NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Electrical and Electronic Engineering

OPTICAL COMMUNICATION

Graduation Project
EE- 400

Student: Aneel Shahzad Baig (971305)

Supervisor: Prof. Dr. Fakhreddin Mamedov

Nicosia - 2001

ACKNOWLEDGMENT

I wish to thank to my parents who supported and encouraged me at every stage of my education and who still being generous for me as they are ever.

My sincere thanks and appreciation for my supervisor Prof. Dr. Fakhreddin Mamedov who was very generous with his help, valuable advices and comments to accomplish this research and who will be always my respectful teacher.

All my thanks goes to N.E.U. educational staff especially to electrical & electronic engineering teaching staff for their generosity and special concern of me and all E.E. students.

Final acknowledgments go to my class mates and friends Kashif, Umair, Zubair, Salman, Usman and Taufiq who provided me with their valuable suggestions throughout the completion of my project.

ABSTRACT

A communication network consists of interconnected links, each of which has three basic elements: transmitter, channel and receiver. An optical fiber consists of a core and a cladding layer. According to core diameters there are single mode and multimode fibers. Fiber attenuation and dispersion limit the transmission capacity and distance. Attenuation is caused by photon absorption, scattering, fiber bending and coupling. Different propagation velocities of components in the lightwave cause fiber dispersion.

Photdetectors convert incident light into current from photon absorption and electron hole pair (EHP) generation. Noise and distortion are important performance limiting factors in signal detection. Thermal noise is white Gaussian noise due to random thermal radiations. Shot noise is caused by random EHP generations in photodiodes. APD noise is due to both random primary an secondary EHP generations. RIN is caused by intensity fluctuation. Phase noise is the phase fluctuation. Mode partition noise is caused by random power distribution among longitudinal modes. Transmission performance in digital communications is measured by the BER, which is in turn determined by the received signal power, noise power, intersymbol interference and crosstalk.

In time domain medium access, stations send data over different time intervals. In TDMA time is divided into frames, which are fixed in size. TDM has fixed frame size consisting of the overhead and fixed number of slots. WDoMA systems differ in various aspects. According to the relative size between channel separation and bandwidth. WDoMA systems are either dense or sparse. There is WDM, where signals from set of transmitters are multiplexed for transmission over a long distance. Photonic switching can avoid the electronics speed bottleneck by providing optical domain switching. A photonic switch needs to be able to configure optical paths according to the traffic patterns. External modulation is used to modulate the refractive index of a waveguide through which a constant frequency lightwave passes. In contrast to direct detection, coherent detection has the capability to detect the phase, frequency, amplitude and even the polarization of the incident light signal. When transmission system power is limited, optical amplifiers are used to improve the received SNR and increase the repeater distance.

INTRODUCTION

In this project optical communications is studied with intensive care. Optical communication is a new technology, which will have a large impact on telecommunications as well as fast data transmission and computer interconnections. The project consists of five chapters.

Chapter 1 gives a short introduction to optical communications by considering the historical development, the general system and the major advantages provided by this technology.

Chapter 2 gives the basic idea of a communication system and a communication network. Important elements in the communication process are introduced, which consist of Light sources, Optical Fibers, and Light Detection devices.

Chapter 3 will cover important and advanced topics in optical communications. On the transmitter side direct modulation is discussed in detail. On the receiver side Coherent Detection and timing recovery are explained. Finally, on the subject of light signal transmission technology, recent technology developments on optical amplifiers are discussed. These recent breakthroughs were developed to overcome the transmission loss and dispersion problems over a very long distance. It evaluates the important aspect of signal transmission, which is noise analysis. It also gives the brief idea of Noise and Distortion, their Reduction Techniques, Shot Noise in Pin Diodes, Relative intensity noise in Laser Diodes, Other types of noise and Total Noise of a system.

Chapter 4 covers the important Optical Networking aspects and supporting optical devices. Specifically, various multiplexing and switching techniques in optical communication are explained, and various innovative devices are described.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION TO OPTICAL COMMUNICATIONS

1.1 Introduction

Communication is an important part of our daily lives. It helps us to get closer to one another and exchange important information. The communication process involves information generation, transmission, reception, and interpretation. As needs for various types of communication, such as images, voice, video, and data communications, increase, demands for large transmission capacity also increase. This need for large capacity has driven the rapid development of light-wave technology to support worldwide digital telephony and analog cable television distribution.

Communications using light is not a new science. Old Roman records indicate that polished metal plates were sometimes used as mirrors to reflect sunlight for long range signaling. The military used similar sunlight powered devices to send telegraph information from mountaintop to mountaintop in the early 1800s. For centuries the navies of the world have been using and still use-blinking lights to send messages from one ship to another. Back in 1880, Alexander Graham Bell experimented with his "Photo phone", which used sunlight reflected off a vibrating mirror and a selenium photo cell to send telephone like signals over a range of 600 feet. During both world wars some light wave communications experiments were conducted, but radio and radar had more success and took the spotlight. It wasn't until the invention of the laser, some new semiconductor devices and optical fibers in the 1960s that optical communications finally began getting some real attention.

During the last thirty years great strides have been made in Electro-optics. Light beam communications devices are now finding there way into many common appliances, telephone equipment and computer systems. On-going defense research programs may lead to some major breakthroughs in long-range optical communications. Ground-station to orbiting satellite optical links have already been demonstrated. Today, with the recent drop in price of some critical components, practical through-the-air communications systems are now within the grasp of the average experimenter. You can now construct a system to transmit and receive audio, television or even high-speed computer data over long distances using rather inexpensive components.

1.2 Communication systems

An optical communication system is a communication system that uses the light waves as the carrier for transmission. A communication system can be referred to as point-to-point transmission link. When many transmission links are interconnected with multiplexing or switching functions as described in Figure 1.1, they are called a communication network. A communication network allows us to communicate with one another via shared transmission facilities.

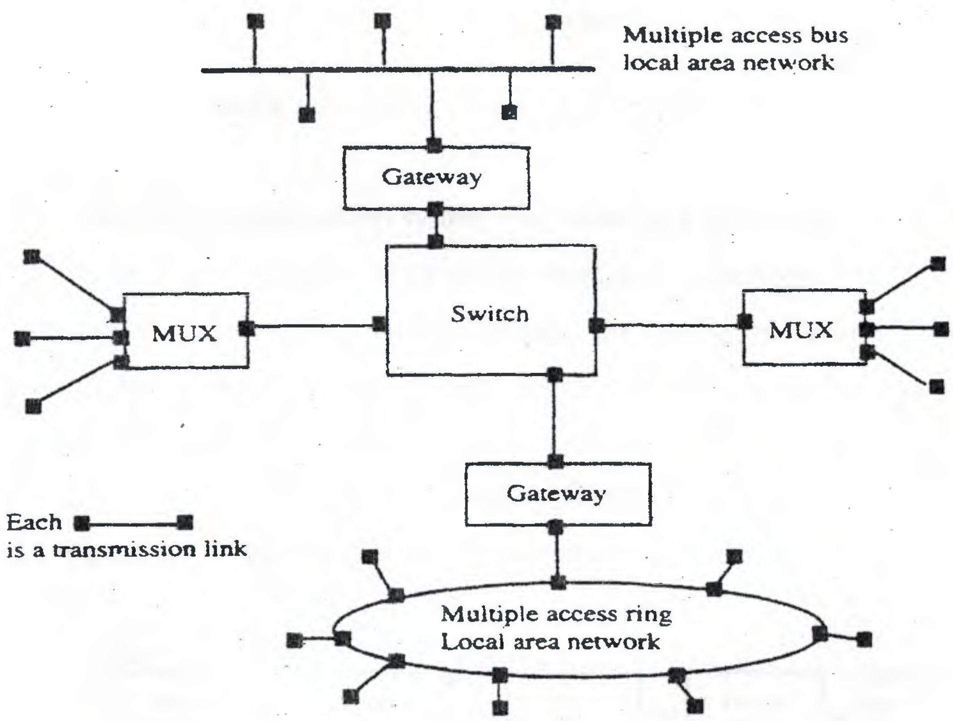

Figure 1.1 A communication network of interconnected links and subnetworks.

Each communication link shown in figure 1.1 consists of three basic components: a transmitter, a channel, and a receiver. As illustrated in figure 1.2 the transmitter converts the input message to a form suitable for transmission through the communication channel, which is a medium guiding the transmitted signal to the receiver. In most cases, the final received signal is corrupted by noise. In Figure 1.2, the noise comes from the channel, but this is only illustrative. In optical fiber communication systems, noise from the fiber channel is negligible. On the other hand, there are multiple noise sources from both the transmitter (light source) and the optical receiver. In addition to noise, the received signal can also be corrupted by distortion from a nonideal channel. Therefore, the challenge of the receiver design is to recover the transmitted signal from the corrupted form.

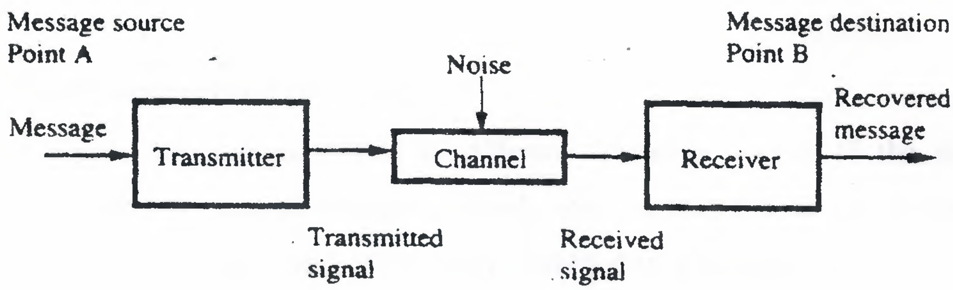

Figure 1.2 A point-to-point transmission link.

In a practical communication system, the transmitter and receiver can be further broken down into many smaller blocks. For example, a transmitter may consist of blocks performing source coding, channel coding, line coding, modulation, and signal amplification. And a receiver may include blocks performing equalization, retiming, detection, demodulation, and decoding. These blocks are illustrated in Figure 1.3. In optical communications, because transmission speeds are usually very high, simple coding, retiming, and equalization are important in the system design

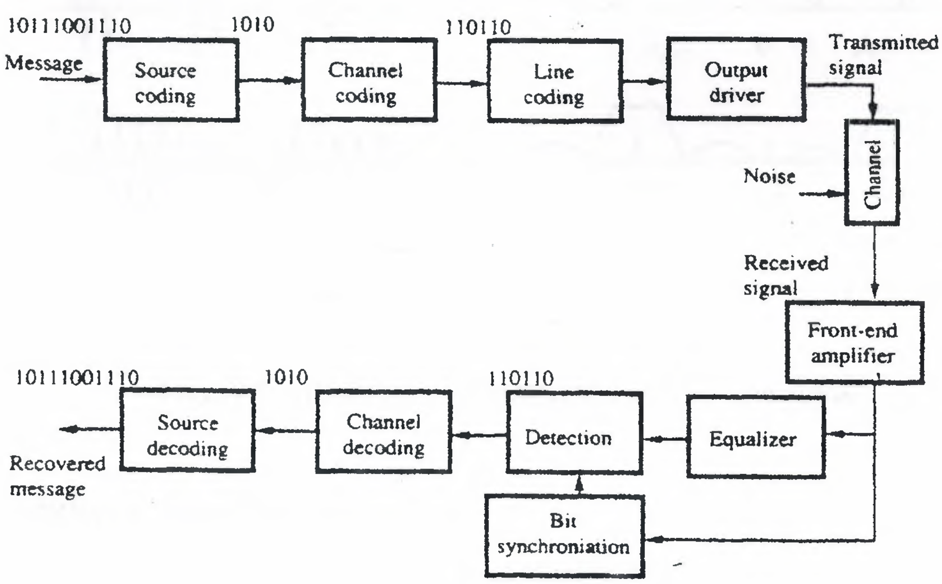

Figure 1.3 A more detailed look at the communication system.

There are many different communication systems in which signals are transmitted and detected in different ways. Differences in communication systems can be characterized in three aspects: (1) baseband versus passband, (2) analog versus digital, (3) coherent versus incoherent detection.

1.3 Baseband versus passband

A signal can be transmitted in different frequency bands. If the signal is transmitted over its original frequency band, the transmission is called *baseband transmission.* On the other hand, if the signal is shifted to a frequency band higher than its original baseband, it is called *passband transmission.* Some baseband and passband signals are illustrated in Figure 1.4.

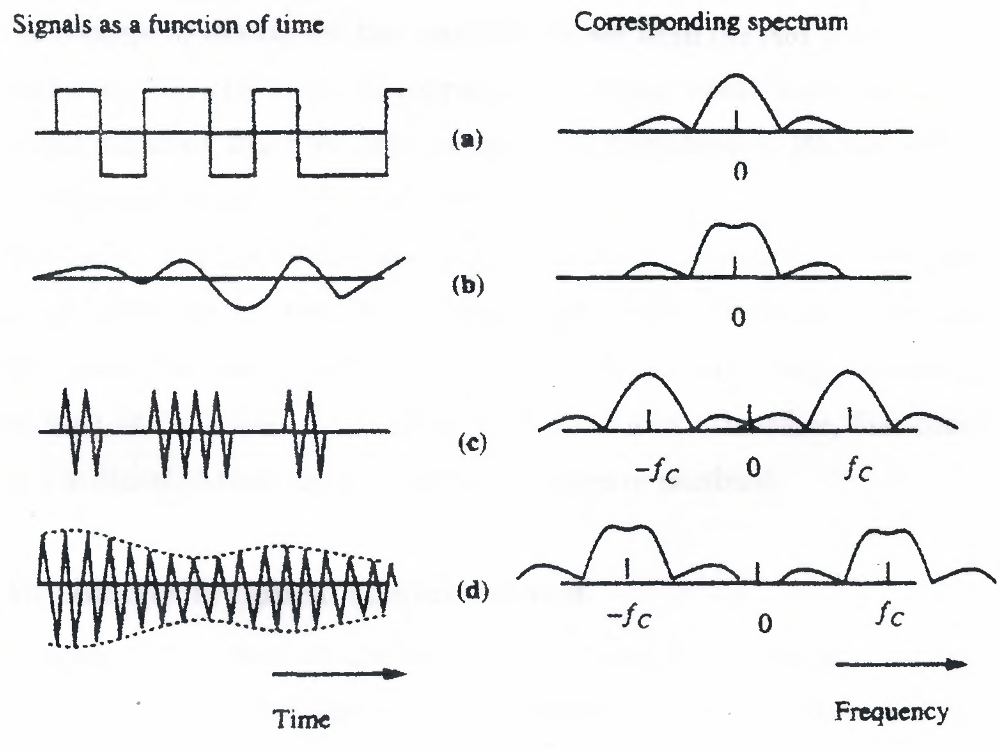

Figure 1.4 Illustration of baseband and passband signals.

1.4 Analog versus digital

Another important characteristic in communications is the discreteness of a message that is transmitted. For example, if a digitized image is transmitted in discrete binary bits. This is called digital communication. On the other hand, in case of AM and FM, input signals have a continuous waveform. This is called analog communication.

1.5 Coherent versus incoherent

When a transmitted signal is received, we need to detect what was originally transmitted. If the signal is passband, it must be shifted back to the baseband. There are two different ways to do this: coherent and incoherent detections. In coherent detection, a different carrier source at the receiver side is used to demodulate the received signal or to shift the passband signal back to the baseband. This carrier is generally called the *local carrier* and is synchronized to the received signal in frequency and phase. On the other hand, in incoherent detection, there is no use of the local carrier. Instead, some nonlinear processing is used to extract the amplitude or envelope of the passband signal.

1.6 Modulation and line coding

The concept of modulation has been well known from the AM and FM. Similar to modulation that converts a baseband signal to a passband signal, line coding converts a binary input sequence into a suitable waveform for transmission. Because of similar functions, line codes are also called *modulation codes.*

Modulation and line coding are used for different communication systems. In general, modulation can be used in both digital and analog communications. On the other hand, line-coding maps a finite set of signals to another set of signals with certain properties such as dc balance or frequent level transitions. Therefore, line coding is always used in digital communication, whether baseband or passband.

1.7 Advantages of optical communication

Because of objective of communication system is to transmit messages, a communication system is evaluated by its transmission capacity, distance, and fidelity. Recent active research and development in optical communications has been driven by its many advantages, such as large capacity and long distance transmission. Some important advantages are discussed below.

1.7.1 Large transmission capacity

As mentioned earlier, when higher frequency carriers carry signals, more information can be transmitted. For example, the capacity in microwave communication is around several hundred MHz, and it is several thousand GHz in optical communication. As a result, transmission capacity in optical communication is not limited by the optical channel but by the electronic speeds. This motivates the use of parallel transmissions such as WDM.

1.7.2 Low loss

Another important advantage is the low transmission loss in optical fibers. Because of recent advancements, fiber attenuation can be as low as 0.2 dB/km at a wavelength of 1.55 /xm.[2] In contrast, depending on the- specific carrier frequency and the gauge of the cable, microwave waveguide loss is on the order of 1 dB/km and twisted-pair wires is around 10 dB/km. In other words, if only loss is considered, optical fibers can transmit signals 5 times farther than waveguides and 50 times farther than twisted-pair wires.

1.7.3 Immunity to interference

Because of the waveguide nature and easy isolation, optical signals can be easily confined in a fiber without any external interference. In contrast, twisted-pair and radio transmissions have significant crosstalk and multi-path interference.

1.7.4 High-speed interconnections

Optical communication is also well suited for highspeed interconnections. Unlike electrical signals, which require a careful control of impedance matching, optical signals can be easily transmitted and received through free space or fiber connections.

1.7.5 Parallel transmission

Because optical signals can be transmitted in free space, parallel transmission in three dimensions is possible. This provides powerful ways to interconnect large numbers of processors for parallel processing, photonic switching, and optical computing.

CHAPTER 2

OPTICAL COMMUNICATION DEVICES

2.1 Light Sources

In optical communication, a transmitter consists of a light source and a modulation circuit. The light source generates an optical frequency carrier, and the modulation circuit modulates the carrier according to the transmitted signal.

Carrier is characterized by its amplitude, frequency, and phase. An ideal carrier can be expressed by

$$c(t) = A\cos(\omega_c t + \phi) \tag{2.1}$$

Where its amplitude is A, circular frequency ω_c, and phase ϕ are time-invariant. In practice, however, a light source driven by a constant current source has the following form:

$$c(t) = \sum_i A_i \cos[\omega_{c,i} t + \phi_i(t)] \tag{2.2}$$

Where each carrier at frequency $\omega_{c,i}$, represents a *longitudinal mode* and $\phi_i(t)$ is a random phase noise. The carrier with the largest amplitude A_i is called the *main mode* and the others are called *side modes*. A multimode source is not good for communication.

2.1.1 Semiconductor Light Sources

Semiconductor light sources are the most important kind of light sources used in optical communications. Their small size, low power consumption are among the reasons for the popularity. More significantly, they can generate light signals at wavelengths $1.3\mu m$ and $1.55\mu m$, where the minimum fiber attenuation and/or dis-persion can be achieved. There are two types of semiconductor light sources: light emitting diodes (LEDs) and laser diodes (LDs).

(a) *P-N junction*

The core of a semiconductor light source is its p-n junction, often referred to as the *active layer*. The p-n junction is an interface between an N-type doped layer and a P-type doped layer. Therefore, a semiconductor light source has essentially the same p-n junction as any semiconductor diode.

The first difference is that the material used must have a direct energy bandgap. In a direct bandgap material such as GaAs, electrons and holes have their minimum energies at the same momentum.

In an indirect bandgap material such as silicon, on the other hand, the energy bands have different momenta at their minimum energies. Due to conservation of momentum, an electron and a hole at the same momentum can directly recombine and generate a photon.

From the above observation, the significance of direct bandgap is twofold. First, a high direct electron and hole recombination (EHR) rate can be achieved because electrons and holes have the highest population at their minimum energy states. Second, because the energy release from each direct EHR is around the energy bandgap E_g, according to the Planck-Einstein relationship, photons generated by direct EHRs are around the same frequency

$$f = \frac{E_g}{h} \tag{2.3}$$

where h is the Planck constant. If EHRs are not direct but take place through an intermediate interaction with the material lattice, each recombination can generate two particles (not necessarily photons) of smaller energies ΔE_1 and ΔE_2 with $E_g = \Delta E_1 + \Delta E_2$.*Be*-cause ΔE_1 and ΔE_2 are undefined and random, indirect EHRs result in an undesirable wide spectrum.

2.1.2 Light Emitting Diodes

Light-emitting diodes are semiconductor diodes that emit *incoherent light* when they are biased by a forward voltage or current source. Incoherent light is an optical carrier with a rapidly varying random phase. This random phase results from independent EHRs. That is, the phase and frequency of a photon generated from an EHR are different from those of photons from other EHRs. Therefore, incoherent light has a broad spectrum.

The linewidth of a light source can be defined in different ways. One common definition is called *full-width half-maximum* (FWHM), which is the width between two 50 percent points of the peak intensity.

$\lambda f = c$

where c is speed of light, taking the total derivative, we have

$f\partial\lambda + \lambda\partial f = 0$

For a given linewidth $\Delta\lambda$, we thus have

$$\frac{|\Delta\lambda|}{\lambda} = \frac{|\Delta f|}{f}, |\Delta\lambda| = c\frac{|\Delta f|}{f^2}, |\Delta f| = c\frac{|\Delta\lambda|}{\lambda^2} \qquad (2.4)$$

where Δf is the corresponding spectral width.

The spectrum width of LEDs depends on the material, temperature, doping level, and light coupling structure. For AlGaAs devices, the FWHM spectrum width of LEDs is about $2kT/h$, where k is Boltzmann constant and T is temperature in Kelvin. For InGaAsP devices, it is about $3kT/h$. As the doping level increases, the linewidth also increases.

The spectrum width also depends on the light coupling structure of the LED. The light coupling structure couples photons out of the active layer. There are two different light coupling structures: surface emitting and edge emitting. The first type couples light vertically away from the layers and is called surface emitting or Burrus LED. The second type couples light out in parallel to the layers and is called an edge emitting LED.

Because of self-absorption along the length of the active layer, edge emitting LEDs have smaller line widths than those of surface-emitting diodes. In addition, because of the transverse waveguiding, the output light has an angle aroung 30^o vertical to the active layer. On the other hand, because surface-emitting LEDs have a large coupling area, it is easier to interface them with fibers. Also they can be better cooled because the heat sink is close to the active layer.

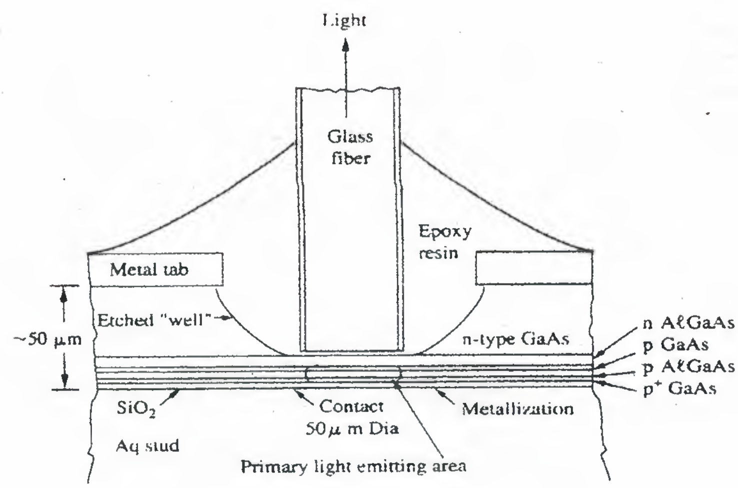

Figure 2.1 Illustration of a surface-emitting diode.

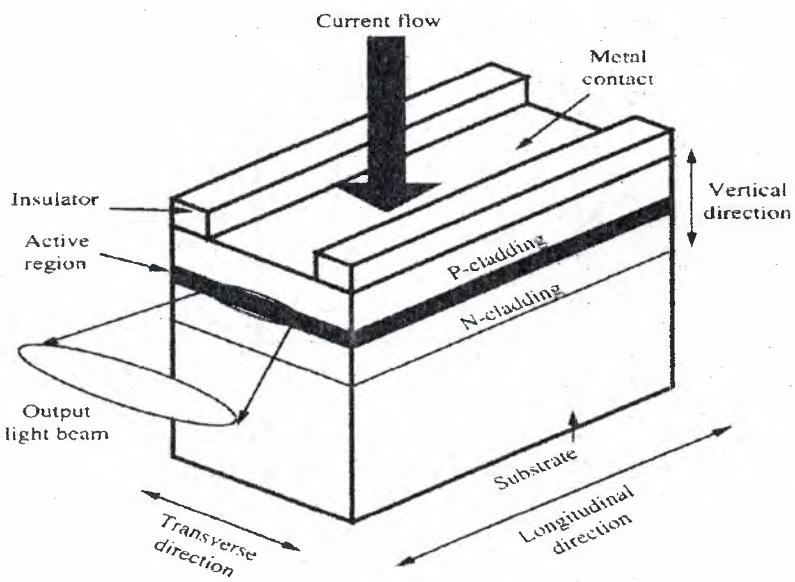

Figure 2.2 Illustration of an edge-emitting diode.

2.1.3 Laser Principles

Another important type of light source used in optical fiber communications is the laser diode (LD). A basic laser diode structure is similar to that of the edge-emitting LED illustrated in Figure 2.2. By adding an additional structure for transverse photon confinement, a coherent carrier close to what is expressed in Equation (2.1) can be generated. Like any other lasers, the principles of semiconductor lasers are based on *(a) external pumping* and *(b) internal light amplification.*

(a) External pumping

When a laser has several energy states, external pumping excites carriers to a higher energy state. When they return to the ground state, they release energy and generate photons. As an illustration, Figure 2.3 shows a simple two-level system in which carriers (e.g., electrons) can stay at one of the two energy states E_1 and E_2. When there is no external pumping, most carriers stay at the ground state E_1 because of thermal stability. When there is external pumping, on the other hand, carriers at E~ can be excited to E_2.

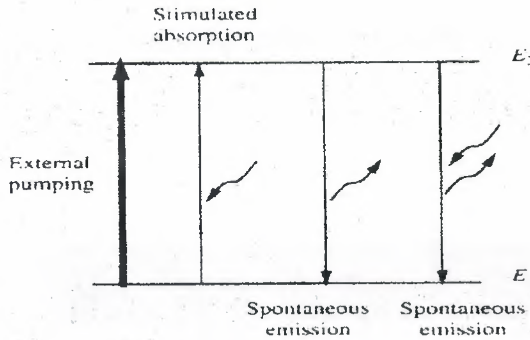

Figure 2.3 A two energy level system

(b) Light amplification

Photon generation from external pumping is not sufficient for coherent light generation. An additional amplification mechanism is needed to multiply photons of the same frequency and phase. In a laser, this is made possible by a quantum phenomenon called *stimulated emission.* When the optical loss in a laser cavity is small from a good optical confinement, a net positive amplification gain from stimulated emission can be achieved. As a result, coherent photons can be built up.

As shown in figure 2.3, several photon emission and absorption processes exit in a two-level atomic system. When a carrier is pumped to the upper state, it can comeback to the ground state either spontaneously or by stimulation. The corresponding photon generation processes are called *spontaneous emission* and *stimulated emission,* respectively. In spontaneous emission, photons generated have random phases and frequencies. As a result, the light is incoherent. On the other hand, photons generated from stimulated emission have the same phase and frequency as the stimulating photons.

Therefore, for coherent light generation, stimulated emission needs to dominate over spontaneous emission. In addition to photon emission, figure 2.3 shows that photons can also be absorbed to excite carriers from the ground state to upper state. This is called *stimulated absorption.*

Electrons and holes in a semiconductor have emission and absorption processes similar to those discussed above. As shown in figure 2.4, electron hole pairs (EHPs) are generated by external current injection and stimulated photon absorption. They can later recombine either spontaneously or by stimulated emission. Because of leaky current at the p-n junction, there is also non-radiative carrier recombination in semiconductors.

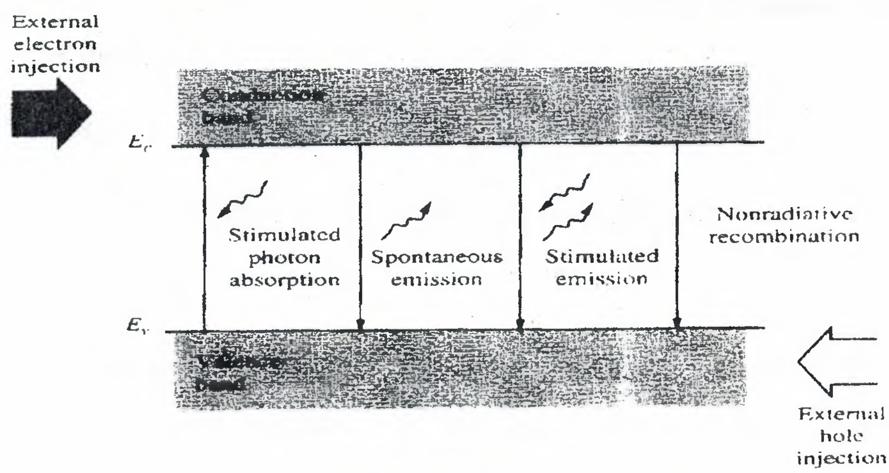

Figure 2.4 Carrier recombination and photon emission in a semiconductor laser diode.

(c) Positive optical gain

In the equilibrium state, carriers have the same rate between the two states. Thanks to Einstein, we have

$$A_{21}N_2 + B_{21}N_2S = B_{12}N_1S + R_P \qquad (2.5)$$

Where S is the photon energy density, N_1 and N_2 are carrier densities at E_1 and E_2 respectively, B_{12} is stimulated absorption rate coefficient, B_{21} is the stimulated emission rate coefficient, A_{21} is the spontaneous emission rate coefficient, and R_P is the external pumping rate in a unit volume. Therefore, the left-hand side of equation (2.5) is the total rate from E_2 to E_1, and the right-hand side is the total rate from E_1 to E_2.

From equation (2.5), when there is no external pumping $(R_P = 0)$, the photon energy density S in equilibrium is

$$S = \frac{A_{21}/B_{21}}{(B_{12}N_1/B_{21}N_2)-1} = \frac{8\pi hf^3n^3}{c^3}\left(\frac{1}{e^{hf/kT}-1}\right) \quad [joule/m^3.Hz] \qquad (2.6)$$

where the last equality comes from the Planck-Boltzmann's black-body radiation distribution. Equation (2.6) implies $B_{12} = B_{21}$ and $(A_{21}/B_{21}) = (8\pi hf^3n^3/c^3)$, where n is the refractive index of the medium.

To increase the stimulated emission rate, the light intensity S must be increased. In general, the light intensity increases at a rate proportional to $(B_{21}N_2 - B_{12}N_1)S$, which is the photon stimulated emission rate minus the photon stimulated absorption rate. Building up the light intensity thus requires that

$$B_{21}N_2 > B_{12}N_1 \qquad (2.7)$$

This condition is commonly referred to as *population inversion* and requires that

$N_2 > N_1$ when $B_{21} = B_{12}$.

To achieve population inversion, the external pumping rate (R_P) needs to be greater than the spontaneous emission rate. This can be seen by rearranging equation (2.5) as

$$S = \frac{R_P - A_{21}N_2}{B_{21}N_2 - B_{12}N_1} \qquad (2.8)$$

Because $S > 0$, R_p needs to be higher than the spontaneous emission rate $(A_{21}N_2)$ when

$$B_{21}N_2 > B_{12}N_1$$

2.1.4 Fabry-Perot Laser Diodes

A basic LD that has a rectangular cavity is equivalent to a Fabry-Perot (FP) resonator. An LD that has a simple rectangular cavity is thus called an FP LD.

(a) Power loss and gain

When photons are reflected back and forth between the two FP cavity ends, they experience both gain and loss. The gain comes from stimulated emission, and the loss comes from medium absorption and partial reflections at the two cavity ends. In the steady state, the gain and loss are equal. From this condition, the required gain from stimulated emission can be derived.

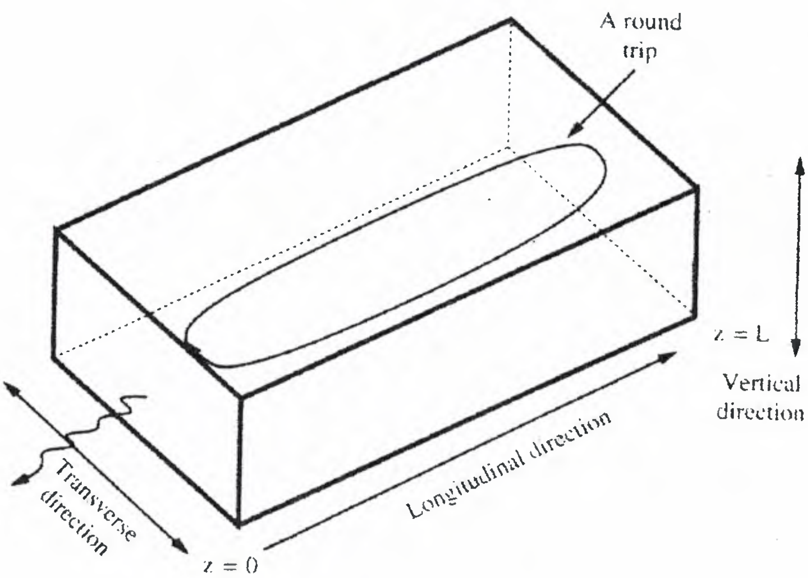

Figure 2.5 Illustration of a laser cavity

For simple discussion, assume the wave function of a light wave is a scalar and has a value A (a complex variable) at point $z=0$. From Figure 2.5, the wave traveling to the right can be expressed as

$$E^{+}(z) = Ae^{(j\beta_z - \alpha/2 + g/2)z}$$

where β_z is the propagation constant along the z-axis, and α and g are the distributed medium loss and gain, respectively. The time dependence factor e^{-jwt} of the field is dropped for its irrelevance, and the power of the traveling wave is proportional to $e^{(g-\alpha)z}$. At $z = L$, the wave function has value $Ae^{(j\beta_z - \alpha + g)L}$ Assuming the cavity end on the right has a reflection coefficient of r_2, the reflected wave traveling to the left can be written as:

$$E^{-}(z) = r_2 Ae^{(j\beta_z - \alpha/2 + g/2)(2L-z)}$$

The round trip is complete after the left-traveling wave is reflected back by the left mirror with reflection coefficient r_1. Therefore, the conditions of a unit round trip gain in the steady state are

$$r_1 r_2 e^{(g-\alpha)L} = 1 \qquad (2.9)$$

and

$$2L\beta_z = 2m\pi \qquad (2.10)$$

where m is an integer. Equation (2.9) can be expressed as

$$g = \alpha - \frac{1}{L} * \ln(r_1 r_2) = \alpha + \alpha_m = \alpha_{tot} \qquad (2.11)$$

where α_m accounts for the reflection loss at the cavity ends. Equation (2.11) is the gain-loss condition at the steady state, and equation (2.10) is the phase condition for the laser wavelength. This second condition is the basis for the longitudinal modes.

The gain-loss condition in equation (2.11) is only applicable to the steady state. Before the laser reaches its steady state, the gain is greater than the total loss. This builds up the radiation field in the laser by stimulating more photons from carrier recombination. As the laser field is being built up, there are more stimulated emissions or EHRs. This brings down the carrier density and also the gain. Finally a steady state is reached where the stimulated emission rate is in equilibrium with the carrier supply or generation rate. Therefore, the output power is determined by the injected current supply.

(b) Longitudinal modes

Substituting $\beta_z = 2\pi n/\lambda$ into the roundtrip phase condition given by equation (2.10), we have

$$\lambda_m = \frac{2Ln}{m} \tag{2.12}$$

where n is the refractive index of the gain medium and λ_m is the mth longitudinal mode. Where m is the large integer, the longitudinal mode separation between λ_m and λ_{m+1} is

$$\Delta\lambda_{long} = \lambda_m - \lambda_{m+1} = 2Ln\left\{\frac{1}{m} - \frac{1}{m+1}\right\} \approx 2Ln\frac{1}{m^2} = \frac{\lambda^2}{2Ln} \tag{2.13}$$

When an externally injected current generates carriers, depending on their energy distributions in the conduction and valence bands, they contribute to stimulated emission at different longitudinal modes.

2.1.5 Single Mode Laser Diodes

FP laser diodes generate undesirable multiple longitudinal modes. A different kind of laser diode called the single mode laser can suppress side modes and generate only one main longitudinal mode. Important single mode lasers include: DFB (distributed feed-back) lasers, DBR (distributed Bragg reflector) lasers, and cleaved-coupled-cavity (C^3) lasers.

(a) DFB lasers

DFB lasers use Bragg reflection to suppress undesirable modes. As illustrated in figure 2.6, there is a periodic structure inside the cavity with the period equal to Λ. Because of the periodic structure, a forward traveling wave has interference with a backward traveling wave. To have constructive interference, the roundtrip phase change over one period should be $2\pi m$, where m is an integer and is called the order of the Bragg diffraction. With $m = 1$, the first order Bragg wavelength is

$$2\pi = 2\Lambda \frac{2\pi n}{\lambda_B}$$

or

$$\lambda_B = 2\Lambda n \tag{2.14}$$

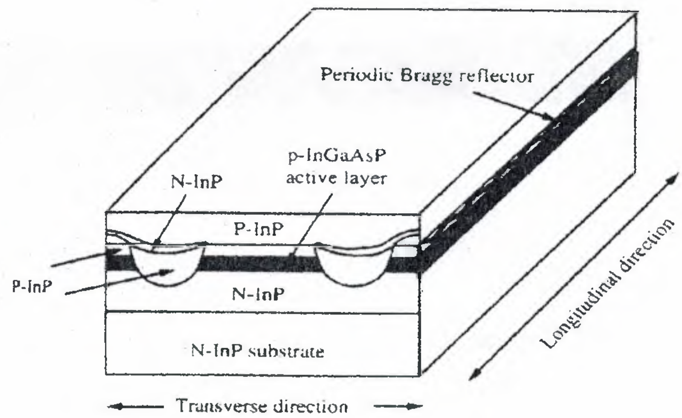

Figure 2.6 Illustration of a buried-hetrostructure DFB LD

Therefore, the period of the periodic structure determines the wavelength of the single-mode light output. In reality, a periodic DFB structure generates two main modes with symmetric offsets from the Bragg wavelength; a phase-shift of $\lambda / 4$ can be used to remove the symmetry. As illustrated in figure 2.7, the periodic structure has a phase discontinuity of $\pi / 2$ at the middle, which gives an equivalent $\lambda / 4$ phase shift.

(b) DBR lasers

DBR lasers use the same Bragg reflection principle to generate only one longitudinal mode. The difference between DBR and DFB lasers is that DBR lasers have the diffraction structure outside the laser cavity, as shown in figure 2.8. With this arrangement, the laser control and the frequency control can be done separately.

(c) Coupled cavity lasers

A coupled-cavity laser has two FP resonant cavities, which can be both active or one active and one passive. These lasers are illustrated in figure 2.9. In either case, the basic principle to generate only a single longitudinal mode is illustrated in figure 2.10.

$$\lambda = \frac{2L_1 n}{m_1} = \frac{2L_2 n}{m_2} \tag{2.15}$$

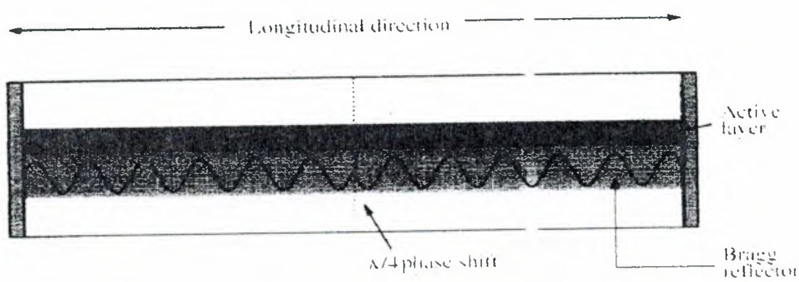

Figure 2.7 Illustration of a $\lambda/4$ phase-shift DFB LD.

Where m_1 and m_2 are two integers. With λ satisfying equation (2.15) and using equation (2.13), the modal separation is

$$\Delta\lambda_{long} = M_1 \frac{\lambda^2}{2L_1 n} = M_2 \frac{\lambda^2}{2L_2 n} \tag{2.16}$$

Where M_1 and M_2 are two mutually prime integers such that

$$\frac{L_1}{M_1} = \frac{L_2}{M_2} \overset{def}{=} L_0 \tag{2.17}$$

$$\Delta\lambda_{long} = \frac{\lambda^2}{2nL_0} \tag{2.18}$$

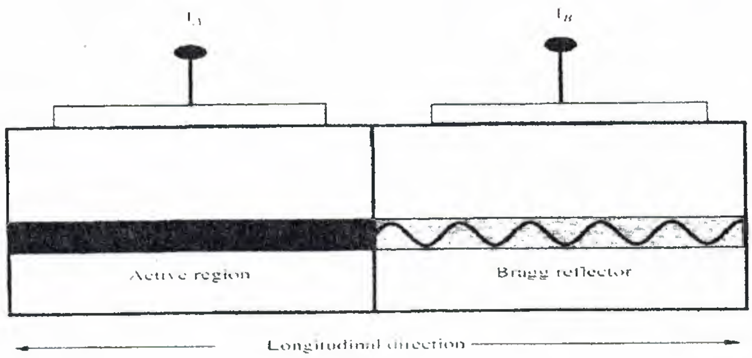

Figure 2.8 Illustration of a DBR LD.

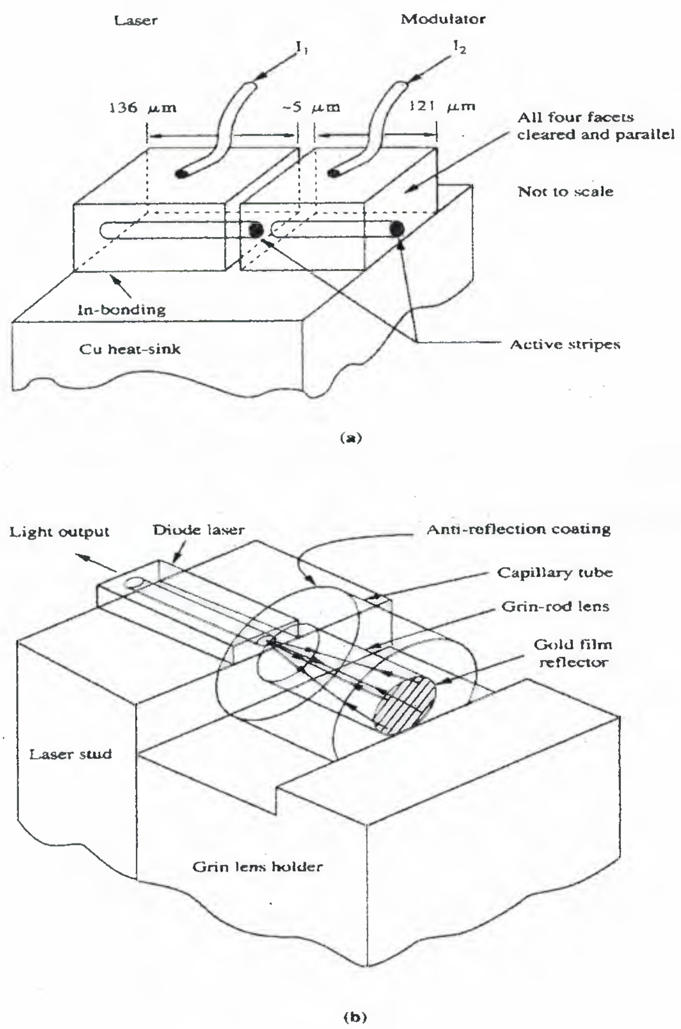

Figure 2.9 Illustration of coupled-cavity LDs (a) active-active, (b) active-passive

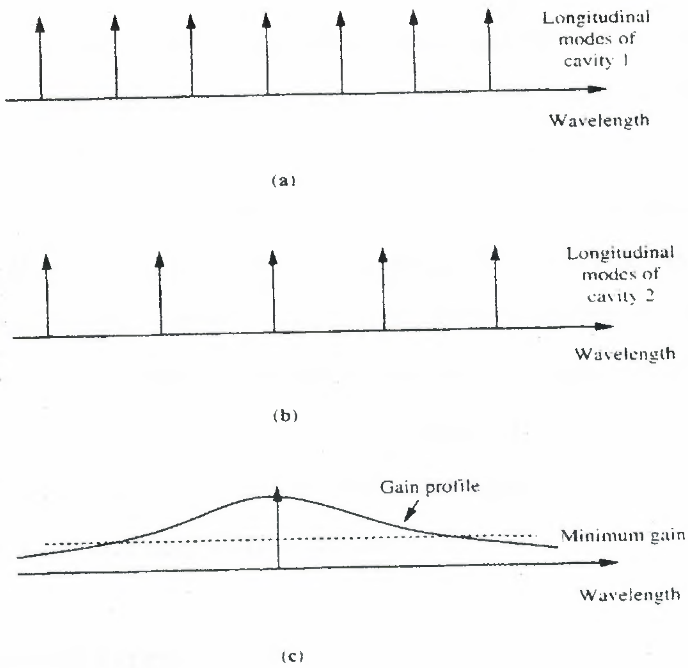

Figure 2.10 Principle of coupled-cavity diodes

2.2 Optical Fibers

Our current "age of technology" is the result of many brilliant inventions and discoveries, but it is our ability to transmit information and the media we use to do it, that is perhaps most responsible for its evolution. Progressing copper wire of a century ago to today's fiber optic cable, our increasing ability to transmit more information, more quickly and over longer distances has expanded the boundaries of our technological development in all areas.

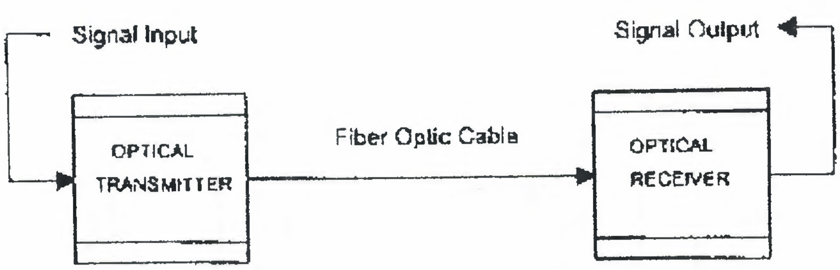

Figure 2.11 Illustration of Optical Fiber Communication

Today's low-loss glass fiber optic cable offers almost unlimited bandwidth and unique advantages over previously developed transmission media. The basic point-to-point fiber optic transmission system consists of three basic elements: the optical transmitter, the fiber optic cable and the optical receiver (figure 2.11).

Fibers are a common transmission medium used in optical communications. Compared with other transmission media such as space and wires, optical fibers provide low attenuation and strong immunity to electromagnetic interference (EMI). Because of these advantages, optical fibers have been used in long-haul undersea and interoffice communications Optical fibers are the communication channel in which light propagates. Like any other transmission medium, signal attenuation and distortion in optical fibers are important degradation factors.

2.2.1 Structure and Types

A bas c opt cal f ber has a c rcular cross sect on as dep cted n Figure 2.12 Although a practical fiber has many layers, only the core and cladding are important to light propagations. Both the core and cladding are typically made of silica glass, however, the core has a higher refractive index to confine light inside.

As light propagates inside a fiber, most of its power is confined in the core region, which is surrounded by the cladding. The cladding has a slightly lower optical density (or refraction index), typically between a fraction of 1 percent and a few percent. Most fibers have the cladding diameter around 125 μ m. Its size is generally not important to light propagations. Outside the cladding are several layers of protection jackets. The jackets prevent the fiber surface from being scratched or cut by mechanical forces.

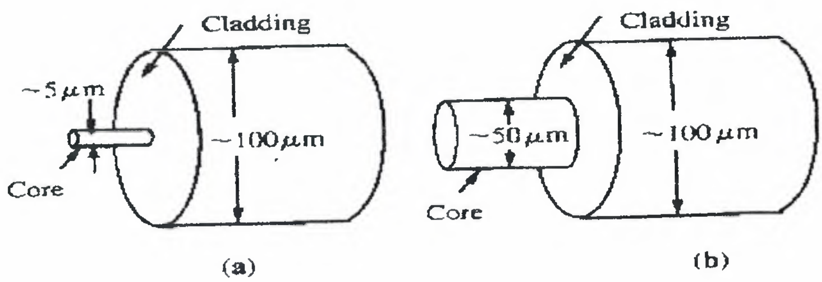

Figure 2.12 Basic structures of (a) Single mode (b) Multiple mode optical fibers.

(a) Single mode and multimode fibers

When a lightwave propagates inside the core of a fiber, it can have different EM field distributions over the fiber cross-section. Each field distribution that meets the Maxwell equations and the boundary condition at the core-cladding interface is called a *transverse mode.*

Several transverse modes are illustrated in Figure 2.13. As shown, they have different electric field distribution over the fiber cross-section. In general, different transverse modes propagate along the fiber at different speeds. Fibers that allow propagation of only one transverse mode are called *single-mode fibers* (SMF). Fibers that allow propagation of multiple transverse modes are called *multimode fibers* (MMF).

The key in fiber design to having single-mode propagation is to have a small core diameter. This can be understood from the dependence of the *cutoff wavelength* λ_C of the fiber on the core diameter. The cutoff wavelength is the wavelength above, which there can be only one single transverse mode. λ_C is expressed as

$$\lambda_c = \frac{2\pi a}{V}\left(n_1^2 - n_2^2\right)^{1/2} \qquad (2.19)$$

Where V=2.405 for step-index fibers. A is the core radius, n_1 and n_2 are the refractive indices of the core and cladding, respectively. This expression shows that fibers of a smaller core radius have a smaller cutoff wavelength.

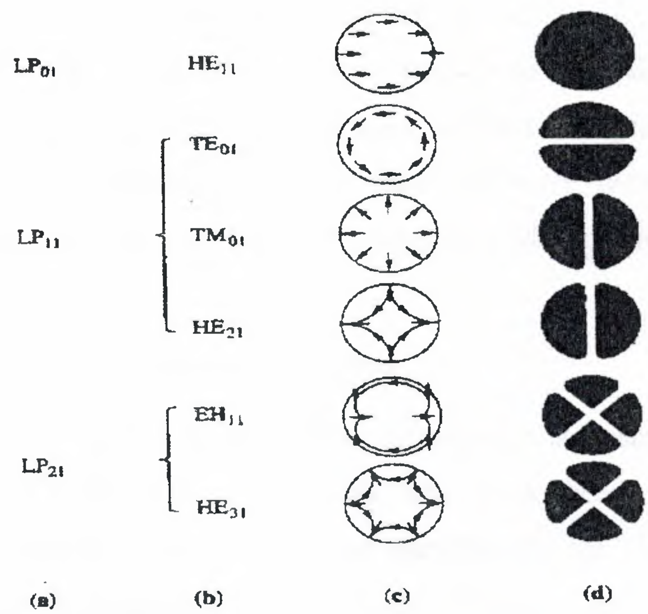

Figure 2.13 Some examples of low-order transverse modes of step-index fiber, (a) Linear polarized (LP) mode designations, (b) Exact mode designations, (c) Electric field distribution, (d) Intensity distribution of electric field component E_x

When the core diameter of a single-mode fiber is not much larger than the wavelength, there is a significant power portion or field penetration in the cladding. Therefore, it is necessary to define another parameter called *mode field diameter* (MFD). Intuitively, it is the "width" of the transverse field. Specifically, it is the *root mean square* (RMS) width of the field if the field distribution is Gaussian.

2.2.2 Light Propagation in Fiber Optics

In addition to attenuation, fiber dispersion is another limiting factor in lightwave transmission. Dispersion is a phenomenon that photons of different frequencies or modes propagate at different speeds. As a result, a light pulse gets broader as it propagates along the fiber. This section first explains the basic physics of light propagation in optical fibers.

Signal propagation in optical fibers can be described by either geometrical optics or Maxwell's equations. Geometrical optics is a good approximation when the wavelength of light is relatively small compared-to the system's dimensions. On the other hand, solving Maxwell's equations can tell the exact story but is much more mathematically complex. This section uses both methods to study light propagation, but avoids complex mathematics such as solving wave equations in optical fibers. Instead, it stresses the relationship between geometrical optics and wave functions. This approach shows the insight of the physics of lightwave propagation with minimal mathematics.

(a) Signal propagation by geometrical optics

The geometrical optics model for fiber propagation is illustrated in Figure 2.14, where incident light from a light source emits "rays" to a fiber in different directions. From Snell's refraction law, each ray will go partially into the cladding region or be totally reflected back depending on its incident angle θ_1 at the core-cladding boundary.

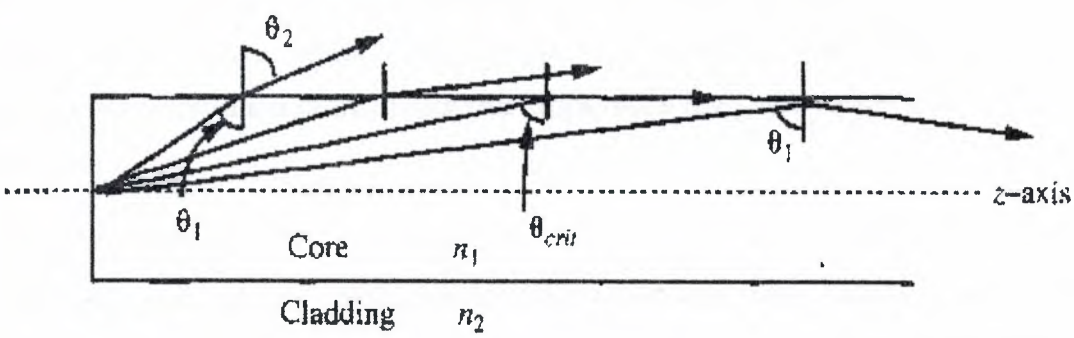

Figure 2.14 Light propagation using geometrical optics.

More spec f cally, a ray will go partially into the cladding if a θ_2 exists such that

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \tag{2.20}$$

Where n_1 and n_2 are the refractive indices of the core and cladding respectively. Because $n_1 > n_2$, a complete internal reflection is possible if

$$\theta_1 > \sin^{-1}\left(\frac{n_2}{n_1}\right) \overset{def}{=} \theta_{crit.} \tag{2.21}$$

That is, the corresponding ray will be totally reflected back to the core. As a result, a ray will propagate along the fiber without loss (other than various attenuation sources discussed earlier) if the incident angle θ_1 satisfies Equation (2.21).

Figure 2.15 provides two further, important observations first, rays with different $\theta_1 > \theta_{crit}$ have different z -component velocities. Specifically, for a ray of θ_1 its z component velocity is

$$v_z = \frac{c}{n_1}\sin(\theta_1) \tag{2.22}$$

This velocity dependence on θ_1 results in different propagation delays or dispersion. To reduce dispersion, graded-index fibers mentioned earlier can be used. Ray propagation in a graded-index fiber is illustrated in Figure 2.15. As shown, although rays of larger θ_1 propagate in a shorter distance, they travel through a higher refractive index region. On the other hand, rays of smaller θ_1 travel a longer distance but through a lower refractive index region. As a result, graded-index fibers can equalize the propagation delay of different propagation rays, and greatly reduce the fiber dispersion.

Another important observation is that rays at larger θ_1's have larger z-component velocities, and consequently smaller radial velocities. Intuitively, we see that the larger the radial velocity, the higher the penetration of the light power into the cladding region. As the wave analysis will demonstrate shortly, rays of larger radial velocities correspond to higher propagation modes. When the radial velocity becomes too large or $\theta_1 > \theta_{crit}$, the ray will propagate into the cladding and never come back.

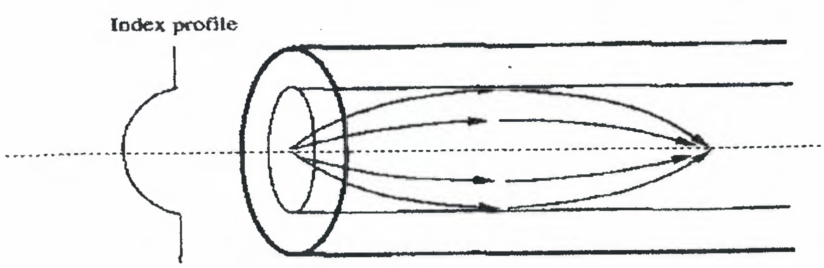

Figure 2.15 Ray propagation in the graded-index fiber.

The critical angle condition of equation (2.21) from geometrical optics gives the condition of propagatable rays. When θ_{crit} is large or n_2 is very close to n_1, the incident ray has to be closely parallel to the fiber axis. Therefore, it is more difficult to inject light into the fiber for propagation. To quantify the ease of coupling light into a fiber for propagation, a parameter called numerical aperture (NA) defined by

$$NA \overset{def}{=} \sqrt{n_1^2 - n_2^2} \tag{2.23}$$

has been used. Because n_1 is close to n_2 in optical fibers, $n_1^2 - n_2^2 = (n_1 + n_2)(n_1 - n_2) \approx 2n_2^2[(n_1 - n_2)/n_2]$. NA in equation (2.23) can thus be approximated by

$$NA \approx \left[\left(2n_2^2\right)\left(\frac{n_1 - n_2}{n_2}\right) \right]^{1/2} = n_2 (2\Delta)^{1/2} \tag{2.24}$$

is the ratio of the refractive index difference.

The physical meaning of NA can be seen as follows. An incident ray that can propagate in the fiber should be within the solid angle given by

$$\Omega = \frac{cone\_area}{d^2} = 2\pi[1 - \cos(\theta_{in})] = 4\pi \sin^2\left(\frac{\theta_{in}}{2}\right) \approx \pi \sin^2(\theta_{in}) \tag{2.25}$$

when $\theta_{in} << 1$. From figure 2.17 and equation (2.21),

$$\sin(\theta_{in}) = n_1 \cos(\theta_{crit}) = n_1 [1 - \sin^2(\theta_{crit})]^{1/2} = (n_1^2 - n_2^2)^{1/2} = NA \tag{2.26}$$

Therefore,

$$\Omega \approx \pi (n_1^2 - n_2^2) = \pi NA^2 \tag{2.27}$$

This shows that larger the NA, the larger the solid angle within which incident light can propagate.

(b) Fiber dispersion

If an optical pulse consists of components at different frequencies and different propagation modes, different propagation delays from these components will result in a broader pulse at the other end of the fiber. This phenomenon is known as Fiber Dispersion.

2.3 Light Detection

Light detection is a process that converts incident light into an electrical photocurrent. Alight detection device is used at the front end of every optical receiver to generate a photocurrent proportional to the incident light.

There are two main types of light detection devices: *photoconductors* and *photodiodes.* They are both semiconductor devices.

2.3.1 Photoconductors

There are two main types of photoconductors: *intrinsic* and *extrinsic.* An intrinsic photoconductor is an intrinsic semiconductor. Its conductivity increases when electrons are excited from the valence band to the conduction band. When there is incident light onto the photoconductor, electrons and holes are excited. As a result, the conductivity increases. On the other hand, an extrinsic photoconductor is a semiconductor with either N-type or P-type doping. Its conductivity increases when electrons (or holes) are excited from there N-type (or P-type) impurity levels. Because extrinsic photoconductors have free carriers, they have low resistance. This is undesirable from the thermal noise consideration. Therefore, the conductance of a photoconductor should be made as small as possible. From this consideration intrinsic photoconductors are better than extrinsic ones.

The current gain and frequency response of an intrinsic photoconductor are derived as follows. At an incident light power of P_{in} the amount of the carrier density increase (electrons and holes) is

$$\Delta n = \Delta p = \eta \frac{P_{in}}{hf} \tau_e \frac{1}{V} \tag{2.28}$$

Where τ_e is the relaxation time of the photoconductor and V is the volume of the absorption layer. If the photoconductor is in series with a bias resistance R_F as shown in Figure 2.16, the amount of increased current is

$$\Delta i = \frac{V_{DD}}{R_F + R_c - \Delta R_c} - \frac{V_{DD}}{R_F + R_c} \approx \frac{R_c^2 V_{DD}}{(R_c + R_F)^2} \left[\frac{q\eta}{hf} P_{in} (\mu_n + \mu_p) \tau_e \frac{1}{d^2} \right] \tag{2.29}$$

where μ_n and μ_p are the mobilities of electrons and holes, respectively.

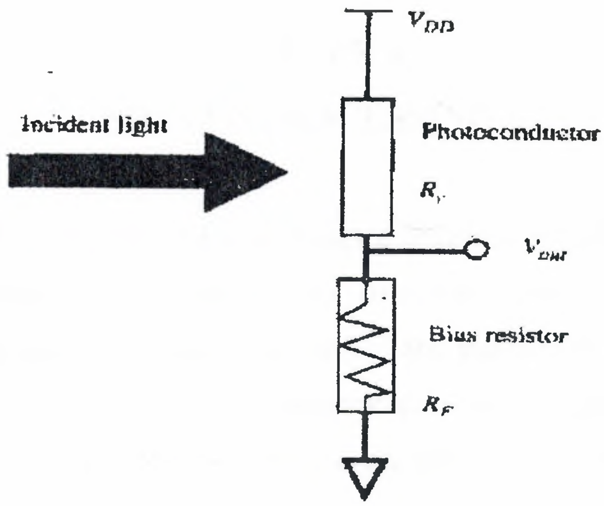

Figure 2.16 Circuit connection of a photoconductor for light detection.

2.3.2 Photodiodes

There are two types of photodiodes: PINs and APDs. A photon with sufficient energy can excite an electron-hole pair. If the pair is in the presence of a large electric field, the electron and hole will be separated and move quickly in opposite directions, resulting in a photocurrent.

Because one absorbed photon generates one EHP in PINs, the photocurrent is a linear function of the input optical power P_{in} :

$$I_{ph} = \eta \frac{q}{hf} P_{in} = \left(\frac{\eta\lambda}{1.24} \right) P_{in} \tag{2.30}$$

where η is quantum efficiency.

For APDs, because of the current gain from EHP multiplications, the generated photocurrent is

$$I_{ph} = M_{apd} \left(\frac{\eta\lambda}{1.24} \right) P_{in} \tag{2.31}$$

CHAPTER 3
SIGNAL PROCESSING

To provide high bit rate and long distance transmission, there have been many significant breakthroughs. For example, to overcome power loss due to fiber attenuation, optical amplifiers have been developed and installed in practical systems; to increase receiver sensitivity, coherent communication and external modulation have been used; and to overcome fiber dispersion, single frequency light sources and soliton transmission have been developed.

3.1 Direct Modulation

The light output must be modulated in either amplitude, frequency, or phase to transmit information. As illustrated in Figure 3.1, according to position where modulation is performed, light modulation can be classified as either direct or external modulation. With direct modulation, light is directly modulated inside a light a source. External modulation, on the other hand, uses a separate, external modulator.

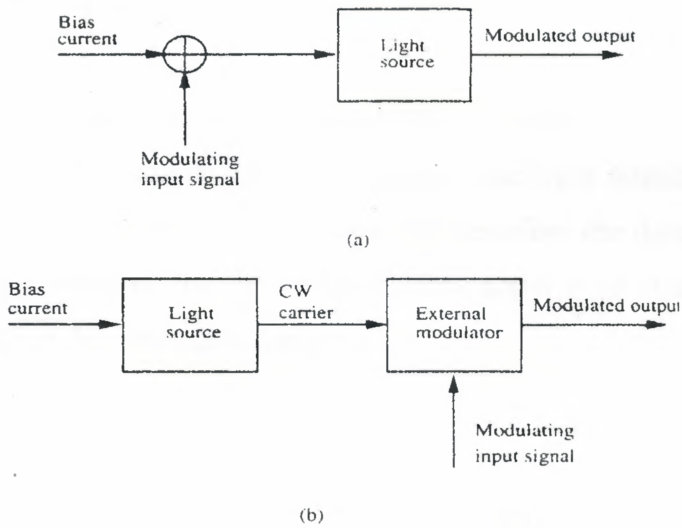

Figure 3.1 (a) Direct modulation (b) external modulation.

Direct modulation is used in most optical communication systems for its simpler implementation than external modulation. Because of physical limitations, the light output under direct modulation cannot respond to the input signal instantaneously.

Instead, there can be delays and oscillations when the input has large and rapid variations. Finding the light output under direct modulation requires using rate equations, which describe the dynamics of photons and carriers inside the active layer of the light source. The light output and the modulation bandwidth under certain bias condition can be determined from the rate equations.

Direct modulation has several undesirable effects such as frequency chirping and line width broadening. In frequency chirping, the spectrum of the light output is time varying because of refractive index modulation of the light source, and in line width broadening, the spectrum width is broader than that in the steady state. These effects can be the limiting factors in long distance and high-speed communications. To reduce these effects, external modulation can be used. In this case, the bias current to the light source is constant, and a separate modulator is used to modulate the continuous wave (CW) output of the light source.

3.1.1 Direct Modulation for LEDs

Direct modulation uses the input current to modulate the output light intensity. Because the light output of an LED is incoherent, the phase or frequency content of the output is not of interest. Therefore, the following analysis considers only the intensity or power of the modulated light.

(a) Rate equation for laser diodes and steady state solution

Whether in steady state or under modulation, the light intensity output of an LED can be derived from a single rate equation that describes the dynamic relationship between the carrier density and the input current. If the active layer of an LED is N-type of donor concentration N_D, the rate equation is

$$\frac{dN}{dt} = \frac{J}{qd} - B(NP - n_i^2) - \frac{N - N_D}{\tau_{nr}} \qquad (3.1)$$

Where N is the electron density, P is the hole density, J is the current density, d is the thickness of the active layer, τ_{nr} in the nonradiative time decay constant for the electrons, and B is the electron- hole recombination coefficient.

The above rate equation (3.1) can be understood from charge conservation. On the left hand, dN/dt is the net charge rate of the electron density of the p-n junction. On the right hand side are various sources for the rate change. The first term is the carrier-pumping rate due to external current density. The second term is the net radiative electron-hole recombination (EHR) rate. Finally the third term is the net nonradiative carrier recombination rate inside the active layer.

Because it is the EHR of the second term on the right hand side of equation (3.1) that contributes to the spontaneous emission, the optical power output is

$$P(t) = \eta_{ext} hfVB(NP - n_i^2) \tag{3.2}$$

Where η_{ext} is the external quantum efficiency that represents the fraction of photons coupled to the output, hf is the photon energy and V is the active layer volume.

In the steady state, $dN/dt = 0$. There fore,

$$\frac{J}{qd} = B(NP - n_i^2) + \frac{(N - N_D)}{\tau_{nr}}$$

According to charge neutrality,

$$NP - n_i^2 = N(N - N_D + \frac{n_i^2}{N_D}) - n_i^2 \approx N^2 - N_D N \tag{3.3}$$

It is convenient to define

$$\tau_{rr} \overset{def}{=} \frac{1}{BN} \tag{3.4}$$

Where τ_{rr} is called the radiative recombination time constant. It depends on the carrier density.

$$\frac{1}{\tau_r} \overset{def}{=} \frac{1}{\tau_{rr}} + \frac{1}{\tau_{nr}} = BN + \frac{1}{\tau_{nr}} \tag{3.5}$$

The rate equation in the steady reduces to

$$\frac{J}{qd} = \frac{N - N_D}{\tau_r} \tag{3.6}$$

From equation (3.3),

$$B(NP - n_i^2) = \frac{N - N_D}{\tau_{rr}} = \frac{\tau_r}{\tau_{rr}} \frac{J}{qd} \tag{3.7}$$

The ratio

$$\frac{\tau_r}{\tau_{rr}} = \frac{1/\tau_{rr}}{1/\tau_{rr} + 1/\tau_{nr}} \overset{def}{=} \eta_{\text{int}} \tag{3.8}$$

is called the internal quantum efficiency, meaning the ratio of the radiative recombination rate to the total recombination rate. From equations (3.2), (3.7), and (3.8), the optical power output can be related to the input current by

$$P = \eta_{ext} hfVB(NP - n_i^2) = \eta_{ext}\eta_{int} hfV \frac{J}{qd} = \eta \frac{hf}{q} I = \eta \frac{1.24}{\lambda(\mu m)} I \qquad (3.9)$$

Where

$$\eta = \eta_{ext}\eta_{int}$$

is the total quantum efficiency.

(b) Digital signal modulation

On-off keying (OOK) is a common signaling scheme in digital communications. In this scheme a pulse is send to convey a bit "1" and nothing is sent to transmit bit "0" for one bit interval.

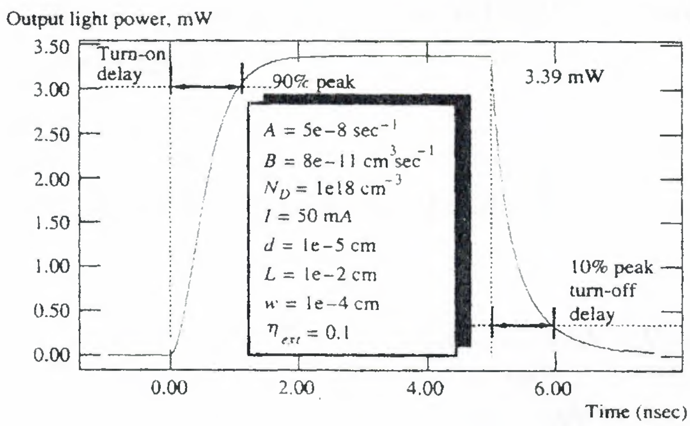

Figure 3.2 A typical pulse output of an LED.

To generate a pulse, the bias current is first increased from a low value I_1 to a high value I_2 and then reduced back to I_1 after certain duration. In general, when the input current to an LED goes up and down, there is a delay before the output light intensity follows. This is illustrated in figure 3.2. From equation (3.2), the basic reason for this delay is it takes time to build up the carrier density product NP from external current injection. Because these turn-on and turn-off can limit the final achievable transmission bit rate.

To derive these delays we have to solve the rate equation (3.1) for a step two input: one goes from low to high, and one goes from high to low. Consider the first case, in which a step input goes from low J_1 to high J_2. The initial condition at time zero for J is

$$J(0) = J_1 = qd\left[B(N_1P_1 - n_i^2) + \frac{N_1 - N_D}{\tau_{nr}}\right]$$

Where

$$NP\Big|_{t=0} \overset{def}{=} N_1P_1$$

Similarly, the steady state condition for $J = J_2$ at $t = \infty$

$$J(\infty) = J_2 = qd\left[B(N_2P_2 - n_i^2) + \frac{N_2 - N_D}{\tau_{nr}}\right]$$

Where

$$NP\Big|_{t=\infty} \overset{def}{=} N_2P_2$$

From charge neutrality, the number of electrons and holes generated are always the same. It is thus possible to define

$$\Delta N(t) \overset{def}{=} N(t) - N_1 = P(t) - P_1 \tag{3.10}$$

As the carrier density increases from the initial value at t = 0. From equations (3.1) and (3.10), for t > 0,

$$\begin{aligned} \frac{d[\Delta N(t)]}{dt} &= \frac{J_2}{qd} - B(N_1 + \Delta N)(P_1 + \Delta N) + Bn_i^2 - \frac{1}{\tau_{nr}}(\Delta N + N_1 - N_D) \\ &= \frac{J_2 - J_1}{qd} - B(N_1 + P_1 + \Delta N)\Delta N - \frac{\Delta N}{\tau_{nr}} = \frac{J_2 - J_1}{qd} - \frac{\Delta N}{\tau_{step}} \end{aligned} \tag{3.11}$$

Where

$$\frac{1}{\tau_{step}} \overset{def}{=} B(N_1 + P_1 + \Delta N) + \frac{1}{\tau_{nr}} \tag{3.12}$$

(c) *Analog signal modulation*

In analog communications, the information signal is added to a DC bias to form the total current input to a light source. From the nonlinear rate equation at a large current the signal component needs to be small compared to the DC bias to maintain high end-to-end waveform linearity. Therefore, analog communication can be considered as small signal modulation.

Previously we have analyzed turn-on and turn-off delays with pulse modulation. In analog communications, on the other hand, the modulation bandwidth for the small signal input is of interest. The time-frequency duality suggests that the delay and bandwidth analysis yield LED in pulse modulation, the larger the 3 dB bandwidth in small signal modulation.

To proceed, consider a single tone input frequency ω :

$$J(t) = J_0[1 + k_m \cos(\omega t)] \tag{3.13}$$

Where J_0 is the bias current density and k_m is the amplitude modulation index. Let N_1 be the carrier density at J_0, and let

$$N(t) \overset{def}{=} N_1 + \Delta N \cos(\omega t - \phi) \tag{3.14}$$

Where the second term is small signal term due to the single tone input. From the above two equations, the rate equation (3.1) reduces to

$$-\Delta N \omega \sin(\omega t - \phi) = \frac{J_0}{qd} k_m \cos(\omega t) - \frac{1}{\tau_{ac}} \Delta N \cos(\omega t - \phi) \tag{3.15}$$

Where

$$\frac{1}{\tau_{ac}} \overset{def}{=} B(N_1 + P_1) + \frac{1}{\tau_{nr}} \tag{3.16}$$

The solution to equation (3.15) is

$$\Delta N = \frac{J_0}{qd} k_m \tau_{ac} \frac{1}{(1 + \omega^2 \tau_{ac}^2)^{1/2}} \tag{3.17}$$

And

$$\phi = \cos^{-1}\left(\frac{1}{(1 + \omega^2 \tau_{ac}^2)^{1/2}}\right) \tag{3.18}$$

From equation (3.17), the output power is

$$P(t) = \eta h f \frac{I_0}{q}\left[1 + \frac{k_m}{(1 + \omega^2 \tau_{ac}^2)^{1/2}} \cos(\omega t - \phi)\right] \tag{3.19}$$

Therefore, the signal part is

$$\Delta P(t) = \eta h f \frac{I_0}{q} k_m \frac{1}{(1 + \omega^2 \tau_{ac}^2)^{1/2}} \cos(\omega t - \phi) \tag{3.20}$$

From equation (3.20), the 3 dB optical signal bandwidth is

$$B_{opt,3dB} = \frac{\sqrt{3}}{2\pi\tau_{ac}} = \frac{0.28}{\tau_{ac}} \tag{3.21}$$

Based on the final photocurrent signal on the receiver side, the 3 dB electrical signal bandwidth is

$$B_{elec,3dB} = \frac{1}{2\pi\tau_{ac}} = \frac{0.16}{\tau_{ac}} \tag{3.22}$$

The results show that the 3 dB frequency response is determined solely by the time constant τ_{ac}.

3.1.2 Rate Equation for Laser Diodes

Similar to LEDs, laser diodes can be directly modulated by varying the bias current. Because carriers interact with photons in the laser generation process, there are separate rate equations to describe the dynamics of carriers and photons. These equations can quantify and explain various important phenomena such as turn-on delay, relaxation oscillation, chirping, and linewidth broadening.

(a) Laser rate equation for carrier density

Based on the laser generation dynamics, the rate equation for the carrier density can be written. Similar to the rate equation of LEDs, it can be expressed as

$$\frac{\partial N}{\partial t} = \frac{J(t)}{qd} - R(N, N_{ph}) = \frac{J(t)}{qd} - \frac{N}{\tau_e(N)} - \Gamma v_g g(N) N_{ph} \tag{3.23}$$

Where $R(N, N_{ph})$ is the total carrier recombination rate and is a function of the carrier density N and photon density N_{ph}. Specifically, for a lightly doped active layer, $R(N, N_{ph})$ can be expressed as

$$\begin{aligned} R(N, N_{ph}) &= AN + BN^2 + CN^3 + \Gamma v_g g(N) N_{ph} \\ &= \frac{N}{\tau_e(N)} + \Gamma v_g g(N) N_{ph} \end{aligned} \tag{3.24}$$

Where

- $AN = N/\tau_{nr}$ is the nonradiative recombination rate.
- BN^2 is the radiative spontaneous emission rate.

- CN^3 is called the Auger recombination rate.
- $1/\tau_e(N) \overset{def}{=} A + BN + CN^2$ is the effective recombination rate.
- $\Gamma v_g g(N) N_{ph}$ is the stimulated emission rate.

In the stimulated emission term given above, v_g is the group velocity, Γ is the cavity confinement factor, and

$$g(N) = a(N - N_0) \tag{3.25}$$

is the optical gain, with a being the gain constant.

To include the gain suppression effect, the gain constant can be expressed as

$$a = \frac{a_0}{1 + \eta_g N_{ph}} \approx a_0(1 - \eta_g N_{ph}) \tag{3.26}$$

Where the parameter η_g is called the gain suppression coefficient. The gain constant a decreases as the optical intensity increases.

(b) Rate equation for photon density

Similar to the carrier density rate equation, the photon density rate equation can be formulated according to the sources of its rate change. Because the photon density in a cavity is determined by the stimulated emission rate, spontaneous emission rate, and cavity loss, the following rate equation holds:

$$\frac{dN_{ph}}{dt} = \Gamma v_g g(N) N_{ph} - \frac{N_{ph}}{\tau_{ph}} + \beta_{sp} BN^2 \tag{3.27}$$

Where the parameter β_{sp} represents the percentage of the spontaneous emission that happens to be coherent with and in phase to that of the stimulated emission, and τ_{ph} is the photon decay constant related to the total cavity loss α_{tot}.

The photon decay constant τ_{ph} can be expressed in terms of the total cavity loss α_{tot}. Because the photon decay over one trip in the cavity has a factor of $e^{-\alpha_{tot} L}$ and takes a time of L / v_g,

$$e^{-L/v_g \tau_{ph}} = e^{-\alpha_{tot} L}$$

Therefore,

$$\frac{1}{\tau_{ph}} = v_g \alpha_{tot} \tag{3.28}$$

3.1.3 Pulse Input Response for Laser Diodes

The rate equations given earlier can be used to study the pulse input response of laser diodes for digital transmission. A typical pulse input response from numerically solving equations (3.23) and (3.27) is shown in figure 3.3. Clearly laser diodes respond quite differently from LEDs. On the rising edge, a turn-on delay is followed by an oscillation. On the falling edge, there is a sharp drop or negligible turn-off delay.

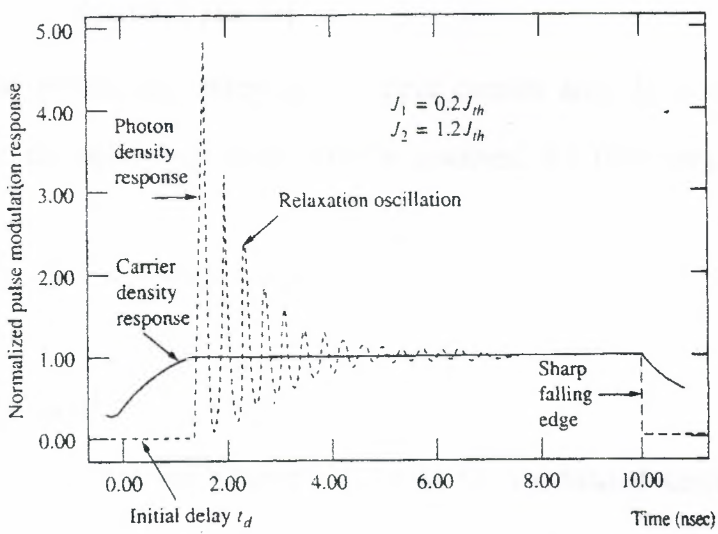

Figure 3.3 Pulse modulation response.

The behavior illustrated in figure 3.3 can be qualitatively understood as follows. First, careful examination of equation (3.27) shows that the rate of increase dN_{ph}/dt is essentially zero when N_{ph} is small. It will not become significantly positive until the net gain is positive:

$$\Gamma v_g g(N) - \frac{1}{\tau_{ph}} > 0$$

Therefore, when the laser diode is turned on, the photon density due to stimulated emission will stay essentially zero until N reaches its threshold N_{th}. This can be clearly seen from figure 3.3, where the normalized carrier density N/N_{th} is shown. The delay for the carrier density to reach N_{th} is called the turn-on delay.

3.1.4 Small Signal Response for Laser Diodes

Similar to LEDs, small signal analysis of laser diodes is useful for analog communications. A similar analysis can help explain the modulation bandwidth of a single tone current input.

For convenience, a single tone current input at the circular frequency ω_m is expressed by

$$J(t) = J_0\left[1 + k_{m,j}\exp(j\omega_m t)\right] \tag{3.29}$$

Where $k_{m,j}$ is the modulation index for the input current and J_0 is the bias current density. Similarly the following forms can be assumed for the carrier and photons densities:

$$N(t) = N_{n,0}\left[1 + k_{m,n}\exp(j\omega_m t)\right] \tag{3.30}$$

And

$$N_{ph}(t) = N_{ph,0}\left[1 + k_{m,ph}\exp(j\omega_m t)\right] \tag{3.31}$$

Where $k_{m,n}$ and $k_{m,ph}$ are modulation indices of the modulated carrier and photon densities, respectively. In general, they can be complex to represent possible phase shifts with respect to $J(t)$.

3.2 External Modulation

To avoid undesirable chirping and mode partitioning in direct modulation, external modulation is a good alternative. A basic external modulator is general consists of an optical waveguide in which the incident light passes through and the refractive index of the medium is modulated by the information bearing signal. According to the way the refractive index is modulated, there are two primary types of modulation: electro-optic (EO) and acousto-optic (AO), where the refractive index is modulated by either a voltage input or an acousto wave, respectively.

3.2.1 Principles of External Modulation

As described earlier, there are two primary types of external modulators: EO and AO modulators. In EO modulators, a constant refractive index change is introduced by an external voltage signal. In AO modulators, a periodic change of the refractive index in introduced by an acoustic wave.

(a) Electro-optic modulator

An EO modulator can be as simple as an optical channel or waveguide through which incident light passes. For many media used in EO modulators, the incident light "sees" different refractive indices at different polarizations.

Let β_a and β_b be two propagation constants, both of which can be modulated by an external voltage signal. The electric field at the waveguide output can be written as

$$\overline{E}_{out} = A_a e^{-j\beta_a L}\hat{s}_a + A_b e^{-j\beta_b L}\hat{s}_b \tag{3.32}$$

(b) Acousto-optic modulator

In AO modulators, the periodic modulation of the refractive index forms a grating. As a result, the incident light is diffracted. Thus the grating can be controlled to modulate either the through light or the diffracted light.

In general, the grating structure created in AO modulation is similar to the periodic Bragg reflector structure of DFB lasers. As a result, there are two waves inside the modulator, one in the same direction as the incident wave and the other diffracted by the grating. These two waves interact as they propagate along the modulator and are similarly governed by the coupled mode equations. As a result, by modulating the coupling constant through the input acoustic wave, the light out put can be modulated.

3.2.2 Wave Propagation in Anisotropic Media

A medium that responds differently to the different polarizations of incident light is called anisotropic. A medium that has the same response to different polarizations is called isotropic. When the medium of an external modulator is modulated, it can turn from isotropic to anisotropic. In fact, to achieve large modulation depth and power efficiency, an external modulator should have a large response to either the EO or AO effect.

(a) Anisotropic materials

The response of a medium to incident light is characterized by its permittivity. For an anisotropic medium, its response depends on the polarization of the light. As a result, its permittivity is no longer a constant. Instead, it is a tensor or a three by three matrix through which the electric flux density $\overline{D}$ is related to the electric field $\overline{E}$ by

$$\overline{D} = \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} = \overline{\varepsilon E} \tag{3.33}$$

From basic physics, it can be verified that $\varepsilon_{ij} = \varepsilon_{ji}^{*}$. Therefore, it can be transformed to a diagonal matrix.

$$\overline{D} = \begin{bmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{bmatrix} \overline{E} \tag{3.34}$$

The axes that result in the above equation are called principle axes.

(b) Normal surfaces

To find the two permissible plane waves propagating in a given direction requires using Fresnel equation (3.35) to solve the propagation constants and using equation (3.36) to find the polarizations. These results, allow one to find the birefringence and design the external modulator.

$$\frac{s_x^2}{n^2 - n_1^2} + \frac{s_y^2}{n^2 - n_2^2} + \frac{s_z^2}{n^2 - n_3^2} = \frac{1}{n^2} \tag{3.35}$$

$$\overline{E} \propto \begin{bmatrix} \dfrac{s_x}{n^2 - n_1^2} \\ \dfrac{s_y}{n^2 - n_2^2} \\ \dfrac{s_z}{n^2 - n_3^2} \end{bmatrix} \tag{3.36}$$

$$\begin{aligned} &\left(n^2 - n_2^2\right)\left(n^2 - n_3^2\right)n^2 s_x^2 + \left(n^2 - n_3^2\right)\left(n^2 - n_1^2\right)n^2 s_y^2 \\ &+ \left(n^2 - n_1^2\right)\left(n^2 - n_2^3\right)n^2 s_z^2 = \left(n^2 - n_1^2\right)\left(n^2 - n_2^2\right)\left(n^2 - n_3^2\right) \end{aligned} \tag{3.37}$$

The first method is the use of the so-called normal surface, which is defined as the set of all points (ns_x, ns_y, ns_z) with n satisfying equation (3.35). Therefore, for a given propagation direction $\hat{s}$, n_a and n_b can be found by generating a ray in the same direction $\hat{s}$ from the origin. The distance of the intersection of the ray with the normal surface gives the corresponding n value.

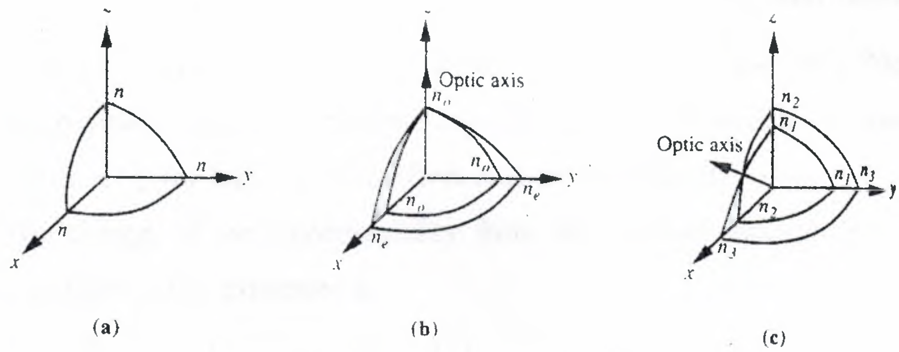

Figure 3.4 Normal surfaces (a) isotropic (b) uniaxial (c) biaxial crystal.

The normal surface can be constructed as follows. For the special case that $s_z = 0$ or $\overline{\beta}$ is in the x-y plane, equation (3.37) reduces to

$$n^2 = n_3^2 \tag{3.38}$$

And

$$\frac{1}{n^2} = \frac{s_x^2}{n_2^2} + \frac{s_y^2}{n_1^2} \tag{3.39}$$

Equations (3.38) and (3.39) allow the two curves to be drawn on the x-y plane as illustrated in figure 3.4. Note that in figures 3.4b and 3.4c, there is a special axis called the optic axis, which is defined to be the propagation direction in which the two n values are the same.

3.2.3 Electro-Optic Modulation

This kind of modulation deals with the basic principles of EO modulation.

(a) Modulation of refractive indices

When an external voltage is applied to an EO crystal, the electric field present in the crystal can modulate its refractive index tensor. The electro-optic effect can be characterized by the change of the impermeability tensor:

$$\Delta\eta_{ij}(\bar{E}) = \sum_{k} r_{(ij)k} E_k + \sum_{kl} s_{(ij)(kl)} E_k E_l \tag{3.40}$$

The linear coefficients $r_{(ij)k}$ are called the Pockels coefficients, and the quadratic coefficients $s_{(ij)(kl)}$ are called the Kerr coefficients. In general, the Kerr effect is insignificant unless the Pockels coefficients are all zero. Crystals that have zero Pockels effect have a centrosymmetric crystal structure. For most optical EO modulators that are not centrosymmetric, only the linear Pockels effect must be considered.

Given the change of the impermeability from the external electric field, the index ellipsoid equation can be expressed as

$$\begin{aligned} &\left(\frac{1}{n_1^2} + \Delta\eta_{11}\right)x^2 + \left(\frac{1}{n_2^2} + \Delta\eta_{22}\right)y^2 + \left(\frac{1}{n_3^2} + \Delta\eta_{33}\right)z^2 \\ &+ 2\Delta\eta_{23}yz + 2\Delta\eta_{31}zx + 2\Delta\eta_{12}xy = \frac{1}{n^2} \end{aligned} \tag{3.41}$$

In general, from the symmetry of $\Delta\eta_{ij} = \Delta\eta_{ji}$, the new index ellipsoid can be transformed into the canonical form given by following equation

$$\frac{x^2}{n_1^2} + \frac{y^2}{n_2^2} + \frac{z^2}{n_3^2} = \frac{1}{n^2} \tag{3.42}$$

This means that the principle axis of the crystal can rotate in the presence of an external field.

(b) Longitudinal and transverse modulators

According to the relative direction of the bias electric field and the light wave propagation, an EO modulator can be classified into two categories: *transversal* and *longitudinal*. As illustrated in figure 3.5, the bias electric field is perpendicular to the wave propagation direction in transverse modulation; and in longitudinal modulation, the bias electric field is in the same direction as the wave propagation.

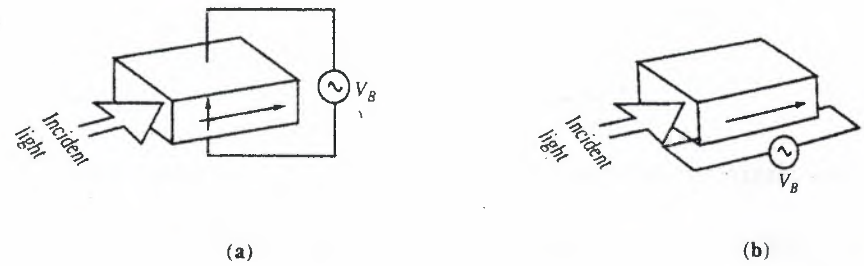

Figure 3.5 (a) Transverse (b) Longitudinal modulators.

For a given birefringence B, the two polarizations over a propagation distance L will have a phase difference $\Delta\phi$ given by

$$\Delta\phi = (\beta_a - \beta_b)L = (n_a - n_b)\frac{\omega}{c}L = \frac{\omega}{c}\mathrm{B}L \qquad (3.43)$$

For longitudinal modulators, from equation (3.43) the phase change of a given permissible polarization is linearly proportional to the applied electric field E_z. Therefore, if the lightwave propagates in the z direction,

$$\Delta\phi = \frac{2\pi L}{\lambda}\Delta n = \frac{2\pi r}{\lambda}LE_z = \frac{2\pi r}{\lambda}V_B \qquad (3.44)$$

Where V_B is the external bias voltage and r is the EO coefficient according to the wave polarization.

3.2.4 Acousto-Optic Modulation

Similar to EO modulation, AO modulation uses sound waves to modulate the refractive index of the medium. The difference is that sound waves have periodic compression and refraction patterns that result in periodic variations of the medium's refractive index. This periodic variation of the refractive index acts as a phase grating that diffracts part or all of the incident light.

(a) Electro-optic coefficients

Similar to electro-optic coefficients, there are a so-called elasto-optic coefficient that quantifies the change of the permeability tensor under an acoustic wave.

For an acoustic wave $\overline{U}(\Omega t - \overline{K}.\overline{r})$ in a medium with propagation constant $\overline{K}$, the strain tensor is defined as

$$S_{(ij)} = \frac{\partial U_i}{\partial r_j} \tag{3.45}$$

Because the acoustic wave $\overline{U}$ is in units of meters, the strain tensor $S_{(ij)}$ is dimensionless. When the strain tensor due to an acoustic wave is known, the change of the permeability is described by

$$\Delta\eta_i = \sum_{j=1}^{6} p_{ij} S_j \tag{3.46}$$

The interaction of the incident light and acoustic wave depends on their polarizations because the AO medium can become anisotropic.

3.3 Coherent Detection

Coherent detection is simple in implementation; it cannot detect the phase and frequency of the received signal. In other words, it can detect only amplitude-modulated signals, when phase modulation or frequency modulation is desirable, such as when intensity noise is strong, coherent detection becomes a better choice. This is familiar from radio communications, where FM is much better than AM in transmission quality.

Coherent detection is also important in applications such as wavelength division multiplexing (WDM), where multiple channels are transmitted at the same time. Although passive tunable filters can be used to avoid coherent detection, a larger channel separation is necessary because of limited filter resolution.

Optical amplifiers developed over the last few years provide another attractive alternative. For example, Erbium-doped fiber amplifiers can be easily inserted into regular optical fibers for a power gain of 20-30 dB and at a pumping efficiency of 5-10 dB/mW. One disadvantage is that the amplifier also introduces noise because of amplified spontaneous emission (ASE).

3.3.1 Signal and Noise Formulation

After the two light signals are mixed by the hybrid, there are two main configurations used in photodetection: *single detection* and *balanced detection.* As illustrated in figure 3.6, single detection uses only one photodiode. This is the same as in incoherent detection. In this case, one of the hybrid's output is not used and can be used for carrier recovery. Balanced detection feeds the two outputs to two photodiodes whose current ouputs are subtracted. One major advantage of balanced detection is that it cancels the relative intensity noise from local oscillator.

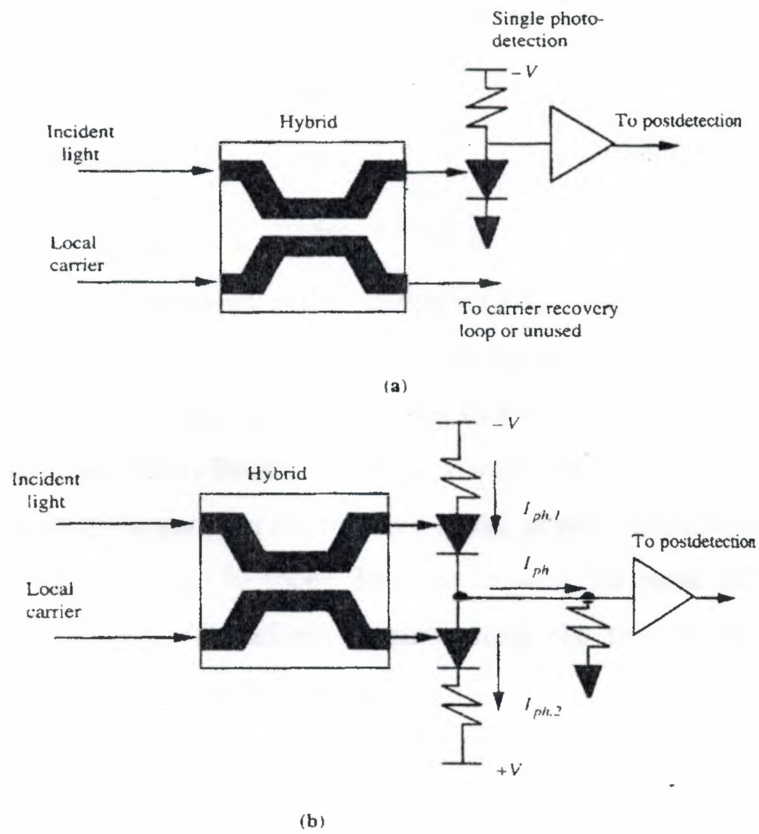

Figure 3.6 (a) Single detection versus (b) balanced detection.

Without loss of generality, consider the use of a 180^0 hybrid. The two outputs from the hybrid can thus be expressed as

$$E_{o1} = \frac{1}{\sqrt{2}}(E_{inc} + E_{loc})$$

$$E_{o2} = \frac{1}{\sqrt{2}}(E_{inc} - E_{loc})$$

After photodetection

$$I_{ph,1} = \frac{1}{2}\Re\left\{P_{inc} + P_{loc} + 2\sqrt{P_{inc}P_{loc}}\cos[[\omega_{inc} - \omega_{loc}]t + \phi(t)]\right\} \qquad (3.47)$$

$$I_{ph,2} = \frac{1}{2}\Re\left\{P_{inc} + P_{loc} - 2\sqrt{P_{inc}P_{loc}}\cos[[\omega_{inc} - \omega_{loc}]t + \phi(t)]\right\} \qquad (3.48)$$

Where P_{inc} is incident light power and P_{loc} is the local carrier power. With balanced detection, the difference between the photocurrent is

$$I_{ph} = I_{ph,1} - I_{ph,2} = 2\Re\sqrt{P_{inc}P_{loc}}\cos[(\omega_{inc} - \omega_{loc})t + \phi(t)] \tag{3.49}$$

Base on the balanced detection, when homodyne detection is used or $\omega_s = \omega_{loc}$

$$I_{ph,\mathrm{hom\,o}}(t) = 2\Re\sqrt{P_{inc}P_{loc}}\cos[\phi(t)] \tag{3.50}$$

Similarly, when heterodyne detection is used,

$$I_{ph,hetero}(t) = 2\Re\sqrt{P_{inc}P_{loc}}\cos[\omega_{IF}t + \phi(t)] \tag{3.51}$$

The current outputs given in the equations (3.48) and (3.49) contain only signal terms. In practice, there are additional noise terms that need to be added. In addition to receiver noise, two important noise terms are the shot noise from photodetection and the RIN from the local oscillator. Because the RIN power is proportional to the local optical power, which is much larger than the received signal power, the RIN can greatly affect detection performance. When balanced detection is used, the same RIN occurs at the two photodiodes outputs. Therefore, by subtracting the two current outputs from balanced detection, the RIN can be cancelled.

3.3.2 On-Off Keying

As figure 3.7 shows, OOK transmits a pulse when the input bit is one, and transmits no pulse when the input bit is zero. This is the simplest signaling method because it can be implemented by direct modulation.

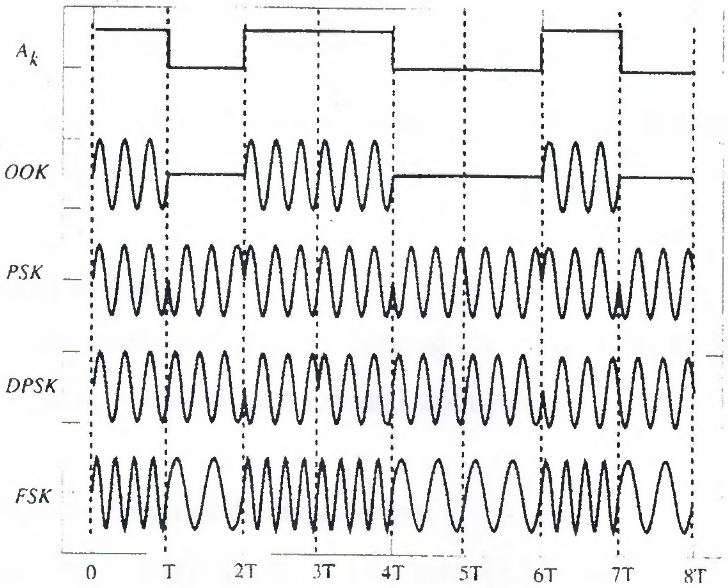

Figure 3.7 Various digital signaling formats in coherent communications.

(a) Signal representation and demodulation

When OOK signaling used, the signal power, $P_{inc}(t)$, in equations (3.50) and (3.51) can be expressed as

$$P_{inc}(t) = \sum_k A_k p(t - kT) \tag{3.52}$$

If homodyne detection is used, the photocurrent is a base band NRZ pulse train, and the transmitted data can be recovered by matched filtering and simple threshold detection. On the other hand, if heterodyne detection is used, the incoherent post-detection method can be used.

In case of coherent detection, a squarer followed by a low-pass filter is commonly used. In this case,

$$[A\cos(\omega_{IF}t)]^2 = \frac{1}{2}A^2 + \frac{1}{2}A^2\cos(2\omega_{IF}t)$$

Where

$$A = 2\Re\sqrt{P_{inc}P_{loc}}$$

In the case of coherent post-detection, an IF recovery loop is needed to generate the same carrier $\cos(\omega_{1F}t)$ of the IF filter output.

$$[A\cos(\omega_{IF}t)]^2 = \frac{1}{2}A + \frac{1}{2}A\cos(2\omega_{IF}t)$$

Where the first term is the term of interest, and the second term will be rejected be the low-pass filtering.

3.3.3 Phase-Shift Keying

PSK transmits information by modulating the phase term in equation (3.50) or (3.51), which can be achieved through external modulation.

(a) Signal representation and demodulation

The phase term $\phi(t)$ given in equations (3.50) and (3.51) in PSK can be generally expressed as

$$\phi(t) = \sum_k A_k p(t - kT) + \theta \tag{3.53}$$

Where θ is a constant term and A_k can be either 0 or π for binary PSK.

In homodyne detection, similar to OOK, the photocurrent output can be matched filtered and detected with a local carrier, $\cos(\omega_{loc}t + \theta)$. Similarly, in heterodyne detection, the IF filter output can be demodulated by an IF carrier, $\cos(\omega_{IF}t + \theta)$.

Because both the envelope and frequency of a PSK signal are constant, incoherent post-detection such as envelope detection cannot be used.

3.3.4 Differential Phase-Shift Keying

In PSK heterodyne detection, although there is no need to recover the original phase and frequency of the transmitter carrier in the optical loop, it must still be done in IF loop. In order to simplify the carrier recovery, differential PSK (DPSK) can be used.

Different from PSK, information in DPSK is transmitted through the phase difference of adjacent symbols. Specifically, if the transmitted bit is "0" A_k in equation (3.53) is the same as the previous value A_{k-1}. If the transmitted bit is "1" there is a phase change of π between A_k and A_{k-1}.

(a) DPSK demodulation

With differential phase modulation, the absolute phase θ is not of concern. Instead, only the phase changes over two adjacent symbols must be detected.

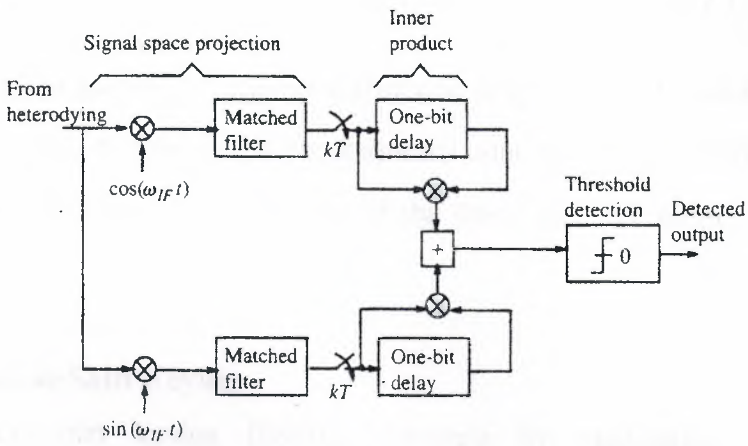

Figure 3.8 Block diagram of DPSK detection.

A block diagram of DPSK demodulation is illustrated in Figure 3.8. As shown it consists of three stages (1) signal space projection (2) inner product calculation and (3) threshold detection.

3.3.5 Frequency-Shift Keying

In FSK, the frequency of the transmitted optical carrier is not a constant but is modulated according to the transmitted data. It is a common modulation technique in coherent communication because of its incoherent post-detection ability. Specifically, instead of generating an IF carrier, two pass-band filters followed by envelope detection can be used in post-detection. When pulse noise is important, this incoherent post-detection becomes even more critical to maintaining good detection performance.

(a) Signal representation and demodulation

For binary FSK, the angular frequency of the incident light can be expressed as

$$\omega_{inc} = \omega_0 \pm \frac{\Delta\omega_d}{2} = \omega_0 + \frac{\Delta\omega_d}{2}\sum_k A_k p(t-kT) \quad (3.54)$$

Where $\Delta\omega_d$ is the frequency separation of the two frequencies and A_k is either 1 or −1, according to the transmitted data. Because the frequency is not constant, only heterodyne detection is used in coherent detection, and the photocurrent in equation (3.51) can be expressed as

$$I_{ph}(t) = 2\Re\sqrt{P_{inc}P_{loc}}\cos\left[\omega_{IF}t + (\Delta\omega_d/2)t \times \sum_k A_k p(t-kT)\right] \qquad (3.55)$$

To detect the above FSK current signal, one simple way is to use two band-pass filters centered at $\omega_{IF} \pm \Delta\omega_d/2$. If the frequency band of the two band-pass filters is smaller than the frequency separation, one of the filters has high output and the other has a low output.

3.3.6 Polarization-Shift Keying

Polarization-shift keying (PolSK) transmits by modulating the state of polarization (SOP) of the light carrier. Thus, for accurate detection, it is important to maintain the SOP transmission.

PolSK can be attractive when the multilevel signaling is used. PolSK has three-dimensional signal space, which allows a more compact constellation design at a given number of signal points and a given signal separation.

3.4 Incoherent Detection

Photodetection converts incident light into photocurrent, which is proportional to the power of incident light and carries no information about the phase of the incident light. As a result, this detection is called incoherent detection or direct detection. This is in contrast to coherent detection, which detects both power and phase of the incident light.

Because incoherent detection only detects the power of the incident light, it is used primarily for intensity or amplitude modulation transmission. Coherent detection is necessary when phase or frequency modulation is used.

3.4.1 Analog Signal Detection

A receiver block diagram for analog communications is shown in Figure 3.9. In addition to the photocurrent generated, noise from the front-end amplifier such as thermal noise and transistor junction noise is added. Because there is a dc bias in analog communications, ac coupling is used to reject the dc component.

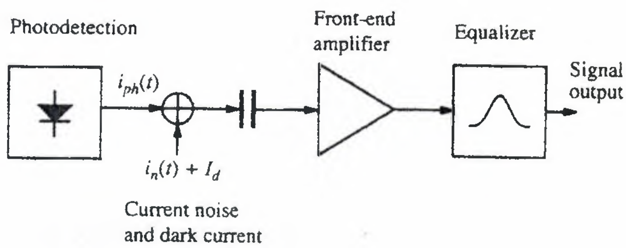

Figure 3.9 Block diagram of an analog receiver.

In analog communications, the amplitude-modulated signal at the output of the photodiode can be expressed as

$$i_{ph}(t) = I_0[1 + k_m m(t)] \tag{3.55}$$

Where I_0 is the dc current, $m(t)$ is the message signal, and k_m is the amplitude modulation index. The SNR is given by

$$SNR = \frac{k_m^2 I_0^2 \overline{m(t)^2}}{\sigma^2_{n,out}} \tag{3.56}$$

Where $\sigma^2_{n,out}$ is total noise power. Specifically, for a given receiver equalizer $H(\omega)$.

$$\sigma^2_{n,out} = \int \left[q(I_0 M_{apd} + I_d M^2_{apd}) F_{apd} + \frac{RIN}{2} I_0^2 + S_a \right] |H(\omega)|^2 \frac{d\omega}{2\pi} \tag{3.57}$$

Where I_d is the dark current and M_{apd} is the current gain of the photodiode, F_{apd} is the excess noise factor of the photodiode, RIN is the relative intensity noise factor, and S_a is the PSD of the equivalent front-end amplifier input noise.

3.4.2 Binary Digital Signal Detection

A basic block diagram for digital signal detection is shown in Figure 3.10. As illustrated, it consists of a photodetector. Affront-end amplifier, an equalizer, a slicer (threshold detector), and a bit timing recovery circuit.

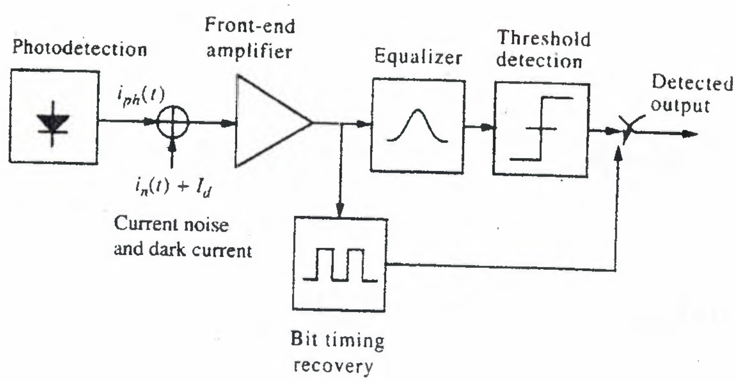

Figure 3.10 Block diagram of a typical digital receiver.

The photodetector converts incident light into photocurrent. The front-end amplifier amplifies the photocurrent with minimal added noise. The equalizer is used in combination with the front-end amplifier to achieve a certain receiver transfer function. For example, it can be used to compensate the low-pass response of the front-end amplifier, and it can also be designed to reduce ISI and maximize the SNR. The slicer performs threshold detection. In the case of binary transmission, it detects the equalized output as either high or low. To regenerate the original bit stream, the bit timing recovery circuit recovers the original transmitter clock from the received signal.

3.4.3 Intersymbol Interference and Noise Formulation

If the front-end amplifier and equalizer have a combined transfer function, the output of equalizer is

$$y_{out}(t) = i_{tot}(t) \otimes h(t)$$

The signal component of the output signal is

$$y_s(t) = i_{ph}(t) \otimes h(t) = \sum_k A_k p(t - kT_0) \otimes h(t) \overset{def}{=} \sum_k A_k p_{out}(t - kT_0) \quad (3.58)$$

To detect the transmitted amplitude A_k, y_{out} at the equalizer output is sampled at the bit rate and compared with a threshold. From equation (4.58)

$$y_{out,k} \overset{def}{=} y_{out}(kT_0 + \tau) = A_k p_{out}[0] + ISI_k + y_{n,k} \quad (3.59)$$

Where the constant dark current term has been dropped for its irrelevance.

$$ISI_k \overset{def}{=} \sum_{k' \neq k} A_{k'} p_{out}[k - k'] \quad (3.60)$$

$$y_{n,k} \overset{def}{=} y_{n,out}(kT_0 + \tau)$$

The error detection probability is

$$P_E = p_0 P(y_{out,k} > y_{th} \mid A_k = A_L) + p_1 P(y_{out,k} < y_{th} \mid A_k = A_H) \quad (3.61)$$

The two-sided PSD at the photodiode output is

$$S_{n,out}(\omega,t) = q\left[i_{ph}(t)M_{apd} + I_d M_{apd}^2\right]F_{apd} + \frac{RIN}{2} i_{ph}^2 + S_{MPN}(\omega) \quad (3.62)$$

The total noise power at the equalizer output is

$$\sigma_{n,out}^2(t) = S_{n,ph}(0,t) \otimes h(t)^2 + \int S_a(\omega)\left|H(\omega)\right|^2 \frac{d\omega}{2\pi} \quad (3.63)$$

3.4.4 Bit Error Rate Neglecting Intersymbol Interference

To understand BER evaluation, let us first consider the simple case where ISI is neglected. A power penalty can be subsequently added when ISI is included.

Using equation (3.59) and dropping the ISI term. We find the sampled output at the equalizer output is

$$y_{out,k} = A_k p_{out}[0] + y_{n,k}$$

From Gaussian approximation and equation (3.61), BER is

$$P_E = p_1 Q\left(\frac{y_H - y_{th}}{\sigma_H}\right) + p_0 Q\left(\frac{y_{th} - y_L}{\sigma_L}\right) \quad (3.64)$$

Where $y_H = A_H p_{out}[0]$, $y_L = A_L p_{out}[0]$, y_{th} is threshold, and σ_H^2, σ_L^2 are the corresponding noise power at $A_k = A_H$ and A_L, respectively.

3.4.5 Bit Error Rate Including Intersymbol Interference

The BER is calculated by neglecting the ISI. Now the effective ISI is included and a more rigorous BER expression is derived.

Because both ISI and noise are signal dependent, the detection error probability depends on input sequence. For accurate error probability calculation, the detection error probability must be computed for each possible input sequence case and a statistical average is taken, When ISI is a function of many adjacent symbols, this can be tedious process. For simplicity, one may consider only the worst case. That is, consider only the input sequence that gives the large error probability.

If the $y_{n,k}$ is assumed to be Gaussian, the BER reduces to

$$P_E = E\left[p_1 Q\left(\frac{A_H p_{out}[0] - y_{th} - ISI_k}{\sigma_H}\right) + p_o Q\left(\frac{y_{th} - A_L p_{out}[0] - ISI_k}{\sigma_L}\right)\right] \tag{3.65}$$

The output noise power can be decomposed as

$$\sigma^2_{n,out} \overset{def}{=} \sigma^2_{sd} + \sigma^2_{si} \tag{3.67}$$

3.5 Optical Amplification

Power loss in the fiber transmission places a fundamental limit on the speed and distance product. When various couplings and splittings are used in an optical network or a photonic switch, power loss can also limit the network size and switch throughput.

To overcome power loss problem, optical amplifiers can be used. Compared to electronic amplifiers that amplify electrical signals, optical amplifiers have several important advantages. First, they have a much larger amplification bandwidth. An optical fiber amplifier, for example, can have a bandwidth of several thousand GHz. Therefore, they are attractive to all-optical networking, where speed bottlenecks from electronics are removed by optics implementation. Second, optical amplifiers can amplify multiple optical inputs at different wavelengths simultaneously. Therefore, they are attractive to applications such as wavelength division multiplexing (WDM).

3.5.1 Semiconductor Amplifiers

As mentioned earlier semiconductor amplifiers are laser diodes that are biased below the threshold current. To provide amplification, the active layer of a semiconductor amplifier has positive medium gain but not large enough for laser emission.

(a) External pumping and rate equation

Similar to laser diodes, a positive optical gain in semiconductor amplifiers comes from external current injection. From the rate equation of the carrier density given by equation (3.23),

$$\frac{\partial N(t)}{\partial t} = R_p(t) - R_s(t) - \frac{N(t)}{\tau_r} = \frac{J(t)}{qd} - \Gamma v_g a[N(t) - N_{th}]N_{ph}(t) - \frac{N(t)}{\tau_r} \tag{3.68}$$

Where

$$R_p(t) = \frac{J(t)}{qd} \tag{3.69}$$

is the external pumping rate from current injection.

$$R_s(t) = \Gamma a v_g [N - N_{th}] N_{ph} \tag{3.70}$$

Where

$$N_{ph,sat} \stackrel{def}{=} \frac{1}{\Gamma a v_g \tau_r} \tag{3.71}$$

is called the saturation photon density.

And

$$g(N) = \frac{g_0}{1 + N_{ph} / N_{ph,sat}} \tag{3.72}$$

is the gain of the amplifier.

Where

$$g_0 = \Gamma a \tau_r \left[\frac{J}{qd} - \frac{N_{th}}{\tau_r} \right] \tag{3.73}$$

is called medium gain at zero photon density.

(b) *Amplifier gain*

When the medium gain is known, the light power $P(z)$ is determined by following differential equation

$$\frac{dP(z)}{dz} = g(N)P(z) \tag{3.74}$$

From gain saturation is considered from equations (3.72) and (3.74),

$$dP = g(z)P(z)dz = g_0 dz \frac{P(z)}{1 + P(z)/P_{sat}}$$

Where $P_{sat} = N_{ph,sat}(hf)(wd)v_g$ is the saturation optical power.

So

$$g_0 dz = \left[\frac{1}{P(z)} + \frac{1}{P_{sat}} \right] dP$$

integrating above equation from z=0 to z=L gives

$$G_0 = 1 + \frac{P_{sat}}{P_{in}} \ln\left(\frac{G_{0,no-sat}}{G_0}\right) \tag{3.75}$$

Equation (3.75) can be rearranged as

$$10\log_{10}\left(\frac{G_{0,\,no-sat}}{G_0}\right) \approx 4.34\frac{P_{in}}{P_{sat}} G_0 \tag{3.76}$$

This given equation shows the gain penalty in dB is linearly proportional to the actual amplifier gain G_0 and the power ratio P_{in} / P_{sat}.

(c) Fabry-Perot amplifiers

Because the two cavity facets of an amplifier an cause reflections, incident light can be bounced back and forth within the amplifier. Amplifiers that have strong internal reflections are called Fabry-Perot (FP) amplifiers.

To find the amplifier gain of FP amplifiers, assume a certain optical power distribution P(z) in the cavity.

$$P(z) = (1 - r_L^2) \,|\, E^+(z) + E^-(z) \,|^2 \tag{3.77}$$

Equation (3.77) can be simplified as

$$P(z) = P_{in}(1 - r_L^2) \left| \frac{1}{1 - G_0 r_L r_R e^{j\theta_0}} \right|^2 \left\{ G(z) + \frac{G_0^2 r_R^2}{G(z)} + 2r_R G_0 \cos(2\beta z - \theta_0) \right\} \tag{3.78}$$

with

$$G(z) \stackrel{def}{=} e^{\int_0^z g(z')dz'} \tag{3.79}$$

Once P(z) and G(z) are solved, from equation (4.78), the net amplification gain of the FP amplifier is

$$G_{FP} = (1 - r_L^2)(1 - r_R^2)\frac{\left|E^+(L)\right|^2}{P_{in}} = (1 - r_L^2)(1 - r_R^2)\frac{G_0}{\left|1 - G_0 r_R r_L e^{j2\beta L}\right|^2} \tag{3.80}$$

3.6 Noise in Optical Communication

When the transmission channel is not ideal, the waveform of transmitted signal is distorted. As a result, the transmitted signal cannot be perfectly recovered, and it is an important task to minimize the effects of noise and distortion at the receiver end. In analog communications, this means maximizing the *signal-to-noise ratio* (SNR); in digital communications, this means minimizing the *bit error rate* (BER).

In optical communications, noise can come from both the transmitter and receiver. In addition to *thermal noise,* which occurs in essentially every electronic circuit, there are *phase noise, relative intensity noise* (RIN), and *mode partition noise* (MPN) from the light source at the transmitter side, and *shot noise* and *excess (avalanche gain) noise* from the photo detector at the receiver side.

There are additional noise sources in advanced systems. For example, when optical amplifiers are used to overcome power loss, they add so-called amplified spontaneous emission (ASE) noise to the amplified signal. In the wavelength-division multiplexing (WDM) and subcarrier multiplexing (SCM) systems in which multiple channels are transmitted through the same optical fiber, there can also be adjacent channel interference (ACI) or crosstalk, which is the interference from adjacent channels because of the power spectrum overlap. Because adjacent channels are statistically independent of the channel tuned to, they can be considered as another noise source.

In addition to noise and crosstalk, there can be signal distortion because of a no ideal channel. In optical communications, distortion can come from fiber dispersion and device nonlinearity. Depending on the format of the signal transmission, channel distortion results in different effects.

Unlike thermal noise, most sources in optical communications are signal dependent. That is, when the signal level increases, the noise level also increases. For example, shot noise power is linearly proportional to the photocurrent generated. Relative intensity noise and mode partition noise power are even worse, being proportional to the photocurrent squared.

3.6.1 Effects of Noise and Distortion

To know the noise effects quantitatively, consider a basic point-to point communication system. Let the transmitted signal be *s(t)*, the channel impulse response be *h(t),* and the channel noise be *n(t).* The received signal *r(t)* is thus given by

$$r(t) = s(t) \otimes h(t) + n(t) \overset{def}{=} q(t) + n(t) \tag{3.81}$$

Where $\otimes$ denotes the convolution and $q(t) \overset{def}{=} s(t) \otimes h(t)$.

If the channel is ideal, it introduces only a certain delay and loss. Therefore, the impulse response of an ideal channel is given by

$$h(t) = a\delta(t-\tau) \tag{3.82}$$

Where a is a constant factor representing transmission loss (if a<1) and τ is the propagation delay. When h(t) is not ideal, equation (3.81) shows that received signal is corrupted by both noise and channel distortion.

(a) Effect in analog communications

In analog communications, the received signal quality can be characterized by the following ratio:

$$\gamma_Q \overset{def}{=} \frac{E\left[s(t)^2\right]}{E\left(\left|s(t)-r(t)\right|^2\right)} \tag{3.83}$$

Where $E[x]$ denotes the expectation or average of signal x. Therefore, $E\left[s(t)^2\right]$ is the average signal power and $E\left(\left|s(t)-r(t)\right|^2\right)$ is the mean square error (MSE) with respect to the original signal s(t).

When the channel is ideal, γ_Q reduces to the signal-to-noise ratio (SNR) given by

$$\gamma_Q = SNR = \frac{E\left[s(t)^2\right]}{E\left[n(t)^2\right]} \tag{3.84}$$

Where $E\left[n(t)^2\right]$ is the average noise power.

(b) Effect in digital communications

In digital communications, the consideration is a little bit different. Instead of minimizing the MSE, the objective is to recover the original bits transmitted with a minimal error detection probability. Consider a pulse amplitude modulated (PAM) signal transmitted over a channel. The received signal is given by

$$r(t) = \sum_k A_k p(t-kT_0) + n(t) \tag{3.85}$$

Where A_k is the amplitude of the kth pulse, p(t) is the received pulse, and T_0 is the interval between two consecutive pulses. To detect the transmitted amplitude A_k, the received signal is first sampled at $(kT+\tau)$ for a certain τ within $(0,T_0)$. From equation (3.85), the sampled output is

$$r_k \overset{def}{=} r(kT+\tau) = \sum_i A_i p[(k-i)T+\tau] + n_k \tag{3.86}$$

Where $p_i \overset{def}{=} p(iT+\tau)$ and $n_k \overset{def}{=} n(kT+\tau)$. Equation (3.86) shows that the sampled output consists of three terms: signal (A_k), noise (n_k), and distortion ($\sum_{i\neq k} A_i p_{k-i}$). In digital communications, the last distortion item is called the intersymbol interference (ISI) because adjacent symbols and pulses cause it.

Because of noise and ISI, r_k and A_k are not necessarily the same. A digital receiver thus needs to guess what amplitude is transmitted from the received r_k. The process of finding a likely $\hat{A}_k$ from the sampled value r_k is called detection. If $\hat{A}_k = A_k$, detection is correct; if $\hat{A}_k \neq A_k$, detection is incorrect. Because both noise and ISI are unpredictable, the detection performance is characterized by the error detection probability:

$$P_E = P(\hat{A}_k \neq A_k | r_k) \tag{3.87}$$

Therefore, the smaller the error detection probability, the better the transmission performance.

3.6.2 Noise and Distortion Reduction Techniques

When the power spectral densities (PSD) of a signal and noise are known, we can use filtering and equalization techniques to reduce the effects of noise and distortion.

(a) Wiener filtering in analog communications

In analog communications, the objective is to maximize the performance measure given by equation (3.83). To achieve this, a filter called an *equalizer* following the receiver front-end is commonly used. From the received signal given by equation (3.81), the equalized output is

$$r_{out}(t) = r(t) \otimes h_{eq}(t) = q(t) \otimes h_{eq}(t) + n(t) \otimes h_{eq}(t) \stackrel{def}{=} s_{out}(t) + n_{out}(t)$$

Where $h_{eq}(t)$ is the impulse response of the equalizer. Therefore, the MSE is

$$MSE = E\left(|r_{out}(t) - s(t)|^2\right) = E\left(|s_{out}(t) - s(t)|^2\right) + E\left(n_{out}^2\right) \stackrel{def}{=} MSE_s + E\left(n_{out}^2\right)$$

Where the noise and signal are assumed to be statistically independent. The two terms on the right hand side can be written as

$$E\left(n_{out}^2\right) = \int S_n(\omega)\left|H_{eq}(\omega)\right|^2 \frac{d\omega}{2\pi}$$

and

$$MSE_s \stackrel{def}{=} E\left(|s_{out} - s(t)|^2\right) = \int S_s(\omega)\left|H(\omega)H_{eq}(\omega) - 1\right|^2 \frac{d\omega}{2\pi}$$

Where $S_s(\omega)$ is the PSD of the original signal and $H(\omega)$ is the transfer function of the channel. Let H_{opt} be the optimum transfer function of the equalizer that gives the minimum $MSE_{s,\min}$. If $H_{eq} = H_{opt} + \delta H$, by expanding the factor $\left|H(\omega)H_{eq}(\omega) - 1\right|^2$, the MSE reduces to

$$MSE = MSE_{\min} + \Re\left\{\int \left[S_n H_{opt}^* + S_s |H|^2 \left(H_{opt}^* - 1\right)\right]\delta H \frac{d\omega}{2\pi}\right\} + \int \left(S_n + S_s |H|^2\right)|\delta H|^2 \frac{d\omega}{2\pi}$$

Where $\Re\{z\}$ denotes the real part of z. Because the last term on the right hand side of the above equation is always positive, the middle term should always be zero to ensure $MSE \geq MSE_{\min}$ for any δH. As a result, the optimum $H_{opt}(\omega)$ satisfies the following condition:

$$S_n H_{opt} + S_s |H|^2 (H_{opt} - 1) = 0$$

or

$$H_{opt}(\omega) = \frac{S_s(\omega)|H(\omega)|^2}{S_s(\omega)|H(\omega)|^2 + S_n(\omega)} \tag{3.88}$$

The optimum filter by equation (3.88) may not necessarily be implementable (i.e., its inverse Fourier transform may not necessarily have a finite duration). Therefore, it is called noncausal Wiener filter.

(b) Matched filtering in digital communication

In digital communications, instead of minimizing the MSE as in analog communications, one must minimize the error detection probability P_E. If a similar equalizer $h_{eq}(t)$ is used after the receiver front end, the filtered output from equation (3.85) is

$$r_{out}(t) = \sum_k A_k p_{out}(t - kT_0) + n_{out}(t)$$

Where $p_{out}(t) = p(t) \otimes h_{eq}(t)$. If we sample the signal at every T_0, the sampled output at $(k+1)T_0$ is

$$r_{out,k} = A_k p_{out}(T_0) + n_{out,k}$$

Where we define $r_{out,k} = r_{out}[(k+1)T_0]$ and $n_{out,k} = n_{out}[(k+1)T_0]$. In the above expression, $p_{out}(jT_0)$ terms for $j \neq 1$ are ignored. Because the filtered noise $n_{out,k}$ is still Gaussian, so

$$P_E = Q\frac{Ap_{out}(T_0)}{\sigma_{n_{out}}}$$

To minimize P_E, $p_{out}(T_0)/\sigma_{n_{out}}$ must be maximized or

$$\left[\frac{p_{out}(T_0)}{\sigma_{n_{out}}}\right]^2 = \frac{\left|\int P(\omega)H_{eq}(\omega)e^{j\omega T_0}d\omega/(2\pi)\right|^2}{\int S_n(\omega)\left|H_{eq}(\omega)\right|^2 d\omega/(2\pi)} \tag{3.89}$$

Where $P(\omega)$ is the Fourier transform of $p(t)$. If $H_{eq}(\omega) = H_{pre}(\omega)H_{post}(\omega)$ such that

$$\left|H_{pre}(\omega)\right|^2 = \frac{1}{S_n(\omega)} \tag{3.90}$$

then

$$\left[\frac{p_{out}(T_0)}{\sigma_{n_{out}}}\right]^2 = \frac{1}{2\pi}\frac{\left|\int P(\omega)H_{pre}H_{post}e^{j\omega T_0}d\omega\right|^2}{\int\left|H_{post}(\omega)\right|^2 d\omega}$$

According to Schwartz inequality,

$$\left[\frac{p_{out}(T_0)}{\sigma_{n_{out}}}\right]^2 \leq \frac{1}{2\pi}\frac{\int\left|P(\omega)H_{pre}e^{j\omega T_0}\right|^2 d\omega\int\left|H_{post}(\omega)\right|^2 d\omega}{\int\left|H_{post}(\omega)\right|^2 d\omega} = \int\frac{\left|P(\omega)\right|^2}{S_n(\omega)}\frac{d\omega}{2\pi}$$

The equality holds when $H_{post}(\omega) = KP^*(\omega)H^*_{pre}(\omega)e^{-j\omega T_0}$ for some constant K. Therefore, the optimum $H_{opt}(\omega)$ is

$$H_{opt}(\omega) = KP^*(\omega)H^*_{pre}(\omega)e^{-j\omega T_0}H_{pre}(\omega) = K\frac{P^*(\omega)e^{-j\omega T_0}}{S_n(\omega)} \tag{3.91}$$

This optimum filter is called the *matched filter*.

3.6.3 Shot Noise from PIN Diodes

The photocurrent from a PIN diode is constant when the incident light power is constant. In practice, because of random EHP generation, the photocurrent has a random fluctuation from its average value. This random fluctuation is called shot noise and is the most fundamental noise in optical communications.

(a) *Power spectral density of shot noise*

Shot noise $n_{shot}(t)$ as a function of time at the photodiode output is defined to be

$$n_{shot}(t) \overset{def}{=} i_{ph}(t) - I_{ph} \tag{3.92}$$

Where $i_{ph}(t)$ is the photocurrent and I_{ph} is its average. The two-sided PSD of a shot noise is given by

$$S_{shot} = q(I_{ph} + I_d)|H_{pin}(\omega)|^2 \approx q(I_{ph} + I_d) \tag{3.93}$$

Where I_d is the dark current and $H_{pin}(\omega)$ is the Fourier transform of the impulse response of the PIN diode due to an HER. Because $H_{pin}(\omega)$ is generally flat over a large frequency range, it can be dropped from equation (3.93). In other words shot noise can be considered as a white noise over most relevant frequency ranges. If this is the case, the shot noise power over a bandwidth B is

$$\overline{n^2_{shot}} = \int S_{shot}(\omega) \times \frac{d\omega}{2\pi} \approx 2q(I_{ph} + I_d)B = 2q(RP_{in} + I_d)B \tag{3.94}$$

3.6.4 Thermal Noise

Thermal noise, a *white Gaussian noise*, is one of the most common kinds of noise encountered in communication systems. Thermal noise is caused by radiation from random motion of electrons. Because it is a Gaussian noise, the probability density function (PDF) of thermal noise is Gaussian. This Gaussian distribution comes from the fundamental central limit theorem, which states that if the number of noise contributor is large and they are statistically independent, the combined noise distribution is Gaussian. From thermodynamics, the PSD of thermal noise is given by

$$S_T(\omega) = \frac{h\omega}{2\pi}\left(\frac{1}{2} + \frac{1}{e^{h\omega/2\pi kT} - 1}\right) \tag{3.95}$$

Where k is the Boltzmann constant and T is the temperature in Kelvin. The first term in equation (3.95) is from quantum mechanics. When $kT >> h\omega/2\pi$, the power spectrum is almost a constant and equal to kT. From this approximation, thermal noise is a white noise with the following PSD:

$$S_T(\omega) = kT \tag{3.96}$$

The inverse Fourier transform gives the following autocorrelation for thermal noise:

$$R_T(\tau) = E[n_T(t)n_T(t+\tau)] = kT\delta(\tau) \tag{3.97}$$

If the noise is filtered over a finite frequency band B, the filtered power spectrum will be zero outside the frequency band, and the average power is

$$\sigma^2 = R_T(0) = \int_{\text{frequency bands}} kTdf = 2kTB \tag{3.98}$$

Thermal noise can be modeled as a voltage source of bandwidth B by:

$$\frac{\overline{v_{thermal}^2}}{2R} = 2kTB \tag{3.99}$$

Therefore, we have

$$\overline{v_{thermal}^2} = 4kTRB \tag{3.100}$$

In equation (3.99), the factor of 2 in the denominator on the left-hand side is to account for the optimum power transfer efficiency. That is, 50 percent of the noise power from the equivalent voltage source contributes to the measurable noise power 2kTB. The thermal current source can be similarly expressed as

$$\overline{i_{thermal}^2} = 4kTGB \tag{3.101}$$

Where G=1/R is the conductance.

If the thermal noise is included with the shot noise discussed earlier, the SNR at the photodiode output can be expressed as

$$SNR = \frac{(RP_{in})^2}{2qB(RP_{in} + I_d + 2V_TG)} = \frac{I_{ph}^2}{2qB(I_{ph} + I_d + 2V_TG)} \tag{3.102}$$

Where $V_T = kT/q$ is the thermal voltage and G is the conductance of the load resistor.

3.6.5 Relative Intensity Noise in Laser Diodes

Relative intensity noise is the intensity fluctuation at a laser diode output. Similar to phase noise, it is primarily caused by random spontaneous emission. Different from phase noise, which is critical to coherent communication and soliton transmission, RIN is important in analog communications, such as in SCM (subcarrier multiplexing). If incident light to a light detector has power $P_{in} + \Delta P_{in}$, where ΔP_{in} is the zero mean intensity noise, RIN is defined as

$$RIN \stackrel{def}{=} \frac{E\left[\Delta P_{in}^2\right]}{BP_{in}^2} = \frac{2S_{rin}}{P_{in}^2} \tag{3.103}$$

Where B is the bandwidth of the signal and S_{rin} is the two-sided PSD of $\Delta P_{in}(t)$. Because the photocurrent at the photodetector output is proportional to the incident light power, the mean square of the photocurrent fluctuation can be written as

$$E\left[\Delta i_{rin}^2\right] = RIN\left(M_{apd} RP_{in}\right)^2 B \tag{3.104}$$

In addition to spontaneous emission, laser RIN can also be caused be modal instabilities and reflection at the diode-fiber interface. Therefore, RIN can be factored into two components: intrinsic RIN, which is caused by spontaneous emission and internal modal instabilities, and extrinsic RIN, which is caused by reflection.

3.6.6 Phase Noise from Laser Diodes

As mentioned earlier, phase noise is caused by random spontaneous emission in a laser diode. If there is no spontaneous emission, the output light spectrum consists of delta functions. When there are random spontaneous emissions, the spectrum is no longer a sum of delta functions. Instead, the spectrum is broadband and has a finite nonzero linewidth around each ω_i.

(a) Linewidth broadening because of phase noise

To see the effect of linewidth broadening, consider a continuous wave light output of a single-mode laser given by

$$x(t) = e^{j[\omega_c t + \phi(t)]} \tag{3.105}$$

Where ω_c is the center frequency and $\phi(t)$ is the phase noise. If $\phi(t) = 0$, x(t) is at a single frequency and its PSD is a delta function. When there is phase noise, the PSD due to phase noise is given by

$$S_x(\omega) = \frac{2t_{coh}}{1+t_{coh}^2\left(\omega-\omega_c+\alpha_{l\omega}R_{sp}/2I\right)^2} \tag{3.106}$$

Where R_{sp} is the spontaneous emission rate (1/sec), $\alpha_{l\omega}$ is the linewidth broadening factor (1/sec), I is total number of photons in the laser cavity, and t_{coh} is the *coherence time* given by

$$t_{coh} \overset{def}{=} \frac{4I}{R_{sp}(1+\alpha_{l\omega}^2)} \tag{3.107}$$

Equation (3.107) shows that t_{coh} is proportional to I, or the output optical power. Therefore, the larger the optical power, the larger the coherence time. The spectrum given by equation (3.106) is commonly referred to as a *Lorentzian spectrum.* Its FWHM bandwidth is

$$\Delta\omega_{3dB} = \frac{2}{t_{coh}} = \frac{2R_{sp}}{4I}\left(1+\alpha_{l\omega}^2\right) \tag{3.108}$$

Therefore, the larger the output optical power, the smaller the 3dB linewidth. Good single-frequency laser diodes today can have a 3dB bandwidth around 100 kHz.

(b) Phase noise in coherent communications

In addition to the linewidth broadening effect discussed above, phase noise can directly corrupt the phase or frequency of a modulated carrier in coherent communications. The phase noise $\phi(t)$ given by Equation (3.105) consists of two components: one deterministic that results in a frequency shift from ω_c and one random that broadens the linewidth. Because a constant frequency shift has no effect on phase or frequency modulated carriers, only the random phase component must be considered. Denote it by $\phi_r(t)$. The autocorrelation of the time derivative of $\phi_r(t)$ is given by

$$R_{\dot{\phi}_r}(t) = \Delta\omega_{3dB}\delta(\tau) \tag{3.109}$$

By taking Fourier transform, the corresponding PSD is

$$S_{\dot{\phi}_r}(\omega) = \Delta\omega_{3dB} \tag{3.110}$$

Because $\dot{\phi}_r$ is the time derivative of $\phi_r(t)$, the PSD of the phase noise is

$$S_{\phi_r}(\omega) = \frac{S_{\dot{\phi}_r}(\omega)}{\omega^2} = \frac{\Delta\omega_{3dB}}{\omega^2} \tag{3.111}$$

Since the time derivative of a phase is frequency, ϕ_r represents a *frequency noise.* Equation (3.110) shows that the frequency noise caused by shot noise is a white noise, and when FSK is used the frequency separation must be large enough compared to $\Delta\omega_{3dB}$. On the other hand, equation (3.111) shows that the phase noise is proportional to $1/f^2$, which means a strong noise in the low-frequency range. When phase modulation is used, the bit rate must be high enough that the effect of phase noise can be minimized.

3.6.7 Mode Partition Noise

Mode partition noise (MPN) is caused by mode competition inside a multimode FP laser cavity. As a result, even though the total power is constant, the power distribution over different modes is random. Because different modes have different propagation delays in fiber transmission, random power distribution results in random power variation at the receiving end. This power fluctuation due to mode competition is called MPN.

Because the power competition among all longitudinal modes is not fully understood, an exact description of the PDF is not available. However, similar to RIN, it is well known that the noise power of MPN is proportional to the signal power. As a result, an error floor can be reached when MPN becomes dominant.

Suppose a given laser diode has N longitudinal modes and each has a relative power a_i, $i = 1,......,N$. By definition, the sum of these a_i's satisfies

$$\sum_{i-1}^{N} a_i = 1 \tag{3.112}$$

Because each a_i at a certain time is random variable, the average relative power for mode i is given by

$$\overline{a_i} = E[a_i] = \int a_i \times PDF(a_1,......,a_N)da_1....da_N$$

If the waveform of mode i received is $f_i(t)$, the combined received signal is

$$r(t) = \sum_i a_i f_i(t) \tag{3.113}$$

If the signal is sampled at time t_0, the variance of the sampled signal is

$$\sigma^2 = E[r(t_0)^2] - E[r(t_0)]^2 \tag{3.114}$$

From equations (2.33) and (2.34),

$$\sigma^2 = \sum_{i,j} f_i(t_0) f_j(t_0)(\overline{a_i a_j} - \overline{a_i}\,\overline{a_j}) \tag{3.115}$$

To simplify the expression, Ogawa introduces a constant k_{mpn}, which is defined as

$$k^2_{mpn} \overset{def}{=} \frac{\sum_{i,j}[f_i(t_0) - f_j(t_0)]^2(\overline{a_i a_j} - \overline{a_i}\,\overline{a_j})}{\sum_{i,j}[f_i(t_0) - f_j(t_0)]^2\,\overline{a_i a_j}} \overset{def}{=} \frac{N}{D} \tag{3.116}$$

Where

$$\begin{aligned} D &\overset{def}{=} \sum_{i,j}[f_i(t_0) - f_j(t_0)]^2\,\overline{a_i a_j} \\ &= \sum_i f_i(t_0)^2\overline{a_i} - 2\sum_i\sum_j f_i(t_0)f_j(t_0)\overline{a_i a_j} + \sum_j f_j(t_0)^2\overline{a_j} \\ &= 2\sum_i f_i(t_0)^2\overline{a_i} - 2\left[\sum_i f_i(t_0)\overline{a_i}\right]^2 \end{aligned} \tag{3.117}$$

and

$$\begin{aligned} N &\overset{def}{=} \sum_{i,j}[f_i(t_0) - f_j(t_0)]^2(\overline{a_i a_j} - \overline{a_i}\,\overline{a_j}) \\ &= \sum_{i,j}[f_i(t_0)^2 - 2f_i(t_0)f_j(t_0) + f_j(t_0)^2](\overline{a_i a_j} - \overline{a_i}\,\overline{a_j}) \\ &= 0 - 2\sum_{i,j} f_i(t_0)f_j(t_0)(\overline{a_i a_j} - \overline{a_i}\,\overline{a_j}) \\ &= 2\sigma^2 \end{aligned} \tag{3.118}$$

Therefore, from equations (3.115) and (3.118),

$$\begin{aligned} \sigma^2 &= k^2_{mpn} \times \frac{1}{k^2_{mpn}} \times \sigma^2 = k^2_{mpn}\frac{D}{N}\sigma^2 = k^2_{mpn}\frac{D}{2} \\ &= k^2_{mpn}\left\{\sum_i f_i(t_0)^2\overline{a_i} - \left[\sum_i f_i(t_0)\overline{a_i}\right]^2\right\} \end{aligned} \tag{3.119}$$

Because $\overline{a_i}$ is the average power distribution, which can be measured experimentally, the last expression for σ^2 is a convenient form to compute σ^2 at a given k_{mpn}.

3.6.8 Avalanche Noise in APDs

Because of multiplication and avalanche process, the photocurrent generated by an APD has an even larger fluctuation. This larger fluctuation comes from the random secondary EHP generation by each primary EHP. The total noise is called the APD noise, and the contribution from the random secondary generation is called the *excess noise*, which is quantified by a factor called the *excess noise factor*.

The APD noise power over frequency range B is

$$\overline{n_{apd}^2} = 2q(I_{ph}M_{apd} + I_d M_{apd}^2)BF_{apd} \tag{3.120}$$

Where M_{apd} is APD gain and F_{apd} is excess noise factor.

3.6.9 Noise from Optical Amplification

Because spontaneous emission comes together with optical amplification, noise is added when a light signal passes through an amplifier. As mentioned earlier this noise is called the amplified spontaneous emission (ASE) noise because it is also amplified.

(a) Amplified spontaneous emission noise

Photons that contribute to the ASE noise can have different transverse modes in the amplifier and different polarizations. The one-sided power spectral density of the ASE noise that is of the same mode and polarization as the incident light signal was first derived by Kogelnik and Yariv and can be expressed as

$$S_{ASE}(f) = (G-1)n_{sp}\chi hf \tag{3.121}$$

In the expression, G is the amplifier gain, hf is the photon energy, and

$$n_{sp} = \frac{N_2}{N_2 - N_1} \tag{3.122}$$

is a parameter describing the external pumping that depletes the ground state population.

3.6.10 Total Noise

The above discussion presented various kinds of noise in optical communications. Now let us check their total effect on final signal detection. As illustrated in figure 3.11, the total noise at the photodetector output and front-end amplifier input is

$$n_{tot}(t) = n_{mpn}(n) + n_{rin}(t) + n_{apd}(t) + n_{th}(t) \tag{3.123}$$

If a PIN diode is used instead of an APD, $n_{apd}(t)$ should be replaced by $n_{shot}(t)$. Because phase information is not detected in direct detection, phase noise is not included in equation (3.123).

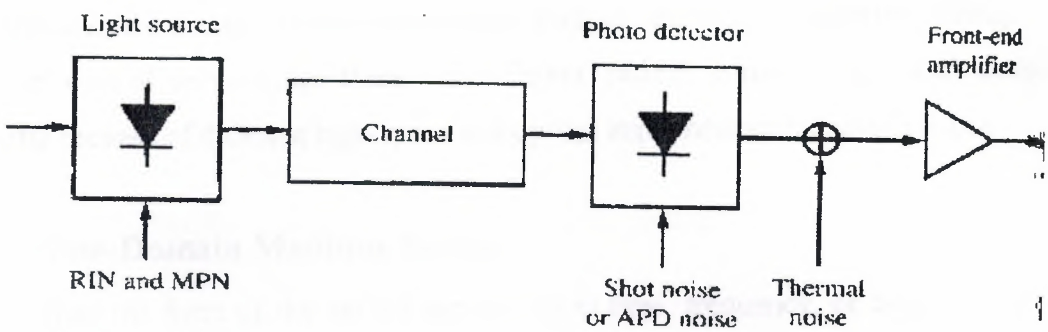

Figure 3.11 Total current noise at the photodetector output.

CHAPTER 4
NETWORKING

Because of the fast growing needs of networking services and rapid development of supporting technologies, light wave technology research and development have expanded from point-to-point transmission to high speed networking over the last decade. This expansion is very critical to many broadband services, such as interactive video, medical imaging, and distributed computing that need not only high-speed transmission but also networking support such as medium access and switching.

In optical networking, there are different system considerations and design trade-offs because of different high-speed and optical implementation characteristics.

2.1 Time-Domain Medium Access

When the form of the shared resource is in time, frequency, or code, we have time-domain, frequency-domain, and code-domain access, respectively. To access the shared medium in the time-domain, a transmission node needs to know when and how long it can send its data. One sample approach is to divide time into frames and partition each frame into a certain number of time slots. In this way, communication nodes can share the same medium by sending data in different slots. Once a slot is allocated during the call setup, a node can repeat the transmission in the same slot of each frame throughout the call duration. This is called *deterministic* access and there is no traffic contention after the call setup.

Data transmission in two adjacent slots can be continuous in bit timing or separated by guard time. To transmit data in different slots continuously, bit timings from different sources first need to be synchronized. This medium access scheme is called *time division multiplexing* (TDM). When a network is used for multiple access, it is impossible to synchronize the timings of all nodes distributed over the network. As a result, guard time that separates consecutive slot transmissions is necessary, and the medium access scheme is called *time division multiple access* (TDMA).

Because TDM requires bit timing synchronization, its implementation is more involved. To synchronize input bit rates, electronic implementation has to be used and direct optical domain multiplexing is not practically feasible. A TDM standard called Synchronous Optic Network (SONET) has been designed for optical transmission. To avoid the electronic bottleneck in high speed multiplexing, SONET introduces an innovative floating payload concept that can synchronize input signals at low speeds.

Because of no need for bit timing synchronization, TDMA can be done directly in the optical domain. An optical transmitter in this case gets access to the medium by transmitting an optical burst consisting of binary bits over a time slot. When all optical networking is concerned, TDMA is an attractive choice compared to TDM. However, TDMA has a worse access efficiency than TDM because of the guard time.

In addition to TDMA and TDM, token passing and random access can do time-domain medium access. An important token passing fiber based local area network is called the Fiber Distributed Data Interface (FDDI).

4.1.1 Time-Division Multiple Access

As illustrated in Figure 4.1, a TDMA frame consists of a *reference burst* and a certain number of time slots. The reference burst is used for frame synchronization, signaling, and the time slots are used to carry data.

The reference burst consists of three parts: a preamble, a start code, and control data. The preamble is a periodic bit sequence for bit timing synchronization. Depending on how rapid synchronization can be achieved, the preamble length can range from 10 to several hundred symbols.

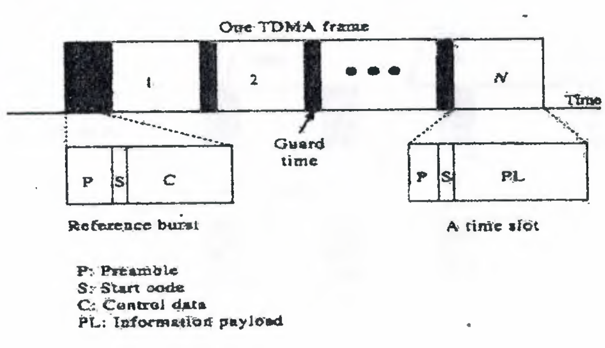

Figure 4.1 A TDMA frame consisting of N slots

Once the bit timing is established, the content in the rest of the reference burst can be read. Following the preamble is a unique start code indicating the end of the preamble and the start of the information portion of the reference burst. By recognizing the word, control data can be interpreted correctly; In general, control data carries information such as station timing, call setup status, and signaling information.

The reference burst in a TDMA frame is the overhead and occupies only a small portion of the frame. The rest of the frame is divided into time slots separated by guard time. Similar to the reference burst, each time slot consists of a preamble, a unique start code, and the information payload. Because of different propagation delays between stations, the guard time between time slots is necessary to avoid overlap between two consecutive time slots.

(a) *Shared medium*

Consider a TDMA network made of a shared medium where a node can immediately receive what it just transmits. A single Ethernet coaxial cable and an ALOHA radio network are examples of such a shared medium.

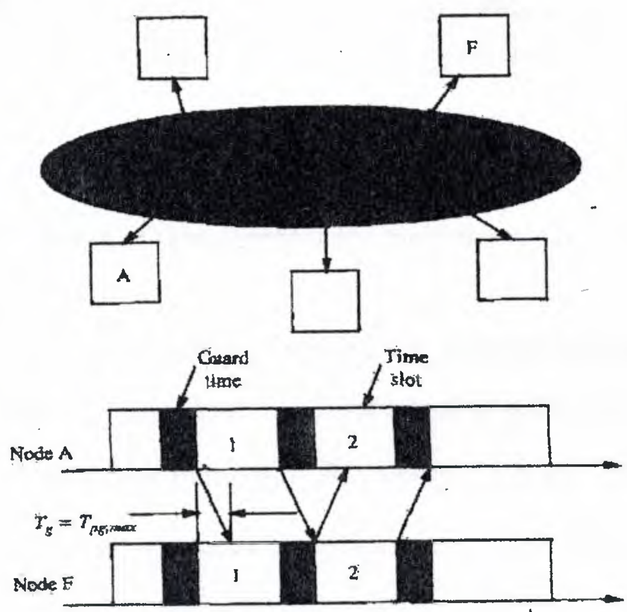

Figure 4.2 Required guard time of a shared TDMA network.

(b) *Medium access and time compression*

When nodes attached to a TDMA network have data to transmit, they need to wait for available time slots. In TDMA, a slot is obtained through a call setup process.

Once a node is granted with an available slot, it can use the same slot in every frame throughout the call. As mentioned earlier, this medium access is called deterministic.

Because of the deterministic access, TDMA is primarily used for constant bit transmission. To support such transmission, input bits of a constant bit-rate signal are tip; Stored in a transmitter buffer. When the assigned time slot arrives, all the bits stored will be transmitted at a much higher speed. This process is illustrated in Figure 4.3. Clearly, data bits are compressed in time during the high-speed transmission. For this reason, TDMA also called *time compression multiplexing*.

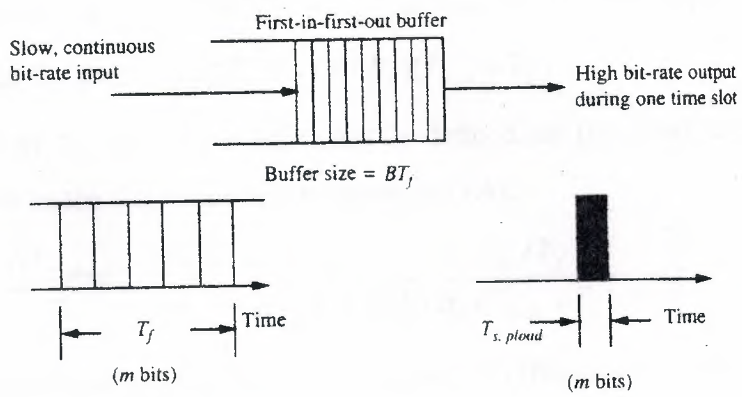

Figure 4.3 Concept of time compression multiplexing in TDMA.

From the compression mechanism described, the payload size of a time slot can be determined for a constant bit-rate input. For input signal of a bit rate B_S b/s, the total number of input bits during a time frame is $T_f B_S$, where T_f is the TDMA frame size. To transmit bits over the payload of time slot, payload size $T_{S,pload}$ should be at least

$$T_{f,pload} = T_f \frac{B_S}{R} \tag{4.1}$$

(c) Slot size, compression delay and access efficiency

In addition to its payload, a time slot has an overhead for synchronization and signaling. Therefore, the time slot size T_S is

$$T_S = T_{S,oh} + T_{S,pload} \tag{4.2}$$

Where $T_{S,oh}$ is the overhead size for preamble, signaling and control.

From equation (4.2) and (4.1) the time slot size is

$$T_S = T_{S,oh} + T_f \frac{B_S}{R} \tag{4.3}$$

At a given reference burst size T_{ref} and guard time T_g, the total number of time slots that can be accommodated within a frame is

$$N = \left\lfloor \frac{T_f - T_{ref}}{T_S + T_g} \right\rfloor = \frac{R}{B_S} \frac{1 - T_{ref}/T_f}{1 + (R/B_S)(T_{S,oh} + T_g)/T_f} \tag{4.4}$$

From equation (4.4), note that N has an upper bound of R/B_S and can be improved by increasing T_f. At a given $T_g, T_{ref}, T_{S,oh}, R/B_S$, the upper bound can be approached when $T_f >> T_{ref}$ and $T_f >> (R/B_S)(T_{S,oh} + T_g)$.

At a given N, the access efficiency is defined as the ratio of the total data transmission time to the frame size. From equation (4.4),

$$\eta_{TDMA} = \frac{NT_{S,pload}}{T_f} = \frac{NB_S}{R} = \frac{B_S}{R}\left\lfloor \frac{R}{B_S} \frac{1 - T_{ref}/T_f}{1 + (R/B_S)(T_{S,oh} + T_g)/T_f} \right\rfloor \tag{4.5}$$

When T_f is large enough, the upper limit of the access efficiency is $(B_S/R)\lfloor R/B_S \rfloor$.

4.1.2 Optical Domain TDMA

In spite of the difficult trade-off mentioned above, optical domain TDMA is still attractive because of no need to synchronize bit timings among transmission nodes. As a result, as long as the receiver knows the transmission bit rate, each transmission node can have its independent bit clock. Because the bit rate is usually high in optical domain TDMA, this clock independence that avoids the need of bit clock synchronization among transmission nodes is an important characteristic.

The star topology is preferred to minimize the effect of guard time. A block diagram of such an optical domain TDMA network is shown in figure 4.4. To synchronize the access, master frame timing is needs to be distributed to all nodes. To achieve this one of the nodes in the network is called master node, which generates a reference burst every T_f.

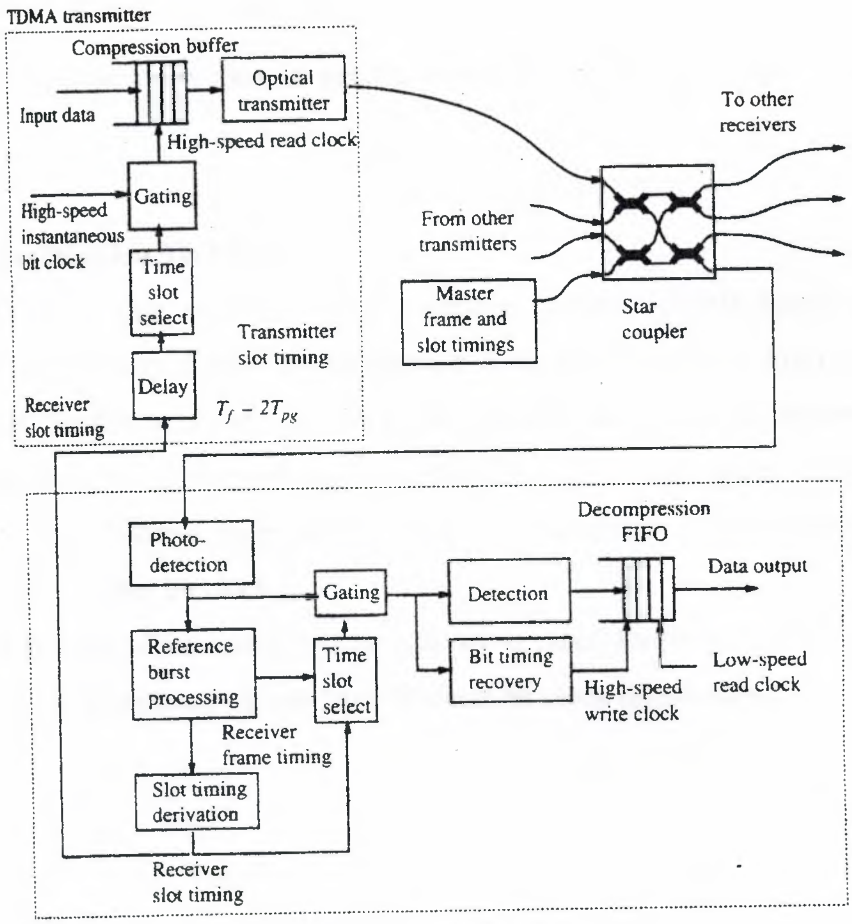

Figure 4.4 An optical domain TDMA implementation.

To receive data over a certain time slot, a gating signal that is high during the slot interval is generated from the derived slot timing. As illustrated in figure 4.4, data in this slot interval can pass through the gate, be detected, and then be stored in the decompression buffer. The received slot timing derived is also sent to the local transmitter to determine its slot timing for transmission.

Therefore, to have same frame timing at the star coupler, the additional delay D_A is added so that

$$\lfloor T_{M,pg} + T_{A,pg} + D_A \rfloor + T_{A,pg} = T_{M,pg} \bmod T_f \tag{4.6}$$

Where $T_{M,pg} + T_{A,pg} + D_A$ is the relative delay of the transmitter frame timing at the user node with respect to the master frame timing. From equation (4.6)

$$D_A = mT_f - 2T_{A,pg} \tag{4.7}$$

Where m is the smallest integer such that $mT_f - 2T_{A,pg} > 0$. Equation (4.7) shows that a large delay line is needed when T_f is large to have a high access efficiency.

4.1.3 Time-Division Multiplexing

TDM was first used in digital telephony, where multiple lower rate digital streams called *tributary signals* are interleaved in the time domain to form a higher rate digital signal. Similar to TDMA, TDM is time domain medium access scheme and each of its frames consists of a certain number of time slots. However, different from TDMA, data carried by different time slots are first synchronized in bit timing, and then interleaved by a higher bit clock.

The process of bit timing synchronization is called *frequency justification*, which is necessary when upstream signals have different bit clock frequencies.

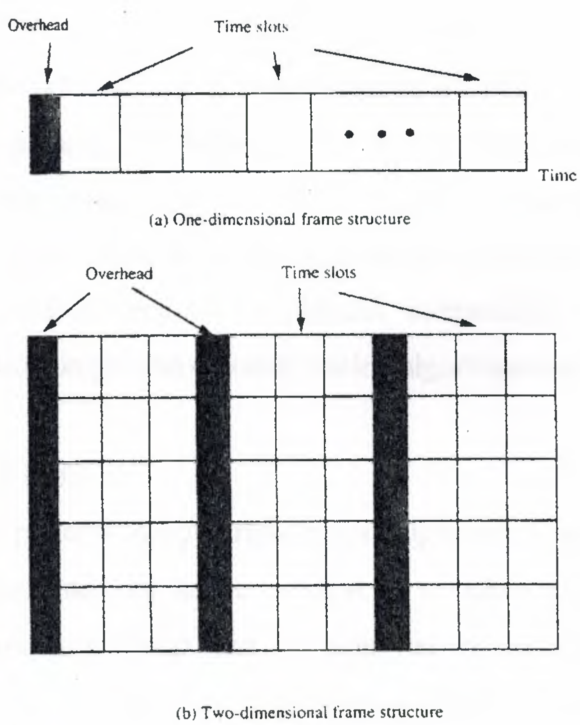

Figure 4.5 TDM frame formats.

The process of bit timing synchronization is called *frequency justification*, which is necessary when upstream signals have different bit clock frequencies. Because all laser diodes can operate only up to tens of GHz, parallel transmission from WDoMA is a key to high utilization of fiber bandwidths.

4.2.1 Wavelength-Division Multiple Access

In WDMA, wavelength channels are the shared resources. To access the network, each source node needs to first acquire a wavelength channel. To ensure proper transmission to the final destination, either the destination receiver must be tuned to the wavelength channel or there must exist an efficient routing algorithm that can forward data to the destination. Over the last few years, many access protocols and routing algorithms have been proposed. In general, a good WDMA access protocol should both perform satisfactorily and be easy to implement. That is, from the system prospective, the protocol should achieve a high access efficiency, high throughput, and low transmission delay. From the implementation prospective, the required tuning speeds of tunable devices should be feasible, the number of total wavelength channels needed should be attainable, and the channel allocation algorithm should be simple and fast.

(a) Logical configurations and routing in fixed tuning WDMA

When the traffic pattern in a WDMA network is stationary, fixed tuning is a good choice to simplify the design and relax the fast tuning requirement. By choosing the tuning wavelengths of individual transceivers properly, a WDMA network can have a logical configuration independent of its physical connections. Once the logical configuration is determined, simple and efficient routing algorithms can be derived.

(b) Dynamic tuning WDMA

When the traffic pattern changes rapidly and high speed-tuning technology is available, a more efficient medium access control is to allow transceivers to tune dynamically. As a result, single-hop can be achieved by setting up a common wavelength channel between the source transmitter and destination receiver.

From the logical configurations, note that a wavelength channel can be mapped logically to spatial channel. Therefore, dynamic tuning in WDMA is in principle logically equivalent to dynamic spatial switching. Consequently, borrowing switching techniques used in packet switching networks can perform dynamic tuning.

(c) *Frequency reuse*

In a WDMA network where all wavelength channels are coupled in the same-shared medium, the number of wavelengths required is

$$N_{\lambda} = pN \tag{4.8}$$

Where N is the total number of nodes and p is the number of transceiver pairs per node. Therefore, when N and p are large, the number of channels can be too large to be implmentable.

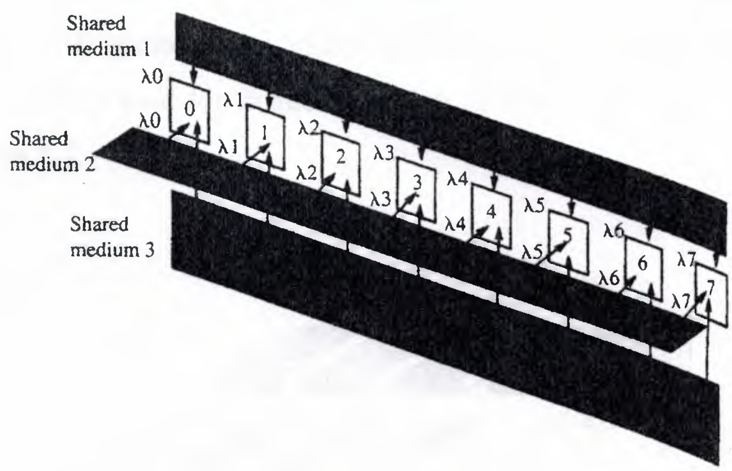

Figure 4.6 Frequency reuse by wavelength partitioning.

To reduce the number of wavelengths required, the concept of frequency reuse has been proposed. There are several approaches to achieving this. First, multiple shared media can be used for wavelength isolation, As illustrated in figure 4.6, a logical hypercube configuration can be implemented by partitioning wavelength channels into three physical networks. As shown for a network size of 8, the number of wavelength channels can be reduced from 24 to 8. As a result, frequency resolution is reduced by a factor of 3. However, three separate sets of fibers, couplers, and splitters are needed.

4.2.2 Tunable Sources

WDoMA requires various advanced opto-electronic devices to implement carrier generation, coupling, and tuning. In the last few years, many interesting and innovative designs have been proposed.

In principle, discrete DFB, DBR lasers can be used as tunable sources inWDoMA. However, in either WDM or WDMA, it is desirable to build a large tunable laser array for cost effective implementation. In WDMA it is desirable to use a large number of transceivers per node for high throughput and low transmission delay. In WDM, a large laser array can also be integrated with multiplexing devices fir high-speed transmission.

(a) Tunable laser arrays

For WDM applications, high speed and tunable laser array designs of sizes up to 20 lasers have been reported recently. Most laser arrays for WDM applications are based on MQW and DBR combinations, where MQW is used as the active layer for low threshold current a DBR is used for frequency tuning. A typical MQW-DBR laser array is illustrated in figure 4.7.

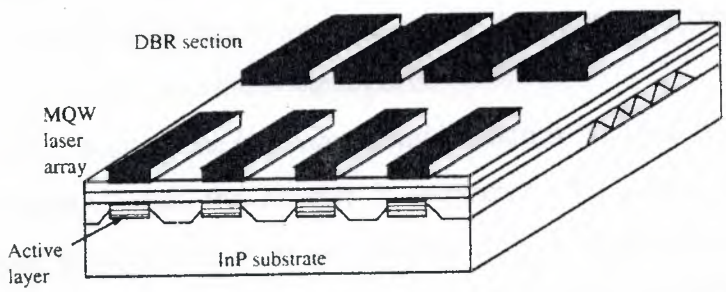

Figure 4.7 An MQW-DBR laser array.

(b) Waveguide grating router and optical amplifier

Another innovative design of tunable light sources has been proposed by forming a positive feedback loop between optical amplifiers and the wave-guide grating router (WGR). In this way, the WGR serves as an optical filter. When the net gain of positive feedback loop is greater than unity or 0 dB, optical carriers with wavelengths determined by the WGR can be generated.

(c) Frequency tuning and stabilization

As pointed out earlier, accurate wavelength generation is important to proper wavelength multiplexing, interchannel interference, and tuning. The output wavelength of a tunable laser, however, can vary because of fabrication errors, temperature, bias current, and aging. A typical temperature coefficient of the output wavelength of a single-frequency laser is around 0.09 nm/C°.

One way to maintain frequency stabilization is to use an absolute frequency source that is stable against various external conditions. In this approach, both channel separations and absolute frequencies can be maintained for proper channel multiplexing and tuning.

When the absolute frequencies are not required, it is only necessary to ensure constant channel separation to avoid a large ICI. In this case, a single FPI can be used as the calibrator.

4.2.3 Frequency Independent Coupling

Wavelength channels need to be coupled to the shared medium for transmission. To allow a node in WDMA to tune its wavelength channels for transmission, coupling should be made insensitive to the wavelength. For this purpose, a star coupler is commonly used. A star coupler has N input ports and N output ports, with each output port consisting of all signals from the N input ports.

In star coupler design, one important consideration is the total power loss. From conservation of energy, when the coupler is symmetric, the output power of each channel at each port is attenuated by at least a factor of N, the size of star coupler. This 1/N coupling loss is called distribution loss because the incident light from one input port is distributed to N output ports. In addition to distribution loss, there is excess loss due to many couplings with loss such as internal power attenuation. As a result, the total coupling loss in a star coupler is

$$\alpha_{tot} dB = 10 1\log_{10}(N) + \alpha_{excess} dB \tag{4.9}$$

4.2.4 Frequency Dependent Multiplexing

In addition to star couplers, frequency-dependent multiplexers can be used to mix wavelength channels. In this case there is no distribution loss. However, mixed outputs are strongly dependent on wavelengths of input signals. Different from star couplers, multiplexers are thus primarily used in WDM. In WDMA, where nodes can tune their transmission wavelengths, use of frequency-dependent multiplexers is not suitable.

To implement a frequency-dependent multiplexer, there are two primary designs: grating and Mach-Zehnder interferometry. In the first case, light from different spatial directions is coupled together. In the later case, light of different wavelengths is coupled together, with coupling coefficients dependent on the input wavelengths.

Grating is an arrangement that imposes on an incident light a periodic phase variation. When many phase-shifted versions of the incident light are superimposed, a spatially dependent intensity pattern is formed.

Another emerging and important design of frequency-dependent multiplexers is based on Mach-Zehnder interferometry. In this approach, a $N:1$ multiplexer is composed of $N-1$ 2×2 Mach-Zehnder interferometers in m stages, where $N=2^m$.

4.2.5 De-Multiplexing and Optical Filtering

When incoherent detection is used, a mechanism is needed to select one of the wavelength channels optically before photodetection. In general, this can be done in either the spatial domain, frequency domain, or polarization domain. In the spatial domain, wavelength channels are split into different spatial paths for detection. In the frequency domain, either an optical band-pass filter is used to suppress out-of-band channels or a narrow-band optical amplifier is used to amplify only the in-band channels. In the polarization domain, a polarizing beam splitter separates a wavelength channel if its polarization is orthogonal to that of other wavelength channels.

As mentioned earlier, there are two design criteria in de-multiplexing wavelength channels: a large free spectral range (FSR) and a high spectral resolution. The fundamental reason that two wavelengths at one FSR separation have the same de-multiplexed output is because wave interference has a periodic effect in phase. To avoid ambiguity, a large FSR is required. In general, the FSR of a de-multiplexer is inversely proportional to the period of its grating or diffracting structure.

The spectral resolution of a de-multiplexer, on the other hand, determines how closely two wavelength channels can be separated. In general, the resolution is inversely proportional to the device size that introduces the wave interference. This can be understood from the Fourier transform, which says that the larger the size of an object in the spatial domain, the smaller the width of its Fourier transform in spatial spectral domain.

4.3 Photonic Switching

Switching is a network function that routes traffic from one channel to another. To provide high switching throughputs and to avoid the electronics speed bottleneck, photonic switches that implement the I/O interface and switching fabric in optics, as shown in figure 4.8, have been actively studied over the last decade. For fast packet switching such as in B-ISDN/ATM that needs high speed switching control, self-routing photonic switches that integrate switching control with switching fabrics in optics have also been demonstrated.

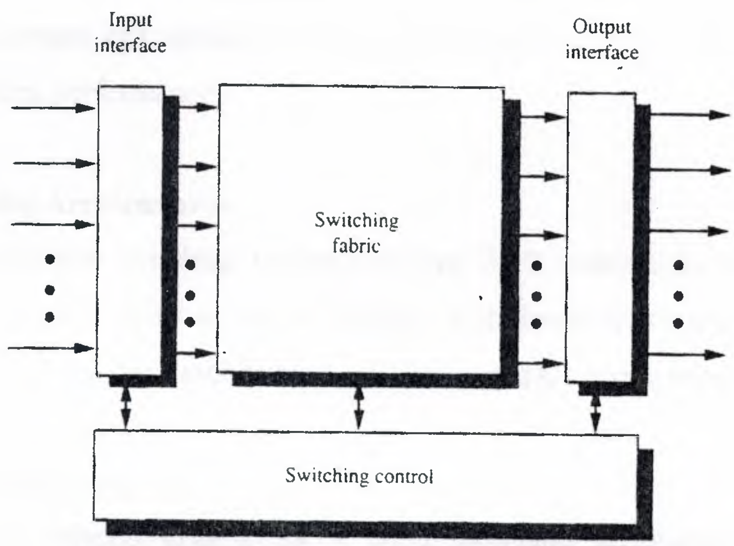

Figure 4.8 A block diagram of a switch.

To simplify implementation of large photonic switches, 2-to-2 switches can be used as the building block. In general, with a multistage interconnection architecture, any switch size of 2^k can be built. In addition, with a good interconnection architecture, the number of 2-to-2 switching elements can be reduced while maintaining good switching performance, such as zero blocking probability.

Similar to multiple access, photonic switching can be done in different domains such as space, time, wavelength and polarization, where traffic is switched from one spatial line, time slot, wavelength channel, or polarization state to another. These different domain switchings have different system characteristics and require different physical implementation. For example, spatial domain switching can configure different physical light paths but require complete spatial interconnections. Wavelength domain switching, on the other hand, is attractive to WDoMA but requires precise wavelength control and tuning.

Performing switching in multiple domains can reduce its complexity. For example, by performing time-space-time (TST) switching, the number of spatial cross-points can be greatly reduced. In WDoMA network, spatial switching can be similarly combined with wavelength channel interchange (WCI) to form a large switch. In general, switching in different domains and their combinations can be logically equivalent. For example, WCI is logically equivalent to time-slot interchange (TSI). As a result, the most convenient technology can be used to implement the same logical switching architecture and similar switching control algorithm can be used to achieve the same switching performance.

4.3.1 Switching Architectures

Some important switching architectures that 2×2 switches as the basic building block. Although 2×2 switches can be realized in different technologies, they can be interconnected with the same architecture and similar system considerations.

(a) System considerations

There are several criteria for a good switching architecture from system considerations. First, for a given switch size, N, the number of cross-points for 2×2 switches should be as small as possible. When the number is large, implementation is expensive and the optical path is subject to large power loss and cross talk. As a result, one primary design objective is to minimize the number of cross-points.

To evaluate a switch design from its number of cross-points, it can be compared with the lower bound, $\log_2(N!)$. This lower bound can be obtained from the following consideration. Because a switch works as a permutator that routes inputs to outputs according to a certain permutation, it needs to provide at least $N!$ different configurations. Because each cross-point can provide two different configurations, a minimum number of $\log_2(N!)$ cross-points are needed to configure $N!$ different routings or permutations.

Many switching architectures have been designed to minimize the number of cross-points. The three-stage Clos switch is one example. Another switching architecture, called the Benes network, can approach the lower bound asymptotically. If internal blockings are acceptable, using a self-routing network can further reduce the number of cross-points.

Optical paths should go through a minimal and equal number of cross-points. Because each cross-point introduces a certain power loss and cross talk, a large number of cross-points along a configured path require a large incident light power. Many multistage switches have thus been designed to have as few as $\log_2 N$ cross-points along any configured path. Furthermore, to reduce power variation at the switch output and to avoid the near—far problem, optical paths need to have the same number of cross-points. When a switch is designed to reduce the number of cross-points in total and in each configured path, it can have a large internal blocking probability. In a multistage switch, internal blocking occurs when there is no available route from the existing configuration. To reduce the blocking probability, the number of possible routes between inputs and outputs can be increased. The Clos switch, for example, can have a zero blocking probability when the number of middle-stage switches, which equals the number of possible routes, is large enough.

In some switching architectures, the internal blocking probability can be completely reduced to zero by using a good switching control or rearranging the current switching configuration. These cases are called wide-sense non-blocking and re-arrangeably non-blocking, respectively. These different non-blocking conditions are defined below.

Strictly Non-blocking: A switch is strictly non-blocking if a connection path can always be found no matter what the current switching configuration is or what switching control algorithm is used.

Wide-Sense Non-blocking: A switch is wide-sense non-blocking if a connection path can always be found no matter what the current switching configuration is if a *good* switching control algorithm is used. In other words, the switch can have internal blockings if a poor switching control is used but will have no blocking if a good switching control is used.

Re-arrangeably Non-blocking: A switch is re-arrangeably non-blocking if a connection path can be found by rearranging the existing switching configuration.

4.3.2 Spatial Domain Photonic Switching

In spatial switching, physical optical paths between input ports and output ports are configured according to the requested connections. Three types of spatial switching elements are then considered: mechanical switches, directional couplers, and bistable optical gates. These switches configure optical paths by moving optical fibers, changing the optical coupling characteristics of wave-guides, and modulating the states of bistable gates, respectively.

(a) Mechanical switches

A mechanical switch configures light propagation paths through a mechanical mechanism. As shown in figure 3.12, an array of 1-to-2 photonic switches has a movable portion so that input fibers can be coupled to one of the two output fiber arrays. In general, multiple 1-to-2 and 2-to-1 switching elements can be interconnected to form a larger switch.

Because there is no need for advanced optics, a mechanical switch is attractive for its simple implementation. As a result, it is good for use in cross-connects and automatic protection switches (APSs) where the slow mechanical speed is not a concern. Mechanical switches have some limitations.

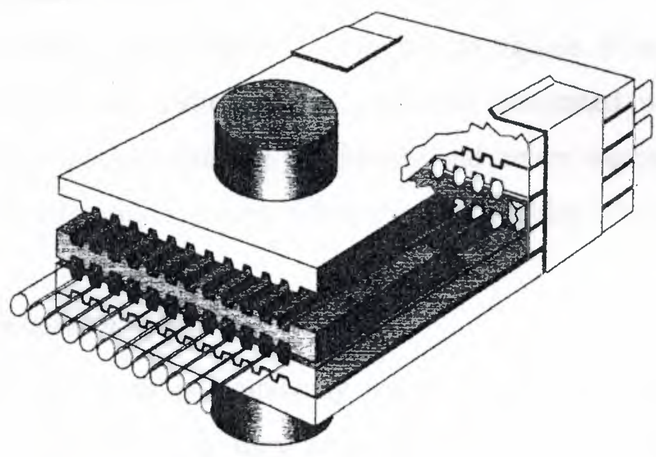

Figure 4.9 A mechanical photonic switch connecting a fiber array to one of the two output arrays.

When an input fiber is switched to different fiber, for example, there can be larger coupling loss from misalignment.

(b) Waveguide switches

A more popular switch design is based on the directional couplers depicted in Figure 4.10. As shown, coupled wave-guides provide the mechanism to switch incident light from one wave-guide to another. An external voltage can be used to modulate the refractive index of the wave-guides through the EO effect, which can in turn modulate the light coupling between the wave-guides.

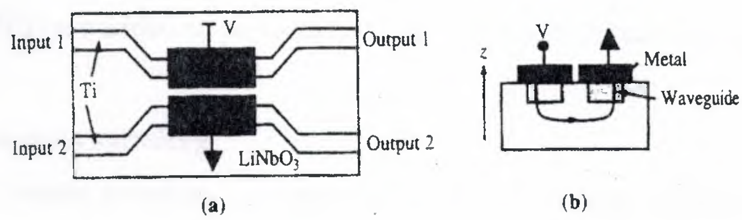

Figure 4.10 A 2-to-2 waveguide switch using directional couplers.

4.3.3 Multidimensional Photonic Switching

In addition to spatial domain switching, traffic can be switched in the wavelength domain and time domain. To switch light signals from different physical lines, these different domain switching are commonly integrated with spatial domain switching. For lower spatial switching complexity and better match with TDoMA or WDoMA traffic, this integration into multidimensional switching is attractive.

(a) Wavelength domain switching

Wavelength switching is commonly used in WDoMA. According to how wavelength channels are switched, there are two primary types: *broadcast and select* and *wavelength routing*. As shown in figure 4.11, the first type uses a star coupler to mix all wavelength inputs and broadcast them to all output ports.

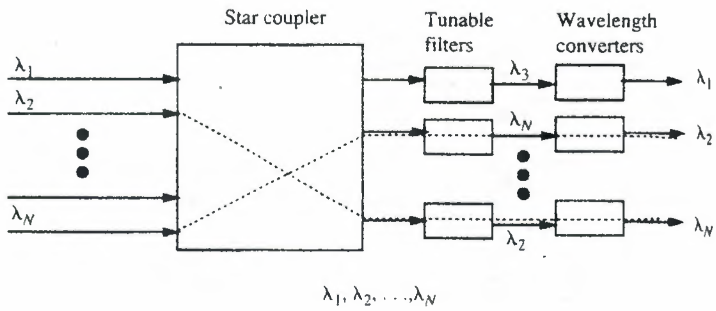

Figure 4.11 A broadcast-and-select wavelength switch.

Using optical filters at the star coupler outputs allows non-blocking wavelength switching. To allow several wavelength switches to be used in series, wavelength converters (WCs) are added to perform necessary wavelength permutation.

(b) Time domain switching

Time domain switching is attractive to TDoMA where traffic is multiplexed in the time domain. Because photons cannot be easily stored and retrieved after a programmable delay, implementation of time domain switching or time slot interchange is not an easy task. A programmable time delay line that uses a fiber loop and a 2×2 photonic switch is shown in figure 4.12a. By designing a unit delay T in the fiber loop, a programmable delay of multiple $T's$ can be achieved by changing the state of the 2×2 switch. To do this, the switch is first set in the CROSS state for a duration of T to switch the input packet to the delay loop. After that, the switch is set in the BAR state for duration of $(k-1)T$, which then keeps the packet for another $(k-1)T$ delay. At the end, the switch is reset to the CROSS state, and a total delay of kT is introduced.

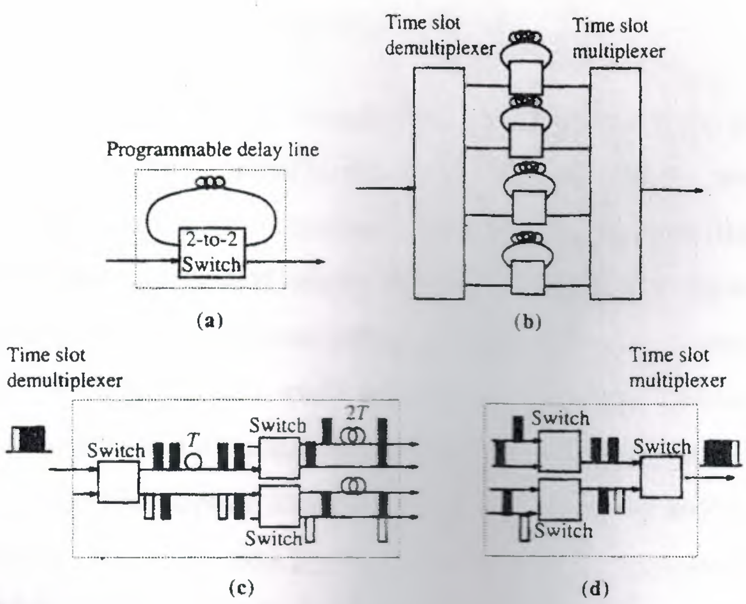

Figure 4.12 An implementation of a photonic time-slot change. (a) a basic programmable delay line, (b) the TSI, (c) implementation of a time-slot de-multiplexer, (d) implementation of a time-slot multiplexer.

CONCLUSION

The development of optical communication technology has passed through several generations. The two primary objectives of these efforts have been larger transmission capacity and longer transmission distance. To improve the capacity and distance, higher output power and smaller fiber attenuation are essential to longer transmission distance, and good spectral coherence is key to higher transmission speeds. Therefore, most efforts have been made to improve the output power and spectral coherence of light sources and to reduce fiber attenuation and dispersion.

In early times, when optical communications was not well known, signals were detected incoherently. But as the time passes coherent detection was used to enhance the receiver's sensitivity. With coherent detection, received signals are amplified by the local carrier, which makes the system performance limited by shot noise.

To eliminate the attenuation and dispersion limits, optical amplifiers have been developed. Optical amplifiers amplify optical signals directly in the optical domain. Now a days, research in optical amplifiers have become sufficiently mature, and repeaterless systems have been demonstrated over hundreds of kilometers. Research in nonlinear soliton technology has demonstrated negligible dispersion over thousands of kilometers.

Optical fiber communication has been essentially used in point-to-point long distance transmission such as telephone networks. As the technology advances, it also becomes attractive to use optical fiber transmission for networking applications. For example, optical communication has been used for local area networks such as fiber distributed data interface (FDDI) and various wavelength division multiple access (WDMA) networks. In space modulators for high data rate laser links are being used. Indeed, at the turn of the century, literally thousands of intersatellite links - radio-frequency (RF) and optical - are expected to be in operation in commercial multi-satellite constellations providing mobile communications, video conferencing and multimedia services. Optical technology offers too many advantages in terms of mass, power, system flexibility and cost, to leave the field entirely to RF. In the near future, we will see optical fibers wired near or even in our homes. These are the so-called fiber-in-the-loop and fiber-to-the-home.

REFERENCES

[1] John M. Senior, *"Optical Fiber Communication Principles and Practice"*, 2nd ed., Prentice-Hall Inc., 1992.

[2] Wim van Etten, Jan van der Plaats, *"Fundamentals of Optical Fiber Communications"*, Prentice-Hall Inc., 1991.

[3] Max Ming, Kang Liu, *"Principles and Applications of Optical Communications"*, IRWIN, 1996

[4] John Wilson, John Hawkes, *"Optoelectronics"*, 3rd ed., Prentice-Hall Inc., 1998

[5] Martin S. Roden, *"Analog and Digital Communication Systems"*, 3rd ed., Prentice-Hall Inc., 1991

[6] Simon Haykin, *"Digital Commumuications"*, John Wiley & sons Inc., 1988