

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: Review what random processes are a little bit now. And remember, in what we're doing here, we have pointed out that random processes in general, have a mean. The mean is not important. The mean is not important for random variables or random vectors either. It's just something you add on when you're all done. So the best way to study random variables, random vectors, and random processes, particularly when you're dealing with things which are based on base things being Gaussian, is to forget about the means, do everything you want to with the fluctuations, and then put the means in when you're all done. Which is why the notes do most of what they do, in terms of zero mean, random variables, random vectors, and random processes, and simply put in what the value of the mean is at the end.

OK so some of this has the mean put in. Some of it doesn't. So to start out with, a random process is defined by its joint distribution at each finite set of epochs. OK, that's where we started with all of this. How do you define a random waveform? I mean, you really have to come to grips with that question before you can see how you might answer the question. If you think it's obvious how to define a random process, then you really have to go back and think about it. Because a random process is a random waveform. It's an uncountably infinite number of random variables. So that defining what it is, is not a trivial problem. And people have generally agreed that the way to define it, be the test for whether you have defined it or not, is whether you can find the joint distribution at each finite set of epochs.

Fortunately most processes of interest can be defined in a much, much simpler way in terms of an orthonormal expansion. When you to find it in terms of an orthonormal expansion, you have the orthonormal expansion which is the sum over k , of a set of random variables, $Z_{sub 1}$, $Z_{sub 2}$ and so forth $Z_{sub k}$, times a set of

orthonormal functions. So all the variation on t is stuck into the set of orthonormal functions. All the randomness is stuck into the sequence of random variables. So at this point, you have a sequence of random variables, rather than a waveform.

Why is it so important to have something countable instead of something uncountable? I mean, mathematicians understand, in a deep mathematical sense, why that's so important. For engineers the argument is a little bit different. I think for engineers the argument is, no matter how small the interval of a waveform you're looking at, even if you shrink it, no matter what you do with it, you can't approximate it in any way and get rid of that uncountably infinite number of random variables. As soon as you represent it in terms of an orthonormal expansion though, at that point you have represented it as a countable sum. As soon as you represent it as a countable sum, you can approximate it by knocking off the tail of that sum. OK, and at that point you have a finite number of random variables, instead of an infinite number of random variables. We all know how to deal with a finite set of random variables. You've been doing that since you were in 6.041 OK. So if you have a countable set of random variables, the way to deal with them is always the same. It's to hope that they're defined in such a way that when you look at enough of them, all the rest of them are unimportant.

That sort of is the underlying meaning of what countable means. You can arrange them in a sequence, yes. But like when you count in school, after you name the first hundred numbers you get tired of it, and you realize that you understand the whole thing at that point. OK in other words, you don't have to count up to infinity to understand what it means to have a countable set of integers. You all understand the integers. You understand them intuitively. And you understand that no matter how big an integer you choose, somebody else can always choose one bigger than you've chosen. But you also understand that that's not important. OK?

So this is the way we usually define a stochastic process, a random process to start off with. The other thing that we have to do, is given a random process like this, we very often want to go from this random process to other random processes which are defined in terms of this random process. So almost always we start out with a

random process which is defined this way. And then we move from there to various other processes which we define in terms of this. So we never really have to deal with this uncountably infinite number of random variables. If we really had to deal with that, we would be in deep trouble. Because the only way we could deal with it, would be to take five courses in measure theory. And after you take five courses in measure theory, it's not enough. Because at that point, you're living in measure theory. And most of the mathematicians that I know can't come back from measure theory to talk about real problems. So that every time they write a paper, which looks like beautiful mathematics, it's very difficult to interpret whether it means anything about real problems. Because that's where things get hard. OK so the point is, if you define random processes in this way, then you always have some sort of grounding about what you're talking about. Because you have a countable number of random variables here. You know that if you're dealing with anything that makes any sense, only a finite number of them are going to be important. The thing you don't know is how many of them you need to be important. OK, so that's why we take an infinite sum here. Because you don't know ahead of time, how many you'd need to deal with.

OK, so we then started to talk about stationarity. The process Z of t is stationary if Z of t sub 1 up to Z of t sub k and Z of t sub 1 plus τ up to Z of t sub k plus τ have the same distribution. OK, so you take any finite set of random variables in this process, you shift them all to some other place, and you ask whether they have the same distribution. How do you answer that question? God only knows. I mean, what you need to be able to answer that question is some easy way of finding those joint probabilities. And that's why you deal with examples. That's why we deal with jointly Gaussian random variables. Because once we're dealing with jointly Gaussian random variables, we can write down those joint distributions just in terms of covariance matrices. Well it means, of course, if you want to deal with things that are not a zero mean. And after you learn how to deal with covariance matrices, at least you're dealing with a function of a finite number of variables. So it's not so bad.

OK so the argument is, this is what you need to know if you're going to call it

stationary. But we have easier ways of testing for that. And in fact this Wide Sense Stationary we said, if the covariance matrix, mainly the expected value of Z of t sub 1 times the expected value of Z of t sub 2 is equal to some function of just t sub 1 minus t sub 2. In other words, where the expected value of this value times this value is the same, if you shift it over by some amount. OK, you see the difference between? I mean the difference between these two things is one, that in the definition for stationarity, you need an arbitrarily large set of random variables here. You shift them over to some other point here. And you need the same joint distribution over this arbitrarily large set of random variables. When you get into dealing with the covariance matrix, all you need for Wide Sense Stationarity is you only have to deal with two random variables here, and the shift of those two random variables here. Which really comes out to a problem involving just two random variables t sub 1, t sub 2, and the fact that it's the same no matter where you shift the t sub 1 to. It's only a function of the difference between t sub 1 and t sub 2. So it's a whole lot easier.

And then we pointed out that since jointly Gaussian random variables, and since the Gaussian random process, it's completely determined by its covariance function, a zero mean Gaussian process. If you'll forgive me, I'm not going to keep coming back and saying that. I will just be thinking solely today in terms of zero mean processes. And you should think solely in terms of zero mean processes. You can sort out for yourselves whether you'd need a mean or not. That's sort of a trivial addition to all of this.

OK, so it's Wide Sense Stationary. Well we already said that. Well, here I put in the mean. OK, Wide Sense Stationary implies that it's stationary for Gaussian processes. OK, so in other words, this messy looking question of asking whether a process is stationary or not, with all of these random variables here and all of these random variables here, becomes trivialized for the case of a Gaussian random process, where all you need to worry about is the covariance function. Because that's the thing that specifies the whole process. OK. That's part of what we did last time.

We set an important example of this. And you don't quite know how important this is yet. But this is really important because all of the Gaussian processes you want to talk about can be formed in terms of this process. So it's nice in that way. And the process is Z of t is the sum of an orthonormal expansion. But the orthonormal functions now are just the time shifted sinc functions. We know the time shifted sinc functions are orthogonal. So we just have this set of random variables. And we say OK, what we're going to assume here is that these random variables, well here I put in the mean. Let's forget about the mean. The expected value of V sub k times V sub i , namely any two of these random variables expected value of the pair, is equal to zero if k is unequal to i . OK, in other words, the random variables are in some sense orthogonal to each other. But let's save the word orthogonal for functions, and use the word correlated or uncorrelated here for random variables. So these random variables are uncorrelated. So we're dealing with an expansion where the functions are orthogonal, and where the random variables are uncorrelated.

And now, what we're really interested in, is making these random variables Gaussian. And then we have a Gaussian random process with this whole sum here. We're interested in making these have the same variance in all cases. And therefore what we have is process which is stationary. And its covariance function is just σ^2 times this sinc function here. There's an error in the notes, in lecture 16, about this. It will be corrected on the web. This is in the notes. This quantity here is left out. So you can put that back in if you want to. Or you can get the new notes off the web after a few hours or so. This is not put on the web yet. I just noticed yesterday. In fact I noticed it this morning.

OK, the sample functions of a Wide Sense Stationary non-zero process are not L^2 . When I talk about a non-zero process I'm not talking about zero mean, I'm saying this peculiar random process which is zero everywhere. In other words, a random process which really doesn't deserve to be called a random process. But unfortunately by the definitions, it is a random process. Because with probability one, Z of t is equal zero everywhere. And that just means you have a constant, which is zero everywhere. So you're not interested in it.

OK, if you take the sample functions of any non-trivial Wide Sense Stationary random process, they dribble on forever. And they have the same-- no matter how far you go out in time-- the V sub k 's which are the samples of this process, sample random variables of the process, weigh out. These things all have the same variance. So that this is stationary because it just keeps going on forever. OK and as we said last time, that runs into violent conflict with our whole idea here of trying to understand what L sub 2 functions are all about. Because what we would like to be able to do is stick the L sub 2 functions as sample values of these processes, have the waveforms that we send the L sub 2. When you add up two L sub 2 functions, you get an L sub 2 function. And therefore, all the sample values of all the things that happen with probability one are all L sub 2 functions. As soon as you do the most natural and simple thing with random processes, you wind up with something which cannot be L sub 2 anymore. But it's not L sub 2 in only a trivial way. OK in other words, it's not L sub 2 because it keeps going forever. It's not L sub 2 for the same reason that $\sin x$ is not L sub 2, or 1 is not L sub 2.

OK, and we've decided not to look at those functions as reasonable functions for the waveforms that we transmit or receive. But for random processes, maybe they're not so bad. OK. But we don't know yet. We started to talk about something called effectively Wide Sense Stationary, which the notes talk about quite a bit. Which to say, OK we will assume that the process is stationary over some very wide time interval. We don't know how wide that time interval is, but we'll assume that it's finite. In other words, we're going to take this random process, whatever it is, you can define it, a nice random process this way. And then we're going to get our scissors, and we're going to cut it off over here and we're going to cut it off over here. So it's then, time-limited over this very broad band.

Why do we want to do this? Because again, it's like the integers which are countable. You don't know how far out you have to go before you're not interested in something anymore. But you know if you go out far enough, you can't be interested in it anymore. Because otherwise you can't take limits. You can't do anything interesting with sequences. OK, so here the idea is. If you go out far enough, you don't care. This has to be the way that you model things because any

piece of equipment that you ever design is going to start being used at a certain time, and it's going to stop being used at a certain time. And about halfway after when it stops being used, the people who make it will stop supporting it. And that's to hurry on the time at which you have to buy something new.

I bought a new computer about six months ago. And I find that Microsoft is not supporting any of the stuff that it loaded into this damn computer anymore. Six months old, so you see why I'm angry about this matter, of things not being supported. So in a sense they're not stationary after that. So you can only ask for things to be stationary over a period of six months or so. OK. But for electronic times, when you're sending data at kilobits per second or megabits per second, or gigabits per second as people like to do now, that's an awful long time. You can send an awful lot of bits in that time. So stationary means stationary for a long time. The notes do a lot of analysis of this. I'm going to do a little of that analysis here in class again. Not because it's so important to understand the details of it, but to understand what it really means to have a process be stationary, and to understand that doesn't really destroy any of the L^2 theory that we built up.

OK. So the covariance function is L^2 in cases of physical relevance. That's the thing. That's one of the things we need. We want the sample functions also to be L^2 in cases of physical relevance. If you have a function here, this is just a function of time now at this point. There's nothing random about this. It's just a statistic of this random process. But it's a nice well-defined function. We're talking about real random processes here. So what can you say about this function? Is it symmetric in time? How many people think it must be symmetric? I see a number of people. How many people think it's not symmetric? Well it is symmetric. It has to be symmetric because its expected value of Z of t sub 1 times Z of t sub 2, and if you flip the roles of those two, you have to get the same answer. So it is symmetric. It is real.

We're going to talk about the Fourier transform of this before trying to give any relevance or physical meaning to this thing called spectral density. We'll just say this is the Fourier transform of the covariance function for a Wide Sense Stationary

process. So we have some kind of Fourier transform. We'll assume this is L^2 and everything. We won't worry about any of that. So there is some function here which makes sense. Now if k is both real and symmetric, what can you say about its Fourier transform? If you have a real function, what property does this Fourier transform have?

AUDIENCE: [UNINTELLIGIBLE]

PROFESSOR: It's conjugate symmetric, yes. And furthermore, if k is symmetric itself, then you can go back from this thinking you have a symmetric transform, and realize that this has to be real. OK. So spectral densities of real processes are always both real and symmetric. We will define what we mean by spectral density later for complex processes. And spectral densities are always real. They aren't always symmetric if you have a complex process. But don't worry about that now. There are enough peculiarities that come in when we start dealing with complex processes, that I don't want to worry about it at all.

So spectral density is real and symmetric. And so far it's just a definition. OK, but the thing we found about this sinc process, is that the sinc process in fact, is a stationary process. It's a Wide Sense Stationary process for whatever variables you want to put in here. And if these variables are IID and they're Gaussian and zero mean, then this process is a zero mean, stationary Gaussian random process. And it has a spectral density. This turns out to be a sinc function. The Fourier transform of it is a rectangular function. So this process has a spectral density which is constant out to a certain point, and then drops off to zero. So it's nice in that sense. Because you can make this be flat as far as you want to, and then chop it off wherever you want to. And we'll see that when we start putting a process like this through filters, very nice things happen. OK.

So we're familiar, relatively familiar, and conversant with at least one Gaussian random process. OK, we talked about linear functionals last time before the quiz. And we said a linear functional is a random variable, V , which is simply this integral here. We've talked about this integral a lot where Z of t is a function, and g of t is a

function. If the sample values of Z of t are L sub 2, then this integral is very well-defined. It turns out, also, that if g of t is L sub 2, and these sample values are bounded, then all of this works out very nicely also. And we'll find other ways to look at this as we move on. OK but what this means is that for all sample points in this entire sample space, in other words we're dealing with a probability space where we have a bunch of processes running around, we have data coming in. We're transmitting the data. We have data that gets received. We're doing crazy things with it. All of this stuff is random. We might have other processes doing something else. Big complicated sample space here for all of the elements in it, the sample value of the random variable, V , is just the integral of the sample value of the process here. In other words for each ω , this process is simply a waveform. So it's this waveform times the function g of t , which we think of like that. So if g of t is time limited in L sub 2, if g of t is time limited as far as these sample functions are concerned, we don't give a fig about what Z of t is outside of that interval. And if we're going to only deal with linear operations on this process where those linear operations are constrained to some interval, then we don't care what the process does outside of that interval. And since we don't care what the process does outside of that interval, we can simply define the process in terms of what it's doing from some large negative time to some large positive time. And that's all we care about.

And so we wanted to find effective stationarity, as something which obeys the law of stationarity over that interval. And we don't care what it does outside of that interval, OK? In other words, if Microsoft stops supporting things after six months, we don't care about it. Because everything we're going to do we're going to get done within six months. And if we don't get it done within six months we're going to have a terrible crash. And we're going to throw everything away and start over again. So it doesn't make any difference at that point.

OK, so we have a process now we'll assume is effectively stationary within two time limits. And our definition is if the covariance function, which is a function of two values, t and τ , namely it's the expected value of Z of t times expected value of Z of τ . And our definition of effective stationarity is that this is equal to a function of t minus τ , whenever t and τ are in this box. OK, and we drew figures last time for

what that meant. We drew this box, you know, bounded by $T \pm T_0/2$. And what this effective stationarity means, is that on all these diagonal lines, the covariance function is constant on those diagonal lines. And that gives us the same effect whenever we're passing this process any time we're calculating the inner product of a sample value of the process with some function, which is contained within those limits. All we need to know about is just what the process is doing within $T \pm T_0/2$. Nothing else matters. And therefore, something is effectively stationary within those limits means that we get the same answer here, whether or not it's stationary. Why is that important? It's important because the things that you can do, the simple things that you can do with stationary processes are so simple that you'd like to remember them, and not remember all these formulas with capital T 's floating around in them. Because having a capital T in it is just as crazy as assuming that it's stationary, because you never know how long it's going to be until Microsoft stops supporting its software. I mean, if you knew you'd never buy their products, right? So you can't possibly know. So you assume it's going to go on forever. But then, when we try to ask what does that mean, we say what it really means is over some finite limits, which are large compared to anything we're going to deal with that at all of these results hold. So eventually we're trying to get to the point where we can leave the T out of it. But we're going through this so we can say, OK there's nothing magical that happens in the limit as T goes to infinite. Mainly we're trying to really derive the fact that all these results can be established over a finite interval of time. And therefore you don't care.

OK. So if this covariance function, single variable covariance function, is less than infinity, what does that mean? What is a single variance, single value, covariance function of zero mean? It's the expected value of Z of T , times Z of T . In other words, it's the variance of the process at T for any T within $T \pm T_0/2$, or $T \pm T_0/2$. OK? So if that variance is finite, then you can just integrate over the sample values of the process, and you get something finite. In other words this is what you need for the sample functions to be L^2 with probability one. As soon as you have this, then you're in business with

all of your L^2 theory. And what does your L^2 theory say? It really says you can ignore L^2 theory.

OK, that was the nice thing about it. That was why we did it. OK in other words, the whole thing we're doing in this course is we're going through complicated things sometimes so that you can know how far you can go with the simple things. And we always end up with the simple things when we're all done. OK, and that's the whole principle. I mean we could just do the simple things like most courses do. Except then, you would never know when it applies and when it doesn't apply. So, OK. So there we are.

Let's talk about linear filtering of processes. The notes don't do quite enough of this. So when I get to the place where you need it I will-- and some of it's on this slide-- I'll tell you. We're taking a random process and we're passing the sample waveform of that random process through a linear filter, a linear time and variant filter, so some other random process comes out. OK, and now the output at some time τ , is just the convolution of the input Z of t times this filter response h of τ minus t . And you can interpret this in terms of these sample functions the way we did before. But now we already sort of understand that. So we're just interpreting in terms of if there's a random variable V , which is the value of the process at time at epoch τ , which is given in this way. It's a random variable when you pass the process through the filter.

OK, if Z of t is effectively Wide Sense Stationary with L^2 sample functions, and if h of t is non-zero only within finite limits, minus A to plus A , what's going to happen? When you pass the sample waveforms through a linear filter, which is bounded between minus A and plus A , that linear filter cannot do anything with the inputs. Namely the output that comes out of there at some time, T , can't depend on the input anymore than A away from that output. OK in other words, let me draw a picture of that. Here's some time where we're observing V of t . And V of t can depend on Z of t only in this region here, Z of t minus A up to Z of t plus A . OK that's what the convolution says. That's what those filter equations say. It says that this output, here at this time, depends on the input only over these finite limits. And

again remember we don't care about realizability here. Because the timing at the receiver is always different from the timing at the transmitter.

OK so what that says then is, what does it say? It says that we know that Z of t is Wide Sense Stationary over these big limits minus T sub 0 to t sub 0 from six months ago until six months from now. And this linear filter is non-zero only over one millisecond. It says that the process which comes out, is going to be Wide Sense Stationary over minus six months plus one millisecond to six months minus one millisecond. And that's just common sense. Because the filter isn't moving the process any more than that little bit.

OK. So V of t then, is going to be Wide Sense Stationary in L sub 2 within those slightly smaller limits. The covariance function is going to be the same thing that it was before if you had a completely stationary process, if you only worry about what's going on within those limits, T sub 0 minus A to T sub zero plus A . You can take that big mess there, and view it more simply as a convolution. Let's see, this part of it is a convolution of h of t with K tilde, which is the covariance, which is a function of one variable. And then it's convolved with-- and here I put in the complex conjugate of h , because it's easier to do it here. Because at some point, we want to start dealing with complex random processes. And for filters it's easy to do that. For linear functionals it's a little harder. So I want to stick to real processes. If we're dealing with filtering, I can simply define this covariance function as the expected value of Z of t times Z complex conjugate of τ . And when you do that, we're taking this integral, which is the convolution of h of t with K tilde. And then we're convolving it with h complex conjugate of minus t . Because the t gets turned around in there. And what that says when you take the Fourier transform of it, at least what I hope it says-- I'm not very good at taking Fourier transforms. And you people, some of you are very good at it. We get the spectral density of the process, Z of t times the magnitude of \hat{h} of f .

Now this is the formula, which is not in the notes I just found out. And it's a very important formula. So you ought to write it down. It says what the spectral density is on the output of the filter, if you know the spectral density of the input of the filter.

And it's a kind of a neat, simple formula for it. It says, you pass a random process. I mean, it's like the formula that you have when you take a waveform and you pass it through a linear filter. And you know it's kind of easier to look at that in the frequency domain where you multiply, than it is to look at it in a time domain where you have to convolve. Here things become even simpler because all we have to do is take the spectral density, which is now a function of frequency, multiply it by this magnitude squared, and suddenly we get the spectral density at the output.

Now what happens when you take this nice sinc filter, this sinc Gaussian process that we have, which is flat over as large a bandwidth as you want to make it. And you pass that process through a linear filter. What do you get out? You get a process out which has any spectral density that you want within those wide limits. OK, so just by understanding the sinc Gaussian process and by understanding this result, you can create a process which has any old spectral density that you want. And if you can create any old spectral density that you want, you know you can create any old covariance function that you want. And all of this is good for Wide Sense Stationarity so long as your filter doesn't extend for too far. Which says all of our theory works starting at some negative time going to some positive time, taking filters that only exist for a relatively small time. If it's Wide Sense Stationary and it's Gaussian, then everything works beautifully. You can simply create any old spectral density that you want. If you have some waveform that's coming in, you can filter it and make the output process look whatever you want to make it look like, so long as you have the stationarity property. Now you know the secret of why everybody deals with stationary processes. It's because of this formula. It's an incredibly simple formula. It's an incredibly simple idea. You see now we know something else. We know that it also applies over finite, but large time intervals.

OK. So just to spell out the conclusions once more, a Wide Sense Stationary process is Wide Sense Stationary after filtering. So if you start out with Wide Sense Stationary, it's Wide Sense Stationary when you get done. If it's effectively Wide Sense Stationary, it's effectively stationary with a reduced interval of effective stationarity after you filter. OK in other words, if you start out with a process which is effectively stationary from minus T sub 0 to plus T sub 0, then after you filter it with

any filter whose impulse response is limited in time, you get something which is effectively stationary within that six months minus one millisecond period of time. Yes?

AUDIENCE: [UNINTELLIGIBLE]

PROFESSOR: What?

AUDIENCE: Don't you differentiate [UNINTELLIGIBLE]

PROFESSOR: Effectively Wide Sense Stationary, thank you. I try to differentiate between them, but sometimes I'm thinking about Gaussian processes, where it doesn't make any difference. Is effectively Wide Sense Stationary with reduced interval after filtering. OK, good.

The internal covariance, in other words, the covariance function within these intervals, the spectral density, and the joint probability density aren't affected by this interval of effective stationarity. In other words, so long as you stay in this region of interest, you don't care what $T_{sub 0}$ is. And since you don't care what $T_{sub 0}$ is, you forget about it. And for now on we will forget about it. But we know that whatever we're dealing with, it's effectively stationary within some time period, which is what makes the functions $L_{sub 2}$. We don't care how big that is. So we don't have to specify $T_{sub 0}$. We don't have to worry about it. We just assume it's big enough and move on. And then if it turns out it's not big enough, if your system crashes when one of Microsoft's infernal errors come up, then you worry about it then. But you don't worry about it before that.

OK so now we can really define white noise and understand what white noise is. White noise is noise that is effectively Wide Sense Stationary over a large enough interval to include all time intervals of interest. Also on the other part of white noise, is that the spectral density is constant in f over all frequencies of interest. OK in other words, white noise, really you have to define it within terms of what you're interested in. And that's important to understand. If you're dealing with wireless channels, you sometimes want to assume that the noise you're seeing is actually

white noise. But sometimes you do things like jumping from one frequency band to another frequency band to transmit. And when you jump from one frequency band to another frequency band, the noise might in fact be quite different in this band, than it is in this band. But you're going to stay in these bands for a relatively long time. So as far as any kind of analysis goes which is dealing with those intervals, you want to assume that the noise is white. As far as any overall analysis that looks at all of the intervals at the same time, in other words, which is dealing with some very long-term phenomenon, you can't assume that the noise is white. OK in other words, white noise is a modeling tool. It's not something that occurs in the real world. But it's a simple modeling tool. Because what do you do with it? You start out by assuming that the noise is white. And then you start to get some understanding of what the problem is about. And after you understand what the problem is about, you come back and say, is it really important for me that the noise is going to have a constant spectral density over this interval of interest? And at that point you say, yes or no.

OK. It's important to always be aware that this doesn't apply for times and frequencies outside the interval of interest. And if the process is also Gaussian, it's called white Gaussian noise. So white Gaussian noise is noise which has a flat spectral density over this effective stationarity period that we're interested in, over some very large period of time, minus T sub 0 to plus T sub 0. And over some very large frequency interval, minus W up to plus W or even better over some limited frequency band in positive and negative frequencies, the spectral density is constant. You can't define the spectral density until after you look at this effective period of stationarity. But at that point, you can then define the spectral density and see whether it looks constant over the period you're interested in. You can either see whether it is, or if you're an academic and you write papers, you assume that it is. OK. The nice thing about being an academic is you can assume whatever you want to assume.

OK let's go back to linear functionals. The stuff about linear functionals is in fact in the notes. This will start to give us an idea of what spectral density really means. So if I have a Wide Sense Stationary process, a linear functional, well this is what a

linear functional is in general. But for a Wide Sense Stationary process, we have the single variance, single variable covariance function. The expected value of the product of two random variables of this sort is just this integral here. This is what we did before. If you don't remember this formula, good. It's just what you get when you take the definition of $V_{sub\ i}$, which is the integral of $g_{sub\ i}$ of t with Z of t . And then you multiply it by $V_{sub\ j}$. And you take Z of t and Z of τ . And you take the expected value of the two. And then you integrate the whole thing. And it sort of comes out to be that, if I've done it right. You can express this in this terms, namely as the integral of $g_{sub\ i}$ of t times a convolution of this function times this function. You can then use Parseval's identity. Because this convolution gives you a function of t , OK. When you integrate this over τ , this whole thing is just a function of t . OK. So what we're dealing with is the integral of a function of t times a function of t . And Parseval's relation says that the integral of a function of time, times a function of time is equal to the integral of the Fourier transform times the Fourier transform. OK. You all know this. This is almost the same as a trick we used with linear filters, just slightly different. So that says that this expected value is equal to this product here. And here I've used the complex conjugate for the Fourier transform of $g_{sub\ j}$, because I really need it there if g is not a symmetric function.

OK so if these Fourier transforms, $g_{sub\ i}$ of t and $g_{sub\ j}$ of t are non-overlapping in frequency, what does this integral go to? And this is absolutely wild. I mean if this doesn't surprise you, you really ought to go back and think about what it is that's going on here. Because it's very astonishing. It's saying that stationarity says something you would never guess in a million years that it says. OK it says, that if this and this do not overlap in frequency, this integral is 0. OK in other words, if you take a linear functional over one band of frequencies, it is uncorrelated with every random variable that you can take in any other band of frequencies. So long as the function is Wide Sense Stationary, this uncorrelated effect takes place. If you're dealing with the Gaussian random process, it's also stationery. Uncorrelated means independent. And what happens then, is that expected value of $V_{sub\ i}$ and $V_{sub\ j}$ is zero. $V_{sub\ i}$ and $V_{sub\ j}$ are statistically independent random variables. And it says that whatever is happening in one frequency band of a stationary Gaussian

process is completely independent of what's happening in any other band.

Now if any of you can give me an intuitive reason for why that is, I would love to hear it. Because I mean it's one of those-- well after you think about it for a long time, it starts to sort of make sense. If you read the appendix in the notes, the appendix in the notes sort of interprets it a little bit more. If you take this process, if you limit it in time and you expand it in a Fourier series, what's going to happen to those coefficients in the Fourier series? If the process is stationary, each one of them, if you think of them as being complex coefficients, the phase in each one has to be random and uniformly distributed between zero and 2π . In other words, if you have a phase on any sinusoid, which is either deterministic or anything other than uniform, then that little slice of the process cannot be Wide Sense Stationary. And there's nothing in any other frequencies that can happen to make the process stationary again. OK in other words, when you expand it in a Fourier series, you have to get coefficients at every little frequency slice, which have uniform phase to them. In other words, the Gaussian random variables that you're looking at, are things that are called proper complex Gaussian random variables which have the same variance for the real part as for the imaginary part. When you look at them as a probability density, you find circles. In other words, you find circularly symmetric random variables at every little frequency interval. And that's somehow, what stationarity is implying for us. So it's a pretty important thing. And it doesn't just apply to white noise. It applies to any old process with any old spectral density, so long as we've assumed that it's Wide Sense Stationary.

OK. So, different frequency bands are uncorrelated for Gaussian Wide Sense Stationary processes. Different frequency bands are completely independent of each other. OK, as soon as you have stationary noise, looking at one frequency band to try to get any idea of what's going on in another frequency band, is an absolute loser. You can't tell anything from one band about what's going on in another band. So long as you're only looking at Gaussian noise. If you're sending data which is correlated between the two bands, then of course it's worthwhile to look at both bands. But if you're sending data that's limited to one frequency band, then at that point, there's no reason to look anywhere else. We're going to find that

out when we start studying detection. It'll be a major feature of understanding detection at that point. OK, and one of the things you want to understand with detection, is what things can you ignore and what things can't you ignore. And this is telling us something that can be ignored. It's also starting to tell us what this spectral density means. But I'll interpret that in the next slide I hope.

OK if you take this set of functions $g_{sub j}$ of t , in the notes I now convert that set of functions to $V_{sub j}$'s instead of $g_{sub j}$'s. If I start out with that set of functions these-- oh dear, this is a random variable-- $V_{sub j}$ is the component of the process in the expansion, in that orthonormal set. So in other words, you can take this process, you can expand it into an orthonormal expansion using random variables and these orthonormal functions. If $g_{sub j}$ of t is a narrow band, and $S_{sub Z}$ of f is constant within that band-- in other words we don't need white noise here. All we need is a spectral density which changes smoothly. And if it changes smoothly, we can look at a small enough interval. And then what happens? When we take the expected value of $V_{sub j}$ squared, what we get is this times this times this-- OK the product of these three terms and since this-- that's supposed to be a j there.

OK so what's happening is these are narrow band, this is just a constant within that narrow band. So we can pull this out and then all we have is the integral $g_{sub i}$ of f times $g_{sub i}$ complex conjugate of f , which since these functions are orthonormal, is just one. So what we find is that random variable there is simply the value of the spectral density. Now we don't use that to figure out what these random variables are all about. We use this to understand what the spectral density is all about. Because the spectral density is really the energy per degree of freedom in the process at frequency f . In other words, if we use orthonormal functions which are tightly constrained in frequency, they just pick out the value of the spectral density at that frequency. $S_{sub Z}$ of $f_{sub 0}$ is just sort of a central frequency in this band, or integrating over here, OK. So the interpretation of the spectral density is that it's the energy per degree of freedom in any kind of orthonormal expansion you want to use, so long as those functions are tightly limited in frequency.

So that's what spectral density means. It tells you how much energy there is in the

noise, at all these different frequencies. It tells you, for example at this point, that if you start out with a sinc Gaussian process like we were talking about before, pass it through a filter, what the filter is going to do is change the spectral density at all the different frequencies. And the interpretation of that which now makes perfect sense when you're going through a filter, is the amount of energy in each of those frequency bands is going to be changed exactly by the amount that you filtered it.

OK so we have a nice thing there. S_{ZZ} of f is the energy per degree of freedom in the process at frequency f . For a Gaussian Wide Sense Stationary process, this is the energy per degree of freedom at f . And the noise at all different frequencies is independent. That's what we were saying before. And after filtering, this independence is maintained. In other words, you pass a Gaussian random process through a filter and you still have independence between all the different frequency bands. It's an absolutely remarkable property. Just this property of the process acting the same at each time as it acts at each other time, that this leads to this amazing property of things being independent from one frequency to another is really worth thinking about. And it's one of the things that people use when they actually design systems. I mean it's almost natural to use if you don't think about it. If you start thinking about it then it becomes even more worthwhile. OK should I start detection? Let me just start to say a few things about detection, then I want to stop in time to pass the quiz back. OK let's talk about where detection fits in with our master plan of what we've been doing. We started the course talking about what was over on this side of the channel business, namely all the source coding and all of those things. We then started to talk about how do you do signal encoding. In other words, how do you map from binary digits coming in, into signals. Well we never talked about this problem. We only talked about this problem, when you didn't have any noise. We carefully avoided dealing with this problem at all. Then we talked about baseband modulation for PAM and for QAM. For PAM it was real, for QAM it was complex. Then we talked about how you go from baseband frequencies up to passband frequencies. Now we've been talking about what happens when you add white Gaussian noise, or any other noise at passband. Then you go from passband down to baseband. When I do this, we're going to have to come back at

some point and deal with the question of what happens when you take passband white Gaussian noise and convert it down to baseband. Because some kind of funny things happen there. And we have to deal with it with a little bit of care. But let's forget about that for the time being, and suppose everything works out all right.

Then we go through our baseband demodulator. Finally something comes out of the baseband demodulator, which we hope is the same as what came out of the signal encoder. But what we're trying to do at this point now, is when the noise is here to say, this is in some sense, going to be what we put in plus noise. And for the time being, in terms of looking at detection, we will just assume that. We will assume that this thing down here is what went in, plus noise. And assume that all this process of passing white Gaussian noise through this junk gives us just a signal plus noise.

OK, so what a detector does is it observes a sample value of this random variable V , or it observes a vector, or observes a random process. It observes whatever it's going to observe. And it guesses the value of another random variable. And we'll call the other random variable H . We had been calling it other things for the input. It's nice to call it H because statisticians always call this a hypothesis. And I think it's easier to think of what's going on if you think of it as a hypothesis type random variable. Namely the detector has to guess. If we have a binary input, it has to guess at a binary output. It has to flip a coin somehow and say, I think that what came in is a zero, or I think what came in is a one.

The synonyms for detection are hypothesis testing, decision making, and decoding. They all mean exactly the same thing. There's no difference whatsoever between them, except the fields that they happen to deal with. And each field talks about it in a different sense. The word detection really came mostly from the radar field where people were trying to determine whether a target was there or not. And hypothesis testing has been around for a great deal longer. Because all scientists from time immemorial have had to deal with the question of have you collect a bunch of data about something. And it's always noisy data. There's always a bunch of junk that comes into it. And you have to try to make some conclusions out of that. And

making conclusions out of it is easier if you know there's only a finite set of alternatives which might be true. In a presidential election, after you know who the candidates are, it's a binary decision. The people who try to guess who people are going to vote for, are really doing binary decision making. They don't have very good probability models. And it certainly isn't a stationary process. But in fact, it is a detection problem, the same as the detection problem we have here.

So all these problems really fall into the same category. We're fortunate here at having one of the cleanest probabilistic descriptions for detection that you will ever find. So if you want to study decision making, you're far better off trying to learn about it in this context and then using it in all these other contexts, than you are doing something else. If you go back and look at old books about this, you will be amazed at the philosophical discussions that people would go through about what makes sense for hypothesis testing, and what doesn't make sense. And all of these arguments were based on a fundamental misapprehension that scientists had until relatively recently. And the misapprehension that they had, was that they refused to accept models as taking an intermediate position between physical reality, and then you have models, and then you have doing something with a model. And you do something with the model instead of doing something with the physical reality. And people didn't understand that originally. So they got the analysis of the model terribly confused with trying to understand what the physical reality was.

In other words, statisticians and probabilists were both doing the same thing. Both of them were trying to simultaneously determine what a reasonable model was, and to determine what to do with a model. And because of that, they can never decide on anything. Because they can never do what we've been doing, which says take a toy model, analyze the toy model. Figure out something from it about the physical reality. Then go back and look at reality and see what you need to know about reality in order to make these decisions, which is the way people do it now I think.

Particularly decision theory is done that way now. Because now we usually assume both that we know Apriori probabilities, it's called, for these inputs here. For the binary communication problem, we usually want to assume that these inputs are

equiprobable, in other words, if P_0 is equal to a half, and P_1 is equal to a half. And that's the Apriori probability of what's coming in. And we then want to assume some probabilistic model on v . Usually what we do, is we figure out what v is in terms of a conditional probability density, conditional probability density of v , conditional on either of these two outputs. You see for this model here, this is just made in heaven for us. Because this thing is an input plus a Gaussian random variable. So if we look at the conditional probability density of the output, which is signal plus noise, conditional on the signal is just a Gaussian random variable shifted over by the input.

OK. So we have these two probabilities, the input probability, this thing, which statisticians call a likelihood, which is the probability of this conditional on this. And in terms of that, we try to make the right decision. Now, I think I'm going to stop now so we can pass back the quizzes and so on. Before Wednesday, if you don't read the notes before Wednesday, spend a little bit of time thinking about how you want to make the optimal decision for this problem as we've laid it out now. What's surprising is that it really is a trivial problem. Or you can read the notes and accept the fact that it's a trivial problem. But there's nothing hard about it. The only interesting things in detection theory are finding easy ways to actually solve this problem which is conceptual trivial. OK.