

OPTICAL COMMUNICATIONS



Chris Lee
Fontys University of Applied Sciences

Optical Communications

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (<https://LibreTexts.org>) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by [NICE CXone Expert](#) and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptations contact info@LibreTexts.org or visit our main website at <https://LibreTexts.org>.

This text was compiled on 05/15/2026

TABLE OF CONTENTS

[Licensing](#)

[1: Introduction](#)

- [1.1: The long tail](#)
- [1.2: SONET/SDH](#)
- [1.3: The importance of clocks](#)

[2: Optical Waveguides](#)

- [2.1: Introduction to waveguides](#)
- [2.2: Spatial modes](#)
- [2.3: Modal dispersion](#)
- [2.4: Material and waveguide dispersion](#)
- [2.5: Different types of optical waveguide](#)

[3: Lasers and amplifiers](#)

- [3.1: Rate equations](#)
- [3.2: Optical cavities](#)
- [3.3: Gain materials](#)
- [3.4: Diode lasers](#)
- [3.5: Bandwidth control](#)
- [3.6: Fiber amplifiers](#)

[4: Integrated optical components](#)

- [4.1: Introduction](#)
- [4.2: Mach-Zehnder Interferometers](#)
- [4.3: Add-drop multiplexers](#)
- [4.4: Arrayed waveguide grating](#)
- [4.5: Photodiodes](#)

[Index](#)

[Glossary](#)

[Detailed Licensing](#)

[Detailed Licensing](#)

Licensing

A detailed breakdown of this resource's licensing can be found in [Back Matter/Detailed Licensing](#).

CHAPTER OVERVIEW

1: Introduction

[1.1: The long tail](#)

[1.2: SONET/SDH](#)

[1.3: The importance of clocks](#)

This text is designed to give a very compact introduction to optical communications. It begins at a very high level, discussing networks and their constraints. Then it dives into the component level, covering the basics of waveguides, sources (lasers and amplifiers) and how to modulate them, and finishes with an introduction to integrated optical components.

This book is not designed to be comprehensive and is specifically designed for a 6 week course, catering to electrical engineering students with no previous studies in optics.

This page titled [1: Introduction](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

1.1: The long tail

1.1.1: Uneven upgrading

Communication, in general, is about agreement. In order to convey information, two parties have to agree on how they will do the conveying. In conversation, we agree on a common language, for instance. But, should a third party join the conversation, it may be necessary to switch languages to include them. This can be done very quickly (assuming there is a common language between the three). However, for a telecommunications system to function, these agreements need to be in place from the beginning, and any upgrades should take into account previous agreements. Figure 1.1.1 shows a network in which a variety of technologies from different eras are able to interconnect with each other. As can be seen, a simple landline, maybe even with a rotary dial, must still be able to connect to a modern cellphone. Likewise, internet service must scale from dial up modems through to optical broadband connections, and much more. This does not mean that a network will support a technology in perpetuity, but that the scaling back and removal of support is usually carefully planned well in advance so that users have the time and resources to move to a supported technology.

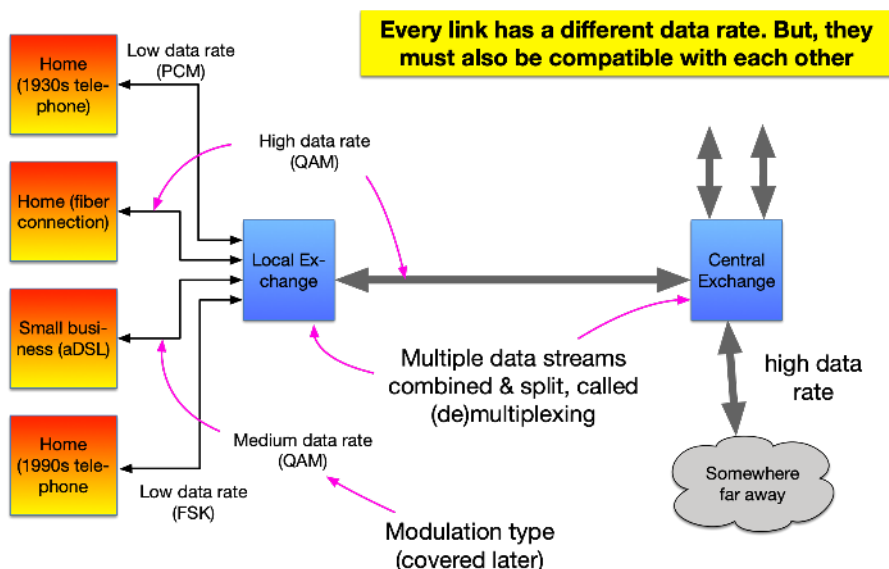


Figure 1.1.1: An impression of the variety of devices and their different data rates that might be connected to a network.

Within a single network, maintaining support for the long tail of devices can be complicated. However, there is no single world-wide network: every country has its own set of networks, often privately owned. At the same time, interconnection between the different networks is still required. This problem was identified early in the development of digital communications, and, as a result, modern telecommunications provides a hierarchical data structure (SONET/SDH and now OTN) that enables long-tail support.

This page titled [1.1: The long tail](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

1.2: SONET/SDH

SONET (Synchronous optical network) and SDH (synchronous digital hierarchy) is the foundation of modern telecommunications. By examining Figure 1.1.1, an analogy can be made with a river. The river is created by the confluence of many tributaries. In communications, each sender wants to communicate with one or more receivers. It is very inefficient to create a few billion independent data streams linking different senders and receivers. Instead, the data from different senders is multiplexed (combined) several times to create streams with a very high data rate. It is then also necessary to remove the right data from that stream (demultiplexing) so that the right data is sent to the right receiver. SONET/SDH enables different data streams with different base transmission rates to be efficiently multiplexed and demultiplexed.

At the same time SONET/SDH maintains a kind of "physical" compatibility with older technology (see Figure 1.2.1). The original telephone network operated at a frequency of 8 kHz (period = 125 μ s). As digital data became more significant, it became necessary to increase the data rate, so clock frequencies that are multiples of 8 kHz were introduced, allowing more data to be packed into the 125 μ s window. For instance, the E1 standard carries 30 telephone channels (each 125 μ s analog pulse is converted to a single 8 bit number) and two overhead channels, called an E1 frame. A single frame takes 125 μ s to transmit, meaning that the base data rate is 2.048 Mb/s. E1 channels can be combined to create higher-order E numbers (E2, for instance has 4 E1 tributaries), which also have a frame that is 125 μ s long. Likewise, SONET/SDH data frames are 125 μ s long, but operate at much higher data rates (STM-1, the smallest SDH frame, is 155.520 Mb/s). This universal quality makes it easier for new technology to accommodate legacy equipment.

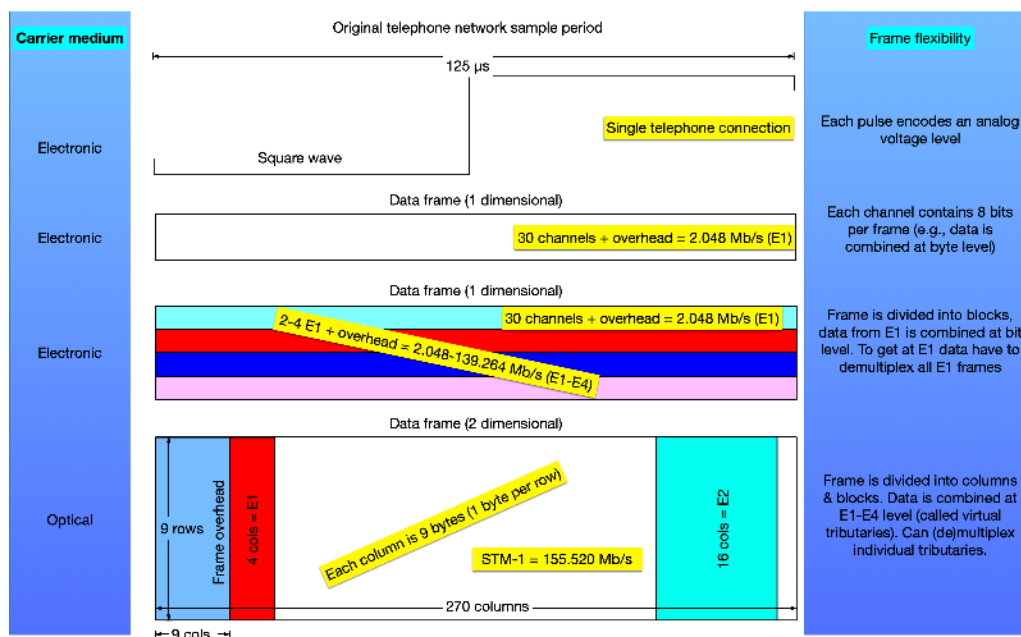


Figure 1.2.1: A visualisation of the relationship between data carrying capability and flexibility of different data carriers.

On the other hand, multiplexing creates an additional level of complexity. Imagine a phone call originating in New York and being received in Amsterdam: that tiny bit of data is multiplexed with many other phone calls and data streams so that it can be efficiently sent to Europe. However, at some point that one phone call needs to be demultiplexed. It might even be the only bit of data that needs to be removed from a much larger frame. Higher order E-frames do not support this. When E1 tributaries are multiplexed into E2, this is performed at the bit level, so a single byte has bits from each tributary. To demultiplex a single channel from a given E1 frame, all four E1 streams have to be separated, the relevant channel removed (from a single E1 frame), a new channel added, and then all four E1 tributaries are multiplexed again.

SONET/SDH accommodates efficient demultiplexing by including significantly more overhead. The payload of a SONET/SDH frame consists of virtual tributaries (denoted as VC), each of which are defined to match common lower data rates, as shown in Table 1.2.1.

Table 1.2.1: data rates for different standards and their corresponding virtual tributaries and their corresponding data rates.

Original standard and data rate	Virtual tributary and data rate
T1 (1.544 Mb/s)	VC-11 (1.728 Mb/s)
E1 (2.048 Mb/s)	VC-12 (2.304 Mb/s)
E3 (34.368 Mb/s)	VC-3 (48.960 Mb/s)
ATM (599.040 Mb/s)	VC-4-4C (601.344 Mb/s)

Notice that the virtual tributary data rate is always higher? This is for several reasons. First there is an internal pointer that keeps track of the location of the payload, even if it spreads across two frames. Second, the additional room allows for differences in clocking between different nodes by shifting bits within the frame.

This page titled [1.2: SONET/SDH](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

1.3: The importance of clocks

SONET/SDH is designed to operate without a central clock. Nevertheless, the importance of good clocks cannot be overemphasized. SONET/SDH allows for a certain amount of desynchronization between the clocks at each node, but the clocks still have to be quite precise.

In a typical network, there is a hierarchy of clocks and each node will use the highest ranking clock. Thus, the network is robust against individual clock failures. Individual clocks are expected to be accurate to within 1 ps or better and be synchronized with other clocks on the network. This can be achieved by deriving the clock signal from a GPS source, or by communicating clock data within the network (for example).

This page titled [1.3: The importance of clocks](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

CHAPTER OVERVIEW

2: Optical Waveguides

[2.1: Introduction to waveguides](#)

[2.2: Spatial modes](#)

[2.3: Modal dispersion](#)

[2.4: Material and waveguide dispersion](#)

[2.5: Different types of optical waveguide](#)

This page titled [2: Optical Waveguides](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

2.1: Introduction to waveguides

In free-space communication, the received signal power falls off as the square of the distance. At the same time, the error rate is inversely proportional to the received signal power. Thus, communication via free space can be quite power inefficient. Furthermore, free-space communication is, generally, limited to line of sight, so communication from Europe to New Zealand, for instance, would require receiving and re-transmitting the signal from at least one satellite.

Waveguides provide a way to confine the power of a wave within a limited space and guide the wave. Hence, we can send signals over longer distances with lower power. In this chapter, the basic principles of optical waveguides will be introduced and the consequences those properties have on communication will be discussed.

2.1.1: Guiding light

Light, in most optical fibers, is guided due to *total internal reflection*. When light tries to travel from one medium to another, it must adjust its speed, which, if the light hits the boundary at an angle, causes the direction of travel to change. This bending is described by Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad (2.1.1)$$

where n_1 and n_2 are the refractive indices of the two materials and θ_1 and θ_2 are the angles of incidence and refraction, as shown in Figure 2.1.1. The refractive index is greater than one for all materials and is unity for vacuum. If $n_2 < n_1$, then for some angles of incidence, $\sin \theta_2 > 1$, which is not possible. For these angles of incidence the light is entirely reflected. The angle at which total internal reflection starts to occur is called the critical angle (θ_c) and for $\theta \geq \theta_c$ all light is reflected.

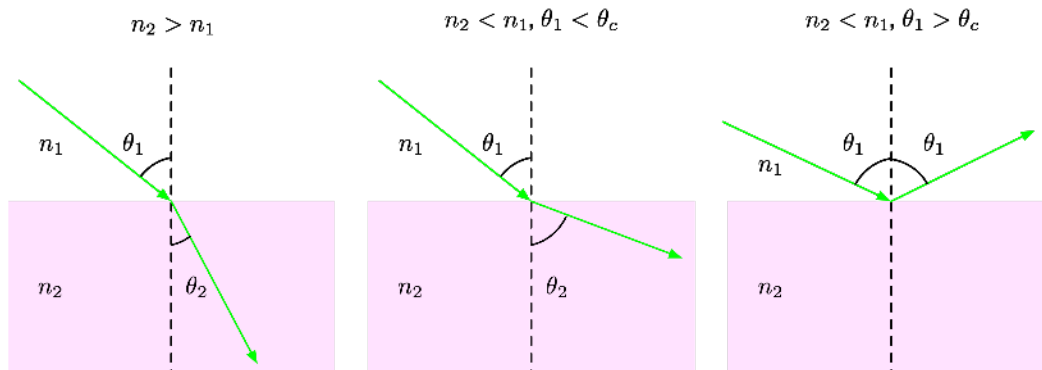


Figure 2.1.1: Snell's law for three cases. In the left example the ray of light (green) is traveling from a less dense medium to a more dense medium. In this case, the light ray bends towards the normal (dotted line). In the middle example, the light ray bends away from the normal because it is traveling from more dense medium to a less dense medium. In the right example the angle of incidence is so high that the light cannot enter the more dense medium and, therefore, reflects.

Total internal reflection is a reflection of 100% of the optical power, compared to between 80-90% for reflectivity from household mirrors. Thus a slab of glass ($n = 1.5$), surrounded by air ($n = 1$) can act as a waveguide. As shown in 2.1.2, only light rays that fall within the critical angle have the potential to be guided. The cone of rays that are accepted define the *numerical aperture* (NA) of the waveguide. The NA of an optical fiber or other waveguide is used to determine the best focal conditions (choice of lens) to optimally couple light into the waveguide.

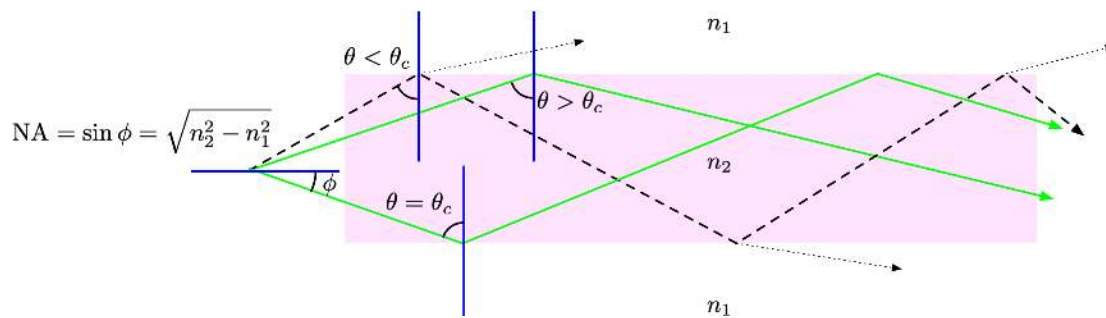


Figure 2.1.2: Light is coupled into a waveguide (pink box). The green rays represent light rays that are incident at angles greater than the critical angle. The cone of angles that meet the condition $\theta \geq \theta_c$ is given by ϕ , which, in turn, defines the NA of the waveguide. The NA of a waveguide can also be calculated by the refractive index of the core, n_2 and the cladding, n_1 . The dashed line represents a ray of light that is not guided. At each reflection, some of the light leaks from the waveguide (dotted arrows).

This page titled [2.1: Introduction to waveguides](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

2.2: Spatial modes

Total internal reflection is not the only condition for guiding, however. Light is a wave that has physical extent in all dimensions. Thus, as demonstrated in Figure 2.2.1, light that has reflected from the top interface interferes with light that has reflected from the bottom interface. If the interference is not constructive, then the amplitude of the light wave will decrease with distance traveled, and eventually disappears entirely. More specifically, after two reflections, the phase change of the wave along the axis of the waveguide should be an integer multiple of π . When this condition is satisfied, the light occupies a *spatial mode* of the waveguide.

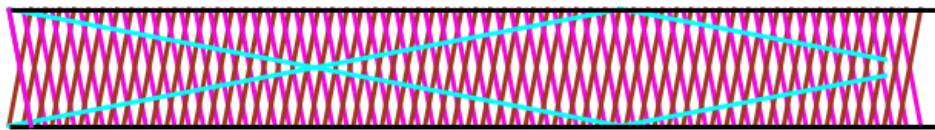


Figure 2.2.1: The brown wavefronts are for a ray traveling from top to bottom, and the pink wavefronts are for a ray traveling from bottom to top. Note how there are three lines down the waveguide where the pink and brown wavefronts cross. Also note how the wavefront at a reflection point overlaps perfectly with the wavefront from the parallel ray before the reflection. These are all conditions to be a spatial mode of the waveguide.

The picture in Figure 2.2.1 is complicated by the fact that total internal reflection introduces an angle-dependent phase shift to the light wave on reflection. However, in what follows, we first apply a constant phase shift of π (thus treating the reflection as mirror-like), and then afterwards account for an angular dependent phase shift.

Note

If you are not familiar with the wavefront picture above, it can be a bit confusing. The arrowed line represents the direction of travel for light (a light ray). The lines at right angle represent the peaks in amplitude, which are separated by one wavelength. This picture, then represents a **frozen moment in time** and allows us to compare the phase relationship between multiple waves, or, in this case, the wave with itself.

2.2.1: Mirror-like reflections

A geometric solution to the phase consistency problem is presented in Figure 2.2.2. The line $\overline{AB}$ represents the path the ray would have taken if there was no reflection, while path $\overline{AC}$ represents the path the light has actually taken. For self consistency the two paths can only differ by integer multiples of 2π . The accumulated phase over $\overline{AB}$ is $\phi_{AB} = \frac{2\pi\overline{AB}}{\lambda}$, where λ is the wavelength of the light. Likewise, the accumulated phase over $\overline{AC}$ is $\phi_{AC} = \frac{2\pi\overline{AC}}{\lambda}$. The difference between the two is $\Delta\phi = \frac{2\pi\overline{AC}}{\lambda} - \frac{2\pi\overline{AB}}{\lambda} - 2\pi$, where the 2π accounts for the additional phase shift due to two reflections. The difference must be $\Delta\phi = 2m\pi$ where m is an integer and it doesn't really matter where we start m , so we can subtract the 2π and reset m , leaving $m = \frac{1}{\lambda} (\overline{AC} - \overline{AB})$, where $m = 1, 2, \dots$

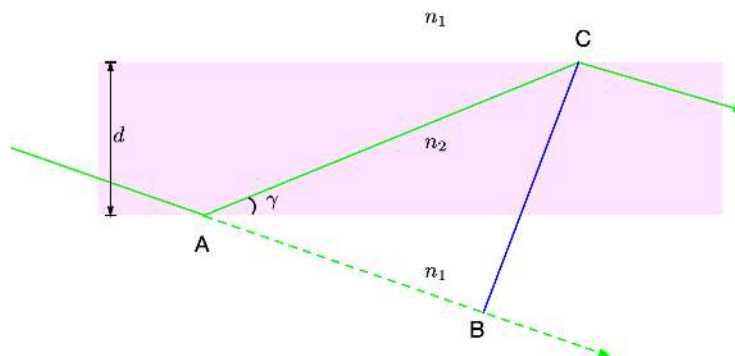


Figure 2.2.2: A geometric view of the self consistency problem. The solid green lines represent the path of the ray in the waveguide. The dashed green line represents the path that the ray would have taken if it had not been reflected. The two paths $\overline{AB}$ and $\overline{AC}$ should be an integer number of wavelengths.

$$\overline{AC} = \frac{d}{\sin \gamma} \quad (2.2.1)$$

where γ is the complementary angle to θ . The triangle $\overline{ABC}$ allows us to solve for the angles that meet the self-consistency condition.

$$\cos 2\gamma = \frac{\overline{AB}}{\overline{AC}} \quad (2.2.2)$$

$$\cos 2\gamma = 1 - 2 \sin^2 \gamma \quad (2.2.3)$$

$$1 - 2 \sin^2 \gamma = \frac{\overline{AB}}{\overline{AC}} \quad (2.2.4)$$

$$\overline{AC} - 2\overline{AC} \sin^2 \gamma = \overline{AB} \quad (2.2.5)$$

$$\overline{AC} - \overline{AB} = 2\overline{AC} \sin^2 \gamma \quad (2.2.6)$$

$$\overline{AC} - \overline{AB} = \frac{2d \sin^2 \gamma}{\sin \gamma} \quad (2.2.7)$$

$$m\lambda = 2d \sin \gamma \quad (2.2.8)$$

$$\sin \gamma_m = m \frac{\lambda}{2d} \quad (2.2.9)$$

Equation 2.2.9 represents the set of angles that satisfy the self consistency relationship for any given wavelength. By setting $\sin \gamma_m = 1$ and solving for m , we find that the maximum number of modes supported by the waveguide is

$$M = \frac{2d}{\lambda} \quad (2.2.10)$$

where M is rounded down to the nearest integer. Note that this also means that, for this type of waveguide, there is a minimum dimension to enable a single mode to exist. If $M = 0$ that indicates that, for that wavelength, no propagating modes exist and no light will be transported by the waveguide. In this case, an attempt to couple light into the waveguide would result in a very strong scattering from the front face.

Later on, we will find that the number of modes is a severe limiting factor in the performance of an optical communications system. Indeed, for long distance communications, $M = 1$, otherwise known as single mode operation, is an important criteria.

2.2.2: Angle-dependent phase shift

Snell's law does not tell the full story for refraction and reflection. When light is reflected from an interface, it also undergoes a phase shift, as illustrated in Figure 2.2.3.

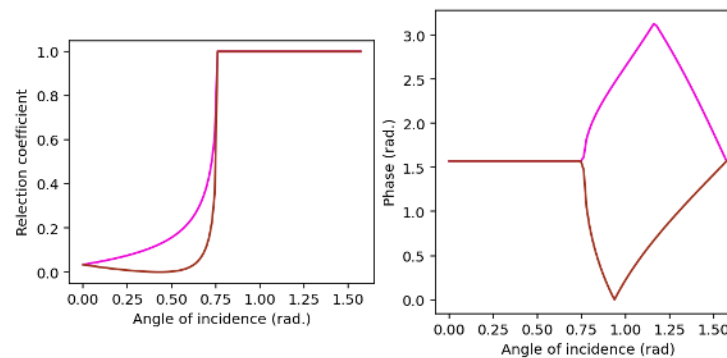


Figure 2.2.3: Reflection coefficient and phase change for the case of the electric field parallel (brown) to the interface (TE mode) and perpendicular (purple) to the interface (TM mode).

Note that the phase shift also depends on the orientation of the electric field with respect to the interface. In what follows, we only discuss the case for the electric field parallel to the interface, called transverse electric (TE) modes (when the magnetic field is parallel to the interface, this is called the transverse magnetic (TM) mode). The phase shift due to reflection is given by

$$\tan \frac{\phi_r}{2} = \sqrt{\frac{\sin^2 \gamma_c}{\sin^2 \gamma} - 1} \quad (2.2.11)$$

where γ_c is the complementary angle to the critical angle, θ_c , and ϕ_r is the phase shift due to reflection. Note that equation 2.2.11 is not valid for $\gamma > \gamma_c$, where, instead, one must use the Fresnel amplitude reflection coefficients, that were used to in Figure 2.2.3.

The self-consistency expression given in equation 2.2.9 becomes

$$\tan\left(\frac{\pi d}{\lambda} \sin \gamma - \frac{m\pi}{2}\right) = \sqrt{\frac{\sin^2 \gamma_c}{\sin^2 \gamma} - 1} \quad (2.2.12)$$

where $m = 0, 1, 2, \dots$. Equation 2.2.12 is best solved graphically by looking for intercepts between the left hand side and right hand side, as shown in Figure 2.2.4. The total number of TE modes supported by a waveguide is given by

$$M \doteq \frac{2d \sin \gamma_c}{\lambda} \quad (2.2.13)$$

which is related to the NA by $M \doteq \frac{2d}{\lambda_0} \text{NA}$, where M is now always rounded up, and λ_0 denotes the wavelength in vacuum. In contrast to the mirror-like case, there is no cutoff for any wavelength. However, to achieve single mode operation, the dimensions have to be quite close to the operational wavelength.

In order to know precisely which mode is being discussed, a labeling scheme is used that tells us the polarization of the mode and its spatial character. For a slab waveguide, as discussed here, the modes are denoted as TE₀, TE₁, TE₂, ... and TM₀, TM₁, TM₂, ... for the transverse electric and transverse magnetic modes.

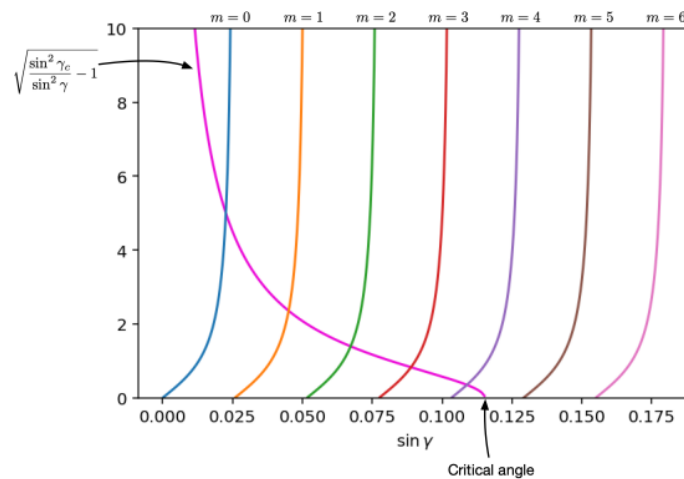


Figure 2.2.4: Graphical solution to identifying spatial modes in a dielectric waveguide. The crossing points between the purple (right hand side of equation 2.2.12) curve and the other curves (left hand side of equation 2.2.12).

2.2.3: Going beyond slab waveguides

So far, we have looked at a waveguide that is an infinite wide slab so that we only have to deal with a single dimension. For square waveguides, the mathematics remains the same, but now there are modes associated with both spatial dimensions, thus, if the axis of propagation of a waveguide is along the z axis, then the two dimensions d_x , and d_y each support M_x and M_y modes (calculated from equation 2.2.13). The modes along each axis are *independent* of each other, so the total number of modes, M is all possible combinations $M = M_x M_y$.

Optical fibers are cylinders, meaning that the geometric approach used above cannot be used. The electric and magnetic fields are never quite parallel to the interface of an optical fiber, resulting in so-called linearly polarized (LP) modes that are denoted by two integers l and m . As with the dielectric slab waveguide, it is simplest to find the modes by graphically finding the intercepts between two equations (both Bessel's functions) that represent the electric field in the core and the cladding of the fiber.

Even the total number of modes supported by an optical fiber can be difficult to calculate. However, we are mostly interested in whether the fiber is single mode, multi-mode, and occasionally few-mode. To estimate this, the V parameter, which (sort of) represents the mismatch between light propagation speed in the core and cladding materials.

$$V = \frac{2\pi r_{co}}{\lambda_0} \text{NA} \quad (2.2.14)$$

where r_{co} is the radius of the core. Single mode operation is guaranteed for $V < 2.405$. For multimode operation, the total number of modes can be estimated from $M \approx \frac{4V^2}{\pi^2}$.

This page titled [2.2: Spatial modes](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

2.3: Modal dispersion

Modal dispersion is, essentially, the idea that each spatial mode has a different group velocity. This means that if you were to divide an input light pulse into two different spatial modes of a waveguide, the result would be that the two copies of the pulses would emerge from the waveguide at different times. This idealized picture indicates that it is possible to control which spatial modes of a waveguide are used. However, in general this is not true, and even if the coupling to the waveguide only put light into a single spatial mode, imperfections in the waveguide would scatter light into all other modes. So, in a communications system, the short sharp pulses of light required for high data rates spread out and overlap with each other. Ultimately, modal dispersion limits the product of distance and data rate: at short distances, high data rates are still possible, while over long distances only low data rates are possible.

2.3.1: Modal dispersion in a slab waveguide

In a first approximation, it is possible to understand modal dispersion simply by examining the path taken by a ray of light that propagates along the axis of a waveguide, compared to a ray of light that propagates at the critical angle, as shown in Figure 2.3.1. These two cases represent the highest and the lowest group velocity supported by the waveguide.

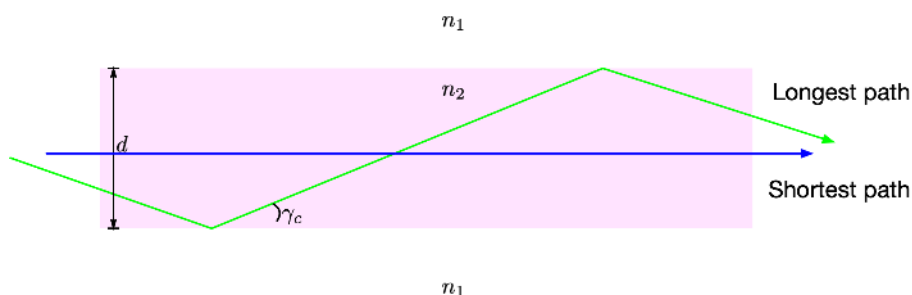


Figure 2.3.1: Shortest and longest paths through a waveguide with mirror-like reflections.

Here it is helpful to define the wavevector, $k_0 = \frac{2\pi}{\lambda_0}$ of a wave, which is modified by the material through which the wave travels: $k = nk_0$. So far, we have only discussed the magnitude. The wavevector for the light paths in Figure 2.3.1 have a component along the axis of the waveguide k_z and a component that is transverse k_t , which must vector sum to k . The transverse wavevectors are given by

$$k_t = nk_0 \sin \gamma_m \quad (2.3.1)$$

$$k_t = \frac{2\pi}{\lambda} \sin \gamma_m \quad (2.3.2)$$

$$k_t = \frac{m\pi}{d} \quad (2.3.3)$$

and the wavevectors along the axis are the wavevectors of the mode k_m

$$k_m^2 = k^2 - k_t^2 \quad (2.3.4)$$

$$k_m^2 = k^2 - \frac{m^2 \pi^2}{d^2} \quad (2.3.5)$$

$$k_m^2 = \left(\frac{\omega}{c}\right)^2 - \frac{m^2 \pi^2}{d^2} \quad (2.3.6)$$

Note that when m is large, k_m becomes imaginary, and the light cannot propagate in the waveguide. Note also, that for a given waveguide size, there is a lowest (angular) frequency that will propagate $\omega_c = \frac{\pi c}{d}$. Equation 2.3.6 is known as a **dispersion relation** and can be used to determine the variation of group velocity with frequency and mode number, which, in turn, dictates the maximum data rate.

Figure 2.3.2 shows the change in k_m with mode number.

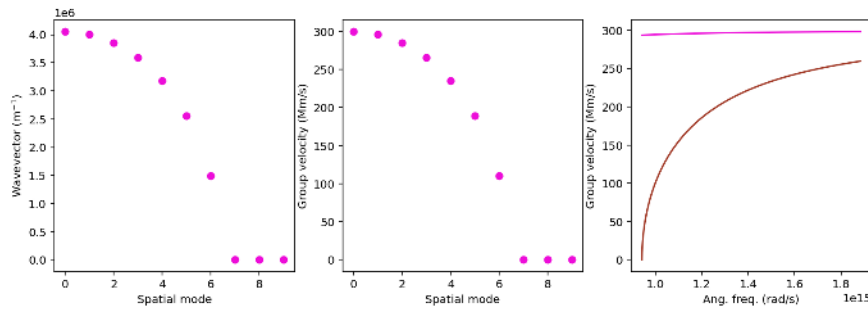


Figure 2.3.2: Modal wavevector with increasing mode number (left hand side), and corresponding group velocity (center) for a fixed frequency. Variation of group velocity with frequency for fundamental and 5th modes (right hand side).

The group velocity (the speed at which information is transferred) is given by $v_g = \frac{d\omega}{dk}$ and can be obtained by taking the derivative of equation 2.3.6

$$2k_m \frac{dk_m}{d\omega} = \frac{2\omega}{c^2} \quad (2.3.7)$$

$$\frac{k_m c^2}{\omega} = \frac{d\omega}{dk_m} \quad (2.3.8)$$

$$\frac{d\omega}{dk_m} = c \sqrt{1 - \frac{m^2 \omega_c^2}{\omega^2}} \quad (2.3.9)$$

Note that the group velocity for a given mode depends on both the waveguide dimensions and the frequency of the light wave. The variation of group velocity with mode number and frequency is shown in Figure 2.3.2.

2.3.2: Modal dispersion in "real" waveguides

In the discussion in Section 2.3.1, it was assumed that the entire wave traveled within the core of the waveguide (the reflections are mirror like), while in real waveguides, the spatial mode is distributed between the cladding and the core. The wavevector along the the axis of the waveguide is given by $k_m = n_2 k_0 \cos \gamma_m$, where $\gamma_c \geq \gamma_m \geq 0$. At $\gamma_m = \gamma_c$, $\cos \gamma_c = \frac{n_1}{n_2}$, so a substitution shows that k_m goes from $n_1 k_0$ to $n_2 k_0$. Or, put more qualitatively, the spatial modes go from looking like a wave traveling in the core material to looking like a wave traveling in the cladding material.

The dispersion relation for a dielectric slab waveguide is obtained from the self-consistency condition $2k_t d - 2\phi_r = 2\pi m$, and can be expressed in the following way

$$\frac{\omega}{\omega_c} = \frac{\sqrt{n_2^2 - n_1^2}}{\sqrt{n_2^2 - n_{eff}^2}} \left(m + \frac{2}{\pi} \tan^{-1} \sqrt{\frac{n_{eff}^2 - n_1^2}{n_2^2 - n_{eff}^2}} \right) \quad (2.3.10)$$

where $n_{eff} = k_m c_0 / \omega$ is the effective refractive index of the waveguide and $\omega_c = \frac{2\pi c_0}{2dNA}$ is the angular frequency at which the mode cuts off.

Figure 2.3.3 shows the dispersion relation for different modes. Note that the traces track between the two light lines: one for the cladding ($\omega = c_1 k_m$) and one for the core ($\omega = c_2 k_m$). Tracking a mode from the moment the point at which it can begin to propagate (e.g., when the frequency is just above the cut off) onwards, notice that it starts near the light line for the cladding. In other words, near the cut off, the mode behaves as if it is propagating in the cladding, while at much higher frequencies the mode behaves as if it is propagating in the core material.

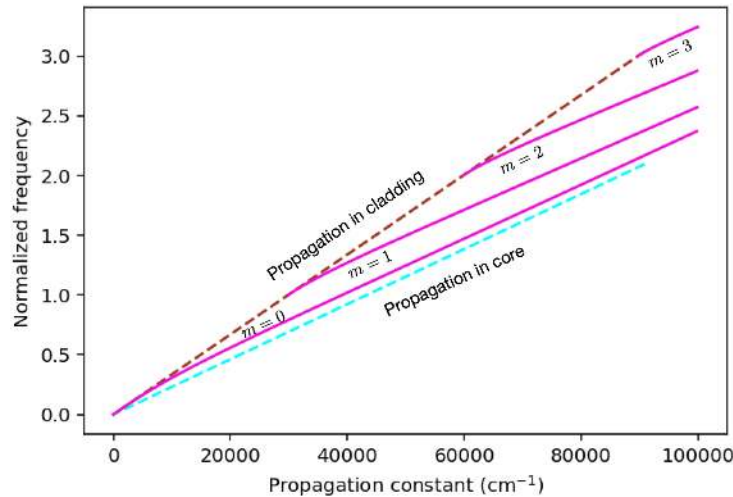


Figure 2.3.3: How the mode wavevector behaves with frequency.

The group velocity is obtained by taking the total derivative of equation 2.3.10, it is possible to obtain an approximate expression for the group velocity. Again, this is not derived in this text.

$$v_g = \frac{d \cot \theta + \Delta z}{\frac{d}{c_1} \csc \theta + \Delta \tau} \quad (2.3.11)$$

where Δz and $\Delta \tau$ take into account that total internal reflection is not instantaneous--the light penetrates slightly into the cladding material before returning. This can be thought of as a slight displacement along the waveguide (Δz) that takes some amount of time ($\Delta \tau$). The upshot is that the group velocity of a mode at low frequency is close to that of the group velocity in the cladding (c_2), and as the frequency increases the group velocity reduces to that of the core (c_1).

2.3.3: Modal dispersion in optical fibers

For optical fibers with a large V number (recall that single mode operation occurs for $V < 2.405$), we have

$$k_{lm} \approx \sqrt{n_2^2 k_0^2 - \frac{2(l+2m)^2 \Delta}{M}} \quad (2.3.12)$$

$$k_{lm} \approx n_2^2 k_0^2 - \frac{(l+2m)^2 \Delta}{M} \quad (2.3.13)$$

where $\Delta = \frac{n_2^2 - n_1^2}{2n_2^2} \approx \frac{n_2 - n_1}{n_2} \ll 1$ is the refractive index contrast, and $M \approx \frac{4V^2}{\pi^2}$ is the total number of modes supported by the fiber. The group velocity is obtained by finding the derivative, $\frac{d\omega}{dk_{lm}}$, of equation 2.3.13, and can be shown to be approximately

$$v_{lm} \approx c_2 \left(1 - \frac{(1+2m)^2 \Delta}{M} \right) \quad (2.3.14)$$

2.3.4: Dispersion limits on communications

So far, we have a number of results showing that the group velocity is not the same for different optical modes in a waveguide or fiber. But, in a practical sense this is not hugely helpful. What we really want to know is that, if I have a pulse of duration τ seconds, and I pass it through an optical fiber of length L meters, what will be the exit pulse duration?

For a highly multimode fiber (so large V), the minimum group velocity is $v_{min} \approx c_2(1 - \Delta)$ and the maximum group velocity is $v_{max} = c_2$, where c_2 is the speed of light in the core of the fiber. The transit times are $t_{min} = \frac{L}{c_2(1-\Delta)}$ and $t_{max} = L/c_2$ the pulse spreading is half the difference between these $\delta t = \frac{L}{2c_2} \left(\frac{1}{1-\Delta} - 1 \right)$. For $\Delta \ll 1$ (which is typical for optical fibers) $\frac{1}{1-\Delta} \approx 1 + \Delta$, so

$$\delta t \approx \frac{L\Delta}{2c_2} \quad (2.3.15)$$

To generalize the result further, the dispersion of an optical fiber is usually expressed by the modal dispersion parameter, as time per unit length $D_m = \frac{n_{co}\Delta}{2c_0}$, where, for clarity, c_2 , the speed of light, has been replaced with speed of light in the vacuum, c_0 , and the

refractive index of the core, n_{co} .

Note

Not all optical fibers use a step index as discussed above. There are also graded index fibers (GRIN fibers) that smoothly change the refractive index from the core to the cladding. GRIN fibers are always multimode, but the modal dispersion, D_m can be much less (ideally $D_m = \frac{n_{co}\Delta^2}{4c_0}$). Recall that $\Delta \ll 1$.

2.3.5: Limits to data rate

Now that it is possible to calculate how much a pulse will spread over a given distance, this provides two limits on the data rate. The first is that the light pulses spread from one time slot into the next causing low (or zero) intensity pulses to be mis-identified as bright pulses. The second is that the peak amplitude of the bright pulses becomes lower, so the contrast between low and high intensity reduces and increases the chance of a mis-identification. If we have a series of optical pulses in a fiber with an even duty cycle, the pulse spreading must be limited to about one quarter of the pulse time to ensure that the dark time-slots are still identified. So if the data rate is N b/s, then the bit time is $1/N$ and $\delta t \leq \frac{1}{4N}$. Substituting this result into equation 2.3.15 gives $NL \approx \frac{c_0}{2n_{co}\Delta}$ for a step index fiber. Note that for a given fiber (and ignoring material properties for now), everything on the right hand side is constant, so the data rate times the distance is a constant factor. Depending on an acceptable error rate and the modulation technique that is used, the factor 4 can change, changing the constant factor, but this is factors of two rather than orders of magnitude.

✓ Example 2.3.1

Two optical fibers, one a step index fiber with $\Delta = 0.0034$ has a core refractive index of 1.46, and the second is a GRIN fiber with the same parameters what are the dispersion parameters for each fiber? Which is better?

In a communications application, the fiber must transport data over 100 m at a data rate of 10 Gb/s. Which, if either fiber can support the data rate?

Solution

The dispersion of the step fiber is $D_m = \frac{n_{co}\Delta}{2c_0}$, where

- $c_0 = 300 \times 10^6$ m/s
- $n_{co} = 1.46$
- $\Delta = 0.0034$

putting these numbers in leads to $D_m = 8.3$ ns/km for the step index fiber

The dispersion for the GRIN fiber is $D_m = \frac{n_{co}\Delta^2}{4c_0}$. Plugging the numbers gives $D_m = 14$ ps/km. The GRIN fiber is about 800 times better in terms of dispersion.

For $N = 100$ Gb/s and $L = 100$, we have $NL = 10$ Tm/s, while $\frac{c_0}{2n_{co}\Delta}$ for the step index fiber is 0.03 Tm/s and for the GRIN fiber it is $\frac{c_0}{n_{co}\Delta^2} = 18$ Tm/s. Thus, we can conclude that the step index fiber cannot support the desired data rate over the required distance. The GRIN optical fiber does meet the requirements with room to spare.

2.3.6: Summary

Optical fibers and waveguides that support multiple modes are subject to modal dispersion. The dispersion spreads optical pulses, reducing their amplitude and causing communications pulses that are expected to fall within a certain time window to spread beyond their allotted window. If unaccounted for, modal dispersion will increase the error rate severely, and even prevent data transmission entirely. If the modal dispersion parameter D_m is known, modal dispersion can be accounted for in the system design.

Single mode devices are not effected by modal dispersion.

This page titled 2.3: Modal dispersion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Chris Lee.

2.4: Material and waveguide dispersion

Modal dispersion (discussed in the previous section) is not the only form of dispersion. In this section, material (chromatic) dispersion will be introduced, however, it should be noted that there are *many* other forms of dispersion, and all forms of dispersion contribute to the degradation of an optical signal.

Material dispersion is due to the fact that the refractive index of all materials (including air) vary with wavelength. In Figure 2.4.1, the refractive index for fused silica (a highly pure form of glass) and lithium niobate, a material used in optical modulators is shown. It is important to note that the refractive index changes *nonlinearly* with wavelength.

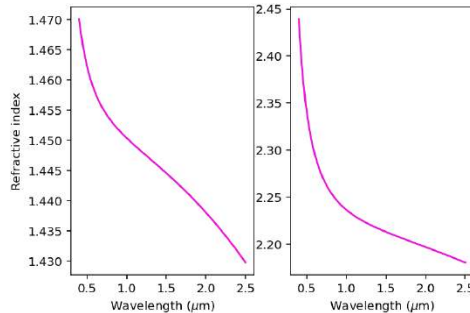


Figure 2.4.1: Refractive index of fused silica (left) and lithium niobate (right).

For various reasons of usability, refractive index is usually measured, tabulated, and calculated in terms of the wavelength of light, so it makes sense to calculate group velocity and dispersion in terms of wavelength. The group velocity is defined as $\frac{1}{\frac{dk}{d\omega}}$, by use of the chain rule, the derivative becomes $\frac{dk}{d\omega} = \frac{dk}{d\lambda} \frac{d\lambda}{d\omega}$. The wavevector is $k = \frac{2\pi n(\lambda_0)}{\lambda_0}$, where λ_0 is the wavelength in free space.

$$\frac{dk}{d\lambda} = \frac{2\pi}{\lambda_0^2} \left(\lambda_0 \frac{dn}{d\lambda} - n(\lambda_0) \right) \quad (2.4.1)$$

$$\lambda_0 = \frac{2\pi c_0}{\omega} \quad (2.4.2)$$

$$\frac{d\lambda}{d\omega} = -\frac{2\pi c_0}{\omega^2} \quad (2.4.3)$$

$$\frac{d\lambda}{d\omega} = -\frac{\lambda^2}{2\pi c_0} \quad (2.4.4)$$

$$\frac{dk}{d\lambda} = \frac{1}{c_0} \left(n(\lambda_0) - \lambda_0 \frac{dn}{d\lambda} \right) \quad (2.4.5)$$

Inverting equation 2.4.5 gives the group velocity as

$$v_g = \frac{c_0}{\left(n(\lambda_0) - \lambda_0 \frac{dn}{d\lambda} \right)} \quad (2.4.6)$$

As with modal dispersion, changes to the group velocity with frequency (or wavelength) can cause pulse spreading. Consider a material, with a length L and two pulses with central (angular) frequencies, $\omega - 0.5\delta\omega$ and $\omega + 0.5\delta\omega$ that propagate through the medium (both with the same starting time). The transit time is $\tau = L/v_g$. The time difference between the two pulses is

$$\delta t = \frac{d\tau}{d\omega} \delta\omega \quad (2.4.7)$$

$$= \frac{d}{d\omega} \left(\frac{1}{v_g} \right) L \delta\omega \quad (2.4.8)$$

$$\delta t = |D_\lambda| L \delta\omega \quad (2.4.9)$$

D_λ is the material dispersion coefficient given by

$$D_\lambda = \frac{d}{d\omega} \left(\frac{1}{v_g} \right) \quad (2.4.10)$$

but this more conveniently expressed in terms of wavelength. By applying the chain rule, we obtain

$$D_{\lambda} = -\frac{\lambda_0}{c_0} \frac{d^2 n}{d\lambda_0^2} \quad (2.4.11)$$

The dispersion coefficient depends on the second derivative of the change of refractive index with wavelength. This means that materials with no change or a linear change of refractive index with wavelength exhibit no material dispersion. But, quadratic and higher order changes with wavelength will introduce dispersion.

A waveguide, however, consists of a cladding and a core with different refractive indices. This results in an additional contribution called waveguide dispersion, given by

$$D_w = -\left(\frac{1}{2\pi c_0}\right) V^2 \frac{d^2 k}{dV^2} \quad (2.4.12)$$

These two components of dispersion can be used to manipulate the total dispersion to meet system requirements.

2.4.1: Managing dispersion

Figure 2.4.2 shows the group velocity dispersion for a typical single mode fiber. Note that at a wavelength of about 1300 nm, there is no dispersion. However, information requires bandwidth, so there is still a tiny amount of residual dispersion around 1300 nm. However, the wavelength of 1300 nm is not normally used for long-distance links because of nonlinear effects, as discussed in Section 2.4.3. Instead, a longer wavelength (1550 nm) is used, where there is a small amount of positive dispersion ($D_{\lambda} \approx +20$ ps/km.nm). As a result, some form of dispersion compensation is required, which is achieved using optical fibers that have impurities added so that the dispersion coefficient is negative. Typically, a *dispersion compensating fiber* (DCF) has a very strong dispersion $D_{\lambda} \approx -160$ ps/km.nm because these fibers are more expensive and have higher losses.

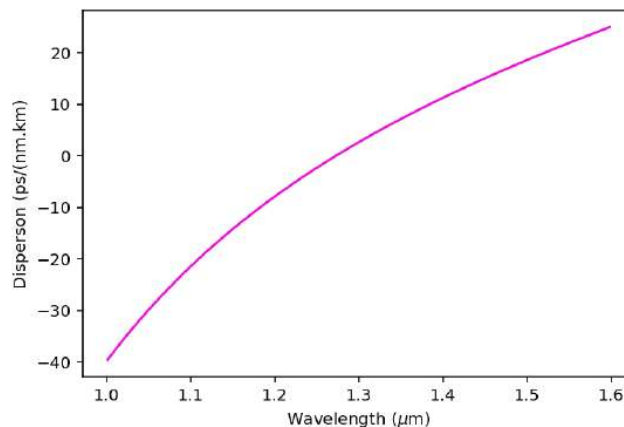


Figure 2.4.2: Group velocity dispersion of a single mode fiber. (Copyright; author via source)

✓ Example 2.4.1

You are designing an optical link between New York and Amsterdam with a single channel capacity of 40 Gb/s, operating at a wavelength of 1550 nm. Ignoring any electronic repeaters, what is the accumulated pulse spreading? Is the pulse spreading within the specification for communications? About how much (DCF) will be required to compensate for the pulse spreading?

Solution

- distance, L from Amsterdam to New York is 5790 km
- required bandwidth is $\Delta f \sim 80$ GHz (twice the data rate), which in wavelength is $\Delta \lambda = \frac{\lambda^2}{c_0 \Delta f} \approx 0.64$ nm

Putting these numbers in leads to a pulse spreading of $\delta t = D_{\lambda} \Delta \lambda L = 74$ ns, while the time-per-bit is 25 ps (even if modern modulation techniques are used, the bit time will not be increased by 1000 times). So, dispersion compensation is required. The length of the fiber is $\frac{\delta t}{D_{\lambda} \Delta \lambda} = 80$ m.

2.4.2: Nonlinear effects

Most optical systems are *linear*. A simple way to think of the difference between linear and nonlinear is that, in a linear system, the light direction is modified by the medium, but the medium is not modified by the light. In a nonlinear system, the light modifies the medium, which then changes the frequency (wavelength) and intensity of the light. Linearity holds for almost all situations, however, when the light intensity is high (high means kW/cm^2 or higher), then nonlinear effects become important. A typical single mode optical fiber has a mode field diameter (the area in which the vast majority of the light power is concentrated) of $12\text{ }\mu\text{m}$, corresponding to an area of about $1.1 \times 10^{-6}\text{ cm}^2$, so a 1 W laser already provides an intensity of $1\text{ MW}/\text{cm}^2$. More realistically, optical communications systems work with powers in the 10-100 mW range, which leads to intensities of 1-100 kW/cm^2 , chosen to *minimize* nonlinear effects.

The most immediately relevant nonlinear effect is called four-wave mixing. Four wave mixing creates new frequencies of light based on sums and differences of input light signals. To give an example: to increase the capacity of an optical communications system, it is standard to divide the data into channels that are separated by wavelength/frequency (the separation between channels is about 100 GHz). In a linear system, this is fine, because the two wavelengths will never mix with each other and can be separated using standard techniques that are presented later. So we have two signals at ω_1 and $\omega_2 = \omega_1 + \delta\omega$. These can mix to create two more signals at $2\omega_1 - \omega_2 = \omega_1 - \delta\omega$ and $2\omega_2 - \omega_1 = \omega_1 + 2\delta\omega$. These two signals fall in adjacent channels that might also be in use. The new signals are a mixture of the information carried in channels one and two, so data from (potentially) three channels have been mixed uncontrollably in one channel.

It gets worse though. Now we have four signals that can all mix with each other, and this process can repeat endlessly, eventually corrupting all available channels (see Figure 2.4.3). What prevents this cascade? Dispersion prevents the cascade. The new signals generated in a four-wave mixing process (and many other nonlinear processes) grow in amplitude as $E_{\text{new}} = E_1^2 E_2 e^{i\Delta k z}$, where E_i is the electric field amplitude z is the distance traveled and the phase-mismatch, $\Delta k = 2k_1 - k_2$, is the difference in wavevectors for the input waves. If the refractive index varies nonlinearly with wavelength $|\Delta k| > 0$ otherwise $\Delta k = 0$. If the phase-mismatch is zero, then the new signal will grow in amplitude until all signals are of about equal amplitude. Otherwise, the growth of the new signal is prevented by the oscillation of the complex exponential term. Even with dispersion slowing the growth of nonlinear effects, the intensity of the light still has to be low enough that processes like four-wave mixing are tiny over a half oscillation of the exponential term.

Increasing time

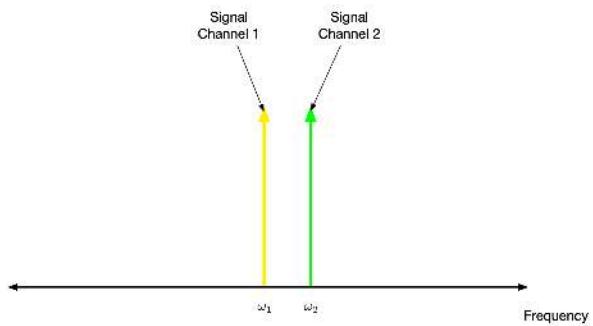
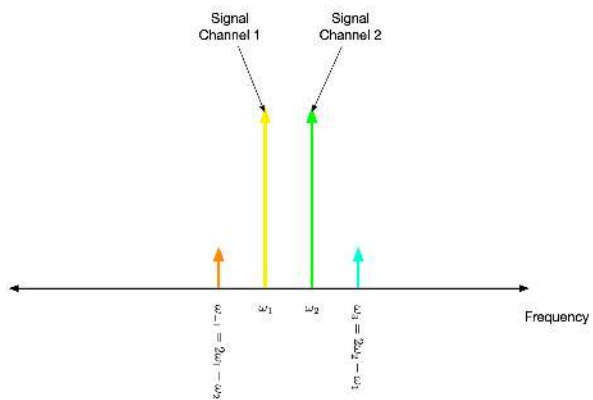
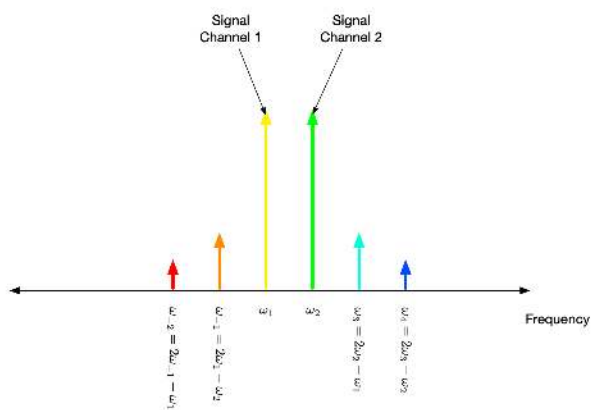
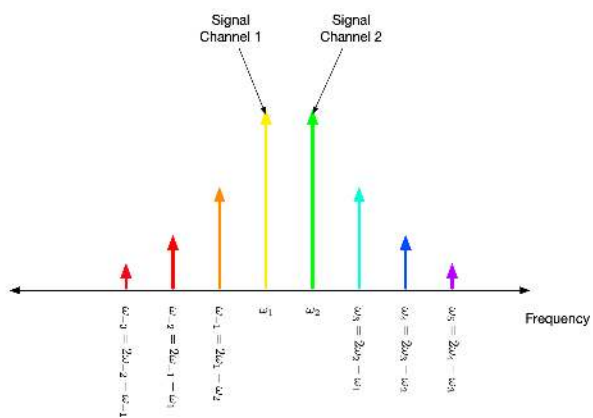


Figure 2.4.3: An illustration of how four-wave mixing can corrupt a data transmission. In the bottom picture, two channels are present, which then mix and add signal to two adjacent channels (which may or may not already have a signal). This process then repeats, spreading the signal across multiple channels (moving from bottom to top in the figure).

This page titled [2.4: Material and waveguide dispersion](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

2.5: Different types of optical waveguide

In this last section, a few different types of waveguides are presented, but without detailed explanation of how they work

2.5.1: Types of optical fiber

Optical fibers come in four basic types: step index single mode fibers, step index multimode fibers, graded index multimode fibers, and photonic crystal fibers. The first three been discussed extensively and are pictured in Figure 2.5.1.

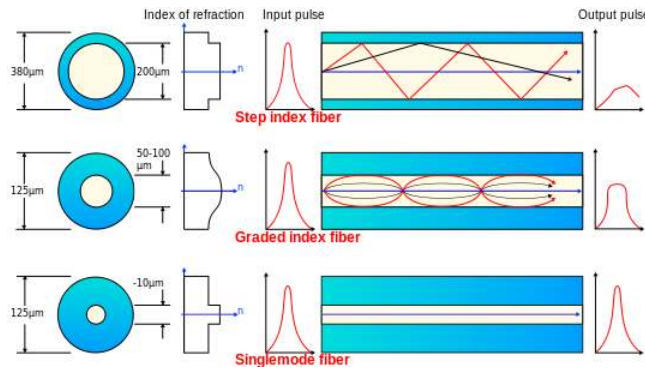


Figure 2.5.1: The three main types of optical fiber.

Photonic crystal fibers (see Figure 2.5.2) are quite different. Instead of relying on total internal reflection, the glass structures that surround the core (which may be glass or air) cause the light to reflect of multiple surfaces. The surfaces are spaced such that light traveling outward from the core experiences destructive interference (this is called a photonic bandgap), and, therefore, cannot leave. Photonic crystal fibers provide designers with additional freedom to control dispersion, prevent multimode operation, and enhance (or suppress) nonlinear behavior. On the other hand, photonic crystal fibers are (much) more expensive and have higher losses, so are not suitable for signal transport.

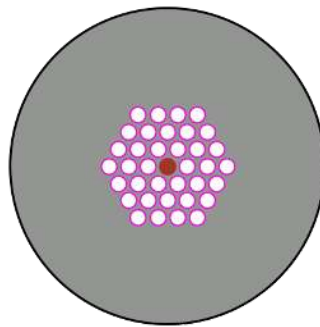


Figure 2.5.2: Diagram of a photonic-crystal fiber. The orange center is solid, while the white circles are hollow tubes running the length of the fiber. The honey comb structure prevents the light from leaving the core (central solid region). To give a sense of scale, the core is about 5 μm in diameter.

2.5.2: Integrated optics

Optical fibers are very useful over long distances, much like copper wires are for electronics. However, like electronics, processing optical data requires a more compact form factor, like a integrated circuit. These chips also use waveguides to guide light and control light. A number of different examples are shown in Figure 2.5.3.

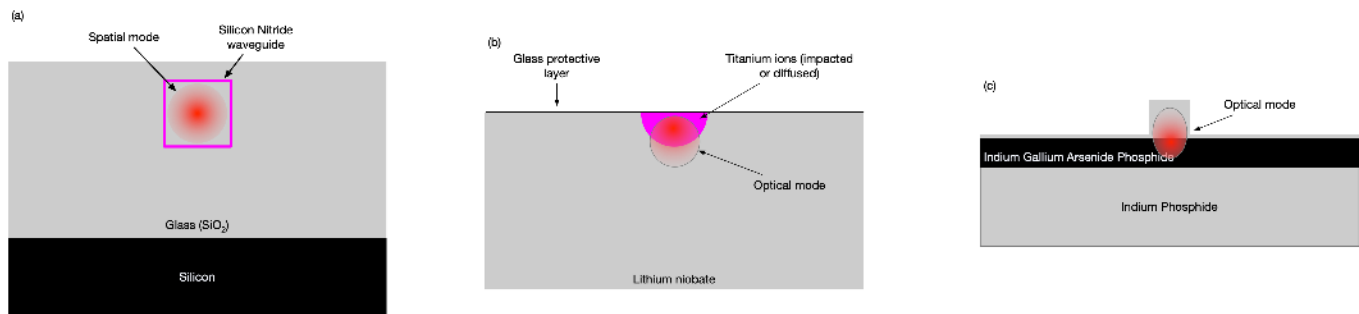


Figure: 2.5.3: Examples of waveguides used in integrated optical chips. Silicon nitride (a): a buried strip (or box, as shown here) of silicon nitride, surrounded by glass provides strong guiding and very low loss. Diffusing titanium ions into a lithium niobate crystal creates enough of a refractive index difference to guide light, however, the optical mode is only barely confined, and will leak out if the waveguide has any strong bends in it. Indium phosphide can be doped with gallium and arsenide to create a confining layer (c).

A ridge of indium phosphide on top provide lateral confinement to the mode. Indium phosphide can also be used as an *active medium*, meaning that waveguides can be directly built into lasers and photodiodes (detectors).

This page titled [2.5: Different types of optical waveguide](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

CHAPTER OVERVIEW

3: Lasers and amplifiers

[3.1: Rate equations](#)

[3.2: Optical cavities](#)

[3.3: Gain materials](#)

[3.4: Diode lasers](#)

[3.5: Bandwidth control](#)

[3.6: Fiber amplifiers](#)

This page titled [3: Lasers and amplifiers](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

3.1: Rate equations

A laser consists of two elements: a gain medium that amplifies light, and a resonator that confines the light to the amplifier. A very simplified view of gain in an optical amplifier is shown in Figure 3.1.1. The takeaway is that the amount of amplification supplied by the gain medium is not necessarily constant in time or space, because every photon of light generated reduces the gain at the time and location where that photon was produced. As a result, laser gain is highly dynamic, which can make it quite difficult to control.

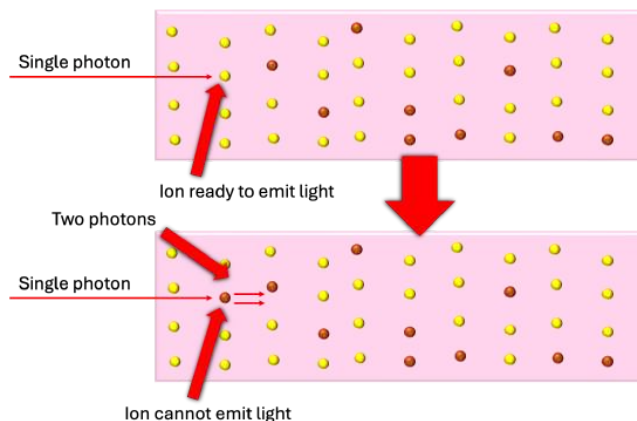


Figure 3.1.1: A simplified view of optical gain. The pink area represents a gain medium consisting of ions. The yellow spheres are ions that are able to emit a photon of light, while the dark red spheres are ions that cannot emit light. The top picture is the state of the medium before a single photon encounters an ion. The bottom picture is the state of the medium after a single photon encounters an ion.

In Figure 3.1.1 the amplification of a single photon is shown. In the top picture, a single photon encounters an ion capable of emitting light. On emitting (bottom picture), the ion cannot emit *another* photon, so it has become dark red, while the number of photons has doubled. As this process proceeds, the number of ions available to emit light decreases, reducing the potential for further light amplification. The balance between light emission and gain is described by rate equations

$$\frac{dN_1}{dt} = I(N_2 - N_1) - PN_1 + kN_2 \quad (3.1.1)$$

$$\frac{dN_2}{dt} = -I(N_2 - N_1) + PN_1 - kN_2 \quad (3.1.2)$$

$$\frac{dI}{dt} = I(N_2 - N_1) - \alpha I - TI \quad (3.1.3)$$

Equations (3.1.1)-(3.1.3) are a set of coupled differential equations that describe the balance between gain and light intensity in an amplifier. N_2 is the fraction of emitters available to emit, and N_1 is the fraction of emitters unavailable. The total number of emitters is a constant ($N_1 + N_2 = 1$). N_2 is increased by the amount of power supplied to the amplifier (P), and is reduced by loss processes (k) and producing light. The intensity of the light, I , is increased by amplification, which is proportional to $N_2 - N_1$, hence gain is only possible when $N_2 > N_1$. The light intensity is also reduced by some internal losses (α), and emission to the outside world, which is given by T .

To understand the behavior of the rate equations, let's simplify by assuming an amplifier with no output ($T = 0$) and no internal losses ($\alpha = k = 0$). Let us start with $N_2 = 1$ and $N_1 = P = 0$. The initial light intensity is some tiny value, corresponding to a single photon of light. The intensity of the light initially grows exponentially, but, then the gain drops and the amplifier saturates (see Figure 3.1.2). Note that everything stabilizes at $N_2 = N_1$. Because there is no emission and no optical losses, the light intensity also stabilizes at a non-zero value. If we add loss processes ($k > 0$, $\alpha > 0$), then N_2 and I will reduce to zero.

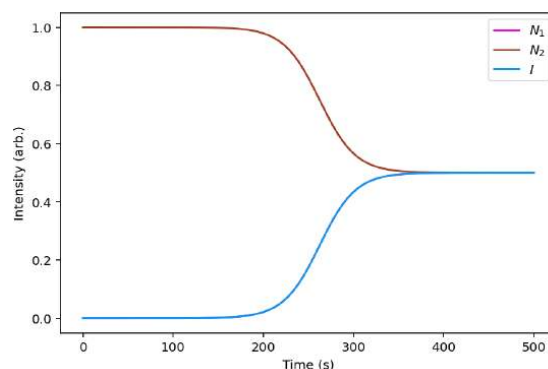


Figure 3.1.2: A gain medium with initial conditions $N_2 = 1$, and no additional excitation.

A laser should have an output. When we do this, the intensity will grow slower, and, once the gain is exhausted, the intensity will drop back to zero, as shown in Figure 3.1.3.

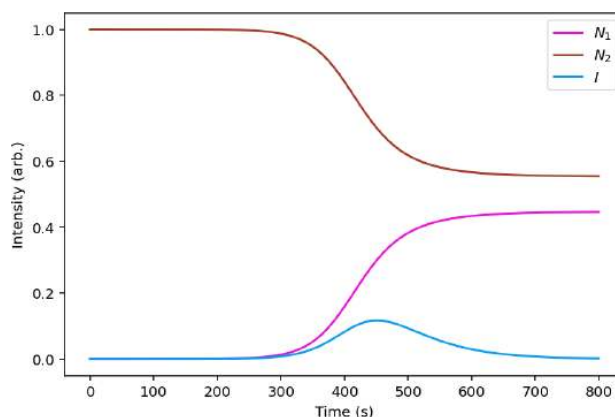


Figure 3.1.3: A gain medium with initial conditions $N_2 = 1$, and some output ($T > 0$).

Now to complete the picture, we will set $P > 0$ and $k > 0$ so that the gain is refreshed and there are some internal losses. Because we are exciting the gain, we will start with $N_2 = 0$. As shown in Figure 3.1.4, the laser only switches on when $N_2 > N_1$ for long enough.

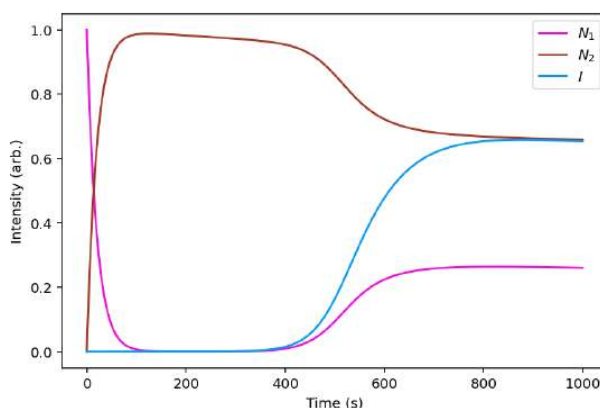


Figure 3.1.4: In this case, the gain medium is excited continuously. $N_2 = 0$ at $t = 0$ and rapidly grows. Only once it is very high does the intensity start to grow, causing the N_2 population to reach a steady state just greater than N_1 .

As seen in Figure 3.1.5, the gain output stabilizes at a constant output, known as steady state conditions. These are the operating conditions for most lasers used in communications systems. One might be tempted to modulate the laser output by modulating the input power $P = f(t)$. In early optical communication systems, this was, indeed, the case. To understand why this is no longer the

case, consider the change of intensity in Figures 3.1.3 and 3.1.4. Note that the slopes are not very steep, indicating that the rise and fall times are quite slow. The time axis in these figures is, of course, related to the settings used in the calculation and should not be taken literally. However, the atomic processes and influence of the optical cavity (see next section) place quite stringent limits on how fast a pulse can build up or decay away in a laser.

? Exercise 3.1.1

The script below was used to generate Figures 3.1.2 – 3.1.4. Adjust the input power, P , the loss $k \leq 1$, and the output, $T \leq 1$, to see how fast you can reach steady state at an output intensity greater than 0.2. Adjust only one parameter at a time while holding the rest fixed to observe how they influence the steady state behavior of the laser

- How do changes in P change the turn on time and output intensity?
- How do changes in k change the turn on time and output intensity?
- How do changes in T change the turn on time and output intensity?

```

1 import numpy as np
2 from matplotlib import pylab as plt
3 from scipy.constants import c
4 from scipy.integrate import solve_ivp
5
6 #conditions for the simulation
7 in_cond = [1, 0, 1e-6] #initial conditions. iorder is N1, N2, I
8 sigma = 0.05           #sigma: material property, leave as is
9 T = 0.02               #T: output 0-1
10 P = 0.05               #P: input power 0-infinity
11 loss = 0.0001          #k: losses 0-1
12 t_span = [0,1000]      #time over which the simulation will run. Adjust as necessary
13                        #but always start at 0
14
15
16 #These are the coupled differential equations
17 def coupledDE(t, out, sig, pump, trans, k):
18     ret = [0,0,0]
19     ret[0] = sig*(out[1]-out[0])*out[2]-pump*out[0]
20     ret[1] = pump*out[0] - sig*(out[1]-out[0])*out[2] - k*out[1]
21     ret[2] = sig*(out[1]-out[0])*out[2] - trans*out[2]
22     return ret
23
24 #this line solves the differential equation

```

run run all restart restart & run all

<matplotlib.legend.Legend at 0x7f089a117a60>

Answer

Increasing the input power changes the output power but does not change the rise time. Increasing the output causes the startup to delay, but the rise time is steeper. Once the output is too high, the laser never switches on. Increasing the loss reduces the output power and delays start up. It does not shorten the rise time. When the loss is high enough, the laser never reaches steady state and will eventually turn off.

This page titled [3.1: Rate equations](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

3.2: Optical cavities

In addition to gain, a laser also requires feedback, which is provided by mirrors that recirculate the light through the gain medium. This is referred to as an optical cavity. The simplest version, shown in Figure 3.2.1, consists of two mirrors facing each other.

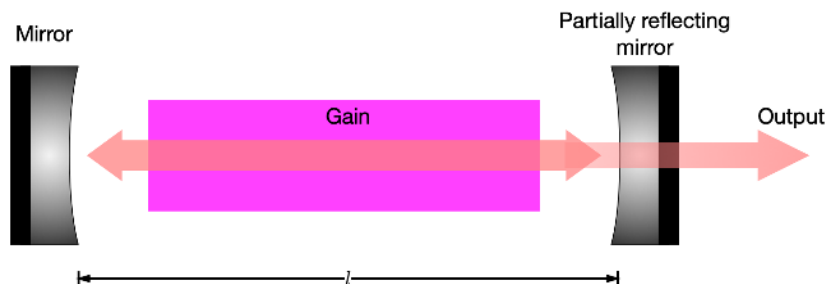


Figure 3.2.1: simplified diagram of a laser.

In Figure 3.2.1, the mirrors are separated by a length, l . If we follow a light wave starting at the left hand mirror, it will travel to the partially reflecting mirror and return to the left hand mirror again (a roundtrip). That will introduce a phase shift of $\delta\phi = 2\pi\frac{2l}{\lambda}$, where λ is the wavelength of the light. If $\delta\phi = 2m\pi$, where m is an integer, then there is no apparent phase change, and constructive interference occurs, allowing the light intensity to build up in the cavity. On the other hand, for any other value, each round trip introduces a phase change that reduces the build up of the intensity. Thus, there is a regularly spaced interval of frequencies ($f = \frac{c}{\lambda}$) that can *resonate* in the cavity given by

$$f_m = \frac{mc}{2l} \quad (3.2.1)$$

where c is the speed of light. The resonant frequencies are called the longitudinal modes of the cavity and the spacing between longitudinal modes is called the free spectral range (FSR). Like optical fibers, there are also spatial models, called transverse modes, but, we will consider that the laser is always operating in the fundamental spatial mode. If the mirrors were perfectly reflecting and there were no other losses, only wavelengths that exactly satisfy equation (3.2.1) would resonate. However, apart from losses, the partially reflecting mirror ensures that light constantly leaves the cavity. Thus, it is possible to define an average number of round trips that a photon will make, meaning that $\delta\phi$ also has a maximum value.

The consequence is that each longitudinal mode has a finite width that plays a role in defining how sharply the laser emission is defined. Figure 3.2.2 shows an example of the longitudinal modes for a Helium-Neon (HeNe) laser operating at 632 nm (length 20 cm, refractive index 1.0), and a laser diode operating at 1550 nm (length 5 mm, refractive index 3.2).

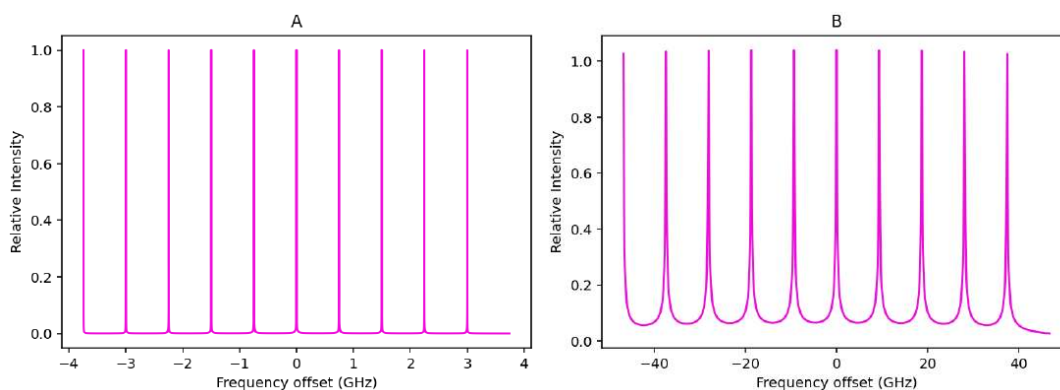


Figure 3.2.2: Longitudinal modes around the central frequency of a Helium Neon laser (A) and a laser diode (B). Note that the central frequency has been subtracted to make the horizontal axis readable.

The HeNe laser has a much closer mode spacing (smaller FSR) than the laser diode. The longitudinal modes of the HeNe laser appear to be delta functions because the reflectivity of the output coupling mirror must be quite high (~98%), while the gain of a laser diode is so high that the natural reflection from the end facets of the semiconductor material is sufficient.

When we combine the fact that the gain material also has a finite bandwidth with the longitudinal mode spectrum of the optical cavity, we get the potential output spectrum of a laser (shown in Figure 3.2.3). However, when we measure the spectrum of a continuous wave laser (one that emits with a constant power), it is highly unlikely that a spectrum like that in Figure 3.2.3 will be observed.

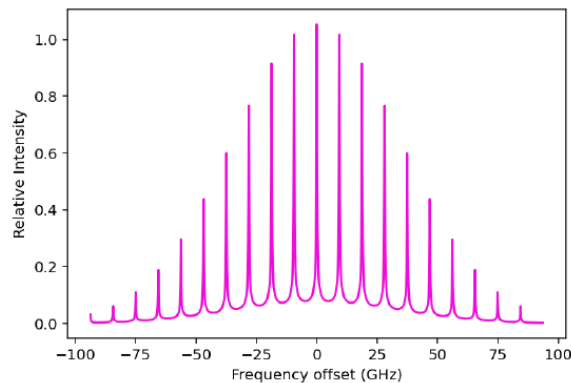


Figure 3.2.3: The gain spectrum of a laser. The gain medium has a finite bandwidth, which reduces the gain of longitudinal modes that do not coincide with the peak of the gain spectrum.

There are two reasons for the observed emission spectrum being different from the gain spectrum. First, the amount of gain at any one moment is finite, and is dependent on the *intensity* of light. The mode with the most gain will start first, and because it has a higher intensity, will also grow fastest. This will reduce the gain available to the other modes, slowing their growth even more. This suppresses most modes. This is illustrated in Figure 3.2.4 that shows the results of a rate equation calculation for three competing longitudinal models with different gains.

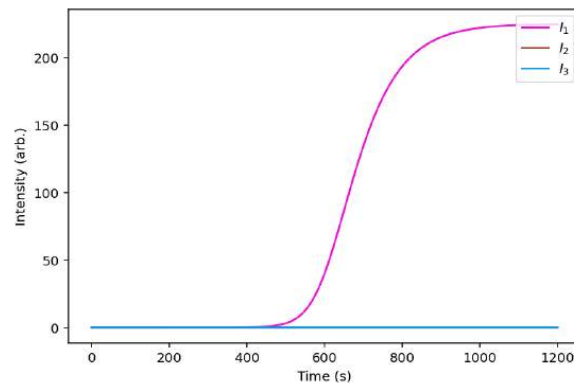


Figure 3.2.4: Rate equation calculation for three longitudinal modes. All three modes have the same starting conditions, but experience slightly different gains.

Gain competition can be useful, since we can, by cavity design, increase the gain of the longitudinal mode that we want, then gain competition will ensure that it is the mode most likely to be observed at the output. The second reason that we do not observe the full gain spectrum at once, but also don't observe a stable single longitudinal mode is due to the fact that the gain is distributed in space. Each longitudinal mode forms a standing wave between the mirrors. A standing wave has nodes and anti-nodes in fixed spatial locations, which means spatial locations where the intensity is always high and spatial locations where the intensity is always low. Thus, there is not full competition for the same gain, and more than a single mode can oscillate (see Figure 3.2.5). However, the overlap between the different modes is sufficient to suppress *most* modes at any one time.

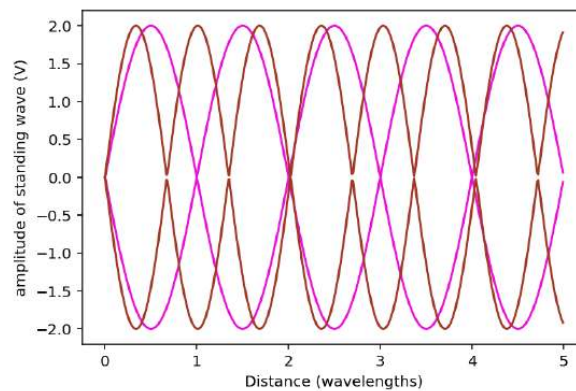


Figure 3.2.5: Two standing waves. Note that the brown trace has peaks at 1 and 3 wavelengths. At these locations, the gain for the purple mode is lower than that of the brown.

Furthermore, the competition between modes rarely leads to a stable emission spectrum. As a mode grows, it exhausts the gain where its standing wave intensity is highest, reducing the gain that it experiences. The gain elsewhere is available, but the mode currently dominant does not have the highest gain in that location because the standing wave intensity is low. Thus, the mode begins to die away, and a new mode begins to gain strength. This leads to a dynamic emission spectrum with longitudinal modes constantly growing and dying.

? Exercise 3.2.1

The code to calculate the output of a laser with gain competition between three modes is below. Is it possible to get more than a single mode to start? How different can the gains (σ in the code) be?

```
1 import numpy as np
2 from matplotlib import pylab as plt
3 from scipy.constants import c
4 from scipy.integrate import solve_ivp
5
6 #conditions for the simulation
7 in_cond = [1, 0, 1e-6, 1e-6, 1e-6] #initial conditions. iorder is N1, N2, I
8 ##### these are the gain factors for the three different modes
9 sigma1 = 0.05 #sigma1: gain of mode 1
10 sigma2 = 0.03 #sigma2: gain of mode 2
11 sigma3 = 0.02 #sigma3: gain of mode 3
12 #####
13 T = 0.02 #T: output 0-1
14 P = 15 #P: input power 0-infinity
15 loss = 0.00000 #k: losses 0-1
16 t_span = [0,1500] #time over which the simulation will run. Adjust as necessary
17 #but always start at 0
18
19
20 #These are the coupled differential equations
21 def coupledDE(t, out, sig1, sig2, sig3, pump, trans, k):
22     ret = [0,0,0,0,0]
23     ret[0] = sig1*(out[1]-out[0])*out[2]-pump*out[0] +sig2*(out[1]-out[0])*out[3] +sig3*(out[1]-out
24     ret[1] = -sig1*(out[1]-out[0])*out[2] - k*out[1] -sig2*(out[1]-out[0])*out[3] -sig3*(out[1]-out
```

run

run all

restart

restart & run all

<matplotlib.legend.Legend at 0x7f2f5f225460>

Answer

It depends quite a bit on the settings of the numerical solver. When the gain factors are identical, all three modes will run. When they are within about 0.1% of each other, the modes will start up, but those with less gain will slowly decay with

time. What this shows is that the steady state solution has not been found yet, and after a very long time, the mode with lower gain will still stop emitting.

This page titled [3.2: Optical cavities](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

3.3: Gain materials

When describing the gain of a laser, it was abstracted as a population of emitters with an excited state, N_2 capable of emitting light and a state that the emitter returned to after emitting light, N_1 . The condition for net gain is $N_2 > N_1$, called population inversion, which seems simple enough. In reality very few materials allow this condition to be satisfied. Consider just one aspect of the rate equation: to obtain population inversion, we must input energy: $\frac{dN_2}{dt} = P(N_1 - N_2)$. But the process to drive emitters from N_1 to N_2 is symmetrical, which means P will also drive population from N_2 to N_1 . Therefore, the best that we can hope for is that $N_2 = N_1$. To obtain population inversion, we must search for materials that have a rather special energetic structure, such as those shown in Figure 3.3.1. In Figure 3.3.1, the vertical scale is energy, and the horizontal lines represent allowed energy states, indicated by $|i\rangle$, of the emitter. The arrows indicate the different processes from the rate equations. Each allowed state has a characteristic lifetime, τ_i , which is the time it takes to halve the number of emitters in state $|i\rangle$.

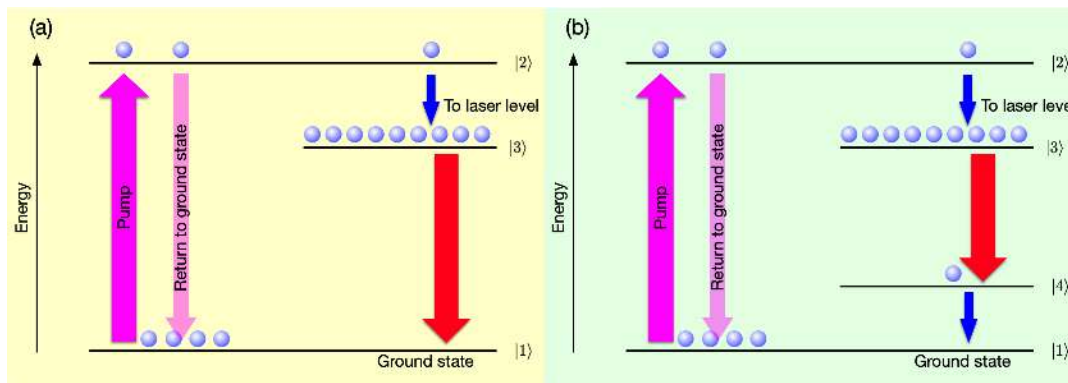


Figure 3.3.1: Three level (a) and four level (b) laser systems. The pump (purple) drives ions (blue spheres) from the ground state to $|2\rangle$. A small number of ions end up in $|3\rangle$ (blue arrow), where they are trapped. When population inversion occurs, then can emit together to generate laser light (red arrow).

The three level system in Figure 3.3.1a achieves population inversion because $|3\rangle$ has a very long lifetime, so even though the driving excitation from $|1\rangle$ to $|3\rangle$ is symmetric (so with a very high level of excitation $N_2 = N_1$), there is always a small fraction of emitters that go from $|2\rangle$ to $|3\rangle$, where they are trapped for a long time. If the excitation rate is fast enough, eventually population inversion is achieved ($N_3 > N_1$). This, however, can be rather difficult since $|1\rangle$ is the lowest energy state of the emitter, so N_1 is usually large. As a result, a three-level laser is generally less energy efficient in terms of excitation compared to a four-level laser (see Figure 3.3.1b), but the so-called quantum efficiency can be quite high. The quantum efficiency is a measure of the energy lost within the emitter before emission. In the case of a three-level laser, energy is lost going from $|2\rangle$ to $|3\rangle$, and the rest contributes to laser output. The energy that is lost to internal processes usually results in the generation of heat.

A four level system makes achieving population inversion easier because the laser emission is not directly to the lowest energetic state, so as long as the lifetime of $|4\rangle$ is very short, then $N_4 \approx 0$. Excitation proceeds in the same way as for a three-level laser: emitters are driven from $|0\rangle$ to $|2\rangle$, where some fraction of N_2 goes to $|3\rangle$ rather than returning to $|1\rangle$. There, population can accumulate, and, since $N_4 \approx 0$, laser action can be achieved at much lower excitation rates. As a result, four-level lasers are often more energy efficient, but have lower quantum efficiency because energy is lost internally in two transitions: $|2\rangle \rightarrow |3\rangle$ and $|4\rangle \rightarrow |1\rangle$. Three and four level lasers are not always distinctly separate: the same material can sometimes operate as either a three level laser or a four level laser, depending on the way the laser is setup.

Table 3.3.1 shows a list of common gain materials (excluding semiconductors), and their emission wavelength range. The materials that are used in telecommunications are also noted.

Table 3.3.1: common non-semiconductor laser gain materials.

Gain material	Wavelength range (nm)	Application
Nd:YAG	1064	Test & Measurement Machining
Er:glass	1500-1600	Telecommunications
Ti:sapphire	700-1050	Microscopy
Yb:glass	1000-1100	Test & Measurement Machining
Alexandrite	700-820	Tattoo removal
Ar ion	515	light shows
KrF	248	Lithography
ArF	198	Lithography
CO ₂	10600	Machining

This page titled [3.3: Gain materials](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

3.4: Diode lasers

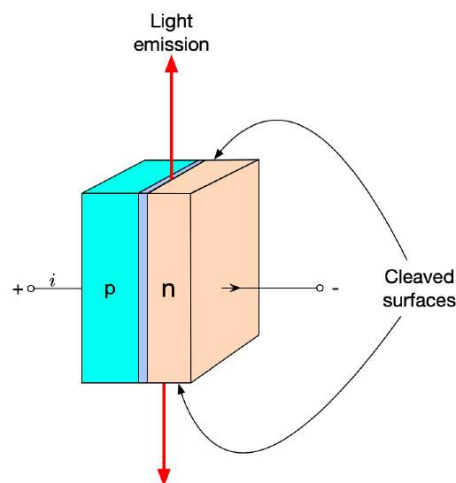
Although there are many different types of lasers, the development of the diode laser enabled cheap optical communications. The most important feature of semiconductor gain materials is their flexibility. Essentially, elements from group III (aluminium, Gallium, and Indium) are alloyed with group V elements (Nitrogen, phosphorous, Arsenic, and Antimony) to obtain the desired range of emission wavelengths. These can be pair-wise or more complicated alloys, where the ratio between different elements is given by $x : (1 - x)$, $y : (1 - y)$. Table 3.4.1 shows the *potential* wavelength range covered by an alloy, however, it is not always possible to cover the full range because the alloying elements may separate out during crystallization.

Table 3.4.1: Combinations of elements commonly used to make LEDs and laser diodes. The wavelength range is approximate.

Group III	Group V	Wavelength range (nm)
$\text{Ga}_x\text{Al}_{(1-x)}$	N	200-400
$\text{Al}_x\text{In}_y\text{Ga}_{(1-x-y)}$	N	200-1600
$\text{In}_{(1-x)}\text{Ga}_x$	$\text{As}_{(1-y)}\text{P}_y$	500-4000

Semiconductors do not have tightly defined energy states, instead, there are two energy bands: the conduction band, and the valence band, which are separated by some amount of energy, called the bandgap. If a current is injected, this creates a population of electrons in the conduction band, and a population of missing electrons (called holes) in the valence band. Light emission occurs when electrons and holes recombine, and together, these populations correspond to N_2 with $N_1 \approx 0$. This mechanism, which is quite different from atomic and ionic emitters, means that population inversion is achieved, despite the fact that the lifetime for excited emitters is very short (picoseconds to nanoseconds). These characteristics also result in a very high gain (up to $\sim 300 \text{ cm}^{-1}$). On the other hand, the internal losses for both emitters (electrons recombining with holes without emitting light) and light (absorption and scattering) is quite high compared to other gain materials.

A simplified diagram of a laser diode is shown in Figure 3.4.2. In more sophisticated designs, the gain material is sandwiched between two related semiconductors. The conduction and valence bands of the outer materials do not match that of the gain material, which traps the electrons and holes in the gain material. This reduces losses and increases the efficiency of the gain. If the laser diode is designed to be edge emitting, as shown in Figure 3.4.2, then the outer materials may also act to confine the guide to the light within the gain material, which also increases the efficiency of the laser.



3.4.2: simplified diagram of a laser diode. A depletion region (blue) forms at the junction between the p and n doped materials. The light emission is in the direction of the edge (red arrows). Mirrors are formed by cleaving the semiconductor along a crystal face to make a smooth a highly reflecting surface. The refractive index difference between the p, n and depletion regions create a slab waveguide.

This page titled [3.4: Diode lasers](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

3.5: Bandwidth control

In a telecommunications system, there are agreements about which wavelengths can be used for communication (see Table 3.5.1). Notice that the channel bandwidths are as small as 12.5 GHz, which means that the wavelength needs to have a precision of about 0.007%. At the same time, a simple laser can operate on several different longitudinal modes with frequency jumps of ~ 10 GHz, which is clearly not acceptable. To obtain the necessary precision, the emission spectrum of the laser must be controlled.

Table 3.5.1: ITU grid spacings for optical communications. Frequencies are specified in THz (1 THz is 10^{12} Hz).

12.5 GHz spacing	25 GHz spacing	50 GHz spacing	100 GHz spacing
195.9375	X	X	X
195.9250	195.925	X	X
195.9125	X	X	X
195.9000	195.900	195.90	195.9
195.8875	X	X	X
195.8750	195.875	X	X

Previously, we looked at how the rate equations and the material gain combine to suppress longitudinal modes: the mode with the highest gain wins. In that case, the difference in gain was due to material characteristics. However, the optical cavity can also be designed so that it provides more feedback to some wavelengths compared to others, as shown in Figure 3.5.1. The same mechanism that leads to suppression of longitudinal modes enhances the selectivity of the cavity. Thus, jumps between longitudinal modes are limited to only those modes within the selected range of the cavity, which, hence, the design of the cavity should only have a single longitudinal mode within the selected range.

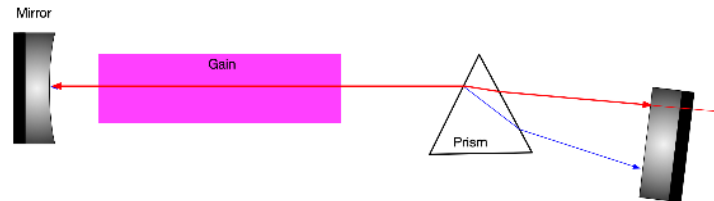


Figure 3.5.1: The prism causes the shorter wavelengths (blue arrow) to be bent more than the red. The red light is reflected back through the gain medium, while the blue light is not. Hence, only the red light experiences gain and the laser outputs red light. Note that the dispersion (degree of bending) is exaggerated in this figure.

In Figure 3.5.1, the two mirrors that create the cavity are separated by a prism which bends the light depending on the wavelength. This type of dispersion limits the longitudinal modes of the cavity to those that fall within a small range of angles from the prism. The optical cavity can be made more selective by choosing a glass that is more dispersive, or by increasing the length of the cavity. Note, however, that as the length increases, the FSR becomes smaller, so, even though the emission bandwidth goes down, the number of longitudinal modes may not decrease.

Clearly, the type of cavity shown in Figure 3.5.1 is unsuitable for optical communications, if for no other reason than the size of the laser. Over the last 30-40 years many different external cavity laser diode designs have been used, but, all modern laser diodes in telecommunications make use of integrated designs.

3.5.1: Distributed feedback lasers

All of the optical cavities discussed so far consist of two mirrors, one of which is partially reflective. However, this does not need to be the case. Consider the optical cavity shown in Figure 3.5.2a. In this three mirror cavity, the center mirror is partially reflective and the output coupler is partially reflective. These three mirrors create three optical cavities (labeled 1, 2 and 3), each with their own set of longitudinal modes, given by $\frac{2l_1}{c}$, $\frac{2l_2}{c}$, and $\frac{2(l_1+l_2)}{c}$. Now consider a light wave incident from the left on the center mirror:

part of the wave will be reflected, and so must (for the amplitude to be maintained) be a mode of cavity 1. However, it is also transmitted, so it also has to be a mode of cavity 3. Likewise, when the wave is reflected from the output coupler and returns to the center mirror from the right, part is transmitted and part is reflected. Thus, the wave considered from this perspective must also be a mode of cavity 2 and cavity 3. Hence, the longitudinal modes of the whole system are only those modes that are common to all three cavities. Examining Figure 3.5.2b, we can see that the mode spacing of the system is much larger than that of each of the individual cavities.

Once we also take into account the reflectivities of the mirrors (recall that highly reflective mirrors define the longitudinal mode more sharply), the longitudinal modes are no longer a set of evenly spaced peaks of equal amplitude, but have a periodic structure in intensity as well. The influence of spacing, reflectivity, and number of mirrors can be explored in Exercise 3.5.1.

? Exercise 3.5.1

The script below calculates the longitudinal mode spectrum for cavities with any number of mirrors. The script is setup to cover the telecommunications band in Table 3.5.1. Play with the script and answer the following questions

- Setup a three mirror cavity. How does changing the spacing between mirror 2 (the center mirror) and mirror 3 change the longitudinal mode spectrum?
- For a three mirror cavity, how does changing the reflectivity of mirrors 2 and 3 change the longitudinal mode spectrum?
- For a three mirror cavity, is it possible to obtain just a single mode with an intensity above 0.1 within the L band if you only adjust the spacing between mirrors 2 and 3?
- Setup a 100 mirror cavity, with reflectivities of 2% and a spacing of 750 nm. What does the longitudinal mode spectrum look like?
- Can you adjust the mirror spacing so that there is only a single mode within the L band?
- Can you tune the cavity so that the single mode is coincident with a specified wavelength in Table 3.5.1?

Code goes here

Answer

Add texts here. Do not delete this text first.

In a distributed feedback (DFB) laser, as shown in Figure 3.5.3, each "mirror" has only a tiny reflectivity (1-5%), but there are many mirrors, creating many different cavities. As a result, the output is confined to a very specific wavelength, defined by the separation of the mirrors. The wavelength can be tuned very slightly by adjusting the temperature of the laser (thermal expansion changes the distance between the mirrors). These designs allow the creation of lasers that operate with the very tight specifications required to satisfy the requirements given in Table 3.5.1. Unfortunately, a single DFB laser cannot be tuned to any and all wavelengths in Table 3.5.1. For that a distributed Bragg reflector (DBR) laser is required.

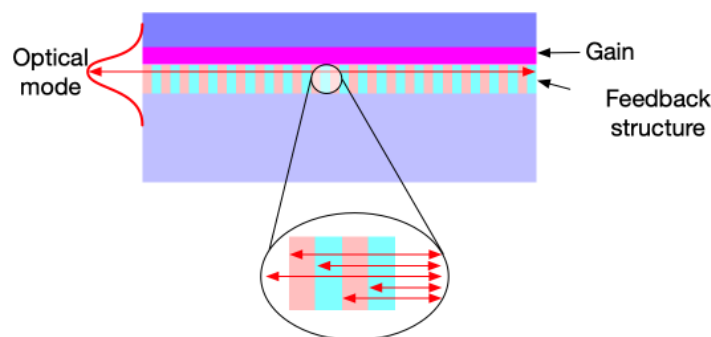


Figure 3.5.3: Distributed feedback laser. The optical mode (red) is amplified by the gain layer, while a periodic structure (light blue and peach) reflects light back and forth. A zoom illustrates an example of this. The condition for a wavelength being a mode of the laser is that all reflections from all interfaces must add up in phase with each other (constructive interference). This is only satisfied for a very narrow range of wavelengths. (b)

3.5.2: Distributed Bragg Reflector Laser

The Bragg reflector operates on similar principles to the optical cavity, but can be more simply described. As shown in Figure 3.5.4, each reflective surface is separated by $\frac{\lambda}{2}$. Light waves at a wavelength of λ constructively interfere with each other and are strongly reflected. In addition to the Bragg reflectors, the DBR laser has a passive phase section, where an applied voltage adjusts the effective length of the laser cavity. Between tuning the Bragg mirrors, and adjusting the phase section, a single laser can be adjusted to emit over ranges between 10-50 nm, depending on the exact construction.



Figure 3.5.4: Distributed Bragg Reflector laser. Operates on a similar principle, but the reflectors are only on the edge and a phase section allows for extended tunability.

This page titled [3.5: Bandwidth control](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

3.6: Fiber amplifiers

In long distance communication, losses in the optical fiber play an important role. Consider that the distance from Amsterdam to New York is 5,880 km, while a typical long haul optical fiber has losses of around 0.17 dB/km, meaning that the total attenuation is about 1000 dB. A typical receiver might be reliable for signals as low as -16 dBm (25 μ W), meaning that the input power would have to be $1000 - 16 = 984$ dBm (2.5×10^{95} W). Even if the receiver had a very high sensitivity, of -60 dBm, which corresponds to a single photon per bit at a 10 Gb/s data rate, then the input power still has to be unreasonably high.

A typical DFB laser has an output power of 10 dBm (10 mW), meaning that 26 dB is the maximum acceptable attenuation. Thus, the maximum length of fiber is about 150 km, which does not take into account losses due to connecting (splicing) lengths of fiber together. Giving ourselves some margin, we need 58 amplifiers between Amsterdam and New York.

Amplification can take place in two ways: the optical signal can be detected, converted to an electrical signal, then returned to the optical domain by modulating an optical source, or an amplifier that directly amplifies the optical signal can be used. In practice both are used for reasons that will, hopefully, become clear later.

Optical amplification works by the same process as a laser: a population inversion is achieved by exciting a medium, and the input light is amplified by the same process, including the competition for gain. A typical amplifier consists of an optical fiber that has been doped with the element erbium, called an erbium doped fiber amplifier (EDFA). Erbium has a relatively high gain over the C band, as shown in Figure 3.6.1. Despite a factor of four difference in emission cross section, the resultant gain is relatively flat. Nevertheless, the wavelength dependence of the gain must be compensated for. In order to ensure that each channel in the fiber gets equal gain, the signal must be spectrally shaped before amplification: essentially a filter should be shaped such that the gain-intensity product is constant across the C band: $G(\lambda)I(\lambda) = k$.

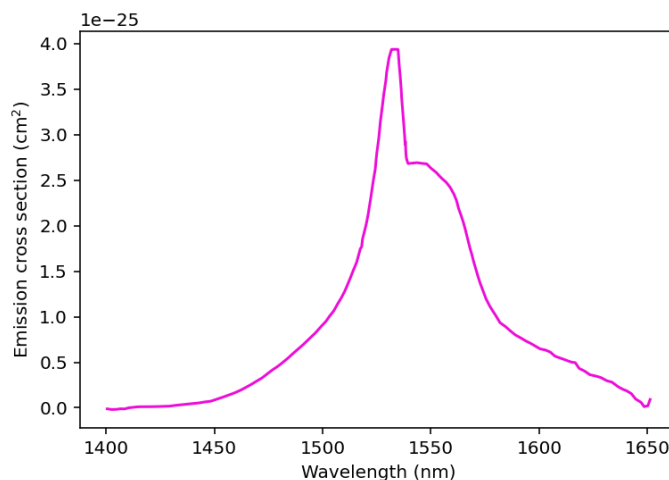


Figure 3.6.1: Emission cross section of Erbium (larger is better). There is significant gain between 1500-1600 nm.

3.6.1: Excitation

EDFAs are optically excited. A pump laser with a wavelength of 980 nm drives erbium ions into the excited state, where emission in the C band can take place. Thus, the gain is proportional to the population inversion as described for laser gain through stimulated emission. There is also a second gain mechanism called Raman amplification. In this case, emission from an excited state population is not involved, instead, a nonlinear optical effect converts pump photons to signal photons via the generation of vibrations in the glass. Since vibrations in solids are low frequency (compared to the frequency of light), Raman amplification requires a pump wavelength much closer the signal wavelength (see Figure 3.6.2).

Both amplification methods have their benefits: at low signal powers, amplification via Erbium ion emission is much more efficient. However, the gain is intrinsically limited by the number of ions available. At a certain pump power, all ions are available, and no more gain is possible. On the other hand, Erbium has many more energetic states available, so at high pump powers, the ions often get driven out of the excited state into other states that do not contribute to gain (called excited state absorption). Raman

gain grows with pump power, but the gain is very low at low signal powers, and, because the pump wavelength is close to the signal wavelength, it can be difficult to filter the pump out of the channel.

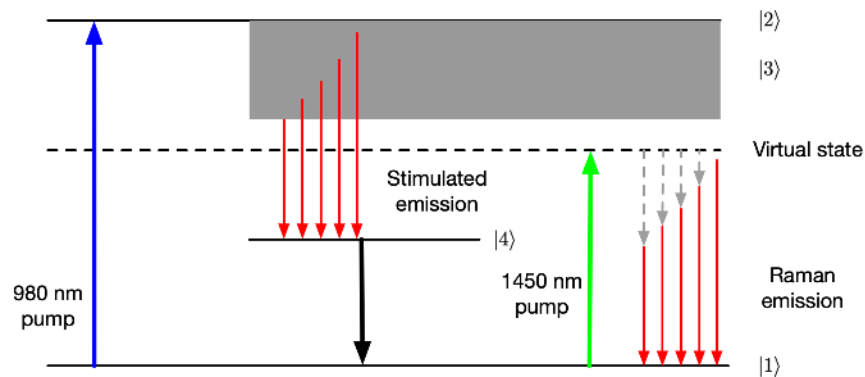


Figure 3.6.2: Conceptual energy level diagram for an EDFA (the real energy level diagram is far more complex). The laser emission state ($|3\rangle$) is a set of states that overlap in energy (grey box), creating a band. At high emission energy (bluer wavelengths), the laser behaves like a three-level laser, while for redder wavelengths, the laser is more like a four-level laser (red arrows). Raman emission is driven by a 1450 nm pump (green arrow), which stimulates Raman emission (red arrows) along with high frequency vibrations in the fiber (grey arrows).

Note that the two amplification techniques complement each other to great extent. It is, therefore, common to combine both in a single amplifier. The basic geometry of a double pumped EDFA is shown in Figure 3.6.3. The Erbium doped gain fiber is ~ 100 m long and *double clad*, with the 980 nm pump traveling in the same direction as the signal. This provides optimum small-signal gain, but, as the signal grows, the gain drops because the Erbium concentration is not large enough. The remainder of the gain fiber is glass without doping. The Raman pump (at 1450 nm) is injected in the opposite direction to the signal, so that the pump power is highest where the signal is most intense, ensuring that the gain grows as the signal intensity grows.

Designing a fiber amplifier involves a balance between pump *intensity*, ion doping concentration, and fiber length.

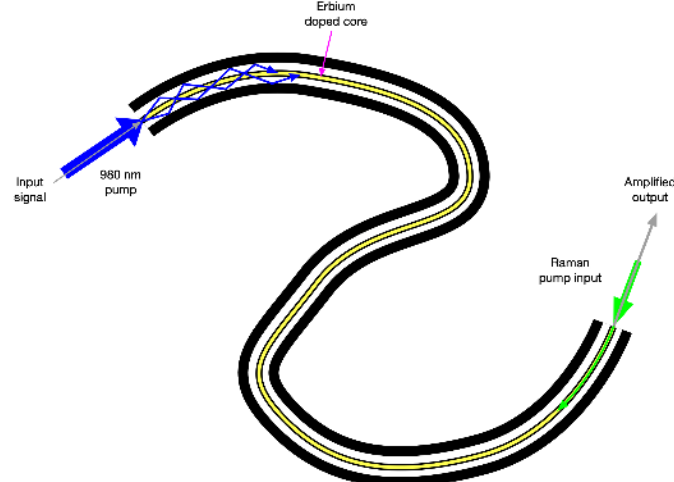


Figure 3.6.3: Double pumped EDFA. The 980 nm pump (for stimulated emission) is guided in the cladding in the same direction as the signal. The Raman pump is injected in the reverse direction and provides gain for the signal once it has already been amplified by stimulated emission. In the configuration shown here, the Raman pump is guided in the core.

3.6.2: Amplified spontaneous emission

The amplification process that we have described so far relates to stimulated emission, where the presence of the signal light wave drives an excited ion to emit light, amplifying the wave. However, if left to its own devices, an excited emitter will eventually emit a photon on its own, *generating* a new light wave. This light wave can then be amplified by other excited ions, in a process called amplified spontaneous emission (ASE).

ASE is generated in parallel to signal amplification, creating a growing noise floor that the signal must overcome. Figure 3.6.4 shows how ASE grows after each amplifier stage. The growth of ASE and optical losses are used to calculate the optimum spacing between amplifiers, and the maximum number of stages before the signal must be converted back to the electrical domain. We begin by stating that over the length of the link, the amplification must exactly compensate for the losses (the total optical power in the signal band is, on average, constant over the length of the link)

$$G = \alpha L_s \quad (3.6.1)$$

where G is the gain in dB, α is the per-km loss (dB/km) and L_s is the length of fiber between amplifiers. The ASE is given by

$$P_n = 2\eta_{sp}hfN(G-1)\Delta f \quad (3.6.2)$$

where $\eta_{sp} \approx 2$ is the inversion factor, h is Plank's constant, f , Δf are the optical center frequency and bandwidth, and N is the number of amplifiers. Since the total power is a constant, the signal to noise ratio is

$$\text{SNR} = \frac{P - P_n}{P_n} \quad (3.6.3)$$

In a communications link, the specified bit error rate defines the minimum acceptable SNR, while the sensitivity of the detector defines the minimum optical power at the detector. Thus, we can combine Equations (3.6.1) – (3.6.3) to find the maximum number of amplifiers.

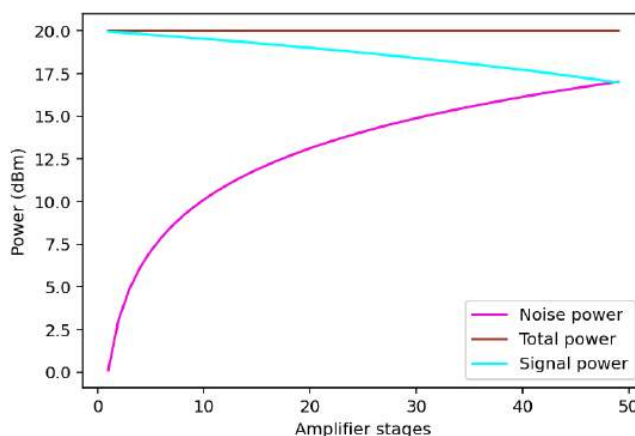


Figure 3.6.4: Noise growing in a fiber link. In this calculation, the noise power for a single channel is calculated. It is assumed that each amplifier stage is set to compensate exactly the *total* losses between each amplifier stage.

✓ Example 3.6.1

An optical fiber link (center wavelength: 1550 nm) with a single channel data rate of 10 Gb/s is to be deployed between Amsterdam and New York. The total length of passive fiber is 6000 km, with an average attenuation of 0.17 dB/km. The input power 10 dBm, and the amount of input noise is negligible. To obtain a bit error rate (BER) of 10^{-9} SNR ≥ 7 is required. The detectors require an average flux of 20 photons per bit to reliably detect a signal. The chosen optical amplifiers have a gain of 50 dB

Calculate the number of optical amplifiers and electronic repeaters required for the link.

Solution

First step: calculate the power (in dBm) at the receiver: 20 photons per bit = $\frac{20 \times 10^9}{2}$ photons per second. The energy per photon is $E_p = \frac{hc}{\lambda} = 1.3 \times 10^{-19}$ J. So, the average signal power at the end detector is $E_p \times (5 \times 10^9) = 0.64$ nW (-62 dBm). This, as can be seen below is not the limiting factor.

The total loss due to the fiber absorption is $0.17 \times 6000 = 1020$ dB, while the gain of the amplifier is 50 dB, indicating that we need 21 amplifier stages at 286 km intervals. The ASE, given by Equation (3.6.2) is 126 μ W per amplifier (-9 dBm). After 21 stages, the ASE is 2.6 mW, thus the SNR is (from Equation (3.6.3)) $\frac{10^{-2.6}}{2.6} = 2.8$, which implies a BER $> 10^{-9}$.

Setting $P_n = NP_{amp}$ where P_{amp} is the ASE from a single stage, equation (3.6.3) can be rearranged to $N = \frac{P}{P_{amp}(SNR+1)}$. Filling in the numbers we get $N \approx 10$. Thus, we will need a single electronic repeater, which replaces amplifier number 10.

Note that this leaves no margin for error and does not take into account other losses. This includes losses when fibers are spliced together, losses for dispersion compensating fiber, losses due to spectral shaping filters, and more. Thus, three electronic repeaters at positions 5, 12, and 19 is a better choice.

Example 3.6.1, illustrates how noise grows in a fiber link. This idea is illustrated in Figure 3.6.4, and is a good place to start when thinking about a *link budget*.

3.6.3: Link Budget

When planning to deploy a fiber link, the requirements will generally give a data rate and a maximum acceptable BER. The data rate, B and the average number of photons per bit $\bar{n}$ required to meet the BER requirement lead give us the minimum acceptable power at the receiver in dBm.

$$P_r = 10 \log \left(\frac{\frac{hc}{\lambda} \bar{n} B}{1 \times 10^{-3}} \right) \quad (3.6.4)$$

It is never a good idea to operate at the minimum, so some margin should be added, P_m . Attenuation losses can be split into two components: power loss per splice/coupling P_c , power loss due to attenuation of the fiber, αL . Thus, the link budget for absorption losses is given by

$$P_s - N_c P_c - \alpha L = P_r \quad (3.6.5)$$

where N_c is the number of splices and couplers.

On top of that the pulse will spread out due to the dispersion of the material (we consider only single model fibers), such that

$$LB = \frac{1}{4|D_\lambda \sigma_\lambda|} \quad (3.6.6)$$

where D_λ is the dispersion coefficient (ps/km.nm), and σ_λ is the the bandwidth of the pulse. If we assume the best case scenario (the minimum bandwidth for a pulse of a given duration), then $\sigma_\lambda = 0.75 \frac{\lambda^2}{c} B$. After a number of substitutions related to the dynamics of pulses that have a Gaussian shape, equation (3.6.6) becomes

$$LB^2 = \frac{\pi}{4} \frac{c}{|D_\lambda \lambda^2|} \quad (3.6.7)$$

From equations (3.6.5) and (3.6.7), it is possible to graphically see how desired bit rate and link length are related to each other (see Figure 3.6.5).

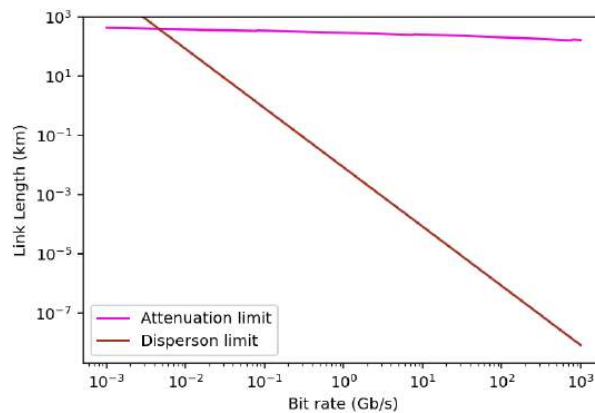


Figure 3.6.5: How dispersion and attenuation limit the maximum bit rate of a link. Without dispersion compensation, high-speed, long distance links are impossible.

The dispersion can be compensated by using two different fibers such that $D_\lambda L + D'_\lambda L' = 0$, where D'_λ is the dispersion coefficient of a second fiber that has *anomalous* dispersion in the transmission band. These fibers have higher loss, but their dispersion is typically much stronger, so only short lengths are required to compensate for the dispersion of the normally dispersive fiber.

In a typical communication system, the entire C and L band might be used, while the dispersion compensating fiber is chosen to only exactly compensate in the middle of the communication band. Thus, there is residual positive and negative dispersion for channels adjacent to the center channel. However, this residual dispersion is still smaller than the uncompensated dispersion, and allows for much longer links.

This page titled [3.6: Fiber amplifiers](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

CHAPTER OVERVIEW

4: Integrated optical components

[4.1: Introduction](#)

[4.2: Mach-Zehnder Interferometers](#)

[4.3: Add-drop multiplexers](#)

[4.4: Arrayed waveguide grating](#)

[4.5: Photodiodes](#)

This page titled [4: Integrated optical components](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

4.1: Introduction

In the previous chapters we have discussed how to guide light and generate light. However, an optical network would not be complete if we were unable to control where the light went, or modulate the light to communicate. Although there are many different ways to do each of these tasks, we will only cover a few solutions that are based on *integrated optics*, which happen to be most common in optical communications. Integrated optical circuits are the photonic equivalent of integrated circuits for electronics. Instead of wires, waveguides control the flow of light, while electronic and thermal variations are used to switch the light.

Integrated optics makes use of three basic fundamental properties: the refractive index of materials is temperature dependent, and, for some materials, the refractive index also changes with an applied DC (or low frequency) voltage. These two properties allow for switching. The third property is that of directional coupling: when two waveguides are in close proximity to each other, the light from one waveguide will flow into the other waveguide. The consequences of these three properties are outlined below, and are used to explain the functional properties of a number of devices, including:

- Mach-Zehnder interferometers
- Add-Drop multiplexers
- Array waveguide gratings
- Modulators

4.1.1: Integrated optics: materials

Integrated optical circuits can be fabricated in a number of materials. Glass (silicon oxide) on silicon wafers is a common choice, as is silicon nitride on silicon wafers. These material choices offer very low loss, and can be controlled using temperature. However, glass devices (in contrast to silicon nitride) have to be quite large. Indium phosphide provides for more compact designs, at the expense of higher losses. Indium phosphide can also provide on-chip sources (lasers and LEDs) and detectors (photodiodes). In addition, Indium phosphide, being a semiconductor can be used to control light via various high-speed mechanisms (electro-optic modulation, for instance). A final option that should be mentioned is lithium niobate. Lithium niobate is closer to glass in terms of losses, but has a strong electro-optic effect, providing for high-speed modulation. Unfortunately, circuit fabrication in lithium niobate is more difficult compared to the other materials (which use standard lithography process developed for the semiconductor industry), thus lithium niobate is mainly limited to high-speed modulation.

4.1.2: Electro and thermo-optical effect

The electro and thermo-optical describe how the refractive index of a material changes with temperature or applied voltage. The change in refractive index with respect to temperature is often given by a Sellmeier equation, which is an empirical fit to measured data, given by:

$$n^2 = A_0 + A_1 f(T) + \sum_{i=0}^{i=N} \frac{B_i + D_i f(T)}{\lambda^2 - (C_i - E_i)^2} \quad (4.1.1)$$

where $f(T)$ is a temperature dependent correction, which may be of the form

$$f(T) = (T - k_0)(T + k_1) \quad (4.1.2)$$

It is important to remember that equations (4.1.1) and (4.1.2) are general examples. The most basic form of the Sellmeier equation can be derived from a general understanding of the interaction of light with matter. However, the goal is to provide the most accurate function of refractive index with wavelength and temperature, and, therefore, deviations from the general form are very common. Table 4.1.1 shows the measured constants for lithium niobate. Even here, care must be taken, because *how* lithium niobate is grown (it is a crystal and must be grown), and what dopants are added will change the Sellmeier equation. Hence, it is important to know the details of the material type before a published Sellmeier equation is used.

Table 4.1.1: coefficients for Sellmeier equation for lithium niobate with an additional term that does not fit the general form of equation (4.1.1).

	$i = 0$	$i = 1$
A coefficients	5.35583	4.629×10^{-7}
B coefficients	0.100473	100
C coefficients	0.20692	11.34927
D coefficients	3.862×10^{-8}	2.657×10^{-5}
E coefficients	0.89×10^{-8}	0
k coefficients	24.5	570.82
Additional term	$1.5334 \times 10^{-2} \lambda^2$	

The electro-optic effect is similar: the refractive index is linearly or quadratically dependent on an applied voltage. The quadratic dependence can be ignored for our purposes, leaving only the linear dependence, called the Pockels effect, which is given by:

$$n(E) \approx n - \frac{1}{2} \tau n^3 E \quad (4.1.3)$$

where n is the refractive index when the applied electric field, $E = 0$, and τ is the Pockels coefficient, which has typical values of between 1-100 pm/V ($p = \text{pico} = 10^{-12}$). With such a small coefficient, it initially seems impossible to make use of the Pockels effect in practical high-speed devices. The field is dependent on the applied voltage divided by the distance between the terminals. In a material like lithium niobate ($\tau = 25$ pm/V), the wafer thickness might be between 100-500 μm , meaning that 5 V results in 10-50 kV/m, and a change in refractive index of 6.7×10^{-6} . Thus, a clever design is required.

To give an impression of how these effects are used, consider two waves that are combined at a 50/50 beam splitter, as shown in Figure 4.1.1. The phase relationship between the two waves determines the balance between the intensities at ports 1 and 2.

$$\begin{aligned}
 E_{p1} &= \frac{E_1 + E_2}{2} \\
 I_{p1} &= E_{p1} E_{p1}^* \\
 &= \frac{1}{4} (E_1 E_1^* + E_2 E_2^* + E_1^* E_2 + E_1 E_2^*)
 \end{aligned} \quad (4.1.4)$$

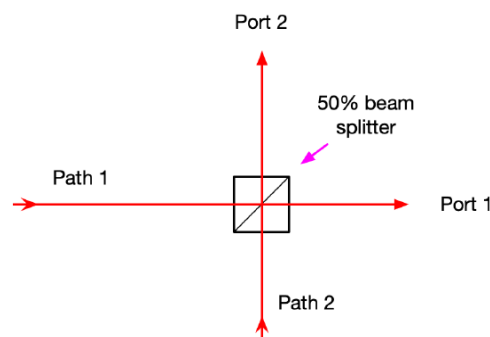


Figure 4.1.1: the combination of two light waves via a 50% reflective mirror.

The first two terms in equation (4.1.4) are constants given by the amplitudes of the individual waves. The second two terms are so-called interference terms and determine at which port the light will exit. To make that clear, let us assume that $E_1 = E_0 \exp(-jk_1 l_0)$ and $E_2 = E_0 \exp(-jk_2 l_0)$, where $k_1 = 2\pi n_1/\lambda$ and $k_2 = 2\pi n_2/\lambda$. We can, thus, see that $E_1 E_1^* = E_0^2$ and that $E_2 E_2^* = E_0^2$, but the interference terms are

$$\begin{aligned}
 E_1^* E_2 &= E_0^2 \exp(j(k_1 - k_2)l_0) \\
 E_1 E_2^* &= E_0^2 \exp(-j(k_1 - k_2)l_0) \\
 E_1^* E_2 + E_1 E_2^* &= E_0^2 (\exp(j(k_1 - k_2)l_0) + \exp(-j(k_1 - k_2)l_0)) \\
 &= E_0^2 (\cos(k_1 - k_2)l_0 + j \sin((k_1 - k_2)l_0) + \cos((k_1 - k_2)l_0) \\
 &\quad - j \sin((k_1 - k_2)l_0)) \\
 &= 2E_0^2 \cos((k_1 - k_2)l_0)
 \end{aligned}
 \tag{4.1.5}$$

$$\tag{4.1.6}$$

Recombining the results from equation (4.1.6) with equation (4.1.4) and recognizing that the intensity is the square of the amplitude, leads to

$$I_{p1} = \frac{I_0}{2} (1 + \cos((k_1 - k_2)l_0)) \tag{4.1.7}$$

Thus, the intensity at ports one and two depend on the refractive index of the medium through which waves one and two propagated

$$k_1 - k_2 = \frac{2\pi}{\lambda} (n_1 - n_2) = k_0 \delta n \tag{4.1.8}$$

to fully redirect the intensity from port 1 to port 2, the cosine term must change from unity to negative one, which requires a change of π . Thus, in the balanced case considered above, $k_0 \delta n l_0 = 0.5$ is necessary. If we consider that $\lambda_0 = 1550$ nm and $\delta n = 6.7 \times 10^{-6}$, then we see that $l_0 \approx 13$ m. However, this number is without any cleverness in the design, and typical lengths are 2-3 mm.

4.1.3: Waveguide coupling

The field in a waveguide is partially in the core and partially in the cladding. The natural phase velocity of the wave in the cladding is slightly faster than the phase velocity in the core. However, the phase velocity doesn't instantaneously change as the wave crosses from cladding to core because that would cause a discontinuity in the electric field amplitude. Instead, the wave in the cladding gradually speeds up as it moves further from the core, creating a curved wavefront (a line of constant phase) that focuses the wave's energy into the core (this is another way to think about how guiding works), as illustrated in Figure 4.1.2a. Now imagine that there is a second core placed close to the first. The wavefront must slow down again if it enters the core, so the wavefront in the cladding focuses energy into both cores, thus any energy from one core that is beyond a certain point in the cladding will be transported to the other core (see Figure 4.1.2b).

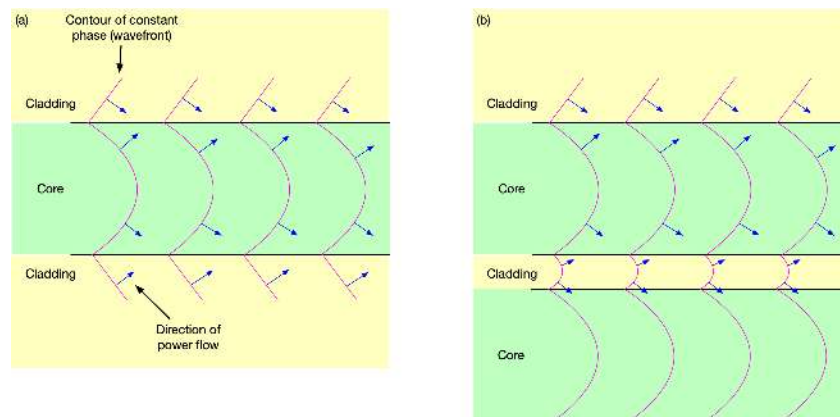


Figure 4.1.2: (a) the wavefronts (pink) in a standard waveguide. Energy always flows at right angle to the wavefront, as indicated by the blue arrows. The outward and inward energy flows balance for a mode of the waveguide. (b) For two waveguides that are close enough to have overlapping fields, the power flow is no longer balanced, and energy will travel from one waveguide to another.

The description above is an *over simplification* of the true situation. In fact, when two waveguide cores are close enough to couple, then in a distance, L_c , all the energy for a specific wavelength in one core will be transported to the other core. If the distance is doubled, all the energy will return to the original core. By controlling the separation between two cores and the distance at which they remain coupled, it is possible to choose how much energy is coupled from one core to another. And, because the required distance is wavelength dependent, it is also possible to separate different wavelengths from each other as well.

The coupling between two waveguides is given by a pair of coupled differential equations:

$$\frac{dE_1}{dz} = -jC_{21} \exp(j\Delta kz) E_2(z) \quad (4.1.9)$$

$$\frac{dE_2}{dz} = -jC_{12} \exp(-j\Delta kz) E_1(z) \quad (4.1.10)$$

where $\Delta k = k_1 - k_2$ is the mismatch in propagation constants between the two waveguides, and C_{12}, C_{21} are the coupling coefficients, which are themselves given by a so-called overlap integral

$$C_{21} = \frac{1}{2} (n_2^2 - n_{cl}^2) \frac{k_0}{k_1} \int_0^{a+d/2} u_1(y) u_2(y) dy \quad (4.1.11)$$

$$C_{12} = \frac{1}{2} (n_1^2 - n_{cl}^2) \frac{k_0}{k_2} \int_{-a-d/2}^0 u_1(y) u_2(y) dy \quad (4.1.12)$$

where a is the separation between the waveguides and d is width of the waveguide. Equations (4.1.11) and (4.1.12) tell us that the coupling coefficient is proportional to how much overlap there is between the modes ($u_1(y)$ and $u_2(y)$) in the two waveguides. Equations (4.1.9) and (4.1.10) have an analytical solution for parallel waveguides, which for the case of transferring power from waveguide one to waveguide two (and no power in waveguide two initially) is

$$P_1(z) = P_1(0) \left(\cos^2 \gamma z + \left(\frac{\Delta k}{2\gamma} \right)^2 \sin^2 \gamma z \right)$$

$$P_2(z) = P_1(0) \frac{|C_{21}|^2}{\gamma} \sin^2 \gamma z \quad (4.1.13)$$

$$\gamma = \left(\frac{\Delta k}{2} \right)^2 + C^2$$

$$C = \sqrt{C_{12}C_{21}}$$

For the case of two identical waveguides, then $n_1 = n_2$ and $\Delta k = 0$, then the transmitted field amplitudes are:

$$E_1(z) = E_1(0) \cos Cz - jE_2(0) \sin Cz \quad (4.1.14)$$

$$E_2(z) = -jE_1(0) \sin Cz + E_2(0) \cos Cz \quad (4.1.15)$$

An example of the power transfer for perfectly matched waveguides is shown in Figure 4.1.3. A 3 dB directional coupler requires a length of about 900 μm for this particular coupling coefficient, while complete power transfer occurs at about 1.8 mm

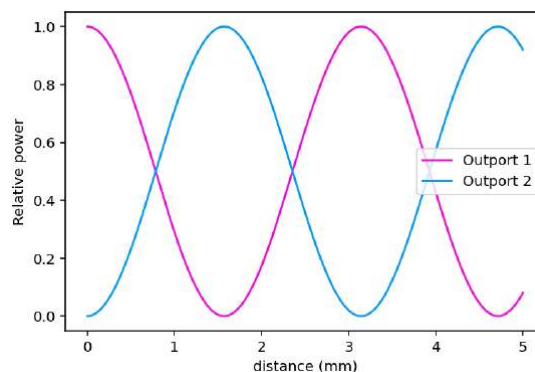


Figure 4.1.3: power transfer between two identical waveguides that are in close proximity to each other. By choosing the length of the coupler, the ratio of power transfer can be controlled. The vertical axis is the amount of power transferred relative to the power incident at input 1.

? Exercise 4.1.1

The code below models the coupling between two waveguides. The phasematch (Δk) and coupling coefficients can be adjusted to control the directional coupler's behaviour. What happens when the phasematch is not zero? What happens with the coupling constants are not symmetric?

Missing code

Answer

As the phase mismatch increases (absolute value), the maximum power transfer reduces. If the coupling coefficients are not identical and $\Delta k = 0$ we get see a kind of power loss: all the power leaves waveguide 1, but does not appear in waveguide 2. However, it is physically very difficult to achieve different coupling coefficients *and* $\Delta k = 0$. *If both* ($\Delta k = 0$ and the coupling coefficients are unequal, the power transfer discrepancy gets smaller, but does not completely disappear.

This page titled [4.1: Introduction](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

4.2: Mach-Zehnder Interferometers

A key component in integrated optical circuitry is the Mach-Zehnder interferometer (MZI). An MZI consists of two beam splitters that first split light so that it travels by two different paths, and is then recombined at the second beam splitter. The length of the two different paths changes the relative phase of the two light fields and, therefore, dictates the intensity balance between the two outputs. A schematic for a generic MZI is shown in Figure 4.2.1(a) and the integrated optical version is shown in Figure 4.2.1(b).

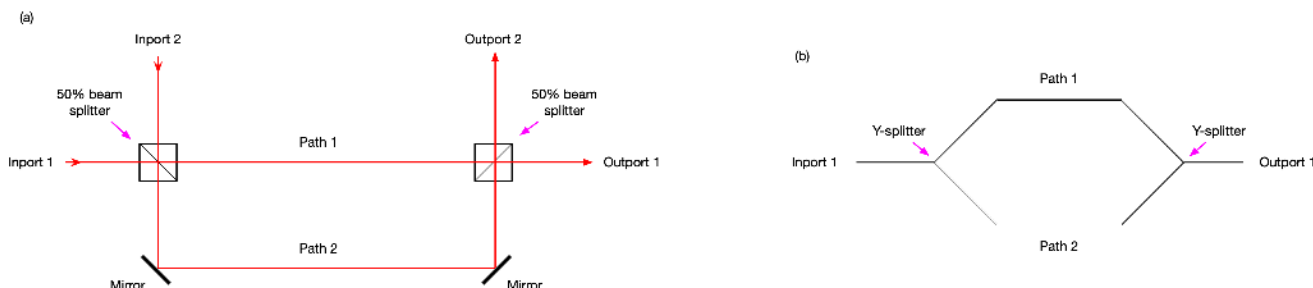


Figure 4.2.1: (a) a classical free space MZI. (b) an integrated optical MZI.

In Figure 4.2.1(a) it is clear that an MZI has two inports and two outports, while for Figure 4.2.1(b) only one inport and one outport are visible. However, the missing ports still exist. When light is supposed to exit the second outport, it is scattered out of the integrated optical circuit, while the second inport is also scattered light coming *into* the integrated optical circuit. Since light scattered into the circuit is tiny, it can be neglected, while the purpose of the MZI is to control the light intensity at the *guided* outport, so we don't need to care about light scattered out of the outport.

The beam splitters are also different. In Figure 4.2.1(a), the beam splitters are 50% reflective mirrors, meaning that they divide the power across the whole mode evenly between the two paths. In Figure 4.2.1(b), the mode is split by Y connectors, which act a bit like a scalpel, sending the left half of the mode along one track, and the right half of the mode along a different track. Thus, to obtain 50/50 splitting/combining, a Y connector must be highly symmetrical and the optical mode must also be symmetric about the splitting point.

MZIs are generally used as modulators and coarse filters. They also form an integral part of the fine filters that are used in multiplexers. From the previous section, we know that the intensity at the output of an MZI is given by

$$I_o = \frac{1}{4} (E_1 E_1^* + E_2 E_2^* + E_1^* E_2 + E_1 E_2^*) \quad (4.2.1)$$

where I_o is the output intensity and E_1, E_2 are the *complex* amplitudes of the light fields that travel by the two different paths to the combiner. At the Y combiner, the two fields are given by

$$E_i = E_0 \exp \left(\frac{2\pi}{\lambda_0} n_i l_i \right) \quad (4.2.2)$$

where the subscript i denotes the path, λ_0 is the wavelength in free space. In this formulation, the phase difference between the two paths is given by the difference in the factor $n_i l_i$. Following the same mathematical arguments as the previous section, we get the same formula for the output, but with an emphasis on the *physical* form and changes of the MZI

$$I_o = \frac{I_i}{2} \left(1 + \cos \left(\frac{2\pi}{\lambda_0} (n_1 l_1 - n_2 l_2) \right) \right) \quad (4.2.3)$$

An MZI with thermally controlled optical path length differences is often used as a highly stable but tunable filter, while an electro-optically controlled optical path length difference is more suitable for modulation, with GHz modulation frequencies are common.

This page titled [4.2: Mach-Zehnder Interferometers](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

4.3: Add-drop multiplexers

In wave division multiplexing (WDM), information is sent in parallel on different wavelengths. In general, we want to be able to be able to multiplex and demultiplex in a controlled way. One way to do this is with an add-drop multiplexer, as illustrated in Figure 4.3.1. An add-drop takes advantage of the linearity and symmetry of most optical systems: if input 1 takes 20 channels in and channel 3 is directed to output 2, then channels 1,2,4-20 go to output 1. That also means that if a signal on channel 3 is sent into input 2, it will also appear on output 1. Hence, we have simultaneously dropped channel 3 and added a new channel 3 to the multiplexed signal on output 2.

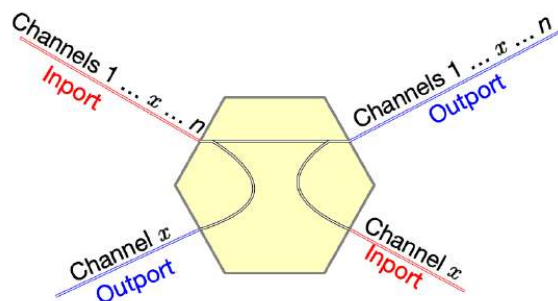


Figure 4.3.1: conceptual diagram of an add/drop multiplexer designed to simultaneously drop and add channel x . Channel x is separated from the other channels at the left hand input and redirected to the left hand output. A new channel x is added via the right hand input and combined with the remaining channels to exit at the right hand output.

WDM channels can be as closely spaced as 12.5 GHz, however, it is much more common to use 50 and 100 GHz spacing. Given that the carrier frequency is about 195 THz, this means that an add-drop must have a frequency selectivity better than 0.05%, which cannot be achieved with something as coarse as an MZI. Instead, microring resonators (MRR) are used. The ring, pictured in Figure 4.3.2 is an optical resonator: light that is coupled in to the ring from the top waveguide must be a longitudinal mode of the resonator. As with a resonator consisting of two mirrors, the phase of the light must be an integer multiple of 2π after a round trip, which, in this case has a physical length of $l = 2\pi r$, where r is the radius of the ring. However, the ring is also a waveguide with a refractive index n_{eff} , so the resonance condition is satisfied when $f = \frac{mc}{2\pi rn_{eff}}$, where m is an integer.

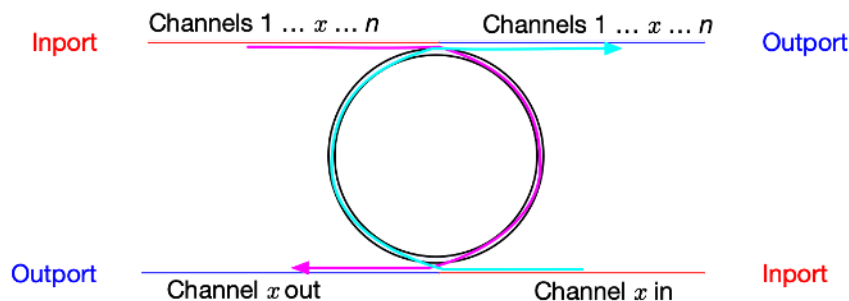


Figure 4.3.2: Microring resonator add/drop multiplexer. The ring waveguide diameter is such that channel x is resonant, allowing it to couple in from the left hand input waveguide and out through the right hand output. Likewise the new channel x couples in from the right hand input, and out through the right hand output. The other channels are **not resonant** with the MRR and travel from the left hand input to the right hand output unaffected.

As with the optical resonators discussed earlier, the FSR of the MRR is a comb of possible frequencies that fall within the transparency range of the material used to fabricate the integrated optical circuit. An example of such a comb is shown in Figure 4.3.3 for a ring with a radius of 10 μm and $n_{eff} = 3$. Note that the spacing is between modes is very large (about 1.5 THz)

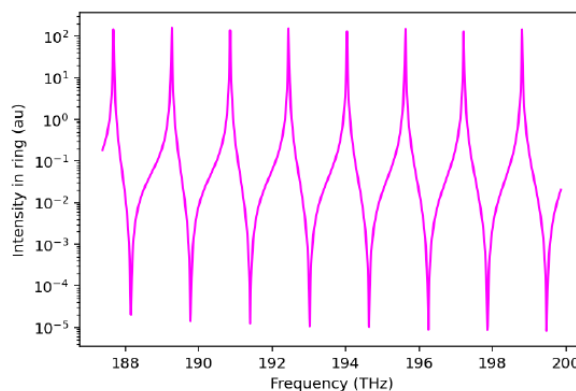


Figure 4.3.3: longitudinal modes of a 10 μm radius MRR in the C and L bands.

The parameters chosen for Figure 4.3.3 are reasonable for high-index materials like InP or silicon. However, for materials like silicon nitride, the radius of the MRR must be larger, leading to a smaller FSR. If we compare Figure 4.3.3 to the ITU grid spacing for 100 GHz DWDM ($D = \text{dense}$), then we expect 5 of the 80 channels to be resonant. Hence, a single MRR add-drop, as shown in Figure 4.3.4a is insufficient to the purpose. There are two possible solutions, shown in Figures 4.3.4b and 4.3.4c. One solution is to take advantage of the fact that the drop port contains 5 channels with a separation of 1.5 THz, so a coarse filter, like an MZI or a directional coupler can be used to separate the desired channel from the five remaining channels.

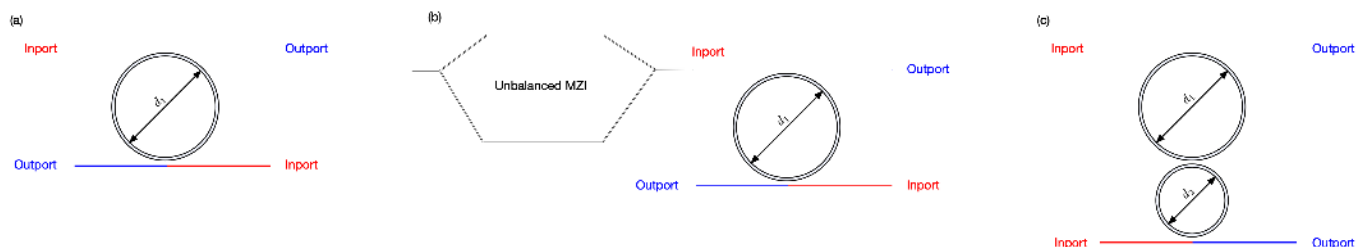


Figure 4.3.4: A standard MRR add/drop multiplexer, but limited control over which channels are added or dropped. By first filtering the input with an MZI (b), a specific channel within the passband of the MZI can be chosen. However, this will also remove channels from outside the passband of the MZI from the output of the MRR add/drop. Using two MRRs with different ring radii (c) allows the add/drop to be tuned such that only a single channel is dropped or added.

Alternatively, one can use two rings in place of a single ring. This works on the same principle as a DFB laser: each ring has a set of longitudinal modes, but, because the light couples from the waveguide to ring 1 to ring 2 and thence to a second waveguide, the light in the system as a whole must be a longitudinal model of *both* rings simultaneously. Thus, the FSR of the system becomes much larger, limiting the transmission to the drop port to a single channel, as shown in Figure 4.3.5.

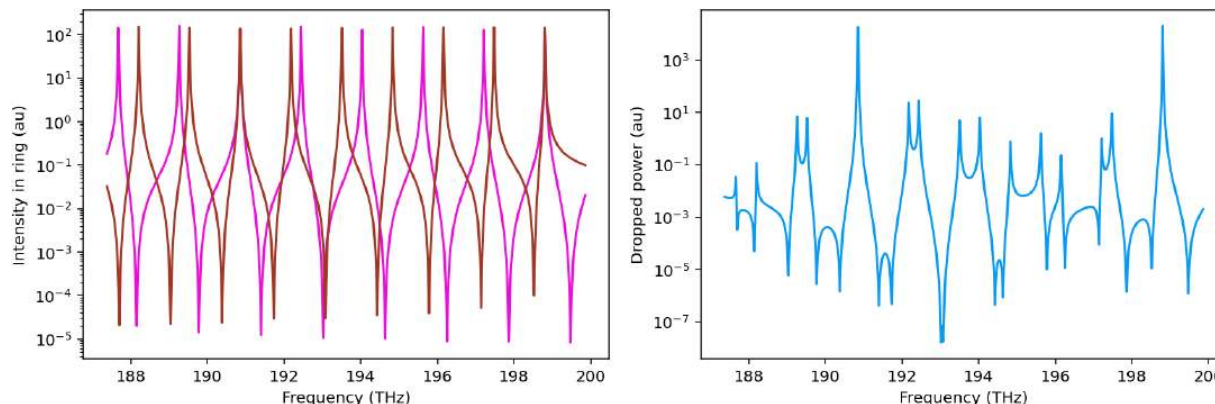


Figure 4.3.5: longitudinal modes of MRRs with radii of $10\text{ }\mu\text{m}$ (fuchsia) and $12\text{ }\mu\text{m}$ (brown). There is only a single common mode for the two rings within the C and L bands. This can be more clearly seen in the longitudinal modes of the system (blue). Note the vertical scale is logarithmic and that the peak on the far right is outside the L band.

So far we have discussed the FSR of MRRs and how combining rings will lead to a greater FSR. However, another aspect is the bandwidth of the add-drop multiplexer. Consider that the data rate of a single channel might be as high as 40 Gb/s, meaning a bandwidth of at least 40 GHz is required. We saw with laser how the reflectivity of the mirrors also defined the bandwidth of the cavity. A MRR has no mirrors, but it does have a per-round-trip loss, and the waveguides that couple light in also couple light out, thus, the bandwidth of an MRR can be tuned to match the bandwidth of the communication system.

? Exercise 4.3.1

The code below allows you to calculate the transfer function of a two ring add-drop multiplexer. You can adjust the radii of the two rings, and the coupling between the rings and the waveguides. Tune (adjust the radii) the multiplexer so that it drops channel 18. How big a change in radii is required to drop channel 19? Zoom in on the next largest peak within the C or L band. What channel is closest? What is the relative cross talk from the chosen channel to channel 19?

This page titled [4.3: Add-drop multiplexers](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

4.4: Arrayed waveguide grating

In the previous section, the add-drop multiplexer was discussed, which is a way to add and/or drop a single channel from a WDM system. It is certainly possible to cascade add-drops to create a WDM link with an ever-changing group of channels. It is also possible to do this in a single device called an arrayed waveguide grating. The arrayed waveguide grating (AWG) looks a *bit* like a very complex MZI, but it is easier to understand it as a type of diffraction grating.

Consider, as illustrated in Figure 4.4.1, light reflected from a grooved surface with period, a . The light in direction θ is the sum of fields from light scattered from each groove, and will be highest when those fields are in phase, which is given by

$$m\lambda = a(\sin \phi + \sin \theta) \quad (4.4.1)$$

where ϕ is the angle of incidence of the light, and m is an integer. What equation (4.4.1) tells us is that light will be separated by wavelength, as each diffracts at a different angle. But, underlying that is that each groove contributes a different "path", and all paths have to result in constructive interference. In Figure 4.4.1, this is illustrated on the right hand side. Two incident light rays are shown, but one of them has to travel an additional distance d_i before it hits the grating surface. Two diffracted rays are also shown, and one of them must travel an additional distance d_r . These two distances must add to an integer number of wavelengths for constructive interference to occur, which results in equation (4.4.1). With a "real" diffraction grating in free space, the paths continuously vary between some minimum value to some maximum value. But, we don't necessarily need all those paths.

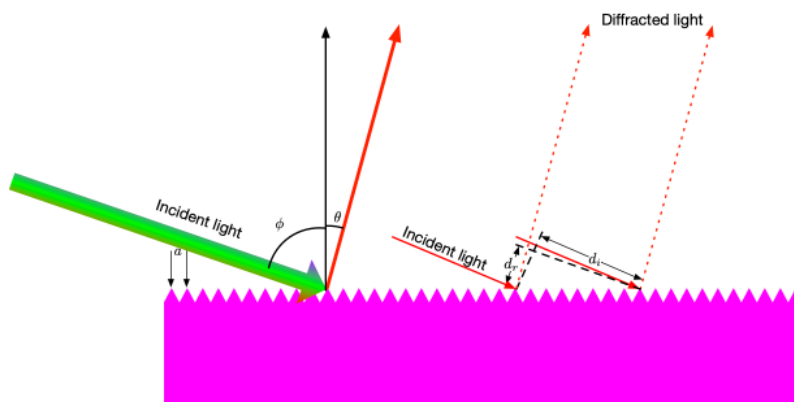


Figure 4.4.1: diffraction grating. White light (multicolored arrow) is incident from the left at angle ϕ and a specific color leaves the grating at angle θ (left hand side). The grooves each act as a source of light (scattering the incoming light) with a phase delay due to their separation (right handside). The result is that each source constructively interferes with all the others at an angle that *depends* on the wavelength of light. This separates the light into its separate colors.

An AWG, as illustrated in Figure 4.4.2, first spreads the light, then captures the light in a number of different waveguides, which are all different length. The light exits these waveguides into a common space where it can recombine. For each channel, constructive interference will occur in a unique direction, so, by placing a receiving waveguide at those angles, the AWG demultiplexes all channels in a single step. Since the AWG is a linear device, the reverse is also true: all channels can be multiplexed into a single waveguide in one step.

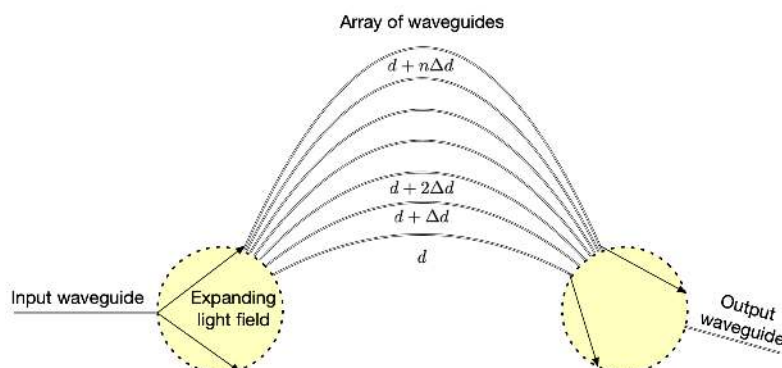


Figure 4.4.2: Arrayed waveguide grating. Light enters from the left and expands evenly in a slab waveguide (light yellow), each waveguide captures a small amount of the light and carries it to a second slab waveguide where the light mixes. Due to the different paths, the interference between the different light fields ensures that each output waveguide (only one is shown) will only receive a small range of wavelengths. Since an AWG is symmetrical, inputs and outputs can be swapped, thus an AWG can demultiplex from a single input, or multiplex from multiple inputs.

The transmission function of an AWG is given by

$$T = \frac{\sin^2\left(\frac{\pi L \Delta d}{\lambda}\right)}{\sin^2\left(\frac{\pi \Delta d}{\lambda}\right)} \quad (4.4.2)$$

where L is the number of waveguides and Δd is the incremental increase in waveguide length. Figures 4.4.3a - 4.4.3d show the transfer function of four AWGs with a varying number of waveguides and a varying incremental increase in length. Note that the transmission appears to increase with wavelength. This is because the grating is not centered. To center the grating, we need to satisfy $\lambda_c = \frac{n_{eff} \Delta d}{m}$, where for simplicity, $n_{eff} = 1$. Not included in

The code that produced Figures 4.4.3a - 4.4.3d is available in Exercise 4.4.1, where you can design your own AWG.

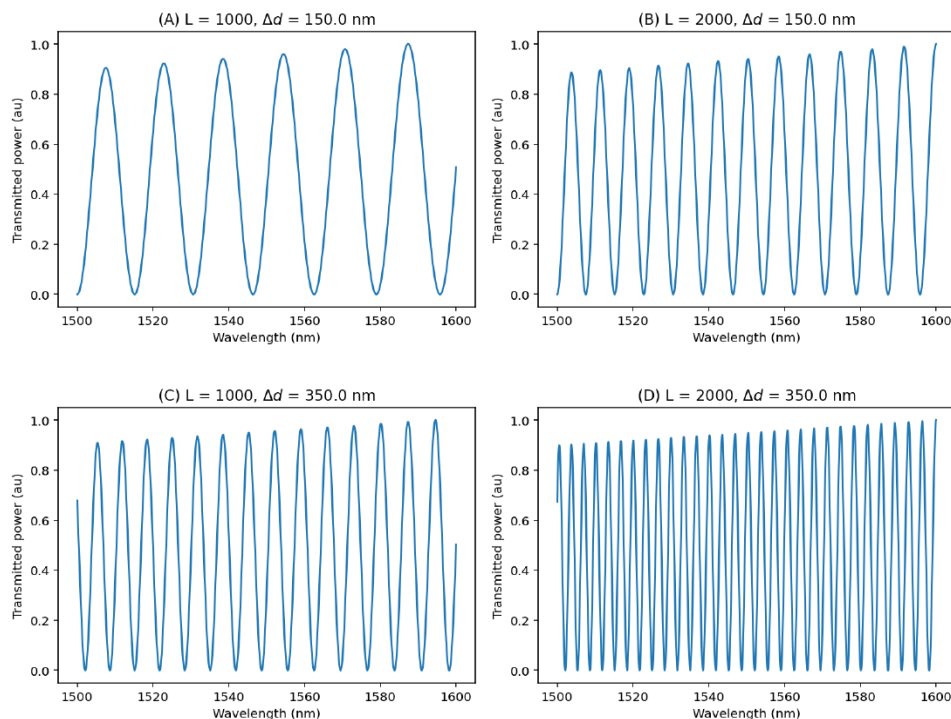


Figure 4.4.3: AWG transfer function for four different combinations of L and Δd

? Exercise 4.4.1

The code below calculates the transfer function of an AWG. Set Δd to 1550 nm. Is the transfer function of the AWG centered on 1550 nm as expected (you may need to adjust L to see whole pattern). In a real AWG, the number of waveguides is limited to about 100. Set the effective refractive index to 2.0 (corresponding to a silicon nitride waveguide) and limit the number of waveguides to 100. Adjust Δd to obtain a multiplexer that can multiplex at least 10 channels from the ITU grid. What is the channel spacing do you obtain?

```
1 import numpy as np
2 from matplotlib import pylab as plt
3 from scipy.constants import c, h
4
5 ld = np.linspace(1500e-9, 1600e-9, 2000) #wavelength range
6 n_eff = 1
7 L = 2000 #number of waveguides in the array
```

```
8 dd = 250e-9 #incremental increase in distance
9 u1 = (np.sin(np.pi*L*n_eff*dd/ld))**2
10 l1 = (np.sin(np.pi*n_eff*dd/ld))**2
11 T = u1/l1
12 T = T/T.max()
13 plt.figure()
14 plt.plot(ld*1e9, T, color="xkcd:fuchsia")
15 plt.xlabel("Wavelength (nm)")
16 plt.ylabel("Transmitted power (au)")
```

run

run all

restart

restart & run all

```
Text(0, 0.5, 'Transmitted power (au)')
```

Answer

Yes, setting Δd to 1550 nm centers the pattern on 1550 nm. To get 10 channels within the ITU channel range, a Δd of about 250 nm is required, then by careful movement of Δd , you can come close to a good spacing and centering.

A basic understanding of AWGs is given by Equation (4.4.2), however, there is another degree of freedom that provides a bit more freedom. The input and output waveguides are placed on the circumference of a circle (radius R) with a spacing of s , which leads to a wavelength resolution of $\delta\lambda = \frac{n_c s^2}{m R^2}$, where n_c is the refractive index of the material that makes up the circle. Note that for a small $\delta\lambda$, a large circle is required, and the number of waveguides between the input and output also needs to be large, making for an unfeasibly large and complicated design. Thus, often more sophisticated designs with cascaded AWGs are often used.

This page titled [4.4: Arrayed waveguide grating](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Chris Lee](#).

4.5: Photodiodes

So far we have discussed how to generate light and amplify, how to modulate the intensity of light to send information, and how to control the flow of light. The final step in the process is to detect the signal. This is done with photodiodes and can be a separate device or part of an integrated optical circuit. Since not all materials support the detection of light (passive materials), this may also lead to hybrid devices, where integrated optical chips with photodiodes are glued to integrated optical devices that are entirely passive. One might choose to do this to manage losses: silicon nitride and glass are very low loss, so demultiplexing might best be performed on a chip made of silicon nitride, while detection requires an active material like silicon, or InP. Thus, the two chips must be bonded to each other.

A photodiode consists of a junction between two materials, one doped with an excess of holes (p-type materials) and one doped with an excess of electrons (n-type materials). Where the two materials join, the holes and electrons recombine to create a neutral region (called the depletion region). Like an ordinary diode, if a reverse bias voltage is applied, then electrons and holes are driven away from the junction, and no current flows. If a photon is absorbed in the depletion region, it will generate an electron and a hole (assuming the photon energy is greater than the bandgap of the material). Due to the reverse bias, the electron and hole are separated from each other, and a current flows (called a photocurrent). This is illustrated in Figure 4.5.1.

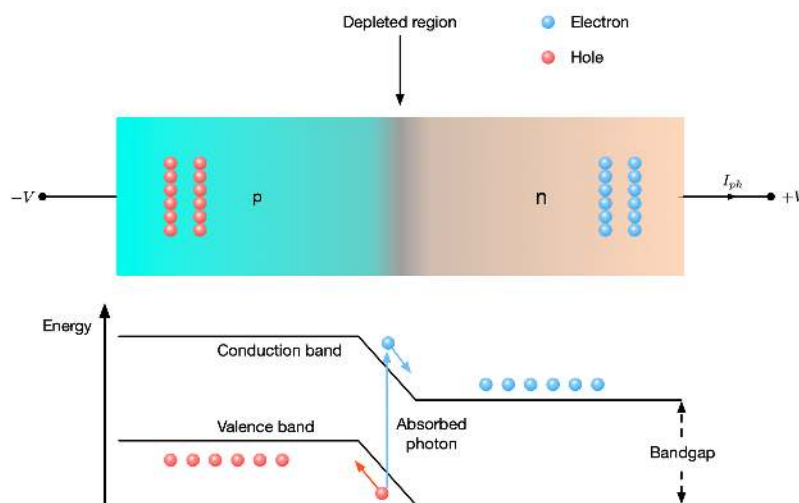


Figure 4.5.1: A photodiode. The p-doped and n-doped materials are brought together to form a junction. The excess positive and negative charges eliminate each other at the junction to create a depletion region (top diagram). When a photon is absorbed in the depletion region, the electron and hole are separated because the applied voltage drives them apart, creating a photocurrent (I_{ph}) that opposes the applied voltage.

Now, assuming that the absorption and conversion of photons to electrons has an efficiency of 100%, we can see that if the average number of photons per bit is 10, then a 10 GB/s data stream will result in an average photocurrent of 16 nA. This is certainly large enough to readily amplify with a transimpedance amplifier to give a digital level. The depletion region is, unfortunately, rather small, meaning that most of the light does not contribute to the photocurrent. Furthermore, the tiny depletion region results in a large junction capacitance (the capacitance is $C = \frac{\epsilon_r \epsilon_0 A}{d}$, where A is the junction area, so a small d means a high capacitance), which acts as a low pass filter. Thus, a photodiode as described is not an efficient transducer for a high-rate optical data signal. To improve performance, the depletion region is artificially extended by separating the p and n materials with an undoped material (called an intrinsic material), creating a pin photodiode. The intrinsic material has the effect of increasing the chance of a photon being absorbed in a neutral region, and reducing the capacitance of the photodiode.

Commercial high speed pin photodiodes have a rise times as short as 25 ps, and responsivities of ~ 1 A/W. Taking the numbers used above, this leads to a photocurrent of 13 nA.

To achieve a voltage of 1.3 V, a transimpedance gain of 100 M Ω is required, which is unrealistic in a single stage. Recall that the gain bandwidth product (GBP) is a constant, so to achieve a gain of 100 M Ω requires a GBP of at least 750×10^{15} Hz, while typical

values for high-speed opamps are between 50-8000 MHz. Thus, for very high speed applications, it may be necessary to use discrete components rather than opamps for the first stages.

Note

It is not necessary to bias photodiodes, since the current is generated by the absorption of photons. However, biasing reduces the junction capacitance and reduces the losses due to electron-hole recombination in the intrinsic region (increased responsivity). On the other hand, biasing increases the dark current (a current that flows even in the absence of light), and increases the noise. Thus, biasing must be carefully chosen to balance responsivity and noise. The connection of the photodiode to the measuring circuit must also be carefully designed. The rise time of the circuit is given by $\sqrt{RC}$, where R is the internal resistance of the photodiode and any external resistances. To keep the rise time high, R must be kept small.

The avalanche photodiode (APD) is a high-gain alternative to the pin diode. When a large bias voltage is applied to a photodiode, the electrons and holes are accelerated to very high energies. The electrons (and holes) will inevitably collide with atoms in the material, but now have enough energy to knock additional electrons free, which are themselves accelerated. This process can cascade (avalanche) such that a single absorbed photon generates thousands of electrons. The APD response time is limited by two factors: the junction capacitance and a dead time, which is the time the APD takes for the avalanche medium to recover. In order to reduce the dead time, most APD circuits quench the avalanche once the photocurrent reaches a certain threshold--in a digital system, the absolute value of the current is unimportant, only that it is above or below a set of threshold values. To increase the efficiency of the avalanche process, the avalanche region is separated from the area where photons are initially absorbed and the collector, as shown in Figure 4.5.2.

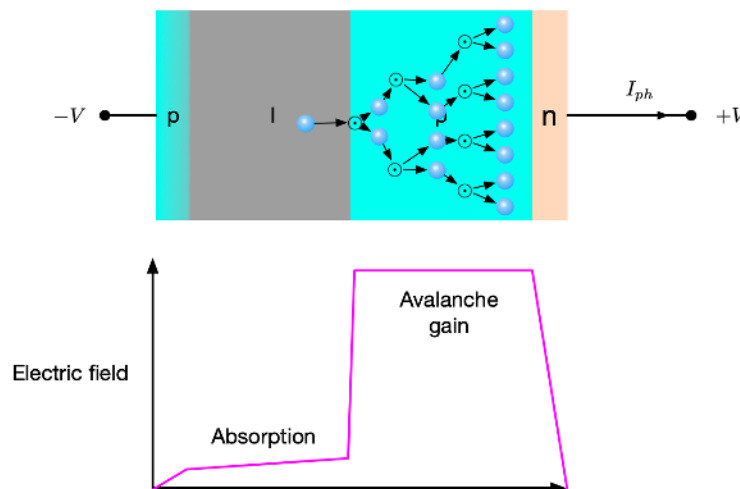


Figure 4.5.2: Avalanche photodiode. Light is absorbed in the intrinsic region (i), and the high electric field in the second p region causes impact ionization and gain. Holes are also generated with every impact, but these are not shown. For some semiconductors the gain for holes and electrons is not the same, while for others it is. In most designs, photons enter through the p-doped region on the lefthand side.

A typical avalanche photodiode (APD) can have a responsivity of more than 10 A/W. More importantly, APDs can be designed to detect single photons, which is important for advanced telecommunication applications, such as communication between satellites and quantum communication channels. It is important to note that using APDs does not allow one to escape the GBP dilemma: quenching is a way to limit the gain in order to increase the bandwidth. For extremely high bandwidth systems (100 Gb/s), APDs must be operated such that their responsivity is less than unity.

This page titled 4.5: Photodiodes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Chris Lee.

Index

1: D

dire

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

1: Overview

Title: [Optical Communications](#)

Webpages: 34

Applicable Restrictions: Noncommercial

All licenses found:

- [CC BY-NC-SA 4.0](#): 94.1% (32 pages)
- [Undeclared](#): 5.9% (2 pages)

2: By Page

- [Optical Communications](#) - [CC BY-NC-SA 4.0](#)
 - [Front Matter](#) - [CC BY-NC-SA 4.0](#)
 - [TitlePage](#) - [CC BY-NC-SA 4.0](#)
 - [InfoPage](#) - [CC BY-NC-SA 4.0](#)
 - [Table of Contents](#) - [Undeclared](#)
 - [Licensing](#) - [CC BY-NC-SA 4.0](#)
 - [1: Introduction](#) - [CC BY-NC-SA 4.0](#)
 - [1.1: The long tail](#) - [CC BY-NC-SA 4.0](#)
 - [1.2: SONET/SDH](#) - [CC BY-NC-SA 4.0](#)
 - [1.3: The importance of clocks](#) - [CC BY-NC-SA 4.0](#)
 - [2: Optical Waveguides](#) - [CC BY-NC-SA 4.0](#)
 - [2.1: Introduction to waveguides](#) - [CC BY-NC-SA 4.0](#)
 - [2.2: Spatial modes](#) - [CC BY-NC-SA 4.0](#)
 - [2.3: Modal dispersion](#) - [CC BY-NC-SA 4.0](#)
 - [2.4: Material and waveguide dispersion](#) - [CC BY-NC-SA 4.0](#)
 - [2.5: Different types of optical waveguide](#) - [CC BY-NC-SA 4.0](#)
- [3: Lasers and amplifiers](#) - [CC BY-NC-SA 4.0](#)
 - [3.1: Rate equations](#) - [CC BY-NC-SA 4.0](#)
 - [3.2: Optical cavities](#) - [CC BY-NC-SA 4.0](#)
 - [3.3: Gain materials](#) - [CC BY-NC-SA 4.0](#)
 - [3.4: Diode lasers](#) - [CC BY-NC-SA 4.0](#)
 - [3.5: Bandwidth control](#) - [CC BY-NC-SA 4.0](#)
 - [3.6: Fiber amplifiers](#) - [CC BY-NC-SA 4.0](#)
- [4: Integrated optical components](#) - [CC BY-NC-SA 4.0](#)
 - [4.1: Introduction](#) - [CC BY-NC-SA 4.0](#)
 - [4.2: Mach-Zehnder Interferometers](#) - [CC BY-NC-SA 4.0](#)
 - [4.3: Add-drop multiplexers](#) - [CC BY-NC-SA 4.0](#)
 - [4.4: Arrayed waveguide grating](#) - [CC BY-NC-SA 4.0](#)
 - [4.5: Photodiodes](#) - [CC BY-NC-SA 4.0](#)
- [Back Matter](#) - [CC BY-NC-SA 4.0](#)
 - [Index](#) - [CC BY-NC-SA 4.0](#)
 - [Glossary](#) - [CC BY-NC-SA 4.0](#)
 - [Detailed Licensing](#) - [CC BY-NC-SA 4.0](#)
 - [Detailed Licensing](#) - [Undeclared](#)

Detailed Licensing

Overview

Title: [Optical Communications](#)

Webpages: 34

Applicable Restrictions: Noncommercial

All licenses found:

- [CC BY-NC-SA 4.0](#): 94.1% (32 pages)
- [Undeclared](#): 5.9% (2 pages)

By Page

- [Optical Communications](#) - [CC BY-NC-SA 4.0](#)
 - [Front Matter](#) - [CC BY-NC-SA 4.0](#)
 - [TitlePage](#) - [CC BY-NC-SA 4.0](#)
 - [InfoPage](#) - [CC BY-NC-SA 4.0](#)
 - [Table of Contents](#) - [Undeclared](#)
 - [Licensing](#) - [CC BY-NC-SA 4.0](#)
 - [1: Introduction](#) - [CC BY-NC-SA 4.0](#)
 - [1.1: The long tail](#) - [CC BY-NC-SA 4.0](#)
 - [1.2: SONET/SDH](#) - [CC BY-NC-SA 4.0](#)
 - [1.3: The importance of clocks](#) - [CC BY-NC-SA 4.0](#)
 - [2: Optical Waveguides](#) - [CC BY-NC-SA 4.0](#)
 - [2.1: Introduction to waveguides](#) - [CC BY-NC-SA 4.0](#)
 - [2.2: Spatial modes](#) - [CC BY-NC-SA 4.0](#)
 - [2.3: Modal dispersion](#) - [CC BY-NC-SA 4.0](#)
 - [2.4: Material and waveguide dispersion](#) - [CC BY-NC-SA 4.0](#)
 - [2.5: Different types of optical waveguide](#) - [CC BY-NC-SA 4.0](#)
 - [3: Lasers and amplifiers](#) - [CC BY-NC-SA 4.0](#)
 - [3.1: Rate equations](#) - [CC BY-NC-SA 4.0](#)
 - [3.2: Optical cavities](#) - [CC BY-NC-SA 4.0](#)
 - [3.3: Gain materials](#) - [CC BY-NC-SA 4.0](#)
 - [3.4: Diode lasers](#) - [CC BY-NC-SA 4.0](#)
 - [3.5: Bandwidth control](#) - [CC BY-NC-SA 4.0](#)
 - [3.6: Fiber amplifiers](#) - [CC BY-NC-SA 4.0](#)
 - [4: Integrated optical components](#) - [CC BY-NC-SA 4.0](#)
 - [4.1: Introduction](#) - [CC BY-NC-SA 4.0](#)
 - [4.2: Mach-Zehnder Interferometers](#) - [CC BY-NC-SA 4.0](#)
 - [4.3: Add-drop multiplexers](#) - [CC BY-NC-SA 4.0](#)
 - [4.4: Arrayed waveguide grating](#) - [CC BY-NC-SA 4.0](#)
 - [4.5: Photodiodes](#) - [CC BY-NC-SA 4.0](#)
 - [Back Matter](#) - [CC BY-NC-SA 4.0](#)
 - [Index](#) - [CC BY-NC-SA 4.0](#)
 - [Glossary](#) - [CC BY-NC-SA 4.0](#)
 - [Detailed Licensing](#) - [CC BY-NC-SA 4.0](#)
 - [Detailed Licensing](#) - [Undeclared](#)