

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](http://ocw.mit.edu).

**PROFESSOR:**

OK. I want to review zero-mean jointly Gaussian random variables. And I want to review a couple of the other things I did last time. Because when I get into questions of stationarity and things like that today, I think it will be helpful to have a little better sense of what all that was about, or it will just look like a big mess. And there are sort of a small number of critical things there that we have to understand.

One of them is that if you have a non-singular covariance matrix, OK, you have a bunch of random variables. This bunch of random variables, each one of them has a covariance. Namely,  $z_{sub\ i}$  and  $z_{sub\ j}$  have an expected value of  $z_{sub\ i}$  and  $z_{sub\ j}$ . I'm just talking about zero mean here. Anytime I don't say whether there's a mean or not, I mean there isn't a mean. But I think the notes are usually pretty careful.

If the covariance matrix is non-singular, then the following things happen. The vector,  $z$ , is jointly Gaussian. And our first definition was, if you can represent it as a -- all the components of it each as linear combinations of IID normal Gaussian random variables, namely independent Gaussian random variables -- so they're all just built up from this common set of independent Gaussian random variables.

Which sort of matches our idea of why noise should be Gaussian. Namely, it comes from a collection of a whole bunch of independent things, which all get added up in various ways. And we're trying to look at processes, noise processes. The thing that's going to happen then is that, at each epoch in time, the noise is going to be some linear combination of all of these noise effects, but is going to be filtered a little bit. And because it's being filtered a little bit, the noise at one time and the noise at another time are all going to depend on common sets of variables. So that's the intuitive idea here.

Starting with that idea, last time we derived what the probability density had to be. And if you remember, this probability density came out rather simply. The only thing that was involved here was this idea of going from little cubes in the IID noise domain into parallelograms in the  $z$  domain. Which is what happens when you go through a linear transformation. And because of that, this is the density that has to come out of there. By knowing a little bit about matrices -- which is covered in the appendix, or you can just take it on faith if you want to -- you can break up this matrix into eigenvalues and eigenvectors.

If the matrix is nonsingular -- well, if the matrix is singular or nonsingular, if it's a covariance matrix, or it's -- a non-negative definite matrix, is what these matrices are -- then you can always break this matrix up into eigenvalues and eigenvectors. The eigenvectors span the space. The eigenvectors can be taken to be orthonormal to each other.

And then when you express this in those terms, this probability density becomes just a product of normal Gaussian random variable densities, where the particular Gaussian random variables are these inner products of the noise vector with these various eigenvectors. In other words, you're taking this set of random variables which you're expressing in some reference system, where  $z_1$  up to  $z_k$  are the different components to the vector. When you express this in a different reference system, you're rotating that space around. What this says is, you can always rotate the space in such a way that what you will find is orthogonal random variables, which are independent of each other. So that this is the density that you wind up having.

And what this means is, if you look at the region of equal probability density over this set of random variables, what you find is it's a bunch of ellipsoids. And the axes of the ellipsoids are simply these eigenvectors here. OK, so this is the third way. If you can represent the noise in this way, again, it has to be jointly Gaussian.

Finally, if all linear combinations of this random vector are Gaussian, that's probably the simplest one. But it's the hardest one to verify, in a sense. And it's the hardest

one to get all these other results from. But if all linear combinations of these random variables are all Gaussian, then in fact, again, the variables have to be jointly Gaussian.

Again, it's important to understand that jointly Gaussian means more than just individually Gaussian. It doesn't mean what you would think from the words jointly Gaussian, as saying that each of the variables are Gaussian. It means a whole lot more than that. In the problem set, what you've done in one of the problems is to create a couple of examples of two random variables which are each Gaussian random variables, but which jointly are not jointly Gaussian.

OK, finally, if you have a singular covariance matrix,  $z$  is jointly Gaussian if a basis of these  $z_i$ 's are jointly Gaussian. OK? In other words, you throw out the random variables which are just linear combinations of the others, because you don't even want to think of those as random variables in a sense. I mean, technically they are random variables, because they're defined in the sample space. But they're all just defined in terms of other things, so who cares about them? So after you throw those out, the others have to be jointly Gaussian.

So in fact, if you have two random variables,  $z_1$  and  $z_2$ , and  $z_2$  equals  $z_1$  -- OK, in other words the probability density is on a straight diagonal line -- and  $z_1$  is Gaussian,  $z_1$  and  $z_2$  are jointly Gaussian in that case. This is not an example like the ones you did in this problem set that you're handing in today, where you in fact have things like a probability density which doesn't exist because it's impulsive on this line, and also impulsive on this line. Or something which looks Gaussian in two of the quadrants and is zero in both the other quadrants, and really bizarre things like that. OK?

So please try to understand what jointly Gaussian means. Because everything about noise that we do is based on that. It really is. And if you don't know what jointly Gaussian means, you're not going to understand anything about noise detection, or anything from this point on. I know last year I kept telling people that, and in the final exam there were still four or five people who didn't have the foggiest

idea what jointly Gaussian meant. And you know, you're not going to understand the rest of this, you're not going to understand why these things are happening, if you don't understand that. OK.

So the next thing we said is  $Z$  of  $t$  -- I use those little curly brackets around something just as a shorthand way of saying that what I'm interested in now is the random process, not the random variable at a particular value of  $t$ . In other words, if I say  $Z$  of  $t$ , what I'm usually talking about is a random variable, which is the random variable corresponding to one particular epoch of this random process. Here I'm talking about the whole process. And it's a Gaussian process. If  $Z$  of  $t_1$  to  $Z$  of  $t_k$  are jointly Gaussian for all  $k$  and for all sets of  $t$  sub  $i$ .

And if the process can be represented as a linear combination of just ordinary random variables multiplied by some set of orthonormal functions, and these  $Z$  sub  $i$  are independent, and if the sum of the variances of these random variables sum up to something less than infinity. So you have finite energy in these sample functions, effectively is what this says. Then it says the sample functions of  $Z$  of  $t$  are L2 with probability one.

OK. You have almost proven that in the problem set. If you don't believe that you've almost proven it, you really have. And anyway, it's true. One of the things that we're trying to do when we're trying to deal with these wave forms that we transmit, the only way we can deal with them very carefully is to know that they're L2 functions. Which means they have Fourier transforms. We can do all this stuff we've been doing. And you don't need anything more mathematically than that. So it's important to be dealing with L2 functions.

We're now getting into random processes, and it's important to know that the sample functions are L2. Because so long as the sample functions are L2, then you can do all of these things we've done before and just put it together and say well, you take a big ensemble of these things. They're all well defined, they're all L2. We can take Fourier transforms, we can do everything else we want to do. OK, so that's the game we're playing.

OK, I'm just going to assume that sample functions are L2 from now on, except in a couple of bizarre cases that we look at. Then a linear functional is a random variable given by the random variable is equal to the integral of the noise process  $z$  of  $t$ , multiplied by an ordinary function,  $g$  of  $t$ ,  $dt$ . And we talked about that a lot last time. This is the convolution of a process -- well, it's not the convolution. It's just the inner product of a process with a function. And we interpret this last time in terms of the sample functions of the process and the sample values of the random variable. And then since we could do that, we could really talk about the random variable here.

This means that for all of the sample values in the sample space with probability one, the sample values of the random variable  $v$  are equal to the integral of the sample values of the process times  $g$  of  $t$   $dt$ . OK, in other words, this isn't really something unusual and new.

I mean, students have the capacity of looking at this in two ways, and I've seen it happen for years. The first time you see this, you say this is trivial. Because it just looks like the kind of integration you're doing all your life. You work with it for a while, and then at some point you wake up and you say, oh, but my god, this is not a function here. This is a random process. And you say, what the heck does this mean? And suddenly you're way out in left field. Well, this says what it means. Once you know what it means, we go back to this and we use this from now on. OK?

If we have a zero-mean Gaussian process  $z$  of  $t$ , if you have L2 sample functions, like you had when you take a process and make it up as a linear combination of random variables times orthonormal functions, and if you have a bunch of different L2 functions,  $g_1$  up to  $g_{j_0}$ , then each of these random variables, each of these linear functionals, are going to be Gaussian. And in fact the whole set of them together are jointly Gaussian. And we showed that last time. I'm not going to bother to show it again.

And we also found what the covariance was between these random variables. And it was just this expression here. OK. And this follows just by taking  $v_i$ , which is this kind

of integral, times  $v_j$ , and interchanging the order of integration, taking the expected value of  $z$  of  $t$  times  $z$  of  $\tau$ , which gives you this quantity. And you have the  $g_i$  of  $t$  and the  $g_j$  of  $\tau$  stuck in there with an integral around it. And it's hard to understand what those integrals mean and whether they really exist or not, but we'll get to that a little later.

We then talked about linear filtering of processes. You have a process coming into a filter, and the output is some other stochastic process. Or at least we hope it's a well defined stochastic process. And we talked a little bit about this last time. And for every value of  $\tau$ , namely for each random variable in this output random process,  $V$  of  $\tau$  is just a linear functional. It's a linear functional corresponding to the particular function  $h$  of that  $\tau$  minus  $t$ . So the linear functional is a function of  $t$  here for the particular value of  $\tau$  that you have over here. So this is just a linear functional like the ones we've been talking about.

OK, if  $z$  of  $t$  is a zero-mean Gaussian process, then you have a bunch of different linear functionals here for any set of times  $\tau_1$  up to  $\tau_k$ . And those are jointly Gaussian from what we just said. And if all of these sets of  $k$  sample -- If based a random process,  $v$  of  $\tau$ , is a Gaussian random process, if for all  $k$  sets of epochs,  $v$ ,  $\tau_1$ ,  $\tau_2$ , up to  $\tau_k$ , if this set of random variables are all jointly Gaussian. So that's what we have here. So the  $V$  of  $\tau$  is actually a Gaussian process if  $Z$  of  $t$  is a Gaussian random process to start with. And the covariance function is just this quantity that we talked about before.

OK, so we have a covariance function. We also have a covariance function for the process we started with. And if it's Gaussian, all you need to know is what the covariance function is. So that's all rather nice.

OK, as we said, we're going to start talking about stationarity today. I really want to talk about two ideas of stationarity. One is the idea that you have probably seen as undergraduates one place or another, which is simple computationally, but is almost impossible to understand when you try to say something precise about it. And the other is something called effectively stationary, which we're going to talk about

today. And I'll show you why that makes sense.

So we say that a process  $Z$  of  $t$  is stationary if, for all integers  $k$ , all shifts  $\tau$ , all epochs  $t_1$  to  $t_{\text{sub } k}$ , and all of values of  $z_1$  to  $z_k$ , this joint density here is equal to the joint density shifted. In other words, what we're doing is we're taking a set of times over here,  $t_1, t_2, t_3, t_4, t_5$ . We're looking at the joint distribution function for those random variables here, and then we're shifting it by  $\tau$ . And we're saying if the process is stationary, the joint distribution function here has to be the same as the joint distribution function there.

You might think that all you need is the distribution function here has to be the same as the distribution function there. But you really want the same relationship to hold through here as there if you want to call process stationary. OK. If we have zero-mean Gauss, that's just equivalent to saying that the covariance function at  $t_i$  and  $t_{\text{sub } j}$  is the same as the covariance function at  $t_i$  plus  $\tau$  and  $t_j$  plus  $\tau$ . And that's true for  $t_1$  up to  $t_{\text{sub } k}$ . Which really just means this has to be true for all  $\tau$ , for all  $t_{\text{sub } i}$  and for all  $t_{\text{sub } j}$ . So you don't need to worry about the  $k$  at all once you're dealing with a Gaussian process. Because all you need to worry about is the covariance function. And the covariance function is only a function of two variables, the process at one epoch and the process at another epoch.

And this is equivalent, even more simply, to saying that the covariance function at  $t_1$  and  $t_2$  has to be equal to the covariance function at  $t_1$  minus  $t_2$  and zero. Can you see why that is? If I start out with this, and I know that this is true, then the thing that I can do is take this and I can shift it by any amount that I want to. So if I shift this by adding  $\tau$  here, then what I wind up with is  $k_z$  of  $t_1$  minus  $t_2$  plus  $\tau$ , comma  $\tau$ , is equal to  $k_z$  of  $t_1$  plus  $\tau$ ,  $t_2$  plus  $\tau$ . And if I change variables around a little bit, I come to this.

So this is the condition we need for a Gaussian process to be stationary. I would defy any of you to ever show in any way that a process satisfies all of these conditions. I mean, if you don't have nice structural properties in the process, like a Gaussian process, which says that all you need to define it is this, this is something

that you just can't deal with very well.

So we have this. And then we say, OK, this covariance is so easy to work with -- I mean no, it's not easy to work with, but it's one hell of a lot easier to work with than that. And therefore if you want to start talking about processes, and you don't really want to go into all the detail of these joint distribution functions, you will say to yourselves that one thing I might ask for in a process is the question of whether the covariance function satisfies this property. I mean, for a Gaussian process this is all you need to make the process stationary. For other processes you need more, but at least it's an interesting question to ask. Is the process partly stationary, in the sense that the covariance function at least is stationary? Namely, the covariance function here looks the same as the covariance function there.

So we say that a zero-mean process is wide-sense stationary if it satisfies this condition. So a Gaussian process then is stationary if and only if it's wide-sense stationary.

And a random process with a mean, or with a mean that isn't necessarily zero, is going to be stationary or wide-sense stationary if the mean is constant and the fluctuation is stationary. Or wide-sense stationary, as the case may be. So you want both of these properties there. So as before, we're just going to throw out the mean and not worry about it. Because if we're thinking of it as noise, it's not going to have a mean. Because of it has a mean, it's not part of the noise. It's just something we know, and we might as well remove it.

OK, interesting example here. Let's look at a process,  $V$  of  $t$ , which is defined in this way here. It's the sum of a set of random variables,  $V$  sub  $k$ , times the sinc function for some sampling interval, capital  $T$ . And I'm assuming that the  $v$  sub  $k$  are zero-mean, and that, at least as far as second moments are concerned, they have this stationarity property between them. Namely, they are uncorrelated from one  $j$  to another  $k$ . Expected value of  $v_j$  squared is  $\sigma^2$ . Expected value of  $v_j v_k$  for  $k$  unequal to  $j$  is zero. So they're uncorrelated; they all have the same variance.

Then I claim that the process  $V$  of  $t$  is wide-sense stationary. And I claim that this

covariance function is going to be  $\sigma^2 \frac{\sin(t - \tau)}{t - \tau}$ . Now for how many of you is this an obvious consequence of that? How many of you would guess this if you had to guess something? Well, if you wouldn't guess it, you should. Any time you don't know something, the best thing to do is to make a guess and try to see whether your guess is right or not. Because otherwise you never know what to do.

So if you have a function which is defined in this way -- think of it this way. Suppose these  $V_k$ 's were not random variables, but instead suppose they were all just constants. Suppose they were all 1. Suppose you're looking at the function  $V$  of  $t$ , which is the sum of all of the sinc functions. And think of what you'd get when you add up a bunch of sinc functions which are all exactly the same. What do you get? I mean, you have to get a constant, right? You're just interpolating between all these points, and all the points are the same. So it'd be very bizarre if you didn't get something which was constant.

Well the same thing happens here. The derivation of this is one of those derivations which is very, very simple and very, very slick and very elegant. And I was going to go through it in class and I thought, no. You just can't follow this in real time. It's something you have to sit down and think about for five minutes. So I urge you all to read this, because this is a trick that you need to use all the time. And you should understand it, because various problems as we go along are going to use this idea in various ways.

But anyway, when you have a process defined this way, it has to be wide-sense stationary. And the covariance function is just this function here. Now if these  $V_k$ 's up here are jointly Gaussian, and they all have the same variance, so they're all IID, then what this means is you have a Gaussian random process. And it's a stationary Gaussian random process. In other words, it's a process where, if you look at the random variable  $v$  of  $\tau$  at any given  $\tau$ , you get the same thing for all  $\tau$ .

In other words, it's a little peculiar, in the sense that when you look at this, it looks

like time 0 was a little special. Because time 0 is where you specified the process. Time  $t$  you specified the process. Time  $2t$  you specified the process. But in fact this is saying, no. Time 0 is not special at all. Which is why we say that the process is wide-sense stationary. All times look the same.

On the other hand, if you take these  $V_{sub k}$  here to be binary random variables which take one the value plus 1 or minus 1 with equal probability, what do you have then? You look at the sum here, and at the sample points -- namely at 0, at  $t$ , at  $2t$ , and so forth -- what values can the process take away? At time 0? If  $v_{sub 0}$  is either plus 1 or minus 1, what is  $v$  of 0 over here? It's plus 1 or minus 1. It can only be those two things.

If you look at  $V$  of capital  $T$  over 2, namely if you look halfway between these two sample points, you then ask, what are the different values that  $v$  of  $T$  over 2 can take on? It's an awful mess. It's a very bizarre random variable. But anyway, it's not a random variable which is plus or minus 1. Because it's really a sum of an infinite number of terms. So you're taking an infinite number of binary random variables, each with arbitrary multipliers. So you're adding them all up. So you get something that's not differentiable. It's not anything nice.

I don't care about that. The only thing that I care about is that  $v$  of capital  $T$  over 2 is not a binary random variable anymore. And therefore this process is not stationary anymore. Here's an example of a wide-sense stationary process which is not stationary. So it's a nice example of where you have wide-sense stationarity, but you don't have stationarity.

So all kinds of questions about power and things like that, this process works very, very well. Because questions about power you answer only in terms of covariance functions. Questions of individual possible values and probability distributions you can't answer the very well.

One more thing. If these variables are Gaussian, and if you actually believe me that this is a stationary Gaussian process, and it really is a stationary Gaussian process, what we have is a way of creating a broad category of random processes. Because

if I look at the sample functions here of this process, each sample function is bandlimited. Baseband limited. And it's baseband limited to  $1/2T$ .  $1/2T$  over the quantity  $2T$ . So by choosing capital T to be different things, I can make these sample functions have large bandwidth or small bandwidth so I can look at a large variety of different things. All of them have Fourier transforms, which look sort of flat in magnitude. And we'll talk about that as we go on.

And when I pass these things through linear filters, what we're going to find is we can create any old Gaussian random process we want to. So that's why they're nice. Yes?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Can we make V of t white noise? In a practical sense, yes. In a mathematical sense, no. In a mathematical sense, you can't make anything white noise. In a mathematical sense, white noise is something which does not exist. I'm going to get to that later today. And it's a good question. But the answer is yes and no.

OK, the trouble with stationary processes is that the sample functions aren't L2. That's not the serious problem with them, because all of us as engineers are willing to say, well, whether it's L2 or not I don't care. I'm just going to use it because it looks like a nice random process. It's not going to burn anything else or anything. So, so what? But the serious problem here is that it's very difficult to view stationary processes as approximations to real processes. I mean, we've already said that you can't have a random process which is running merrily away before the Big Bang. And you can't have something that's going to keep on running along merrily after we destroy ourselves, which is probably sooner in the future than the Big Bang was in the past.

But anyway, this definition of stationarity does not give us any way to approximate this. Namely, if a process is stationary, it is either identically zero, or every sample function with probability one has infinite energy in it. In other words, they keep running on forever. They keep building up energy as they go. And as far as the definitions current is concerned, there's no way to say large negative times and

large positive times are unimportant. If you're going to use mathematical things, though you're using them as approximations, you really have to consider what they're approximations to. So that's why I'm going to develop this idea of effectively wide-sense stationary or effectively stationary.

OK. So a zero-mean process is effectively stationary or effectively wide-sense stationary, which is what I'm primarily interested in, within two intervals, within some wide interval of time, minus  $T_0$  to plus  $T_0$ . I'm thinking here of choosing  $T_0$  to be enormous. Namely if you build a piece of equipment, you would have minus  $T_0$  to plus  $T_0$  include the amount of time that the equipment was running.

So we'll say it's effectively stationary within these limits if the joint probability assignment, or the covariance matrix if we're talking about effectively wide-sense stationary, for  $t_1$  up to  $t_{sub k}$  is the same as that for  $t_1$  plus  $\tau$  up to  $t_k$  plus  $\tau$ . In other words I have this big interval, and I have a bunch of times here.  $t_{sub k}$ . And I have a bunch of times over here,  $t_1$  plus  $\tau$  up to  $t_k$  plus  $\tau$ . And this set of times don't have to be disjoint from that set of times. And I have this time over here, minus  $t_0$ , and I have this time way out here which is plus  $t_0$ . And what I'm saying is, I'm going to call this function wide-sense stationary if, when we truncate it to minus  $t_0$  to plus  $t_0$ , it is stationary as far as all those times are concerned. In other words, we just want to ignore what happens before minus  $t_0$  and what happens after plus  $t_0$ .

You have to be able to do that. Because if you can't talk about the process in that way, you can't talk about the process at all. The only thing you can do is view the noise over some finite interval.

OK. And as far as covariance matrix for wide-sense and stationary, well it's the same definition. So we're going to truncate the process and deal with that.

For effectively stationary or effectively wide-sense stationary, I want to view the process as being truncated to minus  $t_0$  to plus  $t_0$ . We have this process which might or might not be stationary. I just truncate it and I say I'm only going to look at this finite segment of it. I'm going to define this one variable covariance function as  $k_z$  of  $t_1$  and  $t_2$ . And  $k_z$  of  $t_1$  and  $t_2$  is going to be -- blah. The single variable covariance

function is defined as  $k_z$  if  $t_1$  minus  $t_2$  is equal to the actual covariance function evaluated with one argument at  $t_1$  and the other argument at  $t_2$ . And I want this to hold true for all  $t_1$  and all  $t_2$  in this interval.

And this square here gives a picture of what we're talking about. If you look at the square in the notes, unfortunately all the diagonal lines do not appear because of the bizarre characteristics of LaTeX. LaTeX is better than most programs, but the graphics in it are awful. But anyway, here it is drawn properly. And along this line,  $t$  minus  $\tau$  is constant. So this is the line over which we're insisting that  $k_z$  of  $t_1$  and  $t_2$  be constant. So for a random process to be wide-sense stationary, what we're insisting is that within these limits, minus  $t_0$  to plus  $t_0$ , the covariance function is constant along each of these lines along here. So it doesn't depend on both  $t_1$  and  $t_2$ . It only depends on the difference between them.

Which is what we require for stationary to start with. If you don't like the idea of effectively wide-sense stationary, just truncate the process and say, well if it's stationary, this has to be satisfied. It has to be constant along these lines.

The peculiar thing about this is that the single variable covariance function does not run from minus  $t_0$  to plus  $t_0$ . It runs from minus  $2t_0$  to plus  $2t_0$ . So that's a little bizarre and a little unpleasant. And it's unfortunately the way things are. And it also says that this one variable covariance function is not always the same as the covariance evaluated at  $t$  minus  $\tau$  and 0. Namely, it's not the same because  $t$  minus  $\tau$  might be considerably larger than  $t_0$ . And therefore, this covariance is zero, if we truncated the process, and this covariance is not zero.

In other words, for these points up here, and in fact this point in particular -- well, I guess the best thing to look at is points along here, for example. Points along here, what we insist on is that  $t$  minus  $\tau$ ,  $k_z$  of  $t$  and  $\tau$  is constant along this line. But we don't insist on  $k_z$  of  $t$  minus  $\tau$ , which is some quantity bigger than  $t_0$  along this line, being the same as this quantity here.

OK, so aside from that little peculiarity, this is the same thing that you would expect from just taking a function and truncating it. There's nothing else that's peculiar

going on there.

OK, so let's see if we can do anything with this idea. And the first thing we want to do is to say, how about these linear functionals we've been talking about? Why do I keep talking about linear functionals and filtering? Because any time you take a noise process and you receive it, you're going to start filtering it. You're going to take start taking linear functionals of it. Namely all the processing you do is going to start with finding random variables in this particular way. And when you look at the covariance between two of them, what you find is the same thing we found last time. Expected value of  $V_i$  times  $V_j$  is equal to the integral of  $g_i$  of  $t$  times the covariance function evaluated at  $t$  and  $\tau$ , times  $g_j$  of  $\tau$ .

All of this is for real processes and for real functions. Because noise random processes are in fact real. I keep trying to alternate between making all of this complex and making it real. Both have their advantages. But at least in the present version of the notes, everything is real, concerned with random processes.

So I have this function here. If  $g_j$  of  $t$  is zero for  $t$  greater than  $T_0$ , what's going to happen here then? I'm evaluating this where this is a double integral over what region? It's a double integral over the region where  $t$  is constrained to minus  $T_0$  to plus  $T_0$ , and  $\tau$  is constrained to the interval minus  $t_0$  to plus  $T_0$ . In other words, I can evaluate this expected value, this covariance, simply by looking at this box here. If I know what the random process is doing within this box, I can evaluate this thing. So that if the process is effectively stationary within minus  $t_0$  to plus  $t_0$ , I don't need to know anything else to evaluate all of these linear functionals.

But that's exactly the way that were choosing this quantity,  $T_0$ . Namely, we're choosing  $T_0$  to be so large that everything we're interested in happens inside of there. So we're saying that all of these linear functionals for effective stationarity are just defined for what happens inside this interval.

So if  $Z$  of  $t$  is effectively wide-sense stationary within this interval, you make the intervals large as you want or as small as you want. The real quantity is that these functions  $g$  have to be constrained within minus  $T_0$  to plus  $T_0$ . Then  $V_i$ ,  $V_j$ , are

jointly Gaussian. In other words, you can talk about jointly Gaussian without talking at all about whether the process is really stationary or not. You can evaluate this for everything within these limits, strictly in terms of what's going on within those limits.

That one was easy. The next one is a little harder. We have our linear filter now. We have the noise process going into a linear filter. And we have some kind of process coming out of the filter. Again, we would like to say we couldn't care less about what this process is doing outside of these humongous limits. And the way we do that is to say -- well, let's forget about that for the time being. Let us do what you did as an undergraduate, before you acquired wisdom, and say let's just integrate and not worry at all about changing orders of integration or anything else. Let's just run along and do everything as carelessly as we want to.

The covariance function for what comes out of this filter is something we already evaluated. Namely, what you want to look at is  $v$  of  $t$  and  $v$  of  $\tau$ .  $v$  of  $t$  is this integral here.  $v$  of  $\tau$  is this integral here with  $\tau$  substituted for  $t$ . So when you put both of these together and you take the expected value, and then you bring the expected value in through the integrals -- who knows whether you can do that or not, but we'll do it anyway -- what we're going to get is a double integral of the impulse response, evaluated at  $t$  minus  $t_1$ , times the covariance function evaluated at  $t_1$  and  $t_2$ , times the function that we're dealing with,  $h$ , evaluated at  $\tau$  minus  $t_2$ .

This is what you got just by taking the expected value of  $v$  of  $t$  times  $v$  of  $\tau$ . And you just write it out. You write out the expected value of these two integrals. It's what we did last time. It's what's in the notes if you don't see how to do it. You take the expected value inside, and this is what you get.

Well now we start playing these games. And the first game to play is, we would like to use our notion of stationarity. So in place of  $k_z$  of  $t_1$  and  $t_2$ , we want to substitute the single variable form,  $k_z$  tilde of  $t_1$  minus  $t_2$ . But we don't like the  $t_1$  minus  $t_2$  in there, so we substitute  $\phi$  for  $t_1$  and  $t_2$ . And then what we get is the double integral of  $k_z$  of  $\phi$  now. And we have gotten rid of  $t_1$  by this substitution, so we wind up with the integral over  $\phi$  and over  $t_2$ .

Next thing we want to do, we want to get rid of the  $t_2$  also. So we're going to let  $\mu$  equal  $t_2$  minus  $\tau$ . When we were starting here, we had four variables,  $t$  and  $\tau$  and  $t_1$  and  $t_2$ . We're not getting rid of the  $t$  and we're not getting rid of the  $\tau$ , because that's what we're trying to calculate. But we can play around with the  $t_1$  and  $t_2$  as much as we want, and we can substitute variables of integration here. So  $\mu$  is going to be  $t_2$  minus  $\tau$ . That's going to let us get rid of the  $t_2$ . And we wind up with this form here when we do that substitution. It's just that  $h$  of  $\tau$  minus  $t_2$  becomes  $h$  of minus  $\mu$  instead of  $h$  of plus  $\mu$ .

I don't know what that means. I don't care what it means. The only thing I'm interested in now is, you look at where the  $t$ 's and the  $\tau$ 's are, and the only place that  $t$ 's and  $\tau$ 's occur are here. And they occur together. And the only thing this is a function of is  $t$  minus  $\tau$ . I don't know what kind of function it is, but the only thing we're dealing with is  $t$  minus  $\tau$ . I don't know whether these intervals exist or not. I can't make any very good argument that they can. But if they do exist, it's a function only of  $t$  minus  $\tau$ . So aside from the pseudo-mathematics we've been using,  $V$  of  $t$  is wide-sense stationary. OK?

Now if you didn't follow all this integration, don't worry about it. It's all just substituting integrals and integrating away. And it's something you can do. Some people just take five minutes to do it, some people 10 minutes to do it.

Now we want to put the effective stationarity in it. Because again, we don't know what this means. We can't interpret what it means in terms of stationarity. So we're going to assume that our filter,  $h$  of  $t$ , has a finite duration impulse response. So I'm going to assume that it's zero when the magnitude of  $t$  is greater than  $A$ . Again, I'm assuming a filter which starts before zero and runs until after zero. Because again, what I'm assuming is the receiver timing is different from the transmitter timing. And I've set the receiver timing at some convenient place. So my filter starts doing things before anything hits it, just because of this change of time.

So  $v$  of  $t$ , then, is equal to the integral of the process  $Z$  of  $t_1$  times  $h$  of  $t$  minus  $t_1$ . This is a linear functional again. This depends only on what  $Z$  of  $t_1$  is for  $t$  minus  $A$

less than or equal to  $t_1$ , less than or equal to  $t$  plus  $A$ . Because I'm doing this integration here, this function here is 0 outside of that interval. I'm assuming that  $h$  of  $t$  is equal to 0 outside of this finite interval. And this is just a shift of  $t_1$  on it. So this integral is going to be equal to zero from when  $t$  is equal to  $t_1$  minus  $A$  to when  $t$  is equal to  $t_1$  plus  $A$ . Every place else, this is equal to zero.

So  $V$  of  $t$ , the whole process depends -- and if we only want to evaluate it within minus  $T_0$  plus  $A$  to  $T_0$  minus  $A$  -- in other words we look at 10 to the eighth years and we subtract off a microsecond from that region. So now we're saying we want to see whether this is wide-sense stationary within 10 to the eighth years minus a microsecond, minus that to plus that. So that's what we're dealing with here.

$V$  of  $t$  in this region depends only on  $Z$  of  $t$  for  $t$  in the interval minus  $T_0$  to plus  $T_0$ . In other words, when you're calculating  $V$  of  $t$  for any time in this interval, you're only interested in  $Z$  of  $t$ , which is diddling within plus or minus  $A$  of that. Because that's the only place where the filter is doing anything.

So  $V$  of  $t$  depends only on  $Z$  of  $t$ . And here, I'm going to assume that  $Z$  of  $t$  is wide-sense stationary within those limits. And therefore I don't have to worry about what  $Z$  of  $t$  is doing outside of those limits. So if the sample functions of  $Z$  of  $t$  are  $L_2$  within minus  $T_0$  to plus  $T_0$ , then the sample functions of  $V$  of  $t$  are  $L_2$  within minus  $T_0$  plus  $A$  to  $T_0$  minus  $A$ .

Now this is not obvious. And there's a proof in the notes of this, which goes through all this  $L_1$  stuff and  $L_2$  stuff that you've been struggling with. It's not a difficult proof, but if you don't dig  $L_2$  theory, be my guest and don't worry about it. If you do want to really understand exactly what's going on here, be my guest and do worry about it, and go through it. But it's not really an awful argument. But this is what happens.

So what this says is we can now view wide-sense stationarity and stationarity, at least for Gaussian processes, as a limit of effectively wide-sense stationary processes. In other words, so long as I deal with filters whose impulse response is constrained in time, the only effect of that filtering is to reduce the interval over which the process is wide-sense stationary by the quantity  $A$ . And the quantity  $A$  is

going to be very much smaller than  $T_0$ , because we're going to choose  $T_0$  to be as large as we want it to be. Namely, that's always the thing that we do when we try to go through limiting arguments.

So what we're saying here is, by using effective stationarity, we have managed to find a tool that lets us look at stationary as a limit of effectively stationary processes as you let  $T_0$  become larger and larger. And we have our cake and we eat it too, here. Because at this point, all these functions we're dealing with, we can assume that they're L2. Because they're L2 for every  $T_0$  that we want to look at. So we don't have to worry about that anymore.

So we have these processes. Now we will just call them stationary or wide-sense stationary. Because we know that what we're really talking about is a limit as  $T_0$  goes to infinity of things that make perfect sense. So suddenly the mystery, or the pseudo-mystery has been taken out of this, I hope.

So you have a wide-sense stationary process with a covariance now of  $k_{z}$  of  $\tau$ . In other words, it's effectively stationary over some very large interval, and I now have this covariance function, which is a function of one variable. I want to define the spectral density of this process as a Fourier transform of  $k_{z}$ . In other words, as soon as I get a one variable covariance function, I can talk about its Fourier transform. At least assuming that it's L2 or something. And let's forget about that for the time being, because that all works.

We want to take the Fourier transform of that. I'm going to call that spectral density. Now if I call that spectral density, the thing you ought to want to know is, why am I calling that spectral density? What does it have to do with anything you might be interested in? Well, we look at it and we say well, what might it have to do with anything?

Well, the first thing we're going to try is to say, what happens to these linear functionals we've been talking about? We have a linear functional,  $V$ . It's an integral of  $g$  of  $t$  times  $Z$  of  $t$  dt. We would like to be able to talk about the expected value of  $V$  squared. Talk about zero-mean things, so we don't care about the mean at zero.

So we'll just talk about the variance of any linear functional of this form. If I can talk about the variance of any linear functional for any  $g$  of  $t$  that I'm interested in, I can certainly say a lot about what this process is doing.

So I want to find that variance. If I write this out, it's same thing we've been writing out all along. It's the integral of  $g$  of  $t$  times this one variable covariance function now, times  $g$  of  $\tau$   $d\tau$   $dt$ . When you look at this, and you say OK, what is this? It's an integral of  $g$  of  $t$  times some function of  $t$ . And that function of  $t$  is really the convolution of  $k$  sub  $z$ . This is a function now. It's not a random process anymore; it's just a function. That's a convolution of the function  $k$  tilde with  $g$ .

So I can rewrite this as the integral of  $g$  of  $t$  times this convolution of  $t$ . And then I'm going to call the convolution  $\theta$  of  $t$ , just to give it a name. So this variance is equal to  $g$  of  $t$  times this function  $\theta$  of  $t$ , which is just the convolution of  $k$  with  $g$ .

Well now I say, OK, I can use I can use Parseval's relation on that. And what I get is this integral, if I look in the frequency domain now, as the integral of the Fourier transform of  $g$  of  $t$ , times the complex conjugate of  $\theta$  star of  $f$  times  $df$ . I've cheated it you just a little bit there because what I really want to talk about is -- I mean, Parseval's theorem relates this quantity to this quantity, where  $\theta$  is  $\theta$  complex conjugate.

But fortunately,  $\theta$  is real. Because  $g$  is real and  $k$  is real. Because we're only dealing with real processes here, and we're only dealing with real filters. So this is real, this is real. So this integral is equal to this in the frequency domain. And now, what is the Fourier transform of -- OK,  $\theta$  of  $t$  is this convolution. When I take the Fourier transform of  $\theta$ , what I get is the product of the Fourier transform of  $k$  times the Fourier transform of  $g$ . So in fact what I get is  $\hat{g}$  of  $f$  times  $\theta$  star of  $f$ . Which is  $\hat{g}$  complex conjugate of that times  $k$  complex conjugate -- times -- it doesn't make any difference whether it's a complex conjugate or not, because it's real.

So I wind up with the integral of the magnitude squared of  $\hat{g}$  times this spectral density. Well at this point, we can interpret all sorts of things from this. And we now

know that we can interpret this also in terms of effectively stationary things. So we don't have any problem with things going to infinity or anything. I mean, you can go through this argument carefully for effectively stationary things, and everything works out fine. So long as  $g$  of  $t$  is constrained in time.

Well now, I can choose this function any way that I want to. In other words, I can choose my function  $g$  of  $t$  in any way I want to. I can choose this function in any way I want to. So long as its inverse transform is real. Which means  $\hat{g}$  of  $f$  has to be equal to  $\hat{g}$  of minus  $f$  complex conjugate. But aside from that, I can choose this to be anything. And if I choose this to be very, very narrow band, what is this doing? It's saying that if I take a very narrow band  $g$  of  $t$ , like a sinusoid truncated out to some very wide region, I multiply that by this process, I integrate it out, and I look at the variance. What am I doing when I'm doing that?

I'm effectively filtering my process with a very, very narrow band filter, and I'm asking, what's the variance of the output? So the variance of the output really is related to the amount of energy in this process at that frequency. That's exactly what the mathematics says here. It says that  $s$  sub  $z$  of  $f$  is in fact the amount of noise power per unit bandwidth at this frequency. It's the only way you can interpret it. Because I can make this as narrow as I want to, so that if there is any interpretation of something in this process as being at one frequency rather than another frequency and the only way I can interpret that is by filtering the process. This is saying that when you filter the process, what you see at that bandwidth is how much power there is in the process at that bandwidth. So this is simply giving us an interpretation of spectral density as power per unit bandwidth.

OK, let's go on with this. If this spectral density is constant over all frequencies of interest, we say that it's white. Here's the answer to your question: what is white Gaussian noise? White Gaussian noise is a Gaussian process which has a constant spectral density over the frequencies that we're interested in. So we now have this looking at it in a certain band of frequencies.

You know, if you think about this in more or less practical terms, suppose you're

building a wireless network, and it's going to operate, say, at five or six gigahertz. And suppose your bandwidth is maybe 100 megahertz or something. Or maybe it's 10 megahertz, or maybe it's 1 megahertz. In terms of this frequency of many gigahertz, you're talking about very narrowband communication. People might call it wideband communication, but it's really pretty narrowband.

Now suppose you have Gaussian noise, which is caused by all of these small tiny noise affects all over the place. And you ask, is this going to be flat or is it not going to be flat? Well, you might look at it and say, well, if there's no noise in this little band and there's a lot of noise in this band, I'm going to use this little band here. But after you get all done playing those games, what you're saying is, well, this noise is sort of uniform over this band. And therefore if I'm only transmitting in that band, if I never transmit anything outside of that band, there's no way you can tell what the noise is outside of that band. You all know that the noise you experience if you're dealing with a carrier frequency in the kilohertz band is very different than what you see in the megahertz band, is very different from what you see at 100 megahertz, is very different from what you see in the gigahertz band, is very different from what you see in the optical bands. And all of that stuff. So it doesn't make any sense to model the noise as being uniform spectral density over all of that region.

So the only thing we would ever want to do is to model the noise as having a uniform density over this narrowband that we're interested in. And that's how we define white Gaussian noise. We say the noise is white Gaussian noise if, in fact, when we go out and measure it we can measure it by passing it through a filter by finding this variance, which is exactly what we're talking about here, if what we get at one frequency is the same as what we get at another frequency. So long as we're talking about the frequencies of interest, then we say it's white.

Now how long do we have to make this measurement? Well, we don't make it forever, because all the filters we're using and everything else are filters which only have a duration over a certain amount of time. We only use our device over a certain amount of time. So what we're interested in is looking at the noise over some large effective time from minus  $T_0$  to plus  $T_0$ . We want the noise to be

effectively stationary within minus  $T_0$  to plus  $T_0$ .

And then what's the next step in the argument? You want the noise to be effectively stationary between these very broad limits. And then we think about it for a little bit, and we say, but listen. I'm only using this thing in a very small fraction of that time region. And therefore as far as my model is concerned, I shouldn't be bothered with  $T_0$  at all. I should just say mathematically, this process is going to be stationary, and I forget about the  $T_0$ . And I look at it in frequency, and I say I'm going to use this over my 10 megahertz or 100 megahertz, or whatever frequency band I'm interested in. And I'm only interested in what the noise is in that band. I don't want to specify what the bandwidth is. And therefore I say, I will just model it as being uniform over all frequencies.

So what white noise is, is you have effectively gotten rid of the  $T_0$ . You've effectively gotten rid of the  $w_0$ . And after you've gotten rid of both of these things, you have noise which has constant spectral density over all frequencies, and noise way which has constant power over all time. And you look at it, and what happens?

If you take the inverse Fourier transform of  $s_z$  of  $f$ , and you assume that  $s_z$  of  $f$  is just non-zero within a certain frequency band, what you get when you take the Fourier transform is a little kind of wiggle around zero. And that's all very interesting. If you then say well, I don't care about that. I just like to assume that it's uniform over all frequencies. You then take the Fourier transform, and what you've got is an impulse function.

And what the impulse function tells you is, when you pass the noise through a filter, what comes out of the filter, just like any time you deal with impulse functions -- I mean, the impulse response is in fact the response to an impulse. So that as far as the output from the filter goes, it's all very fine. It only cares about what the integral is of that pulse. And the integral of the pulse doesn't depend very much on what happens at enormously large frequencies or anything. So all of that's well behaved, so long as you go through some kind of filtering first. Unfortunately, as soon as you start talking about a covariance function which is an impulse, you're in real trouble.

Because the covariance function evaluated at zero is the power in the process. And the power in the process is then infinite. So you wind up with this process which is easy to work with any time you filter it. It's easy to work with because you don't have these constants capital T0 and capital W0 stuck in them. Which you don't really care about, because what you're assuming is you can wander around as much as you want in frequency, subject to the antennas and so on that you have. And you want to be able to wander around as much in time as you want to and assume that things are uniform over all that region.

But you then have this problem that you have a noise process which just doesn't make any sense at all. Because it's infinite everywhere. You look at any little frequency band of it and it has infinite energy if you integrate over all time. So you really want to somehow use these ideas of being effectively stationary and of being effectively bandlimited, and say what I want is noise which is flat over those regions.

Now what we're going to do after the quiz, which is on Wednesday, is we're going to start talking about how you actually detect signals in the presence of noise. And what we're going to find out is, when the noise is white in the sense -- namely, when it behaves the same over all the degrees of freedom that we're looking at -- then it doesn't matter where you put your signal. You can put your signal anywhere in this huge time space, in this huge bandwidth that we're talking about. And we somehow want to find out how to detect signals there. And we find out that the detection process is independent of what time we're looking at and what frequency we're looking at. So what we have to focus on is this relatively small interval of time and relatively small interval of bandwidth.

So all of this works well. Which is why we assume white Gaussian noise. But the white Gaussian noise assumption really makes sense when you're looking at what the noise looks like in these various degrees of freedom. Namely, what the noise looks like when you pass the noise through a filter and look at the output at specific instance of time. And that's where the modeling assumption is come in.

So this is really a very sophisticated use of modeling. Did the engineers who created

this sense of modeling have any idea of what they were doing? No. They didn't have the foggiest idea of what they were doing, except they had common sense. And they had enough common sense to realize that no matter where they put their signals, this same noise was going to be affecting them. And because of that, what they did is they created some kind of pseudo-theory, which said we have noise which looks the same wherever it is. Mathematicians got a hold of it, went through all this theory of generalized functions, came back to the engineers. The engineers couldn't understand any of that, and it's been going back and forth forever.

Where we are now, I think, is we have a theory where we can actually look at finite time intervals, finite frequency intervals, see what's going on there, make mathematical sense out of it, and then say that the results don't depend on what  $t_0$  is or  $w_0$  is, so we can leave it out. And at that point we really have our cake and we can eat it too. So we can do what the engineers I've always been doing, but we really understand it at this point.

OK. I think I will stop at that point.