Cambridge University Press
978-1-108-42756-2 — Lightwave Communications
George C. Papen , Richard E. Blahut
Excerpt
More Information

1 Introduction

A vast network of optical fiber continues to expand beneath the surface of our planet. This fiber network forms the communications infrastructure upon which sit the omnipresent mobile devices that now bind society in new ways, never seen before. At the heart of this network is the seemingly simple and passive optical fiber made of glass or plastic that is about the width of a human hair yet over hundreds of kilometers long when used in wide-area networks. This revolutionary component, though largely passive, will require the length of this book to describe how it is used to convey information. This introductory chapter presents an overview of digital guided-lightwave communication systems, emphasizing the differences between lightwave and other types of communication systems.

Communication systems convey information (voices, images, video, or data) from a source to a destination. For the purpose of transmission, modern communication systems first map information into electronic signals that can be either analog or digital. Analog systems map information into a continuous physical quantity. A digital system maps or encodes information into a sequence of discrete logical symbols or letters. If the original information source is analog (such as voice), it can be transmitted digitally by first digitizing the continuous waveform into a sequence of digital symbols by sampling and quantizing[1] the continuous waveform to produce a sequence of sample values. The combination of sampling and quantization produces a sequence of digital values.

The set of logical values for each transmitted symbol is called the *channel input alphabet*, with each discrete value being a letter from that alphabet. The most common digital symbol has two possible letter values and is called a *bit*. One letter is called *one* or *high* or *mark*, and the other letter is called *zero* or *low* or *space*. Symbols in other alphabets can have more than two letter values and can be represented by a group of bits. As an example, the keyboard character "$" is commonly mapped into an eight-bit symbol. This keyboard character could be sent as eight separate two-state letters (bits) or it could be sent as a single letter drawn from an alphabet of 256 letters.

Every point-to-point digital communication system conveys data between a source and a destination. Starting with the source, this information is typically handled at multiple conceptual layers of functionality before being transmitted over a communication channel. The most basic of these hierarchical communication layers is known as the *physical layer* or the *modulation layer*, with each higher layer providing additional

[1] The word *quantization* is also used to describe the discrete nature of lightwave signals.

communication functions, such as controlling traffic. The goal of modern physical-layer digital communication system engineering, which is the subject of this book, is to provide the physical layer with specific attributes such as information rate, reliability, and security, to achieve energy-efficient use of the communication channel, and to make the physical layer invisible to the user.

All aspects of the communication system that relate to the point-to-point transmission of a sequence of symbols from a source to a destination are regarded as a part of the physical layer. The differences between lightwave and other types of communication systems are a result of the differing physical-layer transmission mechanisms for the generation, propagation, amplification, and detection of lightwaves.

1.1 Digital Lightwave Communication Systems

A general point-to-point guided lightwave communication system consists of a transmitter, a guided lightwave channel, and a receiver. Each of these system elements is shown in Figure 1.1.

1.1.1 Channel Coding

The input to the transmitter is a sequence of digital symbols called a *dataword*. The dataword is denoted by a sequence $\mathbf{d}$ of length k. To ensure reliable communication, modern digital systems use an *encoder* to modify the dataword to produce a longer sequence of logical symbols $\mathbf{s}$ of length n, called a *codeword*. When the transmitter and receiver use the same alphabet, the codeword blocklength is always larger than the dataword blocklength, with the ratio of the blocklengths known as the *code rate*.

The replacement of the original dataword by a codeword with more symbols can be regarded as a controlled form of memory because the symbols of the codeword depend on the dataword. This deliberate form of memory creates redundancies, with

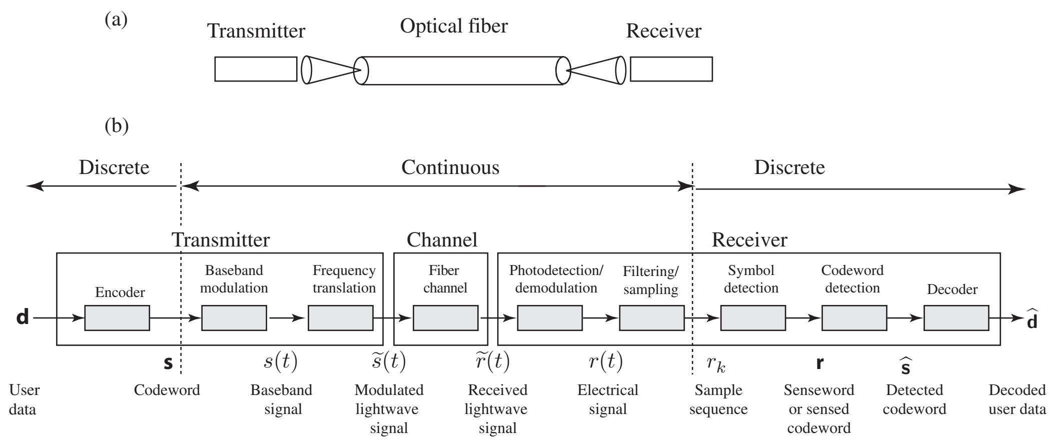

Figure 1.1 (a) Simplified lightwave communication system. (b) The main functional blocks in a digital lightwave communication system.

the information in the dataword "spread out" over the length of the codeword. These redundancies are used to control or correct symbol errors. The process of creating this form of controlled memory is called *channel coding*. The transmitted symbols of a long message are represented by a sequence $\{\mathbf{s}_j\}$ of such codewords.

1.1.2 Modulation

The second function implemented by the transmitter, which may also involve aspects of coding, is the mapping, in turn, of each codeword $\mathbf{s}$, one or several symbols at a time, into a continuous physical quantity $s(t)$ comprising the *baseband signal waveform*. This process is called *modulation*. The baseband signal waveform consists of a superposition of *symbol pulses*. The amplitude of each symbol pulse depends on the present symbol and perhaps on previous symbols. For a lightwave communication system, the baseband signal is transformed, or modulated, using *frequency translation*, into another waveform $\widetilde{s}(t)$ called a *passband signal waveform*. In general, the term "modulation" will be used interchangeably for the process of creating a baseband waveform from a sequence of pulses, or for the process of creating a passband waveform from a baseband waveform by frequency translation, or both of these operations together.

In many modern communication systems, the separation between modulation and coding need not be clear-cut, and we refer to the combination of the two broadly as *coded-modulation*. The coding of the sequence is used to control the dependences of transmitted symbols across the available independent degrees of signaling freedom. These degrees of freedom may include frequency, time, polarization, space, and the number of characteristic spatial patterns or *modes* that the guiding structure supports. Understanding the useable number of independent degrees of freedom is essential to designing an efficient communication system.

Intentional interdependences generated by a code are used to control transmission errors or to manage the spectrum of the transmitted waveform. The modulation process controls the form of the waveform that represents the data symbols. These two aspects are combined to generate specific sequences of modulated symbols that lead to a small probability of a demodulation error. Various versions of this interdependent coded-modulation process are now in common use in modern communication systems.

1.1.3 Types of Lightwave Channels

One or more modulated passband waveforms $\widetilde{s}(t)$ are transmitted over a lightwave channel. In this book, only a guided-lightwave channel based on an optical fiber is studied. Other lightwave channels such as free space or water have different propagation properties.[2] Each waveform in a fiber channel may use a separate guiding structure within the same fiber, a separate mode within a guiding structure, a separate wavelength λ, or a separate polarization mode within a spatial mode.

[2] See Karp, Gagliardi, Moran, and Stotts (2013).

Guided lightwave communication channels are commonly classified by information rate and transmission distance, called *reach*. For low to moderate data rates and short distances, systems often use multimode optical fibers and do not use lightwave amplification. For high data rates and long distances, the lightwave channel is a span of multiple connected segments of an optical fiber that conventionally supports only one spatial mode. Within each segment, the lightwave signal may be amplified to compensate for signal attenuation. This amplification process introduces noise.

In a practical lightwave channel, signal attenuation and other signal impairments affect the performance of the communication system. These impairments adversely redistribute the signal energy and distort the signal waveform as the signal propagates within the fiber. This redistribution or interference creates unintentional dependences between the transmitted symbol intervals, between transmitted polarization components, or between multiple datastreams within the same physical channel. Depending on the lightwave signal power $P(t)$, the physical mechanisms that cause distortion may be linear, nonlinear, or a combination of both.

The most significant linear distortion mechanism is linear dispersion that arises because different frequency components or signaling modes have different propagation velocities. The most significant nonlinear distortion mechanism is an intensity-dependent modification of the propagation characteristics of the lightwave channel, which redistributes the signal energy both within a single datastream and among multiple datastreams within the same physical channel.

Signal propagation in a linear dispersive optical fiber is presented in Chapters 3 and 4. Signal propagation in a nonlinear dispersive optical fiber is presented in Chapter 5. Chapter 8 extends these topics to multiple datastreams for which there may be interchannel interference between the datastreams. The mitigation or accommodation of interference is discussed in Chapter 11, while the estimation of signal and channel parameters is considered in Chapter 12.

1.1.4 Demodulation

The received passband lightwave waveform $\widetilde{r}(t)$ consists of a combination of a distorted replica $\widetilde{s}_{\text{out}}(t)$ of the input passband signal waveform and lightwave noise. Any part of the noise that does not depend on the signal is called *signal-independent noise*. Any part of the noise that depends on the signal is called *signal-dependent noise*. The origin and characterization of noise are discussed in Chapter 6.

The received modulated lightwave signal $\widetilde{r}(t)$ is converted into an electrical signal using a *photodetector*, and is then *demodulated*. For some lightwave communication systems, the combination of the conversion between the lightwave signal and the electrical signal along with the demodulation process produces a baseband electrical signal $r(t)$ that is proportional to the received lightwave signal. For other lightwave communication systems, the conversion between the lightwave signal and the electrical signal has a square-law characteristic that generates an electrical baseband signal $r(t)$ that is proportional to the received lightwave signal power $P(t)$. In such a system, the photodetected electrical signal is proportional to the squared magnitude of the received lightwave signal.

1.1.5 Detection and Decoding

After the lightwave signal has been converted into an electrical signal, a *detection* and *decoding* process is used to determine the most likely sequence of symbols that comprises the transmitted codeword. The first part of this detection process consists of a signal transformation followed by sampling. The signal transformation on the received electrical waveform is typically implemented by a linear filter called a *detection filter*. This filter generates a filtered electrical waveform, also denoted $r(t)$, which is used to generate a sequence $\{r_k\}$ of samples. Alternative signal transformations based on a nonlinear threshold operation are appropriate for the detection and counting of individual photons.

The sequence of sample values is used in the detection process to form the *senseword*, or the sensed noisy codeword, which is then sent to the decoder to determine the most likely transmitted codeword. One method of forming the codeword assigns a logical value to each symbol separately. The detection of the corresponding senseword of logical values is called *hard-decision demodulation*. Alternatively, the senseword can be formed using a quantized value of each sample of the received waveform. The quantized value is a discrete approximation of the analog value of that sample. This is called *soft-decision demodulation*. For either detection process, the estimate $\widehat{\mathsf{s}}$ of the most likely transmitted codeword is then *decoded* or *detected* to recover the codeword or to reconstruct the corresponding dataword. The combination of the frequency translation and the recovery of the user dataword defines the demodulator and the decoder. The combination of the modulator and the demodulator, including the encoder and the decoder, is called the *modem*. Each of these functions will be fully explained at an appropriate place in the book.

When the total distance, or reach, of a span requires the use of multiple segments – each with a lightwave amplifier – one can choose to *amplify-and-forward*, *remodulate-and-forward*, or *recode-and-forward*. These choices differ in performance and cost. The first method simply amplifies both signal and noise. The second method demodulates, then remodulates each symbol to remove the noise, but sometimes makes symbol errors. The third method, called *regeneration*, decodes then recodes each block or frame of data to remove demodulation errors. Regeneration differs from the first two methods because it involves all of the receiver functions and all of the transmitter functions described above, and so is more complex. At a regenerator, the decoded dataword is used as the input to another transmitter, producing a new transmitted signal that is intended to be identical to the original transmitted signal at the source. In this way, noise and distortion are removed at each regenerator, though (rarely) regeneration may destroy an entire block of data, but only if it is already too distorted to recover.

1.1.6 Error Probabilities

The ability of a communication system to reliably convey information, which depends on the method of detection, is measured by the appropriate error probability. If the detection is done on a symbol-by-symbol basis, as is often the case for common lightwave systems, then the *probability of symbol error*, denoted p_e, is used to quantify the performance. If the symbols are binary symbols, then the performance is expressed in terms of the probability of a bit error, usually called the *bit error rate*.

The bit error rate is a fundamental physical-layer performance figure-of-merit for a binary lightwave communication system that uses symbol-by-symbol detection. If a more complex detection technique that processes a sequence of symbols in a codeword as a whole is used, or if the transmission medium has random propagation characteristics, then the errors tend to cluster and the bit error rate may be less useful. In this case, the term *codeword error rate*, *block error rate*, *frame error rate*, or *message error rate* may be more appropriate. Minimizing the probability of a detection error, either for a single bit or for a whole message, subject to a set of system constraints, such as the information rate or the transmitted power, is the fundamental design goal at the physical layer.

1.2 Lightwave Signal Models

The relevant properties of the lightwave signal are expressed using a *signal model*. For some systems, the lightwave signal is adequately modeled using continuous quantities based on geometrical rays or electromagnetic waves. For other systems, particularly at low signal levels, the discrete, quantum nature of a lightwave signal is evident. This unique dual nature of a lightwave signal has led to the development of several signal models, with each signal model being useful in appropriate circumstances. These signal models, which will be discussed in detail throughout the book, can be grouped into *classical optics*, *photon optics*, and *quantum optics*.

Quantum optics is the most complete signal model. Both classical optics and photon optics are limiting forms of quantum optics and do not incorporate all the properties of a quantum-lightwave signal that could be used to convey information measured in bits. Classical optics is based on continuous quantities and includes *geometrical optics* (also called *ray optics*) and *wave optics*. Wave optics models a lightwave signal as a continuous electromagnetic waveform with an amplitude, phase, and polarization that are described by Maxwell's equations. This signal model, which treats the lightwave energy as a continuous quantity, is a limiting form of quantum optics. The application of these equations to an optical waveguide results in characteristic spatial patterns called *modes* and a corresponding *dispersion relationship* that governs the linear propagation characteristics of a lightwave signal.

Within the wave-optics signal model, the *coherence* of the lightwave plays an important role in determining the appropriate form of analysis. An ideal *coherent* lightwave has a deterministic phase. Knowing the phase of a coherent lightwave at one time instant provides complete information about the phase of the lightwave at a different time instant regardless of how widely spaced the time instants are. An ideal *noncoherent* lightwave has a completely random phase. This means that knowing the phase of the lightwave at one time instant provides no information about the phase of the lightwave at a different time instant regardless of how closely spaced the time instants are.

The coherence properties of a lightwave as well as other kinds of randomness such as noise and randomness in a datastream that conveys information are treated within wave optics as forms of *statistical uncertainty*. This form of uncertainty arises from incomplete knowledge about one or more attributes of the system or the user datastream.

Ray optics models a lightwave as a ray of light. The simplicity of ray optics provides an elementary description of lightwave signal propagation and is the limiting form of wave optics whenever the relevant dimensions of the guiding structure are much larger than the wavelength λ of the light in the medium. When the dimensions of the structure are comparable to λ, or smaller than λ, ray theory is no longer adequate. Moreover, ray theory cannot typically be used to assess the phase of a lightwave or to incorporate noise.

Photon optics incorporates the granularity of the energy of a lightwave signal as expressed by the concept of a photon. Photon optics is a limiting form of quantum optics that excludes the phase of the lightwave, but does include the discrete-energy nature of a lightwave signal. Relative to communication systems that use lower frequencies, these discrete-energy effects are evident at lightwave frequencies because the quantum of energy of a lightwave photon is relatively large compared with the quantum of energy of other sources of randomness, as quantified later in this chapter.

Compared with wave optics, photon optics incorporates a distinct source of uncertainty that is fundamentally different than statistical uncertainty. This source of uncertainty is conventionally called *photon noise* and is attributed to the fact that photon arrival times are random. Within quantum optics, this source of randomness is a form of *quantum uncertainty* associated with a photon-counting measurement of a conventional lightwave source. This form of uncertainty is recognized throughout the book, and is analyzed in detail in Section 15.2.5. Whereas statistical uncertainty need not be present, quantum uncertainty is always present when a conventional lightwave is photodetected. Accordingly, the quantum uncertainty as expressed by photon noise is deemed fundamental. A pragmatic combination of wave optics and photon optics is usually adequate for most systems. This combination is called a *semiclassical signal model* for lightwaves.

Quantum optics incorporates additional properties of a lightwave signal that are not incorporated into photon optics or wave optics. One of the unique properties of a quantum-lightwave signal is a quantum correlation or quantum coherence structure. The description and use of quantum coherence is at odds with “common-sense” notions, based on macroscopic observations, of how signals interact. Accordingly, systems based on the quantum-optics signal model that use these properties can seem to be counterintuitive. The quantum-optics signal model is discussed in Chapter 15.

Relative to quantum optics, wave optics is incomplete because it does not incorporate the discrete nature of a lightwave signal, or photon noise, nor can it fully describe quantum coherence effects. Photon optics is incomplete because it is based solely on the discrete energy of a lightwave signal. This model does incorporate photon noise, but does not incorporate phase or modal structure.

1.2.1 Relationship of Wave Optics to Photon Optics

One of the challenges of this book is to smoothly transition between these signal models, presenting the relevant physical conditions when each signal model is appropriate. To that end, we provide a preliminary discussion of the relationship between the wave-optics signal model, which is based on a continuous-energy signal that has an amplitude and a phase, and the photon-optics signal model, which is based on a discrete-energy signal that does not incorporate the phase. Wave optics is necessary to describe propagation. Photon optics is necessary to describe the physics of lightwave generation and photodetection.

Wave optics models a lightwave signal as a continuous electromagnetic wave. The relationship between the continuous power $P(t)$ and the continuous energy $E(t)$ defined over a time interval of duration T is given by

$$E(t) = \int_{t-T/2}^{t+T/2} P(\tau)\mathrm{d}\tau. \tag{1.2.1}$$

Each of these quantities could have statistical uncertainty because of randomness in the lightwave source or in the subsequent continuous sources of noise.

Photon optics models a lightwave signal as a stream of photons. Each photon has an energy given by

$$E = hf = \hbar\omega = \frac{hc}{\lambda}, \tag{1.2.2}$$

where h is *Planck's constant*[3] and $\hbar = h/2\pi$ is the *reduced Planck constant*, f is the *frequency* of the light expressed in hertz, $\omega = 2\pi f$ is the *angular frequency* of the light expressed in radians/second, λ is the wavelength of light in the medium, and $c = f\lambda$ is the velocity of light in the medium. This simple relationship implicitly couples the particle concept of energy with the wave concept of frequency.

Each photon has a *momentum* with magnitude p given by

$$p = \frac{h}{\lambda} = \hbar k, \tag{1.2.3}$$

where $k \doteq 2\pi/\lambda$ is defined as the *wavenumber*. This simple relationship implicitly couples the particle concept of momentum with the wave concept of wavenumber. Moreover, because $k = \omega/c_0$ (cf. (2.3.23)) in free space, $p = E/c_0$ in free space. This expression must be modified for a guided wave.

Photodetection of a Lightwave Signal

A lightwave signal interfaces with an electrical signal both at the transmitter and at the receiver. At the transmitter, a baseband electrical signal is converted into a modulated lightwave signal by a *modulator*. At the receiver, a lightwave signal is converted into an electrical signal by a *photodetector*. The term "photodetection" is used to describe the physical optical/electrical conversion process that transfers the modulation on the lightwave signal onto the electrical signal. The term "detection" is used to describe

[3] Planck's constant is $h = 6.626 \times 10^{-34}$ joule-seconds.

the system-level algorithmic process that determines the most probable transmitted sequence of letters.

Within the wave-optics signal model, photodetectors are sensitive to the lightwave signal power $P(t)$, which is the time-averaged square of a passband lightwave signal $\widetilde{s}(t)$ (cf. Section 1.3), where the time average is long with respect to the temporal period of the lightwave carrier but short compared with the information-bearing frequency components that are modulated onto the carrier. This signal is derived from a spatial integration over the aperture of the photodetector, with the lightwave power being the spatial integral of the lightwave intensity. The direct conversion of the continuous lightwave power $P(t)$ into an electrical signal $r(t)$ using a square-law photodetector is called *direct photodetection*. Within wave optics, the photodetector output $r(t)$, called the *photocurrent*, is related to the lightwave power $P(t)$ by

$$r(t) = \mathcal{R}P(t), \tag{1.2.4}$$

where $\mathcal{R}$ is a constant called the *responsivity* of the photodetector with units of amperes per watt. Often, for pedagogical convenience, we will set $\mathcal{R} = 1$.

Other methods of photodetection convert the lightwave amplitude into an electrical amplitude. Therefore, the optical/electrical conversion process can be linear with respect to the lightwave signal amplitude or it can be linear with respect to the lightwave signal power. The linear conversion of the magnitude and the phase of a passband lightwave signal into an electrical signal $r(t)$ requires the use of an additional reference lightwave signal that is added to the incident lightwave signal before photodetection. This type of optical/electrical conversion, called *balanced photodetection*, is introduced later in this chapter and is discussed in detail in Chapter 7.

Within the photon-optics signal model, a random number of photons is photodetected in any time interval T. This process is mathematically described as a *Poisson counting process* characterized by a *photon arrival rate* $\mathsf{R}(t)$. A complete description of this process is given in Section 6.2.3.

The bridge between wave optics and photon optics is established by associating the continuous wave-optics power $P(t)$, in watts, with the photon arrival rate $\mathsf{R}(t)$, in photons per second, as given by

$$\mathsf{R}(t) \doteq P(t)/hf, \tag{1.2.5}$$

where hf is the energy E of a single photon.

When the wave-optics lightwave power P is a constant, the photon arrival rate R is a constant. The resulting random number of photon counts measured in a time interval of duration T is described by a Poisson probability distribution with mean $\mathsf{R}T$. This probability distribution describes the quantum uncertainty conventionally attributed to photon noise. The randomness caused by photon noise in the continuous photodetected electrical signal $r(t)$ is called *shot noise*.[4]

When the wave-optics lightwave power P is a random variable, there is additional statistical uncertainty overlaid on the photon noise to generate a composite form of

[4] In other contexts, shot noise arises from the discreteness of the electron charge.

randomness that is a mixture of quantum uncertainty and statistical uncertainty. This composite form of uncertainty is described in Section 6.3.

The choice between a wave-optics model and a photon-optics model depends on the particular attributes under discussion. When the mean energy in the lightwave signal over an observation interval T is much larger than the energy of a single photon so that many photons are to be observed, the wave-optics signal model is typically used. When the mean energy over an observation interval is on the order of tens of photons or fewer, the discrete-energy property of a lightwave signal is evident so that the photon-optics signal model must be considered.

Examples of the magnitude of the quantities involved are useful. A wavelength[5] of 1500 nm corresponds to a frequency of $f = c/\lambda = 2 \times 10^{14}$ Hz. A photon at this wavelength has an energy of $E = hf = 1.33 \times 10^{-19}$ joules. A one-milliwatt (mW) lightwave source at this wavelength emits on average approximately $P/hf = 7.5 \times 10^{15}$ photons per second. This large number means that wave optics is an appropriate signal model for this signal level. In contrast, a one-nanowatt lightwave source at this wavelength emits on average approximately 7.5 photons per nanosecond. On this scale, the discrete-energy nature of a lightwave signal can be observed, and the use of a photon-optics signal model is needed for an accurate analysis.

At a radio frequency of 2 GHz, the energy of a single photon is five orders of magnitude smaller than the photon lightwave energy used in the previous example and is over three orders of magnitude smaller than the mean thermal energy kT_0 in the external environment at a normal temperature. Because the energy of a photon at a frequency of 2 GHz is over one thousand times smaller than the average energy of the external environment at a normal temperature, the discrete-energy property of a radio-frequency signal cannot be observed. Accordingly, a wave-optics signal model that ignores the discrete-energy nature of a photon is the conventional signal model used for lower-frequency systems.

1.2.2 Choosing a Signal Model

The use of one signal model does not reject the conclusions of the other signal model. For example, the use of photon optics, which is based on discrete energy, does not inform us about the phase of the signal or the structure of modes, which are features of wave optics. Accordingly, the conclusions drawn from both signal models, such as the discrete energy and the continuous phase, may be appropriate for the same system. This apparent incompatibility is a consequence of the incompleteness of both wave optics and photon optics, each of which fails to model different properties of a lightwave signal. The quantum-optics signal model provides a common framework that incorporates the complete set of properties of a lightwave signal and is discussed in Chapter 15.

A combination of ray optics, wave optics, and photon optics will be used to analyze lightwave communication systems. Ray optics and wave optics are used to describe

[5] This wavelength, defined in free space, is near the attenuation minimum of a conventional optical glass fiber.