

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)

**PROFESSOR:** Hello and welcome. So today is mostly distinguished by what happens tomorrow. Not surprisingly, but of course, you all know that. So tomorrow we have our first quiz. Tomorrow evening 7:30 to 9:30, no recitation. We've been through this several times. I won't spend any time on it other than to ask if there are any questions, so if I don't hear any questions, the idea is going to be at the end of this lecture, the next time we'll see you is in office hours, and, or tomorrow, Wednesday, 7:30 on the third floor, building 26. Questions or comments about the exam? Yep.

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** So that one page of notes, 8 and 1/2 by 11, front and back. You can write as small as you like. In fact, later in today's lecture, I'll show you how a microscope works and you're welcome to use a microscope because they're completely non-electronic. Oh, as long as you use an optical microscope. Other questions about the exam?

OK, then for today, so far, since the beginning of the term, we thought about a number of different kinds of representations for both DT systems-- discrete time-- and CT systems-- continuous time systems-- and we saw that we were interested in that large number of representations, because each of them had some particular aspect that made it particularly convenient sometimes.

So, for example, in both CT and DT we looked at verbal, but not so much. That was mostly in the homework. We looked at difference in differential equations, mostly because they were so compact, so concise, so precise, they told you exactly what the system does. No fluff, this is it. So that was nice. Block diagrams, by contrast, are less concise, but they tell you the way a signal propagates through the system on its way from the input to the output, and that can be very helpful for understanding why certain behaviors occur, especially when we talk about things like feedback. We looked at operator representations. They were nice because we could transform the way we think about systems into the way we think about polynomials, so we reduced a college level thing to a high school level thing.

That's always nice, and then we looked at transforms. The distinguishing feature of transform representation was that we took an entire function of time, and turned it into an algebraic

expression. So we turned a differential system that described a system of differential equations that describes a system into a system of algebraic equations that describes a system. All of those were useful for different ways and what I wanted to talk about today is yet another way to represent a system and that is to represent a system by a single signal.

So in some sense we're going backwards, because we're taking what we would normally think of as an entire system and reducing it, getting rid of the system altogether, so all we're going to have is signals. That turns out to be a particular powerful way to do some sorts of operations, and is actually the first instance, the first step, in that we will take in a major field thought of signal processing. When you reduce the entire behavior of a system to a signal, we then regard the whole processing task as a signal processing task. Signal processing task, not system.

So far, we have focused on the responses to the most elementary kinds of signals-- unit sample signal, unit impulse signal-- but generally speaking, we're interested in much more complicated signals. As you've already seen, I've already asked you to calculate things like unit step responses. So generally, we're going to be interested in much more complicated signals-- the responses of systems to much more complicated signals. That's not hard. The reason we skipped over it was that you can always figure out the response to a more complicated signal at least by falling back on some of our more primitive ways of thinking about systems.

So for example, if we think about a difference equation or a block diagram, we can think about how a more complicated signal excites a response by simply thinking about the system operating on a sample by sample basis. So you, of course, all know that, and to prove that you all know that, answer the following question. Here is a system. Here is a signal that is more complicated than just a unit sample or unit sample signal. Figure out what is the third response of this system to that signal? You should be absolutely quiet. That was sarcasm, just so you know.

Lots of self-satisfied looks so I assume everybody's done. So what's the answer? Raise the number of fingers that corresponds to the answer  $y$  of 3. Raise your hand so that I can see them. Drop the ones with the wrong answers. I don't want to see those. OK, about 50%, so maybe take another 10 seconds. Notice that I'm asking for  $y$  of 3.

So what's the answer  $y$  of 3, and raise your hand. Everybody has the right answer, raise your

hand. Much better. OK, so now the overwhelming majority says the answer is 2. That's not very hard. If we think about the block diagram representation, and if we think about propagating the more complicated signal represented over here through that signal, through that system, we start with the system at rest. That means there's 0's coming out of all the delays, at time  $n$  equals minus 1,  $x$  is 0, combined with the initial 0's, we get the answer is 0. And then, at time  $n$  equals 0, the input becomes 1, so now the output goes to 1, at time 1, it goes to 2. At time 2, it goes to 3. Times 3, it goes to 2. This is obvious, right? Et cetera. So the answer is 2.  $y$  of 3 is 2, and the point is that it's very trivial to think about by thinking about the system in a sort of sample by sample way.

Not surprisingly, the point of today is to not think of the system sample by sample, but to elevate the conversation from samples to signals. The first step in thinking about it as signals is to realize that you can think about the response of the system by decomposing the input into additive parts. I can think about  $x$  which is this 3 sample signal. I can decompose it into single samples, and then think about the response to each of those, and it may not surprise you that if the system were linear, then the response to the sum would be the sum of the responses. So you can sort of see that there would be a way of adding together these rectangular pulses to get a triangle pulse.

So that works simply because the system is linear. The system has the property that the output for a sum is the sum of the outputs for the individual components, and we can write that this way, so a system is linear. We can define linear in a more rigorous mathematical sense by saying that a system is linear if the response to a weighted sum of inputs is the similarly weighted sum of outputs. So imagine that I have a system whose output when the input is  $x_1$  is  $y_1$  and whose output when the input is  $x_2$  is  $y_2$ . We'll say the system is linear if and only if the weighted sum of inputs  $\alpha x_1$  plus  $\beta x_2$  gives the same wakings  $\alpha y_1$  plus  $\beta y_2$  for all possible values of  $\alpha$  and  $\beta$ . If that's true, then we'll say the system is linear.

So then if it's linear, then we're allowed to do this decomposition, because all we're doing is decomposing the input into a sum of inputs. We'll always be able to do that operation of the decomposition if the system is linear according to the definition I already showed. If the system also has the property that we will call time invariance, then the response to the parts will be particularly easy to calculate.

Time invariance has a similar formal definition. We will call a system time invariant if given that the input-- the response to  $x$  of  $n$  is  $y$ . Given that, we'll say the system is time invariant if a

shifted version of the input simply shifts the response. That seems like a kind of gobbledygook sort of thing. It's a very simple minded notion that you should all have from common experience. All it says is that if I do something today, and I get a response, if I do the same thing tomorrow, I should get the same response just delayed by a day. That's all it's saying. So basically the system behaves sort of the same now as it did previously, and as it will in the future. That's what time invariance means.

So if the response is time invariant, then we can compute the response to a shifted unit sample. Notice that this is the unit sample response. This is a shifted unit sample so if the system is time invariant, then the response to a shifted unit sample is the shifted unit sample response which you can see from the picture. So the idea then is that superposition is a very easy way to think about the response of a system. If the system is linear and time invariant, linearity let's us break it up, and think about the response to each part. Shift invariance, or time invariance, allows us to shift the input, and know automatically what's the response going to look like. And there's a formal way we can think about the way you do that operation.

How do you implement this superposition thing? You think about a system as having a unit sample response. This is the unit sample signal. We'll call the unit sample response  $h$  of  $n$ , then a shifted unit sample will give a shifted unit sample response. That's time invariance. Then a weighted shifted impulse gives the same weighted shifted impulse response. Then the sum of such things gives the sum of such things. That's just a formal derivation of a process that we will call convolution.

So the response to an arbitrary DT signal that excites a linear time invariant system can be described by the convolution of the input with the unit sample response. We'll call that formula- - we'll call that operation convolution. Convolution is completely straightforward. For that reason, we try to make it a little bit more confusing by using terribly confusing notation. That too is sarcastic, I mean.

So the only thing that's at all confusing about convolution-- convolution is completely trivial. Here's the way we would write it.  $x$  convolves with  $h$ . The signal  $x$  convolves with the signal  $h$  to give a new signal. Being a signal, I can ask what's the  $n$ th sample look like, and what that symbol means is this sum. The confusing thing is that most people in the field write it this way. The signal  $x$  of  $n$  convolves with the signal  $h$  of  $n$ . The reason that's confusing, and the thing you will never do, because you are here. The thing that you will never do is confuse the meaning of that statement with what looks like an operation on samples.

Had I said multiply, you would have said, if this were a multiply operator instead of a convolution multiplier operator-- if that were multiply instead of convolve, you would have said, oh, that means the  $n$ th sample multiplied by the  $n$ th sample is the product of the  $n$ th samples. This is not true. This is not generally true. The convolution operation means take the whole signal  $x$ . That's why we think about it as an operation on signals, not an operation on samples. Convolution means take the whole signal  $x$ , and convolve it with the whole signal  $h$  to get a brand new signal  $x$  convolved  $h$ , and then take the  $n$ th sample. So the only thing that's at all confusing about convolution is remembering the convolution is an operation that is applied to signals, not samples.

So structure of convolution. So I just showed you a mathematical formula. What I'd like you to have is a little bit of an intuition for what happens when you convolve two signals. So let's think about the structure of this operation. What are we doing? Imagine that we're going back to that original problem. What happens when you take  $x$  of  $n$  and convolve it with  $h$  of  $n$ ? All we need to do is this formula, right? That's all we need to do. What's that formula say? Well let's think about how you would compute the 0-th output.

According to that formula all I did was substitute  $n$  equal 0 every place there was an  $n$ , and what I see is I have to multiply  $x$  of  $k$  times  $h$  of minus  $k$ , but I've got  $x$  of  $n$  in  $h$  of  $n$ . So the first thing I do is I flip the axes and I make them  $k$ 's. That's not hard. Then the  $x$  looks OK, but the  $h$  doesn't. I need  $h$  of minus  $k$ , so I have to flip it. So I'm flipping about the  $n$  equals 0 axis. So that positive  $n$  becomes minus  $n$ , positive  $k$  becomes minus  $k$ , because I want this to be minus  $k$  up here.

Then generally, I have some shift thing here. This 0, because I was looking for the 0-th sample. In general, that might be different. That might be 7 if I wanted to find  $y$  of 7. Then I'd have a 7 over here, and that number represents a shift. In the case of 0, it's a 0 shift. Then I have to multiply these two, so I just place this thing over here so I can multiply. I multiply down. You can see that there's only one sample that in the two is both non-zero. Therefore, I get a single non-zero answer, and then according to the formula, I have to sum so the 0-th answer is flip, shift, multiply, sum, and you just repeat that for all the different answers.

So at time equals 0, the answer at time 0 is flip, shift by 0, multiply, sum. The answer is 1. If I want to find the one answer now the shift is shift by 1. So now instead of having flip which would have put the 3 samples here, I shift by 1 so now they're over there. So now when I do

the multiply, I pick up 2 non-zero answers and the answer is 2. If I wanted  $y$  equals 2, I do the same thing but now I shift by 2. Flip, shift, multiply, sum, that's all I do. It's completely trivial. If I continue, the shift becomes larger and now it's falling off the end. Continue, continue, and I get in general that's the prescription.

So what I've tried to show is two ways of thinking about this convolution thing. The first was by superposition, where I just think about breaking the input into a bunch of samples, thinking about the response to each of those samples and adding. That's an input centric way of thinking about things, because I think of the input being broken up by a bunch of samples. This convolution formula is an output centric way of thinking about things I tell you. I'd like to know the output at time  $p$ , and to compute the output of time  $p$ , you say, well that's easy. Flip, shift by  $p$ , multiply, sum. So input centric, that's the superposition way of thinking about things. Output centric, that's the convolution way of thinking about things.

So now that you know about convolution, find which plot below 1, 2, 3, 4, or none of the above shows the result of convolving the two functions shown above. You're so quiet. I assume you're practicing for the exam tomorrow. You're allowed to talk. So which one's right? 1, 2, 3, 4, or 5? See if I know. It looks good. About I only see one wrong, two wrong, so 95% or so.

So how do I think about this? What's the way that I should think about convolving those? Easiest, most straightforward way, go back to the formula. That will always work. Can somebody tell me a more intuitive, insightful way of thinking about what will be the result of convolving those top two functions? Tell me a property of the result of convolving. Yes, yes, flip, shift, multiply, sum. What's the answer at  $n$  equals 1? 1.

So I've got two things that look kind of like geometrics. Imagine for the moment-- that was intended to be a hint-- imagine for the moment that the sequence looks like 1,  $2/3$ ,  $4/9$ ,  $8/27$ , blah, blah, blah. Imagine that it's a geometric sequence with the base of about  $2/3$ . How would I compute the answer when I convolve that sequence with itself at zero? Flip, shift, multiply, divide, so I started out with two things that were both starting at zero. You flip one of them. How much overlap is there? Just the 1. Just the  $n$  equals 0, what's the answer at  $n$  equals 0? 1, right? So if I imagine that this is the sequence after I flip it. There's a 1 under the 1. The 0's here kill the terms down here. The 0's here kill the terms up there. The only thing that lives is 1 times 1 is 1, so the answer at  $y$  equals 0 is 1. What's the answer at  $y$  equals 1? Flip, shift, multiply, sum. So I flip, shift. So when I shift, the new answer looks like 1,  $2/3$ ,  $4/9$ ,  $8/27$ , et cetera. Multiply and sum, what's the answer?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:**  $4/3$ . So the only non-zero answers are 1 times  $2/3$  plus  $2/3$  times 1--  $4/3$ . If I want to compute  $y$  of 0 1 2, I shift it further. So I do 1,  $2/3$ ,  $4/9$ ,  $8/27$ , blah, blah, blah. So flip, shift, I shift 1 more, multiply, sum, multiply 1 times  $4/9$ , and I get  $4/9$ . Multiply  $2/3$  times  $2/3$ , I get  $4/9$ . Multiply  $4/9$  times 1, I get  $4/9$ . The answer to the sum of those is  $4/3$ . So I get one  $4/3$ ,  $4/3$ . So you can see it's tracing out. This wave form so far, this is the only one that has up and then flat, and if I continue that process it will start to fall off.

If you're exclusively mathematically minded, you can also just do it with math. All you do is think about a mathematical description of the left signal. Say  $2/3$  of the nu of  $n$  and the right signal, and now all I need to do is think about that formula. So do a sum, taking this, the function of  $n$  and turning it into a function of  $k$ . The second one, I want to make a function of  $n$  minus  $k$ . I have to shift both. I have to change the exponent as well as the index into the unit sample signal, same thing over here. Now when I think about multiplying them,  $u$  of  $k$  kills all the terms for  $k$  less than 0, so I can start at 0 instead of minus infinity. This  $u$  kills everything for which  $n$  minus  $k$  is less than 0. That means  $k$  less than  $n$ -- less than or equal to  $n$ . So I end up with this. This product is particularly easy because it's the the  $k$  into the minus  $k$ , so the answer is to the  $n$ , and now I'm summing over  $k$ , but there are no  $k$ 's, so that's summing over 1. And so my answer is just  $n$  plus 1 and  $2/3$  to the nu of  $n$  which is the same thing by thinking about it intuitively.

So the point is that the operation is friendly, and so the idea then the big picture was convolution is a different way to represent a system. Using convolution, we represent an entire system by a single signal. That signal, the unit sample response, is sufficient to characterize the output of the system for any possible input. We just saw how the operation is called convolution, so that enables us the big picture. We've represented an entire system by a single signal, in this case  $h$  of  $n$ . That's what convolution is. It's a new representation.

You can do exactly the same thing for a CT system, and the reason we use delta to represent the unit sample signal and the unit impulse response. The unit impulse signal is clear, because the representation of an arbitrary signal, in terms of delta functions, looks much the same in CT and in DT. You can get there by thinking about the limiting argument for how to interpret the unit impulse. The unit impulse function was a function that is easiest to think about in a limit.

Imagine that I have a signal that is a square pulse whose area, regardless of width, is 1. That's what a unit impulse function is. Imagine how you would construct an arbitrary signal  $x$  of  $t$  by having such a signal. You could take a signal and its shifted version, and come up with a weighted sum of impulse functions or rectangular approximations to impulse functions to represent an arbitrary signal. If you did that, you would get an approximation to the signal  $x$  which could be written as a limit.

So if you think about each of these being of width  $\Delta$ , then the height has to be  $1/\Delta$ . So the area remains one. Then if I want to build an arbitrary function  $x$  out of such signals, I need a sum of them, and each one of these  $p$ 's has to be multiplied by  $\Delta$ . So that when I multiply by the value of  $x$  at one point  $k\Delta$ . I get the right height independent of what  $\Delta$  is. So for a given  $\Delta$ , I get a sum that looks like that, and then in keeping with the idea of thinking about a unit impulse function as a limit, I take the limit of that.

The result is a function that looks very much like the decomposition of a signal in terms of the unit sample. In the previous case, we sum together a weighted version of a unit sample signal. Here the sum is replaced by an integral, and is weighted just like it was before. The point is that the mathematics for CT and DT look very similar. I decompose in the case of the CT, an arbitrary signal  $x$  of  $t$  into an entire row of weighted unit impulse functions.

Once I have it in that form, the argument's precisely the same for CT as it was in DT. Imagine that I have a linear time invariant system. Linear means that I can compute the response to a sum as the sum of the responses. Time invariant means that shifting the input merely shifts the output. Doing the experiment tomorrow is the same as doing the experiment today, except it's now a day later. So if the response of a system is  $h$  of  $t$  when the input is  $\delta$  of  $t$ , if the system is shift invariant, shifting this by  $\tau$  is the same as shifting that by  $\tau$ . A weighted sum of such things is a weighted sum of such things, and a sum of such things is the sum of such things, so I get an expression which we'll think of as convolution for CT that looks just the same.

So in DT, we thought about if you convolve  $x$  with  $h$ , you take the first index,  $x$  of  $n$ , and turn it into  $x$  of  $k$ , a dummy variable. You take the second one, and do  $n$  minus  $k$ . Here we do the same thing.  $t$  goes to  $\tau$ , and the second one goes to  $t$  minus  $\tau$ . The sum up here turns into an integral. Otherwise, it's exactly the same thing. So to show your mastery of such things, what signal would result if you convolved  $e^{-t}$  with  $e^{-t}$ ? 1, 2, 3, 4, or none?



Well, the place is quiet so I assume that means you stopped talking, so that means you've all agreed, yes? So which wave form best represents the convolution of the top two signals, 1, 2, 3, or 4? Almost 100% correct. Most people say 4. How do you get 4? Yeah?

**AUDIENCE:** Same reason [INAUDIBLE]

**PROFESSOR:** So what would I do? What would be my first step? So I imagine that I want to think of flip so this gets multiplied by the flip of the other one. So at time  $t$  equals 0, the answer is--

**AUDIENCE:** 0

**PROFESSOR:** --0, because there's no overlap.

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** OK so that it starts at 0, so that means that this is out. OK, OK, OK, fine. Now shift. What do I shift? Which one do I shift which way?

**AUDIENCE:** Why is there a [INAUDIBLE] flipping is there a value that equals 0?

**PROFESSOR:** That's a valid question. So the question is if I'm integrating over a function that goes from minus infinity to 0, and from 0 to infinity. Let's say the answer right at 0 is 1. You could say that there is a single point whose value is non-zero. What would happen if I integrated a function that is 0 everywhere except at a point? So it's 0 everywhere up to here, then a 0 everywhere after that, and at zero it's not zero. What's the integral of a function that differs from 0 at a single point?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** 0, it's a little bit of a trick question, because we will later have some functions for which that's not true. What kind of a function would that not be true for?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Delta. If I were convolving delta with delta, then you can integrate over an infinitesimal area region, and get something that's not 0. So a little bit of a caveat, so as long as the function that I'm convolving doesn't have an impulse in it or worse. We will talk later in the course about things worse than impulses. If there's nothing as high as an impulse or worse, so that has a

step in it. We would think of a step as a singularity that is better, less ill behaved than an impulse. As long as the function does not have an impulse or a worse, when you flip it, you'll get zero contribution at zero. That all makes sense? So this is zero at zero, but the reasoning is a little bit complicated.

So now what do I get when  $t$  gets a little bit bigger than 0? What's the result of convolving when the time is slightly bigger than time 0? All flip, shift, multiply, integrate. So I have to shift one of those, so now instead of having this one, I might have shifted a little bit to the right. So it might look like that. So now what happens when I multiply? Well you don't get 0 anymore. So as for small  $t$  for  $t$  on the order of  $\epsilon$ , how does the integral grow with  $t$ ? So if I want to make a plot of the convolution, so if I want to think about  $e^{-t}$  convolved with  $e^{-t}$  versus  $t$ , I already know that that's like that for  $t$  small. How will the function grow?

Linear so if this is very small then the deviations from the height which is 1 is very small. So if this distance is small, the deviation is that little triangle which goes like  $t^2$ . So for  $t$  small  $t^2$  is very small compared to  $t$ , I can ignore it, and so the function is going to start going up like  $t$ , and if you work out the details it will eventually roll off. Because as you shift it further and further, the one exponential is in the tail of the other exponential, so one of the exponentials kills the other one, and so the response goes to zero. If you're more mathematically inclined, yes?

**AUDIENCE:** [INAUDIBLE] you were saying that if both functions are left sided [INAUDIBLE] right sided it always starts out  $t=0$  is always 0.

**PROFESSOR:** That's not quite right because right sided just means that the left is 0. So if I tell you that a signal is right sided, all I've said is that all of the non-zero values are on the right, but I haven't told you whether they're impulses or not. Right sided says something about the left. The left is zero. Kind of weird, so if I'm right handed, I might as well not have a left hand when I'm writing so the left is zero. So right sided signals have zeros. The signals on the left of right sided signals are zero, but I haven't told you what was on the right. The right could be an impulse or worse.

So I just inferred some properties of what this convolution is going to look like. If you were mathematically inclined, you could do it by math. The math doesn't look very different from the math for the discrete version. You simply write this as a function of  $\tau$  rather than  $t$ . This one

as a function of  $t - \tau$  rather than  $t$ . Recognize that the  $u$ 's cut off parts of the integral. This  $u$  is 1 only if  $t$  is bigger than 0. So that lops off the  $t$  less than 0 part. This lopped off the part bigger than capital, bigger than  $t$ . So that leaves the integrals 0 to  $t$ . Putting these two together results in the  $\tau$  parts killing each other, so I'm left with only a  $t$  part but the integrals on  $\tau$ . So just like the other one, the integral goes to the integral of 1 over finite limit 0 to  $t$ , so the answer is  $t$ . So the overall answer is  $t e^{-t}$  which is plotted here.

So the point of today then is that this is a different representation for the way systems work. It's often computationally interesting. This is the way. This is a perfectly plausible way of doing discrete time signal processing. Represent the system by a signal  $h[n]$ , and then compute the response to any signal by convolving  $h[n]$  with that signal. Perfectly reasonable way to compute things. Honestly, we'll find better ways of computing things by the end of the course.

The real reason for studying convolution is conceptually, you can think about how a system ought to work by thinking about convolution, and I want to show an example of that by thinking about systems that I'm interested in. I do work on hearing. I do work with microscopes, and we can regard a microscope as an LTI system. A linear time invariant system, and convolution is a very good way of thinking about such optical systems. So the idea is that even the best microscope gives blurry images, and that's very fundamental physics. It has to do with the diffraction limit of optical systems. If you're interested in that sort of thing, take an optics course in the physics department or come to my lab and do a [INAUDIBLE].

So the idea is that even the best microscopes in the world are blurred. We have the best microscopes in the world in my lab, and they fundamentally blur things, and we have to worry about that. The blurring is inversely related to the numerical aperture which has to do with the size of the optic.

Big optics are good. So if you imagine that a target emits a spherical wave of light. So every point on the target emits a spherical wave, then there's some optic that's collecting all of those waves, and relaying them back to a different point. The resolution of the picture goes with how many of those rays the optic system picked up. So if you make the optics smaller, the picture becomes blurrier. If you make the optic even smaller, the picture becomes even smaller, and the way we think about that is by convolution.

We think about the microscope as an LTI system, so we characterize it by its point spread function. We don't like to use any words that come from a different field. Just like every other

field, we like to invent our own. So in optics, the thing that we will call a impulse response is called a point spread function. It just says if you had an ideal point of light, what would the image look like? It's exactly the same as convolving, so you can think of the blurry image as the convolution of the effect of the microscope, the point spread function, the impulse response with the ideal target. So then as you change the size of the optic, it changes the size of the impulse response, the point spread function. Crummy optics, fat point spread motions. Fat point spread functions, blurry pictures, and so here's a picture of how our system works.

This is a representation to scale of a tiny microscopic bead a fraction of a micron. So it's about six times smaller than the image. So this is an image taken with our microscope system, and you can see that most of the energy fits inside a region about half a micron. World's best microscope, you can't do better than this by physics. This is using 500 nanometer light, and the size of that has to do with the length scale of the light. So you end up with this particular microscope not being able to make images with less blurring than that. That's the point spread function.

Now of course, the point spread function of a microscope is three dimensional. In this class, we're only talking about 1-D time. In a microscope is 3D, x, y, and z. So the impulse response has extent in x, y, and z. So here is a picture taken by Anthony Patire, who was a student in my lab of that same tiny little dot. When the in-focus plane shows the dot, and as you go out of focus, the dot gets bigger, smearier, and blurrier. What you can do then is assemble those pictures that were taken one of the time into a 3-D volume, and that 3-D volume then represents the point spread function or the three dimensional impulse response of the microscope.

So the idea then is that convolution is a very good way to think about optical systems because they are very easy to relate to the underlying physics. The blurring is a direct result of a fundamental property of light. The diffraction limit, and you can very accurately represent the effect of the blurring as convolving with a point spread function and impulse response.

Same sort of thing applies to optics at any scale. So going from the microscopic to the rather macroscopic-- that is to say the universe and beyond. We can think about the Hubble Space Telescope. Same thing. Light, that's all that matters, and here the issue is-- the reason they wanted to make a Space Telescope is that there are two principal sources of blurring for a ground based telescope. One is the atmosphere blurring, because we're looking through an atmosphere and other is blurring because of the property of the optic elements in the

telescope, and it turns out pretty easy to show that the combined effect of the atmosphere and the lenses is the convolution of the individual parts.

You can think about atmospheric blurring as convolving with the atmosphere's point spread function, and you can think of the blurring due to the microscope as convolving with the blurring function due to optics. The combined is the convolution of those which means that if you've got some amount of atmospheric blurring and a telescope made out of 12 centimeter optics, then the combined responses showed here not very different from each individual. But if you made a big telescope-- a one meter type telescope-- you might be expecting this much blurring. But because of the atmosphere, you get much more blurring. So the deviation between the small telescope, and what you actually measure is not very different.

The atmosphere makes an enormous difference when you start talking about a high resolution telescope. That's the reason for putting it in space. You get rid of the atmosphere. The Hubble Space Telescope was made principally out of two big mirrors, both were parabolic, both were highly optimized, both were enormous. This is the main lens. The lens is 2.4 meters, about eight feet in diameter, and it was astonishing the thing.

So in order for a lens to work perfectly like I illustrated, it's important that every reflection remain in phase coherence. So the length that every ray travels has to be precisely matched to what it's supposed to be. In the case of the Hubble, they matched the surface. The surface was controlled to within 10 nanometers. That's absurd. So the blurring of my microscope was about half a micron. So 50 times worse. The best I could see with my microscope is 50 times worse than was required for making this mirror.

It was absolutely astonishing feat to make the mirror and they made a mistake. So when they put this in space, they were expecting to see pictures like this. This is a picture of a distant star. They were expecting the distant star would look like this. This is what they actually measured, and the reason was that the feedback system that they used to grind the lens made a mistake by 2.2 microns. 2.2 microns would have been just barely resolvable on my microscope, but barely. So the hair? That's about 100 microns in diameter. They were off by 2.2 microns, and because of that, it was like a complete disaster. So that small error was enough to make the images terrible.

So the solution-- it wasn't very practical to ship up a new lens, so they shipped up eyeglasses. The eyeglasses was another transformation, just like your eyeglasses. Your eyeglasses work

by changing the point spread function that is determined by your retina and your lens into a new point spread function. They shipped up eyeglasses, and the result of putting the eyeglasses into Hubble was to turn this into that, and to give some of the most dazzling pictures we've ever had. So the point is that convolution is a complete way of describing a system. It's a very intuitive way for certain kinds of systems, and it's especially useful for systems like light based systems where blurring is a natural way of thinking about the way the system works. Have a good time. See you tomorrow at 7:30.