

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw@mit.edu](https://ocw.mit.edu).

DENNIS

With that, what I want to do is think about-- finish up thinking about modulation. Last time we thought about modulation in a communications context. That's a very important context. It's a way of thinking about how we can use modulation to better match a signal to the medium.

FREEMAN:

So we saw in particular that if we were trying to transmit human voice via electromagnetic waves, trying to simply launch an electrical representation of the voice into electromagnetic waves, just doesn't work very well. And that's because of the enormous frequency difference. We would rather have frequencies on the order of 2 gigahertz for efficient transmission through the atmosphere and human voices just are not centered on 2 gigahertz.

We can use modulation to bridge that gap. Having done that, we get a number of other advantages too. So for example, we talked last time about the idea of broadcast radio, which was an enormous revolution. The idea of being able to instantly communicate lots of stuff like a newspaper, but instantaneously, not with a day or a week delay, was an enormous deal.

Today what I want to think about-- I mean the way we motivated last time-- the way to do the modulation that we motivated last time was the idea of amplitude modulation. Amplitude modulation is a terribly non-linear process. We multiply two signals. That's pretty non-linear. And you could imagine that if we're willing to open up that can of worms so that we're willing to do a transformation that is non-linear, there's enormous numbers of them. So you don't have to just do amplitude modulation. You could also do-- here are three classics-- you can do phase modulation and frequency modulation as well. All of those systems work.

Phase modulation, the idea is-- Well, first off, amplitude modulation. The idea was you modulate so-- this is the carrier. $\cos \omega_c t$ is the carrier. That's the thing that goes through the medium well. Now we want to somehow embed the message on the carrier. In AM, we did that by multiplying. We take the signal $x(t)$ and we multiply it. We take the signal of interest, $x(t)$, and we multiply it by the carrier and transmit that. And we saw that, for reasons of decoding simplicity, it was convenient to add a constant term, which was essentially sending, not only the modulated carrier, but also the carrier alone, so that you could locally retrieve both the modulated message and the carrier. It simplified the modulation.

But a different kind of alternative is, convey the message in the phase. Right? We all love

phase, right? Nod your head yes. We all love phase. So what we could do instead is take the carrier and add a phase term in proportion to the message. That's a different way of coding the message on a carrier.

Another way to do it is to code the frequency. So take the message, integrate it-- it is kind of a funny way to talk about FM. So you could use the message to modulate the frequency of the carrier, which would be very similar to using the message to modulate the phase. They, in fact, differ by an integral. And I'll talk about that more in a minute.

So the idea then is even though the first thing they tried-- the first successful way of making broadcast radio --was via AM, we should think about the alternatives and try to figure out what are the advantages and disadvantages of different kinds of coding schemes.

And also consistent with the goals of this course, in thinking about these alternatives, we're going to get a lot of practice in thinking about Fourier transforms. Which is good.

So let's think about what would be the difference between AM modulation, where we convey the message by modulating the amplitude of the carrier versus FM, where we convey the message by modulating the instantaneous frequency of the carrier. And if we're just looking at time domain, which is, after all, probably the first place anybody should look-- if we just compare those two schemes in the time domain, there's some very big differences.

You can see that when we amplitude modulate, the power that is used to transmit the message depends on the message. So there are places where there is not much power. There's not much energy in this signal. And there are places where there's lots.

By contrast, down here you see something that has constant power. At least if you integrate over-- say you've got a 2 gigahertz carrier, then as long as you integrate over 10 or 20 cycles, which is a small fraction of a second, you're going to get the same answer, regardless of where in the message you do the integral. So there is obviously a power difference. There's no need to transmit the carrier in order to decode the FM signal, unless of course you're interested in understanding exactly where DC is. The interesting thing about audio is that we're very insensitive to DC. So that's not really very important for conveying speech sounds.

But the original motivation for thinking about alternative ways of coding was in fact bandwidth. So the idea was there's a limited resource. Bandwidth is a limited resource. If person A is using 1 megahertz plus or minus 5 kilohertz, which is one of the broadcast AM frequency

bands, then person B can't use that band. You can have only one transmitter for each of the radio bands. That makes bandwidth-- radio bandwidth --a resource. Now there's lots of it. There's lots of frequencies between a megahertz and 2 gigahertz. On the other hand, there's lots of people who want it.

So your local fire department thinks they ought to be able to talk to each other. Your local police department thinks they ought to be able to talk to each other. Your local ambulance service thinks they ought to be able to talk to each other. So there's lots of people with demands for it. So the idea of-- the original pursuit of FM was to try to think about a scheme that would use less bandwidth. If you're thinking about speech, you need about 3 kilohertz. 3 kilohertz is considered telephone quality speech. It's not perfect, but it's good enough to get very good speech intelligibility across it. So I AM used plus or minus 5 kilohertz bandwidth. They allocated a band of 10 kilohertz for every station.

The initial idea in FM to-- so take this expression for FM and think about an instantaneous frequency ω_i , which is a frequency with a slightly different value from the carrier frequency. It's different because we're doing FM. And you can calculate the instantaneous frequency as the derivative of phase. So the derivative of phase gives you an ω_c term out front. $\omega_c t$ is the carrier phase. And then there's the part of the phase that comes from the message. So the total instantaneous frequency is ω_c plus the time derivative of phase. And since we're modulating frequency-- since we're modulating with the frequency, that turns out to be proportional to x .

So the instantaneous frequency, like you would like for frequency modulation-- the instantaneous frequency --is a linear function of the message x . But it's proportional. So the reasoning was make k small. If the instantaneous deviations of frequency are small, say 10 to the minus 6th hertz-- really small. Well, if you could have 10 to the minus 6th hertz, you could pack 10 to the 6th stations in a hertz. Well that's pretty good. So instead of using 10 kilohertz to send one message, you could get 10 to the 6th messages in one hertz of bandwidth. That was the original motivation. That was an idea that was propagated at Bell Labs, who were one of the commercial entities who were seriously interested in trying to make money on broadcast radio.

So the idea was maybe we could use FM to squeeze more signals in the available bandwidth. Well that turns out to be completely wrong. And that's why studying Fourier transforms is such a good example of how you can use Fourier transforms to figure things out. That argument is

just completely wrong, as two lines of double-0 3 will show you.

So here is our expression for the FM signal. You can see that it's complicated because it's the cosine of a sum. But we all know from trigonometry, the cos of a plus b is cos a cos b minus sine a sine b. So we can get an exact expression for this part-- we can expand this exactly as the cosine a times the cosine of b minus the sine of a times the sign of b. Well that's easy. Now what would happen to that expression if we made k very small? Well if we make k very small-- if k goes close to 0 --the cosine gets arbitrarily close to 1. Fine, that sounds OK.

This sine of l times something does not go-- well the fallacy in the reasoning was that goes to 0. It does go to 0, but it actually goes to 0 slowly. It actually-- the limit approaches k times-- so the limit of the sine of theta, as theta gets small, is theta.

So the idea wasn't quite right. It doesn't go arbitrarily close to 0. It gets arbitrarily close to the message. So that means that this expression, which looks horrible up here, is equivalent to this expression, which says that the signal that I'm transmitting is the carrier. Just like AM minus sine omega ct times the message.

But that's just AM. I took the message times sine. Uh, it's sine. It's not cosine, who cares?

That's a difference of 90 degrees of phase. Who cares? It's AM. The fallacy in the reasoning is that this k does not-- the sine of theta, when theta gets small, does not go to 0. It goes to theta. And so the limiting case for narrowband FM has precisely the same bandwidth as AM. Narrowband FM has the same bandwidth as AM, so therefore, this whole idea that you could use FM to squeeze more channels into a given bandwidth is just completely wrong. So it was the initial motivation. It's just wrong.

In fact, what's good about FM is the other limit. Don't worry about narrowband. Worry about broadband. The value of FM-- and this was Armstrong again. The same guy who did the superheterodyne receiver, which made AM broadcast radio possible. Same guy went on to think about FM and he saw that the value of FM was, in fact, to use lots of bandwidth. Why would you do that? Well, we'll see in a minute. That the reason you want to use a lot of bandwidth is that you generate a robust signal that can be recovered, even when noise gets added to it.

One of the big problems, especially with the early versions of AM broadcast radio, was that it had a lot of noise in the background we called static. Kind of like tsss, but a little bit more

poppy and more irritating. And Armstrong figured out that there was a way to reduce that static by using more bandwidth to make a more robust signal. So that's the idea. And coincidentally, it gives us an excellent opportunity to practice our skills of figuring out Fourier transforms.

So let's figure out the Fourier transform of an FM signal. Right? That'll be fun. Remember what the FM signal looked like? We saw back here, it's kind of horrendous looking. So the goal for the next five minutes is to take the Fourier transform of that.

So let's think about phase coding. Phase and frequency coding are almost the same thing. All you do is, in one case you transmit a phase in proportion to x . That's phase modulation. In the case of frequency modulation, you transmit a phase proportionate to the integral of x , which is the same as modulating the instantaneous frequency by x .

So let's think about phase modulating by $\sin \omega_m t$. This is ω_c carrier and ω_m message. And I'm putting an m out front. The modulation depth. Just so that I can track what happens, as I turn up the amplitude of the signal.

So let's think about, what does this signal look like? So we have $\cos(a + b)$. We expand that as $\cos a \cos b - \sin a \sin b$. And let's start by looking at this first term, which is modulated something. The something is the hard part. So the something is the cosine of $m \sin \omega_m t$. For the particular case, that the message is the $\sin \omega_m t$. So let's think about that.

So we'll start with the message. $\sin \omega_m t$. And we'll say that the modulation depth is 1. Now we have to take the cosine of that. So now we take the cosine of this signal. So this signal is going 0- 1- 0 minus 1- 0. So the cosine of that-- when the sine is 0, the cosine is 1. Then the sine is going toward 1. As the sine goes toward 1, the cosine starts going down. Right? The first part of the cosine waves, cosine t , as you increase from 0. It starts to decay. Same thing's happening here. But then by the time the sine gets up to 1, the sign starts going down. So the cosine starts going back up. Everybody see what I'm what I'm implying? So this waveform is a sine waveform. And this is the cosine of that waveform. That's what an FM signal is.

As I increase the modulation depth, so make the modulation depth now 2, now the deviation from 1 is bigger. Right? Before, the deviation for 1 was caused when the sine wave went up to 1. Now, the biggest deviation occurs when the sine wave gets up to 2. It's bigger. And then I may make it bigger and bigger.

OK. By the time-- The big difference between 3 and 4-- 3 and 4. Why is a big difference between 3 and 4? It's because there's a number between 3 and 4. Pi. By the time you get to pi, it rolls over. Everybody see that? I started out 1, 2, 3, just less than pi, 4, just bigger than pi, 5, 6, 7, 8, 9, 10, 20, 50. Horrendous, right? But I want you to find the Fourier transform of that. Actually, I want you to find the Fourier transform of that! But as a subtopic, I'll accept the Fourier transform of that. OK. Look at your neighbor. Tell me the Fourier transform of that.

So what's the Fourier transform of the bottom thing? OK. Was a hard question. Yes! Yes. Oh nice! It was number 4, got it. [LAUGH] So what should I notice about this signal? Except what? As a sophisticated signal processor at this point, after all it's the penultimate lecture, you're already sophisticated. So as a sophisticated signal processor, what do you notice immediately about that waveform? It just cries out, I feel. It says? Periodic. Exactly. Even though it's horrendous, it's periodic. Why is that interesting? It has a series. Precisely.

So all I need to do to figure out the transform, is figure out the series. OK. So let's do that. Now let's start. The same waveform on the top. m equals 1, just like before. Now I'm going to take the series of this periodic waveform. And I'm going to represent that down here. As you see, as I start to modulate this waveform, I'm getting two bumps, where I had one bump of this-- where I had one period of this. I'm getting two periods here. So that's the reason my first non-0 contribution is k equals 2.

Also notice that if I turn to m the whole way down to 0, I would just get dc k equals 1. So for the small m 's, I'm mostly getting dc and k equals 2.

OK. Now as I turn up the amplitude, now I'm getting something that doesn't look quite as much like a sine wave. And that distortion is manifest as some k whose 4. Then I turn it up higher. Whoops, I hit the button a few times more than I intended to. If I go up to m equals 5, now it's wrapped around like this. And I can see I've got k equals 0, 2, 4, 6, and a little 8. Keep going up, up, up, up. And what you can see is that as I turn up the amplitude, I'm getting bigger and bigger k 's.

In fact, there's kind of a simple relationship. By the time I got up to m equals 50, I got about 50 k 's. Roughly speaking. And that has something to do with the periodicity of the cosine being 6.2π . So the number of terms I'm getting is related to how big that signal is.

I was able to represent this horrendous signal now by a series, but I asked you for-- what I'm

really trying to do is find the transform of this. We just found the series of this, but I really want to find the transform of that. So I'd like to turn this into a transform.

How do I turn a series in a transform? Turn each k into an impulse. So now I get a train of impulses. Each one of them has weights proportional to the length of the lines. Then each impulse gets located in frequency, instead of in k space, each impulse gets located in frequency at a multiple of 2π over ω_m . So when I do that, I get the transform of this mass, which looks like this centered here, but then I modulate by the cosine. So modulation gives me two copies. I get the whole spectrum for the inside here and a duplicate over there. This time centered on ω_c . That's all perfectly clear. Right? So then I just did this term. Now I have to worry about that term, but it's the same thing. Except now my periodicity is a little different. Now I'm taking the sine of the sine rather than the cos of the sine. So now as I crank up the waveforms, the picture looks slightly different. The principle harmonics are now odd because of the odd symmetry of the sine wave.

I had evens when I used cosine. I have odds when I use sine. So now I'm filling in a bunch of odd harmonics, but the general pattern looks very similar. So that now the result looks very similar to the result from before. And the sum is the sum.

Here's the Fourier transform of this waveform y of t . The point is it's huge bandwidth. The advantage of FM is not in conserving bandwidth, it's actually using bandwidth. So we're using-- this is the K , big case. So we're doing wideband FM. We're shoving frequency components all over the spectrum, but the advantage of that is that I get a signal that's very robust.

Imagine if I were to add a small level of noise to this signal. You could still recover the signal. The only thing that's coded here is one sine wave that has one period across that whole length. You could recover that in the absence of, not only a little noise, but an enormous amount of noise. And that was what FM was good for. And that's what Armstrong figured out. And that's why we have FM. And that's why television, even HDTV, uses FM coding of the audio. Because it's resilient to noise.

So that was kind of a motivation for thinking about modulation, in terms of communications. I don't want you to go away thinking about modulation as only useful for communications, so I want to close with a kind of unconventional use of modulation.

I may have mentioned that I like microscopy. I'm going to show you how you can use modulation to improve microscopy. So this was an idea of Michael Mermelstein. Michael was a

PhD student in my lab. He thought of it. Stan and Jay developed it. They all three got PhD's on this topic. Berthold is a professor in CS and five of us worked on this project. So the idea is improving microscopy with 6.003. With modulation in particular.

You've all seen this. This is the double-0 3 model of a microscope. Double-0 3 microscope convolved with blurring function. Done. That's what a microscope is. That's the double-0 3 model of a microscope. Microscope is a low pass filter. Because all of the different spatial frequencies that are available in the target can't get through all the lenses for very fundamental reasons in physics, the high frequencies don't make it. The low frequencies do. The result is a blurred image. So I'm representing this as the target. It passes through the microscope, which is represented by a low pass filter and it comes out a blurry picture.

Michael's idea was instead of illuminating the target with uniform light-- if you tear apart a microscope, a lot of the guts are intended to make a nice uniform light that goes across the target. Nice uniform illumination. Michael's idea was, let's not do that. Let's project stripes on it. That's a little bizarre.

So instead of having a nice, blurry picture of a target, Michael wanted me to generate blurry pictures of stripey targets. But you're a sophisticated double-0 3 people, as well as Michael. Why does he want to do that? Different frequency. Phase modulated microscopy. That's why he wants to do this. We're going to modulate microscopy. OK, well what's that? OK, so Stan Hong was one of our TA's in double-0 3. And he wanted to explain the way this works using purely double-0 3 terms. He also knew that I would be receptive to that and that Bertold Horn, who teaches this course all the time, would be receptive to that kind of an argument.

So Stan made a picture to illustrate that. So here Stan's picture. What do you see? Nice picture, right? What do you see? Great. Excellent. That's exactly right. Some of you close to the front, what do you see? Stripes. So what you see is stripes. And if I were sitting where you're sitting, I wouldn't see the stripes. Why is that? Because my eyes, like any optical system, blurs. Me being old, they blur more than you being young. You can see stripes better than I can. So if I were to sit there or if I read it-- or if you were to sit-in the back row it would be hard to see the stripes would you still see blur. You'd still see gray.

So what do I do? This is double-0 3. We just did phase modulated microscopy, but what should I do next? I have a stripey picture. How did that picture get stripey? It got modulated. What should I do next? Demodulate it. So how do we do that? Do the same process when--

and that process was?

So I put stripes-- I illuminated the picture with stripes in order to make the original. Now what I should do is illuminate the result with stripes. If I multiply the picture by stripes I'm phase modulating it. Multiply by $\cos \omega$ whatever, right? So now I'll put the clicker down. No cheating, right? Nothing up my sleeves. And now I'm going to project that stripey pattern. Cute, huh? It's pretty amazing. So who is that? Fourier. Good.

That's the principle behind Michael's microscope. And by the way, I'm not cheating. So if I roll the phase-- It works just like radio.

Now how's it work? And how well-- so that was a demo. Stan's committee loved it.

Here's the idea. So the poster was phase modulated Fourier. Fourier is a function of x and y . And you modulate the phase of carrier. The carrier now is the y displacement because we were putting this kind of line on it. So we modulated carrier. $\cos \omega y$ was the carrier by the phase of Fourier. So we bumped the lines up and down. We being, Stan. Stan bumped the lines up and down in proportion to the brightness of Fourier. And that put Fourier's content centered on ωc .

It was hard to see from the audience because your eyes-- the projector projected ωc . It's hard for you to see from the audience because your visible frequencies in space don't go that high, but you beat it with the stripey pattern, that modulates in space just the way it modulates in time. And it takes a copy of Fourier, which had been modulated up to ωc and bumps it back down to the visible. That's why you can see it.

Well, we're going to see that in a moment. Good question.

The idea is precisely the same as the superheterodyne radio. If this is the complicated picture that we would like to look at-- this is the Fourier transform or the complicated picture that we would like to look at. If we demodulate with this stripe at ωc , we're able to bounce down this pe house and we're able to see that part of the picture, even though those frequencies are too high for them to go through the microscope. There's a limited number of frequencies that will go through the microscope. Just like there's a limited number of frequencies that'll go through your eye. And we can take a band of frequencies, that don't make it through the microscope, beat it with this stripey pattern into a frequency range that does go through the microscope. Then we can change the stripey pattern. If we make stripey pattern with slightly

higher frequency, we get a different part of the invisible spectrum. And we just keep repeating.

There's a bit of an issue that it's a 2D transform. We have to worry about frequencies in x and frequencies in y, but we are all experts at this sort of thing. The difference between stripes this way and stripes this way is Fourier's this way and Fourier's that way. You rotate the stripe. It just rotates the Fourier transform.

We can think about, in space we can modulate like that or we can modulate like that. And in fact, what we would like to do is modulate by a whole bunch, so that we could take all these little regions-- so say the circle corresponds to the radius of frequencies that get through the optical microscope. What we'd like to do is put that little radius at every possible place, one of the time, to bring down those frequencies in the target, one at a time.

So the idea-- Then the problem becomes, how do you make so many stripey patterns? And you have to generate those stripey patterns at very high spatial frequencies. Very small distances. If you're going to beat a microscope-- Optical microscope resolutions are on the order of 500 nanometers. Wavelength of light. So you're going to have to make these patterns small compared to the wavelength of light.

So Michael's idea was interference. You take two coherent laser beams, point them toward each other but at a slight angle, and they will interfere and make a stripey pattern. Then you turn on a different pair of beams and you get a different stripey pattern. Different. Different. Different. I'm showing on the left, the spatial, and on the right, the Fourier transform. Fourier transform for the stripey pattern. All light-- all pictures have some dc because there's no negative photons. So that's this. And these are coding-- the angle is coding the orientation and the distance is coding the pitch. Different stripey pattern. Different. Different. Different. There's a bunch of them.

With 15 beams, you get 15^2 . Order 15 squared. That's the idea. And that was Michael. This is Stan. Stan figured out a way to build an apparatus to take the beam from a laser, break it into 15 parts, steer it with a bunch of mirrors, and point them all toward the center. So the target-- so the lasers coming in over here. There's these pick-off mirrors steering things around.

This was some complicated optimization that Stan did. Here is the region of interest, which is this big on his microscope. And about 2 centimeters down, the beams converge on a specimen. Which is right there. So if we zoom in and you can see it a little bit better, there is a

conventional microscope objective with 15 beams fired toward it.

Now the idea is you put the specimen between the beams and the objective. And then you view the stripey illuminated target with the microscope objective.

Here is a picture taken by Jay, where he took a bunch of small beads.

It's easy to get plastic beads of very uniform dimension. So these are about 1 micron beads and about a gazillion of them. So he just made a solution of beads, put them on a glass slide, evaporated it so that he would have a random constellation of plastic beads all one micron in diameter. And here's one picture of it. And you can see the pixels. Each one of these squares is one pixel in the camera. I'm zooming in a lot. Keep in mind that these guys are only about a micron in dimension. But then if you change the stripey pattern, you get a slightly different picture. And if you change this stripey pattern again. And again. And again. And again. And again. And again.

So he recorded then, 300 and some odd pictures of the same thing with different stripey patterns. Then he did some signal processing and turned that sequence of pictures into that picture. So you can see the resolution is up a bit. The resolution is very sub-pixel. You can see many resolution elements inside one pixel from the original picture.

In fact, if you compare the original to the reconstruction-- so this is one picture taken with uniform illumination. This is the result of calculating some 300 pictures with structured illumination, with different structure in each picture. And you can see a lot better resolution here. In fact, you can see here there's something that looks like it might be a bead, but over here you can see more clearly it's really two beads. And if you take a solitary bead and plot the brightness through a line, you get this. But if you plot the brightness through a reconstruction, you get something much narrower.

Stan then developed a method for scanning this around. He also developed a method for measuring the point spread function directly.

Here I want to show you a measurement, same sort of apparatus, with a 200 nanometer bead. A much smaller bead. Keep in mind that the apparent brightness goes as the cube of diameter. So changing the diameter by a factor of five, changed the apparent brightness by 5 cubed.

So here is a picture taken with Stan's microscope. And if we take a row of pixels and plot the brightness, this is the reconstructed image by using standing wave illumination, which is what we called it. And it's got an apparent diameter of 290. The 290 is bigger than the 200 because of the blurring of the microscope. The prediction, based on the angle of the laser beams-- the resolution of this microscope depends on the angle of the laser beams because the angle of the laser beams determines the pitch of the fringe. That the pitch of the stripey pattern.

So the prediction, based on the angle of the beams from the apparatus, was that it should have been 250. And the fairest thing to compare that to is, what would it have been if we hadn't used any wave illumination? If we hadn't used any illumination, it would have been 1500 nanometers. So you can see, there's an enormous increase in the resolution by using this phase modulated microscopy.

So the point is that you can use modulation for a lot other things than communications. Because the application and communications is terribly important, but here was a completely different application of modulation to improve microscopy.

Thank you. See you next time. Fill out the subject evaluation.