

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 to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR:

We started to talk about the discrete time Fourier transform last time. This is probably something you've been exposed to before, and as in many cases, we're looking at it in a very different way than what you probably looked at it before. The discrete time Fourier transform is simply the time frequency dual of the Fourier series. Nothing more than that. For any L2 function -- now we're talking about a function of frequency, but a function is a function. So this is a complex valued function of frequency. If it's limited, if it's truncated to the band from minus w to plus w , then, in fact, its inverse Fourier transform is given by this limit in the mean of the sum of the coefficients times v sub k of f . This is exactly the same as the Fourier transform replacing times with frequencies, replacing w for t over 2 , and in the complex frequency you're replacing a to the plus $2\pi i$ here with e to the minus $2\pi i$ in the Fourier series.

So those are the only differences, it's just notational differences, and aside from that, it works exactly the same way. The coefficients are given by this. We showed, when we were talking about the Fourier series, that these coefficients exist as complex numbers, they're always finite. You can calculate them if you want to. This quantity here can be rather fishy. This is this limit in the mean which says that you have to calculate this by looking at the sum here over a finite sum, over a finite sum this is well-defined and behaves very nicely. As you go to the limit funny things can happen, but the thing that we showed is in terms of energy, nothing funny can happen.

I'm going to give you an example of this kind of funny business as we talk about the sampling theorem in just a couple of minutes. So the \hat{u} of f has to be L1 since it is limited in this way. It has a continuous inverse transform, which is this. OK, so you can go from -- blah blah blah blah blah. The discrete Fourier transform is simply a

transform between a function and a sequence of terms. Now this bit here is something we haven't talked about before. Because now we've talked about Fourier transforms also, and if you have an L2 function, \hat{u} of f , that L2 function has a Fourier transform and an inverse Fourier transform in this case, which is u of t , which is given by this expression here, the usual expression for the Fourier transform.

Now, the thing that's peculiar about this Fourier transform is that since \hat{u} of f is limited, is truncated, this transform here exists everywhere. We don't need a limit in the mean here. You can calculate this for every t this exists. This is a point-wise convergent thing. This is the thing that you're used to when you think of functions, because the function is defined everywhere in this case. And since we're going to go into the sampling theorem, it had damned well better be defined everywhere because otherwise the sampling theorem wouldn't make any sense at all.

So, the inverse Fourier transform of the discrete time Fourier transform, as we said, is this, it's a limit in the mean. Namely, you take more and more terms here. You get closer and closer to this in terms of energy. It doesn't say anything about what happens for particular values of f . This is a sampling expansion with t equals 1 over $2w$. OK, let's go back here. We're talking about some function of frequency which has a Fourier series, it also has a Fourier transform. What we're interested in now is what is the relationship between these coefficients here and a discrete time Fourier transform and this function here. What I want to show you is that, in fact, these two things are very closely related. You can now go back to the Fourier series itself and relate the Fourier series coefficients to the Fourier transform and you'll get the same kind of sampling representation that we're getting right now.

So, if we -- there's something missing in what I'm trying to say here. Oh, I see what I'm trying to do, sorry. What I'm trying to do is to take the inverse Fourier transform of \hat{u} of f , which is given by this expression here. Temporarily I'm going to forget about the fact that this is a limit in the mean, throw mathematics to the winds, and simply take this inverse transform. Take the inverse transform of this also. So, when I take the inverse transform of the sum, what I'm going to get is the sum over k of u

sub k , and then in place of this frequency function -- these are orthogonal functions here, the things I listed on the previous page -- I got u of t is the sum of u sub k times these time functions now.

These time functions are just the Fourier transforms of these frequency functions here. This is a set of orthogonal wave forms here, which are truncated sinusoids. I want to take the Fourier transform of these and the Fourier transforms of these -- I should have put them both on the same slide so you could see what they are -- but in fact, in your homework you're going to take the inverse Fourier transform of that and show that it is, in fact, this. That's just a nice exercise in taking rectangular functions and sinc functions and applying shifts in both time and frequency. When you do this, this turns out to have the inverse transform of this. So this is u of t is just this, where this comes from here, and then this turns out to be the inverse transform of this function here, which, in fact, is just this. Sorry for all of that.

So now if we want to try to understand what this means, suppose that you take a function u of t , which is, in fact, the inverse transform of this u hat of f , which is truncated to a finite band limit. If we then take v of t equal to u of t everywhere except on the sample points, and on every sample point we simply add 1 to v -- in other words, v of kt is equal to u of kT plus 1. So we take this nice smooth function that we have here, and at every sample point we simply add 1. So now the question is, is this new function I have still baseband limited or isn't it? You see you can't answer that question because we weren't careful enough to say what we meant by a baseband limited function. Usually when you talk about a baseband limited function, you're talking about a function whose Fourier transform is zero, except in the range minus w to plus w .

Well, this new function v of t here has the property that its Fourier transform is zero outside of minus w to plus w . And therefore if you define baseband limited as functions whose Fourier transform is zero outside of limits, then the sampling theorem doesn't hold. So what do we do about this? Well, the easiest thing to do about it is to change what we mean by baseband limited to what you would have meant if we hadn't going through all of this mathematics. In other words, the Fourier

transform cuts both ways -- u of t has a Fourier transform, \hat{u} of f , \hat{u} of f has an inverse transform, u of t .

What we mean now by baseband limited is that u of t is the inverse transform of a frequency function, which is limited to minus w over plus w . In other words, we will regard u of t here as being baseband limited, we will not regard v of t as being baseband limited. Because we have these two functions, u of t and v of t , which both have the same Fourier transforms, but which are not equal to each other. We want to, at this point, say that the only one which is really baseband limited is the function which is the inverse transform of u of t . If you read the sampling theorem as it's written in the notes, that's exactly what it says. It defines this function u of t to which the sampling theorem applies in that way.

Well anyway, this is the sampling theorem at this point. It says you can take u of t , you can express it in this way. If I go back to the previous slide, what we did here by comparing u of t with this expression for u sub k . u sub k are the coefficients in the discrete time Fourier transform. u of t is this inverse transform. These quantities here are almost the same. This quantity here is just evaluating this at particular frequencies. Namely, if for frequency f I substitute in, if we're time t , I substitute in k over $2w$, then this formula just becomes that formula. In other words, $2w$ times u sub k is equal to u evaluated at k over $2w$. You already saw that when you looked at the Fourier series. When you looked at the Fourier series, you saw these coefficients, which, in fact, look like the Fourier transform terms. And which, in fact, were the same as the Fourier transform except for a scale factor at some particular frequency.

Here we have these coefficients, which are, in fact, scaled versions of the inverse transform of particular times. So the conclusion from that is that, in fact, what the discrete time Fourier transform is is it's simply the Fourier transform of the sampling theorem expansion. The two of them are duals in a very different way than the Fourier series and the dtft are duals. The dtft and the Fourier series are duals in the sense that if you take the expressions for the Fourier series and change frequency for time and time for frequency, you get to the dtft. Here what we're doing is taking

the dtft and simply taking the inverse Fourier transform of it, so that the sampling theorem is, in fact, the Fourier transform of the dtft. It's not the dual, it's the Fourier transform itself.

Well, the discrete time Fourier transform generalizes to arbitrary frequency intervals just as well as to a baseband interval. Namely, you can do exactly the same thing as what we've just done if you're looking not at the frequency at the range of frequencies from minus w to plus w , but you shift that up to any old place you want to and look at δ minus w to δ plus w . And the fact the discrete time Fourier transform, if you don't put the rectangular function in it is going to be periodic anyway. It's exactly the same thing that we had with a Fourier series. With a Fourier series, we could find the Fourier series for a function limited between minus t over 2 to plus t over 2. Now by duality, we can find the dtft for a function limited between minus w plus δ and plus w plus δ . That's what we're doing here. So the dtft in generalized form is now this, and v sub k is now the integral from δ minus w to δ plus w of the same old thing as before. This is equal to this, which is the same old thing as before, except we now have this shifted frequency in the rectangular function. So if we take the inverse Fourier transform of this, again, using the same duality we had before, we get v of t is equal to the sum times the sinc function.

And the only difference now that we're expanding this given frequency band not centered around zero but centered around something else, the only difference in the sampling theorem is now we have this rotating term gyrating around up at this frequency, k over $2w$. That's the only way in which this is different from the sampling theorem that we had before. I wish I could put more things on one slide but you wouldn't see them if I could. So here's a Fourier transform of the sum 1 over t , u of kt time the sinc function. Here it's the same thing. That's one reason for comparing these things sometimes. 1 over t in here. OK, times, so it's the same thing with this rotating term which is the only difference.

Now, how many of you can see that the Fourier transform of this quantity is equal to this sinc function? I can't do that. It's one of the things you're going to do in your homework. I get confused every time I do it, and I got confused enough to leave out

the $1/t$ this time when I did it. You just have to be patient with that and do it a few times and you'll find that it's not -- well, it becomes automatic after awhile. It's one place where you need plug and chug.

So now that we've generalized the dtft to look at any old frequency band instead of just the frequency band around zero, we can do the same thing that we did with time functions. Namely, with a time function -- Yes?

AUDIENCE: [INAUDIBLE].

PROFESSOR: You don't think it should be a $1/t$? Well, you very well might be right because I didn't think it should have been either when I wrote it down. But I don't see --

AUDIENCE: [INAUDIBLE].

PROFESSOR: It was the--.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Oh, that's the difference, yes, of course. It's not the coefficient that I have here, it's the actual -- yeah. Yes. So it should be the same as the-- Oh, I see the problem. I shouldn't have had the $1/t$ here, should I? No. I know one of them couldn't be right. It is right in the notes, so you can sort it out there.

So, the thing we did when we were dealing with a Fourier series is we took an arbitrary function of time, we segmented it into time intervals and then we expanded each one of those time intervals into a Fourier series. By doing that we could take an arbitrary L_2 function and represent it as an orthogonal expansion over this double sum of time shifts and frequency terms. We call that the truncated sinusoidal expansion, the t spaced truncated sinusoids and we made an expansion out of that that would allow us to express any old L_2 function in terms of that.

Here we're going to do the same thing. We can take an arbitrary frequency function, separate into bands of frequencies. You often want to do this in digital communication when you're looking at transmitting information in different bands, which you do all the time in radio. Somebody has a certain part of the spectrum,

they transmit a signal there, somebody else has another part of the spectrum, they transmit a signal there. You can look at those different signals which are in different frequency bands, they're all orthogonal to each other, they don't interfere with each other at all. We're doing the same thing here. We're just saying an arbitrary function can be split into different frequency bands, each one of those frequency bands can be represented both by a dtft, which is the thing we just did on the last slide, and by sampling theorem expression, which is what we get when we take the inverse Fourier transform of the dtft. So when we do that what we get is a perfectly arbitrary frequency function which exists from minus infinity to plus infinity. We can represent it as the sum of all these separate frequency functions. I just threw a limit in the mean here because I'm not being careful about what happens where we separate from one frequency function to the next. Namely, at frequency ω do I use one term or do I use the other term or do I use the sum.

But we don't want to worry about that. We don't want to even think about it. So we put a limit in the mean here. So the $v_{\hat{m}}(\omega)$ then is going to be the part of u of f which is in this particular frequency range. That's completely the analogy of taking a time function, looking at that time function over a particular range of time, and here what we're doing is taking a frequency function, segmenting it into different frequency intervals so that the end frequency interval is then just this with a rectangular function to truncate it. If I take the inverse Fourier transform of this what I'm going to get is u of t , take the inverse transform of all of these terms, so I'll get the sum of v_m of t . Now, v_m of t is the inverse Fourier transform of v_m of f . You take that quantity, take the inverse transform of it, and sure enough you get this kind of expression here. It's a sampling theorem in $v_{\hat{m}}(\omega)$ with this rotating frequency term here, which is just the thing that we had before.

So, all we're doing here is starting out with some arbitrary frequency function, we're segmenting it in frequencies. Each frequency band then has a dtft associated with it. When we take the inverse Fourier transform of that dtft, what we get is a sampling expansion for that particular frequency band. So what this is doing here, finally when we get all done with this, is I'm just combining the sampling theorem

expansion in each frequency range, which is what u of t is. So I've taken u of t , I split it up into different frequency ranges, I've expressed what's in each frequency range in terms of a sampling theorem. The sampling theorem terms are these with these rotating terms in them corresponding to the m th frequency range. This is completely analogous then to the truncated sinc function expansion we had before.

This becomes a little more sensible if we substitute a sampling time, t , for 1 over $2w$. Namely, all of these expressions here are talking about what happens when you sample these individual frequency bands at intervals 1 over $2w$. So we'll just call that capital T to make the formula look a little simpler. Then we get u of t is this limit in the mean of this whole expression there. So it's a double sum, it's a sum over time, over the samples, so there's one term for each time, kT , and there's one term for each frequency interval. Frequencies are indexed by m , time is indexed by k . So the thing we have here is an expansion now in terms of coefficients -- these are just called coefficients again. This expansion which looks suspiciously like the t space truncated sinusoids that we had before. The only difference is that, the terms were truncated in time; here, the terms are truncated in frequency. So the different terms making up this expansion, these orthogonal terms here, in one case what we have is a sinc function which is translated in time and then translated in frequency. In another case we have the rectangular function, which is translated in time and then translated in frequency. So in that sense, these two expansions are almost the same, and you can think of doing expansions perhaps in other things also and we'll talk more about that later.

So, this then is just this thing we're going to call a t spaced sinc weighted sinusoid expansion. So the only thing we have is this one sinc function which is this hat sort of function. The terms in here are those functions shifted in time by some number of sampling intervals, t , and then shifted in frequency by some number of frequency bands, $2w$. See, the original frequency bands that we had went from minus w to plus w . The next one goes from w to $3w$, the next one goes from $3w$ to $5w$. So the frequency bands we're talking about here are of width $2w$. The time intervals we're talking about are of width t . Why do people confuse you that way? Well, because all of this happened a long time before people realized how closely the duality

relationship between time and frequency was.

So people wanted to talk about frequencies, baseband frequencies. You talk about a baseband limited to w and you're talking about positive frequencies, because engineers used to deal with cosines and sine, and there weren't any such thing as negative frequencies. Then they decided everything was easier when they dealt with complex sinusoids, negative frequencies reared their ugly head, but people didn't want to change their notation for what a frequency band was, which is probably good. So we're simply stuck with this incompatibility of dealing with frequencies one way and dealing with time the other way. We can look at that as increments of time t , and increments of frequency, 1 over t , but, in fact, 1 over t is $2w$, so the increments in time we're using in both of these expansions, are t , the increments in frequency we're using are $2w$.

Now, there's a relatively long section in the notes talking about degrees of freedom, which is a pretty important topic. It's a little bit fishy mathematically, but it really makes good engineering sense. It's an idea which is important both in terms of taking source wave forms and representing them in orthogonal expansions, and in taking frequency functions and representing them -- well, it is also important in terms of taking things that we transmit where you have bits coming into an encoder. We're going to turn those bits into signals, we're going to turn those signals into wave forms. Those wave forms will usually be thought of as things that we transmit in time, and we also transmit them in frequency, because we often use some kind of multiplexing between different frequency bands. We want to have a common way of thinking about all of these things, and this is the way that we're going to do it. Namely, if we're thinking in terms of a particular sampling time, t , and we want to look at a very, very large frequency band, and therefore, look at many multiplex frequency bands, we can say how many coefficients can we send on this channel? When we look at it in these terms of different frequency bands, the number of coefficients we can send is over a period of time, t zero. We can send t zero over t different coefficients in time. We now look at frequency. We have some very broad frequency band, w zero. The number of different bands that we have is w zero over

2ω .

As a result of all of this when you add everything up, you get 2τ zero ω zero degrees of freedom over this overall bandwidth of ω zero. Now remember, an overall bandwidth of ω zero in terms of these complex frequencies goes from minus ω to plus ω . Minus ω zero to plus ω zero. The time interval goes from minus τ zero over 2 to plus τ zero over 2. So this factor 2 here, which we always talk about in terms of number of degrees of freedom, is really a consequence of the fact that we measure time intervals and frequency intervals in a slightly different way. But anyway, whether we look at it in terms of one expansion or the other expansion the answer we get is the same. If you take some large time interval, some large frequency interval, tuck as many numbers as you can in that interval, this is what you come up with.

Now, why did I say that this is just slightly fishy? Well, it's slightly fishy because if you take a function and you truncate it in time -- if we take this function and truncate it to minus τ zero over 2 and plus τ zero over 2, even though that might be ten years, how can we limit the frequency? Well, we can't. Because when we take the Fourier transform of a time limited function, it exists for all frequencies. The same thing happens if we try to limit it in frequency, it squirts out forever in time. So you can't get around that. The thing that saves us is that if τ zero and ω zero are both large enough, these functions all dribble away quickly enough that it doesn't make any difference. You know it has to.

If you think in terms of the sampling theorem, and you try to think about it carefully in mathematics, what does it say? If you want to transmit a function by putting these little sine x over x hats around each of the coefficients in the function, when do you have to start transmitting those wave forms? You have to start transmitting them at minus infinity. I mean we turn on our transmitter and we somehow have to have been transmitting for an infinite amount of time before we send the first symbol. Well that's ridiculous, of course. So that we always approximate these sinusoids by sinusoids which are truncated, and we always have some engineering faith that what we're throwing away is not important. The only place that it's important is when

you start talking about things which are really zero everywhere and then it becomes important. But the idea of degrees of freedom is a very sensible idea until you try to express it precisely.

Let's get on to something called aliasing. We're going to spend most of the rest of today talking about aliasing. I want to try to explain why it is that we want to spend time on this, because there are really two things going on here. One of the things that are going on is that if you want to look at a wave form and do some processing on it, the usual way to do it with the digital technology we have today, is to take that wave form and sample it very, very rapidly in time. Then process the hell out of all those samples. We do that hardly caring whether it's band limited, hardly caring about the information in it, hardly caring about anything, we just want to sample it so fast that we essentially approximate the function very well. Now when we do that something's going to get lost, because when we take those samples we're ignoring what happens between the samples. In some approximate sense, wave forms are always smooth because they always get filtered by something before anybody looks at them. And because they're smooth, if you'd sample it fast enough, your fast enough samples are going to look like the wave form. If you just connect them with straight lines you're going to get a very good approximation of what the wave form is.

But then you stop and ask, and when you stop and ask you're in trouble because you say well, how fast do I have to sample, and if I sample that fast, how much error am I going to make? That's the question for which aliasing gives you the answer. So we want to explore it for that reason. The other reason that we want to explore it is when we start talking about modulation, we'll start talking about something called Nyquist criterion, and that's best looked at in terms of aliasing again, and we'll see why that is when we get there. So for both of those reasons, and also for the reason of trying to understand these expansions in terms of source wave forms, we want to understand what this relationship is between the samples of a function that isn't quite band limited and the function itself.

So the thing we're going to do to try to understand that is instead of studying just the

samples, which is what you usually do, if you're just looking at the samples and you're trying to say how much error do I incur by doing that, and we're looking at mean square error, somehow or other we have to get back to wave forms and compare the resulting wave form with the wave form we started with. So the thing that we're going to do is to take our function u of t , we're going to sample it at some very rapid speed, and then we're going to take those samples and recreate a wave form by the sampling theorem. So we're going to call the approximation to u of t some approximation s of t -- s of t is, in fact, baseband limited at this point to a frequency w where t is equal to 1 over $2w$, and we simply have this sampling theorem applied to u of kt , as if u of kt came from a band limited wave form.

So this is, in a sense, an interpolation formula. It's a little better than taking these samples and joining them by straight lines, because we're, in fact, joining them by these smoother sinc functions. The question is how close is s of t to u of t and how do we look at that question? Well, we now have a nice way of looking at it because after going through this t spaced sinc weighted sinusoidal expansion, we have an exact expression for u of t , which is the limit in the mean of these frequency terms in u of t . Namely, to get this expression, remember the thing we did was to take arbitrary u of t , split it up into little frequency bands, each of width w , then apply the sampling theorem to each of those frequency terms, which we can do exactly. So this, in fact, is an exact expansion -- don't worry about the limit in the mean right now, we're going to get rid of that in a while.

So we have s of t , which is this. We have u of t , which is this. Now, we look at this and we say OK, s of kt , namely, the case sample of s of t , is simply u of kt . That's what happens here. You put in any old time which is some integer j times capital T and you look at this expression here, and the sinc function is only non-zero when j capital T over t , when if you take the sinc function of an integer it's zero unless the integer is zero. The sinc function goes through zero at every integer point except for zero itself where it's equal to 1. That's the thing which makes the sinc function nice. It dribbles away -- 2, 3 and so forth minus 1. So it's zero at all those sample times, so that the case sample of s of ts of kt is just equal to u of kt . Namely, what we're doing in this approximation, which is what one usually does if one samples

something, is we're assuming that the approximation is correct at the sample points, and we're arranging it so it's correct at the sample points.

So, s of kt then is equal to u of kt from this. u of kt from this -- OK, now take t and substitute k times capital T in here. And if we substitute k times capital T in here -- let's not, let's substitute j times capital T to avoid a conflict with notation here. Then this T here becomes jT and what we have is sinc of j minus k . So we have a sum here over all k of sinc of j minus k . Sinc of j minus k is zero for all integers k , except for j , therefore, this quantity is zero every time k is not equal to j . Therefore, we just have the sum over m . So, s of kt is equal to u of kt , which is equal to the sum over all frequency bands, m , of v sub m of kt .

Now what is this saying? The thing this is saying is if you take this function u of t , which has a bunch of different frequency bands in it, each of those frequency bands has a sampling theorem associated with it. Each one of those frequency bands is represented by its samples at periods of time, k . But as soon as we look at u of kt , if we only have the samples u of kt , there's no way to tell which frequency band it's coming from. So all of these different frequency bands all get alias together. If I just look at u of kt , what it is is the sum over all these frequency bands of the samples of the individual frequency functions. So that if you tell me what u of kt is, I can't tell you what these samples are. All I know is what the sum of them is. So, in fact, if you start out with a function which instead of being baseband limited to w , in fact, is sitting between w and $3w$, and there's nothing in this baseband, and I look at these samples and then I recreate things this way, what am I doing? I'm just taking that function at w to $3w$ and translating down in frequency to minus w to w . You can't tell. There's no way to tell just from the samples which frequency band we're looking at. So all these things get mixed together.

So, s of t then, since s of t is this times these sinc functions, and this is the sum of all the v_m 's, s of t is just this double sum now where all of these are now down at this baseband frequency interval. There's no way to tell them apart. We have this double sum here so I'm adding up all of these different coefficients and they're all mixed together all down in this one frequency band.

So, u of t is represented this way, double sum, v_m of kt , v_m of kt , sinc of t over t minus k , sinc of t over t minus k . This rotating term up here, an s of t , what I've done effectively is to get rid of all these rotating terms. It simplifies it enormously, but I changed the function. In fact, this function is low pass limited to w and this function has all of the glory of an arbitrary set of frequencies in it. By just looking at these samples, I've lost all of this stuff and it's just back down to a baseband limited function at this point. So, if I look at the difference between u of t and s of t , what I'm going to get -- and this is expressed a little differently than the way it is in the notes but it's the same thing -- just the sum over k and m , and it has these sinc functions in it, which both of these terms have. Then u of t has these rotating terms and s of t doesn't, so s of t just has one in place of the rotating term. So the difference between u of t and s of t is just this big monster sum here, which is looking at all the different frequency bands at all the different sampling times.

If I now try to look at the energy difference between u of t and s of t , because that's what I'm interested in -- how much error have I accumulated, how much mean square error have I gotten? By taking u of t and sampling it and then viewing it as a low pass function. This energy difference, well, we have a set of coefficients here, we have a set of functions here. These functions are all orthogonal to each other. Why are they orthogonal? Well, because sinc functions are orthogonal, which I think we've shown on our homework, and space time functions are orthogonal. So, all of these functions are orthogonal to -- excuse me -- spaced frequency functions are orthogonal to each other, and the sinc weighted functions are orthogonal to each other. So all I'm doing here is expanding all of this in terms of these orthogonal functions. I have to do this separately for these terms -- I know what happens. This works when it's cold and it doesn't work when it's hot. Interesting. So I'm separating this into this term and this term.

Now, look at what this is when m is equal to zero. When m is equal to zero, e to the $2\pi i$, zero t over t minus 1 is zero. So, all of the m terms here are collapsed into zero, so there isn't any error down at baseband in a sense. All of the error occurs due to these frequency terms larger than minus w to w . Well, that's the way it should

be. Because we know that if u of t didn't have any terms outside of minus w to plus w , the sampling theorem would be absolutely rock solid with no error. So the errors are due to two kinds of terms. One, they're due to these terms, which is this thing. Two, they're due to these terms, which is this. The only difference between this is that this is a square of a sum and this is a sum of squares. Sum of squares are nicer, we can deal with them more nicely, and we understand what that is better. A sum and then taking the square of it after we take the sum is a good deal dirtier. The trouble is that this difference doesn't have to be L_2 . In other words, the mean square error we got from doing this can be infinite. One of the problems in the current problem set is a nice simple example of this.

So, s of t need not have finite energy. That's a real kicker, because all of this theory is very nice until this point. I mean the sampling theorem works perfectly because when you're dealing with baseband limited functions -- for up baseband limited function, the sinc functions give you a perfect approximation. Namely, the sampling theorem works with no error at all. As soon as you get a function which spills out into higher frequencies, you can, in fact, have these samples coming up with infinite energy in them, and at that point this difference here can have infinite energy. In fact, you would do much, much better if you simply represented u of t by zero. You'd do infinitely better in terms of mean square error just by throwing it away and saying I'm not going even bother to approximate it. I'll just call it zero and nothing else. If you did that, you would only have the error in u of t and you wouldn't have all this generated stuff from all these terms that you're throwing away.

When we start looking at random processes, and really the only thing that we're interested in here is random processes because noise is a random process, signals are random. So when we start looking at random processes, the thing that we're going to find is that this square of a sum in terms of expected value is going to be approximately the same as this sum of squares. So that this term and this term are going to be roughly equal for most of the stochastic processes that we deal with, so this problem doesn't occur of s of t having infinite energy. What that means is that these two terms are of roughly equal magnitude. Now, we can understand something more about these two terms. The term here is what you get, namely,

these are the parts of u of t which are outside of the frequency range from minus w to w .

A reasonable alternative to simply sampling this function would be to filter the function first. If you filter the function first what's going to happen? These terms will go away. These terms will stay. You still have the aliasing -- you can't get rid of that. But you can get rid of these terms. Excuse me. I take that back. I seem to have a binary problem today, whatever. I seem to be mixing up proof and falsehood. If you filter a function you're clearly going to get rid of all of the aliasing, because after you filtered the function you will have a band limited function sitting there. So the only error you're going to get is the error in what you've thrown away. What you've thrown away is, in fact, this quantity here. Slow down, say it right. What you have thrown away is all of the extra frequency terms. So the error is all of those frequency terms that you've thrown away. These are the frequency terms that you've thrown away. This is the error that you wind up with after you filter. These terms here are the aliasing terms. These are the things which, in fact, could be infinite if you're unlucky enough. These you can get rid of by filtering.

So now the question is should you filter first or should you just sample and say to hell with it? Well, if you think about why you want to sample and use digital signal processing, what you're trying to do is to avoid building very complex analog filters. So if you try to build a very, very sharp filter which is getting rid of all of those out of band terms, you're sort of throwing away a lot of the reason for trying to sample to start with. So the usual conclusion is no, we're not going to filter first, or we're only going to have a very crude filtering operation first, and we'll just sample a little faster if we have to, because sampling faster is easier. So we usually just sample faster and avoid these terms.

Now we want to look at aliasing viewed in frequency terms, and we've almost been doing this. We've been sort of talking around it a little bit. We're viewing an arbitrary function, u of t , in terms of a sum of these frequency limited functions. We're just arbitrarily taking a function, splitting it up into these frequency bands. In terms of frequency what we're doing is just segmenting the Fourier transform u of f . As far

as the samples are concerned, u_f of kt is just equal to the sum of the samples in each of these frequency bands -- that's what we've been saying. That's really what aliasing is at a fundamental level. So when you sample you can't tell which frequency band each of these samples come from, and they usually come a little bit from each one.

So, s of t now, I'd like to split up s of t in the same way that I split up u of t . I'd like to split up s of t , even though I can't do this from looking at s of t , I can do it mathematically. I want to split up s of t into the contributions to s of t from all of these different bands. So I want to view the m th frequency band as the sum of v_m of kt times sinc of t over t minus k . Namely, look at what s of t was to start with, somewhere back here a long time ago. What we found was that s of t is equal to this quantity here, which is the sum over time terms, the sampling terms, and a sum over frequency term. What I'm doing now is I'm just defining $s_{\text{sub } m}$ of t to be this sum in here for one particular value of m . I'm just splitting this double sum into a number of separate terms. So this is the contribution to s of t from the m th frequency band. So, s of t is the sum of all of these different frequency contributions in this quantity. v_m of t , which came from u of t , same quantity here except for the complex exponential here. v_m of t is the same quantity with the rotating term at the end of it. Namely, this is the actual signal at this m th frequency band. So the contribution to it and the sampling approximation is this, the actual term is this.

So we look at these two and we say gee, this looks like -- if we look at this in the frequency domain this looks like just a frequency shift. So the thing we can then say is that $v_{\text{sub } m}$ of f , namely the Fourier transform of this, differs from the Fourier transform of this just by a frequency shift. If we take the frequency shift formula and frequency and look at it in time it just becomes that rotating term there. Namely, frequency shifts look like complex exponential multipliers. We already know that $d_{\text{sub } m}$ of f is the m th frequency band and u of t . So in fact, it's u of f truncated to the m th frequency band. That's what this said. So, \hat{u} of f rect $f t$ minus m is just equal to this term here. So this says that \hat{s} of f , which is the sum of all these terms, is just the sum over m , sum over all the frequency terms of \hat{u} of f plus m

over t times a rectangular function of ft .

All this is saying is that s of f in frequency, it's all down at baseband now. Each of these frequency terms in u of t , up at some band and frequency, when we sample it, we're effectively bringing it down to base then. When we sample it and multiply by the sinc functions we're bringing it down to baseband. So this is saying when you look at this approximation, the space band approximation, what we wind up with is this frequency function evaluated at all of these different -- well, just evaluated at different times. So, \hat{s} of f is, in fact, frequency limited to minus w to plus w , and it has all of these different terms contributing to it. It's just looking at aliasing and frequency instead of looking at aliasing and time. In both cases it's the same thing. In one case, all of the samples get mixed together. In the other case, all of the frequency bands get mixed together. I hope this will be clearer in terms of this.

Let's take some arbitrary function here -- a most amazing thing, these things all get. It seems as if latex -- a thing for drawing pictures is fine on my screen but not fine when I print things. But anyway, this is fine as it is. It's slightly different from the one in the notes, which is also goofed up in the same way. Let's suppose we start out with a frequency function which looks like this. Actually, it should have something else there but that's OK. This is minus 1 over $2t$, this is plus 1 over $2t$. In other words, minus w to w . Here we have another frequency band from w to $3w$, another frequency band from $3w$ to $5w$. What that formula said is that each frequency band gets picked up and stuck down in the baseband area.

So, in fact, what's happening is that this part of the frequency function is going to be picked up, stuck down there. So this goes over there. This quantity here is in the band from $3w$ to $5w$. It gets picked up and put in here also. It's gotten picked up and put down there. So I have this part, this part, and this part just stays where it is -- this is the part that's actually band limited. So that \hat{s} of f now is going to be the sum of this and this and this, and one disappeared term which was supposed to be here, and that's going to come in there. When you add up all of these -- this goes over to there. When you add up all of these you get the total frequency function \hat{s} of f .

All of you understand the mechanics of this? I mean graphically you can just find $\hat{u}(f)$ from taking all of these things and folding them all into this baseband approximation. So that you can look at the error that you get in terms of sampling a function which is not quite band limited in these terms, also. The aliasing then looks like the translation from something which is in this band stuck into the main band between $-\omega$ and ω . Translation of this thing stuck into the main band. Translation of this stuck into the main band. They're all added together in such a way that you can't tell which one came from where. So once you look at this, you can't go back to here, which is why people call it aliasing. These things are all just mixed together and there's no way to get back.

The theorem that corresponds to all of this, and the theorem is pretty much proven in the appendix if you want to read it. If you're not interested in those mathematical details you're certainly welcome not to read it. I'm not going to have any problems on it. I'm not going to have any quiz problems on it, we might have a problem on it. It turns out that just L^1 and L^2 is not enough when you're dealing with aliasing. Aliasing, since you're both sampling and looking at things at arbitrarily large frequencies, the mathematics just gets messy. So the condition that you need is that the frequency function you're dealing with in order for all of these aliasing results to hold true, is that the limit as f goes to infinity of $\hat{u}(f)$ times some function, f to the $1 + \epsilon$ has to go to zero. In other words, this is saying that $\hat{u}(f)$ has to go to zero with increasing frequency fast enough. It has to go to zero a little faster than $1/f$. If it went to zero, it was $1/f$, that would be guaranteed by a thing, L^2 . That's not enough here. You need the stronger condition.

So if it goes to zero fast enough as f gets very, very large -- you know any function you're going to deal with you can always model it so that it does this. Because you simply can't transmit wave forms that have arbitrarily high frequencies in them. I mean no matter what kind of antenna you use to transmit them, if it's an optical antenna you can get to much higher frequencies, but no matter what kind of antenna you use you're going to be limited somehow in frequency, and it's going to drop off much faster with increasing frequency than this. So this isn't any sort of

practical limitation on aliasing, it's just that it's there and it limits the models you can create somewhat.

If you have this condition then it says that the Fourier transform of u of t has to be an L^1 function. The inverse transform, u of t , has to be continuous and bounded. In other words, when we go from \hat{u} of f to u of t we get a bonafide function there. It's not something which is a limit in the mean. Because as soon as we're dealing with samples of something you really need a function to talk about. You can't have something which is -- well, you can't have something which has these little extra things on it. You can't live with that. So the theorem says that this frequency function has to be L^1 . It says the inverse transform has to be continuous and bounded. We've already seen that L^1 functions have continuous and bounded Fourier transforms. For any t greater than zero, the sampling approximation, namely, s of t , which is this, is going to be bounded and continuous. And \hat{s} of f satisfies this relationship here. So this is the frequency version of the aliasing formula. Here you still have the limit in the mean here. As soon as you go back to frequencies you can't say anything precise anymore. In time, everything is precise. These functions are bonafide functions. In terms of frequencies they're not quite.

I want to just start talking about these L^2 functions. As you realize, we've not only been talking about L^2 functions, we've been talking about L^1 functions also, and now these crazy functions which go off to zero fast enough. The thing we're basically interested in, aside from sampling, is the L^2 functions because they're the ones where you can go from function to Fourier transform and back again. They're the ones which work or these orthogonal expansions, and they're the main things we're going to be interested in.

When we try to go further talking about these functions, it turns out that it's much easier to think about them in terms of vector spaces. Some of you probably have thought about wave forms before as vectors, some of you probably haven't. I'm sure all of you are familiar with vectors at least in terms of a notational convenience as a way of representing n tuples of numbers. You can always take an n tuple of numbers and instead of writing out $u_1, u_2, u_3, \dots, u_n$, you can say vector u ,

and you can manipulate those vectors, you can add vectors, you can multiply them by scalars, you can do all of the neat things that people do, even without knowing anything about vector spaces. It's not much of an extension on that to take accountably infinite sequence of numbers and represent it as a vector, especially if you've never done it before. If you've done it before and if you've thought about it before, you will realize there are some problems there.

But at least conceptually there's no problem. You can think about a sequence of numbers as being a vector, you can add those sequences, you can multiply them by scalars, you can do all sorts of neat things with them. Since we've shown how to represent wave forms as sequences of coefficients in orthogonal expansions, it then isn't much of an extension to think about wave forms as being vectors. But the nice thing about doing this is that you're not really stuck to a particular orthogonal expansion when you do this. You can think of wave forms as being vectors in just as nice a sense as you can think about these n tuples as being vectors, and that's what we're trying to get at.

These orthogonal expansions we're going to be looking at are really viewed most easily in terms of vector space. All of the questions about convergence of orthogonal expansions and things like that, limits and all of that, all of these are very natural in vector spaces. They're so natural that people often forget about the fact that they're taking limits even. So it just looks nice there. But as soon as we do this, what we're going to be doing constantly from now on is trying to say what we know when we're looking at two-dimensional vectors, we have all sorts of pretty pictures about how they behave. We are going to be trying to use those pretty pictures in two-dimensional space to understand what happens with wave forms, and in order to do that we have to know a little bit more about vectors than just vectors are things you can add and multiply scalars by.

So I will bore you for two minutes by quickly going through the axioms of a vector space. My reason for going through this is when you define a vector space axiomatically, then you can, in fact, prove that wave forms satisfy all those axioms. Once you prove that wave forms satisfy all those axioms, then everything you know

about two-dimensional and three-dimensional vectors and all of that stuff all applies. Namely, everything which is general for vector spaces applies to these vectors.

So the axioms are the following. You start out with a set of elements which you call vectors. In terms of n tuples, the set of elements is just the set of different n tuples. You also start out -- well, for here we don't have to worry about scalars. So we have the set of n tuples now, or perhaps it might be a set of sequences or a set of wave forms. The things we insist on before we will call this set a vector space are that we have an operation which we call addition. Fortunately for n tuples, addition is what you would think it would be. It's element-wise addition. For sequences, it's element-wise addition again. For wave forms, it's function addition. You have commutativity -- you can add things in either order. I mean look, if nobody pointed this out to you, you'd do it anyway, right? So that it looks like nothing. But in fact, you can invent things, mathematical objects, which you can deal with which are not commutative.

So this is saying to be a vector space it has to have this commutativity operation. Associativity -- again, it's hard to see why you would say something like this, but all we're defining is this addition operation, namely, there's a definition of addition, and we know by axiom that the sum has to be in this vector space. As soon as we assume this associativity axiom, it says OK, this is a vector in the space, this and this are both vectors in the space. Therefore, this sum has to be in the space. Therefore this plus the sum has to be in the space. What associativity says is that that's the same element as this. Once you see this you leave out the parentheses because you can add as many things as you want to. Same thing -- well, here there has to be a unique vector, zero -- v plus zero is equal to v for all v , and there's only one vector that has that property. We're going to see some problems there in a little bit when we start studying these L^2 functions, but we'll do that next time.

Finally, for every v there's a unique negative vector so that the sum is equal to zero. All the operations that you're used to. We'll go through the other things next time.