Chapter 12: Optical Communications

12.1 Introduction to optical communication links

12.1.1 Introduction to optical communications and photonics

Optical communications is as ancient as signal fires and mirrors reflecting sunlight, but it is rapidly being modernized by *photonics* that integrate optics and electronics in single devices. Photonic systems are usually analyzed in terms of individual photons, although wave methods still characterize the guidance of waves through optical fibers, space, or other media. This chapter introduces optical communications and applications of photonics in Section 12.1. It then discusses simple optical waveguides in Section 12.2, lasers in Section 12.3, and representative components of optical communications systems in Sections 12.4, including photodetectors in 12.4.1-2, multiplexers in 12.4.3, interferometers in 12.4.4, and optical switches in 12.4.5.

12.1.2 Applications of photonics

Perhaps the single most important application of photonics today is to optical communications through low-loss glass fibers. Since 1980 this development has dramatically transformed global communications. The advantage of an optical fiber for communications is that it has a bandwidth of approximately one terahertz, and can propagate signals over continental and even global distances when assisted by optical amplifiers. These amplifiers are currently separated more than ~80 km, and this separation is steadily increasing as technology improves. In contrast, coaxial cable, wire-pair, and wireless links at radio frequencies still dominate most communication paths of bandwidth < ~2 MHz, provided the length is less than ~1–50 km.

One broadband global wireless alternative to optics is microwave communications satellites in geosynchronous orbit[66] that can service ships at sea and provide moveable capacity addressing transient communications shortfalls or failures across the globe; the satellites simply point their antenna beams at the new users, who can be over 10,000 km apart. The greatest use of satellites, however, is for broadcast of entertainment over continental areas, either to end-users or to the head ends of cable distribution systems. In general, the limited terrestrial radio spectrum is more efficiently used for broadcast than for one-to-one communications unless there is re-use of spectrum as described in Section 10.4.6. Optical techniques are disadvantaged for satellite-ground links or ground-to-ground links through air because of clouds and fog, which restrict such links to very short distances or to cases where spatial diversity[67] offers clear-air alternatives.

Optical links also have great potential for very broadband inter-satellite or diversity-protected satellite-earth communications because small telescopes easily provide highly focused antenna beams. For example, beamwidths of telescopes with 5-inch apertures are typically one

---

[66] Geosynchronous satellites at 22,753-mile altitude orbit Earth once every 24 hours and can therefore hover stationary in the sky if they are in an equatorial orbit.
[67] Spatial diversity involves use of spatially distinct communications links that suffer any losses independently; combining these signals in non-linear ways improves overall message reliability.

arc-second[68], corresponding to antenna gains of $\sim 4\pi \times (57\times 3600)^2 \cong 5\times 10^{11}$, approximately 5000 times greater than is achievable by all but the very best radio telescopes. Such optical links are discussed in Section 12.1.4.

Optical fibers are increasingly being used for much shorter links too, simply because their useable bandwidth can readily be expanded after installation and because they are cheaper for larger bandwidths. The distance between successive amplifiers can also be orders of magnitude greater (compare the fiber losses of Figure 12.2.6 with those of wires, as discussed in Section 7.1.4 and Section 8.3.1). The bandwidth per wire is generally less than $\sim$0.1 GHz for distances between amplifiers of 1 km, whereas a single optical fiber can convey $\sim$1 THz for 100 km or more. Extreme data rates are now also being conveyed optically between and within computers and even chips, although wires still have advantages of cost and simplicity for most ultra-short and high-power applications.

Optical communication is not the only application for photonics, however. Low-power lasers are used in everyday devices ranging from classroom pointers and carpenters' levels to bar-code readers, laser copiers and printers, surgical tools, medical and environmental instruments, and DVD players and recorders. Laser pulses lasting only 10^{-15} second (0.3 microns length) are used for biological and other research. High power lasers with tens of kilowatts of average power are used for cutting and other manufacturing purposes, and lasers that release their stored energy in sub-picosecond intervals can focus and compress their energy to achieve intensities of $\sim 10^{23}$ W/m^2 for research or, for example, to drive small thermonuclear reactions in compressed pellets. Moreover, new applications are constantly being developed with no end in sight.

12.1.3 Link equations

The link equations governing through-the-air optical communications are essentially the same as those governing radio, as described in Section 10.3. That is, the received power P_r is simply related to the transmitted power P_t by the gain and effective area of the transmitting and receiving antennas, G_t and A_e:

$$P_r = \left(G_t P_t / 4\pi r^2\right) A_e \ \ [W] \hspace{2cm} \text{(optical link equation)} \quad (12.1.1)$$

The gain and effective area of single-mode optical antennas are related by the same equation governing radio waves, (10.3.23):

$$G = 4\pi A / \lambda^2 \hspace{6cm} (12.1.2)$$

Some optical detectors intercept multiple independent waves or modes, and their powers add. In this case, the gain and effective area of any single mode are then less relevant, as discussed in Section 12.1.4.

---

[68] One arc-second is 1/60 arc-minutes, 1/60^2 degrees, 1/(57.3$\times$3600) radians, or 1/60 of the largest apparent diameters of Venus or Jupiter in the night sky.

The maximum bit rate that can be communicated over an optical link is not governed by the $E_b > \sim 10^{-20}$ Joules-per-bit limit characteristic of radio systems, however, but rather by the number of photons the receiver requires per bit of information, perhaps ~ 10 for a typical good system. Each photon has energy $E = hf$ Joules. Thus to receive R bits/second might require received power of:

$$P_r = E_b R \cong 10 hfR \;\; [\text{W}] \qquad\qquad \text{(optical rate approximation)} \qquad (12.1.3)$$

where h is Planck's constant (6.624×10^{-34}) and f is photon frequency [Hz]. Clever design can enable many bits to be communicated per photon, as discussed in the following section.

12.1.4   Examples of optical communications systems

Three examples illustrate several of the issues inherent in optical communications systems: a trans-oceanic optical fiber cable, an optical link to Mars, and an incoherent intra-office link carrying computer information.

First consider a trans-oceanic *optical fiber*. Section 12.2.2 discusses losses in optical fibers, which can be as low as ~ 0.2 dB/km near 1.5-micron wavelength ($f \cong 2 \times 10^{14}$ Hz). To ensure the signal (zeros and ones) remains unambiguous, each link of an R = 1-Gbps fiber link must deliver to its receiver or amplifier more than $\sim 10 hfR$ watts, or $\sim 10 \times 6 \times 10^{-34} \times 2 \times 10^{14} \times 10^9 \cong 1.2 \times 10^{-9}$ watts; a more typical design might deliver $\sim 10^{-6}$ watts because errors accumulate and equipment can degrade. If one watt is transmitted and 10^{-6} watts is received, then the associated 60-dB loss corresponds to 300 km of fiber propagation between optical amplifiers, and perhaps ~ 20 amplifiers across the Atlantic Ocean per fiber. In practice, erbium-doped fiber amplifiers, discussed in Section 12.3.1, are now spaced approximately 80 km apart.

Next consider an *optical link* communicating between Earth and astronauts on Mars. Atmospheric diffraction or "seeing" limits the focusing ability of terrestrial telescopes larger than ~ 10 cm, but Mars has little atmosphere. Therefore a Martian optical link might employ the equivalent of a one-square-meter optical telescope on Mars and the equivalent of 10-cm-square optics on Earth. It might also employ a one-watt laser transmitter on Earth operating at 0.5-micron wavelength, in the visible region. The nominal link and rate equations, (12.1.1) and (12.1.3), yield the maximum data rate R possible at a range of $\sim 10^{11}$ meters (approximate closest approach of Mars to Earth):

$$R = P_r / E_b \cong \left(G_t P_t / 4\pi r^2 \right) A_e / 10 hf \;\; [\text{bits s}^{-1}] \qquad\qquad (12.1.4)$$

The gain G_t of the transmitter given by (12.1.2) is $G_t \cong 4\pi A / \lambda^2 \cong 5 \times 10^{11}$, where $A \cong (0.1)^2$ and $\lambda \cong 5 \times 10^{-7}$ [m]. The frequency $f = c/\lambda = 3 \times 10^8 / [5 \times 10^{-7}] = 6 \times 10^{14}$. Therefore (12.1.4) becomes:

$$R \cong \left\{ [5 \times 10^{11} \times 1] \middle/ \left[4\pi (10^{11})^2 \right] \right\} \left\{ 1 \middle/ [10 \times 6.624 \times 10^{-34} \times 6 \times 10^{14}] \right\} \cong 1 \text{ Mbps} \qquad (12.1.5)$$

Table 11.4.1 suggests that this data rate is adequate for full-motion video of modest quality. The delay of the signal each way is $\tau = r/c = 10^{11}/[3\times10^8]$ seconds $\cong 5.6$ minutes, impeding conversation. This delay becomes several times greater when Mars is on the far side of the sun from Earth, and the data rate R would then drop by more than a factor of ten.

This 1-Mbps result (12.1.5) assumed 10 photons were required per bit of information. However this can be reduced below one photon per bit by using *pulse-position modulation*. Suppose $\sim10^6$ 1-nsec 10-photon pulses were received per second, where each pulse could arrive in any of 1024 time slots because the ratio of pulse width to average inter-pulse spacing is 1024. This timing information conveys ten bits of information per pulse because $\log_2 1024 = 10$. Since each 10-photon pulse conveys 10 bits of information, the average is one bit per photon received. With more time slots still fewer photons per bit would be required. If a tunable laser can transmit each pulse at any of 1024 colors, for example, then another factor of 10 can be achieved. Use of both pulse position and pulse-frequency modulation can permit more than 10 bits to be communicated per photon on average.

The final example is that of a 1-mW laser diode transmitting digitally modulated light at $\lambda = 5\times10^{-7}$ [m] isotropically within a large office over ranges r up to 10 meters, where the light might travel directly to the isotropic receiver or bounce off walls and the ceiling first. Such optical communications systems might link computers, printers, personal digital assistants (pda's), and other devices within the room. In this case $G_t = 1$ and $A_e = G\lambda^2/4\pi = (5\times10^{-7})^2/4\pi \cong 2\times10^{-14}$ [m^2]. The maximum data rate R can again be found using (12.1.4):

$$R = P_r/E_b \cong \left(1\times10^{-3}/4\pi 10^2\right)\left(2\times10^{-14}\right)/\left(10\times6.6\times10^{-34}\times6\times10^{14}\right) \cong 0.004 \; \left[\text{bits s}^{-1}\right] \quad (12.1.6)$$

The fact that we can send 10^6 bits per second to Mars with a one-watt transmitter, but only 4 millibits per second across a room with a milliwatt, may conflict with intuition.

The resolution of this seeming paradox lies in the assumption that the receiver in this example is a single mode device like that of typical radio receivers or the Martian optical receiver considered above. If this room-link receiver were isotropic and intercepted only a single mode, its effective area A_e given by (12.1.2) would be 2×10^{-14} [m^2]. The tiny effective area of such low-gain coherent optical antennas motivates use of incoherent photodetectors instead, which respond well to the total photon flux from all directions of arrival. For example, intra-room optical links of this type are commonly used for remote control of many consumer electronic devices, but with a much larger multimode antenna (photodiode) of area $A \cong 2\times10^{-6}$ [m^2] instead of 2×10^{-14}. This "antenna" is typically responsive to all photons impacting its area that arrive within roughly one steradian. That is, a photodetector generally intercepts all photons impacting it, even though those photons are incoherent with each other. Thus the solution (12.1.6) is increased by a factor of $10^{-6}/10^{-14}$ if a two-square-millimeter photodetector replaces the single-mode antenna, and R then becomes 0.4 Mbps, which is more capacity than normally required. In practice such inexpensive area detectors are noisier and require orders of mangitude more photons per bit. Better semiconductor detectors can achieve 10 photons per bit or better, however, particularly at visible wavelengths and if stray light at other wavelengths is filtered out.

12.2   Optical waveguides

12.2.1   Dielectric slab waveguides

Optical waveguides such as optical fibers typically trap and guide light within rectangular or cylindrical boundaries over useful distances. Rectangular shapes are easier to implement on integrated circuits, while cylindrical shapes are used for longer distances, up to 100 km or more. Exact wave solutions for such structures are beyond the scope of this text, but the same basic principles are evident in dielectric slab waveguides for which the derivations are simpler. *Dielectric slab waveguides* consist of an infinite flat dielectric slab of thickness 2d and permittivity ε imbedded in an infinite medium of lower permittivity ε_o, as suggested in Figure 12.2.1(a) for a slab of finite width in the y direction. For simplicity we here assume $\mu = \mu_o$ everywhere, which is usually the case in practice too.
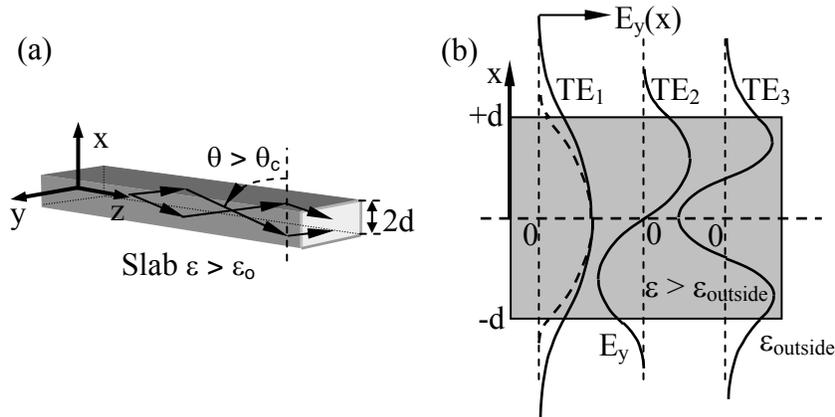
Figure 12.2.1   Dielectric slab waveguide and TE mode structure.

As discussed in Section 9.2.3, uniform plane waves within the dielectric are perfectly reflected at the slab boundary if they are incident beyond the critical angle $\theta_c = \sin^{-1}(c_\varepsilon/c_o)$, where c_ε and c_o are the velocities of light in the dielectric and outside, respectively.   Such a wave and its perfect reflection propagate together along the z axis and form a standing wave in the orthogonal x direction.   Outside the waveguide the waves are evanescent and decay exponentially away from the guide, as illustrated in Figure 12.2.2. This figure portrays the fields inside and outside the lower half of a dielectric slab having $\varepsilon > \varepsilon_o$; the lower boundary is at x = 0. The figure suggests two possible positions for the upper slab boundary that satisfy the boundary conditions for the TE_1 and TE_2 modes. Note that the TE_1 mode waveguide can be arbitrarily thin relative to λ and still satisfy the boundary conditions. The field configurations above the upper boundary mirror the fields below the lower boundary, but are not illustrated here.   These waveguide modes are designated TE_n because the electric field is only transverse to the direction of propagation, and there is part of n half-wavelengths within the slab. The orthogonal modes (not illustrated) are designated TM_n.
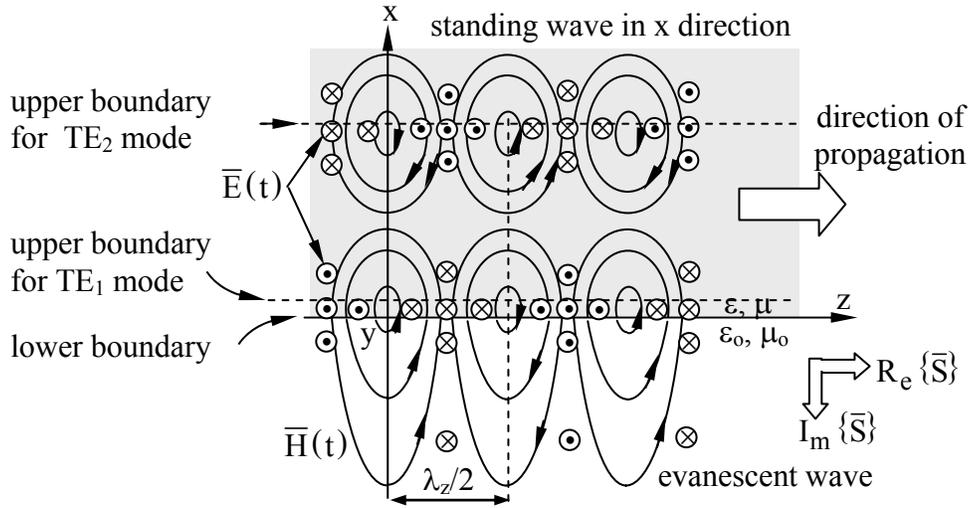
Figure 12.2.2   Fields in dielectric slab waveguides for TE_n modes.

The fields inside a dielectric slab waveguide have the same form as (9.3.6) and (9.3.7) inside parallel-plate waveguides, although the boundary positions are different; also see Figures 9.3.1 and 9.3.3. If we define $x = 0$ at the axis of symmetry, and the thickness of the guide to be 2d, then within the guide the electric field for TE modes is:

$$\overline{E} = \hat{y} \underline{E}_o \{\sin k_x x \ or \ \cos k_x x\} e^{-jk_z z} \qquad\qquad \text{for } |x| \leq d \qquad (12.2.1)$$

The fields outside are the same as for TE waves incident upon dielectric interfaces beyond the critical angle, (9.2.33) and (9.2.34):

$$\overline{E} = \hat{y} \underline{E}_1 e^{-\alpha x - jk_z z} \qquad\qquad \text{for } x \geq d \qquad (12.2.2)$$

$$\overline{E} = \{\text{- } or \text{ +}\} \hat{y} \underline{E}_1 {}^{+\alpha x - jk_z z} \qquad\qquad \text{for } x \leq -d \qquad (12.2.3)$$

The first and second options in braces correspond to anti-symmetric and symmetric TE modes, respectively. Since the waves decay away from the slab, α is positive. Faraday's law in combination with (12.2.1), (12.2.2), and (12.2.3) yields the corresponding magnetic field inside and outside the slab:

$$\overline{H} = \left[\hat{x} \ k_z \{\sin k_x x \ or \ \cos k_x x\} \right.$$
$$\left. + \hat{z} \ jk_x \{\cos k_x x \ or \ \sin k_x x\} \right] (\underline{E}_o / \omega \mu_o) e^{-jk_z z} \qquad \text{for } |x| \leq d \qquad (12.2.4)$$

$$\overline{H} = -(\hat{x} k_z + \hat{z} j\alpha)(\underline{E}_1 / \omega \mu_o) e^{-\alpha x - jk_z z} \qquad\qquad \text{for } x \geq d \qquad (12.2.5)$$

$$\overline{H} = \{+ \ or \ -\}(\hat{x}k_z - \hat{z}j\alpha)(\underline{E}_1/\omega\mu_o)e^{\alpha x - jk_z z} \qquad\qquad \text{for } x \leq -d \qquad (12.2.6)$$

The TE_1 mode has the interesting property that it approaches TEM behavior as $\omega \to 0$ and the decay length approaches infinity; most of the energy is then propagating outside the slab even though the mode is guided by it. Modes with $n \geq 2$ have non-zero cut-off frequencies. There is no TM mode that propagates for $f \to 0$ in dielectric slab waveguides, however.

Although Figure 12.2.1(a) portrays a slab with an insulating medium outside, the first option in brackets $\{\bullet\}$ for the field solutions above is also consistent for $x > 0$ with a slab located $0 < x < d$ and having a perfectly conducting wall at $x = 0$; all boundary conditions are matched; these are the anti-symmetric TE modes. This configuration corresponds, for example, to certain optical guiding structures overlaid on conductive semiconductors.

To complete the TE field solutions above we need additional relations between $\underline{E}_o$ and $\underline{E}_1$, and between k_x and α. Matching $\overline{E}$ at $x = d$ for the symmetric solution [cos $k_x x$ in (12.2.1)] yields:

$$\hat{y}\underline{E}_o \cos(k_x d)e^{-jk_z z} = \hat{y}\underline{E}_1 e^{-\alpha d - jk_z z} \qquad\qquad\qquad (12.2.7)$$

Matching the parallel ($\hat{z}$) component of $\overline{H}$ at $x = d$ yields:

$$-\hat{z}jk_x \sin(k_x d)(\underline{E}_o/\omega\mu_o)e^{-jk_z z} = -\hat{z}j\alpha(\underline{E}_1/\omega\mu_o)e^{-\alpha d - jk_z z} \qquad\qquad (12.2.8)$$

The *guidance condition* for the symmetric TE dielectric slab waveguide modes is given by the ratio of (12.2.8) to (12.2.7):

$$k_x d \tan(k_x d) = \alpha d \qquad\qquad\qquad \text{(slab guidance condition)} \quad (12.2.9)$$

Combining the following two dispersion relations and eliminating k_z can provide the needed additional relation (12.2.12) between k_x and α:

$$k_z^2 + k_x^2 = \omega^2\mu_o\varepsilon \qquad\qquad\qquad \text{(dispersion relation inside)} \quad (12.2.10)$$

$$k_z^2 - \alpha^2 = \omega^2\mu_o\varepsilon_o \qquad\qquad\qquad \text{(dispersion relation outside)} \quad (12.2.11)$$

$$k_x^2 + \alpha^2 = \omega^2(\mu_o\varepsilon - \mu_o\varepsilon_o) > 0 \qquad\qquad \text{(slab dispersion relation)} \quad (12.2.12)$$

By substituting into the guidance condition (12.2.9) the expression for α that follows from the slab dispersion relation (12.2.12) we obtain a transcendental guidance equation that can be solved numerically or graphically:

$$\tan k_x d = \left(\left[\omega^2 \mu_o (\varepsilon - \varepsilon_o) d^2 / k_x^2 d^2 \right] - 1 \right)^{0.5} \qquad \text{(guidance equation)} \quad (12.2.13)$$

Figure 12.2.3 plots the left- and right-hand sides of (12.2.13) separately, so the modal solutions are those values of $k_x d$ for which the two families of curves intersect.
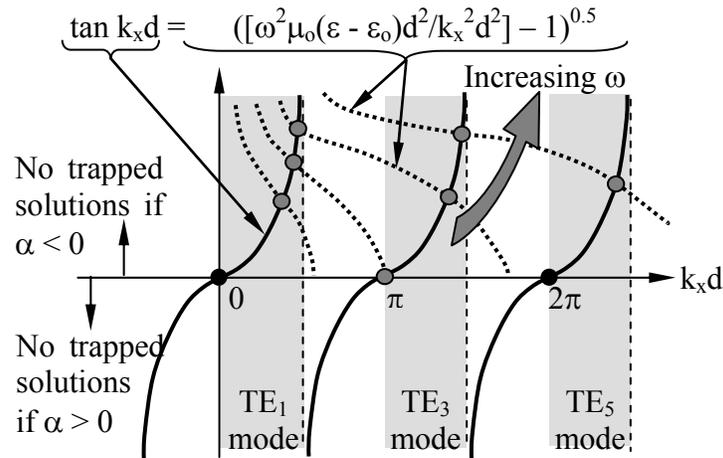
Figure 12.2.3 TE modes for a dielectric slab waveguide.

Note that the TE_1 mode can be trapped and propagate at all frequencies, from nearly zero to infinity. At low frequencies the waves guided by the slab have small values of α and decay very slowly away from the slab so that most of the energy is actually propagating in the z direction outside the slab rather than inside.   The value of α can be found from (12.2.12), and it approaches zero as both $k_x d$ and ω approach zero.

The TE_3 mode cannot propagate near zero frequency however. Its cutoff frequency ω_{TE3} occurs when $k_x d = \pi$, as suggested by Figure 12.2.3; ω_{TE3} can be determined by solving (12.2.12) for this case. This and all higher modes cannot be trapped at low frequencies because then the plane waves that comprise them impact the slab wall at angles beyond θ_c that permit escape. As ω increases, more modes can propagate. Figures 12.2.2 and 12.2.1(b) illustrate symmetric TE_1 and TE_3 modes, and the antisymmetric TE_2 mode. Similar figures could be constructed for TM modes.

These solutions for dielectric-slab waveguides are similar to the solutions for optical fibers, which instead take the form of Bessel functions because of their cylindrical geometry. In both cases we have lateral standing waves propagating inside and evanescent waves propagating outside.

12.2.2   Optical fibers

An *optical fiber* is generally a very long solid glass wire that traps lightwaves inside as do the dielectric slab waveguides described in Section 12.2.1. Fiber lengths can be tens of kilometers or

more. Because the fiber geometry is cylindrical, the electric and magnetic fields inside and outside the fiber are characterized by *Bessel functions*, which we do not address here. These propagating electromagnetic fields exhibit lateral standing waves inside the fiber and evanescence outside. To minimize loss the fiber core is usually overlaid with a low-permittivity glass cladding so that the evanescent decay also occurs within low-loss glass.

A typical glass optical fiber transmission line is perhaps 125 microns in diameter with a high-permittivity glass core having diameter ~6 microns. The core permittivity $\varepsilon + \Delta\varepsilon$ is typically ~2 percent greater than that of the cladding (ε). If the lightwaves within the core impact the cladding beyond the critical angle θ_c, where:

$$\theta_c = \sin^{-1}\left(\varepsilon/(\varepsilon + \Delta\varepsilon)\right) \tag{12.2.14}$$

then these waves are perfectly reflected and trapped. The evanescent waves inside the cladding decay approximately exponentially away from the core to negligible values at the outer cladding boundary, which is often encased in plastic about 0.1 mm thick that may be reinforced. Graded-index fibers have a graded transition in permittivity between the core and cladding. Some fibers propagate multiple modes that travel at different velocities so as to interfere at the output and limit information extraction (data rate). Multiple fibers are usually bundled inside a single cable. Figure 12.2.4 suggests the structure of a typical fiber.
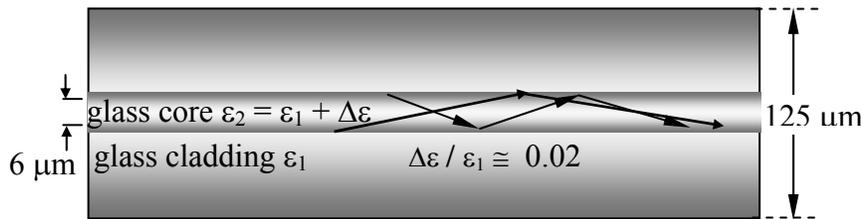
Figure 12.2.4   Typical clad optical fiber.

Figure 12.2.5 shows four common forms of optical fiber; many others exist. The multimode fiber is thicker and propagates several modes, while the single-mode fiber is so thin that only one mode can propagate. The diameter of the core determines the number of propagating modes. In all cylindrical structures, even single-mode fibers, both vertically and horizontally polarized waves can propagate independently and therefore may interfere with each other when detected at the output. If a single-mode fiber has an elliptical cross-section, one polarization can be made to escape so the signal becomes pure. That is, one polarization decays more slowly away from the core so that it sees more of the absorbing material that surrounds the cladding.
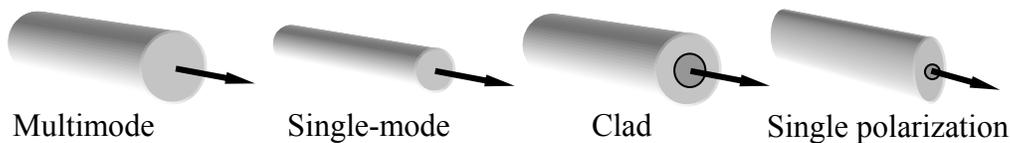
Figure 12.2.5   Types of optical fiber.

The initial issue faced in the 1970's by designers of optical fibers was propagation loss. Most serious was absorption due to residual levels of impurities in the glass, so much research and development involved purification. Water posed a particularly difficult problem because one of its harmonics fell in the region where attenuation in glass was otherwise minimum, as suggested in Figure 12.2.6.
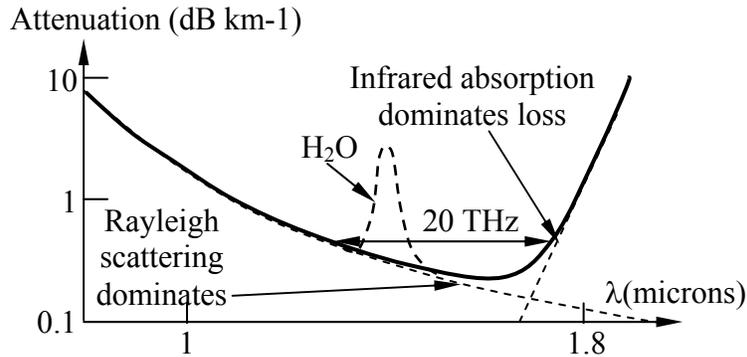
Figure 12.2.6 Loss mechanisms in optical fibers.

At wavelengths shorter than ~1.5 microns the losses are dominated by Rayleigh scattering of the waves from the random fluctuations in glass density on atomic scales. These scattered waves exit the fiber at angles less than the critical angle. *Rayleigh scattering* is proportional to f^4 and occurs when the inhomogeneities in ε are small compared to $\lambda/2\pi$. Inhomogeneities in glass fibers have near-atomic scales, say 1 nm, whereas the wavelength is more than 1000 times larger. Rayleigh scattering losses are reduced by minimizing unnecessary inhomogeneities through glass purification and careful mixing, and by decreasing the critical angle. Losses due to scattering by rough fiber walls are small because drawn glass fibers can be very smooth and little energy impacts the walls.

At wavelengths longer than ~1.5 microns the wings of *infrared absorption* lines at lower frequencies begin to dominate. This absorption is due principally to the vibration spectra of inter-atomic bonds, and is unavoidable. The resulting low-attenuation band centered near 1.5-microns between the Rayleigh and IR attenuating regions is about 20 THz wide, sufficient for a single fiber to provide each person in the U.S.A. with a bandwidth of $20 \times 10^{12}/2.5 \times 10^8 = 80$ kHz, or 15 private telephone channels! Most fibers used for local distribution do not operate anywhere close to this limit for lack of demand, although some undersea cables are pushing toward it.

The fibers are usually manufactured first as a preform, which is a glass rod that subsequently can be heated at one end and drawn into a fiber of the desired thickness. Preforms are either solid or hollow. The solid ones are usually made by vapor deposition of SiO_2 and GeO_2 on the outer surface of the initial core rod, which might be a millimeter thick. By varying the mixture of gases, usually $Si(Ge)Cl_4 + O_2 \Rightarrow Si(Ge)O_2 + 2Cl_2$, the permittivity of the deposited glass cladding can be reduced about 2 percent below that of the core. The boundary between core and cladding can be sharp or graded in a controlled way. Alternatively, the preform cladding is large and hollow, and the core is deposited from the inside by hot gases in

the same way; upon completion there is still a hole through the middle of the fiber. Since the core is small compared to the cladding, the preforms can be made more rapidly this way. When the preform is drawn into a fiber, any hollow core vanishes. Sometimes a hollow core is an advantage. For example, some newer types of fibers have cores with laterally-periodic lossless longitudinal hollows within which much of the energy can propagate.

Another major design issue involves the *fiber dispersion* associated with frequency-dependent phase and group velocities, where the *phase velocity* $v_p = \omega/k$. If the *group velocity* v_g, which is the velocity of the envelope of a narrowband sinusoid, varies over the optical bandwidth, then the signal waveform will increasingly distort as it propagates because the faster moving frequency components of the envelope will arrive early. For example, a digital pulse of light that lasts T seconds is produced by multiplying a *boxcar modulation* envelope (the T-second pulse shape) by the sinusoidal optical carrier, so the frequency spectrum is the convolution of the spectrum for the sinusoid (a spectral impulse) and the spectrum for a boxcar pulse ($\propto [\sin(2\pi t/T)]/[2\pi t/T]$). The outermost frequencies suffer from dispersion the most, and these are primarily associated with the sharp edges of the pulse.

The group velocity v_g derived in (9.5.20) is the slope of the dispersion relation at the optical frequency of interest:

$$v_g = (\partial k/\partial \omega)^{-1} \tag{12.2.15}$$

Figure 12.2.7 illustrates the dispersion relation for three different modes; the higher order modes propagate information more slowly.
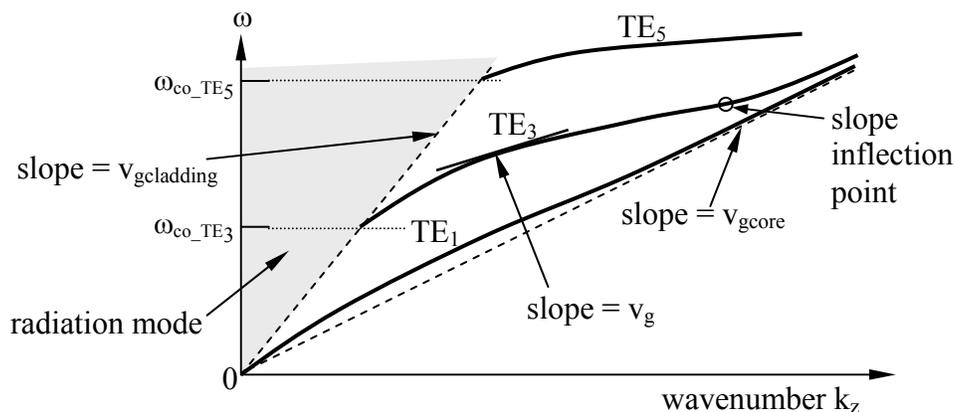
Figure 12.2.7   Group velocities for optical fiber modes.

The group velocity v_g is the slope of the $\omega(k)$ relation and is bounded by the slopes associated with the core (v_{gcore}) and with the cladding ($v_{gcladding}$), where the cladding is assumed to be infinite. The figure has greatly exaggerated the difference in the slope between the core and cladding for illustrative purposes.

A dispersive line eventually transforms a square optical pulse into a long "frequency chirped" pulse with the faster propagating frequencies in the front and the slower propagating frequencies in the back. This problem can be minimized by carefully choosing combinations of: 1) the dispersion n(f) of the glass, 2) the permittivity contour $\varepsilon(r)$ in the fiber, and 3) the optical center frequency f_o. Otherwise we must reduce either the bandwidth of the signal or the length of the fiber. To increase the distance between amplifiers the dispersion can be compensated periodically by special fibers or other elements with opposite dispersion.

Pulses spread as they propagate over distance L because their outermost frequency components ω_1 and $\omega_2 = \omega_1 + \Delta\omega$ have arrival times at the output separated by:

$$\Delta t = L/v_{g1} - L/v_{g2} = L\left[d\left(v_g^{-1}\right)/d\omega\right]\Delta\omega = L\left(d^2k/d\omega^2\right)\Delta\omega \qquad (12.2.16)$$

where v_{gi} is the group velocity at ω_i (12.2.15). Typical pulses of duration T_p have a bandwidth $\Delta\omega \cong T_p^{-1}$, so brief pulses spread faster. The spread Δt is least at frequencies where $d^2k/d\omega^2 \cong 0$, which is near the representative slope inflection point illustrated in Figure 12.2.7.

This natural fiber dispersion can, however, help solve the problem of fiber nonlinearity. Since attenuation is always present in the fibers, the amplifiers operate at high powers, limited partly by their own nonlinearities and those in the fiber that arise because ε depends very slightly on the field strength E. The effects of non-linearities are more severe when the signals are in the form of isolated high-energy pulses. Deliberately dispersing and spreading the isolated pulses before amplifying and introducing them to the fiber reduces their peak amplitudes and the resulting nonlinear effects. This pre-dispersion is made opposite to that of the fiber so that the fiber dispersion gradually compensates for the pre-dispersion over the full length of the fiber. That is, if the fiber propagates high frequencies faster, then those high frequency components are delayed correspondingly before being introduced to the fiber. When the pulses reappear in their original sharp form at the far end of the fiber their peak amplitudes are so weak from natural attenuation that they no longer drive the fiber nonlinear.

___

Example 12.2A
If 10-ps pulses are used to transmit data at 20 Gbps, they would be spaced 5×10^{-10} sec apart and would therefore begin to interfere with each other after propagating a distance L_{max} sufficient to spread those pulses to widths of 50 ps. A standard single-mode optical fiber has dispersion $d^2k/d\omega^2$ of 20 ps²/km at 1.5 μm wavelength. At what distance L_{max} will such 10-ps pulses have broadened to 50 ps?

Solution: Using (12.2.16) and $\Delta\omega \cong T_p^{-1}$ we find:

$$L_{max} = \Delta t/[\Delta\omega(d^2k/d\omega^2)] = 50 \text{ ps} \times 10 \text{ ps }/(20 \text{ ps}^2/\text{km}) = 25 \text{ km}$$

Thus we must slow this fiber to 10 Gbps if the amplifiers are 50 km apart.

___

12.3 *Lasers*

12.3.1 <u>Physical principles of stimulated emission and laser amplification</u>

Lasers (Light Amplification by Stimulated Emission of Radiation) amplify electromagnetic waves at wavelengths ranging from radio to ultraviolet and x-rays. They were originally called *masers* because the first units amplified only microwaves. Lasers can also oscillate when the amplified waves are reflected back into the device. The physical principles are similar at all wavelengths, though the details differ. Laser processes can occur in solids, liquids, or gases.

Lasers have a wide and growing array of applications. For example, optical fiber communications systems today commonly use *Erbium-doped fiber amplifiers* (EDFA's) that amplify ~1.5-micron wavelength signals having bandwidths up to ~4 THz. Semiconductor, gas, and glass fiber laser amplifiers are also used to communicate within single pieces of equipment and for local fiber or free-space communications. Lasers also generate coherent beams of light used for measuring distances and angles; recording and reading data from memory devices such as CD's and DVD's; and for cutting, welding, and shaping materials, including even the human eye. Laser pointers have been added to pocket pens while higher-power industrial units can cut steel plates several inches thick. Weapons and laser-driven nuclear fusion reactions require still higher-power lasers. Peak laser pulse powers can exceed 10^{15} watts, a thousand times the total U.S. electrical generating capacity of ~5×10^{11} watts. The electric field strengths within a focal spot of <100-micron diameter can strip electrons from atoms and accelerate them to highly relativistic velocities within a single cycle of the radiation. The roles of lasers in science, medicine, industry, consumer products, and other fields are still being defined.

Laser operation depends intimately upon the quantum nature of matter and the fact that charges trapped in atoms and molecules generally move at constant energy without radiating. Instead, transitions between atomic or molecular energy states occur abruptly, releasing or absorbing a photon.[69] This process and lasers can fortunately be understood semi-classically without reference to a full quantum description.

Electrons within atoms, molecules, and crystals occupy discrete *energy states*; the lower energy states are preferentially occupied. Energy states can also be vibrational, rotational, magnetic, chemical, nuclear, etc.[70] The number of possible states greatly exceeds those that are occupied.

---

[69] Alternatively, acoustic phonons with energy hf can be released or absorbed, or an additional molecular or atomic state transition can occur to conserve energy. Phonons are acoustic quanta associated with mechanical waves in materials. Optical transitions can also absorb or emit two photons with total energy equal to $E_2 - E_1$, although such *two-photon transitions* are much less likely.

[70] The distances between adjacent nuclei in molecules can oscillate sinusoidally with quantized amplitudes and frequencies characterisitic of each vibrational state. Isolated molecules can spin at specific frequencies corresponding to various rotational energy states. Electron spins and orbits together have magnetic dipole moments that align with or oppose an applied magnetic field to a quantized degree. Atoms bond to one another in quantized ways having specific chemical consequences. Nuclear magnetic moments can also align with other atomic or molecular magnetic moments in quantized ways corresponding to discrete energy states.

For example, as illustrated in Figure 12.3.1(a), an electron trapped in an atom, molecule, or crystal with energy E_1 can be excited into any vacant higher-energy state (E_2) by absorbing a photon of frequency f and energy ΔE where:

$$\Delta E = E_2 - E_1 = hf \; [J] \tag{12.3.1}$$

The constant h is *Planck's constant* (6.625×10^{-34} [Js]), and the small circles in the figure represent electrons in specific energy states.
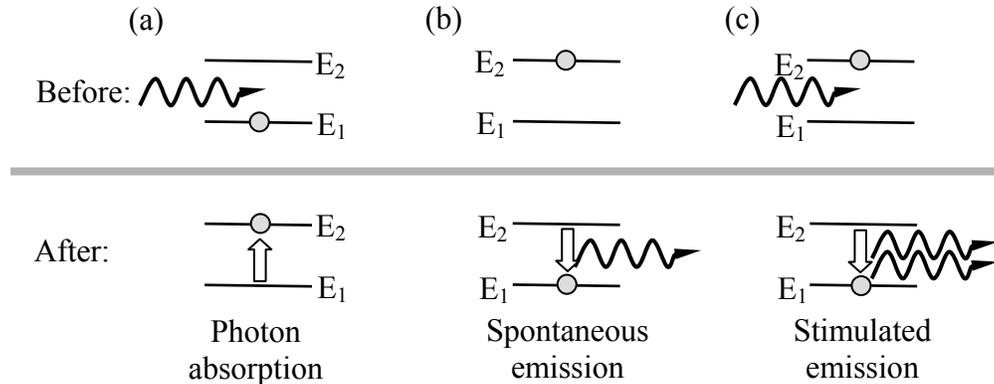
Figure 12.3.1   Photon absorption, spontaneous emission, and stimulated emission.

Figures 12.3.1(b) and (c) illustrate two additional basic photon processes: *spontaneous emission* and *stimulated emission*. *Photon absorption* (a) occurs with a probability that depends on the photon flux density [Wm^{-2}], frequency [Hz], and the cross-section for the energy transition of interest. Spontaneous emission of photons (b) occurs with a probability A that depends only on the transition, as discussed below. Stimulated emission (c) occurs when an incoming photon triggers emission of a second photon; the emitted photon is always exactly in phase with the first, and propagates in the same direction. Laser action depends entirely on this third process of stimulated emission, while the first two processes often weaken it.

The net effect of all three processes—absorption, spontaneous emission, and stimulated emission—is to alter the relative populations, N_1 and N_2, of the two energy levels of interest. An example exhibiting these processes is the Erbium-doped fiber amplifiers commonly used to amplify optical telecommunications signals near 1.4-micron wavelength on long lines. Figure 12.3.2 illustrates how an optical fiber with numerous atoms excited by an optical pump (discussed further below) can amplify input signals at the proper frequency. Since the number of excited atoms stimulated to emit is proportional to the input wave intensity, perhaps only one atom might be stimulated to emit initially (because the input signal is weak), producing two in-phase photons—the original plus the one stimulated. These two then propagate further stimulating two emissions so as to yield four in-phase photons.
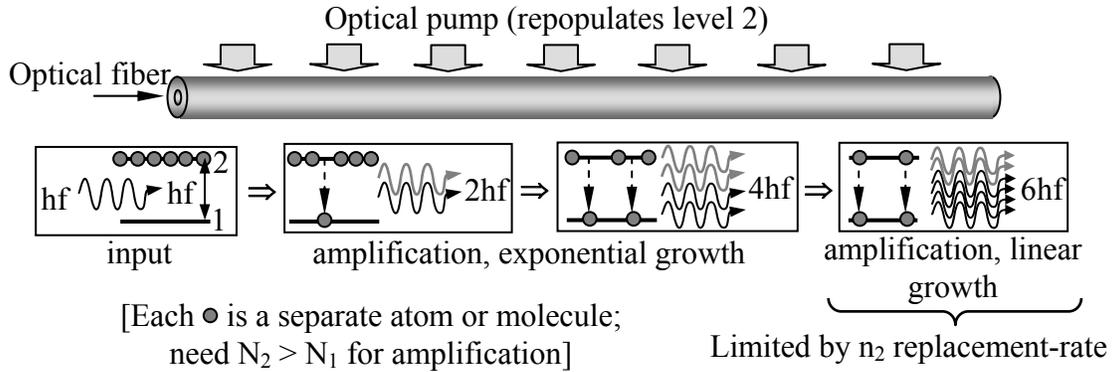
Figure 12.3.2 Optical fiber amplifier with exponential and linear growth.

This exponential growth continues until the pump can no longer empty E_1 and refill E_2 fast enough; as a result absorption $[m^{-1}]$ approaches emission $[m^{-1}]$ as N_1 approaches N_2 locally. In this limit the increase in the number of photons per unit length is limited by the number n_p of electrons pumped from E_1 to E_2 per unit length. Thereafter the signal strength then increases only linearly with distance rather than exponentially, as suggested in Figure 12.3.3; the power increase per unit length then approaches $n_p hf$ $[Wm^{-1}]$.
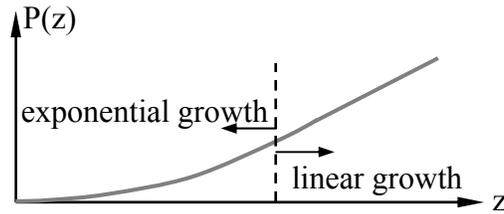
Figure 12.3.3 Exponential and linear growth regimes in optical fiber amplifiers.

Simple equations characterize this process quantitatively. If $E_1 < E_2$ were the only two levels in the system, then:

$$dN_2/dt = -A_{21}N_2 - I_{21}B_{21}(N_2 - N_1) \ \left[s^{-1}\right] \tag{12.3.2}$$

The probability of spontaneous emission from E_2 to E_1 is A_{21}, where $\tau_{21} = 1/A_{21}$ is the $1/e$ lifetime of state E_2. The intensity of the incident radiation at $f = (E_2-E_1)/h$ [Hz] is:

$$I_{21} = F_{21}hf \ \left[Wm^{-2}\right] \tag{12.3.3}$$

where F_{21} is the photon flux [photons $m^{-2}s^{-1}$] at frequency f. The right-most term of (12.3.2) corresponds to the difference between the number of stimulated emissions ($\propto N_2$) and absorptions ($\propto N_1$), where the rate coefficients are:

$$B_{21} = A_{21}\left(\pi^2 c^2 / h\omega^3 n^2\right)\left[m^2 J^{-1}\right] \qquad (12.3.4)$$

$$A_{21} = 2\omega^3 D_{21}^2 / h\epsilon c^3 \left[s^{-1}\right] \qquad (12.3.5)$$

In these equations n is the index of refraction of the fiber and D_{21} is the quantum mechanical electric or magnetic dipole moment specific to the state-pair 2,1. It is the sharply varying values of the *dipole moment* D_{ij} from one pair of levels to another that makes pumping practical, as explained below.

Laser amplification can occur only when N_2 exceeds N_1, but in a two-level system no pump excitation can accomplish this; even infinitely strong incident radiation I_{21} at the proper frequency can only equalize the two populations via (12.3.2).[71] Instead, three- or four-level lasers are generally used. The general principle is illustrated by the *three-level laser* of Figure 12.3.4(a), for which the optical *laser pump radiation* driving the 1,3 transition is so strong that it roughly equalizes N_1 and N_3. The key to this laser is that the spontaneous rate of emission $A_{32} \gg A_{21}$ so that all the active atoms quickly accumulate in the metastable long-lived level 2 in the absence of stimulation at f_{21}. This generally requires $D_{32} \gg D_{21}$, and finding materials with such properties for a desired laser frequency can be challenging.
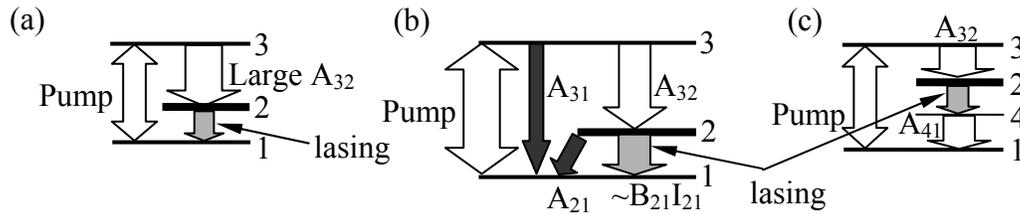
Figure 12.3.4   Energy diagrams for three- and four-level lasers.

Since it requires hf_{13} Joules to raise each atom to level 3, and only hf_{21} Joules emerges as amplified additional radiation, the power efficiency η (power out/power in) cannot exceed the intrinsic limit $\eta_I = f_{21}/f_{31}$. In fact the efficiency is lowered further by a factor of η_A corresponding to spontaneous emission from level 3 directly to level 1, bypassing level 2 as suggested in Figure 12.3.4(b), and to the spontaneous decay rate A_{21} which produces radiation that is not coherent with the incoming signal and radiates in all directions. Finally, only a fraction η_p of the pump photons are absorbed by the transition $1 \rightarrow 2$. Thus the maximum power efficiency for this laser in the absence of propagation losses is:

$$\eta = \eta_I \eta_A \eta_p \qquad (12.3.6)$$

---

[71] Two-level lasers have been built, however, by physically separating the excited atoms or molecules from the unexcited ones. For example, excited ammonia molecules can be separated from unexcited ones by virtue of their difference in deflection when a beam of such atoms in vacuum passes through an electric field gradient.

Figure 12.3.4(c) suggests a typical design for a four-level laser, where both A_{32} and A_{41} are much greater than A_{24} or A_{21} so that energy level 2 is metastable and most atoms accumulate there in the absence of strong radiation at frequency f_{24} or f_{21}. The strong pump radiation can come from a laser, flash lamp, or other strong radiation source. Sunlight, chemical reactions, nuclear radiation, and electrical currents in gases pump some systems.

The ω^3 dependence of A_{21} (12.3.5) has a profound effect on maser and laser action. For example, any two-level maser or laser must excite enough atoms to level 2 to equal the sum of the stimulated and spontaneous decay rates. Since the spontaneous decay rate increases with ω^3, the pump power must also increase with ω^3 times the energy hf of each excited photon. Thus pump power requirements increase very roughly with ω^4, making construction of x-ray or gamma-ray lasers extremely difficult without exceptionally high pump powers; even ultraviolet lasers pose a challenge. Conversely, at radio wavelengths the spontaneous rates of decay are so extremely small that exceedingly low pump powers suffice, as they sometimes do in the vast darkness of interstellar space.

Many types of *astrophysical masers* exist in low-density interstellar gases containing H_2O, OH, CO, and other molecules. They are typically pumped by radiation from nearby stars or by collisions occurring in shock waves. Sometimes these lasers radiate radially from stars, amplifying starlight, and sometimes they spontaneously radiate tangentially along linear circumstellar paths that have minimal relative Doppler shifts. Laser or maser action can also occur in darkness far from stars as a result of molecular collisions. The detailed frequency, spatial, and time structures observed in astrophysical masers offer unique insights into a wide range of astrophysical phenomena.

Example 12.3A

What is the ratio of laser output power to pump power for a three-level laser like that shown in Figure 12.3.4(a) if: 1) all pump power is absorbed by the $1\rightarrow3$ transition, 2) $N_2 \gg N_1$, 3) $A_{21}/I_{21}B_{21} = 0.1$, 4) $A_{31} = 0.1A_{32}$, and 5) $f_{31} = 4f_{21}$?

Solution: The desired ratio is the efficiency η of (12.3.6) where the intrinsic efficiency is $\eta_I = f_{21}/f_{31} = 0.25$, and the pump absorption efficiency $\eta_p = 1$. The efficiency η_A is less than unity because of two small energy losses: the ratio $A_{31}/A_{32} = 0.1$, and the ratio $A_{21}/I_{21}B_{21} = 0.1$. Therefore $\eta_A = 0.9^2 = 0.81$, and $\eta = \eta_I\eta_A\eta_p = 0.25 \times 0.81 \cong 0.20$.

12.3.2 Laser oscillators

Laser amplifiers oscillate nearly monochromatically if an adequate fraction of the amplified signal is reflected back to be amplified further. For example, the *laser oscillator* pictured in Figure 12.3.5 has parallel mirrors at both ends of a laser amplifier, separated by L meters. One mirror is perfect and the other transmits a fraction T (say ~0.1) of the incident laser power. The roundtrip gain in the absence of loss is e^{2gL}. This system oscillates if the net roundtrip gain at any frequency exceeds unity, where round-trip absorption ($e^{-2\alpha L}$) and the partially transmitting mirror account for most loss.
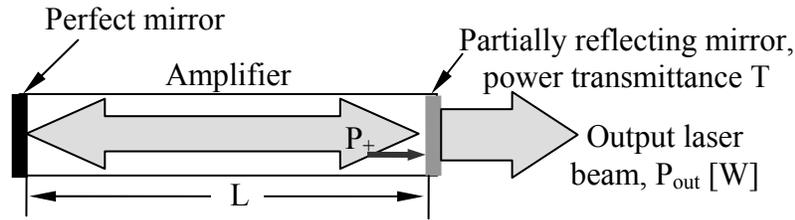
Figure 12.3.5   Laser oscillator.

Amplifiers at the threshold of oscillation are usually in their exponential region, so this net roundtrip gain exceeds unity when:

$$(1-T)e^{2(g-\alpha)L} > 1 \qquad (12.3.7)$$

Equation (12.3.7) implies $e^{2(g-\alpha)L} \geq (1-T)^{-1}$ for oscillation to occur. Generally the gain g per meter is designed to be as high as practical, and then L and T are chosen to be consistent with the desired output power. The pump power must be above the minimum threshold that yields $g > \alpha$.

The output power from such an oscillator is simply $P_{out} = TP_+$ watts, and depends on pump power P_{pump} and laser efficiency. Therefore:

$$P_+ = P_{out}/T = \eta\, P_{pump}/T \qquad (12.3.8)$$

Thus small values of T simply result in higher values of P_+, which can be limited by internet breakdown or failure.

One approach to obtaining extremely high laser pulse powers is to abruptly increase the Q (reverberation) of the laser resonator after the pump source has fully populated the upper energy level. To prevent lasing before that level is fully populated, strong absorption can be introduced in the round-trip laser path to prevent amplification of any stimulated emission. The instant the absorption ceases, i.e. after Q-switching, the average round-trip gain g of the laser per meter exceeds the average absorption α and oscillation commences. At high Q values lasing action is rapid and intense, so the entire upper population is encouraged to emit instantly, particularly if the lower level can be rapidly emptied. Such a device is called a *Q-switched laser*. Resonator Q is discussed further in Section 7.8.

The electronic states of glass fiber amplifiers are usually associated with quantized electron orbits around the added Erbium atoms, and state transitions simply involve electron transfers between two atomic orbits having different energies. In contrast, the most common lasers are *laser diodes*, which are transparent semiconductor p-n junctions for which the electron energy transitions occur between the conduction and valence bands, as suggested in Figure 12.3.6.
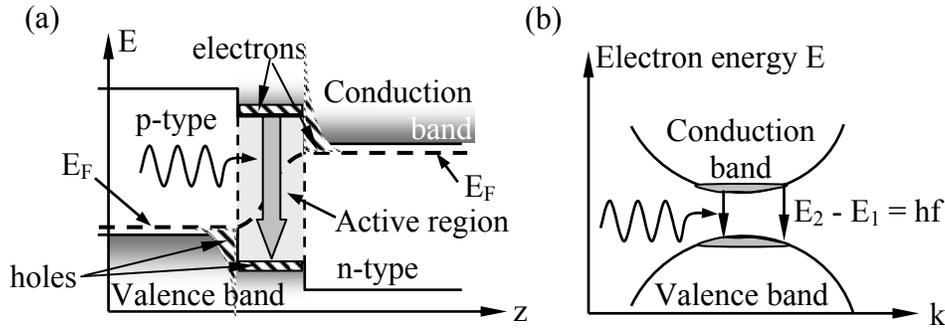
Figure 12.3.6   Laser diode – a forward-biased p-n junction
bounded by mirrors promoting oscillation.

Parallel mirrors at the sides of the *p-n junction* partially trap the laser energy, forming an oscillator that radiates perpendicular to the mirrors; one of the mirrors is semi-transparent. Strong emission does not occur in any other direction because without the mirrors there is no feedback.   Such lasers are pumped by forward-biasing the diode so that electrons thermally excited into the n-type conduction band diffuse into the active region where photons can stimulate emission, yielding amplification and oscillation within the ~0.2-μm thick p-n junction. Vacancies in the valence band are provided by the holes that diffuse into the active region from the p-type region.   Voltage-modulated laser diodes can produce digital pulse streams at rates above 100 Mbps.

The vertical axis E of Figure 12.3.6(a) is electron energy and the horizontal axis is position z through the diode from the p to n sides of the junction.   The exponentials suggest the Boltzmann energy distributions of the holes and electrons in the valence and conduction bands, respectively. Below the *Fermi level*, E_F, energy states have a high probability of being occupied by electrons; $E_F(z)$ tilts up toward the right because of the voltage drop from the p-side to the n-side.   Figure 12.3.6(b) plots electron energy E versus the magnitude of the k vector for electrons (quantum approaches treat electrons as waves characterized by their wavenumber k), and suggests why diode lasers can have broad bandwidths: the energy band curvature with k broadens the *laser linewidth* Δf.   Incoming photons can stimulate any electron in the conduction band to decay to any empty level (hole) in the valence band, and both of these bands have significant energy spreads ΔE, where the linewidth $\Delta f \cong \Delta E/h$ [Hz].

The resonant frequencies of laser diode oscillators are determined by $E_2 - E_1$, the linewidth of that transition, and by the resonant frequencies of the TEM mirror cavity resonator.   The width $\Delta\omega$ of each resonance is discussed further later.   If the mirrors are perfect conductors that force $\overline{E}_{//} = 0$, then there must be an integral number m of half wavelengths within the cavity length L so that $m\lambda_m = 2L$.   The wavelength λ_m' is typically shorter than the free-space wavelength λ_m due to the index of refraction n of the laser material.   Therefore $\lambda_m = 2Ln/m = c/f_m$, and:

$$f_m = cm/2Ln \qquad (12.3.9)$$

For typical laser diodes L and n might be 0.5 mm and 3, respectively, yielding a spacing between cavity resonances of: $c/2Ln = 3\times10^8/(2\times10^{-3}\times1.5) = 100$ GHz, as suggested in Figure 12.3.7(a). The figure suggests how the natural (atomic) laser line width might accommodate multiple cavity resonances, or possibly only one.
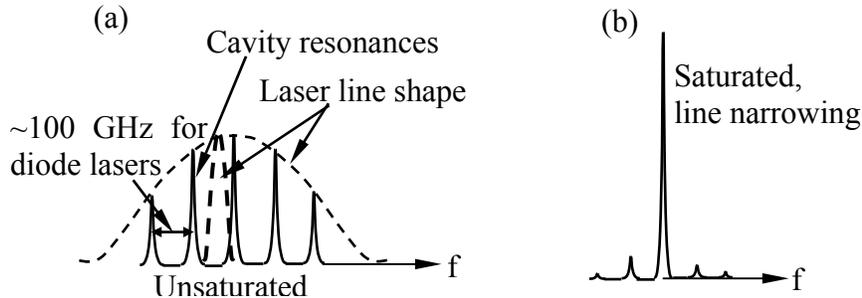
Figure 12.3.7   Line widths and frequencies of the resonances of a cavity laser.

If the amplifier line shape is narrow compared to the spacing between cavity resonances, then the cavity length L might require adjustment in order to place one of the cavity resonances on the line center before oscillations occur. The line width of a laser depends on the widths of the associated energy levels E_i and E_j. These can be quite broad, as suggested by the laser diode energy bands illustrated in Figure 12.3.6(b), or quite narrow. Similarly, the atoms in an EDFA are each subject to slightly different local electrical fields due to the random nature of the glassy structure in which they are imbedded. This results in each atom having slightly different values for E_i so that EFDA's amplify over bandwidths much larger than the bandwidth of any single atom.

Lasers for which each atom has its own slightly displaced resonant frequency due to local fields are said to exhibit *inhomogeneous line broadening*. In contrast, many lasers have no such frequency spread induced by local factors, so that all excited atoms exhibit the same line center and width; these are said to exhibit *homogeneous line broadening*. The significance of this difference is that when laser amplifiers are saturated and operate in their linear growth region, homogeneously broadened lasers permit the strongest cavity resonance within the natural line width to capture most of the energy available from the laser pump, suppressing the rest of the emission and narrowing the line, as suggested in Figure 12.3.7(b). This suppression of weak resonances is reduced in inhomogeneously broadened lasers because all atoms are pumped equally and have their own frequency sub-bands where they amplify independently within the natural line width.

In gases the width of any spectral line is also controlled by the frequency of molecular collisions. Figure 12.3.8(b) illustrates how an atom or molecule with sinusoidal time variations in its dipole moment might be interrupted by collisions that randomly reset the phase. An electromagnetic wave interacting with this atom or molecule would then see a less pure sinusoid. This new spectral characteristic would no longer be a spectral impulse, i.e., the Fourier transform of a pure sinusoid, but rather the transform of a randomly interrupted sinusoid, which has the *Lorentz line shape* illustrated in Figure 12.3.8(a).   Its half-power width is Δf, which is approximately the collision frequency divided by 2π. The limited lifetime of an atom or

molecule in any state due to the probability A of spontaneous emission results in similar broadening, where $\Delta f \cong A/2\pi$; this is called the intrinsic *linewidth* of that transition.
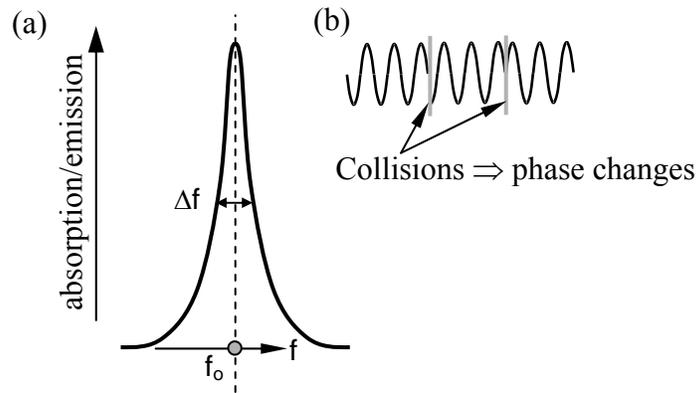
Figure 12.3.8   Lorentzian line shape and origins of intrinsic line width.

Example 12.3B
A Q-switched 1-micron wavelength laser of length L = 1 mm is doped with 10^{18} active atoms all pumped to their upper state. When the Q switches instantly to 100, approximately what is the maximum laser power output P [W]? Assume $\varepsilon = 4\varepsilon_o$.

Solution: The total energy released when the Q switches is $10^{18}hf \cong 10^{18} \times 6.6\times10^{-34} \times 3\times10^{14} =$ 0.20 Joules. If the laser gain is sufficiently high, then a triggering photon originating near the output could be fully amplified by the time the beam reaches the rear of the laser, so that all atoms would be excited as that reflected pulse emerges from the front of the laser. A triggering photon at the rear of the laser would leave some atoms unexcited. Thus the minimum time for full energy release lies between one and two transit times τ of the laser, depending on its gain; $\tau = L/c' = 2L/c = 6.7\times10^{-12}$. Lower laser gains may require many transit times before all atoms are stimulated to emit. Therefore $P < \sim 0.2/(6.7\times10^{-12}) \cong 30$ GW.

12.4   *Optical detectors, multiplexers, interferometers, and switches*

12.4.1   Phototubes

Sensitive radio-frequency detectors typically require at least 10^{-20} Joules per bit of information, which roughly corresponds to thousands of photons of energy hf, where Planck's constant h = 6.625×10^{-34} Joules Hz^{-1}. This number of photons is sufficiently high that we can ignore most quantum effects and treat the arriving radio signals as traditional waves. In contrast, many optical detectors can detect single photons, although more than five photons are typically used to distinguish each pulse from interference; this requires more energy per bit than is needed at radio wavelengths. The advantage of long-range optical links lies instead in the extremely low losses of optical fibers or, alternatively, in the ability of relatively small mirrors or telescopes to focus

energy in extremely small beams so as to achieve much higher gains than can practical radio antennas.

Typical photon detectors include phototubes and semiconductors. A *phototube* detects photons having energies hf > Φ using the *photoelectric effect*, where Φ is the *work function* [J] of the metal surface (*cathode*) that intercepts the photons. Photons with energies above this threshold eject an electron from the cathode with typical probabilities η (called the *quantum efficiency*) of ~10-30 percent. These ejected electrons are then pulled in vacuum toward a positively charged *anode* and contribute to the current I through the load resistor R, as illustrated in Figure 12.4.1(a). Although early phototubes ejected electrons from the illuminated surface, it is now common for the metal to be sufficiently thin and transparent that the electrons are emitted from the backside of the metal into vacuum; the metal is evaporated in a thin layer onto the interior surface of the tube's evacuated glass envelope.
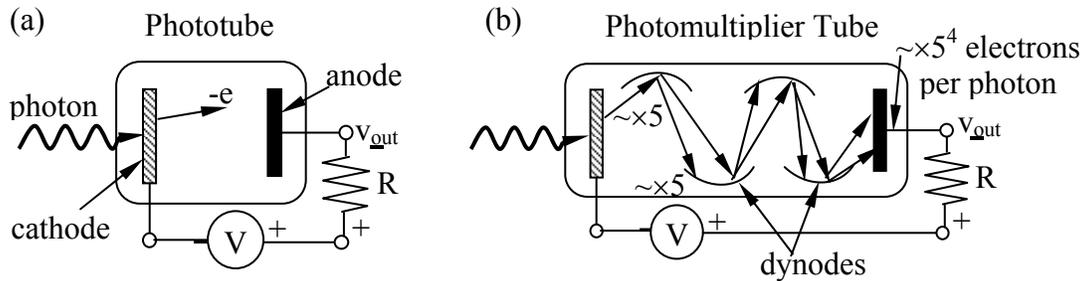
Figure 12.4.1   Phototube and photomultiplier tube detectors.

The current I is proportional to the number N of photons incident per second with energies above Φ:

$$I = -\eta Ne \ [A] \tag{12.4.1}$$

The work functions of most metals are ~2-6 electron volts, where the energy of one *electron volt* (e.v.) is $-eV = 1.602\times10^{-19}$ Joules[72]. Therefore phototubes do not work well for infrared or longer wavelengths because their energy hf is too small; 2 e.v. corresponds to a wavelength of 0.62 microns and the color red.

Because the charge on an electron is small, the currents I are often too small to induce voltages across R (see Figure 12.4.1) that exceed the thermal noise (Johnson noise) of the resistor unless the illumination is bright. *Photomultiplier tubes* release perhaps 10^4 electrons per detected photon so as to overcome this noise and permit each detected photon to be unambiguously counted. The structure of a typical photomultiplier tube is illustrated in Figure 12.4.1(b). Each photoelectron emitted by the cathode is accelerated toward the first *dynode* at ~50-100 volts, and gains energy sufficient to eject perhaps five or more low energy electrons from the dynode that are then accelerated toward the second dynode to be multiplied again. The

---

[72] Note that the energy associated with charge Q moving through potential V is QV Joules, so QV = 1 e.v. = e×1 = 1.602×10^{-19} Joules.

illustrated tube has four dynodes that, when appropriately charged, each multiply the incident electrons by ~5 to yield $~5^4 \cong 625$ electrons at the output for each photon detected at the input. Typical tubes have more dynodes and gains of $~10^4$-10^7. Such large current pulses generally overwhelm the thermal noise in R, so random electron emissions induced by cosmic rays or thermal effects at the cathode dominate the detector noise. The collecting areas of such tubes can be enhanced with lenses or mirrors.

12.4.2 Photodiodes

Phototubes are generally large (several cubic inches), expensive, and fragile, and therefore semiconductor *photodiodes* are more commonly used. Photodiodes also respond better to visible and infrared wavelengths and operate at much lower voltages. Figure 12.4.2(a) illustrates the *energy diagram* for a typical short-circuited *p-n junction* between p-type and n-type semiconductors, where the vertical axis is electron energy E and the horizontal axis is distance z perpendicular to the planar junction.
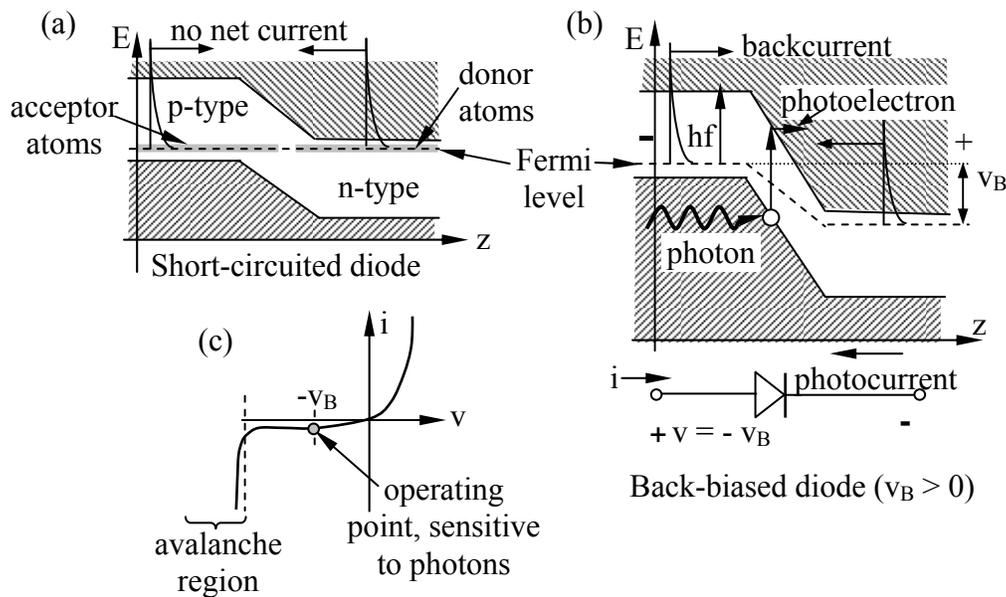
Figure 12.4.2 Semiconductor photodiodes.

The lower cross-hatched area is the *valence band* within which electrons are lightly bound to ions, and the upper area is the *conduction band* within which electrons are free to move in response to electric fields. The *band gap* between these regions is ~1.12 electron volts (e.v.) for silicon, and ranges from 0.16 e.v. for InSb (indium antimonide to ~7.5 e.v. for BN (boron nitride). In metals there is no such gap and some electrons always reside in the conduction band and are mobile. Additional discussion of p-n junctions appears in Section 8.2.4.

Electrons move freely in the conduction band, but not if they remain in the valence band. Most photons entering the junction region with energy greater than the bandgap between the Fermi level and the lower edge of the conduction band can excite electrons into the conduction

band to enhance device conductivity. In semiconductors the *Fermi level* is that level corresponding to the nominal maximum energy of electrons available for excitation into the conduction band. The local Fermi level is determined by impurities in the semiconductors that create electron donor or acceptor sites; these sites easily release or capture, respectively, a free electron. The Fermi level sits just below the conduction band for n-type semiconductors because *donor atoms* easily release one of their electrons into the conduction band, as illustrated in Figure 12.4.2(a) and (b). The Fermi level sits just above the valence band for p-type semiconductors because *acceptor atoms* easily capture an extra electron from bound states in nearby atoms.

If the p-n junction is short-circuited externally, the Fermi level is the same on both halves, as shown in Figure 12.4.2(a). Random *thermal excitation* produces an exponential *Boltzmann distribution* in electron energy, as suggested in the figure, the upper tails of which lie in the conduction band on both halves of the junction. When the device is short-circuited these current flows from thermal excitations in the p and n halves of the junction balance, and the external current is zero. If, however, the diode is back-biased by V_B volts as illustrated in Figure 12.4.2(b), then the two exponential tails do not balance and a net back-current current flows, as suggested by the I-V characteristic for a p-n junction illustrated in Figure 12.4.2(c). The back current for an un-illuminated photodiode approaches an asymptote determined by V_B and the number of thermal electrons excited per second into the conduction band for the p-type semiconductor. When an un-illuminated junction is forward biased, the current increases roughly exponentially.

When a p-n junction is operated as a photodiode, it is back-biased so that every detected photon contributes current flow to the circuit, nearly one electron per photon received. By cooling the photodiode the thermal contribution to diode current can be reduced markedly so that the diode becomes more sensitive to dim light. Cooling is particularly important for photodiodes with the small bandgaps needed for detecting infrared radiation; otherwise the detected infrared signals must be bright so they exceed the detector noise.

If photodiodes are sufficiently back-biased, they can enter the avalanche region illustrated in Figure 12.4.2(c), where an excited electron is accelerated sufficiently as it moves through the semiconductor that it can impact and excite another electron into the conduction band; both electrons can now accelerate and excite even more electrons, exponentially, until they all exit the high-field zone so that further excitations are not possible. In response to a single detected photon such *avalanche photodiodes* (APD's) can produce an output pulse of $\sim 10^4$ electrons that stands out sufficiently above the thermal noise that photons can again be counted individually. The number of photons detected per second is proportional to input power, and therefore to the square of the incident electric field strength.

12.4.3   Frequency-multiplexing devices and filters

The major components in fiber-optic communications systems are the fibers themselves and the optoelectronic devices that manipulate the optical signals, such as detectors (discussed in Sections 12.4.1–2), amplifiers and sources (Section 12.3), multiplexers and filters (this section), modulators, mixers, switches, and others (Section 12.4.4). These are assembled to create useful communications, computing, or other systems.

A typical wave-division multiplexed (WDM) amplifier is pictured in Figure 12.4.3; narrowband optical signals of different colors are aggregated at a point of departure and merged onto a single long fiber by a *frequency multiplexer* (MUX). Along this fiber extremely broadband optical amplifiers (OAMPs) are spaced perhaps 80 km apart to sustain the signal strength. OAMPs today are typically erbium-doped fiber amplifiers (EFDA's). At the far end the signal is de-multiplexed into its spectral components, which are then directed appropriately along separate optical fibers. Before broadband amplifiers were available, each narrow band had to have separate amplifiers and often separate fibers.
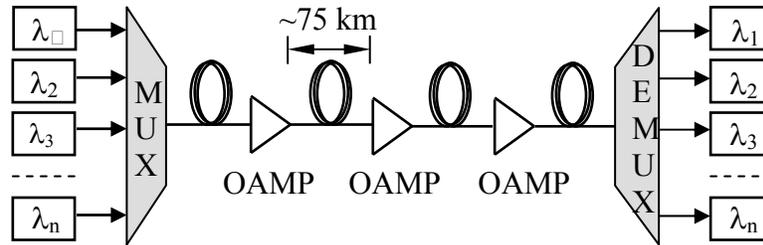
Figure 12.4.3 Wave-division multiplexed amplifier.

Such multiplexers can be made using prisms or diffraction gratings that refract or diffract different colors at different angles, as suggested in Figure 12.4.4(a) and (b); by reciprocity the same devices can be used either for multiplexing (superimposing multiple frequency bands on one beam) or demultiplexing (separation of a single beam into multiple bands), depending on which end of the device receives the input.

The diffraction grating of Figure 12.4.4(b) is typically illuminated by normally incident uniform plane waves, and consists of closely spaced ruled straight lines where equal-width stripes typically alternate between transmission and reflection or absorption. Alternate stripes sometimes differ only in their phase. Each stripe must be more than $\lambda/2$ wide, and λ is more typical. In this case the rays from each transparent stripe 2λ apart will add in phase straight forward ($\theta = 0$) and at $\theta = \sin^{-1}(\lambda/2\lambda) = 30^o$, exactly analogous to the grating lobes of dipole array antennas (Section 10.4). Since the stripe separation (2λ here) is fixed, as the frequency $f = c/\lambda$ varies, so does θ, thus directing different frequencies toward different angles of propagation, much like the prism.

Another useful optical device is the *Fabry-Perot resonator*, which is the optical version of the TEM resonator illustrated in Figure 7.4.3(a) and explained in Section 7.4.3. For example, an optical TEM resonator can be fabricated using parallel mirrors with uniform plane waves trapped between them; the allowed resonator modes have an integral number n of half-wavelengths in the distance L between the parallel conductors:

$$n\lambda_n/2 = L \qquad\qquad\qquad (12.4.2)$$
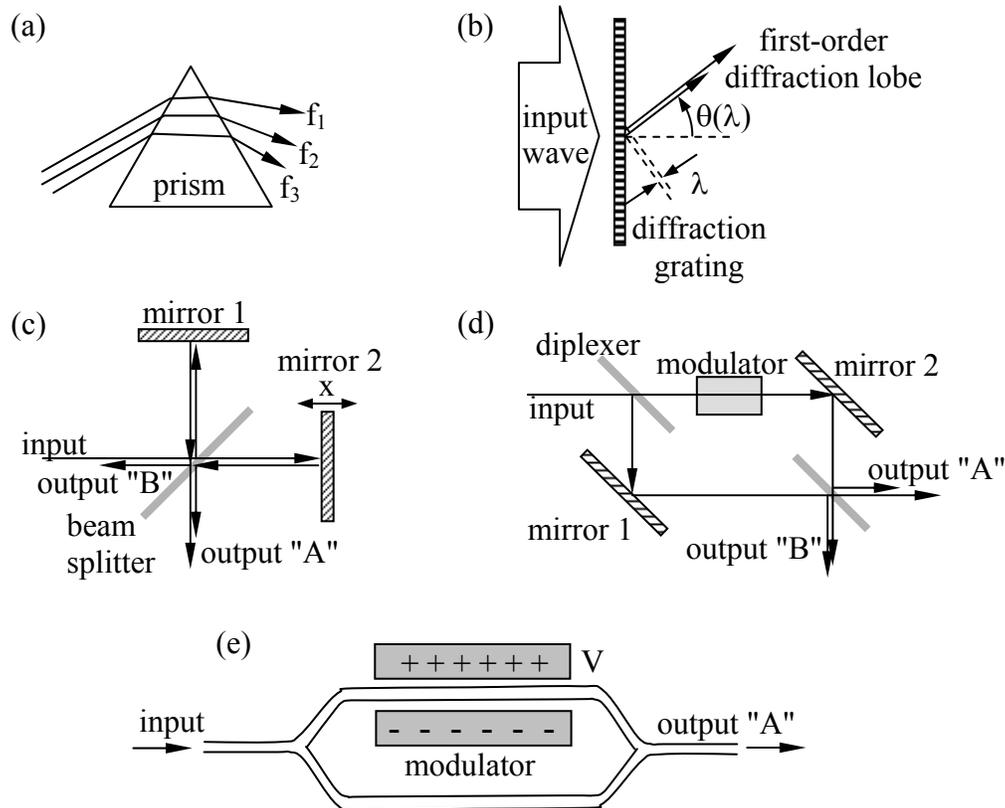
Figure 12.4.4 Optical frequency multiplexers, interferometers, and modulators:
(a) prism, (b) diffraction grating, (c) Michelson interferometer, (d) Mach-Zehnder interferometer,
(e) waveguide Mach-Zehnder interferometer.

Thus the frequency f_n of the nth TEM resonance is:

$$f_n = c/\lambda_n = nc/2L \quad [Hz] \tag{12.4.3}$$

and the separation between adjacent resonances is $c/2L$ [Hz]. For example, if L is 1.5 mm, then the resonances are separated $3 \times 10^8 / 0.003 = 100$ GHz.

    If the input and output mirrors transmit the same small fraction of the power incident upon them, then the "internal" and external Q's of this resonator are the same, where the "internal Q" here (Q_I) is associated with power escaping through the output mirror and the "external Q" (Q_E) is associated with power escaping through the input mirror. As suggested by the equivalent circuits in Figures 7.4.4–5, there is perfect power transmission through the resonator at resonance when the internal and external Q's are the same, provided there are no dissipative losses within the resonator itself. The width of the resonance is then found from (7.4.45–6):

$$\Delta\omega = \omega_o/Q_L = 2\omega_o/Q_E = 2P_E/w_T \tag{12.4.4}$$

The loaded $Q_L = Q_E/2$ when $Q_E = Q_I$, and P_E [W] is the power escaping through the input mirror when the total energy stored in the resonator is w_T [J]. With high reflectivity mirrors and low residual losses the bandwidth of such a resonator can be made almost arbitrarily narrow. At optical frequencies the ratio of cavity length L to wavelength λ is also very large. This increases the ratio of w_T to P_E proportionally and leads to very high Q_L and narrow linewidth.

If the medium in a Fabry-Perot interferometer is dispersive, then it can be shown that the spacing between resonances is $v_g/2L$ [Hz], or the reciprocal of the round-trip time for a pulse. Thus such a resonator filled with an active medium could amplify a single pulse that rattles back and forth in the resonator producing sharp output pulses with a period $2L/v_g$. The Fourier transform of this pulse train is a train of impulses in the frequency domain with spacing $v_g/2L$, i.e., representing the set of resonant frequencies for this resonator. The resonant modes of such a *mode-locked laser* are synchronized, so they can usefully generate pulse trains for subsequent modulation.

Example 12.4A

What is the ratio of the width $\Delta\omega$ of the passband for a Fabry-Perot resonator relative to the spacing $\omega_{i+1} - \omega_i$ between adjacent resonances? What power transmission coefficient $T^2 = |\underline{E}_t|^2/|\underline{E}_i|^2$ is required for each mirror in order to produce isolated sharp resonances? What is the width Δf [Hz] of each resonance?

Solution: The resonance width $\Delta\omega$ and spacing are given by (12.4.4) and (12.4.2), respectively, so:

$$\Delta\omega/(\omega_{i+1} - \omega_i) = (2P_E/w_T)/(\pi c/L) = \left[(2P_+T^2)/(2LP_+/c)\right](L/\pi c) = T^2/\pi <\sim 0.3$$

Therefore we require $T^2 <\sim 1$ so that $\Delta\omega <\sim 0.3(\omega_{i+1} - \omega_i)$.

$\Delta f = \Delta\omega/2\pi = (T^2/\pi)(\omega_{i+1} - \omega_i)/2\pi = (T^2/\pi)c/2L$ [Hz]; it approaches zero as T^2/L does.

12.4.4 Interferometers

The *Michelson interferometer* and the *Mach-Zehnder interferometer* are important devices illustrated in Figure 12.4.4(c) and (d), respectively. In both cases an input optical beam is split by a *beam-splitter* into two coherent beams that are reflected by mirrors and then recombined coherently in a second beam-splitter to form two output beams. The intensity of each output beam depends on whether its two input components added in-phase or out-of-phase. The beam-splitters are typically dielectric mirrors coated so that half the power is reflected and half transmitted from the front surface; the rear surface might be anti-reflection coated. Half-silvered mirrors can also be used.

As the position x of a Michelson mirror varies, the output power varies sinusoidally from zero, which results when the two beams cancel at the output, to its peak value when the two beams add in phase. Cancellation requires that the two beams have equal strength. It is interesting to ask where the input power goes when output "A" is zero; the figure suggests the answer. The missing power emerges from the other output; the sum of the powers emerging

from the two outputs equals the input power, less any dissipative losses. This requirement for power conservation translates into a requirement for a specific phase relationship between the various beams.

Consecutive peaks in output strength occur as the mirror moves $\lambda/2$ (typically $\sim 3 \times 10^{-7}$ m); the factor of 1/2 arises because of the round trip taken by the reflected beam. The sinusoidal output power can generally be measured with sufficient accuracy at optical wavelengths to determine relative mirror positions x with accuracies of an angstrom (10^{-10} m), or tiny fractions thereof; thus the Michelson interferometer is a powerful tool for measuring or comparing wavelengths and distances. Another important application is measurement of optical spectra. Since each optical wavelength λ in the input beam produces an additive sinusoidal contribution to the output power waveform A(x) of period $\lambda/2$, the input optical power spectrum is the Fourier transform of A(2x). Because this technique was first used to analyze infrared spectra, it is called *Fourier transform infrared spectroscopy* (FTIR).

If the two output beams in a Mach-Zehnder interferometer add in phase, the output power is maximized and equals the input power, much like the Michelson interferometer. In either type of interferometer the phase of the optical beam in one arm can be modulated by varying the effective dielectric constant and delay of its propagation medium; certain dielectrics are tunable when biased with large electric fields. In this fashion the output beam power can be modulated by varying the voltage V across the propagation medium, as illustrated in Figure 12.4.4(d) and (e). If the device operates near a transmission null, very little change in refractive index is required to produce a large increase in output power. Such devices can modulate optical power at frequencies of 10 GHz or more.

The Mach-Zehnder interferometer configuration in Figure 12.4.4(e) is widely used for modulators because the waveguides can be integrated on a chip together with other optical components. The output is maximum when the two arms have equal phase delays. When the two merging beams are out of phase the excited waveguide mode is not trapped in the output waveguide but radiates away; the radiated wave corresponds to output B in Figure 12.4.4(d). The same integrated configuration can alternatively be used as a notch filter, eliminating an undesired optical wavelength for which the two arm lengths differ by exactly $\lambda/2$, while passing nearby wavelengths.

12.4.5   Optical switches

Optical switches redirect optical beams just as electrical switches redirect currents. One approach is to use MEMS devices that mechanically move mirrors or shutters to redirect the light beams, which usually are narrow, coherent, and laser-produced. Such devices can switch light beams at rates approaching 1 MHz.

Another approach is to direct the light beam at right angles to a dielectric within which ultrasonic acoustic waves (at radio frequencies) are propagating transverse to the light so as to produce a dynamic phase grating through which the light propagates and diffracts. The configuration is that of Figure 12.4.4(b). The acoustic waves compress and decompress the medium in a wavy pattern; the compressed regions have a slightly higher permittivity and

therefore a slightly lower velocity of light. By making the dielectric sufficiently thick, the cumulative phase variation of the light passing through the device can be $\lambda/2$ or more, thus producing strong diffraction at an angle θ corresponding to the wavelength of light λ and the acoustic wavelength λ_a, where $\theta = \sin^{-1}(\lambda/\lambda_a)$ and $\lambda_a \cong 2\lambda$. In practice the cumulative phase variation is often much less than $\lambda/2$ because of improved simplicity, linearity, and the availability of high input powers that can compensate for the reduced diffractive power efficiency. In this fashion the diffracted beam can be steered among several output ports at rates up to ~1 MHz or more, limited largely by the time it takes the acoustic wave to traverse the diffraction zone. Acoustic velocities in solids are roughly 1000-3000 m/s.

A more important method, however, is the use of Mach-Zehnder interferometers (see Section 12.4.4) to modulate input optical streams, varying their intensity by more than 15 dB at rates up to ~10 GHz, limited by the time it takes the signals modulating the electrical phase length modulator of Figure 12.4.4(e) to propagate across that modulator (e.g. nanoseconds). The detected modulator output signal is the product of the optical and modulator signals. The spectrum of this product contains the convolution of the two input spectra, which exhibits upper and lower sidebands that correspond to the radio frequency signal being communicated.

6.013 Electromagnetics and Applications
Spring 2009