The Complex Propagation of Light Explained Visually: How to Make a Hologram

Bruno Ray Becher

Bard College

Follow this and additional works at: https://digitalcommons.bard.edu/senproj_s2021

Part of the Atomic, Molecular and Optical Physics Commons, Geometry and Topology Commons, Harmonic Analysis and Representation Commons, and the Optics Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.

Recommended Citation

https://digitalcommons.bard.edu/senproj_s2021/80

This Open Access is brought to you for free and open access by the Bard Undergraduate Senior Projects at Bard Digital Commons. It has been accepted for inclusion in Senior Projects Spring 2021 by an authorized administrator of Bard Digital Commons. For more information, please contact digitalcommons@bard.edu.
The Complex Propagation of Light Explained Visually:
How to Make a Hologram

Senior Projected Submitted to The Division of Science,
Mathematics, and Computing of Bard College

By
Bruno Ray Becher

Annandale-on-Hudson, New York
May, 2021
Acknowledgements

Thank you to my parents Andrea and Max, for being models of perseverance and compassion.

-I would like to thank all of my physics professors at Bard for their constant support, attention, and wisdom throughout my time here. A special thank you to Antonios Kontos for teaching me how to learn, while being a friend. You've played a large part in shaping the way that I think.

-Sam Glass, and Luca.

-Maxamiano, for being a humorous companion, bandmate, editor, and hobbit.

-Richard Murphy, thank you for being the inspirational engineer that you are and always will be.

-Chris Lafratta! You have been an outstanding senior project advisor. It is rare that one finds a professor with such similar interests. I hope to be like you one day.

-Lo you are the magnetism to my electricity, I am the luckiest man alive to have found you. Thank you for the infinite love and sunlight you radiate.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>6</td>
</tr>
<tr>
<td>Ray Optics</td>
<td>9</td>
</tr>
<tr>
<td>Intro to waves</td>
<td>17</td>
</tr>
<tr>
<td>Wave mechanics</td>
<td>30</td>
</tr>
<tr>
<td>Diffraction</td>
<td>48</td>
</tr>
<tr>
<td>Intro to fourier transform</td>
<td>59</td>
</tr>
<tr>
<td>Fourier transform in two dimensions</td>
<td>64</td>
</tr>
<tr>
<td>Optical Correlator</td>
<td>70</td>
</tr>
<tr>
<td>Digital Holograms</td>
<td>80</td>
</tr>
<tr>
<td>Two-photon polymerization</td>
<td>97</td>
</tr>
<tr>
<td>Bibliography</td>
<td>99</td>
</tr>
</tbody>
</table>
Abstract

The complexity of light’s wave nature is shown in the paths that light takes. In this project I will show several useful ways to imagine and predict how light will travel from one place to another. Once light is produced it does not immediately fill a room, instead it undulates through free space as if the space itself was a fluid. Once we understand the way light flows and interacts with its environment not only can we predict and control its shape with a hologram, but also discover clues which give secrets about where the light has been. Telescopes and microscopes reveal hidden packets of information, coded into the light itself, that would normally be lost to eyes. Using a tool called the Fourier Transform I will show how anyone can intuitively “feel out” what motivates light to move through space in its unique way.
Introduction

There are several levels with which you can understand light. In the following diagram each circle models experimental observations of light, getting closer to the truth as you move inward by adding new explanations as well as explaining the observations in the circles it encompasses.

Though I won’t provide an in-depth explanation of the electromagnetic essence of light, or even light’s quantum mechanical behavior, I aim to give a sturdy understanding of waves and tools we can use to describe them.

Background: [Light diagram]

Initially, light was explained through ray optics, a theory in which light traveled from a source to the eye in straight lines. This ray explanation of light helped to understand things like shadows, reflections, scattering, and even lensing but failed to explain more complicated questions.
In this theory, a shadow is the absence of those raylike particles of light as they get obstructed by an obstacle. Thus reflection and scattering are explained as the bouncing of rays at complementary angles. This makes flat surfaces appear shiny in small areas, as what you are really seeing is an image or images of the light source, and makes rough surfaces appear uniformly bright regardless of the angle at which you view it.
Lenses are still a strange case in this domain since it seems that the light is clearly bending as it exits the glass, but this too can be explained by ray optics.

The path of least time
Light will always travel along the path of least time, even if that is not the path of shortest distance. Light travels at different speeds through different materials, moving most quickly through a vacuum. This is the only place that the famous speed of light quantity “c”, is actually a true depiction of how fast light is traveling. In air it travels slightly slower, and in water even more so. Light travels through every material at a different speed and this quality of a medium is called its refractive index “n”. It is usually represented by a number between 1 and 2.

Here is the equation that describes a refractive index:

\[ n = \frac{c}{v} \]

Where c is the speed of light and v is the speed of light in that material. This makes n a ratio between the maximum speed of light, and the speed of light as it gets affected by a material.

In as few words as possible the factors that will determine an object’s refractive index are dependent on the fact that light has an electric field, and the way that the electric charges are distributed in the molecules that make up that material, as well as how packed together they are (density) will alter the lights electric field.

The refractive index of a vacuum is 1, and the refractive index of water is 1.333. Meaning that light travels 1.333 times slower in water than in empty space, which is quite a large difference. Glass has an even larger refractive index of 1.5, making it a great candidate for lenses.
If light travels through air and then glass, in order to follow the path of least time, it must minimize its travel through the glass and maximize its travel through the air. This will “cause” light to travel at the perfect angle that minimizes time between the two mediums.

Here’s an example:

Let’s say that you kicked a soccer ball into a swamp by accident. In order to get the ball you will have to walk on grass and then on some amount of mud that is between you and that ball. Assuming you walk slower through the mud than the grass, you will want to maximize grass walking time and minimize your mud walking time in order to get the ball as soon as possible. Often the shortest path between you and the ball will not be the fastest. This is why light doesn’t always travel in straight lines as it changes mediums. This bending of light as it leaves or enters a medium is known as refraction.
The angle at which light refracts depends on the different speeds of motion through the two mediums. This is called Snell's Law.

Snell’s law describes the angles that light will approach the surface of a medium with respect to a line normal to that surface, and the ratio between the speed in the two materials

\[
\frac{\sin \theta_2}{\sin \theta_1} = \frac{v_1}{v_2} = \frac{n_1}{n_2}
\]
I learned the most beautiful explanation of this from my grandmother before she passed away. She said that you can think of the light interacting with the lens as a tank with two wheels on either side. As it approaches a material that causes the tank to travel at a slower speed, if one wheel will be slowed down before the other, the angle of the tank will shift. Once both wheels are inside the material and experiencing no difference in speed, the tank travels straight ahead. So as a tank driving on grass approaches mud at an angle, its right wheel will hit the mud first and begin to slow the right side of the tank. The left side of the tank is now covering a larger distance than the right side since it is still driving on grass. This causes the tank to turn to the right in this example.

As light hits the top of the lens, the bottom side of the light ray slows before the top, pitching the light downward. Similarly at the bottom, the top most part of the ‘ray’ will lag behind the ‘ray’ at the bottom, which tilts the light up. Exactly in the center both ‘rays’ experience the change in speed at the same time, so that the light continues on without delay. All of these bent rays converge in the center at some distance away from the lens called the focal length.
Air has a refractive index of 1.0003, so light travels through it at nearly the true speed of light. Though still capable of bending the light nonetheless. Heat lines and mirages can be explained by refraction as well. When air gets colder it has a higher density which can cause abrupt changes in the refractive index of the air and a scattering of light rays traversing through it. This is because a stark difference between materials with unique refractive indices such as hot and cold air, such as that rising off a hot road into colder air, bends the light travelling through it. So to an observer, looking at the bent light rays above the surface of the road, they will really be seeing the bent light rays from the horizon which bend directly into your eye without bouncing off the road. Leading to the appearance of a bright puddle on the ground.

When you look more deeply you will notice that light is not traveling as rays. For example if you look at the shadow of your hand on the ground you will see that it has defined edges. You can tell precisely where the border of the shadow starts and stops. But if you move your hand farther away from the ground you can see that the edges of your shadow become more and more ambiguous, tainted by blur as light creeps around the edges of your hand. From this we can say that
light seems to be bending, even without a difference in medium. Just the presence of your hand seems to alter the direction of the light.

Everything became a bit more clear in 1801 when Thomas Young created his now famous Double-Slit Experiment which shed light on the situation. Light was projected through 2 small holes in a dark empty box. He then peered into the box to see how the light rays were acting.
If light acted as a ray that traveled from source to destination then one might expect to see two bright areas beyond the holes where the light passed through and carried on forward, perhaps with a bit of distortion caused by the light bending or expanding after passing through the holes.
Instead what we see is this:

There is a bright area in the center, where there was no hole at all, and slightly diminishing bright spots symmetrically on either side. This result is puzzling, but points to a behavior that science has noticed in other parts of nature.
It seems that the light traveling into our eyes is acting exactly like a disturbance on the surface of a pond. A light source creates a light wave that oscillates up and down as it travels outward. Where we might see light, and how bright it is at that location is determined by the height or amplitude of a wave.

In mathematics a wave is often represented as a sine function.

\[ Sin(\theta) \]

For a given position, it will give a height of the wave.

Sine and cosine functions are generally taught in trigonometry, usually relating to the height or length of a right triangle. The sine of an angle gives the opposite side (height) divided by the hypotenuse. If we fix the hypotenuse to a constant value we will call \( r \), and look at it for any angle \( \theta \), the triangle will now draw out a circle of radius \( r \). The angle \( \theta \) corresponds to a specific point on the circle, while the sin of that angle gives you the height of that point.
Now for the reason we use sine to describe a wave:

If you increase the value of theta at a constant rate, the point will spin around the circle over and over again. If you look at the sine of that angle, the height of the point on the circle will draw out a perfect wave as the circle travels forward and the point loops around and around the perimeter of the circle.

This is one extremely useful way of seeing why a sine function might draw a wave.

The radius of the circle is the amplitude of the wave and the rate at which it spins about the center is its frequency.
Frequency is a quantity that describes the speed of the oscillations as the wave travels through space. As it moves, each revolution corresponds to a distance. The distance between two peaks will be how long it takes for one full revolution of the dot on the circle, also called a wavelength.

The way I like to think of this is that a stationary thing can have a frequency but cannot have a wavelength. Once you move that thing at a constant speed it has a frequency and a wavelength.

Imagine you have two clocks and one of them breaks such that it is now spinning twice as fast. This means that one of the clocks has a higher frequency than the other because it is completing oscillations faster. If you move them both through space at the same speed they will each have a wavelength that corresponds to their frequency. A higher frequency will correspond to a smaller wavelength and a lower frequency will correspond to a longer wavelength. Both of which are right several times a day.

The wavelength and frequency are tied to each other by the speed of the wave which is determined by the material the wave is going through and never changes. This means that in general
you should never expect the frequency of a wave to change, but do note that the amplitude will change as it encounters and phases through other waves.

Here is a general equation of a sine wave:

\[ Asin(\theta) \]

\[ \theta = w * t \]

The A represents the amplitude, and radius of a circle drawn by a spinning point. Angular frequency is represented by w, and the t is increasing time.

If you know the velocity and position of a particle in 3d space in (x,y,z) coordinates, you can know everything there is to know about that particle. How it will behave, where it came from and where it is going.

Since it doesn’t make sense to quantify the “position of a wave”, and the speed of a wave is fixed, we use its wavelength, amplitude, and phase to describe it fully.
PHASE

A unique quality that is present in light is that two waves can interfere with each other constructively or destructively. If two sound waves have the same frequency but opposing amplitudes (one is peaking while the other is troughing) they will destructively interfere, making zero sound. This is how noise canceling headphones work, they listen for a sound and produce an inverse wave (180 degrees out of phase) that cancels it out at your eardrum.

If two waves are produced at the same time and place such that they are in phase, but one wave travels a slightly longer distance than the other then the two of them will now have a small offset in when they are peaking and troughing.
Because one of them is peaking before the other, it’s intensity will decrease since it has to lift the other wave up. Due to this, the amplitude of the combination of the waves will be lower than it was when the two were aligned perfectly. This feature of a wave’s position relative to another wave’s is called phase. Represented by the letter ϕ, the phase offset between two waves is only important to understanding light in that it affects the amplitude of the waves while they oscillate on top of each other at a point at which we are observing the brightness. This makes for a brightness that is dependent on the relative phase of the two.

We can make a connection back to circles here because the extra lag between two oscillations can be represented as an angle offset in the lines from the center of a circle to two points on its perimeter. As you spin the circle they will draw out two waves with that phase offset ‘phi’ One will perpetually chase the other.

If we were to observe or measure the two waves as they passed through each other we would notice that they simply make a third sine wave, the amplitude of which is dependent on the phase of the composite waves that make it up.

Note that the phase of a wave is purely relational, meaning that phase describes the difference between two waves. A single wave only has phase when it is compared to another wave.

At 0 degrees the waves are aligned, making for a maximum brightness. At 180 degrees the two waves are perfectly anti-aligned such that they cancel out. When you go farther out in phase past 180 degrees you actually become more in phase. This is because on a circle there is no distinction between 0 and 360 degrees, they both describe the same point. Similarly if you start at 0 degrees and
go around the circle as many times as you want, it looks the same as 0 degrees and is equivalent to
you not changing anything in your position at all. This is a very strange quirk of waves.

If n is an integer number, then:

\[ \sin(\theta) = \sin(n \cdot \theta) \]

Where \( n = . 0, 1, 2, 3, 4.. \)

Another useful way to represent a wave is with this lovely expression:

\[ e^{i\theta} \]

This is Euler’s equation and it is incredibly useful once we actually start doing calculations to
predict amplitudes of waves. This ends up holding information about the amplitude while describing
the phase of the wave at the same time.

The other quality about a wave which is not intuitive is that although waves interfere as they
pass through one another, the moment that they leave each other’s presence they each carry on as
they were before, as if they didn’t interact at all. They have no push on the other wave and lose or
gain no energy, they pass through each other unscathed. This is different from particles where two
beams of particles would collide and deflect, while two laser beams pointed through each other have
no effect on the individual beams after the intersection. If you measure them there would be no way
of knowing it ever encountered anything at all.
A wave can bounce off of a mirror just like a ray would, and it can be blocked by an object to produce a shadow as well, while the curving around objects is something to do with a wave behavior known as diffraction.

**DIFFRACTION**

Diffraction is when a wave seems to bend or diffract after being constricted. This only applies to a wave in 2 dimensions.

**Waves in two dimensions!**

A good first example of a two dimensional wave is a plane wave. This is like the waves produced by a wave machine at an amusement park. If something wide and flat pushes the water back and forth it will produce flat waves that extend outward, the front of this wave makes up a line in 2 dimensions and a plane in 3 which is why it’s called a plane wave. This wave is really just a stack of 1 dimensional waves that we have already seen. These are nice to use since they travel in neat straight paths.

A useful tool for imagining why waves bend around obstacles is “Huygens principle”. It allows you to model the front of a wave as being made up of many point sources. All of these point sources radiate their own waves out in every direction, and the combination of all of these waves creates a wavefront.
Imagine you were to shrink down the sun to an infinitely small point, this theoretical object would emanate light in a perfect sphere. This is a point source.

The moment you turn on a flashlight, a collection of waves emanates out and fills the room. If you were fast enough would see the earliest part of this wave expanding into this empty space, like galloping horses. The surface of the expanding light is called a wavefront.

The shape of a wave front will define how the wave moves.

A wave front can be asynchronous if it was emitted from an uneven surface, disturbed while passing through an opaque shape, or bent/delayed by a lens.

Wave explanation for lensing:

As a plane light wave like that of a laser beam or sunlight hits a lens, the waves in the center slow first.
Immediately after the lens (and assuming the lens is very thin such that the universal phase delay caused by the width of the lens overall after the curved section is minimal) the stack of linear waves that made up that plane wave will have a concave shape due to the nonuniform delaying face of the glass.

The top and bottom are ahead of the middle and the ends of all these linear waves now make up a new wavefront. From Huygen's principle we can predict what this waveform will do next. If we were to line this new curved wavefront with many point sources emitting circular waves, the addition of all of these circular wavefronts can predict the shape and movement of the original waveform. These Huygens particles emit waves in all directions, but since each point source is surrounded by more point sources on either side, a cancelation occurs and only the section of the wave heading off perpendicular from the surface of the wavefront will remain.
These add up to make a new wavefront which heads towards a point. It is only at the focal point that the wavefront becomes flat again, then from a small flat surface of Huygens particles, a convex wavefront surface emerges after the focal point. This convex surface will produce a wavefront that continues to grow.

Why does Huygens principle work? This has to do with the reason a wave experiences a delay in a material in the first place. As a light wave encounters the surface of glass or water, it does not immediately pass through. Instead it becomes completely absorbed by the molecules in that substance. The electrons in the molecules will then hold on to that energy by oscillating at that frequency, and as it turns out an oscillating electron will produce an oscillating electric field which is exactly the thing that produces light. Each molecule in a material absorbs the electromagnetic wave completely in its charges, and then re-radiates it. This means that after light interacts with any material, it transforms that material into a symphony of point sources which re-radiate that light. This whole process takes some finite amount of time which is why the light appears to slow down. In reality the light is always traveling at the speed of light “c” in between molecules, but the absorption stalls its progress.
The strange part about this is that light acts like it is being produced by a combination of point sources even as it travels through a vacuum. Even without a material to re-radiate the light wave from many molecules, the wave still travels in free space such that you can predict its motion by assuming it is constantly being propagated by stationary points producing spherical waves of its frequency at its wavefront.

A shadow being blurred at its edge can also be explained by Huygen’s principle.

As a plane wave front collides with an obstacle, the point source waves in the center get absorbed while the sources on either side freely emit circular waves. These mostly add up to waves going straight ahead, but the point sources near the obstacle do not have waves on both sides to cancel out, and will send some portion of their circular wave at an angle into the area behind the obstacle. This bending of light around obstacles is called diffraction, while a bending from a phase delay inside of a material is called refraction.

Huygens principle can visually and mathematically describe how these waves will travel by showing that a wavefront acts exactly like an infinite amount of point sources lined up with each of the points emanating its own point source wave.
Most wavefronts are not perfectly flat, and a wavefront with any curvature will curve as it travels. A flashlight has a bowed wavefront due to its source being a small area and not a large plane of light emitting particles and so the light wave extends as a cone and not a cylinder.

The best example of a flat wavefront is the light from the sun. This is because if you take an object emitting a circular wavefront and get far enough away from it, you will notice that, in a small section, it appears flat.

Now that we have some math representation for this thing let’s describe a point source of light.

The field produced by a point source of light should oscillate up and down as it travels outward just like our linear wave, but instead of oscillating with an increase in one direction we want it to vary sinusoidally with an increase in any direction radially from a point.
\[ Ae^{-ikr} \text{ or } Asin(kr) \]

“r” is the distance from the point source to wherever you want to know information about the amplitude of the wave. Named “r” for radius.

If you were to plug in an increasing r into this function you would get back a value that ascends and falls over and over again. Really we should account for the wave getting weaker the farther it goes, as this is our reality. Since a wave is made of energy which is a finite resource, as it covers more area the energy in that area must get less and less intense. Meaning from a longer distance away we should be seeing a smaller output for the whole function.

\[ \frac{e^{-ikr}}{r} \text{ or } \frac{Asin(kr)}{r} \]

This is what our point source should look like. One dependence on r for its revolutions/oscillations, and one inverse dependence where it gets weaker per a larger r.

If we were to shine light into an obstructing object what might that look like with huygens principle defining the shape of the wave. As the point sources on either side of the object radiate, a few of the huygens particles (the point sources of light emanating circular waves) are now missing in the middle. The remaining point sources radiate and the sum of all of their wavefronts creates a slight bend in towards the shadow of the object. The smallest the features are with respect to the wavelength of light, the more pronounced these effects are.

Now that we understand interference, let’s look back at the Double-Slit Experiment, the experiment that changed the world and explain why this pattern is made by light.

A flat or plane wavefront is sent towards a wall and dissipates besides at two locations where there are slits cut in the wall of the box.
Let’s take these small slits in the box to be infinitely small. As there are an infinite amount of huygens sources along the wavefront, but two infinitely small slits we can model the wavefront as it passes through the two slits as just letting two huygen particles through. This makes for two circularly emanating waves of the same wavelength being emitted by two point sources. These two waves propagate inside the box and interfere with each other. A bright area on the back wall is formed when the distance from that spot on the wall to one of the holes matches the distance from the same spot to the other hole plus any integer number times that wavelength. Meaning there will be one bright spot where they are lined up exactly (at the center) and other bright spots where one wave from one slit travels an extra distance which amounts to an integer number of wavelengths more or less than the other.
Here is a nice geometric proof to determine the path length difference for waves after they pass through the two slits and meet at a common point on the back wall. When the screen is a large distance away compared to the distance between the slits ‘d’, then the angle ($\theta$) between the path of a line from the slits to the screen (see the picture) can be approximated as being the same for both of the waves. This is the paraxial approximation. The difference between the waves is shown by trigonometry to be $d \sin \theta$. The angle $\theta$ on the black line from the wall is the same as the angle from the red line to the horizontal, for the following reason. Let’s call the top theta the red theta and the angle from the black line to the blue wall the black theta, The angle in between the two of them (not marked on this diagram) we will call dotted theta. We know that the red theta plus dotted theta adds up to ninety degrees, since the black line is made to be perpendicular to the red lines. The dotted theta plus the black theta is also ninety degrees since the dotted line and blue wall are perpendicular by definition. Since red theta and black theta share a common angle - dotted theta, and they each add up to ninety degrees. This means that red theta and black theta are really the same angle.
Where \( d \) is the distance between the two holes. For constructive interference from two slits, the path length difference \( \Delta L \) must be an integer multiple of the wavelength \( n \).

Where \( n \) is any integer number positive or negative.

\[
\Delta L = n \lambda
\]

Then in order to see a bright spot on the back wall we should be looking for

\[
\Delta L = n \lambda
\]

Where the difference in path length is given by \( dsin\theta \) from our trig proof.

So constructive interference happens at

\[
ds in \theta = n\lambda
\]

For destructive interference from two holes or slits, the path length difference must be an integer multiple of half of a wavelength.

Or

\[
ds in \theta = n * (\lambda/2), \text{ for } n = -2, -1, 0, 1, 2..
\]

This will give a perfectly dark spot.

We call \( n \) the order of the interference. For example, \( n=3 \) is the third-order interference.

The symmetry about the central bright spot is explained by the fact that \( n \) can be a positive or negative integer. After all, light waves diffract up just as much as they diffract down.
For a fixed $\lambda$ and $n$, the smaller the distance is between the two holes $d$, the larger the $\theta$'s that the bright and dark spots will be at. Since

$$\sin \theta = \frac{n\lambda}{d}$$

(not necessarily brighter, but it is hard to scale the wave in only one direction in this illustration software.)
This fits with the previous statement that as features get smaller and smaller, their wave effects become more noticeable.

Double slit image which I captured with my 4F Correlator setup

The intensity of light can be affected by traveling different distances in space, and that specific shapes and patterns constricting a wavefront, like the two slits, can yield a unique set of paths in space which will create a unique image.

Imagine that instead of two slits on our box wall we have an arbitrary number of slits in our wall at random locations. This will create many wavefronts and thus a jungle of interference inside of the box. On the back wall you will now see a complicated and unique interference pattern projected where all the wavelets collide, compare phase, and decide a brightness.
Now what if I told you that with the right dot placements you could project any pattern you want by predicting the interference patterns on the back wall. It’s not too difficult to imagine that you could do this in two dimensions, where you would place holes up and across the wall vertically and horizontally, and that you could produce 2d patterns, of images, by only using holes. This is exactly what I intend to do by the end of this project.

A hologram is a surface that holds a pattern which controls the interference of a wavefront at a screen a distance ‘d’ away from the hologram. This allows you to form a drawing with light, by predicting where all of the light waves will be in their revolution.

An interesting note about holograms is that because they depend on the interference of all the wavelets from all the different holes in the material they have some weird properties. Every part of the image that a hologram projects is stored in every location of the hologram. This makes it so that if you have a hologram that projects an image of a person and you were to physically break the hologram into two halves, each half would still independently project a whole image of a person. The information is encoded in the relationships between each hole and every other hole. The image is encoded at every point in the hologram. The more points on the hologram you use the better the resolution of your output image. There is no center to the hologram. If you were to shine a laser through it at any location it would still project the same image even if the laser misses many of the holes.

How might one predict where best to put these holes? This is where the Fourier transform comes in. It describes how light moves through anything (including holograms), can predict how a
vibration on a string will evolve over time, the way heat moves through a surface, and the notes in a piece of music.

In the mathematical toolbox the Fourier transform should be the screwdriver, because it’s elegant, well known, and classic.

Introduction to the Fourier Transform

Below is the Fourier transform equation

\[ F(w) = \int_{-\infty}^{\infty} f(t)e^{-iwt} \, dt \]  \hspace{1cm} (1)

As well as the inverse Fourier transform.

\[ f(t) = \frac{1}{2} \int_{-\infty}^{\infty} f(w)e^{iwt} \, dw \]  \hspace{1cm} (2)

In short, this Fourier transform equation can take a function, and tell you what frequency waves are inside of that function.

So for example the Fourier transform of a single sine wave will return a function that peaks and exists at one value only, that value is the frequency of your wave.
This function which peaks at one value, called the dirac delta function, exists only at a single point and is a peculiar thing. It is a function that has an infinite height, an infinitely small width and an area of one. To envision how this thing is made mathematically without breaking any rules, imagine that you started with a square with side lengths of one, and increased the height while shrinking the width until its height became infinity. This incredibly thin spike essentially has the purpose of representing a single number rather than a continuum of values. The Dirac Delta function is rarely seen alone though, often it is found in integrals, since an integral will tell you how much area exists under a function, it yields an area of one at a precise location along the number line. The most common way to use this thing is to multiply the delta function by a function you wish to probe, and then integrate this product. It may sound like a cheat, but the delta function will be zero at every single value aside from one, so if you want to know what your function is at some value you can just move the delta function to that particular point where it sits pretty giving you an output of one. Now when you cycle through the integral the integrand gives zero until you hit the delta function where it multiplies your function by one, this integral will give you the value of the function at the point where the delta function lies. As a Bard physics student once put it, “it is the function slayer!”

Back to the transform-
If you take the inverse fourier transform of a delta function which lies at a certain point it will give you a single wave of that frequency back. With the frequency of this wave corresponding to how far up the number line the delta function was sitting. The Fourier transform takes a function that has axes corresponding to distance, and carries it into a different graph which has a horizontal axis that tells you the frequency, and a vertical axis which represents how much of that frequency does the original function contain.

If you had a function which was the sum of two waves of two different frequencies and then took the fourier transform of that function, it would be able to decode and give you back the original frequencies which made it up. Not only does it tell you what frequency waves are present, but it tells you exactly how much each frequency contributes to the function, giving back the amplitudes of all of the waves that summed up to make that function.

Ukulele examples
Let’s imagine this as a physical situation. If you have an instrument that plays a chord, like a ukulele, you can imagine after it is strummed each of the strings oscillate at different frequencies due to different amounts of tension on each string which pulls the strings back to their original position at different speeds. As the strings swing, they slosh the air around them, pushing and pulling the air particles. When the string swings forward, it packs the air in front of it such that it becomes a high pressure area, from which the air dissipates into the low pressure areas around it. This dissipation fills the area around those particles with a new, higher density which will then, itself dissipates and so on. Similarly when the string goes in the opposite direction it evacuates that area of air, causing the nearby air particles to rush in to fill in that empty zone. The new low pressure ring will then be filled in by the particles around them and so on. These high and low pressure rings that travel radially away from the string are the sound waves that eventually push and pull on our eardrum creating electrical signals that are sent to our brain.
When there are several strings pushing and pulling the air particles at the same time one might think that the pressure waves in the air from one string might make it harder for the other to manipulate the air particles at its own frequency, but this is not the case. Since if a string pushes into a volume of air that is already packed with high pressure it will feel more resistance as it swings through, but will be pushed in the direction of its motion on the way back, getting a boost in speed. This means that it will just add the pressure from the two string impulses without affecting what either of the strings would have been doing independently. This is thanks to the fact that two waves will not permanently have an effect on each other. Instead, while they pass through each other they will add up their pressures for a moment and continue on like nothing happened. When you strum a ukulele, four sets of pressure rings are created at the same time. Since there are four different rates of pushing and pulling on your ear drum, your ear drum will move in a chaotic way that is the sum of all the forces of the pressure waves.
The movement of your ear drum or a microphone’s membrane is now a function that tells you what the net force of all of the sound waves are over time. This net force is different from the forces of each wave, so how are we able to enjoy music if we hear all the notes at the same time? What comes to our ear is a garbled wave, but like a fourier transform our ear is able to pick through and hear a frequency spectrum within it.

A microphone that records a ukulele chord, can then give that electrical signal of the net force in time to a speaker. This speaker is just one flat membrane that pushes and pulls air in front of it, so how is a speaker able to create four notes at the same time? It turns out that the function of
our eardrum that represents four notes added up on one flat area played out from the speaker, is identical to playing four notes at once. So essentially you can use one flat membrane or string, to produce the sounds of 4, 5 or an infinite number of strings or membranes all playing different frequencies.

This chaotic net force function can be drawn purely by adding frequencies of perfect sin waves, and it turns out that it is no exception. You can actually draw any function and any shape with a combination of sine waves.

My question to this originally was, “besides the case where we add up several frequencies to make a function, why would there be any frequencies in a function?” For example, does f= (t)^2 have frequencies associated with it?

The answer is yes.

Much like the basis elements x and y can be combined to draw out any point in two-dimensions. You can use frequencies as another set of basis elements to form a linear combination and draw out any function. This basis is an infinite set of sine and cosine waves of frequencies from zero to infinity, and by adding up a scaled combination of these you can draw any time dependent function that could exist.

How does this equation (1.) do this?

If you have a bag of blue and red marbles and you want to know the percent of red marbles in the bag, you could take the total number of marbles and divide by the number of red marbles. This would give you how much of the bag is composed of red marbles.
So, if we have a function and we are curious how much of the function is composed of a particular frequency, let’s say 120 Hz, could we simply divide the function by that frequency? Let’s see how we would go about doing this...

If we have a function that varies in time \( f(t) \), how would we divide it by 120 Hz to see how much of that frequency is in this function \( f(t) \)?

\[
\frac{f(t)}{120Hz} = F(t) \tag{3}
\]

Let’s divide \( f(t) \) by a 120 hz wave and see if that works.

The way we represent a frequency is through Euler’s equation. Where theta is our frequency.

Using \( e^{i\theta} \) will allow us to hold information in the complex plane.

\[
e^{i\theta} = \cos(\theta) + i\sin(\theta) \tag{4}
\]

\[
\frac{f(t)}{e^{i\theta}} = \text{fourier transform}
\]

Since we want to see how our function plays out in time, we should use an integral. This integral will basically cycle through time adding up heights of the function at every time, from \(-\infty\) to \(+\infty\).

\[
\int_{-\infty}^{\infty} \frac{f(t)}{e^{i\theta}} \, dt = F(t)
\]

Now a small trick:
We are integrating symmetrically from -infinity to infinity.

Since sin(theta) is an odd function, meaning that the right side of it is the inverse of the left side, we know that it will integrate to zero, as the areas of the right and left sides will cancel out.

\[ \int_{-\infty}^{\infty} \cos(\theta) + i\sin(\theta) \, d\theta = \int_{-\infty}^{\infty} \cos(\theta) \, d\theta \]  

(5)

So in this integral, \( e^{i\theta} \) is just equal to the \( \cos(\theta) \) without the \( i\sin(\theta) \)

\[ \int_{-\infty}^{\infty} e^{i\theta} \, d\theta = \int_{-\infty}^{\infty} \cos(\theta) \, d\theta \]  

(6)

How much of 120hz is in f(t)?

Now our wave which is a real cos wave of 120 theta can be represented in the integral as such.

\[ \frac{f(t)}{\cos(120t)} = \frac{f(t)}{e^{120t}} \]  

(7)

Since a negative exponential is equivalent to switching the denominator to the numerator or vice versa.
This is starting to look quite familiar as the fourier transform.

\[ F(w) = \int_{-\infty}^{\infty} f(t)e^{-iw_0t} dt \]

Now F(w) is giving you how much of a certain frequency w, you have in f(t).

All you need to do is divide f(t) by a wave of a particular frequency \( e^{i\omega_0 t} \). Then you no longer need to use an integral to cycle through time, you can just divide a function by a frequency to know how much your frequency is contributing to the overall function.

\[ \frac{f(t)}{e^{i\omega_0 t}} = F(w_0) \]

Now to get a full breakdown of all the frequencies and how much each contributes to the function, you measure it for all time.

\[ \int_{-\infty}^{\infty} f(t)e^{-iw_0t} = F(t) \]

Let’s go through a quick example to show how to actually perform the fourier transform.
In this demonstration we will compute a Fourier transform of a rectangular pulse. This is a function that has a constant height from \(-a\) to \(a\), and is zero everywhere else.

\[
x(t) = 1 \text{ when } -a < t < a \text{ and } f(t) = 0 \text{ when } t \leq -a \text{ and } t \geq a
\]

Our Fourier transform equation is

\[
F(w) = \int_{-\infty}^{\infty} x(t)e^{-iwt}dt
\]

And since our function only exists from \(-a\) to \(a\) we can change our bounds

\[
F(w) = \int_{-a}^{a} x(t)e^{-iwt}dt
\]

In this bound we know that our function is just one, so we will plug that in for \(x(t)\)

\[
F(w) = \int_{-a}^{a} e^{-iwt}dt
\]

Now this is a pretty straightforward integral

\[
F(w) = \frac{1}{-i} \left[ e^{iwa} - e^{-iwa} \right] \quad \text{← How to write this bound Chris?}
\]

Multiplying through by negative one yields

\[
F(w) = \frac{2}{i} \left[ -e^{iwa} + e^{-iwa} \right]
\]

Now we are going to make a weird leap and multiply the top and bottom by 2. The reason we do is this is because we are aiming for a certain form in our final equation. You will see what I mean in a second.

\[
F(w) = \frac{2}{i} \left[ e^{iwa} - e^{-iwa} \right]
\]

This resembles Euler's relation for a sine wave:
where,

\[
\frac{1}{2i} \left( e^{iwa} - e^{-iwa} \right) = Sin(wa)
\]

So now we can plug in a sine wave for this section in our equation. We will also multiply the top and bottom by \( a \).

\[
F(w) = \frac{2a}{wa} \cdot \frac{1}{2j} \left[ e^{iwa} - e^{-iwa} \right]
\]

\[
F(w) = 2a \frac{sin(wa)}{wa}
\]

This function actually comes up enough that it has its own name, that name is the sinc function.

Where in general

\[
sinc(x) = \frac{\sin(x)}{x}
\]

Finally we have this as our Fourier transform

\[
F(w) = 2a * sinc(wa)
\]

This plots the following graph as a function of frequency \( w \)

In optics this function shows itself often, as it turns out that the Fourier transform of an aperture will give you the amplitude of the light wave in the far field which traveled through that aperture.
The intensity of light is equal to its amplitude squared, so the following function would be the intensity of light after passing through a small pinhole.

Most of the light is concentrated in the center, but as you look closer you will notice oscillations in intensity symmetrically about the center.

This is the bullseye pattern that a clean gaussian laser beam will have, also called an airy disk. We will notice that a smaller aperture yields a larger spread, this is in agreement with the fact that as you get closer to the wavelength of light diffraction effects become larger.

Airy disk (left) and Fourier transform of a circle (right).
The next amazing attribute of the Fourier transform is that it can predict where and how a distortion will travel.

If you have a single string in a specific tension, and you slap the string with a force at a point and then soon after paused time such that the string is paused in a contorted position, the Fourier transform can tell you what its next position will be. It can do this without knowing how intense the force was or where it was initially distributed.

If we take our captured moment of the string at $t=1$ and give it to the Fourier transform it will simply tell us what frequencies will make up an equivalent version of that wave shape.
Now we have a set of frequencies but how does that help us predict the waves motion?

We know exactly how each of the sine and cosine waves will oscillate in time...

These waves all move in a predictable way: up and down. However, what if we were to take this set and use them to remake an equivalent replica of our disturbed string function that is frozen in time.

Now we hit play…

As every sine and cosine wave wobbles up and down in its predictable way, they add up to a sum of movement which describes the complex behavior of the string as it wriggles out the force of the initial slap. In this way, the fourier transform can tell the future of a disturbed wave without knowing any information about what previously happened to that wave.

Similarly the Fourier Transform can take the previous state of a light wave and tell you what it was doing in the past, or given a set of absorbing points/obstacles tell you how the light will diffract.
But what, you may ask, does the fourier transform have to do with a hologram? To answer that question let's look at what the pattern on the back wall of a box would be if light was entering from a single hole or an aperture. What would an equation that takes an aperture of a certain size and gives the amplitude of the wave when it hits the back wall look like?

**Fraunhofer Diffraction**

How does the fourier transform come up in diffraction?

There is a plane wavefront approaching a wall. This wall has a single slit in the center of it that is a distance of “a” across. The wavelength is $\lambda$ and the plane wave can be represented by $e^{ikz}$ where $z$ is the horizontal distance in the image. Once the wavefront hits the wall we can use Huygens principle. The previously mentioned way to predict a waves behavior, by approximating it as many point sources that spherically radiate wavefronts.

A spherical wavefront radiates as a function of $k$ and $r$. Its field is given by the equation below.
Since this slit is a length of “a” across we can integrate over an infinite amount of point sources from 0 to a to predict how this wave will propagate. This is Huygen’s principle

\[ \int_0^a \frac{e^{-ikr}}{r} \, dx \]  

(12)

To understand how the length of the slit affects the wave when it projects onto the screen, we will need to introduce a few more length variables. The distance from the wall to the screen in the z axis is called ‘d’ in this scenario. The distance from the start of the slit we will call x, with x=0 being the beginning of the slit. The radial length r will always start from a point x. The distance up the screen from x to the point where r contacts the screen is xs.
To find \( r \) as a function of \( d \), \( x_s \), and \( x \) we can use the pythagorean theorem.

The base of the triangle is \( d \) while the height is \( x_s - x \).

\[
 r = \sqrt{(x_s - x)^2 + d^2} \tag{13}
\]

This is quite the tricky integral to do.

Now we can make an approximation!

1.) **Paraxial approximation**

\[
 r = d \sqrt{1 + \frac{(x_s - x)^2}{d^2}} \tag{14}
\]

Let’s call this quantity on the right epsilon.

\[
 \epsilon = \frac{(x_s - x)^2}{d^2} \tag{15}
\]
We can expand this and approximate \( r \) to be

\[
    r = d\sqrt{1 + \epsilon}
\]

We can do this because \( x_s - x \) is very small compared to the distance \( d \). This lets \( \epsilon \) be very small.

Said another way, the angle associated with \( r \) from the \( z \) axis is very small due to the fact that we are projecting our image so far away.

This is the paraxial approximation.

Now we can integrate over the point sources from 0 to \( a \).

\[
    \int_{0}^{a} \frac{e^{ik(d + \frac{1}{2} \frac{(x_s - x)^2}{d})}}{(d + \frac{1}{2} \frac{(x_s - x)^2}{d})} \]

(*missing a \( k \) in the exponential on the top right)
But we still need another approximation to clean this equation up. This quantity in the denominator on the right is just epsilon.

Let’s say that epsilon is $0.005 \cdot d$. Then taking out this term and leaving the denominator as just $d$ will lead to an overall error of only $0.5\%$ in our answer.

$$
\int_0^a \frac{e^{-ikd} \cdot e^{-\frac{1}{2} \cdot \frac{(x_s-x)^2}{d^2}}}{d + \frac{1}{2} \cdot \frac{(x_s-x)^2}{d}} \, dx
$$

This lets us confidently take out epsilon from the denominator. Allowing us to factor it out of the integral.

$$
\frac{d - (d + d \cdot 0.005)}{(d + d \cdot 0.005)} = 0.004975 \cdot 100 = 0.4975
$$

So

$$
d \approx (d + \frac{1}{2} \cdot \frac{(x_s-x)^2}{d})
$$

We can also factor out the exponential on the left.

Our integral becomes the Fresnel integral.
With just one assumption. The paraxial approximation.

Now we can expand

$$ (x_s - x)^2 = x_s^2 - 2x_s x + x^2 $$

Next approximation

Now we are going to make the assumption that $a<<xs$ which is to say

$$ x << x_s $$

All this means is that the image on the screen is much larger than the aperture itself.

This makes the last term $x^2$ quite small.

Taking it out leaves us with this.

$$ (x_s - x)^2 \approx x_s^2 - 2x_s x $$

This approximation cannot be used all the time. For example, when dealing with an aperture function of a lens, the “aperture” is often very close to the same size as the next thing you are projecting onto (sometimes another lens of equal size).

Plugging in this for $(x_s-x)^2$ get’s us with the following equation.

$$ e^{-ikd} \frac{1}{d} \int_0^a e^{\frac{1}{2} \frac{ik}{d} (x_s-x)^2} \, dx $$

Now to deal with the $k$ vector.
K is no longer just in the z direction as the wavefront is no longer a plane wave after diffraction has taken place.

The K vector parallel to r is the hypotenuse of a right triangle whose base is Kz and whose height is Kx.

These vectors have the same proportions as the right triangle made by r, d, and Xs.

So we have an expression for kx when the wave hits a point xs up the screen:

$$k_x = \frac{k x_s}{d} \quad (27)$$

Rearranging this for k

$$k = \frac{d k_x}{x_s} \quad (28)$$

We can plug this in for k in the integral. If we want our slit to not just be in the center or to be multiple slits we will have to integrate over the whole wall. This gets our final equation.
Where \( g(x) \) is the “aperture function”.

This is where we get the term “Fourier Optics” from, because in the far field, or if we assume our diffraction is fraunhofer diffraction, then we get a Fourier transform of our aperture function \( g(x) \) on the screen.

\[
e^{ikd} \star e^{ik \frac{x^2}{d}} \int_{-\infty}^{\infty} g(x) e^{-ik_x x}
\]

Where \( g(x) \) is the “aperture function”.

In our first example, \( g(x) \) is a piecewise function that is 1 from 0 to a and 0 everywhere else. A value of 1 would represent total transmission while a value less than that would mean partial transmission. Of course the aperture function doesn’t need to be binary. If you were to model dirty glass, you could represent translucent areas as having any value between 0 and 1. This is very useful because you can take any complicated surface, do a fourier transform of it and find the diffraction pattern it would produce in the far field.

What is even more powerful about the aperture function is that we can introduce an amplitude and allow the function to let in all or some or none of the light in whatever pattern we want, while also adding a phase component.
Fourier transform in 2 dimensions with the 4f Corellator

In electronics and signal processing we often look at a signal which varies which might appear in the voltage of its output terminal at a fixed location in space, and Fourier transform it in order to get its frequency components. Where the frequency corresponds to how fast the signal is changing in time. In our spatial 2-dimensional Fourier transforms we will be looking at a fixed image at a fixed time, and checking for regularities or periodic patterns in the image. This will give us a set of oscillating waves which can be added up to make that image. To get a feel for how a 2d
Fourier transform might work we will be looking at an optical system that tangibly images a physical shape into a Fourier Transform on a camera.

This setup only requires that a shape or “cutout” be put in the path of a laser beam, and at the focal point of the beam a Fourier transform of that 2d shape is imaged.

4f correlator image

![Diagram of 4f correlator image](image)

Let's imagine what happens to a section of a collimated beam when a lens focuses it into a focal point. Then that beam expands to be focused into a collimated beam once more. After the beam hits the cutout a shape of light flies through the air in parallel lines. After getting focused and collimated the shape is flipped over. The top is now the bottom and the left is now right. This means that at some point along its path this light must have crossed over itself to flip to the other side. The point where the flip happens is the focal point. This is sort of strange though, because, since every part of the image on the top is transferring to the bottom and similarly with the left and right, this means that every part of the image is exactly in the center. Clearly you cannot go from left to right without going through the center, and due to the fact that this flip happens abruptly at the focal point this must mean that at the focal point everything in that image is
somehow stored in the center. If we were to place a cutout in a beam to produce an image and observe it at the focal point what would we see?

Well, we have made a crucial incorrect assumption. This is that the light near the borders of the cutout will travel in parallel lines along with the rest of the light. This is false, as we have seen, light going past an obstruction will diffract and will act as a small segment of a plane wave of an angle theta away from the optical axis. This light will then travel a distance f, and spread out according to its original diffraction angle such that when it reaches the lens it maps onto it with a position associated with its angle or ‘k vector’ after the cutout. The k vector describes the direction that the wavefront is oriented in. The lens will then focus or refract the plane wave which comes in at a particular angle and location to a unique point in the focal plane. This action of converting a plane wave of a certain angle to a point in a plane, is known as a spatial fourier transform, where frequency information is similar to angular distribution. This conversion of a wave to the location of a point is very much what the Fourier Transform is all about. A plane wave coming in at an angle will create a higher frequency periodic pattern on the surface it is projected on, than that same plane wave coming in at an angle perpendicular to that surface. This is one way of thinking about why low frequencies are in the middle of this 2-dimensional fourier transform. Waves at the center are coming in parallel to the surface, while waves coming in from the edges have a large angle with respect to the surface.

Proof of the Fourier Transform Property of the lens(from fundamentals of photonics book).
This proof takes the following four steps

1) A plane wave tilted at a specific angle in the x and y dimensions can be represented so
\[ \theta_x = \lambda v_x \text{ and } \theta_y = \lambda v_y \]

This wave will have a complex amplitude of

\[ U(x, y, d) = F(v_x, v_y) e^{-j2\pi(v_x x + v_y y)} \]

This is in the \( z = 0 \) plane.

At the \( z = d \) plane, Right before crossing the lens this wave becomes

\[ U(x, y, d) = H(v_x, v_y) F(v_x, v_y) e^{-j2\pi(v_x x + v_y y)} \]

Where

\[ H(v_x, v_y) = H_0 e^{2j\pi d(v_x^2 + v_y^2)} \]

Is the transfer function of a distance \( d \) of free space and

\[ H_0 = e^{-jkd} \]

2.) While crossing the lens, the wave is slowed down by the refractive index of the glass such that it picks up a phase lag. This is represented by the complex amplitude being multiplied by

the lens phase factor:

\[ e^{j\pi f(x^2 + y^2)} \]

and the phase overall phase factor of the lens with width \( \Delta \) is \( e^{-jk\Delta} \), and is ignored thanks to the thin lens approximation.

So now

\[ U(x, y, d + \Delta) = H_0 e^{j\pi f(x^2 + y^2)} * e^{2j\pi d(v_x^2 + v_y^2)} F(v_x, v_y) e^{-j2\pi(v_x x + v_y y)} \]

This expression can be simplified by writing

\[ -2v_x x + \frac{x^2}{\lambda f} = \frac{(x - 2v_x \lambda f)}{\lambda f} \]

As

\[ \frac{(x - x_0)^2 - x_0^2}{\lambda f}, \text{ with } x_0 = \lambda v_x f \]

With a similar expression for \( y \) written as
\[ y_0 = \lambda v_y f \]

This makes it so that

\[ U(x, y, d + \Delta) = A(v_x, v_y) e^{\frac{j\pi}{\lambda f} (x-x_0)^2 + (y-y_0)^2} \]

Where

\[ A(v_x, v_y) = H_0 e^{j\pi(\lambda(d-f)(v_x^2 + v_y^2))^2} F(v_x, v_y) \]

\( U_{xyd+\Delta} \) is the complex amplitude of a paraboloid wavefront wave converging toward the point \((x_0, y_0)\) in the lens focal plane.

3.) In order to find the wave at the focal plane where it is imaged we examine the propagation in the free space between the lens and output plane.

The complex wave at the focal plane \( U(x, y, d + \Delta + f) \) we use the relation

\[ \int e^{j2\pi(x-x_0)x'} dx' = \lambda f \delta(x - x_0), \text{ and obtain} \]

\[ U(x, y, d + \Delta + f) = h_0(\lambda f)^2 A(v_x, v_y) \delta(x - x_0) \delta(y - y_0) \]

Where

\[ h_0 = \frac{j}{\lambda f} e^{-j kf} \]

The plane wave is focused into a single point at \( x_0 = \lambda v_x f \) and \( y_0 = \lambda v_y f \)

Converting a wave to a point is precisely what the Fourier transform is. In two-dimensions that means mapping out a two-dimensional wave of a specific angle \( v \) onto a specific point in 2-D space.

4.) The final step is to integrate over all the plane waves \((all \ v_x \ and \ v_y)\).
From this we can do our first real test!

To see if this works I will calculate the fourier transform of a two dimensional shape in Matlab, and compare it to the optical system working on the same shape.

In order to make accurate cutouts I’ll be using a 3d printer with a tolerance of around +/- 0.01mm and designing shapes in CAD Fusion 360.

To suspend the cutouts in the collimated beam in a way that is stable, adjustable, and allows for easy switching between cutouts, I designed the following device:

It includes a neodymium magnet in the base so that I don’t need to fasten it down to the optics table for a firm connection. A threaded connection between the two pieces allows for
adjustment in height, while adjusting the position in the horizontal plane is as easy as sliding it across the table.

I left a slit at the top so that a circular cutout could be easily slid in, as well as removed.

One important change in this design in my second model, was to increase the diameter of the piece holding the cutout so that any peripheral beam which doesn’t interact with the cutout gets absorbed and doesn’t travel into the rest of the optical setup.

This setup is called 4f correlator because of the necessary focal distances to convert the image into a fourier transform and back again.

Note that in order to see the fourier transform you only need one collimated beam big enough to cover the cutout and one lens, really just a 1f setup. The lens and image plane after this is just to get back your original image.

The first lens and pinhole is used to clean up the quality of the beam from the Helium-Neon laser.

The second lens is used to expand the beam such that it will be larger than the cutout when it reaches the next lens.

This collimating lens curves the wave front of the beam to be flat and creates a 1” cylinder of light that doesn’t converge or diverge. To collimate the beam perfectly, I held a paper in front of the collimating lens and scored marks on the perimeter of the projected laser dot on the paper. Then
I followed the beam to the edge of the room (a distance of about 8 feet) and checked again how the dot diameter compared to its original size. If it was larger I would move the collimating lens forward so as to converge the beam, and if the dot was smaller at a farther distance I would move the lens backward.

Once the beam was collimated, I placed a second lens to focus the beam down to a point on the camera with an optical filter to dim the light and ensure that the camera didn’t get damaged by the high intensity spot. This focal point spot would be more of an oval at first, but after aligning all the lenses properly it appeared as a perfect circle at the focal point.

Then I placed the cutout apparatus 1 focal length before the focusing lens and observed the image captured by the camera.

Comparison between the captured image of cutout vs the computed fourier transform:
MATLAB code-

In order to take the theoretical fourier transform of a cutout to predict what I should see on the camera I used this code. If you input an image it will turn it into a matrix and return 2 images. It gives you the original image and the inverse Fourier Transform. I will be 3d printing every image on a 25mm circle with a depth of 2mm. Due to unavoidable error, these will come out with slight flaws in the way they look. In order to account for this in my fourier transformation code, I don't use the original image, instead I use an image of the cutout itself. To do this I collimated the light, sent it through the mask/cutout, and then sent it through a set of lenses to collimate it as a much smaller beam that can project onto the camera’s ccd screen. It is important to use an even amount of lenses
to do this so that they fourier transform the light back again into the original cutout image before it gets to the camera.

Some cutouts for the 4f system to transform

One downside to using cutouts for my images is that I cannot produce any image with floating points, for example a ring would be impossible to print in this way. To get around this I laser cut plexiglass circle templates and adjusted the 3d-printers print-height to produce images on this transparent plate. Here are some examples, the third is a 3d-printed hologram that didn't work.
Here is my highest quality cutout. It is the chinese character 光 guang(1) meaning light. The first image is the model in Fusion 360 while the second is the printed version.
Those lines have a width of 1mm. Here is its projection captured in the previously mentioned optics setup:
You can see some small differences between the digital 光 and the final 3d-printed 光 in the optical setup that this code will now take into account.

Now I load this image into MATLAB and run my code

Code:
clear all
close all
t = Tiff('guang.tif', 'r');
allimagedata = read(t);
imagedata = allimagedata(:,:,1);
%mimagedata = imagedata > 100;
firmagedata = abs(fftshift(fft2(imagedata))).^2;
figure(1)
imagesc(imagedata,[0 255])
%xlim([400 600])
title('X')

figure(2)
imagesc(firmagedata,[0 1e9])
xlim([300 1000])
ylim([150 800])
title('transform')
imagedata2 = abs(ifftshift(ifft2(firmagedata))).^2;
figure(3)
imagesc(imagedata2,[0 1e8])
title('transform back')

%inverse of original
imagedata3 = abs(ifftshift(ifft2(imagedata))).^2;
figure(4)
imagesc(imagedata3,[0 0.0008])
%xlim([450 850])
%ylim([300 700])
title('transform')
This code takes the projected image and returns several transforms. As it turns out, the Fourier transform of a Chinese character is actually its English translation... just kidding.

The image of the cutout is originally stored as a matrix with 4 layers of 2-dimensional data. The first few lines selects one of those four sheets of that matrix to process as a two-dimensional image. In line 8 there is the option to clean up the image with a threshold where every pixel above a certain brightness will be treated as equally bright, and anything dimmer will not make it into the final image matrix for processing. To take a Fourier transform I used the `fft` command, and since it is two-dimensional I used `fft2`. `Fftshift`, makes sure the transform is centered. The code then gives me 4 images, the input image matrix, Fourier Transform of input image, Inverse Fourier Transform of input image, and then the Inverse Fourier transform of the forward Fourier transform of the input image. This last one is to see if you can get back your original image after transforming it forward and then back again, more about this later.

Here is the input image and transform:

![Input Image and Transform](image)

The Forward and Inverse Fourier Transforms ended up being the same, just with different magnitudes. I was curious to see if I could print out the Fourier transform of the character and put it
back into my optical system to get the guang character back. This would be the hologram of the character. Unfortunately it is not so simple.

Here is what you get after transforming an image forward and back again:

Although it clearly has some information of the original, much of it is lost in translation. This is due to the fact that transforms give complex expressions and are squared in my code. Similar to a squared real number losing the information on whether it was positive or negative before, these complex expressions lose all of the information held in their imaginary components when they are squared. It is necessary to square them though as this is what gets us the intensity of the light, which is what we see on our camera (not to mention the fact that a complex expression is impossible to graph). The real component: amplitude squared is equal to the intensity, while the imaginary component corresponds to the phase of the wavefront and squaring a complex expression completely cancels out the imaginary components. The inverse transform is really telling us where to affect the intensity of the light, along with how much to delay it at each point. The latter is lost in this process.
This problem is the main reason making a hologram of an object is often so difficult, the amplitude and phase of a wave at the image plane are tied to its intensity at the fourier plane. We will get back to this problem in the next chapter.

For now let's look at what our camera sees in the fourier plane when a simple cutout is placed in the beam's path. This device is sometimes referred to as a coherent optical computer. It allows us to insert an obstruction (like a mask) into the image plane and partially block or filter out certain spatial frequencies. Altering the frequency spectrum of an image and thus getting the optics to compute a Fourier transform. We will first see a cutout of a square as seen through the optics setup, and its transformation.
We can see a bright center and four legs with rippling intensities. This transform is quite similar to the 1 dimensional example we did earlier in the fourier transform introduction, where the input is zero everywhere except for some length in the center where there is a constant value - tophat function. The tophat function lies along the x and y axes such that it creates two symmetric transformations in those axes.

To make sure that this is the correct transform let's compute the Fourier transform of this shape on a computer.
It appears that the optics setup is accurately taking Fourier transforms of the 3d printed cutouts.

Let's try a circle:

We get a similar rippling effect, but instead of legs it oscillates out in all directions. This is because like the tophat function from before, our input image is zero except for a segment of light in the center of equal intensity. Instead of being two tophat functions in each dimension, it is a tophat function rotated 360 degrees such that the fourier transform of the tophat is distributed radially.

For a smaller circle cutout we get this pattern:
This circle is half the size of the previous one, and due to this its ripples have twice the wavelength of the larger circle. The reason for the asymmetry in the rings is that the 3d printer had a much more difficult time producing a perfectly round circle at this size.

These diffraction patterns from a circular aperture are known as airy disks.

Far from an aperture, the angle from the optical axis at which the first minimum occurs is given by:

\[
\sin \theta \approx 1.22 \times \frac{\lambda}{D}
\]

This pattern is very common in optics, where after being sent through a pinhole a laser will present this bullseye pattern.

Let's see the diffraction pattern of a triangle:
It is strange that a 3 sided shape has 6 legs while a square only had 4 legs. This is a trend with images that have an odd number of features. I believe it comes from the fact that a flat line will cast two legs symmetrically on both sides. Since a square has parallel lines next to each other vertically and horizontally the legs get cast on top of each other and double up in brightness, while for a triangle each of its 3 sides cast two legs which extend into areas that aren’t occupied by another leg.

Hexagon:
This explanation shows how a triangle and a hexagon can have the same number of legs while having double the amount of sides. Since like the square each side casts two legs, but with an even number of sides the legs cast on top of each other.

A pentagon should have more legs than a hexagon.

Pentagon:

You could see how if you increase the number of sides on a polygon the more you approach the transform of a circle as the legs fill out more space and eventually become continuous concentric rings around the center.

The odd number rule is still holding up as this pentagon has 10 legs, 2 for each side.

This rule has an interesting connection to astronomy where images of stars captured by telescopes will have 4 legs extending from them. These are from 4 cables that hold a camera above a parabolic mirror that captures light. In this system the mirror is essentially acting as a lens and will take the fourier transform of the 4 wires holding up the sensor. One might think that using 3 wires would be
the better option, since 3 wires blocks less light, but due to the fact that 3 wires produces 6 legs on every star in your image this makes 4 the better option.

The transform is just dependent on the spatial frequencies so the position of the cutout in the beam doesn’t matter. Rotations do rotate the transform though. Something interesting happens due to this lack of dependence on frequency. If you put a circle in the cutout and the transform is centered in the fourier plane, what happens when you put in more than one circle. This is 3 dots equally spaced in the cutout. The airy disk pattern is the same as if only one dot were present, but there is a new honeycomb pattern which emerges.
This optical computer has an incredible way of transforming large amounts of data instantaneously. Aside from Fourier transforms you can train this system to do image filtering. One example is taking out vertical or horizontal lines by incorporating a grating at the Fourier plane. It can also do an optical lowpass filter, where a pinhole is placed in the center of the beam at the focal plane, because high frequencies are located far from the optical axis this causes sharp boundaries to vanish and can cause shades of gray to fill in between black and white. Similarly you can create a high pass optical filter by putting an obstruction in the center of the beam at the Fourier plane. This blocks the low spatial frequencies and alters your image so that its edges appear much harsher than before. I noticed that this setup is incredibly sensitive to airflow one day while I was cleaning the mirrors with an air canister. If any stream of air passes through the beam at the image plane the camera picks up an exaggerated movement at the Fourier plane. Later on while reading through uses for this optical computer, I came across something this setup can do called Schlieren Methods. You can use a variation on this optical system to capture beautiful images of small changes in the density of air, for example it allows you to see the texture of the rising hot air from a candle. If I were to continue this project I would likely try to show its image recognition capabilities, where you can use this setup to create a pattern at the second imaging plane only when the mask in your beam matches a reference image at the Fourier plane. This reference image can be anything from a bacteria to a military jet plane.
These are some of the more notable transforms I captured.

Fourier gallery:
Digital Holograms

Holograms are a unique and useful application of wave diffraction. They can be convenient solutions to issues regarding information storage, anti-fraud measures, and medical mapping. I will first try to make a tactile 3D printed hologram that diffracts light into the camera. Next I will use a standard printer on a piece of transparency paper and see if the ink is enough to diffract the light.
In Matlab I was noticing that the shapes that transformed forward and back the best had some element of symmetry. I think this has something to do with a connection between the imaginary component of the Fourier component and symmetry. Similar to the way a positive real square root has two values symmetric in the positive and negative directions, while the square root of a negative number only has one imaginary component. I think this might make a non-symmetric shape require more imaginary components to describe it. I tried making a code in Matlab that took an image and turned it into a 4x4 grid of squares with a set of the original image and a symmetric version in the top left and bottom right respectively. Then fourier transforming that image forward and back, but this had similar problems as the code before.

Here what I got from transforming an image of two symmetric rectangles forward and back again.

It is relatively close, but still no cigar.
Finally I found a paper from 2015 that describes a solution to exactly this problem. It uses the delta function and some clever matrix manipulation to set the phase at the image plane to be random white noise after the transformation. This makes it so that at the mask, it holds most of the information in its amplitude, and less of it is dependent on the phase at the mask.

I used their solution and this allowed me to get my first working hologram!

I printed this onto a transparency and projected the light through the hologram and into the camera. Here is the image of the hologram next to the input data I put into the hologram code.

The dots are 1 inch apart for reference.
This hologram yielded the following results.

On the next holograms I tweaked the setting on the image in adobe illustrator, and put them back into my fourier transform code to predict how good they might turn out. The printer adds another layer of noise onto the image, but it helps to know if the transform is good or not.
**Double photon polymerization**

To make a hologram people often use a transparent substrate whose surface has raised areas which delay the light so that the wavefront is shaped accordingly. To make a pattern that can affect the waves properly we will need features near the size of a wavelength of light itself. In order to make features this small in such a precise way, we will use nothing other than light!

To sculpt with light all you need to do is find a liquid sensitive to light such that it becomes a rigid plastic when exposed, then when you are done the areas you haven’t exposed can be easily washed away leaving behind your transparent print.

One problem to this is that this whole system is 3 dimensional.

This means that although we only intend to cure one point in the resin with the laser, we will inadvertently draw out a column of cured resin above and below the target point.
Focusing the laser to a point doesn’t really help either, as now we are curing two large cones above and below the target point.

The solution is to use a mechanism called two photon polymerization. This is where we use a special resin that has some probability of taking two red photons, and doubling them up into a single photon of twice the frequency- blue light. This has a very low probability of happening though, so it takes very high intensities (large numbers of photons) for the method to take effect. This can be utilized where at the center of the focal point of a highly focused beam, you will see these photons “turn blue”.

Since this only occurs at the focal point, we are able to draw with blue light at a single point in the volume of resin.
This method is ideal for creating nanostructures to affect the phase of a wavefront of light, but unfortunately, in the end I had very little to show for my effort here.

Summary

At first it seems that light travels as linear rays, but upon closer examination light follows complicated, but incredibly predictable paths. One it was understood that light is a wave physicists were able to apply all of their tools previously used to describe other waves in nature to predict how light will travel. The Fourier transform is an incredibly useful tool to bolster our understanding of waves. I hope that I have shown that it is not just a mathematical abstraction, but rather a tactile and ever present entity in our daily lives.
Bibliography


