

Notes on Geometric Optics and Interference effects of Electromagnetic Radiation

We just finished discussing Maxwell's Equations and the free space solutions of electromagnetic radiation. Now we examine some practical applications for electromagnetic radiation in the visible region. We start with geometric optics, which is followed by interference effects.

Ray Tracing

Electromagnetic radiation can reflect off surfaces, transmit through materials, interact with charge, etc. Applying Maxwell's equations directly would involve solving for $\vec{E}(\vec{r}, t)$ and $\vec{B}(\vec{r}, t)$ for the appropriate boundary conditions of the situation. Carrying out such calculations can be difficult, however for many situations one can follow the "rays" of the radiation. A ray is a line that is perpendicular to the plane of the electric and magnetic fields of the radiation. It is a line that is in the direction of Poynting's vector, \vec{S} . To determine what the radiation does, we will follow the rays as they reflect off of, or travel through materials.

How accurate is ray tracing? We have all seen a laser beam, or seen a thin light beam after it has passed through a small hole in piece of paper. It is as if the light stays as a small thin line as it travels. So one can imagine breaking up light into thin lines, and following what each line does as it travels. One has to be careful with this picture. As light travels, it's cross sectional area cannot remain constant. We will discuss this more later under the topic of diffraction. I mention one point now. The amount of spreading, that is how much the cross section area increases, depends on the ratio:

$$\text{spreading angle} \sim \frac{\lambda}{d} \quad (1)$$

where λ is the wavelength of the radiation, and d is the diameter of the beam. The wavelengths for visible light range from $400 - 700 \text{ nm}$. So if a beam of light is 1 mm in diameter, then $(7 \times 10^{-7}) / (10^{-3}) \approx 7 \times 10^{-4}$ and one not notice the spreading unless the beam travels hundreds of meters. So, for visible light, it is a good approximation to break up the radiation into thin rays and follow each ray as it travels through the environment as long as it doesn't travel hundreds of meters. Next we will take a "ray tracing" approach in analyzing how light reflects off surfaces and transmits through different media.

Geometric Optics and Mirrors

If a beam of light strikes a flat mirror, or smooth reflecting surface, the angle that the light makes when it is reflected from the mirror equals the angle that it struck the mirror. We know this from experience, and we will demonstrate in lecture that the incident and reflected angles are equal. When we discuss refraction, we will understand why this is true in general for waves reflecting off of a flat surface.

What if the mirror is not flat, but is curved. In this case, the situation is not difficult if the beam is very narrow. That is, if the diameter of the beam is much smaller than the radius of the curved surface where it strikes the mirror. If the beam is very narrow, then the mirror is essentially flat where the beam hits it. So, as with the completely flat mirror, the incident and reflected angles will be equal. In order to handle curved surfaces, the best angles to use are the angles with respect to a line that is perpendicular to the surface that the beam hits.

In the figure we show how the incident and reflected beam, or ray, is measured. Go to where the beam, or ray, strikes the surface. Then draw a line perpendicular (or normal) to the surface. The incident angle, θ_i , is measured from the incoming ray to the normal. The reflected angle, θ_r , is measured from the outgoing ray to the same normal. For reflection:

$$\theta_i = \theta_r \tag{2}$$

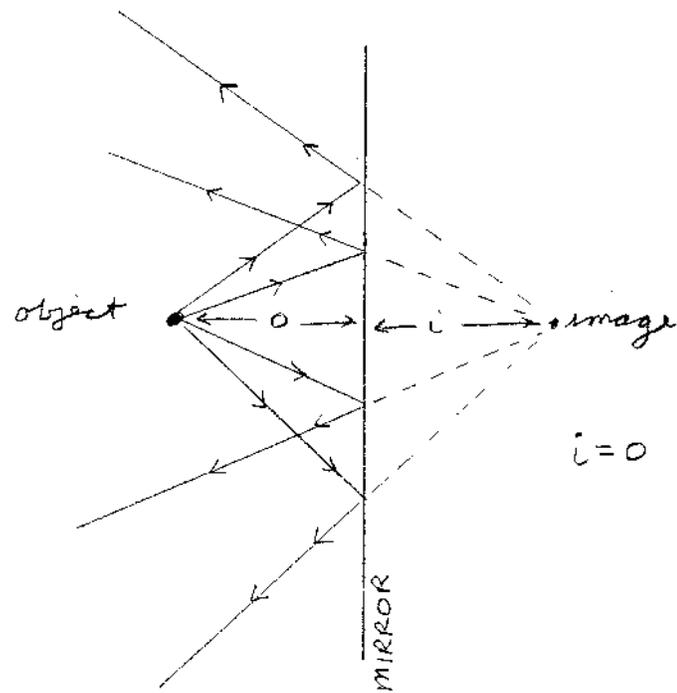
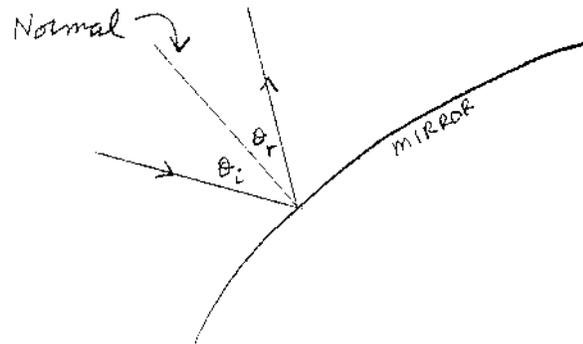
This equation is true even if the surface is curved.

Flat Mirror

Consider an object that is in front of a flat mirror. Let the distance from the object to the mirror be o . The object emits light in all directions, which we show by rays. Some of the light will reflect off the mirror. Let's follow the rays that reflect off the mirror. Each ray will reflect such that $\theta_r = \theta_i$ as shown in the figure.

We can extend the reflected ray back into the mirror, as is drawn in the figure as dashed lines. The dashed lines intersect at a point. *This point is where one would see the object if one looked in the mirror.* That is, if you looked in the mirror, you would see the object at that position no matter where your eyes are. This position is where the image is located. We will label the distance from the mirror to the image as i . For the flat mirror, you can see that

$$i = o \tag{3}$$



by using geometry and $\theta_r = \theta_i$. An image that is seen by looking into a mirror (or lens) is called a *virtual image*. The image cannot be projected onto a screen, but rather it can only be seen by looking into the mirror.

Curved Mirrors

We first consider a special kind of mirror, one that has axial symmetry and is curved in such a way as to focus light to a point. The mirror has the following property: *rays parallel to its axis focus at a point*, the focal point. The distance from the mirror to the *focal point* is called the *focal length* of the mirror. We label this distance f . To produce the focusing of rays at a point, the mathematical shape of the mirror needs to be a parabola (rotated about its axis).

Note that if we put a light source exactly at the focal point, then the light that reflects from the mirror will emerge parallel to the axis. That is, by putting a light bulb at the focal point, the reflected light will be a nice beam, like a flashlight.

Now, consider an object that is placed between the focal point and the mirror. Let the object have a height h_0 and be located a distance o ($o < f$) from the mirror. The object is represented by an arrow in the figure. We can trace three rays from the tip of the arrow that reflect off the mirror:

A ray parallel to the axis (2 → 3)

The ray that is parallel to the axis will reflect off the mirror and pass through the focal point. The dashed line extends the reflected ray back behind the mirror.

A ray directed away from the focal point (2 → 8)

The ray that is directed away from the focal point will reflect parallel to the axis. The dashed line extends the reflected ray back behind the mirror.

A ray that strikes the center of the mirror (2 → 6)

The ray that strikes the center of the mirror will reflect back at the same angle that it strikes the mirror. This part of the mirror is perpendicular to the axis of the mirror. The dashed line extends the reflected ray back behind the mirror.

All three of the dashed lines intersect at a point. Although we only considered these three rays, the dashed lines for all rays from the tip would intersect there as well. Wherever you look into the mirror, you will see the tip of the arrow at this intersection point. The intersection point is the image of the arrows tip. Since the

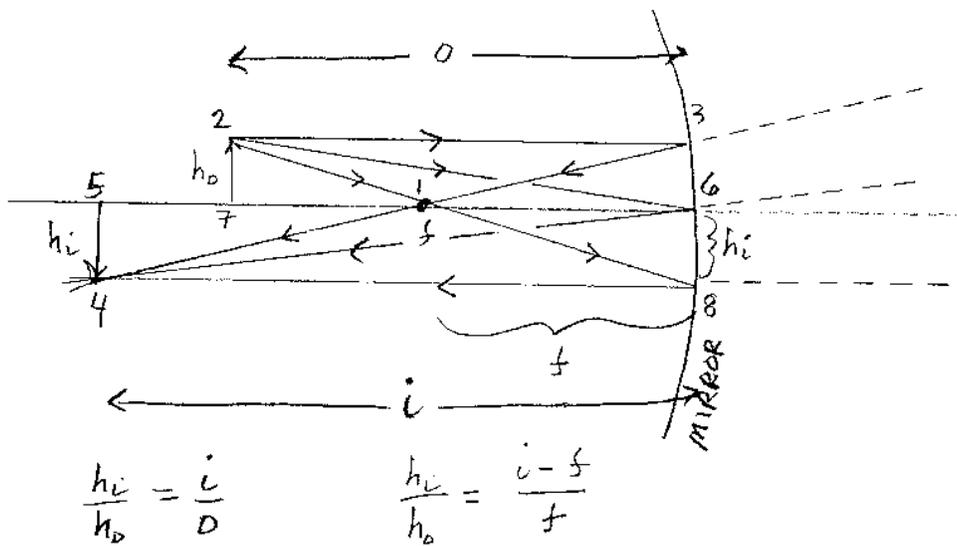
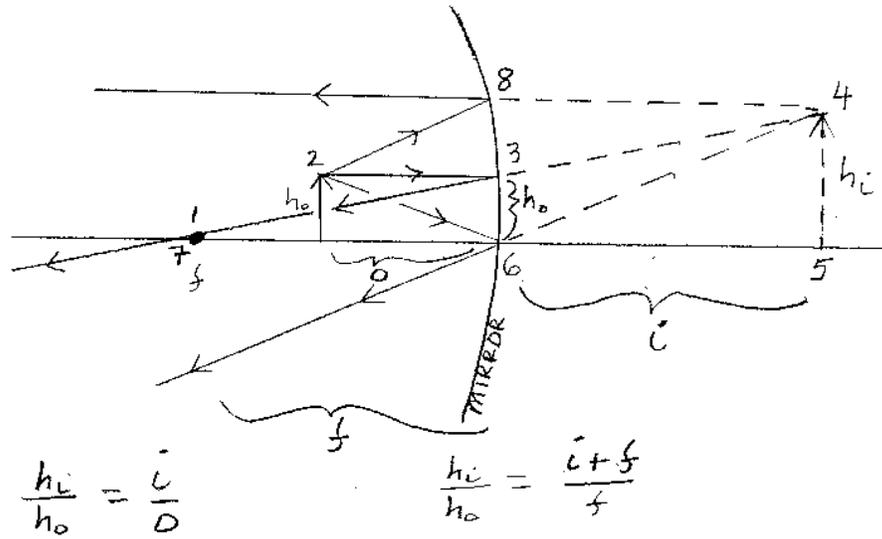


image is "in the mirror", it is a virtual image. We label i as the distance from the image to the mirror, h_i as the height of the image, and h_o has the height of the object.

Using geometry, we can determine the relationship between o , i , f , h_i and h_o .

Triangle 1 – 4 – 5 and triangle 1 – 3 – 6 are similar triangles:

$$\frac{h_i}{h_o} = \frac{f + i}{f} \quad (4)$$

Triangle 2 – 6 – 7 and triangle 4 – 5 – 6 are similar triangles:

$$\frac{h_i}{h_o} = \frac{i}{o} \quad (5)$$

Equating the h_i/h_o we have

$$\begin{aligned} \frac{i}{o} &= \frac{i + f}{f} \\ \frac{1}{o} - \frac{1}{i} &= \frac{1}{f} \end{aligned}$$

The ratio of h_i/h_o is how much the image increased in size, and is called the *magnification* m . For the case just treated, we see that $m = h_i/h_o = i/o$. Since for this case $i > o$, the magnification is greater than 1.

Now, lets use the same mirror, but place the object such that $o > f$. We can follow the same rays from the tip of the arrow as before:

A ray parallel to the axis (2 → 3)

The ray that is parallel to the axis will reflect off the mirror and pass through the focal point. The dashed line extends the reflected ray back behind the mirror.

A ray directed through the focal point (2 → 8)

Since the $o > f$ a ray from the tip can pass through the focal point. The ray that passes through the focal point will reflect parallel to the axis. The dashed line extends the reflected ray back behind the mirror.

A ray that strikes the center of the mirror (2 → 6)

The ray that strikes the center of the mirror will reflect back at the same angle that it strikes the mirror. This part of the mirror is perpendicular to the axis of the mirror.

The dashed line extends the reflected ray back behind the mirror.

From the figure we can see that the dashed lines do not converge. One will not see an image of the arrow in the mirror. However, in this case, the reflected rays do converge. The rays from point "2" do converge at point "4". However, now point "4" is in front of the mirror. Although we just followed three rays, all rays from point "2" that hit the mirror will converge at point "4". If we put a piece of paper, or screen, where point "4" is, we would see an image of the arrow on the screen. An image that can be projected on a screen is called a *real image*. If $o > f$ the image is real.

Once again, we can use geometry to find the relationship between o , i , f , h_0 and h_i :

Triangle 1 – 4 – 5 and triangle 1 – 3 – 6 are similar triangles:

$$\frac{h_i}{h_0} = \frac{i - f}{f} \quad (6)$$

Triangle 2 – 6 – 7 and triangle 4 – 5 – 6 are similar triangles:

$$\frac{h_i}{h_0} = \frac{i}{o} \quad (7)$$

Once again, equating the h_i/h_o we have

$$\begin{aligned} \frac{i}{o} &= \frac{i - f}{f} \\ \frac{1}{o} + \frac{1}{i} &= \frac{1}{f} \end{aligned}$$

As before, the magnification m equals $h_i/h_o = i/o$. For the real image, the magnification can be less than one ($o > 2f$) or greater than one ($f < o < 2f$).

Note that the equation for $o < f$ is similar to the equation for $o > f$. The only difference is the minus sign in front of i . We can combine both equations into one, which will be valid for any value for o if we assign a minus sign to a virtual image. That is, $i > 0$ means the image is in front of the mirror, and $i < 0$ means the image is "behind" or "in" the mirror. The universal equation for all values of o , i , and f is

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f} \quad (8)$$

where $i > 0$ for a real image in front of the mirror, and $i < 0$ for a virtual image behind the mirror. This equation also works for a flat mirror. For a flat mirror, $f = \infty$, and we have $i = -o$.

Non-Focusing (diverging) curved mirror

Now we consider a curved mirror that is curved in the opposite way as the converging mirror we just analyzed. The property that the diverging mirror will have is *rays parallel to the axis diverge from a point behind the mirror*. As before, the "divergence" point is called the focal point. Although parallel rays do not "focus" at this point, they diverge from the point, it is still useful to use the same name. The distance from this focal point to the mirror is the focal length, " f ". Note that if rays are directed at the focal point, after they reflect from the mirror they will travel parallel to the axis. This property is the same as the converging mirror, but now the physical rays cannot go through the focal point.

Lets examine that the image will be for an object of height h_o placed a distance o from the mirror. We will follow the same three rays from the tip of an arrow as we did with the converging mirror, and label the diagram the same way:

A ray parallel to the axis (2 → 3)

The ray that is parallel to the axis will reflect off the mirror and travel away from the focal point. The dashed line extends the reflected ray back behind the mirror.

A ray directed at the focal point (2 → 8)

A ray from the tip of the arrow directed towards the focal point will reflect parallel to the axis. The dashed line extends the reflected ray back behind the mirror.

A ray that strikes the center of the mirror (2 → 6)

The ray that strikes the center of the mirror will reflect back at the same angle that it strikes the mirror. This part of the mirror is perpendicular to the axis of the mirror. The dashed line extends the reflected ray back behind the mirror.

All three of the dashed lines intersect at a point. Although we only considered these three rays, the dashed lines for all rays from the tip would intersect there as well. Wherever you look into the mirror, you will see the tip of the arrow at this intersection point. The intersection point is the image of the arrows tip. Since the image is "in the mirror", it is a virtual image. As before, we label i as the distance from the image to the mirror, and h_i as the height of the image.

Using geometry, we can determine the relationship between o , i , f , h_i and h_o .

Triangle 1 – 4 – 5 and triangle 1 – 3 – 6 are similar triangles:

$$\frac{h_i}{h_o} = \frac{f - i}{f} \quad (9)$$

Triangle 2 – 6 – 7 and triangle 4 – 5 – 6 are similar triangles:

$$\frac{h_i}{h_o} = \frac{i}{o} \quad (10)$$

Equating the h_i/h_o we have

$$\begin{aligned} \frac{i}{o} &= \frac{f - i}{f} \\ \frac{1}{o} - \frac{1}{i} &= -\frac{1}{f} \end{aligned}$$

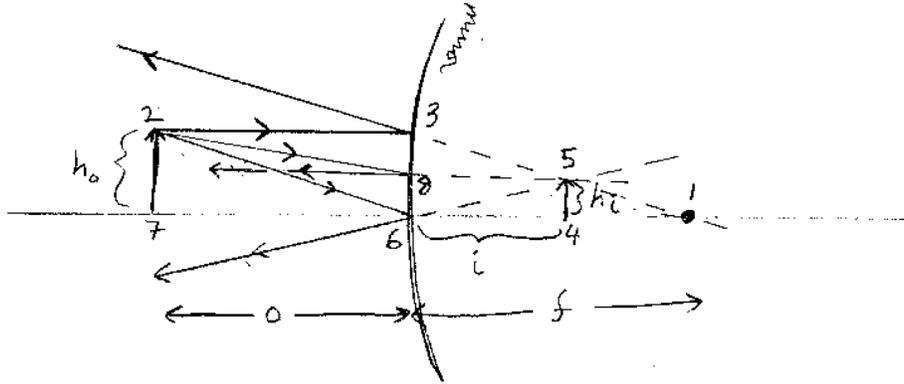
Wow, this equation is very similar to the equation for the converging mirror. If we assign a negative value to the focal length of a diverging mirror, then we can use *one equation* for both converging and diverging mirrors, and the cases which have real or virtual images. Thus, we can produce a *Universal Equation for converging and diverging mirrors*:

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f} \quad (11)$$

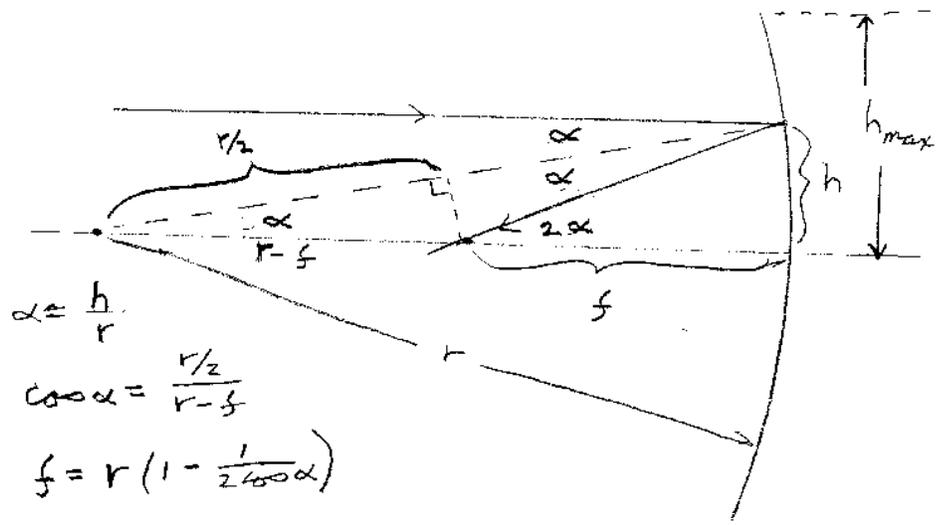
where o and i are greater than zero for real objects and images, and less than zero for virtual objects and images. The focal length f is positive for converging mirrors and negative for diverging mirrors.

There are a few final points that need mentioning:

1. The magnification is defined as $m = h_i/h_o$, which is equal to i/o . Nowever, we are defining the signs of o and i to represent real or virtual quantities. When a real object produces a real image, the image is inverted. So we are can define the magnification as $m = -o/i$, to include the inverted (non-inverted) nature of the image. A negative m means the object and image have opposite orientations. A positive m means that the object and image have the same orientation.



$$\frac{h_i}{h_o} = \frac{i}{o} \quad \frac{h_i}{h_o} = \frac{f-i}{f}$$



$$\alpha \approx \frac{h}{r}$$

$$\cos \alpha = \frac{r/2}{r-f}$$

$$f = r \left(1 - \frac{1}{2 \cos \alpha} \right)$$

- It is much easier to make a spherical mirror than a parabolic mirror. If the light only reflects near the center of a spherical mirror then the rays will focus at a point fairly well. The focusing will not be perfect, as with a parabolic surface. If the radius of curvature of the spherical surface is r , then the focal length will be $f \approx r/2$. This accuracy of this approximation can be determined as follows. From the figure we can see that

$$\frac{r/2}{r-f} = \cos(\alpha) \quad (12)$$

We can solve this equation for f to obtain

$$f = r\left(1 - \frac{1}{2\cos(\alpha)}\right) \quad (13)$$

So far this equation is exact. Now, if $\alpha \ll 1$ we can use the Taylor expansion for $1/\cos(\alpha) = \sec(\alpha)$:

$$\begin{aligned} f &= r\left(1 - \frac{1}{2}\left(1 + \frac{\alpha^2}{2} - \dots\right)\right) \\ &= \frac{r}{2}\left(1 - \frac{\alpha^2}{2} + \dots\right) \\ &\approx \frac{r}{2}\left(1 - \frac{1}{2}\left(\frac{h}{r}\right)^2 + \dots\right) \end{aligned}$$

where we have substituted $\alpha \approx h/r$. So the first correction to $f = r/2$ is of order $(1/2)(h/r)^2$. So parallel rays will focus on the axis of the mirror in a region that is a distance of $r/2$ to $(r/2 - (1/2)(h_{max}/r)^2)$ from the mirror. Thus, the "bluryness" of the focal point will be approximately $(1/2)(h_{max}/r)^2$. If $h_{max} \ll r$ then the focusing is pretty good at $f = r/2$.

- A converging mirror is also called a concave mirror. A diverging mirror is also called a convex mirror. The diverging mirror only has virtual images. Real images (from a real object) can only be produced with a converging mirror, and only when $o > f$.
- Our derivation of the relationship between o , i , f , h_o , h_i , and m was based on geometry. Hence, the analysis is termed "Geometric Optics".

We now consider images formed by lenses. First we need to understand what happens when light passes from one material (medium) to another. We start with Snell's law, then derive the relationships for thin lenses.

Snell's Law

Snell's law deals with how light travels from one material (or medium) to another. Light travels at the speed $c = 3 \times 10^8 \text{ m/s}$ in vacuum. In a transparent material it travels slower. If light travels at the speed v in a medium, we define the *index of refraction* n as $v = c/n$. In vacuum, $n = 1$, and for transparent materials $n > 1$ since light travels slower in materials than in vacuum. We should mention that the index of refraction n will depend on the material and in general on the frequency of the radiation.

Suppose there are two materials (media) which we label as "1" and "2", and that the interface where they are in contact is a flat plane. Let the speed of waves in the respective media be v_1 and v_2 . From experiment, we observe that light "bends" as it travels from one material into another. This phenomena is called refraction. We label the angle that the incident wave makes with the interface as θ_i , and the angle that the refracted wave makes as θ_R . *Both angles are measured from the direction of the wave with respect to a line perpendicular (normal) to the interface.* Both angles are always positive and between 0° and 90° . We want to come up with a relationship between θ_i and θ_R .

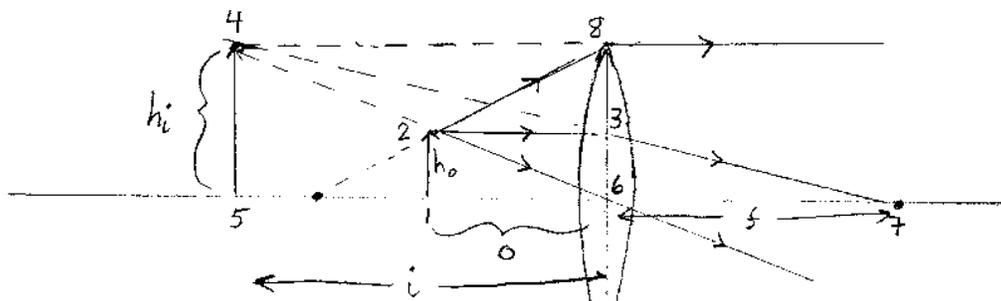
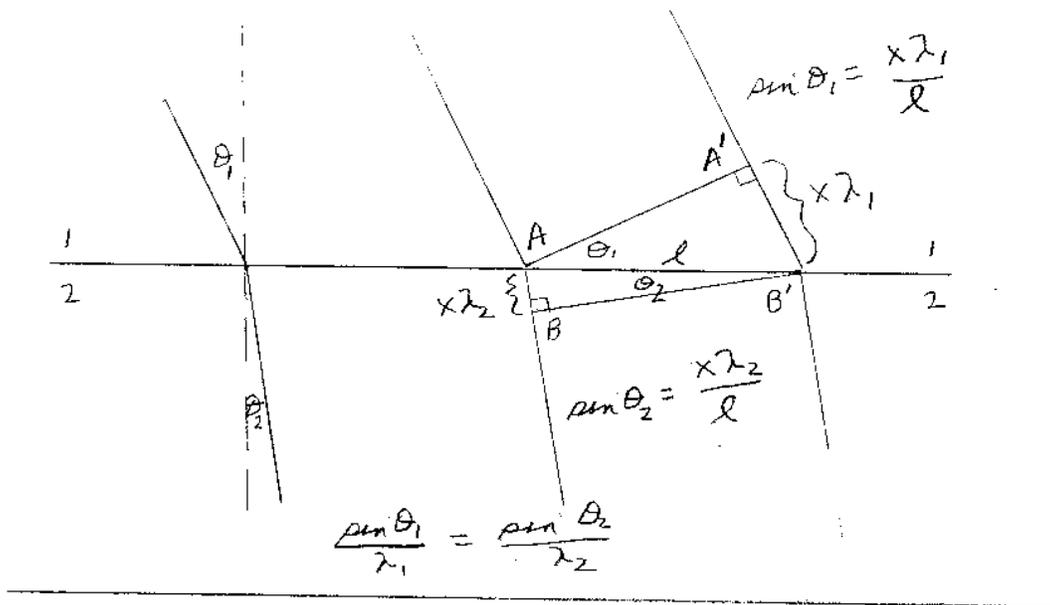
Suppose a plane wave travels from "1" into "2" as shown in the figure. The figure is a snapshot of the wave as it travels from "1" into "2". The lines in the figure are points which have the same phase in the snapshot. If you want, one of the lines can be the crest of the wave when the snapshot is taken. The points A and A' have the same phase. The points B and B' also have the same phase, although different than A and A' . Let $x \equiv (\text{distance } A \rightarrow B)/\lambda_2$. With this assignment, the distance from A to B equals $x\lambda_2$. Since A and A' have the same phase, and B and B' have the same phase, the distance from A' to B' must be $x\lambda_1$. Let l be the distance on the interface between A and B' . From the figure we see that

$$\sin(\theta_1) = \frac{x\lambda_1}{l} \tag{14}$$

and

$$\sin(\theta_2) = \frac{x\lambda_2}{l} \tag{15}$$

equating x/l in each equation we have



$$\frac{h_i}{h_o} = \frac{i}{o} \quad \frac{h_i}{h_o} = \frac{f+i}{f}$$

$$\frac{i}{o} = \frac{f+i}{f} \quad \Rightarrow \quad \frac{1}{o} - \frac{1}{i} = \frac{1}{f}$$

$$\frac{\sin(\theta_1)}{\lambda_1} = \frac{x}{l} = \frac{\sin(\theta_2)}{\lambda_2}$$

$$\frac{\sin(\theta_1)}{\lambda_1} = \frac{\sin(\theta_2)}{\lambda_2}$$

which is true for any value of x . x does not need to be 1.0. The final line should also hold in the limit as $l \rightarrow 0$, and consequently $x \rightarrow 0$.

There is one more bit of physics to refraction: the frequency of the radiation in medium "2" will be the same as the frequency in medium "1". Why is this true? At the interface, medium "1" and medium "2" share the same location. So the time dependence of the fields are the same at the interface and consequently in the different media. That is $f_1 = f_2 \equiv f$. Dividing both sides of the equation by f yields

$$\frac{\sin(\theta_1)}{f\lambda_1} = \frac{\sin(\theta_2)}{f\lambda_2}$$

$$\frac{\sin(\theta_1)}{v_1} = \frac{\sin(\theta_2)}{v_2}$$

This refraction equation is true for waves in general. For light, in terms of the indices of refraction, we have

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \tag{16}$$

This equation is known as Snell's Law.

Is refraction the only thing that can occur when light travels from one medium into another? No, light can (and usually does) reflect as well. For the reflected light, the angle of reflection equals the angle of incidence as with a mirror. Actually, the derivation of Snell's law will work for reflection. In this case $n_1 = n_2$, since the medium is the same. If $n_1 = n_2$ then $\theta_1 = \theta_2$. So, in general when light passes from one medium into another there can be reflection as well as refraction. Usually, both occur.

Some points regarding reflection and refraction:

1. The angles in Snell's law are measured from the ray to a line normal (or perpendicular) to the interface.

- If light travels from a faster medium into a slower medium, rays bend towards the normal. If light travels from a slower medium into a faster medium, rays bend away from the normal.
- Suppose light travels from medium "1" into medium "2", and $n_1 > n_2$. That is, light travels from a slower medium into a faster one. Since

$$\sin(\theta_2) = \frac{n_1}{n_2} \sin(\theta_1) \quad (17)$$

the angle θ_2 is greater than θ_1 . In this case, the refracted light bends away from the normal. If $(n_1/n_2)\sin(\theta_1)$ is larger than one, there is no real solution for θ_2 . The light cannot refract into medium "2". All the light is reflected! We call this phenomenon *total internal reflection*. The critical angle θ_c when this happens is when

$$\begin{aligned} \frac{n_1}{n_2} \sin(\theta_c) &= 1 \\ \sin(\theta_c) &= \frac{n_2}{n_1} \end{aligned}$$

For angles $\theta > \theta_c$ from the normal, all the incident light is reflected.

Snell's law is the "physics" of refraction. Now we can use this understanding in practical applications.

Thin Lenses

Glass, or other transparent materials, can be formed such that parallel rays focus at a point on the other side of the glass. The glass is called a *converging lens*. Alternatively, glass (or other transparent material) can be formed such that parallel rays diverge from a point. In this case we have a *diverging lens*. In lecture we will show that a converging lens is thicker in the middle than on the edges, and a diverging lens is thinner in the middle than the edges. Let's first consider thin converging lenses, then thin diverging lenses. The geometry will be very similar to the mirror analysis, and the final thin lens formula will be the same as the one for mirrors.

Converging Lenses

As with mirrors, we will follow (the same) three rays to examine where the image is formed and if the image is real or virtual. First, consider an object that is placed between the focal point and the lens, $o < f$. Let the object have a height h_0 and be located a distance o ($o < f$) from the lens. The object is represented by an arrow in the figure. We can trace three rays from the tip of the arrow that travel through the lens:

A ray parallel to the axis (2 → 3)

The ray that is parallel to the axis will go through the lens and pass through the focal point. The dashed line extends this ray back behind the lens.

A ray directed away from the focal point (2 → 8)

The ray that is directed away from the focal point will emerge parallel to the axis. The dashed line extends the refracted ray back behind the lens.

A ray that strikes the center of the lens (2 → 6)

The ray that strikes the center of the mirror will pass straight through at the same angle that it strikes the lens. This part of the lens is like a plane window. The dashed line extends the transmitted ray back behind the lens.

All three of the dashed lines intersect at a point. Although we only considered these three rays, the dashed lines for all rays from the tip would intersect there as well. Wherever you look into the lens from the left side, you will see the tip of the arrow at this intersection point. The intersection point is the image of the arrow's tip. Since the image is on the same side as the object, it is a virtual image. We label i as the distance from the image to the mirror, h_i as the height of the image, and h_o as the height of the object as before.

Using geometry, we can determine the relationship between o , i , f , h_i and h_o .

Triangle 7 – 4 – 5 and triangle 7 – 3 – 6 are similar triangles:

$$\frac{h_i}{h_o} = \frac{f + i}{f} \quad (18)$$

Triangle 6 – 1 – 2 and triangle 6 – 4 – 5 are similar triangles:

$$\frac{h_i}{h_o} = \frac{i}{o} \quad (19)$$

Equating the h_i/h_o we have

$$\frac{i}{o} = \frac{i+f}{f}$$
$$\frac{1}{o} - \frac{1}{i} = \frac{1}{f}$$

The ratio of h_i/h_o is how much the image increased in size, and is called the *magnification* m as in the case with the mirror. For the case just treated, we see that $m = h_i/h_o = i/o$.

Now, lets use the same lens, but place the object such that $o > f$. We can follow the same rays from the tip of the arrow as before:

A ray parallel to the axis (2 → 3)

The ray that is parallel to the axis will go through the lens and pass through the focal point. The dashed line extends the refracted ray back behind the lens.

A ray directed towards the focal point (2 → 8)

The ray that is directed towards the focal point will emerge parallel to the axis. The dashed line extends the refracted ray back behind the lens.

A ray that strikes the center of the lens (2 → 6)

The ray that strikes the center of the mirror will pass straight through at the same angle that it strikes the lens. The dashed line extends the transmitted ray back behind the lens.

From the figure we can see that the dashed lines do not converge. One will not see an image of the arrow in the lens. However, in this case, the transmitted rays do converge. The rays from point "2" do converge at point "4". However, now point "4" is on the other side of the lens. Although we just followed three rays, all rays from point "2" that hit the lens will converge at point "4". If we put a piece of paper, or screen, where point "4" is, we would see an image of the arrow on the screen. An image that can be projected on a screen is called a *real image*. If $o > f$ the image is real.

Once again, we can use geometry to find the relationship between o , i , f , h_o and h_i :

Triangle 7 – 4 – 5 and triangle 7 – 3 – 6 are similar triangles:

$$\frac{h_i}{h_o} = \frac{i - f}{f} \quad (20)$$

Triangle 2 – 6 – 1 and triangle 4 – 5 – 6 are similar triangles:

$$\frac{h_i}{h_o} = \frac{i}{o} \quad (21)$$

Once again, equating the h_i/h_o we have

$$\begin{aligned} \frac{i}{o} &= \frac{i - f}{f} \\ \frac{1}{o} + \frac{1}{i} &= \frac{1}{f} \end{aligned}$$

As before, the magnification m equals $h_i/h_o = i/o$. For the real image, the magnification can be less than one ($o > 2f$) or greater than one ($f < o < 2f$).

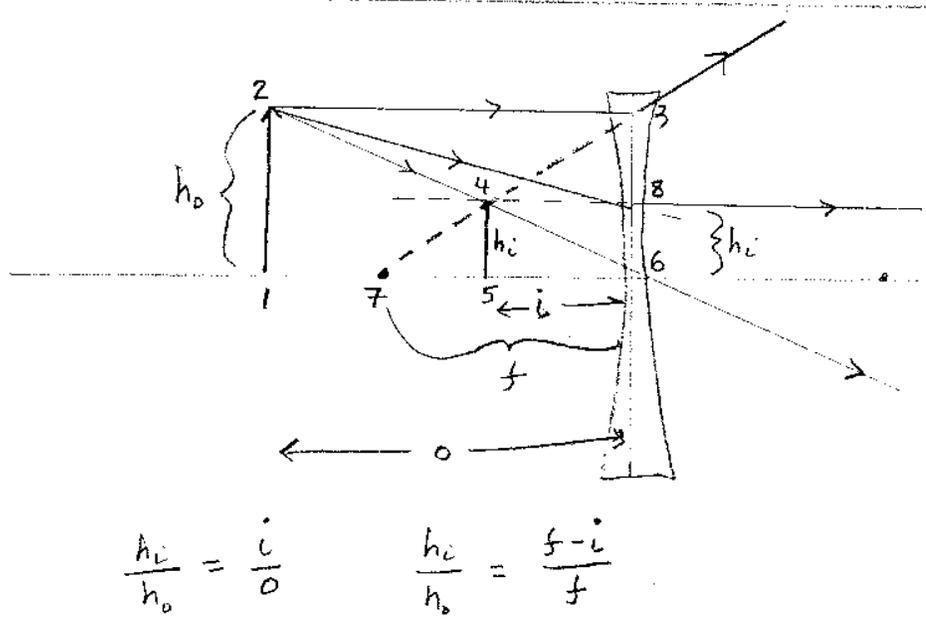
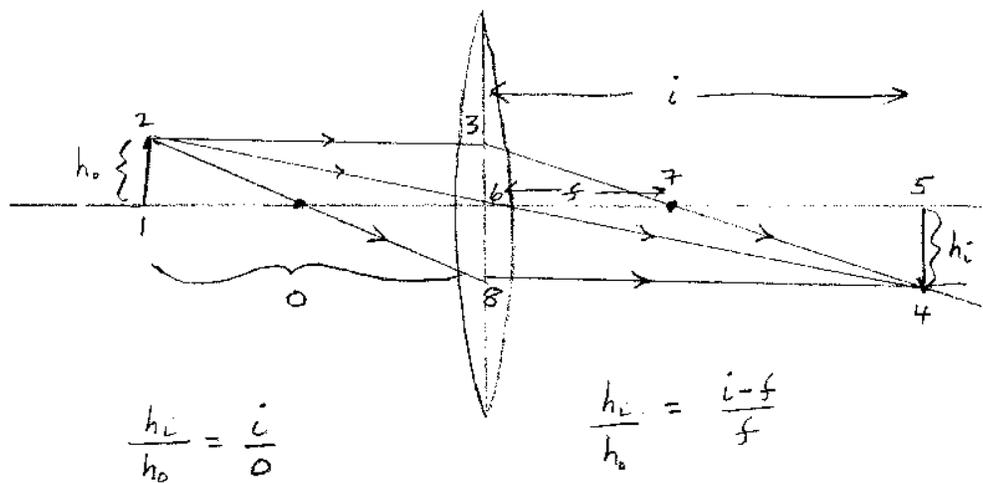
As with the mirrors, the equation for $o < f$ is similar to the equation for $o > f$. The only difference is the minus sign in front of i . We can combine both equations into one, which will be valid for any value for o if we assign a minus sign to a virtual image. That is, $i > 0$ means the image is on the other side of the lens, and $i < 0$ means the image is on the same side of the lens as the object. The universal equation for all values of o , i , and f for a converging lens is

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f} \quad (22)$$

where $i > 0$ for a real image, and $i < 0$ for a virtual image. This equation also works for a flat window. For a flat window, $f = \infty$, and we have $i = -o$.

Diverging Lens

Finally, we consider a diverging lens. This is a lens that is thinner in the center than on the edges. If the lens is formed properly, parallel light rays will diverge on the other side of the lens as if they are coming from a single point. Let's examine the image that is formed with this kind of lens. Place an arrow a distance o from the lens. We trace the same three rays as before:



A ray parallel to the axis (2 → 3)

The ray that is parallel to the axis will go through the lens and diverge from the focal point. The dashed line extends the refracted ray back behind the lens.

A ray directed towards the focal point (2 → 8)

The ray that is directed towards the focal point will emerge parallel to the axis. The dashed line extends the refracted ray back behind the lens.

A ray that strikes the center of the lens (2 → 6)

The ray that strikes the center of the mirror will pass straight through at the same angle that it strikes the lens. The dashed line extends the transmitted ray back behind the lens.

From the figure we can see that the dashed lines do converge. One will see an image of the arrow in the lens. The rays from point "2" do converge at point "4". However, point "4" is on the same side of the lens as the object. Although we just followed three rays, all rays from point "2" that hit the lens will diverge from point "4". If we look into the lens, we will see the tip of the arrow at the point "4", and the arrow will be a virtual image.

Once again, we can use geometry to find the relationship between o , i , f , h_0 and h_i :

Triangle 7 – 4 – 5 and triangle 7 – 3 – 6 are similar triangles:

$$\frac{h_i}{h_0} = \frac{f - i}{f} \quad (23)$$

Triangle 2 – 6 – 1 and triangle 4 – 5 – 6 are similar triangles:

$$\frac{h_i}{h_0} = \frac{i}{o} \quad (24)$$

Once again, equating the h_i/h_o we have

$$\begin{aligned} \frac{i}{o} &= \frac{f - i}{f} \\ \frac{1}{o} - \frac{1}{i} &= -\frac{1}{f} \end{aligned}$$

Wow, this equation is very similar to the equation for the converging lens. If we assign a negative value to the focal length of a diverging lens, then we can use *one equation*

for both converging and diverging lenses, and the cases which have real or virtual images. Thus, we can produce a *Universal Equation for converging and diverging lenses*:

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f} \quad (25)$$

where o and i are greater than zero for real objects and images, and less than zero for virtual objects and images. The focal length f is positive for converging lenses and negative for diverging lenses. This same equation applies universally for mirrors and lenses because the properties of these elements are the same: *Parallel rays converge (focus) or diverge from the focal points of both mirrors and lenses.*

A few final points on lenses and Snell's law:

1. Converging lenses are also called convex lenses. Diverging lenses are also called concave lenses.
2. Real images are only formed with converging lenses, and only when $o > f$.
3. The magnification $m = |h_i/h_o|$ as with mirrors. If we adopt the convention that + means upright and - means inverted, then $m = -i/o$ as with mirrors.
4. The index of refraction depends on the frequency (or color) of the light. For water, a wavelength of $\lambda = 400 \text{ nm}$, has $n = 1.35$. For water, a wavelength of $\lambda = 700 \text{ nm}$, has $n = 1.33$. Thus, for water, red light travels slightly faster than blue light. Blue light will therefore refract more than red in going from air into water, giving us a nice rainbow.

Other topics in Optics

We have covered the main idea used in geometric optics, Snell's law. Considering the different polarizations of light gives interesting results. If there is time in lecture we might discuss:

1. Images using two lenses: In this case, the image of the first lens is the object of the second lens. Applications include telescopes and microscopes.

2. Natural polarization: If light scatters at a special angle, Brewster's angle, it will be polarized. The special incident angle is when the reflected ray makes an angle of 90° with the refracted ray. Then the reflected light will be polarized in a direction parallel to the surface of the interface. Polaroid lenses are designed to greatly reduce this "glare".
3. Polarization in the sky: Light scattered at 90° in the sky will also be polarized. If you have polarized glasses, you can check this out, as well as Brewster's angle.
4. The lens maker's formula. The lens maker's formula relates the radii of curvature of the lens to the focal length.

There are so many topics we can cover in optics, we have two upper division optics courses. If you are interested in optics, I recommend you take one (or both) of these classes: Phy417/Phy417L and/or Phy344(Applied Optics). As a final important topic in optics, we cover interference effects.

Interference Effects

We will have demonstrations in lecture that show similar interference effects with visible light that we saw and discussed in Phy132 with sound. We will shine light at screens with slits, etc. and look at what happens when light goes through the slits or reflects off thin films. We will see that at some positions the transmitted (or reflected) light will be bright, and at other positions dark. We can explain these phenomena using the principle of superposition. In this course we will use the classical approach of treating the electromagnetic radiation as an electric-magnetic field. The net electric field at any point in space will be the superposition of the electric fields from all the sources of the radiation.

We will consider light of one frequency (i.e. one color). Thus, the net electric field will consist of the sum of sinusoidal functions that have the same frequency. We first go over the mathematics of adding sinusoidal functions that have the same frequency, which is the basis for understanding these interference effects.

Adding Sinusoidal Functions that have the same Frequency

The general problem is to add terms of the following form:

$$\sum_i a_i \cos(\omega t + \phi_i) = a_1 \cos(\omega t) + a_2 \cos(\omega t + \phi_2) + \dots \quad (26)$$

where we have measured all phase angles ϕ_i relative to the first term for simplicity. There is no generality lost in taking $\phi_1 = 0$, we have just started our time t when the first sinusoidal function is at a maximum $+a_1$. Before we go into the details, we can guess what this sum will look like. First, after a time equal to the period, $T = 2\pi/\omega$ each term has gone through a complete cycle and the sum is the same as it was. Thus, the sum of the sinusoidals will be periodic with the same period, or frequency, as the terms. We will also show that the sum will be a sinusoidal function as well. The sum will be therefore equal to $A\cos(\omega t + \phi)$:

$$\sum_i a_i \cos(\omega t + \phi_i) = A \cos(\omega t + \phi) \quad (27)$$

Our goal is to find a mathematical way to obtain A and ϕ from the a_i and ϕ_i . I'll present three different ways to do this: using trig identities, "phasors", and complex numbers. I'll also show the special case of two functions with $a_1 = a_2$.

Using Trigonometric Identities

Let's start with the simple case of two sinusoidals where $a_1 = a_2$. We can use the trig identity:

$$\cos(\alpha) + \cos(\beta) = 2 \cos\left(\frac{\alpha + \beta}{2}\right) \cos\left(\frac{\alpha - \beta}{2}\right) \quad (28)$$

Now, if $a_1 = a_2$, $\alpha = \omega t$, and $\beta = \omega t + \phi_2$, we have

$$\begin{aligned} a_1 \cos(\omega t) + a_1 \cos(\omega t + \phi_2) &= 2a_1 \cos(\omega t + \phi_2/2) \cos(-\phi_2/2) \\ &= 2a_1 \cos(\phi_2/2) \cos(\omega t + \phi_2/2) \end{aligned}$$

where we have used $\cos(-\theta) = \cos(\theta)$. So we see that the sum of these two cosine functions is a cosine function with the same frequency ω . The resulting phase is $\phi_2/2$ from the first term and the Net Amplitude A is $A = 2a_1 \cos(\phi_2/2)$.

From the expression for the net amplitude $A = 2a_1 \cos(\phi_2/2)$ we can see how the interference effects can be understood. If $\phi_2 = 0, \pm 2\pi, \pm 4\pi, \dots$, then $\cos(\phi_2/2) = 1$. The net amplitude will be a maximum, the light will be bright, constructive interference will occur. If $\phi_2/2 = \pm\pi, \pm(3/2)\pi, \dots$, then $\cos(\phi_2/2) = 0$. The net amplitude will be zero, the light will be gone, destructive interference will occur. For other values of ϕ_2 the amplitude will be $2a_1 \cos(\phi_2/2)$ and the intensity will be proportional to the square of the amplitude $A^2 = 4a_1^2 \cos^2(\phi_2/2)$.

If $a_1 \neq a_2$ we could perform a similar analysis, but it will be easier to use the phasor method described next.

Using Phasors

Since all the sinusoidal functions have the same frequency, we can add them just like we add vectors. Let's demonstrate this. At time $t = 0$ the sum is

$$a_1 + a_2 \cos(\phi_2) \quad (29)$$

This sum is the projection on the x-axis of the vector sum of the two vectors $\vec{a}_1 = a_1 \hat{i}$ and \vec{a}_2 , where $|\vec{a}_2| = a_2$ and \vec{a}_2 makes an angle of ϕ_2 with respect to the x-axis, as shown in the figure.

At time t , both vectors in the figure have rotated by the same angle ωt . The sum of the sinusoidal functions are, at time t , equal to

$$a_1 \cos(\omega t) + a_2 \cos(\omega t + \phi_2) \quad (30)$$

Once again, this sum is the projection on the x-axis of the vector sum of the two vectors as shown in the figure. The result of the above equation is true for all times.

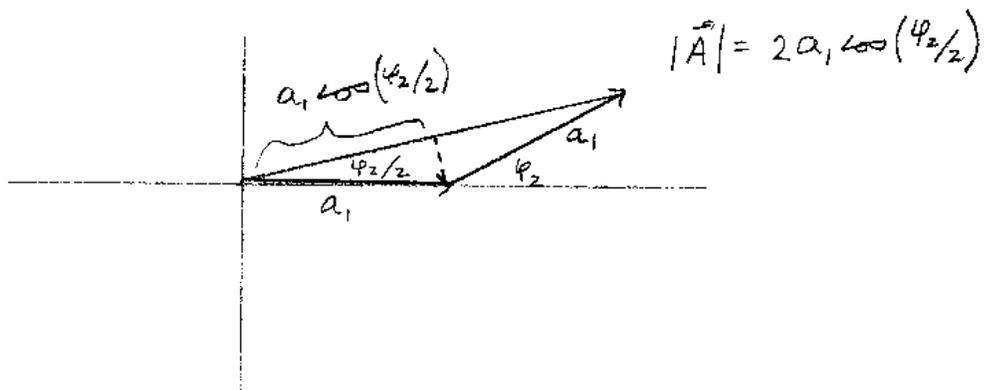
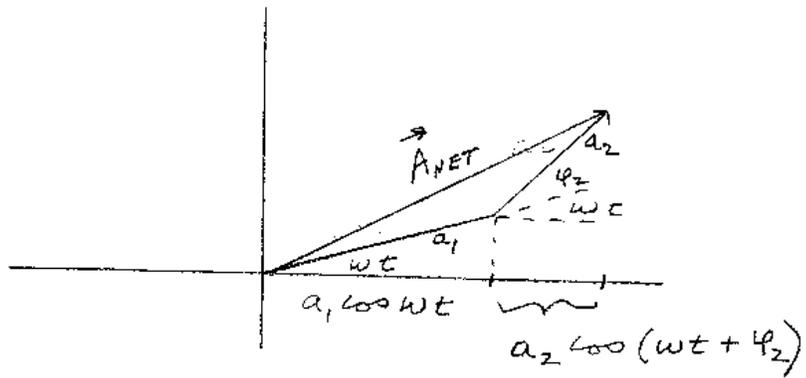
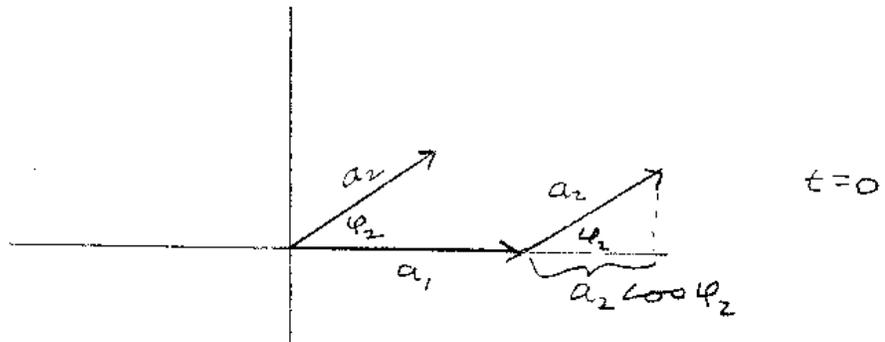
That is, we can define the vector $\vec{a}_1(t)$ to have a magnitude $|\vec{a}_1| = a_1$ and make an angle of ωt with respect to the x-axis, and define the vector $\vec{a}_2(t)$ to have a magnitude or $|\vec{a}_2| = a_2$ and make an angle of $\omega t + \phi_2$ with respect to the x-axis. At any time t , we can add these two vectors. The projection of the resulting vector sum ($\vec{a}_1(t) + \vec{a}_2(t)$) on the x-axis will be the sum of the two sinusoidal functions at time t .

Note the important result: **The magnitude of $\vec{a}_1(t) + \vec{a}_2(t)$ does not change in time!** The vector sum $\vec{a}_1(t) + \vec{a}_2(t)$ just rotates around with angular velocity ω ! Thus the net amplitude of the sum of the sinusoidal functions is just the magnitude of the vector sum: $|\vec{a}_1(t) + \vec{a}_2(t)|$. We can evaluate this sum at any time. It easiest to evaluate the vector sum at $t = 0$, which we have chosen such that $\vec{a}_1(0)$ lies along the x-axis.

In the figure, we evaluate the vector sum for the special case of $a_2 = a_1$. Here we see that the magnitude $|\vec{a}_1(0) + \vec{a}_2(0)|$ is equal to

$$|\vec{a}_1(0) + \vec{a}_2(0)| = 2a_1 \cos(\phi_2/2) \quad (31)$$

which is the same result we obtained using trig identities. The angle that the vector $\vec{a}_1(0) + \vec{a}_2(0)$ makes with respect to the x-axis is $\phi_2/2$. This will be the phase relative of the sum to the first sinusoidal.



We can use this procedure of adding sinusoidal functions of the same frequency to any number of functions. One simply associates a vector with each term in the sum. Then one adds the vectors by the rules of vector addition to obtain a net vector. The magnitude of the net vector is the magnitude of the sum of sinusoidals. The angle of the net vector is the relative phase of the sum of sinusoidals. One can say the same thing mathematically as follows. For a sinusoidal function with frequency ω , amplitude a_k , and phase ϕ_k , associate the vector \vec{a}_k :

$$\begin{aligned} a_k \cos(\omega t + \phi_k) &\leftrightarrow \vec{a}_k \\ &\leftrightarrow a_k \cos(\phi_k) \hat{i} + a_k \sin(\phi_k) \hat{j} \end{aligned}$$

Note that $|\vec{a}_k| = a_k$ and ϕ_k is the angle between \vec{a}_k and the x-axis. Here I am using k as the index for the sinusoidal function, since we are using \hat{i} as the unit vector along the x-axis. The addition of the sinusoidal functions is found by adding the vectors:

$$\begin{aligned} \sum_i a_i \cos(\omega t + \phi_i) &= ? \\ &\leftrightarrow \sum_i \vec{a}_i \quad (\text{Assign the Vectors}) \\ &\leftrightarrow \vec{A}_{net} \quad (\text{add the vectors}) \\ &\leftrightarrow A \cos(\omega t + \phi) \quad (\text{Map back to sinusoidals}) \\ \sum_i a_i \cos(\omega t + \phi_i) &= A \cos(\omega t + \phi) \end{aligned}$$

The addition of sinusoidal functions of the same frequency is isomorphic to vector addition. Instead of using the word vector, phasor is often used. This is because these vectors are all rotating around the z-axis with the same frequency ω . A phasor is a vector that is rotating around an axis with a constant angular velocity.

Using Complex Numbers

Adding complex numbers obeys the same mathematics as adding two dimensional vectors. The real parts add independently to the imaginary parts. For every sinusoidal function we associate a complex number:

$$\begin{aligned}
a_k \cos(\omega t + \phi_k) &\leftrightarrow z_k \\
&\leftrightarrow a_k e^{i\phi_k}
\end{aligned}$$

I use the index k , since i is assigned $\sqrt{-1}$ when using complex numbers. As with the vector (or phasor) method one just adds the complex numbers:

$$\sum_k a_k e^{i\phi_k} = A = |A| e^{i\phi} \quad (32)$$

One just needs to express the complex number A in terms of its magnitude $|A|$ and its phase ϕ . The reason this works for complex numbers is as follows:

$$\begin{aligned}
\sum_k a_k \cos(\omega t + \phi_k) &= \sum_k \operatorname{Re}(a_k e^{i(\omega t + \phi_k)}) \\
&= \operatorname{Re}\left(\sum_k a_k e^{i(\omega t + \phi_k)}\right) \\
&= \operatorname{Re}\left(e^{i\omega t} \left(\sum_k a_k e^{i\phi_k}\right)\right) \\
&= \operatorname{Re}\left(e^{i\omega t} A e^{i\phi}\right) \\
&= \operatorname{Re}\left(A e^{i(\omega t + \phi)}\right) \\
&= A \cos(\omega t + \phi)
\end{aligned}$$

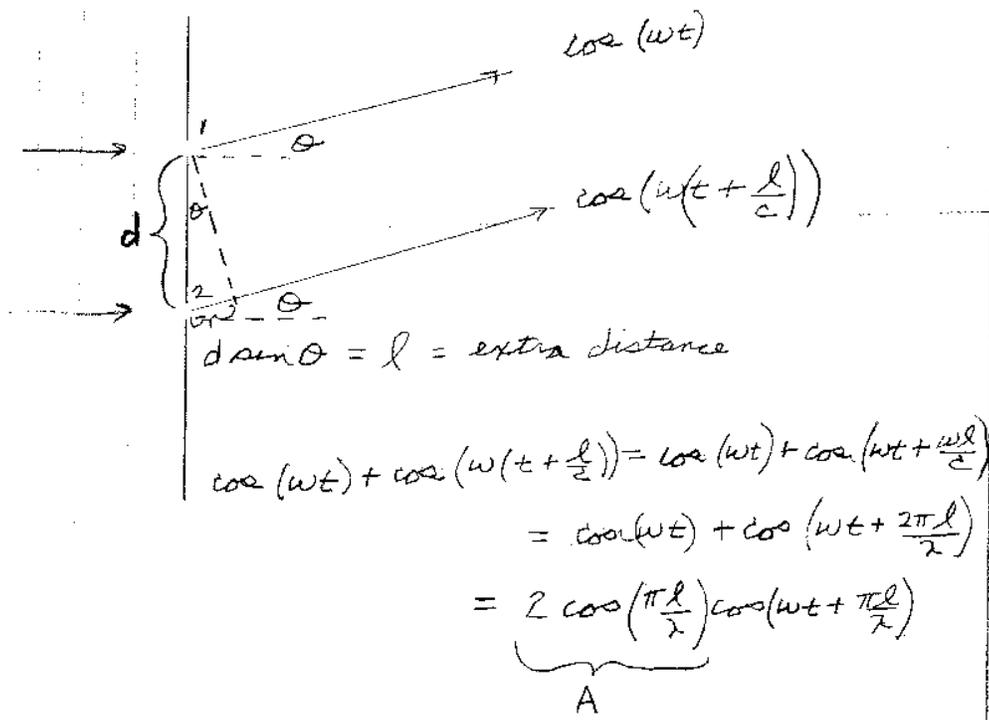
where Re means take the real part of the complex number. Note that it was critical that each function have the same frequency. Otherwise, we could not have removed the common term $e^{i\omega t}$ outside the summation. The addition of sinusoidal functions of the same frequency is isomorphic to the addition of complex numbers.

Now we are ready to see if the "interference" experiments with monochromatic light can be understood using the mathematics of adding sinusoidal functions. We will investigate what happens when light interacts with two slits, many slits, a single slit, thin films, etc.

"Interference" Experiments with light

Double Slit Interference

In the double slit experiment, light is emitted that has a *single frequency* and is from a *single source*. The light hits a screen that has two thin slits, slit "1" and slit



$$\begin{aligned}
 \cos(\omega t) + \cos(\omega(t + \frac{l}{c})) &= \cos(\omega t) + \cos(\omega t + \frac{\omega l}{c}) \\
 &= \cos(\omega t) + \cos(\omega t + \frac{2\pi l}{\lambda}) \\
 &= \underbrace{2 \cos(\frac{\pi l}{\lambda})}_{A} \cos(\omega t + \frac{\pi l}{\lambda})
 \end{aligned}$$

"2". The slits are separated by a distance d . After going through the two slits the light hits a second screen. On the second screen, one sees bright and dark regions.

This phenomena of bright and dark regions can be understood as follows. When the light hits the second screen, the net amplitude is the superposition of the light from slit "1" plus the light from slit "2". At the position on the second screen between the two slits, the distance to each slit is the same. The light from slit "1" has the same phase as the light from slit "2". A bright spot will be formed at this middle position. At positions on the second screen away from the middle, the distance to slit "1" is different than from slit "2". Maxwell's picture of monochromatic light is that it is an oscillating electric field vector. If the distance to slit "1" is different than to slit "2", then the oscillating electric field from slit "1" will have a different phase with respect to the oscillating electric field from slit "2". The net electric field will be the sum of two sinusoidal functions that have a relative phase between them.

Let's calculate the net electric field. If the distance to slit "1" is not too much different than that to slit "2", then the amplitude from each slit is approximately the same, say E_0 . So the net electric field on the second screen is the sum of the electric field from slit "1" plus that from slit "2":

$$\begin{aligned}\vec{E}_{net} &= \vec{E}_1 + \vec{E}_2 \\ &\approx \vec{E}_0(\cos(\omega t) + \cos(\omega t + \phi))\end{aligned}$$

What is the phase difference ϕ ? It is due to the extra distance that the light must travel from one slit compared to the other. Let the extra distance be l . Then the extra time to reach the screen is l/c . So one "wave" arrives at time t , and the other "wave" arrives at a time of $t + l/c$. Substituting into the equation we have

$$\begin{aligned}\vec{E}_{net} &\approx \vec{E}_0(\cos(\omega t) + \cos(\omega(t + l/c))) \\ &\approx \vec{E}_0(\cos(\omega t) + \cos(\omega t + \omega l/c)) \\ &\approx \vec{E}_0(\cos(\omega t) + \cos(\omega t + 2\pi f l/c)) \\ &\approx \vec{E}_0(\cos(\omega t) + \cos(\omega t + 2\pi l/\lambda))\end{aligned}$$

where we have used $c = f\lambda$ to get the last line. Thus, the difference in the sinusoidal phase of the light from the two slits is

$$\phi = 2\pi \frac{l}{\lambda} \tag{33}$$

We know how to add sinusoidal functions from the last section. For the case at hand, two slits, we have

$$\begin{aligned}\vec{E}_{net} &\approx \vec{E}_0(\cos(\omega t) + \cos(\omega t + \phi)) \\ &\approx \vec{E}_0 2\cos(\phi/2)\cos(\omega t + \phi/2) \\ &\approx \vec{E}_0 2\cos(\pi l/\lambda)\cos(\omega t + \phi/2)\end{aligned}$$

Here we see that the amplitude of the light when it strikes the second screen is

$$|\vec{E}_{net}| = 2E_0\cos(\pi\frac{l}{\lambda}) \quad (34)$$

This equation for \vec{E}_{net} enables us to understand the two slit experiment. The intensity of the light is proportional to the square of the electric field strength:

$$I = 4I_0\cos^2(\pi\frac{l}{\lambda}) \quad (35)$$

At the middle position, $l = 0$, and the intensity is $4I_0$. If the extra distance $l = \pm\lambda/2$, $\cos(\pi/2) = 0$, then there is a dark spot. Dark spots will also occur whenever $l = \pm 3\lambda/2, 5\lambda/2, \dots$. Bright spots will occur whenever $l = 0, \lambda, 2\lambda, \dots$. The intensity, or brightness, varies as $\cos^2(\pi l/\lambda)$.

We can express the extra distance l in terms of the "scattering" angle θ shown in the figure. If θ is small, then

$$l \approx d\sin(\theta) \quad (36)$$

So, in terms of the scattering angle θ the intensity on the second screen is

$$I = 4I_0\cos^2\left(\frac{\pi d \sin(\theta)}{\lambda}\right) \quad (37)$$

Some final comments on the two slit experiment:

1. It is important that the light coming from slit "1" is in sych with the light coming from slit "2" at the slit opening at all times. For the case of a single monochromatic source impinging directly on the two slits, the light emanating from each slit is the same (at the slit opening). If the single monochromatic source were closer to one slit than the other, the light from "1" would still be in sych with the light from slit "2". The relative phase would not change in time. The radiation from sources that are in sych is called *coherent*.

2. For interference to occur, the two "sources", slit "1" and slit "2", must be coherent longer than a distance l . The maximum distance that the sources remain coherent is called the *coherence length*. In our case, the coherence length must be longer than l for interference to exist.
3. In a real experiment, the intensity does not decrease exactly as $\cos^2(\pi d \sin(\theta)/\lambda)$. The slits will have a finite opening size, and this will affect the pattern on the second screen. We will consider this effect later.
4. The middle point is often called the central maximum.
5. The two slit experiment was first performed by Young. The experiment is sometimes referred to as "Young's double slit experiment". The experiment was one of the first to show the interference effects of light. It led scientists to the incorrect conclusion that light is a classical wave.

Multiple Slit Interference

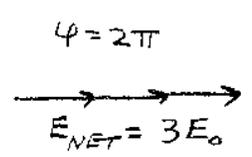
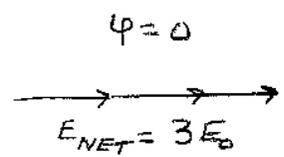
Now let's consider what will happen if light is incident on a screen with three equally spaced thin slits, located on the screen. The light will travel through the slits and strike a wall. On the wall, we will see an interference pattern. We can carry out the same analysis that we did with two slits. Let the spacing between each slit be d . Label the slits "1", "2", and "3". If E_0 is the amplitude on the wall from each slit, then the amplitude at the central maximum on the wall is $3E_0$. On the wall away from the central maximum (upward in the figure), slit "1" will be closer than slit "2" by a length l , and slit "2" will be closer than slit "3" by a length l . The phase difference between adjacent slits will be $\phi = 2\pi l/\lambda$. The net amplitude will be the sum of the amplitudes from each of the three slits:

$$E_{net} = E_0 \cos(\omega t) + E_0 \cos(\omega t + \phi) + E_0 \cos(\omega t + 2\phi) \quad (38)$$

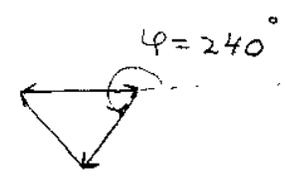
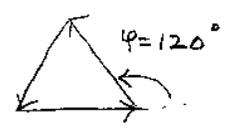
We can add the sinusoids using phasors. The first term is associated with a vector of length E_0 along the x-axis. The second sinusoidal is associated with a vector of length E_0 at an angle of ϕ with respect to the x-axis. The third sinusoidal is associated with a vector of length E_0 at an angle of 2ϕ with respect to the x-axis. The magnitude of E_{net} is the magnitude of the sum of the three vectors.

One of your homework problems is to solve for E_{net} in terms of ϕ . Here we will determine the values of ϕ that result in minima and maxima. Maxima occur when $\phi = 0^\circ$ and when $\phi = 360^\circ$. For a minima to occur, the three vectors must add to

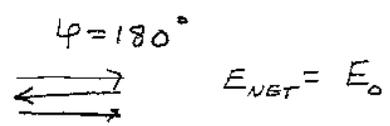
3 slits
 Super
 maxima:



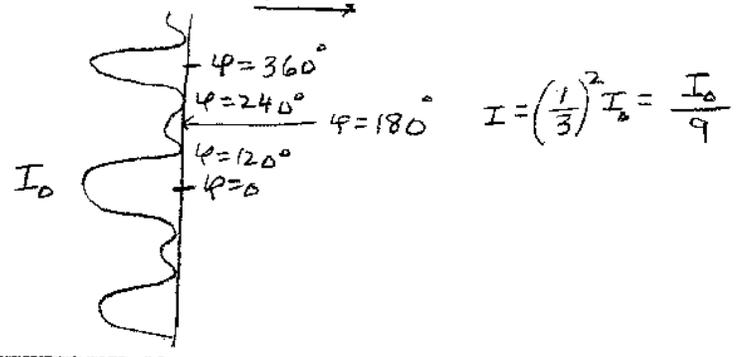
minima



"Local"
 maximum

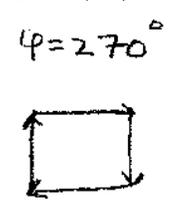
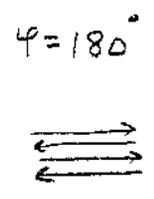
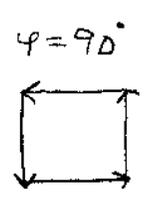


Pattern:

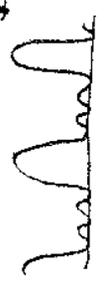


4 slits

minima



Pattern



zero. The first value of ϕ for the sum to be zero is when $\phi = 120^\circ$. The next value of ϕ for the sum to add to zero is when $\phi = 240^\circ$. These are the only values for $0^\circ < \phi < 360^\circ$ where there are minima. In the middle of the minima at $\phi = 120^\circ$ and 240° there must be a maximum, a "local maximum". This will occur at $\phi = 180^\circ$. Substituting in the equation for E_{net} , we have $E_{net} = E_0 - E_0 + E_0 = E_0$. The pattern on the wall will be the following. E_{net} has maxima of $3E_0$ when $\phi = 0, \pm 2\pi, \pm 4\pi, \dots$. E_0 has smaller local maxima at $\phi = \pm\pi, \pm 3\pi, \dots$.

The intensity of the light is proportional to E_{net}^2 . So the intensity at the maxima when $\phi = 0^\circ$ and $\phi = 360^\circ$ will be 9 times the intensity at the local maximum when $\phi = 180^\circ$. As we move across the wall from the central maximum, the intensity will vary as: $I_0, 0, I_0/9, 0, I_0, 0, I_0/9, 0, I_0, \dots$.

We can perform the same analysis with 4 slits as we did with 3. As before, let ϕ be the relative phase between light from one slit and its adjacent slit. The net amplitude will be the sum of four sinusoidals:

$$E_{net} = E_0 \cos(\omega t) + E_0 \cos(\omega t + \phi) + E_0 \cos(\omega t + 2\phi) + E_0 \cos(\omega t + 3\phi) \quad (39)$$

As before we can add the sinusoidals using phasors. The first term is associated with a vector of length E_0 along the x-axis. The second sinusoidal is associated with a vector of length E_0 at an angle of ϕ with respect to the x-axis. The third sinusoidal is associated with a vector of length E_0 at an angle of 2ϕ with respect to the x-axis. The fourth sinusoidal is associated with a vector of length E_0 at an angle of 3ϕ from the x-axis. The magnitude of E_{net} is magnitude of the sum of the four vectors.

Let's try and determine where the minima and maxima are. The central maximum is when $\phi = 0$, and $E_{net} = 4E_0$. Equally large maxima will occur when $\phi = \pm 2\pi, \pm 4\pi, \dots$. Since $E_{net}(\phi \pm 2\pi) = E_{net}(\phi)$, we can limit our analysis for $0 \leq \phi \leq 2\pi$.

To find the minima for $0 \leq \phi \leq 2\pi$, we can slowly increase ϕ and see when the four vectors add up to zero. The first minima will occur at $\phi = 90^\circ$. The vectors form a square going counter-clockwise. The next minima will be when $\phi = 180^\circ$. The vectors cancel each other out. The next minima will be when $\phi = 270^\circ$. The vectors form a square going clockwise. So there are three minima (dark spots) for $0 \leq \phi \leq 2\pi$. In between the minima, there will be local (smaller) maxima. One for $90^\circ \leq \phi \leq 180^\circ$, and another one for $180^\circ \leq \phi \leq 270^\circ$. At the local maxima, the net amplitude will be approximately E_0 by examining how the four vectors add. Since intensity is proportional to the square of E_{max} , the intensity at $\phi = 0$ and $\phi = 2\pi$ will be much larger, around $4^2 = 16$ times larger than at the two local maxima between the bright peaks.

Suppose there are N equally spaced narrow slits that are separated from each other by a distance d . Extending the above analysis, the interference pattern on the wall will have a very bright (and narrow) central maximum. The next very bright (and narrow) maximum will be when $\phi = 2\pi$, since all the vectors will add in a straight line. Between the two bright maxima, there will be $N - 1$ dark spots where $E_{net} = 0$. Between each dark spot there will be a local maxima. The amplitude of the $N - 2$ local maxima will be much smaller than at $\phi = 0$ or $\phi = 2\pi$. Since the intensity is proportional to $|E_{max}|^2$, the intensity at $\phi = 0$ or $\phi = 2\pi$ will be N^2 greater than the local maxima in the middle of the pattern.

The features of the three and four slit interference patterns are enhanced when N is large. If N is large, the interference pattern on the wall will consist of very bright lines that are very narrow, with little light between them. We can determine the narrowness of the lines as follows. The central maximum has its first minimum (i.e. dark spot) when $\phi = 2\pi/N$. At this value of ϕ all the vectors sum to zero. ϕ is related to the scattering angle θ as $\phi = 2\pi l/\lambda = (2\pi d \sin(\theta))/\lambda$. So the first minimum occurs for a value of θ given by

$$\begin{aligned} \frac{2\pi d \sin(\theta)}{\lambda} &= \frac{2\pi}{N} \\ \sin(\theta) &= \frac{\lambda}{dN} \\ \theta &\approx \frac{\lambda}{dN} \end{aligned}$$

where we have used the approximation $\sin(\theta) \approx \theta$ for $\theta \ll 1$. The expression $\lambda/(dN)$ is called the *line width* of the diffraction line. This simple expression is applicable for the central maximum.

Screens with many slits are called *diffraction gratings*. Diffraction gratings can have thousand of slits and produce very narrow interference lines. They are used in spectrometers to measure wavelength to great accuracy. In Phy417, you will use complex numbers to solve for N slit interference. Briefly, the calculation goes as

$$\begin{aligned} I &\propto \left| \sum_{n=0}^N e^{in\phi} \right|^2 \\ &\propto \left| \frac{1 - (e^{i\phi})^N}{1 - e^{i\phi}} \right|^2 \end{aligned}$$

$$I = I_0 \frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)}$$

At $\phi \rightarrow 0$, $I \rightarrow N^2 I_0$. The intensity of the local maxima are around I_0 , much smaller than the central maximum.

Summarizing equally spaced multiple slit interference:

1. The main property of radiation we used was that the net amplitude on the wall is equal to the sum of the amplitudes from each of the slits.
2. The process of adding the amplitudes from each of the slits involves adding sinusoidal functions that have a common frequency. Each sinusoidal function has approximately the same amplitude, but with different phase shifts. The difference in phase from one slit to its neighbor is the same value, ϕ , since the slits are equally spaced.
3. In all our discussions, the phase shift ϕ has always been the relative phase from one slit compared to an adjacent slit. In terms of the extra distance l , $\phi = 2\pi l/\lambda$. In terms of the scattering angle θ , $l = d \sin(\theta)$. So, $\phi = 2\pi d \sin(\theta)/\lambda$.

Single Slit Interference

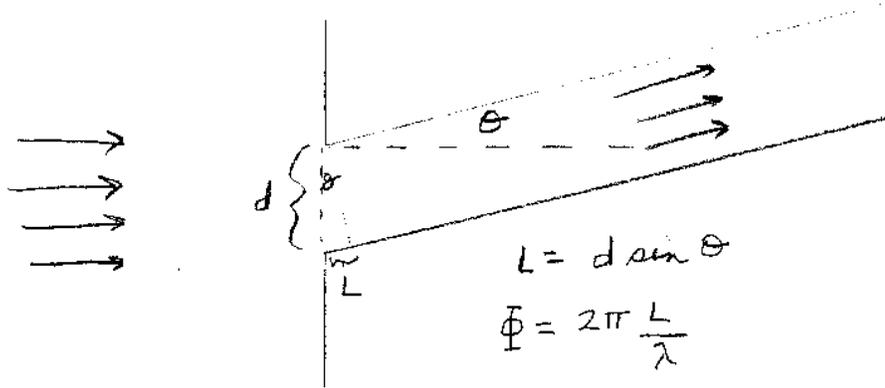
As a final slit interference experiment, we analyze the interference effects when light passes through a single slit. Consider a coherent light beam that hits a screen that has a slit of length d . After the light passes through the slit, it "hits" a wall. In lecture, we will demonstrate this, and notice the interference pattern on the wall. Let the magnitude of the electric field on the wall at the central maximum be E_0 . E_{net} will have its largest value here.

For the case of a thin slit, each slit was a source of light. We can carry this model over to the finite sized slit by dividing the opening of length d into N thin slits separated by a distance of d/N , and letting $N \rightarrow \infty$.

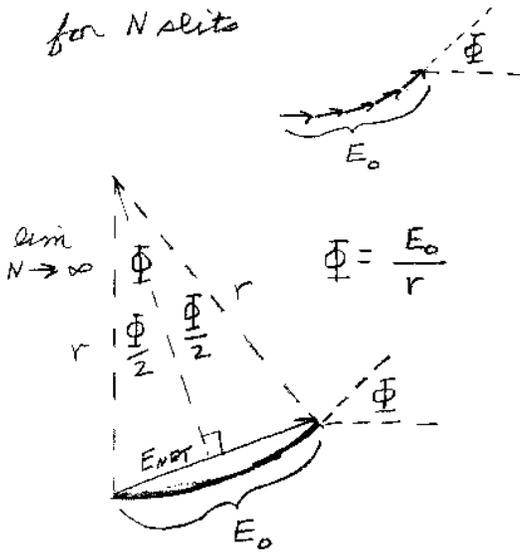
Consider light scattered at an angle θ . The path difference between light emitted at the top of the opening to that at the bottom of the opening is $d \sin(\theta)$, which we label as L , $L = d \sin(\theta)$. The difference in phase between light emitted at the top compared to the bottom of the slit is $2\pi L/\lambda$, which we label as $\Phi = 2\pi L/\lambda$.

If there are N equally spaced sources (of thin slits) in the slit opening, each source will produce an electric field of amplitude E_0/N at the wall. The relative phase between adjacent sources will be $\phi = \Phi/N$. Adding up these N coherent sources at the wall yields a net amplitude E_{net} of

Single Slit



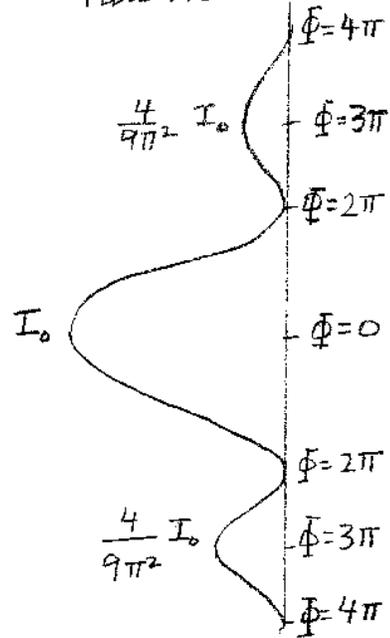
for N slits



$$E_{NET} = 2r \sin \frac{\Phi}{2}$$

$$= E_0 \frac{\sin \Phi/2}{\Phi/2}$$

Pattern



$$E_{net} = \sum_{n=0}^N \frac{E_0}{N} \cos(\omega t + \frac{n\Phi}{N}) \quad (40)$$

We can use phasors (vectors) to add these sinusoidal functions as we did with N slits. If N is large, each vector will be very short and be rotated a small angle, Φ/N , from the adjacent vector. An important property of the sum is that *no matter what N is, the last vector will be rotated by an angle Φ relative to the x -axis*. To find E_{net} , we take the limit as $N \rightarrow \infty$:

$$E_{net} = \lim_{N \rightarrow \infty} \sum_{n=0}^N \frac{E_0}{N} \cos(\omega t + \frac{n\Phi}{N}) \quad (41)$$

Adding the vectors in the limit that $N \rightarrow \infty$ results in a circular arc of length E_0 with the end of the arc rotated by an angle of Φ , as shown in the figure. $|E_{net}|$ is the length of the chord that starts at the origin and ends at the end of the circular arc. From the figure, it is seen that $\Phi = E_0/r$, where r is the radius of the circle. Using geometry, the length of the chord is $2r \sin(\Phi/2)$. Since $r = E_0/\Phi$, we have

$$\begin{aligned} |E_{net}| &= 2r \sin(\Phi/2) \\ &= \frac{2E_0 \sin(\Phi/2)}{\Phi} \\ &= E_0 \frac{\sin(\Phi/2)}{\Phi/2} \end{aligned}$$

The intensity of the light on the wall is therefore

$$I = I_0 \left(\frac{\sin(\Phi/2)}{\Phi/2} \right)^2 \quad (42)$$

where I_0 is the intensity at the central maximum.

Some final points on single slit interference:

1. The first minimum occurs when $\Phi = \pm 2\pi$. The extra distance L traveled from the bottom of the slit is $L = \lambda$. So the first minimum occurs when the radiation emitted from the bottom of the slit is 360° different than that from the top of the slit. This seem strange that cancellation happens, since 360° or 2π radians means that the two sources are in phase. The reason for the total cancellation is that each source in the upper half of the slit cancels with a source in the lower half of the slit. The central maximum region is therefore rather wide.

2. The value of Φ at the first maxima is when $\Phi = \pm 3\pi$. The intensity at these first maxima are

$$I = \left(\frac{2}{3\pi}\right)^2 I_0 = \frac{4}{9\pi^2} I_0 \quad (43)$$

3. In general, there will be minima when $\Phi = \pm 2\pi, \pm 4\pi, \dots$. There will be maxima when $\Phi = \pm 3\pi, \pm 5\pi, \dots$.
4. Another way to obtain the single slit formula is to use the N slit result derived by the complex number way:

$$I = I_0 \frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)} \quad (44)$$

Letting $\phi = \Phi/N$ we have

$$I = I_0 \frac{\sin^2(\Phi/2)}{\sin^2(\Phi/(2N))} \quad (45)$$

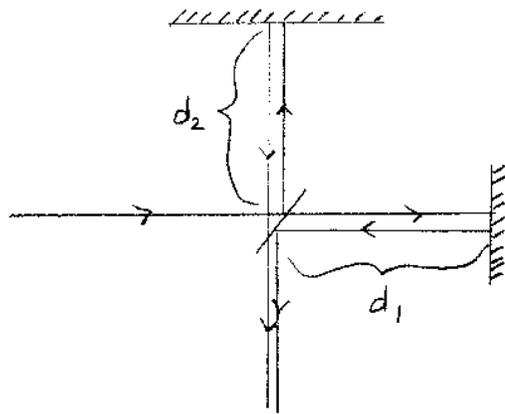
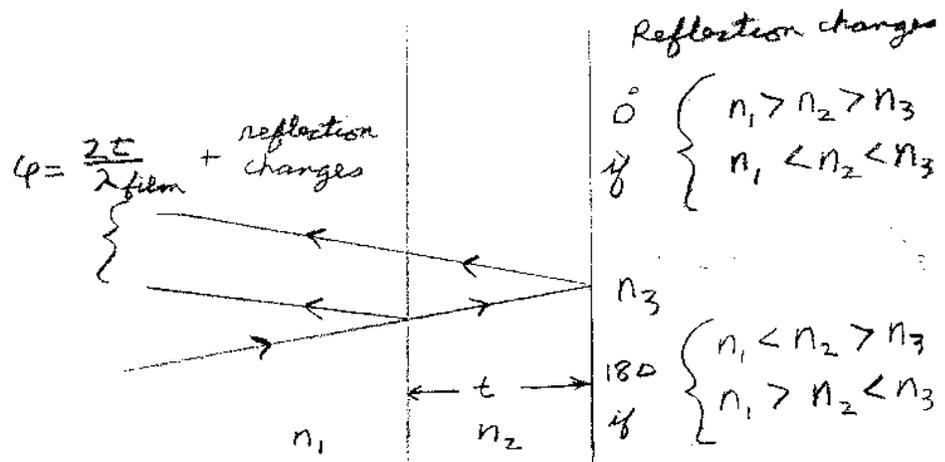
Taking the limit $N \rightarrow \infty$ yields

$$\begin{aligned} I &= (N^2 I_0) \frac{\sin^2(\Phi/2)}{(\Phi/2)^2} \\ &= I_C \frac{\sin^2(\Phi/2)}{(\Phi/2)^2} \end{aligned}$$

where I_C is the intensity of the central maximum, since I_0 is the intensity from one slit. We have used the approximation that $\lim_{N \rightarrow \infty} \sin(\Phi/N) \rightarrow \Phi/N$, since Φ/N is very small as $N \rightarrow \infty$.

Two slits with finite size

Using the results from single slit interference, we can understand a real two slit interference pattern. By real, we mean two slits that have a finite size. Suppose the screen consists of two slits, whose separation is d . Let w be the size of the slit



opening of each slit. Consider a scattering angle θ . As we just derived, the net amplitude for single slit interference at the angle theta is $E_0 \sin(\Phi/2)/(\Phi/2)$, where $\Phi = 2\pi w \sin(\theta)/\lambda$. The net amplitude for the superposition of both finite sized slits is thus

$$E_{net} = E_0 \frac{\sin(\Phi/2)}{(\Phi/2)} \cos(\omega t) + E_0 \frac{\sin(\Phi/2)}{(\Phi/2)} \cos(\omega t + \phi) \quad (46)$$

where $\phi = 2\pi d \sin(\theta)/\lambda$. Note that Φ is the phase difference between sources at the top and bottom of the slit opening. ϕ is the phase difference between the first slit and the second slit. ϕ is larger than Φ . In fact:

$$\frac{\phi}{\Phi} = \frac{d}{w} \quad (47)$$

The equation for E_{net} can be simplified by factoring out the term $\sin(\Phi/2)/(\Phi/2)$:

$$\begin{aligned} E_{net} &= E_0 \frac{\sin(\Phi/2)}{(\Phi/2)} \cos(\omega t) + E_0 \frac{\sin(\Phi/2)}{(\Phi/2)} \cos(\omega t + \phi) \\ &= E_0 \frac{\sin(\Phi/2)}{(\Phi/2)} (\cos(\omega t) + \cos(\omega t + \phi)) \\ &= 2E_0 \frac{\sin(\Phi/2)}{(\Phi/2)} \cos(\phi/2) \cos(\omega t + \phi/2) \end{aligned}$$

So we see that the net interference amplitude is the product of the single slit times the double slit amplitudes. Since the single slit amplitude varies slower than the double slit amplitude, the single slit amplitude squared is an envelope for the double slit pattern. The single slit has its first minimum when $\Phi = 2\pi$. As $\Phi \rightarrow 2\pi$, ϕ passes through $2\pi d/w$ times. Thus, the number of bright spots in the central region is $2d/w - 1$.

Thin Film Interference

Another way interference can occur is when light is reflected off a thin film of transparent material. We show this in the figure. Suppose light is incident near the normal of a thin film. Light can reflect from the near surface of the film, and it can reflect from the far surface of the film. The reflected radiation will be a linear superposition of the radiation reflected from the near plus the far surfaces. If we let t be the thickness of the thin film, then the extra distance that the light reflected from the far surface travels is $2t$:

$$l = 2t \tag{48}$$

So, the difference in phase of the light that reflects from the near versus the far surface is $\phi = l/\lambda_{film} = 2t/\lambda_{film}$. There is one other property of reflection to consider: when light reflects off of a "slower" medium, the reflected radiation is shifted 180° in phase. That is, if light reflects from a medium with $n = n_1$ off a medium with $n = n_2$, then the phase of the reflected radiation is shifted 180° if $n_2 > n_1$. If $n_2 < n_1$, then there is no phase change in the reflected light. We will demonstrate this phenomena in lecture. Let n_1 , n_2 , and n_3 be the indices of refraction for the media in the diagram. Then zero relative phase due to the surfaces will occur if $n_1 < n_2 < n_3$ or $n_1 > n_2 > n_3$. A relative phase change of π will occur is $n_1 < n_2 > n_3$ or $n_1 > n_2 < n_3$.

To summarize: upon reflection from a thin transparent film, the relative phase ϕ between the radiation reflected between the near and far surfaces is

$$\phi = \frac{2t}{\lambda_{film}} + (\text{any additional phase from surfaces}) \tag{49}$$

The wavelength λ_{film} is the wavelength of the light in the medium of the film. The wavelength in the film is λ_0/n_{film} , where λ_0 is the wavelength in vacuum and n_{film} is the index of refraction of the film.

Other Interference Phenomena

An important apparatus that makes use of interference effects is the Michelson interferometer. A monochromatic light source (usually a laser) is incident upon a beam splitter. Half of the light travels through the beam splitter, and half is reflected 90° . Both halves reflect off of mirrors, and return to the beam splitter. One beam travels straight through and the other reflects 90° and they both emerge together. Interference can occur between the two split beams when they "recombine". The phase difference ϕ between the two beams will be

$$\phi = 2\pi \frac{l}{\lambda} \tag{50}$$

where l is the difference in path length of the two beams. If one mirror is a distance d_1 and the other mirror a distance d_2 from the beam splitter, then $l = 2|d_1 - d_2|$. By changing one of the distances d_1 or d_2 , one observes the interference pattern shift. Everytime a new interference fringe is produced, l has changed by one wavelength.

Final thoughts on interference effects

Interference is a remarkable phenomena. Water waves and sound waves (mechanical waves) experience interference when the wave amplitude from different sources combine. Guided by mechanical wave interference, we have explained the double slit, N-slit, single slit interference patterns on the wall by summing an amplitude from each of the individual slits, then squaring the magnitude of the summed (net) amplitude. In the Maxwell or classical picture of electromagnetic radiation, the amplitude represents an electric field vector. The radiation is a sinusoidal wave, with the electric vector field doing the "waving". This classical picture explains many things, but as you will see next quarter one of its failures is the description of what happens when the light is very dim.

As the intensity is reduced, at first the brightness is also reduced. However, at very low intensity the radiation arrives as particles. The particles, or photons, arrive more often in the "bright" regions and never in the dark regions. The modern view is the following. Nothing physical is waving. An amplitude is assigned to each path that the photon can take to get to the wall. After summing up the amplitudes for all possible paths, we take the absolute square of the sum. The interference properties come from *summing the amplitudes*. The absolute square of this sum is the probability (or probability density) that the photon will arrive at the particular region on the wall. Instead of corresponding to an electric field magnitude, the sinusoidal functions represent the probability amplitude (density) for a "free" photon (or particle) to be in a particular location.

Interference effects can be understood without anything waving. The pattern on the wall is also determined from the Fourier transform of the slit openings. The Fourier transform employs the term e^{ikx} , which does not require anything to wave. It took physicists many years to sort these ideas out, and they form the basis of Quantum Mechanics. Next quarter you will cover the experiments and reasoning that occurred from 1900-1930 and lead to our modern formalism.

Electromagnetism and its radiation has taught us many lessons. So far it has taught us about the interaction of charged particles and light. Next, it will teach us about the properties of space and time. We conclude this course with Einstein's theory of special relativity.