# Optics

Diffraction

Diffraction is the phenomenon where waves spread out when passing through a gap or passing an object. It is this phenomenon which, for example, enables you to hear the voice of a person next to you, as the sound coming from the larynx of the person spreads out when passing through his or her mouth. Moreover, if there were no diffraction, radio, television or other electromagentic waves and signals could not pass around mountains to spread into valleys. There are mainly three cases of the amount of diffraction that occurs when a wave passes a gap or an object that should be investigated. During the discussion of these three cases, λ denotes the wavelength of the relevant wave and a the size of the gap through which the or obstacle around which it passes.

1. No significant diffraction occurs if the gap a is very much larger than the wavelength, such that the wavefront passing through the gap is more or less parallel.

$$\frac{\lambda}{a} \ll 1$$

Waves
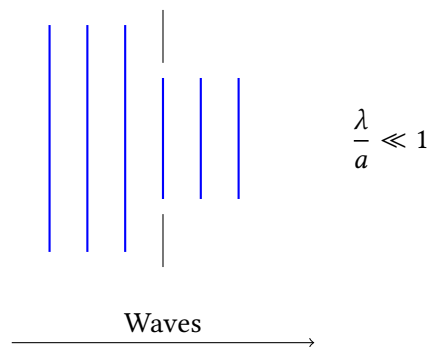
2. Diffraction starts to occur at the edges of the wavefront, the center of which is still mostly parallel, once the wavelength and gap size approach each other more than in case 1:

$$\frac{\lambda}{a} < 1$$
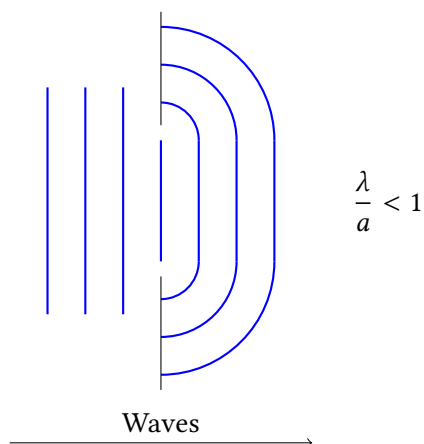
Waves

3. Completely circular or at least very significant diffraction occurs once the wavelength matches the gap size. There is no longer a parallel wavefront:

$$\frac{\lambda}{a} = 1$$
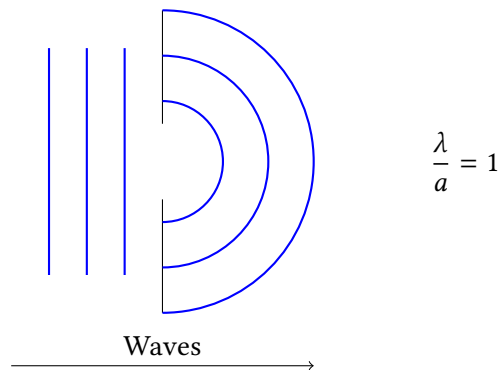
Waves

Double Slit Interference

Waves diffract when passing through a gap or *slit*. This is true for a single slit and is equally true if the number of slits is increased to two. In this case, however, the resultant circular waves passing through the gaps *superpose* and interact with each other, paving the way for another phenomen: *interference*:

Waves
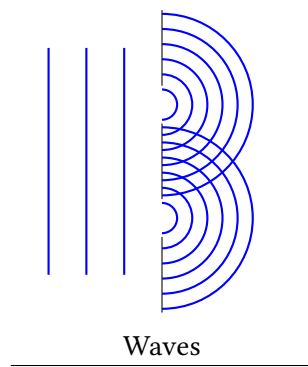
On the one hand, this interference can be *constructive*, where components of the same sign meet such that their intensity is increased, which result in so called *maxima*. For this to happen the path difference τ between the sources of the two (light) waves must be an integer multiple of the their wavelength λ: $\tau_{const} = n \cdot \lambda$. On the other hand, where troughs and peaks meet, *destructive interference* takes place, causing *minima* to be formed where the path difference is an integer multiple of one half the wave length: $\tau_{dest} = (n + \frac{1}{2}) \cdot \lambda$. In the case of light, a minimum would signify darkness, while a maximum results in an increase in brightness and intensity.

It should be noted, however, that the amplitude A at a maximum is double the maximum that would be produced by a single slit and four times the intensity I, given that $I = A^2$. This increase in energy at the maxima balances the lack of energy at the minima and ensures that the energy in the interference pattern is the sum of energies emitted by the two slits. Thus, in total, there are no energy losses, which conforms to the law of energy conservation.

Interestingly, if a screen is set at a certain distance D to the double slit surface with a distance d between the two slits, then, given that this distance D is a lot greater than d, it can be assumed that any two light rays diffracted at either slit, travelling towards a certain point on the opposing screen (at which they superpose) are travelling parallel to one another, as their angles are equally small and thus

relatively equal. This fact can be used to measure the path distance τ between any two given light rays, which ultimately yields a possiblity to measure the light's wavelength. For this, simple trigonometric expressions can be used to, first of all, determine the angle θ between any refracted light ray and the horizontal and then, subsequently, the path difference τ between two light rays. Depending on whether the light rays at the point of measurement on the screen interfere constructively or destructively, the path difference can then be used to calculate the wavelength λ given that $\tau = n \cdot \lambda$ for constructive intereference (at a maximum) and $\tau = (n + \frac{1}{2}) \cdot \lambda$ for destructive interference (at a minimum). The integer n in this case refers to the n-th *order* or *maximum*.

| Double slits at close | τ, θ and d | Double slits at distance + screen |
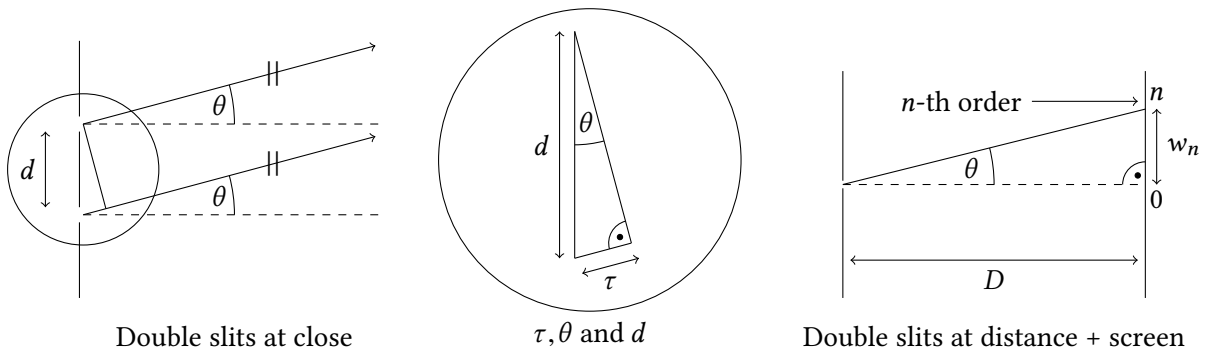
As can be taken from these depictions, the path difference τ is the opposite side in a right triangle in which the distance between the slits d is the hypotenuse and θ the relevant angle. Thus, for the n-th order *maximum* the following equation may be used to calculate τ:

$$\sin \theta = \frac{\tau}{d} = \frac{n \cdot \lambda}{d} \qquad \Rightarrow \qquad n \cdot \lambda = \sin \theta \cdot d$$

And for the n-th order *minimum*:

$$\sin \theta = \frac{\tau}{d} = \frac{(n + \frac{1}{2}) \cdot \lambda}{d}$$

The distance between the double slits d is a known value. The angle θ is calculated from knowledge that θ is also the angle at which (approximately all) the light rays travel towards a certain point. Thus, knowing the distance D between the slits and the screen and from measurement of the distance w_n between the n-th order maximum and the 0-th order (middle of the screen), the inverse tangent of the two can be used to calculate θ:

$$\tan \theta = \frac{w_n}{D}$$

If more than two slits are used, the same principles of interference apply, though with more light rays leading to a different *interference pattern*. In such a case, the surface with many slits through which the light rays pass is referred to as a *diffraction grating*. In case of such a diffraction grating, the distance d between the slits is known as the *grating constant*. Usually, a diffraction grating is defined as having some number of slits n per meter or millimeter, thus d would then be equal to the inverse of of n in the respective unit. There are two main advantages that diffraction gratings with many slits have over gratings with only two (double-slit gratings). The first is that the maxima are much brighter for many slits than for only two, as for every extra slit one more light ray superposes with the others at the maxima. The second is that maxima are much *sharper*, i.e. they are much more well defined with almost no intensity in between invidvidual maxima of order n. The reason for this is that with many slits, only a slight path distance of, say, $\tau = 1.1 \cdot \lambda$, means that at the point at which the light rays coming from the grating meet, one wave will arrive with a path difference of τ, another with 2τ, another with 3τ and so on. In effect, for any light ray arriving with a path difference of $x \cdot lambda$, where x may be any number and not just an integer, there will be another arriving with $(x + \frac{1}{2}) \cdot \lambda$ such that they cancel. The

maxima will therefore not only be brighter because more light rays meet with a path difference of $n \cdot \lambda$ (more constructive intereference), but also sharper because any non-integer deviation from $n \cdot \lambda$ will cause any light ray to be canceled by another with one half more of the wavelength in path difference.

Polarization

Light waves are electromagnetic waves and thus *transverse* waves, meaning that they oscillate and vibrate perpendicular to the direction of the light ray's motion. However, this does not mean that there is only one plane for them to oscillate in space, as they can always be rotated around the axis of motion. These different orientations are referred to as *polarization directions*. If the vibration is parallel to a horizontal plane then the wave is said to be *horizontally plane-polarized*, if it is parallel to a vertical plane such wave is referred to as a *vertically plane-polarized wave*. In general, a transverse wave is *polarized* if it vibrates entirely parallel to a plane in space. In contrast, longitudinal waves cannot be polarized because they have a unique vibration direction parallel to the wave direction. It should is also be mentioned that electromagnetic waves, such as visible light, consist of an electric field $\vec{E}$ and a magnetic field component $\vec{B}$, each vibrating perpendicular to the other and to the direction of the wave. Thus, when discussing polarization, it is convention to define the polarization direction of an electromagnetic wave as the direction in which the electric field varies and not the magnetic field.

However, most conventional sources of light such as any filament lamp or the sun emit *unpolarized* light. This means that rays coming from such sources contain photons of all polarization directions, superposed.

Unpolarized light   Polarized light
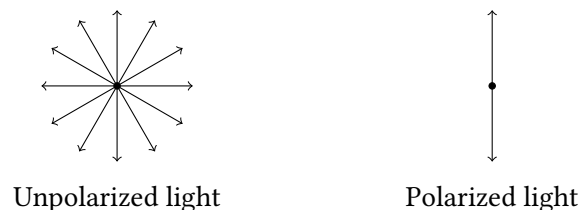
There are two means by which unpolarized light can be polarized, described in the following paragraphs.

Polarization Filters

Unpolarized light may be transmitted through a *polarizing filter* to transform it into plane-polarized light. Polarizing filters are sheets of long-chain, transparent polymer molecules that absorb any light component that is not parallel to its *transmission axis*, transferring it into thermal energy. As a result, the previously unpolarized light exiting the polarization filter is now completely plane-polarized parallel to the transmission axis. The intensity I and amplitude A of the resultant light wave is dependent on its initial amplitude A_0 and intensity I_0 as well as the angle θ between the incident ray and the transmission axis. This yields to the following equation known as *Malus' law*:

$$A = A_0 \cdot \cos \theta \qquad\qquad I = A^2 = (A_0 \cdot \cos \theta)^2 = A_0^2 \cdot \cos^2 \theta \qquad\qquad I = I_0 \cdot \cos^2 \theta$$

As always, the intensity of a light wave is defined as the square of its amplitude. In this case, the cosine of the angle θ is used, as an angle of 0 degrees indicates that the incident light ray already was plane polarized parallel to the transmission axis, such that $\cos 0 = 1$. On the other hand, a light ray that is vertically plane-polarized will be absorbed entirely by a horizontal transmission axis, given that $\cos 90 = 0$. It should be noted here that if entirely unpolarized light is passed through a polarizing filter, the resultant intensity of the polarized light will be one half of the incident intensity. The reason for

this is that, because unpolarized light contains light of all polarization directions, exactly one half of the light will be partially or completely parallel and the other half partially or completely perpendicular to the transmission axis, the first half passing through and latter half being absorbed and transferred into thermal energy.

Given the definition of Malus' law it can be concluded that if two polarizing filters are placed in one line with unpolarized light, in such a way that the transmission axis of the first polarization filter is perpendicular to that of the second, no light will pass through the second filter. The reason why is that light will pass through the first filter completely plane-polarized parallel to the transmission axis of the first filter with an intensity I_1 of one half the incident intensity I_0. Then, it will hit the second polarizing filter at an angle θ of 90 degrees, where all light will consequently be absorbed, yielding an intensity I_2 of 0:

$$I_1 = \frac{I_0}{2} \qquad \Rightarrow \qquad I_2 = I_1 \cdot \cos^2 \theta = \frac{I_0}{2} \cdot \cos^2 90 = 0$$

On the other hand, if a third filter were to be place between the two just mentioned, the light would have to leave the final filter with an intensity greater zero (unless the second filter is also perpendicular to the first). Leaving the first filter, the light would again be reduced to one half of its initial intensity and be now plane-polarized parallel to the transmission axis of the first filter. Then, upon passing through the second filter, the component of the light that is parallel to the second transmission axis would be selected according to the cosine of the angle θ between the now polarized light and the second filter. Given that this second filter is also not perpendicular to the third, some of the light exiting the second filter would again be let through. This is true for all angles between the transmission axes of the first and second filter (except 90°) and reaches its maximum at 45°.

Reflection from Transparent Surfaces

The second possibility to plane-polarize unpolarized light is to reflect it off a transparent surface such a water. When a light ray hits such a transparent surface, the reflected ray is partially plane-polarized parallel to the surface and the refracted ray is partially plane-polarized perpendicular to the surface. At a certain angle of incidence, called the *Brewster angle* θ_b, the refracted and reflected components are perpendicular to each other and entirely plane-polarized. The tangent of this angle is equal to the refractive index $n_{1,2}$ of the incident medium and the transparent medium (the first medium is usually air, such that $n_{1,2}$ is simply equal to the absolute refractive index n_{med} of the medium):

$$\tan \theta_b = n_{1,2}$$
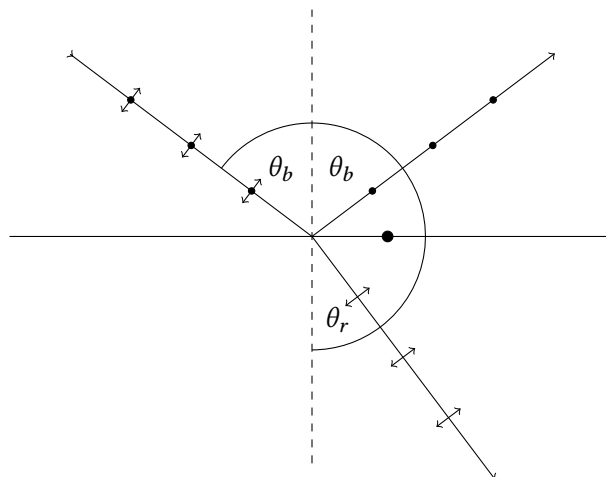
This equation can be derived easily. As is known from Snell's law, the ratio between the sines of the angle of incidence, here θ_b, and the angle of refraction θ_r is equal to the refractive index of the two media:

$$\frac{\sin \theta_b}{\sin \theta_r} = n_{1,2}$$

Moreover, given the definition of the Brewster angle, the angle of the reflected ray (which is equal to the angle of incidence) and the angle of the refracted ray are perpendicular to each other. As the sum of a half-circle is 180°, this means that the remaining 90° are the sum of θ_b and θ_r. This can be reformed:

$$\theta_b + \theta_r = 90° \qquad \Rightarrow \qquad \theta_r = 90° - \theta_b$$

Given that $\sin 90 - \theta_b = \cos \theta_b$ and considering that the sine of any angle divided by the cosine of that angle is defined as its tangent, it follows that the tangent of θ_b is equal to $n_{1,2}$:

$$\frac{\sin \theta_b}{\cos \theta_b} = n_{1,2} \qquad \Rightarrow \qquad \tan \theta_b = n_{1,2}$$

It may now be clear why it is common for sunglasses to be vertically plane-polarized. When looking at a transparent surface such as a lake or the ocean, reflected sun light is either completely (when $\theta_i = \theta_b$) or partially plane-polarized parallel to the surface. Thus, the vertical polarization of the sunglasses ensuret that only a minimum of sun light may pass through them.

Liquid Crystal Displays

Another useful application of polarization are Liquid Crystal Displays (LCDs). LCD screens are made up of many back-lit *pixels*, e.g. 1 024 000 in a typical 1280 by 800 display. To display images in color rather than solely black-and-white, each individual pixel consists of three separate *cells*, each with a different *color filter* to display either red, green or blue at one of 256 levels of brightness. Thus, by mixing and varying the brightness of each of the three color cells, an LCD screen may display any of over 16 million colors and shades. The setup of a single cell as well as the process of varying the brightness by means of polarization, liquid crystals and a variable electric field is outlined and explained in the following paragraphs.

The first component of any liquid crystal cell is a *diffuser*. Light is supplied to individual cells from a central source at the back of the display. The diffuser is a thin sheet of a certain transparent material that spreads out the light coming from the central source equally across all pixels, to ensure that the light reaching each individual cell is uniform.

The basic principle of an LCD cell is then to polarize light rays in one plane, either horizontally or vertically, by passing them through a first polarizing filter spaced at a certain distance to another polarizer, whose transmission axis is perpendicular to that of the first. The intensity and thus brightness of a cell is therefore controlled by rotating the light rays between the first and second polarizer by a particular angle, causing a certain portion of the light to pass through and another portion to be absorbed by the long-chain, transparent polymers of the second polarizer.

This rotation is achieved by means of *liquid crystals* — liquids with crystalline properties. A certain variant of these liquid crystals, called *twisted nematic liquid crystals*, are naturally twisted to a helix shape. These liquid crystals guide light rays through their helix shape to achieve a rotation by a certain angle. If spaced appropriately, this angle is exactly 90°, such that the linearly polarized light exiting the first polarizer of the liquid crystal cell is rotated enough to pass through the second polarizer whose axis of transmission is perpendicular to the first. It should be noted that an incorrect layout or spacing of the liquid crystals may lead to an angle of rotation greater or lower than 90°, which would be equally problematic.

An interesting and essential property of liquid crystals is that if an electric field is applied across them, they untwist. For this, two transparent electrodes of a capacitor, one negative and one positive, are placed at either side of the liquid crystals. The capacitor stores a certain charge, which is supplied periodically by an external system at a frequency referred to as the *refresh rate* of the screen (nowadays usually 60 Hertz). After every interval, a *thin film transistor (TFT)* switches the capacitor into another circuit to let it discharge. Consequently, a new charge is supplied appopriate for the brightness one wishes to set for the cell. The degree of untwisting of the liquid crystals depends on the potential difference between the two electrodes of the capacitor (which, in turn, depends on the charge supplied to it). If no voltage is applied across the two electrodes, the liquid crystals stay twisted in their helix form and rotate the light rays appropriately to be able to exit the second polarizer with full intensity. If a maximum of voltage is applied, the liquid crystals fully untwist, causing no change to the polarization direction of the light and resulting in them being fully absorbed by the perpendicular transmission axis of the second polarizer. Any voltage in between the two extremes results in an appropriate brightness or shade.

Before the light rays pass through the second polarizer, they pass through a color filter to ensure that only the red, green or blue component exits the cell to be mixed with the other two colors of the pixel.

Lasers

The word "laser" is an acronym for *Light Amplification by Stimulated Emission of Radiation*. Laser light finds many practical applications in everyday life and technology, from basic toy laser-pointers to bar-code product identification systems in supermarkets to CD and DVD drives in computers, audio systems or DVD players. The properties of laser light are explained, examined and discussed in the following paragraphs.

Properties of Laser Light

There are four main properties and terms associated with laser light that set it apart from conventional forms of light from filament lamps or the sun. Laser light is . . .

- **Monochromatic**: laser light only has one *chroma*, i.e. a single color. The color of light is related to its frequency f and wavelength λ, thus this property essentially means that all photons in laser light have the same frequency and wavelength. Popular laser colors are red or green.

- **Coherent**: all photons of laser light ray have the same phase — they are *in phase* or *in step*. This is due to the fact that laser photons are not emitted spontaneously as is the case for ordinary, conventional sources of light such as filament lamps, where atoms emit light spontaneously and independently from one another. Rather, laser photons are emitted by explicit stimulation, one after another in a chain-reaction. The coherence adds to the amplitfication of the light due to constructive intereference.

- **Collimated**: the light emitted is highly focused and highly directional and is usually only a relatively narrow beam, with little divergence. By comparison, ordinary light from the sun, a candle or a light bulb is not at all *collimated* or focused, but radiates away in many or all directions. This degree of high collimation results from the setup of the cavity of a laser. The material from which light photons are emitted — often a synthetic ruby crystal — is placed between two parallel mirrors. One of the two mirrors — the back mirror — is entirely reflective, while the other — the front mirror — is only around 99 percent reflective, letting through 1 percent as the laser light beam and reflecting the rest. Only light rays perpendicular to the mirrors (parallel to the path between them) are let through, thus they are all in the same direction — they are *collimated*.

- **Intense**: the light is high and approximately uniform in energy density throughout.

Terminology of Laser Light

It should be explained what the difference between stimulated and spontaneous emission is, as it is this distinction that makes laser light unique compared to ordinary, conventional sources of light. Furthermore, there are a few other terms that carry an important meaning in association with laser light.

- **Spontaneous Emission**

  Spontaneous emission is the phenomenon whereby an excited electron in a higher energy level or quantum state of an atom transitions to a lower state, possibly the ground state of the atom, at a *random* point in time and thereby *spontaneously* emits a light photon whose frequency f depends on the discrete difference in energy ΔE between the energy of the electron in at initial quantum state and its final energy level, according to the Planck relation which states that the energy E of a photon is equal to $h \cdot f$, where h is Planck's constant and f the photon's frequency. Spontaneous emission is the standard form of emission and produces light that is incoherent (out of phase), polychromatic (many wavelengths, frequencies and colors) and divergent (not collimated).

- **Stimulated Emission**

  Stimulated emission is the process whereby an incoming photon, whose energy is equal to the energy difference ΔE of an electron's current quantum state and a lower energy level, can *stimulate* or *trigger* the transition of the electron from its initial, higher energy state to the final, lower quantum level. As a result, the electron emits an exact, identical copy of the incoming photon such that there are now two photons with the same phase, frequency and direction of travel. Under natural circumstances, stimulated emission is a very rare phenomenon. The reason why is that when an electron absorbs a discrete quantum of energy carried by a photon to reach a higher quantum state (energy level), it remains at this excited state for only a short period of time, around 10^{-8} seconds. In this very small time interval, it is highly improbable that a photon carrying just the right, necessary energy will pass the electron to stimulate its transition to a lower quantum state. Population inversion via optical pumping is the method by which laser light increases the chance of stimulated emission. Light produced via stimulated emission is monochromatic (single color and frequency), coherent (equal phase) and collimated (directional and focused) — perfect for lasers.

- **Bottom Heavy**

  A material is referred to as being *bottom heavy* when there is a greater number of electrons at the ground state or generally at lower quantum levels compared to the number of electrons in higher excited states. This is the usual case for the majority of solids and materials, as for there to be electrons in higher states an additional input of energy in form of photons is required, which may then be absorbed by electrons to perform a quantum leap to transition to a higher excited state.

- **Population Inversion**

  Materials are usually bottom heavy, but under certain circumstances there can be more electrons in higher energy states than there are in lower levels, which is atypical and must therefore be stimulated. This situation is referred to as *population inversion*. The achievement of a significant population inversion in atomic energy states is a precondition for the generation of laser

light. For lasers, this population inversion is achieved via *optical pumping*, such that a significant number of electrons is found in the *metastable states* of the material.

- **Metastable States**

  Certain elements such as neon or helium or synthetic ruby happen to have certain electron energy states that are longer-lived than ordinary quantum levels. These metastable states can be seen as temporary energy traps. They stil provide less stability than the ground state of an atom, however the time spent by an electron that absorbs the right quantum of energy to reach such a metastable state is still greater than the time spent in an ordinary energy level by a factor 10^4. All laser materials must have metastable states.

- **Optical Pumping**

  The process whereby a light (optical) source is used to "pump" or elevate electrons from lower energy levels to higher quantum states. The light source emits photons carrying a quantum of energy that is just right for electrons to absorb them and to reach a *metastable state*. Via optical pumping, population inversion is achieved. At some random point in time, one of the optically pumped electrons will undergo spontaneous emission and thereby emit a photon whose energy is just enough to de-excite neighbouring electrons that are also in the metastable state (which are many, because of population inversion). This causes a chain-reaction of stimulated emission of identical light photons with the same phase and frequency, leading to their superposition and thus *light amplification.*

- **Amplifying Medium**

  The amplifying medium refers to the material in which light amplification by stimulated emission of photons takes place. Often, this material may be synthetic ruby, a mixture of helium and neon gases or various possible semiconductors. The essential property of the amplifying medium is that it allows its atoms to have metastable states, which can trap electrons for longer durations and thus make population in version possible.

It should quickly be summarized how the above concepts and phenomena lead to an first amplification of coherent and monochromatic light. In the first step, *optical pumping* is used to elevate electrons within the *amplifying medium* to *metastable states*, where they are trapped for durations longer than normal. This makes *population inversion* possible, which is the situation where there are more electrons in higher quantum states than in lower ones, which is atypical as materials are normally *bottom heavy*. When many electrons are in the metastable states, one will spontaneously transition to a lower energy level and thereby emit a photon. Because this photon carries energy that is just the right, discrete quantum of energy required for the other excited electrons in the metastable state to de-excite, this single spontaneously-emitted photon from the first electron causes a chain reaction of stimulated emissions of radiation from other electrons. The light photons emitted in this process are identical copies of one another and are equal in phase (light is coherent) and frequency (light is monochromatic). This is the basic working principle of stimulated emission within the amplifying medium. The light produced is then further amplified by the *laser cavity*, whose setup is explained below.

Setup of a Laser Cavity

After initial generation of coherent and monochromatic light via stimulated emission within the amplifying medium, the light is further amplified in the *laser cavity*. For this, the amplifying medium is placed between two parallel mirrors, one of which is completely ($\sim 100\%$) reflective — the back mirror — while the other is partially ($\sim 99\%$) reflective — the front mirror. The purpose of these mirrors is to bounce photons back and forth within the laser cavity to achieve more stimulated emission of light photons. About one percent of the light leaves the front, partially reflective mirror as the output of

the laser — the light you see. There are two further advantages of the laser cavity setup with the two mirrors:

1. The mirrors are setup parallel to one one another, at either end of the laser cavity. There is no reflective medium above or below the cavity, thus only light that is perpendicular to the mirrors is reflected, while all other directions of light that would not improve the collimation of the light but rather increase its divergence are simply absorbed by whatever is above or below the cavity.

2. An optical standing wave is created with two nodes at either mirror. This leads to a filtering and selection of the frequency as the light photons emitted by the amplifying medium cover a small range of frequencies — the photons are not perfectly monochromatic. The standing narrow this range to improve the degree of how monochromatic the light is.

Applications of Laser Light

Two real-world applications of laser-light are examined further in this section.

Bar Codes

Bar codes are used in supermarkets, libraries or or warehouses to identify object, products or other items. They are a very simple application of laser light. For a bar code, the thickness of each of the black lines codes for a digit between 0 and 9. A laser scans over the bar code, while a suitable photodetector measures how much of the light emitted from the scanner, hitting each individual line, is reflected back to the scanner. The thicker the line, the more it will absorb the light and the less it will reflect. This determines the digit. The white space between the digits does not absorb light as much.

A Bar Code

Optical Drives

Optical drives are the components of DVD players or computers that read information and data stored on compact discs (CDs) or digital versatile discs (DVDs). They utilize lasers in a most interesting way. Collimated, coherent and monochromatic light from a laser is emitted onto the surface of a CD (everything will also apply to a DVD) that is composed of many small pits spaced apart at a certain distance, but all one quarter of the wavelength of the laser light *deep* (going into the surface layer of the CD by one fourth of the wavelength). When light passes over a plateau, i.e. a region on the CD that is not a pit (always between two pits), or right into a pit, the light will be reflected with full intensity (also constructive interference with incoming light) and this will be detected by a photodetector in the reading device and interpreted as a HIGH signal (binary 1). However, when the light falls right onto

the edge of a pit and plateau, the light travelling down the pit will be reflected and interfere with the light reflected at the plateau 180° later in its cycle. More precisely, the light travels one fourth of a wavelength down the pit and another fourth on the way back to meet up with the plateau light at a phase difference of 180°. This results in destructive intereference and a loss in intensity which is once more detected by the photodetector and this time interpreted as a LOW signal (binary 0). The binary sequence of the data stored on the CD is then converted into analog via a suitable digital-to-analog converter (DAC).