Chapter 1

Introduction to Optical Communication Systems

The purpose of this chapter is to give an overview and a somewhat historical perspective on the field of optical communications. The first section of the chapter describes why there are fundamental reasons that optics is attractive for use in communications. Indeed, the worldwide telephone network, the world's largest communications system, is fiber optic-based. Further, optics is finding progressively more communications applications by the day, not to speak of the numerous sensing applications that are presently being implemented. In some areas, optics is being implemented somewhat more slowly than originally predicted. This is in great part due to cost. In the telephone network, hardware cost was really not an issue compared to a myriad of other costs such as right of way and cable installation, and the initial implementations were carried rapidly rather than cost-effectively. The telecommunications solutions, which were the first to be carried out, could therefore not be taken over directly into other applications. Now, though, the telecommunications market is competitive as are the data communications market and others, and various new cost-effective solutions are appearing. There are numerous applications where optics is competing. The second section of this chapter discusses a very general systems model which can be used to describe essentially any optical system in use at present. The third, fourth, and fifth sections give a somewhat historical perspective on the development of the three most important system components, the source, the fiber, and the detector. Section 6 gives a more broad systems perspective.

1.1 The Electromagnetic Spectrum and What It Tells Us about the Relative Advantages of Optical Communications

Various observations relating to the potential of optical technology for transmission of information can be made from perusal of the frequency line of Table 1.1. The information rates in which one is interested in conventional "modern-day" communications systems generally correspond to audio rates (or multiples thereof) in telephone systems, radio rates (or multiples thereof) in commercial broadcast systems, or digital television rates (or multiples thereof) in the most advanced of video distribution systems. These rates (using a digital television rate of 140 Mbit/sec times perhaps 30 channels) are generally below several gigahertz. If one were to transmit such information without impressing it on an optical carrier but instead on a radio frequency (RF) carrier a bit higher than the maximum rate, the transmission wavelength of the RF carrier would be centimeters or larger. There can, however, be great advantages to using optical carriers. An obvious one is the low loss and directionality of the optical fiber. Clearly, the carrier must have a higher rate than the information rate. A major principle that has, to present, appeared in communications systems has been that the higher the frequency, the greater the technical complexity. Microwaves are harder to handle than are radio waves. As wavelengths decrease to approach the size of circuit components, circuit elements are no longer lumped as such, and leads can act as reflective components and/or antennas and lumped elements as

λ	300km	300m	30cm	3mm	30μm	.3μm	30Å	.03Å
name	audio	radio	μwave	mmwave	IR	OPT/UV	x-ray	gamma
ν	10^3 KHz	10^6 MHz	10^9 GHz	10^{11} THz	10^{13}	10^{15}	10^{17}	10^{20}
ε_p			3μeV	.3meV	30meV	3eV	300eV	.3 MeV

Table 1.1: A frequency line which gives the wavelengths λ, the frequencies ν, and the photon energies $h\nu$ for the various regions of the frequency spectrum named on the line.

electromagnetic resonators (as will be discussed further in Chapter 4). This has generally meant that sending more information would cost more and that there was therefore, in some sense, a cost per bit/sec (bps) of transmitted information in the sense that going to a higher information rate requires a higher frequency. Thus, the first observation from the frequency line would be that, for optical carriers, which have frequencies in the hundreds of terahertz, information bandwidth is in some sense free. That is to say, the optical wavelength is so small compared to most devices that the technology has changed drastically from electrical and microwave. Once we assume that we have such technology, then no matter how high an information rate one might want it will not be necessary to change the carrier, as the carrier frequency is higher than any realistic information rate could become. Bandwidth is not completely free, though, as encoders and decoders must necessarily operate at the information rate, but much of the rest of the system must necessarily handle only the carrier plus modulation. If a component can handle a frequency of 5×10^{14} hertz, an information shift in that frequency of a part in a thousand (corresponding to a 500 gigahertz information rate) will have little or no effect on device performance. Therefore, once the system is already set up, one can upgrade system speed more or less at will without the kind of costs incurred by changing the electromagnetic (i.e. suboptical) carrier in conventional systems. (This is not quite true in long-haul communications at very high bandwidths where fiber dispersion can limit the repeaterless transmission span, but this is a rather special case. However, in any of the non dispersion-limited links, active efforts to increase throughput as traffic increases are being employed. As will be mentioned below, important cases of these are (dense) wavelength-division multiplexing (DWDM—to be discussed in section 11.3 of Chapter 11) and code-division multiple access (CDMA—to be discussed in section 13.?? of Chapter 13).)

A consequence of the size of the optical bandwidth is that the optical carrier can be used to carry many different telephone conversations, television programs, etc., simultaneously. The process by which this is generally carried out (at least in synchronous format) is called time division multiplexing (TDM). The idea is that, if one wishes to multiplex 16 different channels each transmitting at 1 Mbps, one could perform this by dividing each bit period by 16 and then interleaving the bits into a composite 1 μsec bit (1 Mbps rate) which actually carries 16 bits of information on it. With telephone conversations representing a rate of 64 kbps, the 100s of Tbps bandwidth of the optical carrier holds great promise for TDM. Of course, TDM is not the only multiplexing scheme one can imagine using. One could imagine impressing a number of subcarriers, spaced by perhaps some gigahertz, onto the optical carrier. Each of these carriers could then be modulated at an information rate and then reseparated according to their different carrier wavelengths at the output. Such a scheme is referred to as wavelength-division multiplexing (WDM) or subcarrier modulation, depending on the implementation. Many of the present-day schemes for increasing link throughput with increasing traffic involves combining many TDM'd signals onto WDM'd carriers. In fact, the limitation on density of WDM turns out to be not bandwidth but power. That is, each channel requires some amount of power. The more channels, then, the higher the power requirement. At some power level, optical fiber nonlinearity becomes important, and this nonlinearity tends to mix the signals together. There is presently much effort going on in trying to find ways to equalize such nonlinearities.

The high carrier frequency of the optical carrier also has some drawbacks, especially as it relates, through the speed of light, to the optical wavelength (Table 1.1). The optical period corresponds to less than two femtoseconds. This means that phase control corresponds to manipulation of subfemtosecond periods of time. Although techniques to do such are emerging, they are complicated–much more complicated than manipulating microwave or radio frequency waveforms. For this reason, coherent optical reception is still a laboratory technology. The development of the rare earth-doped optical fiber amplifier seems, for the present at any rate, to have obviated the need for coherent techniques in telecommunication as far as improved signal-to-noise ratio goes (as will be discussed in section 11.4 of Chapter 11).

The short period of the optical wave also implies a short wavelength centered around a half of a micron. The smallness of the optical wavelength, therefore, promises to allow for the miniaturization of transmit and receive modules, which should allow considerable reduction in size, weight, and cost of optical communication systems with respect to microwave/radio wave counterparts. One needs to be careful about the size scaling, however. Although the ratio of the optical to the microwave wavelength is roughly 10^{-5}, waveguide dimensions do not quite scale with this factor. A fully metallic (closed) waveguide must always have a dimension of roughly $\lambda/2$. This is not the case with fully dielectric waveguides such as the ones necessary for optical fibers or integrated optical devices. The smallness of the optical wavelength dictates the use of dielectric guides, as even a tiny metallic loss per wavelength would incur a huge loss over propagation distances of millions to billions of wavelengths. In this dielectric waveguide case, the waveguide core dimension is $\lambda/(2\mathrm{NA})$, where NA is the numerical aperture. The numerical aperture is directly related to the index contrast between the core and cladding of the dielectric waveguide. By nature, the index contrast must be low, as if it were not, this would indicate that there was a big difference between the core and cladding material, which would naturally lead to a large scattering loss. A typical NA of 0.2 leads to waveguides with a characteristic dimension of about 5λ, or a factor of 10 larger (per wavelength) than a metallic waveguide counterpart if one could be made to propagate light. This factor of ten would mean that optical dielectric waveguides could still be a factor of roughly 10^4 smaller than a microwave guide. Actually, the difference with electrical is yet smaller. Microstrip and coplanar waveguide circuits are open waveguides combining metal and dielectric characteristics. The characteristic length of an electromagnetic wave is necessarily its wavelength. The characteristic length of a current, however, is the electron wavelength, which can be in the angstrom range. Current can be confined into tiny metallic strips much smaller than the electromagnetic wavelength. As the microwave wavelength is so much longer than the optical wavelength, the metal loss is not nearly so crucial as in the optical case. This is the situation in microstrip and coplanar waveguides, where characteristic stripe dimensions can be tens of microns in extent and therefore comparable to fiber outer diameters. The current, however, only serves to guide the electromagnetic wave which actually carries the signal. Although the most intense, near-field portion of this wave can be confined between conductor and ground plane, the wave clearly can sample its environment out to much greater distances. Therefore, the higher the packaging density of such open microwave channels, the worse the crosstalk. No matter how tightly one packs fiber, the crosstalk is essentially zero if the cladding is properly designed. This leads to the characteristic that fiber is an excellent medium for space division multiplexing (SDM)—that is, packaging a number of channels with different information streams in close proximity.

Although all the advantages of coherent optical communication systems have yet to be brought to fruition, another property of optical radiation has made today's optical communication systems not desirable for applications. The important property here is that of photon energy. As is seen from Table 1.1, the photon energy ranges from roughly 2 eV to roughly 4 eV. As one perhaps recalls from freshman physics, the hydrogen atom, that atom with one of the most tightly bound unpaired outer valence electrons (helium paired outer electrons will have a higher ionization potential), has an ionization energy of 1 rydberg, which is 13.6 eV. As other atomic and molecular transitions must therefore correspond to a fraction of a rydberg, this means that photons can be used for photoionization as well as pumping of atomic transitions. This leads to signal dispersion, as a finite signal linewidth (information impressed on the carrier) will cause one side of the line to lie closer to one transition than to the other, and therefore the two ends of the line will see slightly different media. As the room-temperature phonon energy is 26 meV, single optical photons are detectable with solid-state detectors. Microwave signals are not and require (inefficient by minimally 3 dB) antennas. This would seem to be an advantage in efficiency. However, there is also a penalty to be paid for having such a photon energy. Because single photons are detectable, the emission/reception process must take on a granular nature. As is well-known, even in a steady rain, the probability of a raindrop landing (as a function of time) follows a Poisson distribution, implying that there is raindrop bunching. A raindrop would rather fall right after the one before. Raindrops are impatient and don't like to wait. In much the same manner, a laser likes to spit bunches of photons even under constant bias current. Such behavior leads to a type of noise commonly referred to as shot noise or quantum noise. This is an "almost" fundamental limit[1]More discussion will be given to ways to circumvent fundamental limits in sections 6.3 and 15.3. Such circumvention, however, is not yet of practical value, although it does seem that anytime someone states a fundamental limit, it is for their own ego—to show that they are aware of "fundamental" things—and has little to actually do with science in general. on the emission/detection process, which turns out to be quite serious for analog communications although much more benign in the digital case.

As was mentioned above, the average energy of a thermal phonon is roughly Boltzmann's constant k times the temperature T, which for room temperature is roughly 1/40 of an eV. Optical quantum detectors can operate at room temperature, as single photons are measurable. Therefore, optical direct detection can be quite sensitive if shot noise-limited. Direct detection, further, is totally compatible with intensity modulation schemes—schemes in which the source is essentially just turned on and off. Such modulation schemes are the easiest to implement. When coupled with light's short wavelength which allows for miniature sources and detectors and micron-sized waveguides, direct detection schemes have allowed for small, lightweight, high-bandwidth systems which are competitive in many areas, most notably to the present telecommunications transmission, although a myriad of other applications are continually opening up. As mentioned previously, these applications have tended to open up more slowly than originally predicted, as cost was really not much of a consideration in telecommunications, where equipment costs are swamped by other considerations. With consumer electronics, one need not worry about right of way or installation. At present, the cost of connecting to personal computers (PCs) a few meters from each other optically is so expensive that fiber has not yet come to the consumer market. The high cost of the link in such a case, though, is not fundamental but more historical. Present-day developments in millimeter core plastic is an example of a much cheaper technology than, for example, glass fiber. The costs of components to go into fiber links as well as packaging costs are presently being reduced and new applications are opening up. Some of these other applications will be discussed as the presentation proceeds.

1.2 An Optical System Model

A schematic depiction of the organization of an archtypical optical system is given in Figure 1.1. The model used here, as we will see in the subsequent discussion, is quite general and applies to practically all optical systems that exist, including communications as well as sensing systems. For example, the earliest types of optical communication systems were probably ones in which a lamp was sequentially hidden from sight for fixed periods between periods of being in plain view. Such a system was employed by Paul Revere in 1776, but it is very likely that such systems had been in operation since the times of the early Greeks or Egyptians. In such a system, the information source is the lampholder, the encoder is his arms together with the occluding sheet, the receiver optics are the eyes of the beholder, the detector is the beholder's retinae, the receiver is the brain of the beholder, and the information out is the response of the beholder. One could say that such a system employs digitally encoded on-off keyed intensity modulation with direction detection (a **D**EOOK**IM**DD or DIM system, depending on the receiver design) if one were to classify this system. A similar block diagram description could apply to an Indian on a hill sending smoke signals back to his tribe, a fiber optic gas detection system, a telescope, or, as we will see as we go on, practically any optical system we can devise.

Perhaps a major point in having a model such as in Figure 1.1 is that such a model is so general that we can use it as a basis for comparing widely differing systems that at first may appear incomparable. Certainly, however, any two systems can be compared by comparing the data launched into the system with that exiting the system (bit error rate (BER) counting in digital systems or noise figure measurement in analog systems). Other figures of merit can, indeed, also allow comparison of the specific blocks of the model. Such comparison will be a major occupation of this course—to compare performance of the various

[1] †

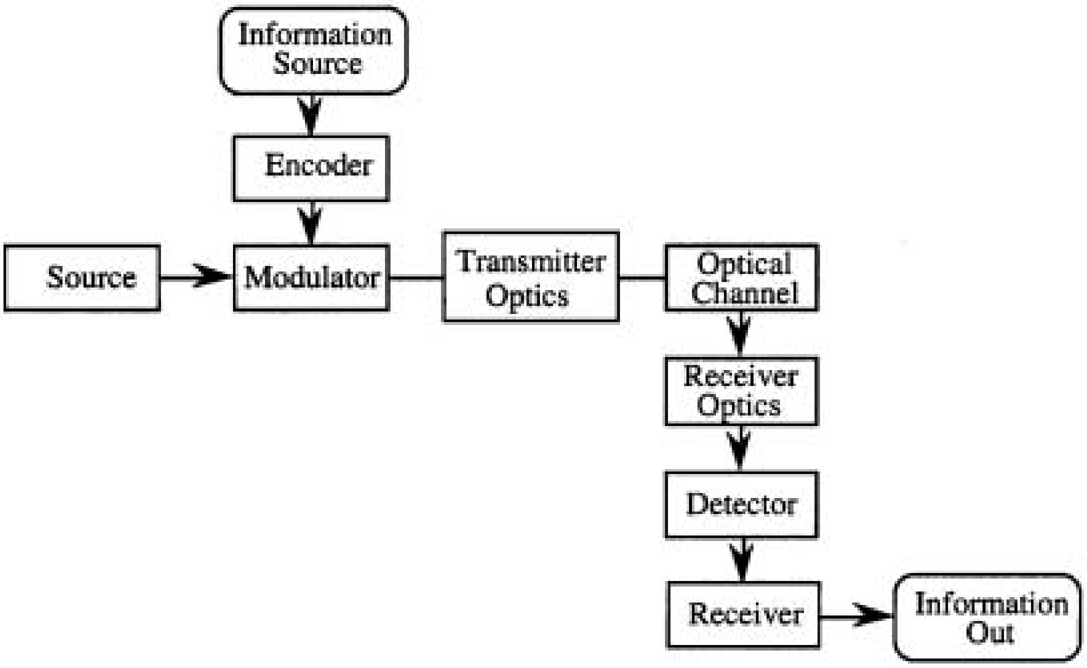

Figure 1.1: A schematic depiction of the organization of an optical communications system in which the square blocks are the optical system itself and the circular blocks denote the system input and output, respectively.

schemes for performing the various functions of the blocks of Figure 1.1. The following four subsections will discuss first sources, then fibers, and then detectors. The last section will give some systems architecture perspective, as the other types of components which make up the blocks of Figure 1.1 in the archtypical “modern-day” optical communication systems are very architecture-dependent. The discussion that follows will be given from a somewhat historical perspective.

1.3 Optical Sources

A discussion of optical sources is the only subject of numerous courses. The present discussion, in comparison with a full course, must by nature be cursory. In the following discussion, we will limit ourselves to sources and salient features of those sources which relate to “modern-day” communications systems. (This topic is the subject matter of those numerous courses as well.) A definition for a “modern-day” optical communications system by nature must be quite arbitrary. A lighthouse beacon comprises the transmitter end of an optical communication system but will not be considered here. Attention here will be placed on compact sources which can be used to transmit at least MHz information rates. Such sources are the ones of interest for telecommunications, data communications, and radio-frequency remoting and sensing. (Sensor systems have bandwidth requirements similar to those of telecommunication systems for other reasons, among them that phase dithers are necessary to place the beat note outside the $1/f$ laser noise peak (to be discussed in section 8.4 of Chapter 8), which may extend to many MHz.)

Perhaps the important milestone date in differentiating “modern day” from early times is the demonstration of the laser in 1960 (Maimon, 1960). One of the most important characteristics of the laser is its temporal coherence—that is, the fact that the laser linewidth is a tiny fraction of its center frequency. Even early lasers (HeNe for example), when operated in a single temporal mode, could exhibit relative bandwidths (that is, linewidth divided by center frequency) of 10^{-12}. In the early 1960s, no other optical source, even with strong filtering, could exhibit relative bandwidths smaller than 10^{-4} or so. But this spectral purity was

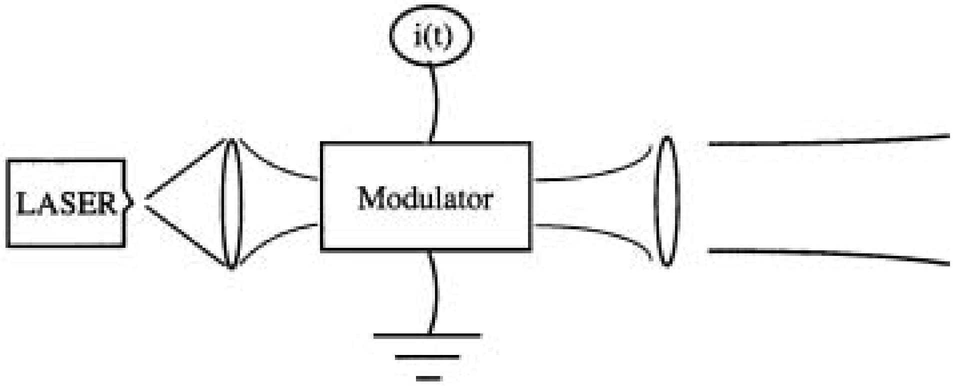

Figure 1.2: Depiction of an archtypical 1960s-style optical transmitter.

not the important laser characteristic when it came to optical communications. Much more important were spatial coherence (brightness) and, eventually, radiation efficiency.

The basic principle behind the laser is to lock the emissions of an ensemble of excited atoms in a cavity into a single direction, thereby locking the individual emissions in both frequency and phase (to be discussed further in section 8.1). That the emissions are locked in direction is an important point, as this locking allows for spatial coherence; that is, the phase across the whole wave front is constant. Such a beam can achieve the minimal beam divergence possible—that is, the divergence required by the diffraction limit. A non spatially coherent beam will diverge at many times this limit. Of course, spatial filtering can be used to eliminate rays which refuse to stay in phase with the beam's "center of gravity". However, this filtering will require an energy loss, as power must be extracted by this passive filter. Therefore, any directivity gain comes at a price. As it turns out, this price is quite high. The second law of thermodynamics tells us that we cannot decrease entropy without adding energy. As it turns out, a measure of the entropy of an electromagnetic wave is the inverse of its brightness, which is defined as the amount of power that the source can radiate from a fixed area into a given solid angle. The second law of thermodynamics, therefore, requires that passive spatial filtering cannot increase a source's brightness. Clearly, a spatially coherent source will have the greatest brightness for a given power output as it radiates into the minimum angle. An isotropic source would need to radiate roughly 10^4 times the power radiated by a laser with a 1 milliradian beam divergence in order to have an equivalent brightness. Further, as the diffraction angle is directly proportional to wavelength, a diffraction limited microwave source would need to radiate roughly 10^5 times the power of a diffraction limited optical source in order to achieve the same brightness. In communications, one generally wants to send information from one point A to a number of points B, C, D, . . . , and not to disperse the transmission throughout all space (broadcast). It is clear from the above discussion that a coherent optical source will be the most efficient source of all for achieving this goal.

Figure 1.2 depicts a 1960s-style optical transmitter. During this early era of "modern-day" optical communications, the available laser sources were either gas or solid-state lasers. These early lasers were also characteristically of low gain per unit length in their cavities and therefore needed to have high Q (quality factor) cavities. Essentially, a high Q means a highly resonant structure which is also a structure which by nature must store energy for a period of time which is long compared to an optical period. A problem with such structures is that of pumping; that is, it is hard to change the state of the fields within the cavity without waiting for many roundtrip times of the cavity. The cavity has, so to speak, a built-in memory of its past state. For this reason, output stability would require stability of the laser pump, and clearly modulation of the output light could not be achieved by pump modulation but only by external modulation. This external modulation would therefore require enough optics to collimate the laser beam, focus it through a modulating crystal, and then recollimate the beam.

Two major developments occurred in 1970 that greatly altered the situation in optics in general and in optical communications in particular. The first was the development of the laser diode which could operate at room temperature (Hayashi and Panish 1970, Alferov et al 1970). The original semiconductor laser diode was developed in 1961 (Basov et al 1961, Hall et al 1962, Nathan et al 1962) but required such high current

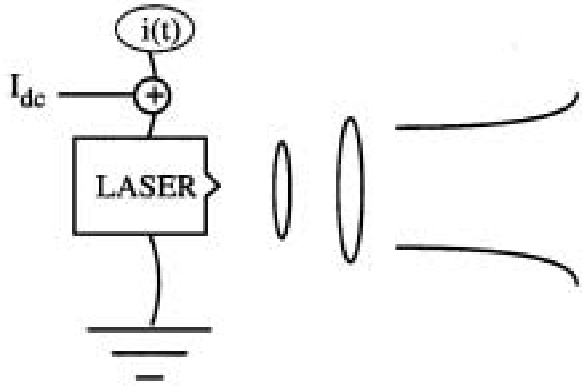

Figure 1.3: Schematic depiction of an optical transmitter which employs direct modulation of the laser bias current I_{dc} by directly summing I_{dc} in a bias T with an information current $i(t)$.

densities to achieve lasing threshold that low-temperature operation was required. (More recently, a different configuration of laser diode, the surface-emitting laser diode or vertical cavity laser diode (as opposed to the original edge-emitting laser diode) has appeared.) However, even these early laser diodes had some striking properties with respect to other lasers. As very high carrier densities are obtainable in semiconductor diode junctions, the gain per length in a laser diode can be very high with respect to, for example, a gas laser medium. For this reason, laser diodes could be made to operate in very low Q, short cavities. Now, a high Q (or high finesse) cavity requires both long length and high reflectivity cavity mirrors, whereas the laser diode cavity could be short (circa 300 μm) and have low reflectivity mirrors (actually, cleaved semiconductor facets with reflectivities on the order of a few percent). For this reason, the laser could be modulated through manipulation of the current flowing through its junction, thereby obviating the need for an external modulator. In a high Q resonator, changes in input power or current would cause a long-lasting ringing which can completely distort any impressed information. An optical transmitter which takes advantage of this direct current modulation characteristic is illustrated in Figure 1.3. It should be noted here, however, that a close cousin of the laser diode developed along with it during this period—that is, the light-emitting diode (LED). A light-emitting diode has essentially the same structure as a laser diode but has essentially zero reflectivity facets and therefore never reaches a lasing threshold. The light emission, therefore, can never become as spatially coherent in an LED as it is in a laser diode. However, due to the small size of the junction and the fact that, due to the finite depth of the device some coherent emission could occur, the emitted light is significantly more bright than that from totally incoherent sources. Indeed, the LED can be directly modulated as the laser diode. Further, the LEDs are extremely inexpensive to manufacture. For these reasons, the LED showed up as a source in many transmitter modules such as the one depicted in Figure 1.3, when system requirements were not so stringent as to require a laser source. A disadvantage, however, of the LED is wall plug efficiency. A laser diode can be close to 100% efficient. As prices drop for laser diodes of all types, they tend to replace LEDs due to the better collimation and wall plug efficiency.

1.4 Optical Paths

The second major development to occur during 1970 was the development of the low loss optical fiber (Kapron, Keck, and Maurer 1970). The idea of communicating through free space is well and good, but as was previously stated, a major advantage of optical transmitters was their directivity. For broadcasting, it is hard to beat radio waves. For point to point, it is hard to beat optics. But there are two problems with using free space as a transmission medium. The first one is very practical. Usually, one is trying to communicate with someone else on the face of the Earth. (Satellite-to-satellite communications will be briefly discussed in some following chapters, in particular 12.1 of Chapter 12 on digital communications, where pulse position modulation (PPM) is mentioned.) Two major problems arise from this. One problem is that the Earth is round. Once line-of-sight distances are exceeded, one needs to increase link length by a quantum leap in

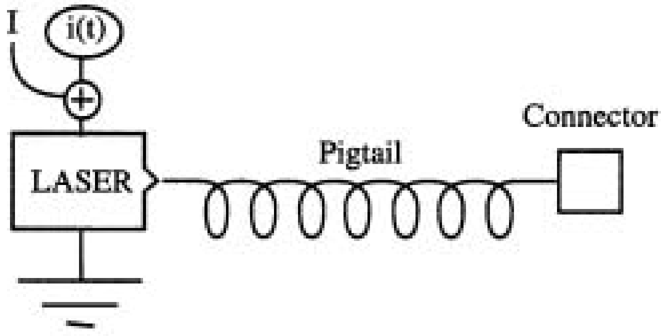

Figure 1.4: Schematic depiction of an optical transmitter as used in the telephone network.

order to include a satellite repeater. This is a problem for both microwave and optical transmission. As was shown by Hertz (Hertz 1983), although the first definitive trans Atlantic demonstration was made by Marconi some years later around the turn of the century, low-frequency waves (AM band, $\leq$ 1 MHz) will cling to the ground for some distance. Already by the shortwave band ($\sim$ 10 MHz), the waves begin to skip off the Earth, although up to roughly 100 MHz they still reflect off the ionosphere. At higher frequencies, one needs an orbiting reflector and/or repeater. Another problem with free-space optical transmission is that, unfortunately, there is a thing surrounding the Earth called the atmosphere. Radio waves don't care too much what is happening in the atmosphere, but optical waves do. Rain, snow, fog, and even wind affect optical transmission. There are free-space optical links still in use, especially between buildings in cities and on campuses, but for any but the shortest, most protected of these links, weather interference does occur. Many sensors, though, by nature, use free space and always will as they are measuring the weather. An example is the LIDAR, or laser radar. (Attention will turn to communication system models of some simple sensor systems in Chapter 14.) A third problem with free-space communication is more fundamental. That problem is diffraction. Coherent waves in free space will expand at an angle that is roughly equal to the wavelength of the radiation divided by the effective radiating aperture. One can minimize the diffraction effect only by using larger and larger focusing lenses. In fact, one can project a 600 m spot on the moon, but this requires using a 2.7 m telescope as the transmitter. The diffraction effect, therefore, puts fundamental bounds on distances and powers necessary in free-space systems. The optical fiber, however, is a solution to both of the above-mentioned problems, at least in commercial telecommunications systems and in some sensor systems as well.

A "modern-day" archetypical telecommunications optical transmitter, such as the one employed in today's telephone network, is depicted in Figure 1.4. The idea is that the laser diode can be "pigtailed" with an optical fiber, therefore obviating the need for any focusing optics whatsoever. The transmitter module therefore needs no alignment. One need only hook up the laser to a current source and hook up the fiber output into a transmission fiber by means of an optical connector.

1.5 Optical Detectors

As was discussed previously in Section 1.3, the photon energy is both a boon and a hindrance to optical communications. That the photon energy is much larger than the room temperature phonon energy, kT, is advantageous, as that means that techniques can be found which are extremely quantum efficient. This, when coupled with the fact that the photon energies correspond closely to various atomic transitions and semiconductor bandgaps, allows for a plethora of different materials to serve as detector elements, each being capable of detecting even single photons. At the same time, however, this high photon energy yields a limitation. As photons are energetic, in order to receive a small number (low-power signal in a short interval), one must put up with a signal granularity; that is, a major source of noise will be "counting" noise. The high energy per photon means also that there is an upper limit on how many photons a detector can receive before it is "fried" from residual thermal energy. This form of counting noise, often referred to as shot noise or quantum noise, forms the technical noise floor in optical communications. Although techniques

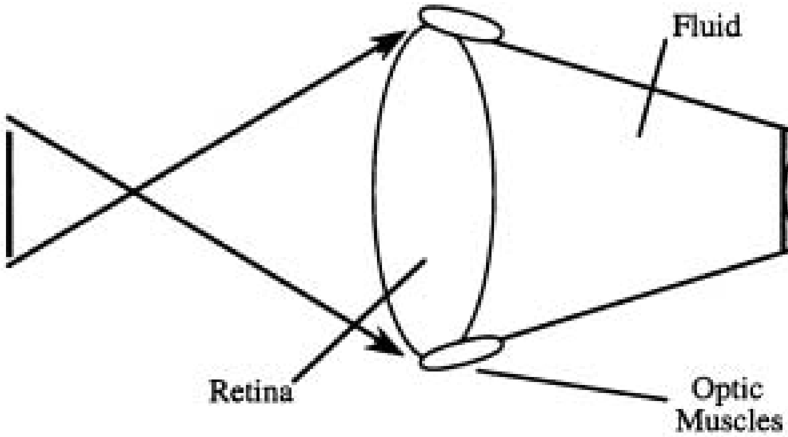

Figure 1.5: Schematic depiction of the workings of the eye.

to circumvent this floor exist (as will be touched upon in sections 6.3 and 15.3), they are not yet viable as practical signal control and processing techniques.

The earliest known and one of the best light detectors is the eye. The workings of the eye are illustrated in Figure 1.5. The retina is a lens whose focal length is controlled by optic muscles. This telephoto lens images an object onto a photosensitive plane called a fovea. Image inversion is performed in the brain. The photosensitive plane of the eye is quite reminiscent of another, less efficacious, but presently quite popular light detector, that of the photosensitive plate or film. The basic workings of light-sensitive film are given in Figure 1.6. In the exposure phase, incident photons break bonds (often silver halide bonds). In the development stage, the film is immersed in a developing bath which contains chemicals which react with the broken bonds to change the nature of one of the bond holders, generally changing a bonded molecule with a dielectric character into one dielectric and one metallic piece. These metal clumps then act as absorption centers, so when reilluminated they form a negative image of what was incident on the film during the exposure phase. Photographic film is very useful for storing large amounts of data; however, development tends to be time consuming, the process tends to not be photon efficient, and it is hard to get very good resolution. It seems at present that electronic imaging arrays reading out into CD ROMs may take over as the recording elements of the next generation of photographic instruments. We will not concern ourselves further with imaging technology, however, in what follows.

It was Einstein in 1905 [3] who first explained an effect known as the photoelectric effect earlier observed by Phillip Lenard. Lenard shone light on a metal electrode in a vacuum tube and noted that the current was proportional to light intensity if and only if the light were of certain colors. The idea is as illustrated in Figure 1.7. If the photon energy is sufficiently high (exceeding the work function of the metal), an electron is ionized from the anode, and if the electric field between the two electrodes is sufficiently high, this electron is accelerated toward the cathode, where it causes a current flow in the external circuit. Such vacuum tube realizations of optical detectors were the most common ones until the 1960s, when solid-state semiconductor detectors became of sufficiently high quality to supplant vacuum tube technology.

A basic circuit for use with a semiconductor p-i-n detector is depicted in Figure 1.8. A basic p-n diode structure, which would develop a depletion region under reverse bias, is modified slightly to include an intrinsic (undoped) region between the p and n regions to enlarge the depletion region sufficiently that it becomes relatively insensitive to reverse bias level and therefore, for example, will not change dimension under moderate illumination. Illumination of the intrinsic region will ionize valence electrons, thereby creating a number of free electrons in the conduction band and holes in the valance band, where the electrons and holes will be swept out of the junction region by the strong fields there. This sweeping out of the electrons and holes will lead to current flow in the external circuit and will thereby temporarily lower the device bias. These temporary lowerings are then picked up as an ac variation on the load resistor R_{L}. The p-i-n structure can be a high-speed structure, but, as with all high-speed structures, it must also be low-power. But it is just to such high-speed, low-power structures that we wish to limit our attention. The PIN will be discussed

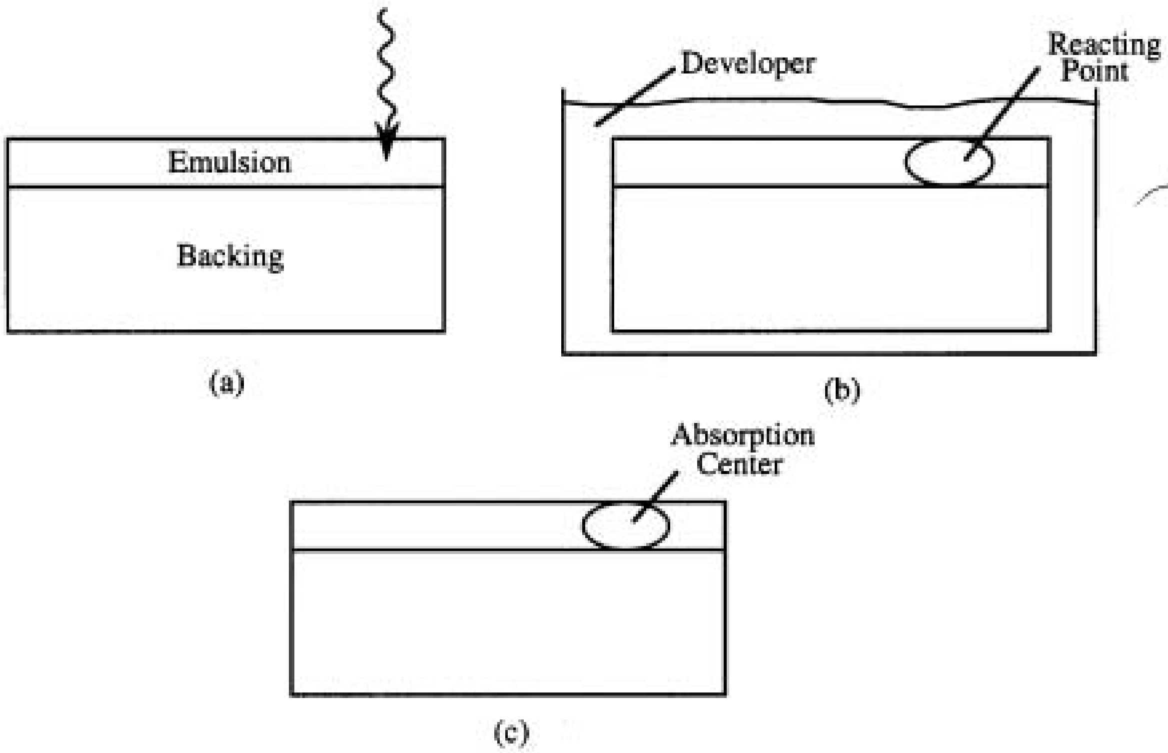

Figure 1.6: Schematic depiction of the working of a photographic film: (a) exposure, (b) development, and (c) the resulting film.

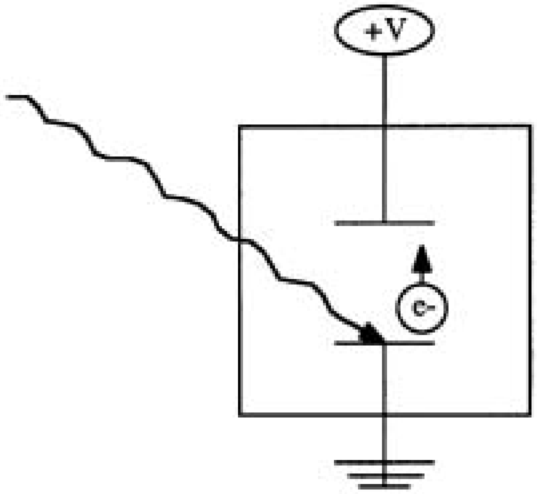

Figure 1.7: Schematic depiction of a vacuum tube diode photodetector.

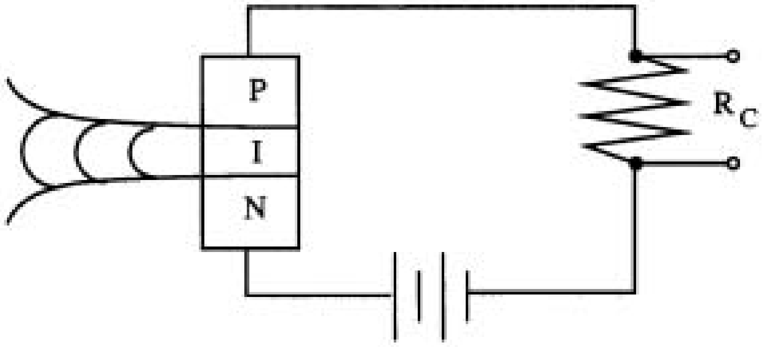

Figure 1.8: Schematic depiction of a p-i-n detector and its associated bias circuit.

further in section 4.6 and will be used as an archetypical square law detector in Chapter 7 and much of the presentation following Chapter 7.

The other detector (in addition to the p-i-n) in common use in optical communication systems is the avalanche photodiode (APD). The APD has an operating principle similar to that of the p-i-n, but it has a much longer propagation path for the electron. By applying a high voltage along this path, the electron impact ionizes a multiplicity of additional carriers, leading to an effective propagation gain. A drawback with APDs was the need for kilovolt voltage supplies, although the situation is improving. There is always a noise penalty, as will be discussed in section 11.4.2.

1.6 Some System Considerations

The demonstration of both the low-loss optical fiber and the room-temperature semiconductor laser was timely, as the need for a new transmission medium for the telephone network was already clear. The original telephone network appeared much as the one depicted in Figure 1.9. (See, for example, Talley 1979 or, for a more fiber optic perspective, Personick 1993.) Central offices, from which the local (subscriber) loops which extended to customers originated, were hooked together by trunk lines of up to a few kilometers in length. These local branch offices would all be linked to a toll office, which might be citywide in smaller cities. These toll offices were the gateway to the long distance network and, through a system of protocols, could hook a subscriber up with the next town, next state, or foreign country. These long distance offices would be linked to each other by much longer interconnects which in this early era were generally radio links. Already by roughly 1960, network planners began to realize that the existing trunk rights of way would eventually be filled with cable, and further expansion of the telephone network would be stopped, as with copper cable, to the lowest order, the aggregate information rate was only a function of the total copper cross-section. During the 1960s, however, digital electronics became cheap and began to become advanced. Switching functions which were originally performed manually or mechanically could now be performed by electronics much more compactly and at much higher rates. Time-division multiplexing (TDM) became a realistic technique for minimizing the horrendous number of cables that linked local branch offices together. However, the higher rates did not help as long as copper was the medium.

The first pre-fiber effort at achieving large-scale TDM was the one in which a standard DS1 transmission rate, which consisted of sixteen interleaved telephone conversations (composite rate of roughly 1.54 Mbps), was dictated to be the trunk rate. That is, local offices would still talk to the local loop on multiple single conversation cables, but interlocal office communications would always be multiplexed. The reasoning for this above all else was to minimize the amount of "stuff" in the transmission ducts, where capacity was rapidly being exceeded. The original standard transmission scheme to perform this function was called T1[2]The nomenclature T1, although originally coined to refer only to one specific means to transmit a DS1 signal, has taken on a life of its own and is now often used synonymously with DS1. and consisted of a cable transmission line with a 2 km repeater spacing. The problem with this T1 scheme was that it was not expandable as was hoped. That is to say, if one were to raise the rate (the number of channels to be TDM'd),

[2]†

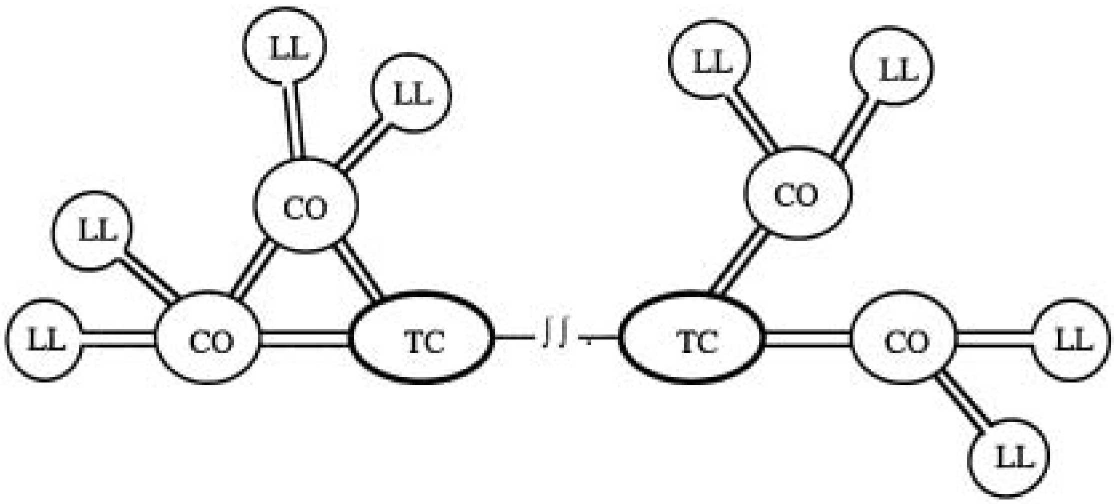

Figure 1.9: Schematic depiction of a highly simplified telephone network showing the typical path from a subscriber (in the local loop) through central offices (for local calls) or to a toll center to enter a new hierarchy in the long lines system. LL is local loop, CO is central office, and TC is toll center.

the copper cable transmission medium, due to signal dispersion, loss, and crosstalk, would require a reduced repeater spacing. Changing equipment in the local office is one thing, but going into the ducts is expensive. Further, if expansion required ever more repeaters, the ducts would eventually fill with repeaters. This situation was threatening to shut down the growth of the national telephone network as of the late 1960s. Although the T1 scheme was found to be (only theoretically at that time, but possible today) extendable to DS1C (=2 DS1) but not beyond, the T1 scheme was considered not sufficiently expandable, and the search for alternative transmission media became frenzied. In the last few years of the 1960s before the low-loss fiber was demonstrated, a myriad of technical systems were researched including a complex millimeter wave waveguide technology and, I understand, even waveguides made from ice.

Originally it was multimode fiber that offered the solution. These original fibers had 50-μm cores and 125-μm outer cladding diameter and had numerical apertures of 0.2. Dispersion was such that 10's to 100's of MHz could be transmitted over the 2 km repeater spacing, and the loss at the original 0.83 μm operating wavelength of roughly 3 dB/km, even including laser and detector coupling, posed no problem at all for lasers capable of 100 μW output. In fact, at the 100 μW level, the link operation could even be shot noise-limited. Multimode fiber provided a greatly expandable T1 solution which could later be expanded from DS1 to DS3 (=28 DS1 =45 Mbps) and upward in data rate. The first field tests of multimode trunk links were installed in 1975 and were a great success, and by 1980 this technology was the chosen one.

The original predictions were that fiber technology would rapidly extend into the local loop. This has yet to happen, in part due to a regulatory morass and in part due to plain economics. The fiber actually took off in the other direction—toward the long haul. In many senses, single-mode fiber is a simpler technology than multimode, despite the more exacting tolerances. If one can simply hold the fiber radius and numerical aperture product to less than some constant in order to achieve single-mode operation, the fiber automatically will have a thousand times less dispersion than one with even two modes. Further, by moving the operating wavelength further into the infrared at 1.3 μm or 1.55 μm, one could decrease the fiber loss by a factor of 10 in dB. The enabling technologies, single-mode fiber production, and single-mode 1.3 μm laser production were in place by 1980, and soon repeater spacings of 10 km at rates of 565 Mbps were realized. These original single-mode fibers had roughly 10-μm cores, 125-μm outer cladding diameter, and numerical apertures of 0.1. The Institute of Electrical and Electronic Engineers (IEEE) synchronous optical network (SONET) standard was put in place around this time, a standard in which universal multiplexing standards had rates ever increasing in factors of 4 (565 Mbps, 2.5 Gbps, to 10 Gbps, ...). The standard has been quite useful in allowing electronics manufacturers to compete on receiver and transmitter electronics, thereby greatly lowering electronics prices. Rates and repeater spacings have ever since been steadily increasing. Twenty kilometer repeater spacing and 2.5 Gbps rates are not at all uncommon on long-haul links, and there are

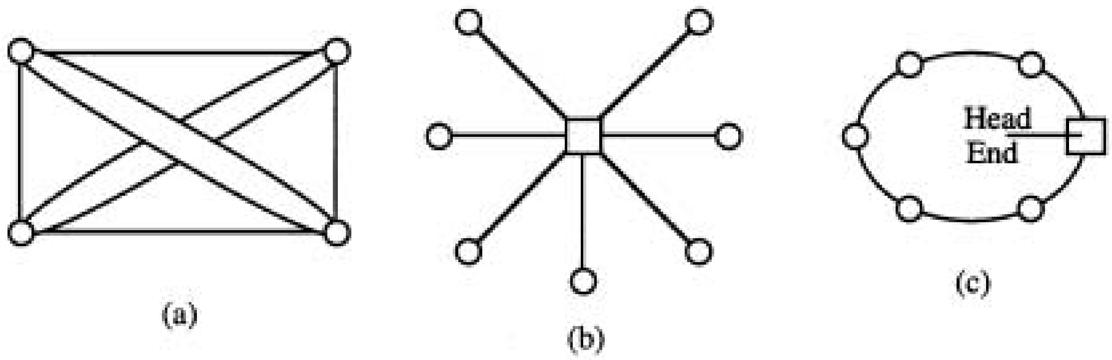

Figure 1.10: Basic interconnection topologies applicable to, for example, trunk lines between local offices: (a) a complete interconnection; (b) a star; (c) a ring.

new transoceanic fiber cables as well. There are some 10-Gbps links already in place, and work on 40-Gbps links is progressing. A problem with these rates of greater than 2.5 Gbps in the present-day market has become cost. With the breakup of telecommunications monopolies essentially worldwide, the companies no longer have the capital necessary to support internal electronics foundries. As the number of components necessary in the telecommunications network is small relative to the commercial electronics market, it is hard to convince electronic components manufacturers to set up production lines and compete on components which have no market yet outside the small telecommunications market.

Our communications needs, however, extend beyond Plain Old Telephone Service (POTS). These days, cable TV as well as two-way (MODEM) computer interconnections are becoming more and more common. Digital television (DTV) transmission, if uncompressed, requires roughly 140 Mbps for a single channel. High-definition television (HDTV), if uncompressed, requires four times the DTV rate or roughly 560 Mbps. Many such channels could be carried by telephone long lines but barely, if at all, by trunks and in no way by the local loop, which has so far been little impacted by the fiber communications revolution. Cable companies provide TV to the home through use of the standard 5 MHz analog transmission. More and more, the cable companies are achieving reliability and fidelity through the use of fiber trunks and only short twisted-pair feeders. MODEM computer connections are notoriously slow. Businesses which have a great enough need can buy bandwidth through setting up their own private branch office (PBX) and hook into the long lines system and provide their own local distribution. Broadband services such as the graphic interfaces to the internet, however, could find a much greater usage could they be brought closer or even directly into the home or small business. Although the technology is there to bring multimedia services to small businesses and homes, the costs and regulations are hard to surmount.

A short discussion of interconnection topologies seems appropriate at present. The basic fully interconnected star and ring configurations have already been matter of factly illustrated in Figure 1.9, but they are individually broken out for study in Figure 1.10. For a configuration where there are N nodes, there are various manners in which they can be interconnected. A maximal interconnect would be one in which each node was interconnected with each other node, requiring each node to have $(N-1)$ interconnects requiring $N(N-1)/2$ interconnection wires, which quadratically becomes a large number of interconnection wires as the number of nodes increases. A star interconnect, as illustrated in Figure 1.10(b), requires only $N-1$ interconnects total but requires a "smart" head end through which all messages must pass for routing and/or broadcast. This function is distributed in the maximal interconnect where each node could serve as a server or router. In the ring (or its duplex version known as a bus), as illustrated in Figure 1.10(c), there are again only $(N-1)$ interconnects, but here the processing can be either localized (in the head end) or distributed through the ring. However, in the ring, everybody's data must pass through each node, whereas in the other two configurations only the data necessary to reach the receiver need pass through any interconnection, ostensibly requiring much less bandwidth.

Any communications system of any great extent is probably going to require a combination of topologies, as was indicated in the discussions surrounding Figure 1.9, therefore becoming some kind of a tree with round leaves perhaps located in a forest of some connectivity. The structure will probably also have built-in redundancy to allow reprogramming to obtain differing virtual rings and busses depending on traffic or down equipment. According to another IEEE set of standards, no matter what equipment is attached to a node, each node is comprised of seven standard layers between the physical layer (network) and upper layer (logical attachment to equipment). This standard has again allowed for reasonably complex, reliable, cheap interconnection chip sets, allowing for a considerable degree of smartness to be built into a given node. Even though the physical layer satisfies an IEEE SONET (synchronous transmission) standard, the logical intervening layers can support asynchronous transmission; that is, the logical network due to traffic constraints or other reasons could decide to route different packets via different routes to different locations. The completely interconnected bar network of Figure 1.10(a) as well as the star network of 1.10(b) can be what are called circuit switched networks. That is, both logically and physically, either network can provide a hard interconnect exists between sender and receiver. The bus network of Figure 1.10(c) must almost necessarily send packets in order to identify who should receive the message. It is not truly packet-switched, though, as there is only one way around the bus, and generally one tries to design a bus so as to remove packets after a single traversal of the bus. A network such as that of Figure 1.10(a), however, could be packet switched. That is, if a node were busy, the packet could go elsewhere and come back. That is, if a receiver were busy, he could send a packet back into the network to receive it later. In fact, by logical operations alone, the network of Figure 1.10(a) could reconfigure itself into either Figure 1.10(b) or Figure 1.10(c), depending on what it wanted to do. The packet switching concept, though, becomes more powerful with system complexity. It is packet switching which makes the World Wide Web possible.

A major point of the above discussion is to point out where fiber optics per se may serve to really change telecommunications. To a great degree, at present, fibers have been used to replace wires or radio links one-for-one without significantly affecting network function. Now a change is in the wind. As was pointed out earlier in this introductory chapter, optics is a good technology for space, time, and wavelength division multiplexing (SDM, TDM, and WDM), making it maximally flexible. In the past, a major reason for using stars for both local area networks (LANs) as well as distribution from the local loop was that the star minimized the number of interconnects of low bandwidth. With fiber, one needs to minimize neither the bandwidth nor the number of interconnects. Further, with smart nodes, one can use sophisticated routing schemes as long as the node has enough bandwidth to match that carried by the fiber. For these reasons, the coming generations of trunk networks will be very high-speed (2.5 Gbps to 10 Gbps to 40 Gbps) local loops which should allow broadband services to come a step closer to the home, if not yet to the local loop and home. A review article by Personick (Personick 1993) points to some directions the telephone system will take, including some discussion of bringing broadband to the trunks. It should be noted here that sensor systems can also profit from combinations of SDM, TDM, and WDM, and this is a prime reason for the steady growth in optical sensing in general and fiber and integrated optic sensing in particular. We will give some further discussion to sensors in Chapter 14.

In the chapters of Part III of this text, the basic principles of optical communications systems will be elucidated, hopefully in such a manner that the reader upon completion of the material will be able to analyze and design future optical systems. At the least, he or she should be able to study the various applications discussed in Part IV of the text and be able to return to the introduction and read through it with better understanding than during the first run-through.

Bibliography

[1] Zh. I. Alferov, M. V. Andreev, E. L. Portnoi, and M. K. Trukan, “AlAs-GaAs Heterojunction Injection Lasers with a Low Room-Temperature Threshold,” *Sov Phys Semiconductors* **3**, pp. 1107–1110 (1970).

[2] N. G. Basov, O. N. Krokhin, and Yu. M. Popov, “Production of Negative Temperature States in P-N Junctions of Degenerate Semiconductors,” *J E T P* **40**, pg. 1320 (1961).

[3] A. Einstein, “Über einen die Erzeugung und Verwandlung des Lichtes Betreffenden Heuristischen Gesightspunkt,” *Annalen der Physik* **17**, series 4, pp. 132–148 (1905).

[4] R. N. Hall, G. E. Fenner, J. O. Kingsley, T. J. Soltys, and R. O. Carlson, “Coherent Light Emission from GaAs Junctions,” *Phys Rev Lett* **9**, pg. 366 (1962).

[5] I. Hayashi and M. B. Panish, “GaAs-Ga_xAl_{1-x}As Heterostructure Injection Lasers Which Exhibit Low Thresholds at Room Temperature,” *J Appl Phys* **41**, pp. 150–163 (1970).

[6] H. Hertz, *Electric Waves* (MacMillan, 1983). This is a reprint of an 1887 original of Hertz.

[7] F. P. Kapron, D. B. Keck, and R. D. Maurer, “Radiation Losses in Glass Optical Waveguides,” *Appl Phys Lett* **17**, pp. 423–425 (1970).

[8] T. H. Maimon, “Optical and Microwave-Optical Experiments in Ruby,” *Phys Rev Lett* **4**, pp. 564–565 (1960).

[9] M. I. Nathan, W. P. Dunke, G. Burns, F. H. Dills, and G. Lasher, “Stimulated Emission of Radiation from GaAs p-n Junctions,” *Appl Phys Lett* **1**, pg. 62 (1962).

[10] Stewart D. Personick, “Towards Global Information Networking,” *Proc IEEE,* **81**, pp. 1549–1557 (November 1993).

[11] David Talley, *Basic Telephone Switching Systems,* Second Edition (Hasbroughk Heights, NJ: Hayden, 1979).