in the spectrum now occur at $\omega_n T + \Delta\phi = 2\pi n$:

$$\omega_n = n\omega_r - \frac{\Delta\phi}{T} \equiv n\omega_r + \omega_0, \quad (14.7.25)$$

with $\omega_0 = -\Delta\phi/T = -(\Delta\phi/2\pi)\omega_r$. Equation (14.7.24) describes the pulse train in the frequency domain, as opposed to the time-domain expression (14.7.23) (Fig. 14.32).

To use the frequency comb as a “ruler” to measure optical frequencies, it is necessary of course to have accurate values of ω_r and ω_0 . Whereas ω_r can be measured from the beat signal when the pulse train is incident on a photodiode, the measurement of the frequency offset ω_0 is not so straightforward. A common way of doing so is to frequency-double the n th comb line (frequency ω_n) and measure the beat signal between its second harmonic ($2\omega_n$) and the $2n$ th comb line (ω_{2n}):

$$2\omega_n - \omega_{2n} = 2(n\omega_r + \omega_0) - (2n\omega_r + \omega_0) = \omega_0. \quad (14.7.26)$$

Our simplified discussion ignores some critical issues in the actual implementation of frequency combs.⁴² We have implicitly assumed in writing (14.7.26), for instance, that the frequency comb from a mode-locked laser covers at least an octave of frequencies, that is, that the highest frequencies in the comb are twice as large as the lowest. Although a full octave is not strictly necessary, it is highly advantageous because it permits straightforward determinations, as described, of both rf frequencies ω_r and ω_0 ; once they are measured and stabilized, optical frequency metrology becomes a going concern.

⁴²For a comprehensive discussion see, for instance, S. T. Cundiff and J. Ye, *Reviews of Modern Physics* **75**, 325 (2003) and the many references therein.

An unchirped, transform-limited Gaussian laser pulse of duration τ has a FWHM spectral width $\Delta\Omega = (4 \ln 2)/\tau$ [Eq. (8.5.22)]. For this width to span an octave around the central frequency ω_c , we require $\Delta\Omega = \omega_c$, or $\tau = (4 \ln 2)/\omega_c = (4 \ln 2)/2\pi$ cycles, i.e., the pulse would have to have a duration of roughly a single cycle or less. Because of its extremely broad spectrum, the primary generator of frequency combs in optical frequency metrology is the Ti:sapphire laser with Kerr lens mode locking and compensation for group velocity dispersion (Fig. 11.25), but even the radiation from this system does not usually span an octave. For this reason octave-spanning frequency combs have been produced by spectrally broadening mode-locked pulses using self-phase modulation in fibers.

Recall from Section 10.4 that a nonlinear refractive index causes spectral broadening due to “self-chirping” or self-phase modulation. Group velocity dispersion of a short pulse causes it to broaden in time, reducing the peak intensity and therefore the self-phase modulation (Section 8.4). It has been found, however, that when a femtosecond pulse propagates in a photonic crystal fiber (Section 11.14) in which the silica core is surrounded by a particular arrangement of air holes, there is nearly zero group velocity dispersion near the 800-nm central wavelength of Ti:sapphire. This allows mode-locked pulses from a Ti:sapphire laser to propagate with little temporal broadening while maintaining the high intensities needed for substantial self-phase modulation and spectral broadening; more than an octave around 800 nm has been obtained in this way.⁴³ This is an example of *supercontinuum generation* in which a very broad (“white light”) spectrum is created from the much narrower spectrum of a laser. Self-focusing and other nonlinear phenomena in addition to self-phase modulation play a role in determining the white-light pulses generated by a mode-locked pulse train, and it is not at all obvious that successive pulses are strongly correlated in phase, as they must be if they are to serve as a broadband frequency comb (Problem 14.21). Interference experiments have demonstrated nevertheless that the supercontinuum white-light pulses generated by mode-locked laser pulses of interest for optical frequency metrology can in fact be phase-locked. In other words, supercontinuum generation can be used to convert a phase-locked laser pulse train into an *octave-spanning* phase-locked train.

A crucial factor for optical frequency metrology is the stabilization of the frequencies ω_r and ω_0 against unpredictable fluctuations in the laser cavity length. A small fraction of the mode-locked Ti:sapphire laser output, typically consisting of ~ 30 -fs pulses, can be used to measure ω_r (or a high harmonic of ω_r) with a photodiode and using a feedback loop to adjust the cavity length and lock ω_r to a frequency referenced to an atomic clock or a GPS-controlled quartz oscillator. For some purposes, such as measuring the difference in frequency between two lasers, locking of ω_r alone suffices, whereas for absolute optical frequency measurements it is also necessary to lock ω_0 . For this purpose the pulse train is injected into the photonic crystal fiber to generate an octave-spanning frequency comb, and part of the output from the fiber is used to measure ω_0 as described above [Eq. (14.7.26)]. One way of varying ω_0 in order to lock it is to swivel the laser feedback mirror of the layout shown in Fig. 11.25. Since the spectrum of the laser radiation with intracavity prisms is not uniform over the mirror, a swivel introduces a frequency-dependent phase shift and results in a group delay and a change in $\Delta\phi$ and therefore ω_0 .

⁴³D. J. Jones, S. A. Diddams, J. K. Ranka, A. Stentz, R. S. Windeler, J. L. Hall, and S. T. Cundiff, *Science* **288**, 635 (2000).

Locking ω_r and ω_0 at particular frequencies allows the synthesis of a comb of optical frequencies referenced to an atomic clock. *Optical synthesizers* providing $\sim 500,000$ absolute frequencies covering most of the optical and near-infrared spectrum, and containing the femtosecond laser and all the other instrumentation needed for comb generation in a unit occupying $\sim 1 \text{ m}^2$ on an optical bench, are now commercially available for absolute frequency measurements, high-precision spectroscopy, precise distance measurements, and other applications. It has been said that having such a synthesizer is like having $\sim 500,000$ extremely stable and precisely tuned lasers all at once. Optical synthesizers will likely be used to make optical clocks by locking a laser to the extremely narrow resonance of a cold atom. Because of the much shorter periods of optical cycles compared to microwaves, optical clocks will in turn result in even more precise timekeeping than is now possible.

- Prior to frequency comb generators absolute optical frequencies were measured by constructing long, complex “harmonic chains” to multiply the cesium clock frequency by harmonic generation in crystals, electronic frequency mixing, and other techniques. Because of their complexity and the large amount of laboratory space required, and because they could only be used to measure a single optical frequency, only a few such harmonic chains were developed. The frequency comb technique has revolutionized optical frequency metrology, and high-harmonic generation might extend the range of frequency metrology to XUV and soft X-ray frequencies.

Knowing that the spectrum of a mode-locked laser consists of a comb of frequencies, one might wonder in hindsight why the idea of a frequency comb as a ruler for optical frequency measurements was not put into practice until relatively recently. One reason, it seems, is that it was not realized just how uniform the frequency combs are; frequency metrology requires that the lines be “exactly” equally spaced [cf. Eq. (14.7.20)]. One might reasonably expect that dispersion effects in the laser would cause the mode spacing to vary slightly across the comb. Experiments in the late 1990s, however, revealed—surprisingly—that the mode spacing in a Kerr lens mode-locked laser with dispersion compensation (Fig. 11.25) is uniform and equal to the pulse repetition frequency to at least one part in 10^{17} , *even after the spectrum of the pulse train is broadened by propagation in a fiber!* Part of the 2005 Nobel Prize in Physics was awarded to J. L. Hall and T. W. Hänsch “for their contributions to the development of laser-based precision spectroscopy, including the optical frequency comb technique.” •

14.8 LASERS IN MEDICINE

Lasers continue to play a large role in medicine. Our brief overview will focus on some qualitative aspects of laser–tissue interactions and on the growing importance of laser science in medical imaging.

The first medical application of lasers was in ophthalmology, just a few years after the first demonstrations of laser oscillation in the early 1960s. It was known for centuries that visual loss can result from prolonged, direct viewing of the sun. This occurs due to a burning and consequent scarring of the macula, the central part of the retina responsible for acute vision (Fig. 14.33). It was suspected for some time that a localized burning and scarring might actually be useful for the treatment of certain visual disorders, and in the late 1940s G. Meyer-Schwickerath demonstrated that burns produced by white-light sources such as the sun and xenon lamps could be used to connect the retina to its sub-stratum tissue. The brightness of lasers made it immediately apparent that they could be

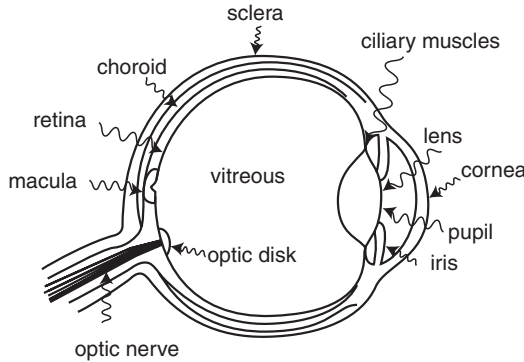


Figure 14.33 The human eye.

used in ophthalmology. In particular, it was apparent that lasers could be used for photo-coagulation of small areas of ocular tissue, and that this could be done with very short exposure times. Furthermore the monochromaticity of laser radiation allows selective coagulation of a particular tissue, since the photocoagulation process is initiated by wavelength-dependent absorption of light. Laser therapy has not only replaced some older surgical procedures for visual disorders but has also been used for conditions where effective treatments were previously unavailable.

- For safety's sake it must be remembered that even a low-power laser can damage the retina. To estimate the intensity focused on the retina when the eye is in the direct path of a laser beam, assume that the lens of the eye has a focal length of 2.3 cm and that the pupil diameter is 2 mm. The diameter of the spot on the retina is $f\theta$, where θ is the divergence angle of the laser beam. [See Eq. (13.2.7) for this formula in the special case of a Gaussian beam.] The fraction of the laser power entering the eye is equal to the square of the pupil diameter d divided by the square of the laser beam diameter D . The intensity of the focused radiation at the retina is therefore

$$I = \text{Pwr} \left(\frac{d}{D} \right)^2 \frac{1}{\pi(f\theta)^2/4} = \frac{4(\text{Pwr})d^2}{\pi D^2 f^2 \theta^2}. \quad (14.8.1)$$

Consider a green laser pointer with $\text{Pwr} = 1 \text{ mW}$, beam diameter $D = 1.5 \text{ mm}$, and divergence angle $\theta = 1.4 \text{ mrad}$. In this example $I \sim 200 \text{ W/cm}^2$. This estimate assumes perfect transmissivity of the eye (not too bad an approximation in the visible) and direct viewing with the eye held fixed and not blinking. But it illustrates the important point: Even very low-power lasers are potentially hazardous.⁴⁴ A 100-W lightbulb would not create nearly such a hazard; it is the small divergence angle θ of a laser that makes the focused intensity so large. •

Laser therapies may be broadly classified as photoabsorptive or photodisruptive. Photoabsorption results in electronic and vibrational excitation, breaking of molecular bonds, and a rise in temperature. Large biological molecules, such as proteins, can undergo conformational changes when the temperature is increased. The result is a *thermal denaturation* in which certain biological functions are lost or impaired due,

⁴⁴The retinal damage threshold in the case of continuous illumination can be as low as $2\text{--}3 \text{ W/cm}^2$. [A. M. Clarke, W. T. Ham, Jr., W. J. Geeraets, R. C. Williams, and H. A. Mueller, *Archives of Environmental Health* **18**, 424 (1969).]

for instance, to changes in cell membranes. Such thermal denaturation is responsible for effects such as inflammation and coagulation of tissue. Through a normal reparative response it leads to a scar, which can serve to connect tissues such as the sensory and pigmented layers of the retina. In other words, coagulation can “weld” disconnected tissue back together. Coagulation can also occlude (close) and destroy blood vessels; this is used in the treatment of diabetic retinopathy, in which fragile blood vessels may appear on the retina. These abnormal vessels tend to break, and the hemorrhaging can cause loss of vision. Since the pigmented cells behind the nerve-containing outer layer of the retina absorb strongly in the blue-green, whereas the neural retina, lens, cornea, and vitreous do not, the green argon ion laser is the primary tool for this procedure.

Photocoagulation begins with absorption of light, the primary absorbers of light in the human eye being melanin, hemoglobin, and xanthophyll. The ocular medium itself is transparent between about 380 and 1400 nm. At shorter wavelengths the lens and cornea are absorbing, and at longer wavelengths the primary absorber is water. Lasers allow a selective energy treatment in that the total energy of irradiation during the exposure can be accurately controlled. This allows a trained ophthalmologist to pre-select a certain energy that, based on experience, will produce a minimal degree of coagulation. The energy can then be increased gradually until a desired degree of coagulation has occurred, with minimal damage or side effects.

Exposure time is of course another very important consideration. For a given total energy, the temperature rise of the irradiated tissue increases with decreasing pulse duration, since there is less time for thermal diffusion to the surrounding tissue. With short pulses, therefore, the temperature rise can be quite large and can lead to vaporization at the irradiated spot; the clinical manifestation of this vaporization is the appearance of gas bubbles near the target. The vaporization can also generate pressure waves strong enough to damage eye tissue.

In contrast to photocoagulation, photodisruptive laser therapies are nonthermal. Photodisruption is initiated by ionization from the intense heat produced by a pulsed laser at the target. The resulting plasma absorbs energy from the laser and becomes very hot, expands rapidly, and produces a shock wave that can blast a hole in an ocular membrane. Such a hole can be several times larger than the waist of the laser beam at the focal point. In laser iridectomy for the treatment of glaucoma, a hole is made in the iris to relieve the elevated intraocular pressure.

A common application of laser photodisruption of tissue is in treating the clouding of the posterior capsule membrane behind the lens, an occasional complication of cataract surgery. In this outpatient treatment, called posterior capsulotomy, a pulsed Nd:YAG photodisruptor tears a hole in the clouded membrane, opening a path for clear vision. This is done without damage to the retina by focusing the laser to a spot just behind the lens. The divergence of the beam beyond the focus then results in a reduced intensity at the retina. The attenuation of the beam by the plasma created also helps to reduce the intensity at the retina.

Laser in situ keratomileusis (LASIK), introduced in the mid-1990s, has become the most common surgical procedure for refractive correction. After a metal blade is used to make a flap in the outer layer of the cornea, an ArF excimer laser (193 nm, ~ 10 -ns pulses, ~ 100 -Hz pulse repetition rate) penetrates the inner part of the cornea and breaks molecular bonds, allowing cells to escape in the form of a tiny “mushroom cloud.” Following this ablative sculpting with the excimer laser, the flap is closed and serves as a natural “bandage.” Most complications from LASIK are flap-related; in a

recent innovation the flap is made with a mode-locked, femtosecond diode-pumped Nd:YAG laser at $1.053\text{ }\mu\text{m}$. The positions and focal depths of successive pulses on the cornea are computer controlled, the pulses creating, by ionization and the formation of an expanding plasma that results in bubbles in the corneal tissue, thousands of holes (“cavitation bubbles”) of diameter $\sim 2\text{--}3\text{ }\mu\text{m}$. The pattern of holes defines the flap and its thickness ($\sim 10^2\text{ }\mu\text{m}$) and “hinge,” and the flap is then lifted prior to the application of the sculpting excimer laser. The flap is created in a matter of seconds, and with the programmed application of the femtosecond laser is done much more precisely than is possible with a hand-held blade. The low-energy femtosecond pulses have no effect on the intracorneal tissue until they reach their programmed depth; they do not cause any temperature rise or shock waves.

For precise corneal sculpting, information about the aberrations in the patient’s eye must be input to the software controlling the excimer laser pulses. This information is obtained by wavefront sensing with an “aberrometer,” which is usually a Shack–Hartmann sensor as described in Section 14.2. An eye-safe laser enters the eye and the aberrations of the originally “flat” wavefront are measured after it propagates through the cornea, vitreous, and lens and then reflects off the retina and makes a return pass through the eye. As in adaptive optics, the displacements of spots focused onto a CCD sensor by the lenslet array correspond to local phase gradients of the wavefront incident on the array, which in turn are determined by the refractive imperfections in the entire optical system of the eye (Fig. 14.34). The measured spot displacements and a computer are used to control the excimer laser pulses that sculpt the cornea to correct for these imperfections.

- There is still much to be learned about laser–tissue interactions. Given that living tissue is about 80% water by weight, it might be expected, for example, that the effects of the UV excimer laser pulses used in LASIK on living tissue can be understood in large part from how they affect water. But there are important differences in the way these pulses interact with water and with living tissue. One is that tissue elasticity limits the growth of cavitation bubbles, so that they

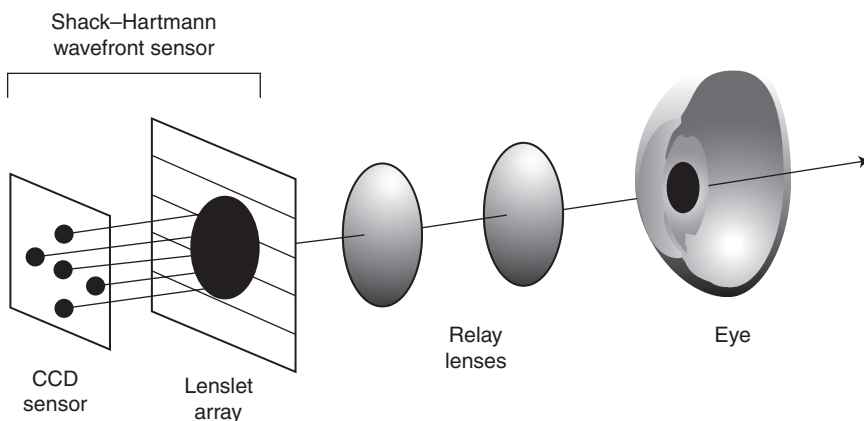


Figure 14.34 An “aberrometer” for determining the aberrations of an eye using a Shack–Hartmann wavefront sensor. The displacements of spots on the CCD array are a measure of the aberrations and are used to control an excimer laser for the corneal sculpting needed for refractive correction. [After L. N. Thibos, *Journal of Refractive Surgery* **16**, S563 (2000).]

are much smaller than in water; the photodisrupted regions of living tissues can be as small as a few hundred nanometers. This is beneficial in that it allows greater cutting precision than might be expected based on experiments with water.

As already noted, photodisruptive laser therapies rely on ionization and plasma formation. As discussed in Section 10.10, this optical breakdown is an avalanche process that proceeds from the presence or creation of a few electrons. In the optical breakdown of pure water these “seed” electrons are created by multiphoton ionization, which requires high intensities. In many biological tissues, however, there are large concentrations of a molecule (NADH) that is ionized by low-intensity UV radiation, yielding the seed electrons necessary for plasma formation; such tissues can be cut with lower laser intensities than tissues lacking this molecule.⁴⁵ ●

The same properties of lasers—brightness, directionality, and monochromaticity—used in eye surgery are employed in many other surgical procedures. Apart from ophthalmology, laser surgery appears to be most well established in dermatology and gynecology, but it is now commonly used in many other specialties. One of its advantages is that blood loss is greatly reduced compared to surgeries with knives. Together with magnification, this greatly facilitates the removal of tiny structures. In laryngeal surgery, for example, a laser and an endoscope make possible a degree of precision difficult to obtain by other procedures. In cancer surgery with lasers there is less danger than with cold-knife surgery of dislodging cancer cells and worsening the malignancy.

The most ubiquitous laser in nonophthalmic surgery has been the CO₂ laser. This is not surprising, given that water strongly absorbs at 10.6 μm ; the absorption coefficient of living tissue is roughly 200 cm^{-1} at 10.6 μm . Thus, CO₂ laser radiation interacts strongly with living tissue, and a focused CO₂ laser beam can produce explosive boiling of tissue fluids, resulting in an incision as the focal point is moved. Hand-held CO₂ surgical lasers, which have for many years been commercially available, are used, for instance, to remove arterial plaque during open-heart surgery.

Photodynamic therapy, which was invented in the early 1900s, came into widespread use in the mid-1980s and is now performed mainly with red diode lasers or diode laser arrays. This procedure begins with the injection into the bloodstream of a “precursor” that results in a photosensitizer, a substance that absorbs light of a particular wavelength. The targeted tissue is irradiated for several minutes with red laser light or LED radiation transported by a fiber. When the photosensitizer molecules absorb the light, they make a singlet–singlet transition and then undergo intersystem crossing to an excited, metastable triplet state (Section 11.11). From this state the excitation energy can be transferred to oxygen molecules in tissues, which in the process are excited from the ground triplet state to an excited singlet state. Singlet oxygen happens to be very reactive and, depending on the photosensitizer used, will react with and destroy undesirable (e.g., cancerous) cells. Photodynamic therapy is usually confined to about a 1-cm layer of tissue, the absorption depth of the light, and is therefore most effective for the treatment of small tumors on the skin or on internal organs accessed with optical fibers.

While the CO₂, argon ion, Nd : YAG, and excimer lasers remain the primary medical lasers, it is expected that diode-pumped fiber lasers, particularly because of their compactness, will in the near future take on increased importance as commercial development proceeds. Similarly, femtosecond lasers, including femtosecond fiber lasers, will

⁴⁵M. S. Hutson and X. Ma, *Physical Review Letters* **99**, 158104 (2007).

likely become more widely used in medicine, having already become effective and commercially viable tools in ophthalmology.

Laser physics and optical science more generally have spawned many medical diagnostic tools such as fiber endoscopes and other instruments that continue to be developed and improved. We will conclude this section with brief descriptions of two relatively new imaging technologies. In each case we focus on the basic principles and refer the reader to the specialized literature for details of clinical implementations.

- The imaging technique discussed below, based on spin-polarized gases, is basically a new way of magnetic resonance imaging (MRI). It is perhaps useful therefore to review briefly some basic principles of MRI.

As noted in Section 14.3, nuclei with an odd number of nucleons have spin. The nucleus of interest for MRI is the hydrogen nucleus (proton)—hydrogen atoms make up about 60% of the human body by number and about 10% by mass. The proton has spin $\frac{1}{2}$, and so in a static magnetic field $\mathbf{B}_0 = B_0 \hat{z}$ it has two allowed energy levels, E_+ and E_- , corresponding to two different spin orientations (“spin up” and “spin down”) with respect to $\hat{z}$, and the transition frequency between these two levels is [Eq. (14.3.1)]

$$\omega_p = \frac{E_+ - E_-}{\hbar} = \left[\left(\frac{1}{2} \mu_N g B_0 \right) - \left(-\frac{1}{2} \mu_N g B_0 \right) \right] / \hbar = \frac{\mu_N g B_0}{\hbar} \equiv \gamma B_0. \quad (14.8.2)$$

μ_N and g are, respectively, the nuclear Bohr magneton and the proton g factor, and $\gamma \cong 2\pi \times 42.576$ MHz/tesla is the *gyromagnetic ratio* of the proton; γ relates the magnetic dipole moment $\mathbf{m}$ and the angular momentum (spin) vector $\mathbf{s}$: $\mathbf{m} = \gamma \mathbf{s}$, with $|\mathbf{s}| = \hbar/2$ for spin $\frac{1}{2}$. An extremely strong field $\mathbf{B}_0$, as large as 1.5 T or more generated with liquid-helium-cooled superconducting coils, is used in MRI; for $B_0 = 1.5$ T, $\omega_p = 63.86 \text{ s}^{-1}$. The torque on a magnetic dipole moment $\mathbf{m}$ in a magnetic field $\mathbf{B}$ is $\mathbf{m} \times \mathbf{B}$, which gives the rate of change of the angular momentum $\mathbf{s}$, so that

$$\frac{d\mathbf{m}}{dt} = \gamma \mathbf{m} \times \mathbf{B}. \quad (14.8.3)$$

This equation says that the magnetic moment $\mathbf{m}$ precesses about $\mathbf{B}_0$, and that the precessional frequency is $\omega_p = \gamma |\mathbf{B}|$.

In a sample exposed to a magnetic field $B_0 \hat{z}$ and in thermal equilibrium at temperature T , there will be a net bulk magnetization $M_0 \hat{z}$, i.e., an imbalance in the densities of spin-up and spin-down protons:

$$\begin{aligned} M_0 &= P_D \gamma \frac{\hbar}{2} \times \frac{e^{-E_-/k_B T} - e^{-E_+/k_B T}}{e^{-E_-/k_B T} + e^{-E_+/k_B T}} = P_D \gamma \frac{\hbar}{2} \tanh\left(\frac{E_+ - E_-}{2k_B T}\right) \\ &= P_D \gamma \frac{\hbar}{2} \tanh\left(\frac{\hbar \gamma B_0}{2k_B T}\right) \\ &\cong P_D \frac{\hbar^2 \gamma^2 B_0}{2k_B T} \end{aligned} \quad (14.8.4)$$

for $\hbar \gamma B_0 / 2k_B T \ll 1$, where P_D is the proton density. For $T = 300\text{K}$ and $B_0 = 1.5$ T, $\hbar \gamma B_0 / 2k_B T = 5.1 \times 10^{-6}$, implying that the thermal magnetization in MRI is very small.

In MRI an rf field propagating perpendicular to the direction $\hat{z}$ of the static field, and having a frequency close to ω_p , induces transitions between the levels E_+ and E_- via the interaction with the proton spins of its time-dependent magnetic field $\mathbf{B}_1$. According to (14.8.3) such a field will result in a component of the magnetization perpendicular to the axial direction $\hat{z}$, i.e., it causes the magnetization $\mathbf{M}_0$ to tilt away from the $\hat{z}$ direction by some tipping angle Θ and therefore to precess about the strong static field $B_0\hat{z}$. This in turn results in a time-dependent magnetic flux perpendicular to $\hat{z}$ that, from Faraday's law of induction ($\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$), generates a voltage in an induction "pick-up" coil. This voltage is the MRI signal.

Considered as a two-state system, the proton can be described by Bloch equations; in fact, as noted in Section 9.5, the Bloch equations were first derived in the context of nuclear magnetic resonance (NMR) theory. When relaxation processes are included, the NMR Bloch equations take the form (9.5.3), where T_1 and T_2' are "longitudinal" and "transverse" relaxation times for the damping of the components of the magnetization parallel and perpendicular to $\mathbf{B}_0$, $\Delta = \omega_p - \omega$, the Rabi frequency $\chi = \gamma B_1$, and $w_0 = M_0$. The "area" of the pulse B_1 , i.e., the integral over time of the Rabi frequency, Eq. (9.5.21), defines the tipping angle Θ . A pulse with $\Theta = \pi$, for example, will cause a magnetic moment to make a 180° flip.

After the magnetization is "tipped" by the rf pulse, it relaxes back to its equilibrium value M_0 . The transverse (x, y) components, proportional to M_0 , decay exponentially at a rate determined by T_2' as well as by an "inhomogeneous broadening" time T_2^* due to spatial inhomogeneities in $\mathbf{B}_0$ and therefore different precessional frequencies of different spins; as was noted in Section 9.11 in connection with the Ramsey method of separated fields, no one has solved the problem of making a magnetic field that is uniform over a large region of space (in the case of MRI, a region very roughly the size of a human body). The decay of the transverse components of the magnetization results in an exponentially damped voltage signal in a pick-up coil, a so-called *free induction decay* signal. The longitudinal (z) component of the magnetization increases after a tipping pulse at the rate $1/T_1$ ($< 1/T_2'$): $M(t) = M_0[1 - \exp(-t/T_1)]$ [see Eq. (9.5.3c) with $\chi = 0$, $w = M$, and $w_0 = M_0$].

Spins in different tissues have different resonance frequencies ω_p , relaxation times T_1 , T_2' , proton densities P_D , and magnetizations M_0 in a magnetic field. These are determined, for instance, by different amounts of water (long relaxation times, ~ 1 s) and fat (short relaxation times, $\sim 10^2$ ms). Thus, the gray matter of the brain has relaxation times $T_1 \sim 760$ ms and $T_2' \sim 77$ ms, whereas for the white matter $T_1 \sim 510$ ms and $T_2' \sim 67$ ms.⁴⁶ They are distinguished by the different signals produced in a pick-up coil following the tipping rf pulses. Various "pulse sequences" with different pulse areas, repetition rates, and other characteristics have been devised to produce MRI signals; a commonly used pulse sequence for eliminating T_2^* effects is the *spin echo* sequence discovered by E. L. Hahn in the early 1950s.⁴⁷

In MRI the voltage signals are converted to a spatial map (image) of proton densities using magnetic field gradients. A "slice" of the sample is selected by a magnetic field gradient G_z along the z direction. Different slices of the sample along z have different precession frequencies $\omega_p = (\gamma B_0 + \gamma G_z z)$, and only one slice will be resonant with the rf field and produce a signal. The thickness of the slice is determined by the bandwidth of the rf pulse, and can be changed by changing G_z . This provides one (z) dimension of localization within the sample. A magnetic field gradient G_x orthogonal to z performs *frequency encoding* by causing different spins within a slice to have slightly different precession frequencies: a spin with spatial coordinate x along this gradient has a precession frequency $\omega_p(x) = \gamma(B_0 + G_x x)$ and therefore oscillates as $\exp[-i\gamma(B_0 + G_x x)t]$. Aside from the rapidly varying factor $\exp(-i\gamma B_0 t)$, therefore, the magnetization of a

⁴⁶J. L. Prince and J. Links, *Medical Imaging Signals and Systems* (Prentice Hall, Upper Saddle River, NJ, 2005).

⁴⁷A conceptually similar phenomenon—the *photon echo*—occurs for atomic transitions. See, for instance, L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms*, (Dover, New York, 1987), Chapter 9.

volume element at x oscillates as $\exp(-i\gamma G_x x t)$, and the signal at a pick-up coil from all the spins along x is

$$S(t) = \kappa \int_{-\infty}^{\infty} M_0(x) e^{-i\gamma G_x x t} dx, \quad (14.8.5)$$

where κ is a calibration constant. Defining $k_x = \gamma G_x t$, we replace (14.8.5) by

$$s(k_x) = \int_{-\infty}^{\infty} M_0(x) e^{-ik_x x} dx. \quad (14.8.6)$$

Inversion of this Fourier transform gives

$$M_0(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} s(k_x) e^{ik_x x} dk_x. \quad (14.8.7)$$

In other words, frequency encoding during the signal readout allows a computation of a one-dimensional map $[M_0(x)]$ of the magnetization in the selected slice of the sample. To obtain a two-dimensional image of each slice, a third magnetic field gradient G_y along the y direction is applied prior to the frequency conversion and signal readout. G_y has the same effect on the spins as G_x , but it is turned off prior to the signal readout. This leaves spins at different points along the y direction with different phases, similar to the way frequency encoding imparts different frequencies. The effect of this *phase encoding* is to replace (14.8.7) by a two-dimensional Fourier transform of the form

$$M_0(x, y) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(k_x, k_y) e^{i(k_x x + k_y y)} dk_x dk_y \quad (14.8.8)$$

giving the two-dimensional map $M_0(x, y)$ for the selected slice: 256 frequency-encoding sequences, each involving 256 phase-encoding steps, produce a 256×256 pixel array that is converted to a gray-scale image on a film. •

Very detailed images of soft tissues are obtained by magnetic resonance imaging. For porous tissues such as the lung and colon, however, the air-filled spaces do not have enough protons to generate a meaningful magnetization signal. Larger magnetizations could be realized by increasing the magnetic field strength B_0 or by lowering the temperature [Eq. (14.8.4)]. But current MRI scanners already use extremely strong magnetic fields, and increasing B_0 further would increase the complexity and the cost beyond their already high levels. And numerical estimates based on (14.8.4) indicate that the temperatures required would freeze patients to death—not a viable option. A different approach, based on the injection into porous tissue of a spin-polarized gas, was proposed in 1994.⁴⁸ The idea is to spin-polarize a gas by optical pumping (Section 14.3) and then inject it into a porous tissue; for the lungs, the patient simply inhales the spin-polarized gas, whose atoms have their spins and therefore magnetic dipole moments aligned. The degree of magnetization per particle is enormously larger than that implied by the thermal-equilibrium expression (14.8.4) for the protons in tissues imaged by standard MRI; this compensates for the low density of the gas, so that MRI as described above

⁴⁸M. S. Albert, G. D. Cates, B. Driehuys, W. Happer, B. Saam, C. S. Springer, Jr., and A. Wishnia, *Nature* **370**, 199 (1994).

can be done with the “breathable magnets” of the spin-polarized gas. For lung imaging, the patient inhales the spin-polarized gas before entering the MRI chamber.

Only certain atoms are suitable as breathable magnets. The atoms should have a nuclear spin in order to have a nuclear magnetic moment, and this eliminates all atoms with an even number of nucleons (Section 14.3). While alkali atoms are easy to spin-polarize with resonant laser radiation, the fact that they react violently with water or oxygen obviously rules them out. The spin-polarized atoms should have no toxic effects and should have long spin relaxation times so that their magnetization is sufficiently long-lived. They should not interact to form molecules that result in strong depolarization of the nuclear spins. Such considerations point to noble gases with nuclear spin $I = \frac{1}{2}$, and ^3He and ^{129}Xe in particular. Aside from the “Donald Duck voice” effect of helium, there are no adverse effects associated with it. Xenon, likewise, has for many years been used as an anesthetic and poses no known health risks, and moreover the small dosages of spin-polarized ^{129}Xe that are sufficient for MRI reduce the anesthetic effects. ^{129}Xe has some advantages over ^3He in that it is more readily and inexpensively available. Unlike ^3He , it occurs naturally at a concentration $\sim 10^{-7}$ in the atmosphere and is a by-product, along with the more abundant ^{131}Xe , of the distillation of air for the commercial production of O_2 and N_2 .

Optical pumping of ground-state noble-gas atoms is generally impractical; for He, for example, the transition from the ground level to the first excited level has a wavelength of 58 nm and cannot be accessed with available lasers. Helium has been spin-polarized by exciting it to a metastable excited level in an electric discharge and then optically pumping it with a laser at 1083 nm. Both ^3He and ^{129}Xe are spin-polarized in large concentrations by spin-exchange collisions (Section 14.3) with optically pumped alkali atoms. The most common way of doing this is to optically pump Rb atoms with 795 nm, circularly polarized laser radiation, usually obtained with a high-power (~ 100 W) diode laser array. The electronic spin polarization of the Rb atoms is then transferred to the ^3He or ^{129}Xe nuclei in collisions involving a magnetic hyperfine interaction. Since the hyperfine interactions are weak, the exchange probability in a collision is small, and it can take seconds or minutes for spin polarization of ^{129}Xe nuclei by this method, and hours to do so for ^3He . Spin relaxation of ^{129}Xe is sufficiently slow that the degree of polarization is limited by the Rb polarization. On the other hand, ^3He has a larger magnetic moment than ^{129}Xe (about three times larger), and thus far greater degrees of spin polarization have been realized with it (~ 20 – 50% compared to ~ 10 – 30% for ^{129}Xe). ^3He is found to produce sharper MRI images and has been used in most lung imaging studies. ^{129}Xe has one extremely attractive property in contrast to ^3He : It dissolves in blood and in many tissues. The circulation time of the blood in a human body is ~ 15 s, whereas ^{129}Xe in the blood maintains its polarization for tens of seconds, enough time for it to be used for brain imaging, for instance.

Unlike conventional MRI, where between rf sampling pulses the magnetization $M_0\hat{z}$ is restored by relaxation to thermal equilibrium in a time T_1 typically ~ 1 s, in MRI with a spin-polarized gas the magnetization is reduced with each successive sampling pulse and can be restored only by adding more polarized gas. The pulse sequencing is also different from conventional MRI; for example, $\pi/2$ pulses are not used in the spin-polarized case because such a pulse would tip the Bloch vector into the xy plane and thereby destroy the polarization with a single pulse. Typically ~ 100 pulses are used to produce an image, each pulse having the same small area. For pulses with an area of 0.20 rad (11.5°), for example, the magnetization after 100 sampling pulses is reduced by a factor $1 - [\cos(0.20)]^{100} = 0.87$, assuming that no other processes act to reduce the

magnetization during the pulse sequence. The fact that 100 rf pulses can be applied without having to wait for the magnetization to be restored by slow relaxation processes is important for lung imaging, for example, where a patient cannot be expected to hold his breath for too long.

Another attractive feature of MRI with spin-polarized gases is that they do not require the huge magnetic fields of conventional MRI systems. In the latter the magnetization is proportional to B_0 [Eq. (14.8.4)] and the voltage signal in a pick-up coil is, by Faraday's law of induction, proportional to the rate of change of the magnetization, that is, to $\omega_p B_0 = \gamma B_0^2$; a large B_0 is required to generate a large enough signal in the presence of noise associated with amplifiers and resistances in the system, since the magnetization is entirely thermal in origin and therefore small. When the magnetization is due to a spin-polarized gas, however, the factor B_0 on the right-hand side of (14.8.4) does not enter into the determination of the signal since the magnetization of the gas is not thermally equilibrated; the signal is then proportional to B_0 , which only determines ω_p , not the magnetization. It has been demonstrated that magnetic fields ~ 20 G can be used for MRI with spin-polarized gases, so that the cost and other disadvantages of extremely strong magnetic fields are avoidable.

Spin polarization of ^3He gas persists for several days in glass containers, and frozen ^{129}Xe maintains its polarization for comparable durations. Compact systems producing spin-polarized gases for clinical lung imaging have been developed. A flask containing ^3He , N_2 , and a small amount of Rb is heated to vaporize the Rb. Circularly polarized laser radiation optically pumps the Rb vapor and results after a few hours in the polarization of most of the ^3He . The flask is then cooled to condense out the Rb, and the spin-polarized ^3He gas is drawn into a plastic bag for later inhalation by a patient.

In the first MRI images demonstrated with spin-polarized ^{129}Xe gas, a xenon density of $\sim 1.2 \times 10^{19} \text{ cm}^{-3}$ filled the excised lung of a mouse,⁴⁸ compared to the $\sim 5 \times 10^{22} \text{ cm}^{-3}$ proton concentrations of tissues imaged in conventional MRI. Since this demonstration many other studies have been reported, and this new approach to MRI imaging appears to be progressing rapidly toward widespread clinical use.⁴⁹

Another novel imaging method is based on the connection between the bandwidth of light and interference. As discussed in Section 13.11, the coherence length of light decreases with increasing bandwidth; the larger the bandwidth, the smaller the arm separation has to be in a Michelson interferometer, for instance, in order to observe interference fringes. Femtosecond laser pulses, for example, have coherence lengths on the order of microns, so that no interference fringes are observed in a Michelson interferometer if the arm separation exceeds, say, a few microns (Section 13.11). This is the basis for *optical coherence tomography* (OCT), and the basic idea behind it is sketched in Fig. 14.35.⁵⁰

⁴⁹Reviews with many references to research papers in both the physics and medical literature have been published by T. Chupp and S. Swanson, *Advances in Atomic, Molecular, and Optical Physics* **45**, 41 (2001) and S. J. Kadlecik, K. Emami, M. C. Fischer, M. Ishii, Y. Jiangsheng, J. M. Woodburn, M. NikKhah, V. Vahdat, D. A. Lipson, J. E. Baumgardner, and R. R. Rizi, *Progress in Nuclear Magnetic Spectroscopy* **47**, 187 (2005).

⁵⁰D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, *Science* **254**, 1178 (1991). A review of optical coherence tomography "from bench to bedside" is given by A. M. Zysk, F. T. Nguyen, A. L. Oldenburg, D. L. Marks, and S. A. Boppart, *Journal of Biomedical Optics* **12**, 051403 (2007).

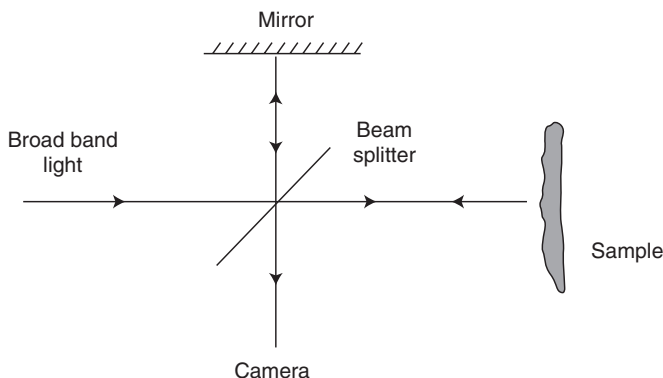


Figure 14.35 Simplified illustration of optical coherence tomography. Light from a broadband source (short coherence length) enters a Michelson interferometer, one arm of which contains a sample S (e.g., a retina or a fingertip). Interference fringes are observed only for path separations, determined by the topography of S , that differ by less than the coherence length. This allows depth resolution of the sample, while the camera provides two-dimensional profiling, so that three-dimensional imaging of S with micron-scale resolution is achieved if the coherence length of the light source is on the order of a micron.

The broadband source in OCT is typically “superfluorescent” (mirrorless) diode laser radiation transported by a fiber, but femtosecond lasers and supercontinuum radiation in fibers are also used in order to obtain extremely small coherence lengths. Compared to imaging techniques such as MRI and X-ray computed tomography (CT), OCT has a much smaller imaging depth (typically a few millimeters) but offers greater resolution of fine structure. Because the light source is nonionizing, it allows long exposure times that are not possible with X rays. Commercially available OCT systems are compact and less complex than MRIs or CTs, and their clinical utility has been demonstrated in oncology, ophthalmology, dermatology, and other areas.

14.9 REMARKS

Lasers have made possible the most accurate determination of several physical quantities, the smallest measured time intervals, the most accurate clocks, the lowest temperatures ever realized, and some of the highest powers ever generated. They have been used to determine distances between Earth and the moon to an accuracy of a few centimeters and to enable ground-based telescopes to produce images comparable in quality to those of a space telescope; to determine concentrations of atmospheric constituents in the atmosphere as well as temperature, density, and wind profiles; to produce ultracold gases and a state of matter in which large numbers of atoms are described by a single quantum state and in which the interference of atom waves can be observed and put to practical use; to time-resolve chemical and biological processes; to determine our grocery bills, print what we read, record our music, and cut and weld materials used in many other aspects of everyday life. Their importance in medicine continues to grow and they are an integral part of modern communications and the Internet. The list of applications has no foreseeable end.

The operating principle of every laser is, of course, stimulated emission of radiation. As discussed in Section 3.6, it was Einstein who first realized that matter can radiate by stimulated emission as well as by spontaneous emission. But no one at the time seems to have imagined the possibility of radiation amplifiers and oscillators, much less their widespread use decades later. While some applications could be confidently predicted when the first lasers were constructed in the early 1960s, many others were not foreseen by even the most imaginative scientists and engineers. The reader might find it interesting to think about which laser applications were the most unpredictable, and to speculate on directions future applications might take.

PROBLEMS

- 14.1. (a)** Using Eqs. (3.7.5) for the case in which the level populations of a two-level atom change by absorption and stimulated and spontaneous emission in a narrowband field, show that the upper-state probability $p(t)$ changes according to the equation

$$\dot{p}(t) = -A_{21}p(t) + \frac{\sigma(\nu)}{h\nu} I(t)$$

if the lower-state probability $\cong 1$.

- (b)** Assuming that the laser pulse duration is long compared to the radiative lifetime, derive Eq. (14.1.3).
- 14.2. (a)** Verify the approximation (14.1.21) for Rayleigh backscattering at frequencies roughly in the visible range.
- (b)** Verify Eq. (14.1.22) for the number of backscattered photons for the assumed lidar system parameters.
- 14.3.** Show that the sodium density at different altitudes in the mesosphere can be determined from the ratio of sodium and Rayleigh photocounts as in Eq. (14.1.26). What assumptions are implicit in this equation?
- 14.4. (a)** Estimate the maximum photon flux (number of photons per unit area per unit time) at ground level that can be obtained when the 589-nm D_2 line of mesospheric sodium is uniformly irradiated with resonant radiation having a 1-m^2 spot size at the mesosphere.
- (b)** What would be the apparent magnitude of a guide star that produces this photon flux?
- (c)** Approximately what fraction of quasi-monochromatic 589-nm laser radiation propagated from ground is absorbed by the mesospheric sodium layer? (Assume that the laser power is sufficiently low that saturation effects are negligible.)
- (d)** Why should the saturation formula (14.2.5) be applicable? Shouldn't the absorption or gain coefficient for a Doppler-broadened transition saturate according to the formula (4.14.7)?
- 14.5. (a)** The reflector placed on the moon by Apollo 11 astronauts consists of 100 corner cubes, each about 1.5 inches across. Estimate the diameter at Earth

of the return ruby laser pulses. (A *corner cube* consists of three reflecting edges intersecting at right angles. Any incident ray is reflected parallel to itself.)

- (b) The laser pulses from a police lidar gun have a beam divergence angle of 3 mrad. Estimate the diameter of these pulses at a distance of 300 m from the policeman. Can you think of ways to thwart this type of “radar”?
- 14.6.** Show that the matrix element for an electric dipole transition in which light is absorbed and an atom goes from a state 1 with magnetic quantum number m to a state 2 with magnetic quantum number m' is given by (14.3.3), and that the reverse, stimulated emission transition is given by the complex conjugate of (14.3.3).
- 14.7. (a)** Show that a detector placed along the x axis, and responding to an average over times long compared to ω_0^{-1} of the power radiated by the electric dipole (14.3.15), will record an intensity with the time dependence (14.3.16) if it is behind a polarizer oriented at an angle φ with respect to the z axis.
- (b) Suppose that the atoms are excited by linearly polarized light as in Fig. 14.12 and that the detector records the time-integrated intensity

$$\mathcal{J} \propto \int_{-\infty}^t dt_0 I(t)$$

with $t \gg 1/\gamma$. Show that $\mathcal{J}$ is proportional to

$$\frac{1}{2\gamma} + \frac{\gamma \cos 2\varphi}{\gamma^2 + 4\omega_L^2} + \frac{2\omega_L \sin 2\varphi}{\gamma^2 + 4\omega_L^2}.$$

Plot this signal vs. ω_L for (i) $\varphi = 0$, (ii) $\varphi = \pi/4$, (iii) $\varphi = \pi/2$, and (iv) $\varphi = 3\pi/4$. What experimental situation does this signal describe? Compare your results with the corresponding experimental results shown in Fig. 14.36.

- 14.8. (a)** Show that optical pumping with circularly polarized light results in complete transparency in the case of the sodium D₁ line in the absence of any spin relaxation effects.
- (b) Calculate the electric dipole matrix element $\langle F'm'|d_{q=1}|Fm\rangle$ for the transition $3P_{3/2} (F=3, m=3) \rightarrow 3S_{1/2} (F=2, M=2)$ of sodium. (You will have to look up the values of the $3j$ and $6j$ symbols for the set of quantum numbers $F'=3, m'=3, F=2, m=2, J'=\frac{3}{2}, J=\frac{1}{2}, I=\frac{3}{2}$.)
- (c) Confirm that the saturation intensity for this transition in the case of pure radiative broadening is 6.3 mW/cm², as stated in Section 4.11.
- (d) Verify Eq. (14.4.5).
- 14.9. (a)** An atom has a transition frequency $\nu_0 = (E_2 - E_1)/h$, where 1 and 2 refer to the ground level and first excited level, respectively, and it is moving with velocity v away from a source of radiation of frequency ν . Using conservation of energy and linear momentum, and assuming that line broadening is

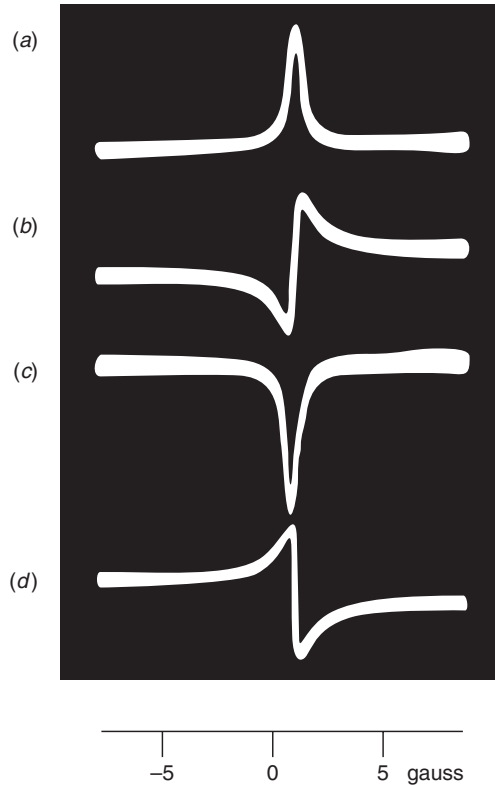


Figure 14.36 Time-integrated Hanle signals from mercury atoms for polarizer angles (a) $\varphi = 0$, (b) $\varphi = \pi/4$, (c) $\varphi = \pi/2$, and (d) $\varphi = 3\pi/4$. (See Problem 14.7.) [From B. P. Kibble and G. W. Series, *Proceedings of the Physical Society of London* **78**, 70 (1961).]

negligible, show that the atom will absorb a photon of frequency ν provided that

$$\nu = \nu_0 \left(1 + \frac{v}{c} \right).$$

Assume that $v \ll c$, so that relativistic effects may be ignored.

(b) Show that there is a *recoil shift*

$$\nu - \nu_0 = \frac{h\nu_0^2}{2Mc^2}$$

in the absorption frequency, where M is the atomic mass.

(c) Derive an expression for the recoil shift in the case of *two-photon* absorption. Calculate the recoil shift in the case of the 1S–2S two-photon transition in atomic hydrogen, and compare your answer with the measured recoil shift of 6.7 MHz reported by D. G. Fried, T. C. Killian, L. Willmann, D. Landhuis, S. C. Moss, D. Kleppner, and T. J. Greytak, *Physical Review Letters* **81**, 3811 (1998).

- (d) Estimate the critical temperature for Bose–Einstein condensation of atomic hydrogen at a density of 1.8×10^{20} atoms/m³.
- 14.10.** Derive the formula (14.4.42) for the potential energy function for the dipole force. [Note: You might use the fact that the dipole force in (14.4.36) applies even under conditions where the polarizability can be saturated. You could also proceed as in the derivation of the light shift (14.4.22).]
- 14.11.** (a) Show that the superposition of an upward-propagating laser beam of frequency $\nu + \Delta\nu$ and a downward-propagating laser beam of frequency $\nu - \Delta\nu$ results in a “moving standing wave” with upward velocity $\lambda\Delta\nu$ ($\lambda = c/\nu$).
- (b) Estimate the length traveled and the time taken for a cesium atom in an atomic fountain to reach its apogee and return to its original position after being launched by a moving standing wave with $\lambda = 852$ nm (Fig. 14.18) and $\Delta\nu = 1.6$ MHz.
- 14.12.** (a) Compare the optical trapping force to the magnetic force acting on atoms in a MOT, and show that the optical force is dominant for atoms at a distance of a wavelength or more from the center of the trap.
- (b) Use Eq. (14.5.9) to estimate the size of the cloud of sodium atoms trapped in a MOT.
- (c) Estimate the zero-point energy of a sodium atom in a magnetic trap with $B = 10$ G/cm.
- 14.13.** A sodium atom in a MOT with a magnetic field gradient of 30 G/cm and counter-propagating laser fields of intensity 1 mW/cm², detuned from the D₂ transition by 30 MHz, is displaced from the center of the trap. Estimate the time it takes for the atom to move to the center of the trap.
- 14.14.** Suppose that sodium atoms are in an optical lattice formed by a laser of intensity 4 mW/cm² detuned by 60 GHz from the D₂ line. Show that the peak value of the light-shift potential energy function is about 10 times the sodium recoil energy. [J. H. Denschlag, J. E. Simsarian, H. Häffner, C. McKenzie, A. Browaeys, D. Cho, K. Helmerson, S. L. Rolston, and W. D. Phillips, *Journal of Physics B: Atomic, Molecular and Optical Physics* **35**, 3095 (2002).]
- 14.15.** (a) Derive the expression (14.5.13) for the (cycle-averaged) force of radiation pressure on a dielectric sphere with a radius small compared to the radiation wavelength. (Hint: Show that this force is equal to $n_b P/c$, where P is the Rayleigh-scattered power.)
- (b) A focused laser beam is to be used to balance the force of gravity on a latex sphere of diameter 5 μ m. Assuming that latex has a density of 1.05 g/cm³ and a refractive index of 1.6 at the laser wavelength, estimate the laser intensity and power required when the sphere is to be held in air, assuming a Gaussian laser beam focused to a spot of diameter 10 μ m. Repeat these estimates when the latex sphere is in water.
- 14.16.** (a) Estimate the critical temperature for Bose–Einstein condensation of liquid ⁴He, for which the density is 2.2×10^{28} atoms/m³.

- (b) Calculate the recoil temperature T_{recoil} for the laser cooling of ^{87}Rb atoms at the 780 nm transition.
- (c) Calculate the recoil temperature for hydrogen atoms at the 122-nm Lyman- α transition, and compare it with the critical temperature obtained in Problem 14.9 for BEC in hydrogen.
- 14.17. (a)** Assuming a wave function of the form (14.7.6), and probability amplitudes $a_v(t) = a_v(0) \exp(-iE_v t/\hbar)$, derive an expression for the expectation value of the interatomic separation R . Compare your formula to that obtained if there is no “off-diagonal coherence,” that is, if collisions or other “decoherence” effects cause the quantities $a_v^*(t)a_{v'}(t)$ to decay rapidly in time for $v \neq v'$.
- (b) Explain how signals such as that shown in Fig. 14.23 can be used to obtain values for vibrational and rotational constants of a molecule.
- (c) Our simplified analysis of the time resolution of molecular vibrations with femtosecond pump and probe pulses in Section 14.7 assumes a single molecule. Under what conditions can we apply such a single-molecule theory to an ensemble of molecules in order to interpret experimental data such as shown in Fig. 14.23?
- 14.18. (a)** Use Eq. (3.14.6) and the formula $n^2(\omega) - 1 = (N/\epsilon_0)\alpha(\omega)$ for the refractive index to obtain

$$\alpha_i(\omega) = \frac{e^2}{m} \sum_j \frac{f_j}{\omega_j^2 - \omega^2}$$

for the polarizability of an atom in state i , where the summation is over all states j connected to state i by oscillator strengths f_j and transition frequencies ω_j .

- (b) What modifications reduce this expression to the approximation (14.4.23)?
- (c) Show that the light shift obtained using the polarizability derived in part (a) reduces to the ponderomotive energy U_p [Eq. (14.7.11)] in the limit in which the field frequency ω is much greater than any of the transition frequencies ω_j .
- 14.19. (a)** Show that the times t_f at which the electron returns to the ion in the model presented in Section 14.7 are given by solutions of the equation

$$\cos \theta_f - \cos \theta_i + (\theta_f - \theta_i) \sin \theta_i = 0,$$

where $\theta_f = \omega t_f$ and $\theta_i = \omega t_i$, t_i being a time at which ionization occurs. Show also that the electron’s kinetic energy when it reaches the ion is $E(t_f, t_i) = 2U_p(\sin \theta_f - \sin \theta_i)^2$, where U_p is the ponderomotive energy.

- (b) Solve numerically for $E(t_f, t_i)$ for θ_i between 0 and $\pi/2$ and verify the results plotted in Fig. 14.27. In particular, verify that the maximum value of $E(t_f, t_i)$ is $3.17U_p$.
- (c) What is the probability that an electron will recollide with its parent during the first cycle of the laser field following ionization?
- (d) How are the predictions of the model changed when the laser radiation is *circularly* polarized?

- (e) Using the Bohr model, compare the amplitude of the electron's quivering in the field to the size of an atom when the Keldysh parameter γ given by (14.7.13) is less than 1.
- 14.20.** Estimate the tunneling ionization rate for a hydrogen atom in (a) a linearly polarized field of intensity $3 \times 10^{16} \text{ W/cm}^2$ and (b) a circularly polarized field of the same intensity.
- 14.21. (a)** Describe the spectrum obtained by the heterodyning of a single-mode laser with a mode-locked laser, assuming the two fields have frequencies as indicated in Fig. 14.31.
- (b)** Explain why there must be phase coherence among pulses equally spaced in time if the pulse train is to serve as a comb for optical frequency metrology.