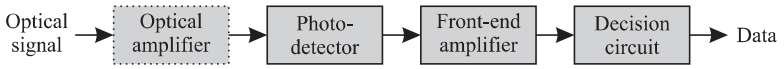

Figure 3.61 Block diagram of a receiver in a digital communication system.

3.6 Detectors

A receiver converts an optical signal into a usable electrical signal. Figure 3.61 shows the different components within a receiver. The *photodetector* generates an electrical current proportional to the incident optical power. The *front-end amplifier* increases the power of the generated electrical signal to a usable level. In digital communication systems, the front-end amplifier is followed by a *decision circuit* that estimates the data from the output of the front-end amplifier. The design of this decision circuit depends on the modulation scheme used to transmit the data and will be discussed in Section 4.4. An optical amplifier may be optionally placed before the photodetector to act as a *preamplifier.* The performance of optically preamplified receivers will be discussed in Chapter 4. This section covers photodetectors and front-end amplifiers.

3.6.1 Photodetectors

The basic principle of photodetection is illustrated in Figure 3.62. Photodetectors are made of semiconductor materials. Photons incident on a semiconductor are absorbed by electrons in the valence band. As a result, these electrons acquire higher energy and are excited into the conduction band, leaving behind a hole in the valence band. When an external voltage is applied to the semiconductor, these electron-hole pairs give rise to an electrical current, termed the *photocurrent.*

It is a principle of quantum mechanics that each electron can absorb only one photon to transit between energy levels. Thus the energy of the incident photon must be at least equal to the bandgap energy in order for a photocurrent to be generated. This is also illustrated in Figure 3.62. This gives us the following constraint on the frequency f_c or the wavelength λ at which a semiconductor material with bandgap E_g can be used as a photodetector:

$$hf_c = \frac{hc}{\lambda} \geq eE_g. \tag{3.19}$$

Here, c is the velocity of light, and e is the electronic charge.

The largest value of λ for which (3.19) is satisfied is called the *cutoff wavelength* and is denoted by λ_{cutoff}. Table 3.2 lists the bandgap energies and the corresponding

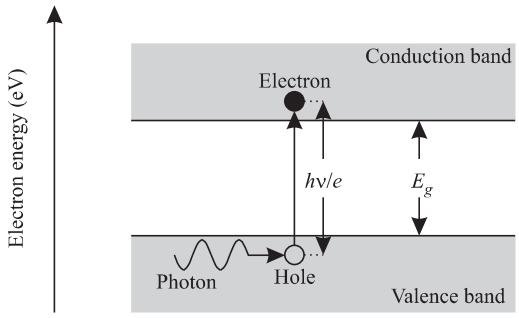

Figure 3.62 The basic principle of photodetection using a semiconductor. Incident photons are absorbed by electrons in the valence band, creating a free or mobile electron-hole pair. This electron-hole pair gives rise to a photocurrent when an external voltage is applied.

cutoff wavelengths for a number of semiconductor materials. We see from this table that the well-known semiconductors silicon (Si) and gallium arsenide (GaAs) cannot be used as photodetectors in the 1.3 and 1.55 μm bands. Although germanium (Ge) can be used to make photodetectors in both these bands, it has some disadvantages that reduce its effectiveness for this purpose. The new compounds indium gallium arsenide (InGaAs) and indium gallium arsenide phosphide (InGaAsP) are commonly used to make photodetectors in the 1.3 and 1.55 μm bands. Silicon photodetectors are widely used in the 0.8 μm band.

The fraction of the energy of the optical signal that is absorbed and gives rise to a photocurrent is called the *efficiency* η of the photodetector. For transmission at high bit rates over long distances, optical energy is scarce, and thus it is important to design the photodetector to achieve an efficiency η as close to 1 as possible. This can be achieved by using a semiconductor slab of sufficient thickness. The power absorbed by a semiconductor slab of thickness L μm can be written as

$$P_{\text{abs}} = (1 - e^{-\alpha L}) P_{\text{in}}, \tag{3.20}$$

where P_{in} is the incident optical signal power, and α is the absorption coefficient of the material; therefore,

$$\eta = \frac{P_{\text{abs}}}{P_{\text{in}}} = 1 - e^{-\alpha L}. \tag{3.21}$$

The absorption coefficient depends on the wavelength and is zero for wavelengths $\lambda > \lambda_{\text{cutoff}}$. Thus a semiconductor is transparent to wavelengths greater than its cutoff

Table 3.2 Bandgap energies and cutoff wavelengths for a number of semiconductor materials. $In_{1-x}Ga_xAs$ is a ternary compound semiconductor material where a fraction $1-x$ of the Ga atoms in GaAs are replaced by In atoms. $In_{1-x}Ga_xAs_yP_{1-y}$ is a quaternary compound semiconductor material where, in addition, a fraction $1-y$ of the As atoms are replaced by P atoms. By varying x and y, the bandgap energies and cutoff wavelengths can be varied.

Material	E_g (eV)	λ_{cutoff} (μm)
Si	1.17	1.06
Ge	0.775	1.6
GaAs	1.424	0.87
InP	1.35	0.92
$In_{0.55}Ga_{0.45}As$	0.75	1.65
$In_{1-0.45y}Ga_{0.45y}As_yP_{1-y}$	0.75–1.35	1.65–0.92

wavelength. Typical values of α are on the order of 10^4/cm, so to achieve an efficiency $\eta > 0.99$, a slab of thickness on the order of 10 μm is needed. The area of the photodetector is usually chosen to be sufficiently large so that all the incident optical power can be captured by it. Photodetectors have a very wide operating bandwidth since a photodetector at some wavelength can also serve as a photodetector at all smaller wavelengths. Thus a photodetector designed for the 1.55 μm band can also be used in the 1.3 μm band.

Photodetectors are commonly characterized by their *responsivity* $\mathcal{R}$. If a photodetector produces an average current of I_p amperes when the incident optical power is P_{in} watts, the responsivity

$$\mathcal{R} = \frac{I_p}{P_{\text{in}}} \text{ A/W}.$$

Since an incident optical power P_{in} corresponds to an incidence of P_{in}/hf_c photons/s on the average, and a fraction η of these incident photons are absorbed and generate an electron in the external circuit, we can write

$$\mathcal{R} = \frac{e\eta}{hf_c} \text{ A/W}.$$

The responsivity is commonly expressed in terms of λ; thus

$$\mathcal{R} = \frac{e\eta\lambda}{hc} = \frac{\eta\lambda}{1.24} \text{ A/W},$$

where λ in the last expression is expressed in μm. Since η can be made quite close to 1 in practice, the responsivities achieved are on the order of 1 A/W in the 1.3 μm band and 1.2 A/W in the 1.55 μm band.

In practice, the mere use of a slab of semiconductor as a photodetector does not realize high efficiencies. This is because many of the generated conduction band electrons recombine with holes in the valence band before they reach the external circuit. Thus it is necessary to sweep the generated conduction band electrons rapidly out of the semiconductor. This can be done by imposing an electric field of sufficient strength in the region where the electrons are generated. This is best achieved by using a semiconductor *pn*-junction (see Section 3.4.5) instead of a homogeneous slab and applying a *reverse-bias* voltage (positive bias to the *n*-type and negative bias to the *p*-type) to it, as shown in Figure 3.63. Such a photodetector is called a *photodiode*.

The depletion region in a *pn*-junction creates a built-in electric field. Both the depletion region and the built-in electric field can be enhanced by the application of a reverse-bias voltage. In this case, the electrons that are generated by the absorption of photons within or close to the depletion region will be swept into the *n*-type semiconductor before they recombine with the holes in the *p*-type semiconductor. This process is called *drift* and gives rise to a current in the external circuit. Similarly, the generated holes in or close to the depletion region drift into the *p*-type semiconductor because of the electric field.

Electron-hole pairs that are generated far away from the depletion region travel primarily under the effect of diffusion and may recombine without giving rise to a current in the external circuit. This reduces the efficiency η of the photodetector. More importantly, since diffusion is a much slower process than drift, the *diffusion current* that is generated by these electron-hole pairs will not respond quickly to changes in the intensity of the incident optical signal, thus reducing the frequency response of the photodiode.

pin Photodiodes

To improve the efficiency of the photodetector, a very lightly doped *intrinsic* semiconductor is introduced between the *p*-type and *n*-type semiconductors. Such photodiodes are called *pin* photodiodes, where the *i* in *pin* is for intrinsic. In these photodiodes, the depletion region extends completely across this intrinsic semiconductor (or region). The width of the *p*-type and *n*-type semiconductors is small compared to the intrinsic region, so that much of the light absorption takes place in this region. This increases the efficiency and thus the responsivity of the photodiode.

A more efficient method of increasing the responsivity is to use a semiconductor material for the *p*-type and *n*-type regions that is *transparent* at the wavelength

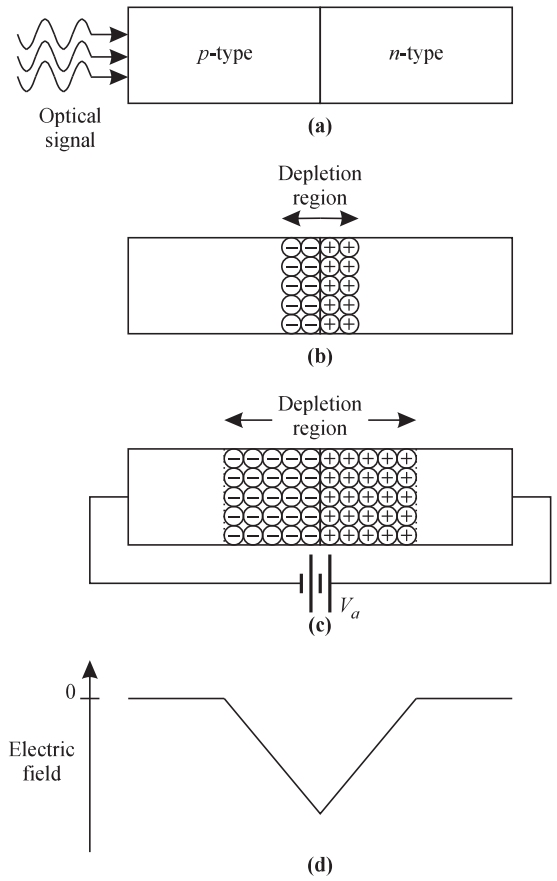

Figure 3.63 A reverse-biased *pn*-junction used as a photodetector. (a) A *pn*-junction photodiode. (b) Depletion region with no bias voltage applied. (c) Depletion region with a reverse-bias voltage, V_a. (d) Built-in electric field on reverse bias.

of interest. Thus the wavelength of interest is larger than the cutoff wavelength of this semiconductor, and no absorption of light takes place in these regions. This is illustrated in Figure 3.64, where the material InP is used for the *p*-type and *n*-type regions, and InGaAs for the intrinsic region. Such a *pin* photodiode structure is termed a *double heterojunction* or a *heterostructure* since it consists of two junctions of completely different semiconductor materials. From Table 3.2, we see that the cutoff wavelength for InP is 0.92 μm and that for InGaAs is 1.65 μm. Thus the

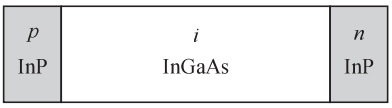

Figure 3.64 A *pin* photodiode based on a heterostructure. The *p*-type and *n*-type regions are made of InP, which is transparent in the 1.3 and 1.55 μm wavelength bands. The intrinsic region is made of InGaAs, which strongly absorbs in both these bands.

p-type and *n*-type regions are transparent in the 1.3–1.6 μm range, and the diffusion component of the photocurrent is completely eliminated.

Avalanche Photodiodes

The responsivities of the photodetectors we have described thus far have been limited by the fact that one photon can generate only one electron when it is absorbed. However, if the generated electron is subjected to a very high electric field, it can acquire sufficient energy to knock off more electrons from the valence band to the conduction band. These secondary electron-hole pairs can generate even further electron-hole pairs when they are accelerated to sufficient levels. This process is called *avalanche multiplication.* Such a photodiode is called an *avalanche photodiode,* or simply an *APD.*

The number of secondary electron-hole pairs generated by the avalanche multiplication process by a single (primary) electron is random, and the mean value of this number is termed the *multiplicative gain* and denoted by G_m. The multiplicative gain of an APD can be made quite large and even infinite—a condition called *avalanche breakdown.* However, a large value of G_m is also accompanied by a larger variance in the generated photocurrent, which adversely affects the noise performance of the APD. Thus there is a trade-off between the multiplicative gain and the noise factor. APDs are usually designed to have a moderate value of G_m that optimizes their performance. We will study this issue further in Section 4.4.

3.6.2 Front-End Amplifiers

Two kinds of front-end amplifiers are used in optical communication systems: the *high-impedance* front end and the *transimpedance* front end. The equivalent circuits for these amplifiers are shown in Figure 3.65.

The capacitances C in this figure include the capacitance due to the photodiode, the amplifier input capacitance, and other parasitic capacitances. The main design issue is the choice of the load resistance R_L. We will see in Chapter 4 that the *thermal*

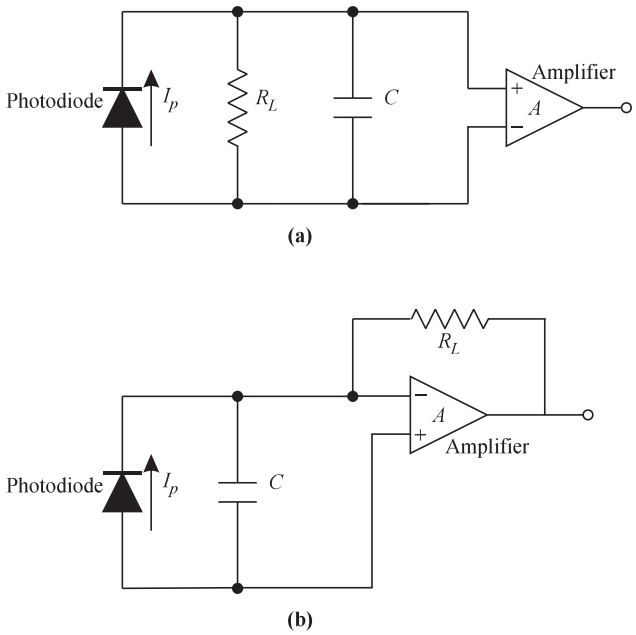

Figure 3.65 (a) Equivalent circuit for a high-impedance front-end amplifier. (b) Equivalent circuit for a transimpedance front-end amplifier.

noise current that arises due to the random motion of electrons and contaminates the photocurrent is inversely proportional to the load resistance. Thus, to minimize the thermal noise, we must make R_L large. However, the bandwidth of the photodiode, which sets the upper limit on the usable bit rate, is inversely proportional to the output load resistance seen by the photodiode, say, R_p. First consider the high-impedance front end. In this case, $R_p = R_L$, and we must choose R_L small enough to accommodate the bit rate of the system. Thus there is a trade-off between the bandwidth of the photodiode and its noise performance. Now consider the transimpedance front end for which $R_p = R_L/(A + 1)$, where A is the gain of the amplifier. The bandwidth is increased by a factor of $A + 1$ for the same load resistance. However, the thermal noise current is also higher than that of a high-impedance amplifier with the same R_L (due to considerations beyond the scope of this book), but this increase is quite moderate—a factor usually less than two. Thus the transimpedance front end is chosen over the high-impedance one for most optical communication systems.

There is another consideration in the choice of a front-end amplifier: *dynamic range*. This is the difference between the largest and smallest signal levels that the

front-end amplifier can handle. This may not be an important consideration for many optical communication links since the power level seen by the receivers is usually more or less fixed. However, dynamic range of the receivers is a very important consideration in the case of networks where the received signal level can vary by a few orders of magnitude, depending on the location of the source in the network. The transimpedance amplifier has a significantly higher dynamic range than the high-impedance one, and this is another factor in favor of choosing the transimpedance amplifier. The higher dynamic range arises because large variations in the photocurrent I_p translate into much smaller variations at the amplifier input, particularly if the amplifier gain is large. This can be understood with reference to Figure 3.65(b). A change ΔI_p in the photocurrent causes a change in voltage $\Delta I_p R_L$ across the resistance R_L (ignoring the current through the capacitance C). This results in a voltage change across the inputs of the amplifier of only $\Delta I_p R_L/(A+1)$. Thus if the gain, A, is large, this voltage change is small. In the case of the high-impedance amplifier, however, the voltage change across the amplifier inputs would be $\Delta I_p R_L$ (again ignoring the current through the capacitance C).

A *field-effect transistor* (FET) has a very high input impedance and for this reason is often used as the amplifier in the front end. A *pin* photodiode and an FET are often integrated on the same semiconductor substrate, and the combined device is called a *pinFET*.

3.7 Switches

Optical switches are used in optical networks for a variety of applications. The different applications require different switching times and number of switch ports, as summarized in Table 3.3. One application of optical switches is in the *provisioning* of lightpaths. In this application, the switches are used inside wavelength crossconnects to reconfigure them to support new lightpaths. In this application, the switches are replacements for manual fiber patch panels, but with significant added software for end-to-end network management, a subject that we will cover in detail in Chapters 8 and 9. Thus, for this application, switches with millisecond switching times are acceptable. The challenge here is to realize large switch sizes.

Another important application is that of *protection switching*, the subject of Chapter 9. Here the switches are used to switch the traffic stream from a primary fiber onto another fiber in case the primary fiber fails. The entire operation must typically be completed in several tens of milliseconds, which includes the time to detect the failure, communicate the failure to the appropriate network elements handling the switching, and the actual switch time. Thus the switching time required is on the order of a few milliseconds. Different types of protection switching are

Table 3.3 Applications for optical switches and their switching time and port count requirements.

Application	Switching Time Required	Number of Ports
Provisioning	1–10 ms	> 1000
Protection switching	1–10 ms	2–1000
Packet switching	1 ns	> 100
External modulation	10 ps	1

possible, and based on the scheme used, the number of switch ports needed may vary from two ports to several hundreds to thousands of ports when used in a wavelength crossconnect.

Switches are also important components in high-speed optical *packet-switched* networks. In these networks, switches are used to switch signals on a packet-by-packet basis. For this application, the switching time must be much smaller than a packet duration, and large switches will be needed. For example, ordinary Ethernet packets have lengths between about 60 to 1500 bytes. At 10 Gb/s, the transmission time of a 60-byte packet is 48 ns. Thus, the switching time required for efficient operation is on the order of a few nanoseconds. Optical packet switching is the subject of Chapter 12.

Yet another use for switches is as external modulators to turn on and off the data in front of a laser source. In this case, the switching time must be a small fraction of the bit duration. So an external modulator for a 10 Gb/s signal (with a bit duration of 100 ps) must have a switching time (or, equivalently, a rise and fall time) of about 10 ps.

In addition to the switching time and the number of ports, the other important parameters used to characterize the suitability of a switch for optical networking applications are the following:

1. The *extinction ratio* of an on-off switch is the ratio of the output power in the on state to the output power in the off state. This ratio should be as large as possible and is particularly important in external modulators. Whereas simple mechanical switches have extinction ratios of 40–50 dB, high-speed external modulators tend to have extinction ratios of 10–25 dB.
2. The *insertion loss* of a switch is the fraction of power (usually expressed in decibels) that is lost because of the presence of the switch and must be as small as possible. Some switches have different losses for different input-output connections. This is an undesirable feature because it increases the dynamic range of the

signals in the network. With such switches, we may need to include variable optical attenuators to equalize the loss across different paths. This *loss uniformity* is determined primarily by the architecture used to build the switch, rather than the inherent technology itself, as we will see in several examples below.

3. Switches are not ideal. Even if input x is nominally connected to output y, some power from input x may appear at the other outputs. For a given switching state or interconnection pattern, and output, the *crosstalk* is the ratio of the power at that output from the desired input to the power from all other inputs. Usually, the *crosstalk of a switch* is defined as the worst-case crosstalk over all outputs and interconnection patterns.
4. As with other components, switches should have a low polarization-dependent loss (PDL). When used as external modulators, polarization dependence can be tolerated since the switch is used immediately following the laser, and the laser's output state of polarization can be controlled by using a special polarization-preserving fiber to couple the light from the laser into the external modulator.
5. A *latching* switch maintains its switch state even if power is turned off to the switch. This is a somewhat desirable feature because it enables traffic to be passed through the switch even in the event of power failures.
6. The switch needs to have a readout capability wherein its current state can be monitored. This is important to verify that the right connections are made through the switch.
7. The reliability of the switch is an important factor in telecommunications applications. The common way of establishing reliability is to cycle the switch through its various states a large number of times, perhaps a few million cycles. However, in the provisioning and protection-switching applications discussed above, the switch remains in one state for a long period, say, even a few years, and is then activated to change state. The reliability issue here is whether the switch will actually switch after it has remained untouched for a long period. This property is more difficult to establish without a long-term history of deployment.

3.7.1 Large Optical Switches

Switches with port counts ranging from a few hundred to a few thousand are being sought by carriers for their next-generation networks. Given that a single central office handles multiple fibers, with each fiber carrying several tens to hundreds of wavelengths, it is easy to imagine the need for large-scale switches to provision and

protect these wavelengths. We will study the use of such switches as wavelength crossconnects in Chapter 7.

The main considerations in building large switches are the following:

Number of switch elements required. Large switches are made by using multiple switch elements in some form or the other, as we will see below. The cost and complexity of the switch to some extent depends on the number of switch elements required. However, this is only one of the factors that affects the cost. Other factors include packaging, splicing, and ease of fabrication and control.

Loss uniformity. As we mentioned in the context of switch characteristics earlier, switches may have different losses for different combinations of input and output ports. This situation is exacerbated for large switches. A measure of the loss uniformity can be obtained by considering the minimum and maximum number of switch elements in the optical path, for different input and output combinations.

Number of crossovers. Some of the optical switches that we will study next are fabricated by integrating multiple switch elements on a single substrate. Unlike integrated electronic circuits (ICs), where connections between the various components can be made at multiple layers, in integrated optics, all these connections must be made in a single layer by means of waveguides. If the paths of two waveguides cross, two undesirable effects are introduced: power loss and crosstalk. In order to have acceptable loss and crosstalk performance for the switch, it is thus desirable to minimize, or completely eliminate, such waveguide crossovers. Crossovers are not an issue with respect to free-space switches, such as the MEMS switches that we will describe later in this section.

Blocking characteristics. In terms of the switching function achievable, switches are of two types: *blocking* or *nonblocking*. A switch is said to be *nonblocking* if an unused input port can be connected to any unused output port. Thus a nonblocking switch is capable of realizing every interconnection pattern between the inputs and the outputs. If some interconnection pattern(s) cannot be realized, the switch is said to be *blocking*. Most applications require nonblocking switches. However, even nonblocking switches can be further distinguished in terms of the effort needed to achieve the nonblocking property. A switch is said to be *wide-sense nonblocking* if any unused input can be connected to any unused output, without requiring any existing connection to be rerouted. Wide-sense nonblocking switches usually make use of specific routing algorithms to route connections so that future connections will not be blocked. A *strict-sense nonblocking* switch allows any unused input to be connected to any unused output regardless of how previous connections were made through the switch.

Table 3.4 Comparison of different switch architectures. The switch count for the Spanke architecture is made in terms of $1 \times n$ switches, whereas 2×2 switches are used for the other architectures.

	Nonblocking Type	No. Switches	Max. Loss	Min. Loss
Crossbar	Wide sense	n^2	$2n-1$	1
Clos	Strict sense	$4\sqrt{2}n^{1.5}$	$5\sqrt{2n}-5$	3
Spanke	Strict sense	$2n$	2	2
Beneš	Rearrangeable	$\frac{n}{2}(2\log_2 n - 1)$	$2\log_2 n - 1$	$2\log_2 n - 1$
Spanke-Beneš	Rearrangeable	$\frac{n}{2}(n-1)$	n	$\frac{n}{2}$

A nonblocking switch that may require rerouting of connections to achieve the nonblocking property is said to be *rearrangeably nonblocking*. Rerouting of connections may or may not be acceptable depending on the application since the connection must be interrupted, at least briefly, in order to switch it to a different path. The advantage of rearrangeably nonblocking switch architectures is that they use fewer small switches to build a larger switch of a given size, compared to the wide-sense nonblocking switch architectures.

While rearrangeably nonblocking architectures use fewer switches, they require a more complex control algorithm to set up connections, but this control complexity is not a significant issue, given the power of today's microprocessors used in these switches that would execute such an algorithm. The main drawback of rearrangeably nonblocking switches is that many applications will not allow existing connections to be disrupted, even temporarily, to accommodate a new connection.

Usually, there is a trade-off between these different aspects. We will illustrate this when we study different architectures for building large switches next. Table 3.4 compares the characteristics of these architectures.

Crossbar

A 4×4 crossbar switch is shown in Figure 3.66. This switch uses 16 2×2 switches, and the interconnection between inputs and outputs is achieved by appropriately setting the states of these 2×2 switches. The settings of the 2×2 switches required to connect input 1 to output 3 are shown in Figure 3.66. This connection can be viewed as taking a path through the network of 2×2 switches making up the 4×4 switch. Note that there are other paths from input 1 to output 3; however, this is the preferred path as we will see next.

The crossbar architecture is wide-sense nonblocking. To connect input i to output j, the path taken traverses the 2×2 switches in row i till it reaches column j and then

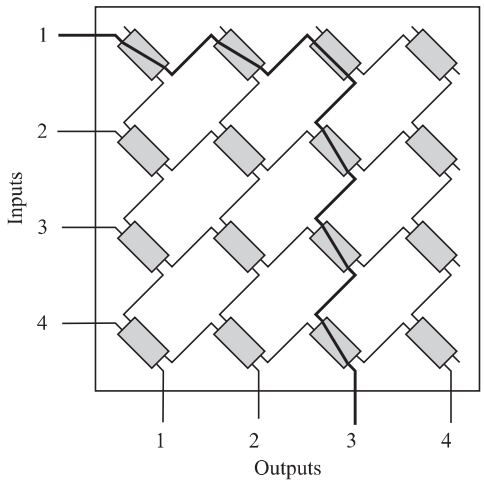

Figure 3.66 A 4 × 4 crossbar switch realized using 16 2 × 2 switches.

traverses the switches in column j till it reaches output j. Thus the 2 × 2 switches on this path in row i and column j must be set appropriately for this connection to be made. We leave it to you to be convinced that *if this connection rule is used,* this switch is nonblocking and does not require existing connections to be rerouted.

In general, an $n \times n$ crossbar requires n^2 2 × 2 switches. The shortest path length is 1 and the longest path length is $2n - 1$, and this is one of the main drawbacks of the crossbar architecture. The switch can be fabricated without any crossovers.

Clos

The Clos architecture provides a strict-sense nonblocking switch and is widely used in practice to build large port count switches. A three-stage 1024-port Clos switch is shown in Figure 3.67. An $n \times n$ switch is constructed as follows. We use three parameters, m, k, and p. Let $n = mk$. The first and third stage consist of k ($m \times p$) switches. The middle stage consists of p ($k \times k$) switches. Each of the k switches in the first stage is connected to all the switches in the middle stage. (Each switch in the first stage has p outputs. Each output is connected to the input of a different switch in the middle stage.) Likewise, each of the k switches in the third stage is connected to all the switches in the middle stage. We leave it to you to verify that if $p \geq 2m - 1$, the switch is strictly nonblocking (see Problem 3.29).

To minimize the cost of the switch, let us pick $p = 2m - 1$. Usually, the individual switches in each stage are designed using crossbar switches. Thus each of the $m \times$

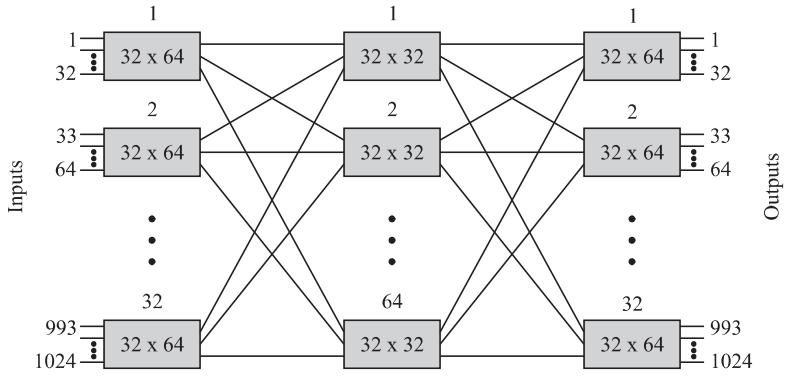

Figure 3.67 A strict-sense nonblocking 1024×1024 switch realized using 32×64 and 32×32 switches interconnected in a three-stage Clos architecture.

$(2m - 1)$ switches requires $m(2m - 1)$ 2×2 switch elements, and each of the $k \times k$ switches in the middle stage requires k^2 2×2 switch elements. The total number of switch elements needed is therefore

$$2km(2m - 1) + (2m - 1)k^2.$$

Using $k = n/m$, we leave it to you to verify that the number of switch elements is minimized when

$$m \approx \sqrt{\frac{n}{2}}.$$

Using this value for m, the number of switch elements required for the minimum cost configuration is approximately

$$4\sqrt{2}n^{3/2} - 4n,$$

which is significantly lower than the n^2 required for a crossbar.

The Clos architecture has several advantages that make it suitable for use in a multistage switch fabric. The loss uniformity between different input-output combinations is better than a crossbar, and the number of switch elements required is significantly smaller than a crossbar.

Spanke

The Spanke architecture shown in Figure 3.68 is turning out to be a popular architecture for building large switches. An $n \times n$ switch is made by combining n $1 \times n$ switches along with n $n \times 1$ switches, as shown in the figure. The architecture is

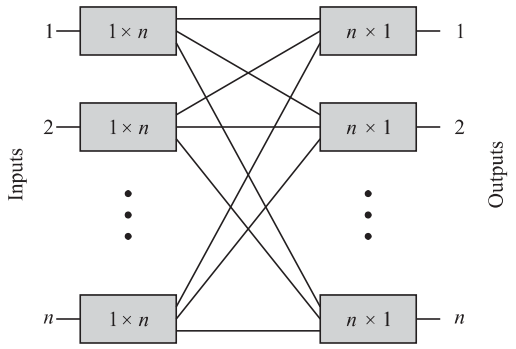

Figure 3.68 A strict-sense nonblocking $n \times n$ switch realized using $2n$ $1 \times n$ switches interconnected in the Spanke architecture.

strict-sense nonblocking. So far we have been counting the number of 2×2 switch elements needed to build large switches as a measure of the switch cost. What makes the Spanke architecture attractive is that, in many cases, a $1 \times n$ optical switch can be built using a single switch element and does not need to be built out of 1×2 or 2×2 switch elements. This is the case with the MEMS analog beam steering mirror technology that we will discuss later in this section. Therefore, only $2n$ such switch elements are needed to build an $n \times n$ switch. This implies that the switch cost scales linearly with n, which is significantly better than other switch architectures. In addition, each connection passes through two switch elements, which is significantly smaller than the number of switch elements in the path for other multistage designs. This approach provides a much lower insertion loss than the multistage designs. Moreover, the optical path length for all the input–output combinations can be made essentially the same, so that the loss is the same regardless of the specific input–output combination.

Beneš

The Beneš architecture is a rearrangeably nonblocking switch architecture and is one of the most efficient switch architectures in terms of the number of 2×2 switches it uses to build larger switches. A rearrangeably nonblocking 8×8 switch that uses only 20 2×2 switches is shown in Figure 3.69. In comparison, an 8×8 crossbar switch requires 64 2×2 switches. In general, an $n \times n$ Beneš switch requires $(n/2)(2 \log_2 n - 1)$ 2×2 switches, n being a power of two. The loss is the same through every path in the switch—each path goes through $2 \log_2 n - 1$ 2×2 switches.

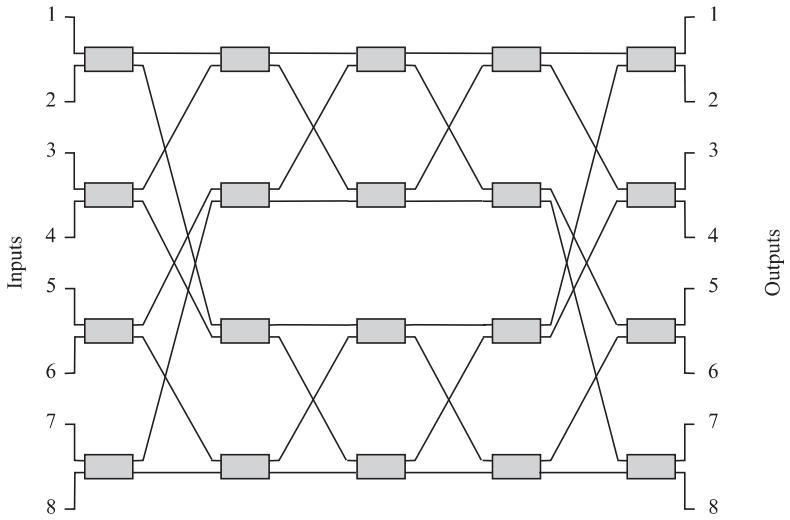

Figure 3.69 A rearrangeably nonblocking 8×8 switch realized using 20 2×2 switches interconnected in the Beneš architecture.

Its two main drawbacks are that it is not wide-sense nonblocking and that a number of waveguide crossovers are required, making it difficult to fabricate in integrated optics.

Spanke-Beneš

A good compromise between the crossbar and Beneš switch architectures is shown in Figure 3.70, which is a rearrangeably nonblocking 8×8 switch using 28 2×2 switches and *no* waveguide crossovers. This switch architecture was discovered by Spanke and Beneš [SB87] and is called the *n-stage planar architecture* since it requires n stages (columns) to realize an $n \times n$ switch. It requires $n(n-1)/2$ switches, the shortest path length is $n/2$, and the longest path length is n. There are no crossovers. Its main drawbacks are that it is not wide-sense nonblocking and the loss is nonuniform.

3.7.2 Optical Switch Technologies

Many different technologies are available to realize optical switches. These are compared in Table 3.5. With the exception of the large-scale MEMS switch, the switch elements described in the next section all use the crossbar architecture.

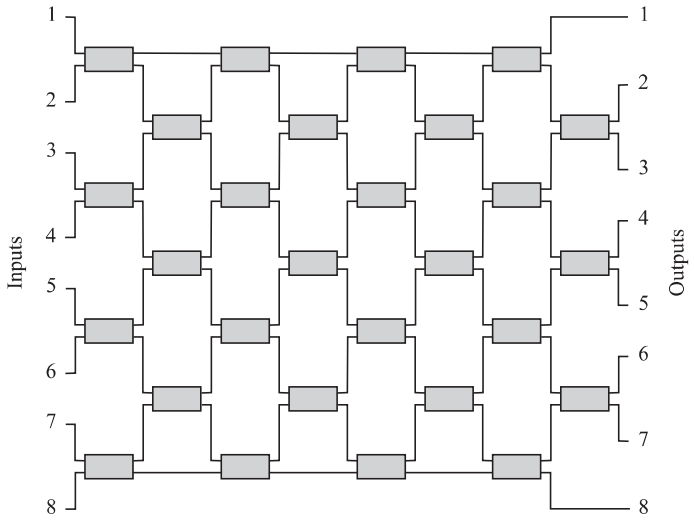

Figure 3.70 A rearrangeably nonblocking 8 × 8 switch realized using 28 2 × 2 switches and no waveguide crossovers interconnected in the *n*-stage planar architecture.

Table 3.5 Comparison of different optical switching technologies. The mechanical, MEMS, and polymer-based switches behave in the same manner for 1.3 and 1.55 μm wavelengths, but other switches are designed to operate at only one of these wavelength bands. The numbers represent parameters for commercially available switches in early 2001.

Type	Size	Loss (dB)	Crosstalk (dB)	PDL (dB)	Switching Time
Bulk mechanical	8 × 8	3	55	0.2	10 ms
2D MEMS	32 × 32	5	55	0.2	10 ms
3D MEMS	1000 × 1000	5	55	0.5	10 ms
Thermo-optic silica	8 × 8	8	40	Low	3 ms
Liquid crystal	2 × 2	1	35	0.1	4 ms
Polymer	8 × 8	10	30	Low	2 ms
Electro-optic $LiNbO_3$	4 × 4	8	35	1	10 ps
SOA	4 × 4	0	40	Low	1 ns

Bulk Mechanical Switches

In mechanical switches, the switching function is performed by some mechanical means. One such switch uses a mirror arrangement whereby the switching state is controlled by moving a mirror in and out of the optical path. Another type of mechanical switch uses a directional coupler. Bending or stretching the fiber in the interaction region changes the coupling ratio of the coupler and can be used to switch light from an input port between different output ports.

Bulk mechanical switches have low insertion losses, low PDL, and low crosstalk, and are relatively inexpensive devices. In most cases, they are available in a crossbar configuration, which implies somewhat poor loss uniformity. However, their switching speeds are on the order of a few milliseconds and the number of ports is fairly small, say, 8 to 16. For these reasons, they are particularly suited for use in small wavelength crossconnects for provisioning and protection-switching applications but not for the other applications discussed earlier. As with most mechanical components, long-term reliability for these switches is of some concern. Larger switches can be realized by cascading small bulk mechanical switches, as we saw in Section 3.7.1, but there are better ways of realizing larger port count switches, as we will explore next.

Micro-Electro-Mechanical System (MEMS) Switches

Micro-electro-mechanical systems (MEMS) are miniature mechanical devices typically fabricated using silicon substrates. In the context of optical switches, MEMS usually refers to miniature movable mirrors fabricated in silicon, with dimensions ranging from a few hundred micrometers to a few millimeters. A single silicon wafer yields a large number of mirrors, which means that these mirrors can be manufactured and packaged as arrays. Moreover, the mirrors can be fabricated using fairly standard semiconductor manufacturing processes. These mirrors are deflected from one position to another using a variety of electronic actuation techniques, such as electromagnetic, electrostatic, or piezoelectric methods, hence the name MEMS. Of these methods, electrostatic deflection is particularly power efficient but is relatively hard to control over a wide deflection range.

The simplest mirror structure is a so-called two-state pop-up mirror, or 2D mirror, shown in Figure 3.71. In one state, the mirror is flat in line with the substrate. In this state, the light beam is not deflected. In the other state, the mirror pops up to a vertical position, and the light beam, if present, is deflected. Such a mirror can be used in a crossbar arrangement discussed below to realize an $n \times n$ switch. Practical switch module sizes are limited by wafer sizes and processing constraints to be around 32×32. These switches are particularly easy to control through digital means, as only two mirror positions need to be supported.

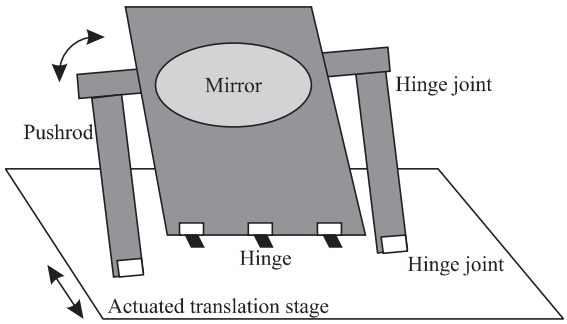

Figure 3.71 A two-state pop-up MEMS mirror, from [LGT98], shown in the popped-up position. The mirror can be moved to fold flat in its other position.

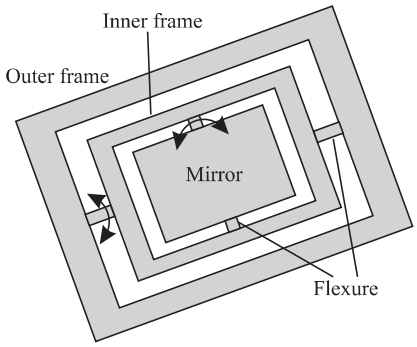

Figure 3.72 An analog beam steering mirror. The mirror can be freely rotated on two axes to deflect an incident light beam.

Another type of mirror structure is shown in Figure 3.72. The mirror is connected through flexures to an inner frame, which in turn is connected through another set of flexures to an outer frame. The flexures allow the mirror to be rotated freely on two distinct axes. This mirror can be controlled in an analog fashion to realize a continuous range of angular deflections. This type of mirror is sometimes referred to as an analog beam steering mirror, a gimbel mirror, or a 3D mirror. A mirror of this type can be used to realize a $1 \times n$ switch. The control of these mirrors is not a trivial matter, with fairly sophisticated servo control mechanisms required to deflect the mirrors to their correct position and hold them there.

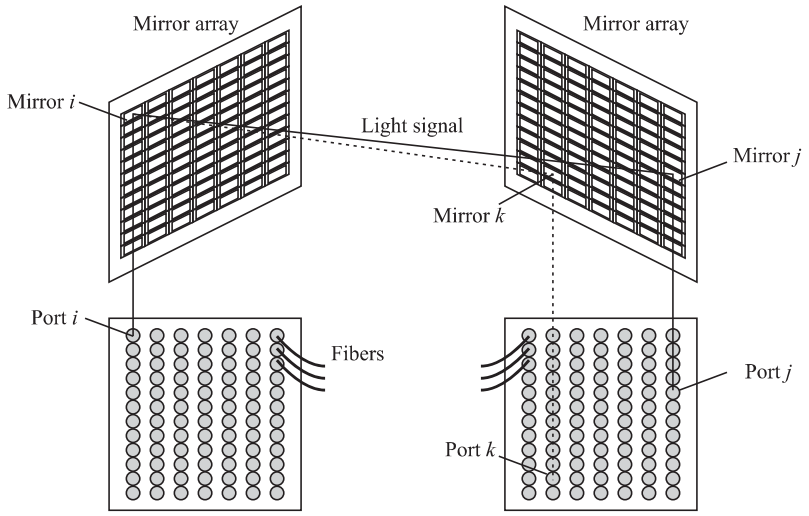

Figure 3.73 An $n \times n$ switch built using two arrays of analog beam steering MEMS mirrors.

Figure 3.73 shows a large $n \times n$ switch using two arrays of analog beam steering mirrors. This architecture corresponds to the Spanke architecture, which we discussed in Section 3.7.1. Each array has n mirrors, one associated with each switch port. An input signal is coupled to its associated mirror in the first array using a suitable arrangement of collimating lenses. The first mirror can be deflected to point the beam to any of the mirrors in the second array. To make a connection from port i to port j, the mirror i in the first array is pointed to mirror j in the second array and vice versa. Mirror j then allows the beam to be coupled out of port j. To make a connection from port i to another port, say, port k, mirror i in the first array and mirror k in the second array are pointed at each other. Note that in order to switch this connection from port i to port k, the beam is scanned from output mirror j to output mirror k, passing over other mirrors along the way. This does not lead to additional crosstalk because a connection is established only when the two mirrors are pointed at each other and not under any other circumstances. Note also that beams corresponding to multiple connections cross each other inside the switch but do not interfere.

There are two types of fabrication techniques used to make MEMS structures: *surface micromachining* and *bulk micromachining*. In surface micromachining, multiple layers are deposited on top of a silicon substrate. These layers are partially

etched away, and pieces are left anchored to the substrate to produce various structures. In bulk micromachining, the MEMS structures are crafted directly from the bulk of the silicon wafer. The type of micromachining used and the choice of the appropriate type of silicon substrate directly influence the properties of the resulting structure. For a more detailed discussion on some of the pros and cons of these approaches, see [NR01]. Today we are seeing the simple 2D MEMS mirrors realized using surface micromachining and the 3D MEMS mirrors realized using bulk micromachining.

Among the various technologies discussed in this section, the 3D MEMS analog beam steering mirror technology offers the best potential for building large-scale optical switches, for example, 256 to 1000 ports. These switches are compact, have very good optical properties (low loss, good loss uniformity, negligible dispersion), and can have extremely low power consumption. Most of the other technologies are limited to small switch sizes.

Liquid Crystal Switches

Liquid crystal cells offer another way for realizing small optical switches. These switches typically make use of polarization effects to perform the switching function. By applying a voltage to a suitably designed liquid crystal cell, we can cause the polarization of the light passing through the cell either to be rotated or not. This can then be combined with passive polarization beam splitters and combiners to yield a polarization-independent switch, as shown in Figure 3.74. The principle of operation is similar to the polarization-independent isolator of Figure 3.5. Typically, the passive polarization beam splitter, combiner, and active switch element can all be realized using an array of liquid crystal cells. The polarization rotation in the liquid crystal cell does not have to be digital in nature—it can be controlled in an analog fashion by controlling the voltage. Thus this technology can be used to realize a variable optical attentuator (VOA) as well. In fact, the VOA can be incorporated in the switch itself to control the output power being coupled out. The switching time is on the order of a few milliseconds. Like the bubble-based waveguide switch, a liquid crystal switch is a solid-state device. Thus, it can be better manufactured in volume and low cost.

Electro-Optic Switches

A 2×2 electro-optic switch can be realized using one of the external modulator configurations that we studied in Section 3.5.4. One commonly used material is lithium niobate ($LiNbO_3$). In the directional coupler configuration, the coupling ratio is varied by changing the voltage and thus the refractive index of the material in the coupling region. In the Mach-Zehnder configuration, the relative path length

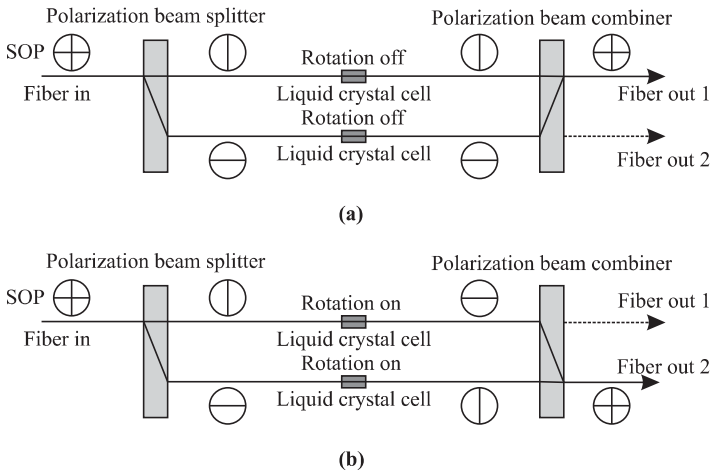

Figure 3.74 A 1 × 2 liquid crystal switch. (a) The rotation is turned off, causing the light beam to exit on output port 1. (b) The rotation is turned on by applying a voltage to the liquid crystal cell, causing the light beam to exit on output port 2.

between the two arms of the Mach-Zehnder is varied. An electro-optic switch is capable of changing its state extremely rapidly—typically, in less than 1 ns. This switching time limit is determined by the capacitance of the electrode configuration.

Among the advantages of lithium niobate switches are that they allow modest levels of integration, compared to mechanical switches. Larger switches can be realized by integrating several 2 × 2 switches on a single substrate. However, they tend to have a relatively high loss and PDL, and are more expensive than mechanical switches.

Thermo-Optic Switches

These switches are essentially 2 × 2 integrated-optic Mach-Zehnder interferometers, constructed on waveguide material whose refractive index is a function of the temperature. By varying the refractive index in one arm of the interferometer, the relative phase difference between the two arms can be changed, resulting in switching an input signal from one output port to another. These devices have been made on silica as well as polymer substrates, but have relatively poor crosstalk. Also the thermo-optic effect is quite slow, and switching speeds are on the order of a few milliseconds.

Semiconductor Optical Amplifier Switches

The SOA described in Section 3.4.5 can be used as an on-off switch by varying the bias voltage to the device. If the bias voltage is reduced, no population inversion is achieved, and the device absorbs input signals. If the bias voltage is present, it amplifies the input signals. The combination of amplification in the on state and absorption in the off state makes this device capable of achieving very large extinction ratios. The switching speed is on the order of 1 ns. Larger switches can be fabricated by integrating SOAs with passive couplers. However, this is an expensive component, and it is difficult to make it polarization independent because of the highly directional orientation of the laser active region, whose width is almost always much greater than its height (except for VCSELs).

3.7.3 Large Electronic Switches

We have focused primarily on optical switch technologies in this section. However, many of the practical "optical" or wavelength crossconnects actually use electronic switch fabrics.

Typically, a large electronic switch uses a multistage design, and in many cases, the Clos approach is the preferred approach as it provides a strict-sense nonblocking architecture with a relatively small number of crosspoint switches. Two approaches are possible. In the first approach, the input signal at 2.5 Gb/s or 10 Gb/s is converted into a parallel bit stream at a manageable rate, say, 51 Mb/s, and all the switching is done at the latter bit rate. This approach makes sense if we need to switch the signal in units of 51 Mb/s for other reasons. Also in many cases, the overall cost of an electronic switch is dominated by the cost of the optical to electrical converters, rather than the switch fabric itself. This implies that once the signal is available in the electrical domain, it makes sense to switch signals at a fine granularity.

The other approach is to design the switch to operate at the line rate in a serial fashion without splitting the signal into lower-speed bit streams. The basic unit of this serial approach is a crossbar fabricated as a single integrated circuit (IC). The practical considerations related to building larger switches using these ICs have to do with managing the power dissipation and the interconnects between switch stages. For example, suppose a 64×64 switch IC dissipates 25 W. About 100 such switches are required to build a 1024×1024 switch. The total power dissipated is therefore around 25 kW. (In contrast, a 1024×1024 optical switch using 3D MEMS may consume only about 3 kW and is significantly more compact overall, compared to an equivalent electrical switch.) Cooling such a switch is a significant problem. The other aspect has to do with the high-speed interconnect required between switch modules. As long as the switch modules are within a single printed circuit board,