PROFESSOR: OK. We are talking about jointly Gaussian random variables. One comment through all of this and through all of the notes is that you can add a mean to Gaussian random variables, or you can talk about zero-mean random variables. Here we're using random variables mostly to talk about noise. When we're talking about noise, you really should be talking about zero-mean random variables, because you can always take the mean out. And because of that, I don't like to state everything twice once for variables of processes without a mean, and once for variables of processes with a mean. And after looking at the notes, I think I've been a little inconsistent about that. I think the point is you can keep yourself straight by just saying the only thing important is the case without a mean. Putting the mean in is something unfortunately done by people who like complexity. And they have unfortunately got in the notation on their side. So anytime you want to talk about a zero-mean random variable. You have to say zero-mean random variable. And if you say random variable, it means it could have a mean or not have a mean. And of course the way the notation should be stated is you talk about random variables as things which don't have means. And then you can talk about random variables plus means as things which do have means, which would make life easier but unfortunately it's not that way.

So anytime you see something and are wondering about whether I've been careful about the mean or not, the answer is I well might not have been. And two, it's not very important. So anyway. Here I'll be talking all about zero-mean things. A k-tuple of zero-mean random variables is said to be jointly Gaussian if you can express them in this way here. Namely as a linear combination of IID normal Gaussian again random variables. OK. In your homework, those of you who've done it already, have realized that just having two Gaussian random variables is not enough to make

those two random variables be jointly Gaussian. They can be individually Gaussian but not jointly Gaussian.

This is sort of important because when you start manipulating things as we will do when we're generating signals to send and things like this, you can very easily wind up with things which look Gaussian and are appropriately modeled as Gaussian, but which are not jointly Gaussian.

Things which are jointly Gaussian are defined in this way. We will come up with a couple of other definitions of them later. But you've undoubtedly been taught that any old Gaussian random variables are jointly Gaussian. You've probably seen joint densities for them and things like this. Those joint densities only apply when you have jointly Gaussian random variables. And the fundamental definition is this. This fundamental definition makes sense, because the way that you generate these noise random variables is usually from some very large underlying set of very, very small random variables. And the Law of Large Numbers says that when you add up a very, very large number of small underlying random variables, you normalize the sum so it has some reasonable variance then that random variable is going to be appropriately modeled as being Gaussian.

If you at the same time look at linear combinations of those things for the same reason, it's appropriate to model each of the random variables you're looking at as a linear combination of some underlying set of noise variables which is very, very large. Probably not Gaussian. But here when we're trying to do this, because of the central limit theorem, we just model these as normal Gaussian random variables to start with. So that's where that definition comes from. It's why when you're looking at noise processes in the real world, and trying to model them in a sensible way this jointly Gaussian is the thing you almost always come up with.

OK. So one important point here, and the notes point this out. If each of these random variables Z sub i is independent of each of the others, and they have arbitrary variances, then because of this formula, the set of Z i's are going to be jointly Gaussian. Why is that? Well you simply make a matrix here A, which is

diagonal. And the elements of this diagonal matrix are sigma 1 squared, sigma 2 squared, sigma 3 squared, and so forth. If you have a vector Z, which is then sigma 1 squared down to sigma k squared times a noise vector N1 to N sub k. What you wind up with is Z sub i is going to be equal to -- I guess I don't want those squares in there. Sigma 1 up to sigma k -- Z sub i is going to be equal to sigma i, N sub i; N sub i is a Gaussian random variable with variance one and therefore Z sub i as Gaussian random variable with variance sigma sub i squared. OK.

So anyway, one special case of this formula is anytime that you want to deal with a set of independent Gaussian random variables with arbitrary variances, they are always going to be jointly Gaussian. Saying that they're uncorrelated is not enough for that. You really need the statement that they're independent of each other. OK. That's sort of where we were last time.

When you look at this formula, you can look at it in terms of sample values. And if we look at a sample value what's happening is that the sample value of the random vector Z, namely little z, is going to be defined as some matrix A times this sample value for the normal vector N. OK. And what we want to look at is geometrically what happens there. Well this matrix a is going to map a unit vector, E sub i into the i'th column of A. Why is that? Well you look at A, which is whatever it is, A sub 1, 1, blah, blah, blah up to A sub k, k. And you look at multiplying it by a vector which is zero only in the i'th position. And what's this matrix multiplication going to do? It's going to simply pick out the i'th column of this matrix here. OK. So a is going to map e sub i into the i'th column of A. OK.

So then the question is what is this matrix A going to do to some small cube in the n plane? OK. If you take a small cube in the n plane from 0 to delta along the n1 line is going to map into 0 to this point here, which is A e sub 1. This point here is going to map into A times e sub 2. This is just drawing for two dimensions of course. So in fact all the points in this cube are going to get map into this little rectangle here. OK. Namely, that's what a matrix times a vector is going to do.

Anybody awake out there? You're all looking at me as if I'm totally insane. OK.

Every everyone following this? OK. So perhaps this is just too trivial. I hope not. OK.

So unit cubes get mapped into rectangles here. If I take a unit cube up here, it's going to get mapped into the same kind of unit rectangle here. If I visualize tiling this plane here with little tiny cubes, delta on a side, what's going to happen? Each of these little cubes is going to map into a parallelogram over here, and these parallelograms going to tile this space here. OK. Which means that each little cube here maps into one of these rectangles. Each rectangle here maps back into a little cube here. Which means that I'm looking at a very special case of a here. I'm looking at a case where a is non-singular.

In other words, I can get the any point in this plane by starting with some point in this plane. Which means for any point here, I can go back here also. In other words, I can also write n is equal to A to the minus 1 times Z. And this matrix has to exist. OK. That's what you mean geometrically by a non-singular matrix. It means that all points in this plane get mapped into points in this plane. And get mapped into only one point in this plane, and get mapped into only one point in this plane. Where every point in this plane is the map of some point here. In other words you can go from here to there. You can also go back again. OK.

The volume of a parallelepiped here, and this is in an arbitrary number of dimensions is going to be the determinant of A. And you all know how to find determinants. Namely you program a computer to do it, and the computer does it. I mean it used to be we had to do this by an enormous amount of calculation. And of course nobody does that anymore. OK.

So we can find the volume of this parallelepiped, it's this determinant. And what does all of that mean? Well first let me ask you the question. What's going to happen if the determinant of a is equaled to zero? What does that mean geometrically? What does that mean here in terms of this two dimensional diagram?

AUDIENCE: [INAUDIBLE]

PROFESSOR: What?

AUDIENCE: Projection onto a line.

PROFESSOR: Yeah. This little cube here is simply going to get projected onto some line here. Like for example that. In other words it means that this matrix is not invertible for one thing, but it also means everything here gets mapped onto some lower dimensional sub-space here in general. OK. Now remember that for a minute, and we'll come back to that when we start talking about probability densities.

OK because of the picture we were just looking at, the density of the random variables Z at A times N, namely the density of Z at this particular value here is just the density that we get corresponding to the density of some point here mapped over into here. And what's going to happen when we take a point here and map it over into here? If you have a certain amount of density here, which is probability per unit volume. Now when you map it into here, and the determinant is bigger than zero, what you're doing is mapping a little volume into a big volume. And if you're doing that over small enough region where the probability density is of such essentially fixed, what's going to happen to the probability density over here? It's going to get scaled down, and it's going to get scaled down precisely by that determinant. OK.

So what this is saying is the probability density of the random variable z, which is linear combination of these normal random variables, is in fact the probability density of this normal vector N, and we know what that probability density is, divided by this determinant of a. In fact this is a general formula for any old probability density at all. You can start out with anything which you call a random vector N and you can derive the probability density of any linear combination of those elements in precisely this way. So long as this volume element is non-zero, which means you're not mapping an entire space into a sub-space.

When you're mapping an entire space into a sub-space and you define density as being density per unit volume, of course the density in this map space doesn't exist anymore. Which is exactly what this formula says. If this determinant is zero, it means this density here is going to be infinite in the regions where this z exists at

all, which is just in this linear sub-space, and what do you do about that? I mean do you get all frustrated about it? Or do you say what's going on and treat it in some sensible way? I mean the thing is happening here. And this talks about it a little bit. If a is singular, then A is going to map Rk into a proper sub-space. Determinant A is going to be equal to 0. The density doesn't exist.

So what do you do about it? I mean what does this mean if you're mapping into a smaller sub-space. What does it mean in terms of this diagram here? Well in the diagram here it's pretty clear. Because these little cubes here are getting mapped into straight lines here. Yeah? What?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Some linear combinations of this are being mapped into 0. Namely if this straight line is this way any Z which is going in this direction is being mapped into 0. Any vector Z which is going in this direction has to be identically equaled to 0. In other words some linear combination of n1 and n2 is a random variable which takes on the value 0 identically.

Why do you try to represent that in a probabilistic sense? Why don't you just take it out of consideration altogether? Here what it means is that z1 and z2 are simply linear combinations of each other. OK. In other words once you know what the sample value of z1 is, you can find the sample values z2. In other words z2 is a linear combination of z1. It's linearly dependent on z1, which means that you can identify it exactly once you know what z1 is. Which means you might as well not call it a random variable at all. Which means you might as well view this situation where the determinant is 0 as where the vector Z is really just one random variable, and everything else is determined. So you throw out these extra random variables, pseudo-random variables, which are really just linear combinations of the others. So you deal with a smaller dimensional set. You find the probability density of the smaller dimensional set, and you don't worry about all of these mathematical peculiarities that would arise otherwise. OK.

So once we do that, A is going to be non-singular. Because we're going to be left

with a set of random variables, which are not linearly dependent on each other. They can be statistically dependent on each other, but not linearly dependent, OK. So for all z then, since determinant A is not 0, the probability density at some arbitrary vector Z is going to be the normal joint density evaluated at a to the minus 1z divided by the determinant of A. OK. In other words we're just working the map backwards. This formula is the same as this formula, except instead of writing An here, we're writing z here. And when An is equal to z then little n has to be equal to A to the minus 1z.

And what does that say? It says that the joint probability density has to be equal to this quantity here. Which in matrix terms looks rather simple, it looks rather nice. You can rewrite this. This is a norm here, so it's this vector there times this vector. Where in fact when you want to multiply vectors in this way, you're taking inner product of the vector with this cell. These are real vectors we're looking at. Because we're trying to model real noise, because we're modeling the noise on communication channels. So this is going to be the inner product of A to the minus 1z with itself, which means you want to look at the transform of a to the minus 1z.

Now what's the transform of a to the minus 1z? It's z transform times a to the minus 1 transform. So we wind up with a to the minus 1 transform times a to the minus 1 times z. So we have some kind of peculiar bilinear form here. So for any sample value of the random vector Z we can find the probability density in terms of this quantity here. Which looks a little bit peculiar, but it doesn't look too bad. OK.

Now we want to simplify that a little bit. Anytime you're dealing with zero-mean random variables -- now remember I'm going to forget to say zero-mean half the time because everything is concerned with zero-mean random variables. The covariance of Z1 and Z2 is expected value of Z1 times Z2. So if you have a k-tuple Z, the covariance is a matrix whose i,j element is expected value of Zi times Zj. And what that means is that the covariance matrix, this is a matrix now, it's going to be the expected value of z times z transpose. Z is a random vector, which is a column random vector, Z transpose is a row-random vector, which is this simply turned upside down, turned by 90 degrees. Now when you multiply the components of this

vector together, you can see that what you get is the elements of this covariance matrix. In other words this is just standard matrix manipulation, which I hope most of you or at least partly familiar with. OK.

When we then talk about the expected values Z times Z transpose we can write this as the expected value of A times N, which is what z is. N transpose times A transpose, which is what Z transpose is. And miraculously here the N and the N transpose are in the middle here, where it's easy to deal with them, because these are normal Gaussian random variables. And when you look at this column times this row, since all diagonal elements are independent of each other, and all of them have variance one, the expected value of N times N transpose is simply the identity matrix. All the randomness goes out of this which it obviously has to because we're looking at a covariance which is just a matrix and not something random. So you wind up with this covariance matrix is equal to a rather arbitrary matrix A, but not singular, times the transpose of that matrix. OK.

We've assumed that A is non-singular and therefore it's not too hard to see the k sub Z is non-singular also. And explicitly case of Z mainly its co-variance matrix, the inverse of it, is A to the minus 1 transpose times A to the minus 1. Namely you take the inverse and you start flipping things and you do all of these neat matrix things that you always do. And you should review them if you've forgotten that. So that when we write our probability density, which was this in terms of this transformation here, what we get is the density of Z is equal to, in place of determinant of A, we get the square root of the determinant in the k sub Z. You probably don't remember that, but is what you get. And it's sort of a blah.

Here this is more interesting. This is minus 1/2 z transpose times Kz to the minus 1 times z. What does this tell you? Look at this formula. Is it just a big formula or what does it say to you? You got to look at these things and see what they say. I mean we've gone through a lot of work to derive something here.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Well it is Gaussian. Yes. I mean that's the way we define jointly Gaussian. But

what's the funny thing about this probability density? What does it depend on? What do I have to tell you in order for you to calculate this probability density for every z you want to plug in here? I have to tell you what this covariance matrix is. And once I tell you what the covariance matrix is, there is nothing more to be specified. In other words, a jointly Gaussian random vector is completely specified by its covariance matrix. And it's specified exactly this way by its covariance matrix. OK. There's nothing more there. So this says anytime you're dealing with jointly Gaussian, the only thing you have to be interested in is this covariance here. Namely all you need to have jointly Gaussian is somebody has to tell you what the covariance is, and somebody has to tell you also that it's jointly Gaussian. Jointly Gaussian plus a given covariance specifies the probability density. OK.

What does that tell you? Well let's look at an example where we just have two random variables, Z1 and Z2. then expected value of Z1 squared is the upper. Left hand element of that covariance matrix, which we'll call sigma 1 squared. The lower, right hand side of the matrix is K22, which we'll call sigma 2 squared. And we're going to let rho be the normalize covariance. We're just defining a bunch of things here, because this is the way people usually define this. So rho will be the normalized cross covariance. Then the determinant of the Kz is this mess here. For A to be non-singular, we have to have rho less than 1. If rho is equal to 1 then this determinant is going to be equal to 0, and we're back in this awful case that we don't want to think about.

So then if we go through the trouble of finding out what the inverse of K sub z is, we find this 1 over 1 minus row square times this matrix here. The probability density plugging this in is this big thing here. OK what does that tell you? Well the thing that it tells me is that I never want to deal with this, and I particularly don't want to deal with it if I'm dealing with three or more variables. OK. In other words the interesting thing here is the simple formula we had before, which is this formula. OK. And we have computers these days which say given nice formulas like this, their standard computer routines to calculate things like this. And you never want to look at some awful mess like this. OK. And if you put a mean into here, which you will see in every textbook on random variables and probability you ever look at, this thing becomes

so ugly that you were probably convinced before you took this class that jointly Gaussian random variables were things you wanted to avoid like the plague.

And if you really have to deal with explicit formulas like this, you're absolutely right. You do want to avoid them like the plague, because you can't get any insight from what that says, or at least I can't. So I say OK, we have to deal with this. But yet we like to get a little more insight about what this means. And to do this, we like to find a little bit more about what these bilinear forms are all about. Those of you who have taken any course on linear algebra have dealt with these bilinear forms. And played with them forever. And those of you who haven't are probably puzzled about how to deal with them. The notes have an appendix, which is about two pages long which tells you what you have to know about these matrices. And I will just sort of quote those results as we go. Incidentally those results are not hard to derive. And not hard to find out about. You can simply derive them on your own, or you can look at Strang's book on linear algebra, which is about the simplest way to get them. And that's all you need to do. OK.

We've said the probability density depends on this bilinear form z transpose times Kz to the minus 1 time z. What is this? Is this a matrix or a vector or what? How many people think it's a matrix? How many people think it's a vector? You think it's a vector? OK. Well in a very peculiar sense it is. How many people think it's a number? Good. OK. It is a number, and it's a number because this is a row vector. This is a matrix. This is a column vector. And if you think of multiplying a matrix times a column vector, you get a column vector. And if you take a row vector times a column vector, you got a number. OK. So this is just a number which depends on little z. OK.

Kz is called a positive definite matrix. And it's called a positive definite matrix, because this thing is always non-negative. And it always has to be non-negative because this refers to the -- if I put capital Z in here, namely if I put the random vector in here Z, then what this is, is the variance of a particular random variable. So it has to be greater than or equal to zero. So anyway K sub z is always non-negative definite. Here it's going to be positive definite, because we've already

assumed that the matrix A was non-singular, and therefore the matrix Kz has to be non-singular. So this has to be positive definite. So it has an inverse, K sub z minus 1 is also positive definite which means this quantity is always greater than zero, if z is non-zero.

You can take these positive definite matrices and you can find eigenvectors and eigenvalues for them. Do you ever actually calculate these eigenvectors and eigenvalues? I hope not. It's a mess to do it. I mean it's just as bad as writing that awful formula we had before. So you don't want to actually do this, but it's nice to know that these things exist. And because these vectors exist, and in fact if you have a matrix which is k by k, little k by little k, then there are k such eigenvectors and they can be chosen orthonormal. OK. In other words each of these Q sub i are orthogonal to each of the others. You can clearly scale them, because you scale this and scale this together. And it's going to maintain equality.

So you just scale them down so you can make them normalize. If you have a bunch of eigenvectors with the same eigenvalue, then the whole linear sub-space formed by that set of eigenvectors all have the same eigenvalue lambda sub i. Namely you take any linear combination of these things which satisfy this equation for a particular lambda. And any linear combination will satisfy the same the same equation. So you can simply choose an orthonormal set among them to satisfy this. And if you look at Q sub i and Q sub j, which have different eigenvalues, then it's pretty easy to show that in fact they have to be orthogonal to each other. So anyway when you do this you wind up with this form becomes just the sum over i of lambda sub i to the minus 1. Namely these eigenvalues to the minus 1. These eigenvalues are all positive. Times the inner product of z with Q sub i. In other words, you take whatever vector Z you're interested in here, you project it on these k orthonormal vectors Q sub i. You get those k values. And then this form here is just that sum there. So when you write the probability density in that way -- we still have this here we'll get rid of that in the minute -- you have e to the minus sum over i, these inner product terms squared divided by 2 times lambda sub i. That's just because this is equal to this. It's just substituting this for this in the formula for the probability

density.

OK. What does that say pictorially? Let me show you a picture of it first. It says that for any positive definite matrix and therefore for any covariance matrix, you can always find these orthonormal vectors. I've drawn them here for two dimensions. They're just some arbitrary vector q1; q2 has to be orthogonal to it. And if you look at the square root of lambda 1 times q1, you got a point here. You look at the square root of lambda 2 times q2, you got a point here. If you then look at this probability density here, you see that all the points on this ellipse here have to have the same sum of z times Q sub i. OK. It looked a little better over here. Yes. OK. Namely the points little z for which this is constant are the points which form this ellipse. And it's the ellipse which has the axes square root of lambda i times Q sub i. And then you just imagine it if it's lined up this way. And think of what you would get if you were looking at the lines of equal probability density for independent Gaussian random variables with different variances.

I mean we already pointed out if the variances are all the same, these equal probability contours are spheres. If you now expand on some of the axes, you get ellipses. And now we've taken these arbitrary vectors, so these are not in the directions we started out with, but in some arbitrary directions. We have q1 and q2 are orthornormal to each other, because that's the way we've chosen them. And then the probability density has this form which is this form. And the other thing we can do is to take this form here and say gee this is just a probability density for a bunch of independent random variables, where the independent random variables are the inner product of the random vector Z with Q sub 1 up to the inner product of z with Q sub k. So these are independent Gaussian random variables. They have variances lambda sub i. And this is the nicest formula for the probability density of an arbitrary set of jointly density of jointly Gaussian random variables. OK.

In other words what this says is in general if you have a set of jointly Gaussian random variables and I have this messy form here, all you're doing is looking at them in a wrong coordinate system. If you switch them around, you look at them this way instead of this way, you're going to have independent Gaussian random

variables. And the way to look at them is found by solving this eigenvector eigenvalue equation, which will tell you what these orthogonal directions are. And then it'll tell you how to get this nice picture that looks this way. OK. OK so that tells us what we need to know, maybe even a little more than we have to know, about jointly Gaussian random variables.

But there's one more bonus we get. And the bonus is the following. If you create a matrix here B where the i'th row of B is this vector Q sub i divided by the square root of lambda sub i, then what this is going to do is the corresponding element here is going to be a normalized Gaussian random variable with variance one and all of these are going to be independent of each other. OK. That's essentially what we were saying before. This is just another way of saying the same thing. That when you squish this probability density around and look at it in a different frame of reference, and then you scale the random variables down or up, what you wind up with is IID normal Gaussian random variables. OK. But that says that Z is equal to B to the minus 1 times N prime. Well so what? Here's the reason for so what. We started out with a definition of jointly Gaussian, which is probably not the definition of jointly Gaussian you've ever seen if you've seen this before. Then what we did was to say, OK if you have jointly Gaussian random variables and they're not linearly independent and none of them are linearly dependent on the others, then this matrix A is invertible. From that we derive the probability density. From the probability density we derive this. OK.

The only thing you need to get this is the probability density. And this says that anytime you have a random vector Z with this probability density that we just wrote down. Then you have the property that N prime is equal to BZ, and Z is equal to B to the minus 1 N prime. Which says that if you have this probability density, then you have jointly Gaussian random variables. So we have an alternate definition of jointly Gaussian. Random variables are jointly Gaussian if they have this probability density. You can somehow represent them as linear combinations of normal random variables. OK.

Then there's something even simpler. It says that if all linear combinations of a

random vector Z are Gaussian, then Z is jointly Gaussian. Why is that? well If you look at this formula here, it says take any old random vector at all that has a covariance matrix. From that covariance matrix, we can solve for all of these eigenvectors. If I find the appropriate random variables here from this transformation, those random variables are then uncorrelated from each other, they are all statistically independent of each other, And it follows from that, that if all these linear combinations are Gaussian then Z has to be jointly Gaussian. OK.

So we now have three definitions. This one is the simplest one to state. It's the hardest one to work with. That's life you know. This one is probably the most straightforward, because you don't have to know anything to understand this definition. This original definition is physically the most appealing because it shows why noise vectors actually do have this property. OK.

So here's a summary of all of this. It says if Kz is singular, you want to remove the linearly independent random variables. You just take them away because they're uniquely specified in terms of the other variables. And then you take the resulting non-singular matrix, and Z is going to be jointly Gaussian if and only if Z is equal to AN for some normal random variable N. If Z has jointly Gaussian density or if all linear combinations of Z are Gaussian all of this for 0 mean would mean it applies to fluctuation. OK.

So why do we have to know all of that about jointly Gaussian random variables? Well because everything about Gaussian processes depends on jointly Gaussian random variables. You can't do anything with Gaussian processes without being able to look at these jointly Gaussian random variables. And the reason is that we said that Z of t is a Gaussian process if for every k and every set of time instance, every set of e, that's t1 to t sub k. If Z of t1 up to Z of tk is a jointly Gaussian random vector. OK. So that directly links the definition of a Gaussian process to Gaussian random vectors. OK.

Supposed the sample functions of Z of t or L2 with probability one. I want to say a little bit about this because otherwise you can't sort out any of these things about

how L2 theory connects with random processes. OK. So I'm going to start out by just assuming that all these sample functions are going to be L2 functions with probability 1. One way to ensure this is to look only at processes of the form Z of t equals some sum of Z sub i times ti of t.

OK remember at the beginning we started looking at this process the sum of a set of normalized Gaussian random variables times sinc functions times displaced sinc functions. You can also do the same thing with Fourier coefficients or anything. You got a fairly general set of processes that way. unfortunately they don't quite work, because if you look at the sinc functions and you look at noise, which is independent and identically distributed in time, then the sample functions are going to have infinite energy. I mean that kind of process just runs on forever. It runs on with finite power forever. And therefore it has infinite energy. And therefore the simplest process to look at, the sample functions are non-L2 with probability one, which is sort of unfortunate. So we say OK we don't care about that, because if you want to look at that process a sum of Z sub i times sinc functions, what do you care about? I mean you only care about the terms in that expansion, which run from the big bang until the next big bang. OK. We certainly don't care about it before that or after that. And if we look within those finite time limits, then all these functions are going to be L2. Because they just last for a finite amount of time.

So all we need to do is to truncate these things somehow. And we're going to diddle around a little bit with a question of how to truncate these series. But for the time being we just say we can do that. And we will do it.

So we can look at any processes the form sum of Zi times phi i of t, where the Zi are independent and the phi i of t are orthonormal. And to make things L2, we're going to assume that the sum over i of Zi squared bar is less than infinity. OK. In other words we only take a finite number of these things, or if we want to take infinite a number of them, the variances are going to go off to zero.

And I don't know whether you're proving it in the homework this time or you will prove it in the homework next time, I forget, but you're going to look at the question

of why this finite variance condition makes these sample functions be L2 with probability one.

OK. So if you had this condition, then all your sample functions are going to be L2. I'm going to get off all of this L2 business relatively shortly. I want to do a little bit of it to start with. Because if any of you have start doing any research in this area, at some point you're going to be merrily working away calculating all sorts of things. And suddenly you're going to find that none of it exists, because of these problems of infinite energy. And you're going to get very puzzled. So one of the things I tried to do in the notes is to write them in way that you can understand them at a first reading without worrying about any of this. And then when you go back for a second reading, you can pick up all the mathematics that you need. So that in fact you won't have the problem of suddenly finding out that three-quarters of your thesis has to be thrown away, because you've been dealing with things that don't make any sense. OK.

So, we're going to define linear functionals in the following way. We're going to first look at the sample functions of this random process Z. OK. Now we talked about this last time. If you have a random process Z than really what you have is a set of functions defined on some samples space. So the quantities you're interested in is what is the value of the random process at time t for sample point omega. OK.

If we look at that and make it for a given omega, this thing becomes a function of t. In fact for a given omega, this is just what we've been calling a sample element of the random process. So if we take this sample element, look at the inner product of that with some function g of t. In other words we look at the integral of Z of t omega times g of t, dt. And if all these sample functions are L2 and if g of t is L2, what happens when you take the integral of an L2 function times an L2 function, which is the inner product of something L2 with something L2 which says something with finite energy inner product with something with finite energy. Well the Schwarz inequality tells you that if this has finite energy and this has finite energy, the inner product exists. That's the reason why we went through the Schwarz inequality. It's the main reason for doing that.

So these things have finite value. So V of omega the results of doing this namely V as a function of the sample space is a real number. And it's a real number for the sample points of omega with probability one, which means we can talk about V as a random variable. OK. And now V is a random variable which is defined in this way. And from now on we will call these things linear functionals which are in fact the integral of a random process times a function. And we can take that kind of integral. It sort of looks like the linear combinations of things we were doing before when we were talking about matrices and random vectors. OK.

If we restrict the random process to have the following form, where these are independent and these are orthonormal, then one of these linear functionals is given by the random variable V is going to be the integral of Z of t times g of t, but Z of t is this. And at this point we're not going to fuss about interchanging integrals with summations. You have the machinery to do it, because we're now dealing with an L2 space. We're not going to fuss about it. And I advise you not to fuss about it. So we have a sum of these random variables here times these integrals here. These integrals here are just projections of g of t on this space of orthonormal functions. So whatever space of orthonormal functions gives you your jollies, use it talk about the inner products on that space. This gives you a nice inner products space of sequences of numbers. And then if the z i are jointly Gaussian, then V is going to be Gaussian.

And then to generalize this one little bit further, if you take a whole bunch of L2 functions, g1 of t g2 of t and so forth, you can talk about a whole bunch of random variables V1 up to V sub j, 0. And V sub j is going to be the integral of Z of t times gj of tdt. Remember this thing looks very simple. It looks like the -- like the convolutions you've been doing all your life. It's not. It's really a rather peculiar quantity. This in fact is what we call a linear functional, and is the integral of a random process times this. Which we have defined in terms of the sample functions of the random process. And now we said OK now that we understand what it is, we will just write this all the time. But I just caution you not to let familiarity breed contempt. Because this is a rather peculiar notion. And a rather powerful notion.

OK.

So these things are jointly Gaussian. We want to take the expected value of V sub i times V sub j, and now we're going to do this without worrying about being careful at all. We have the expected value of the integral of Z of t, gi of t, td times the integral of Z tau gj of tau, d tau. And now we're going to slide this expected value inside of both of these integrals. And not worry about it. And therefore what we're going to have is a double integral of gi of t expected value of z of t time z of tau times gj of tau dt, d tau, which is this thing here. Which you should compare with what we've been dealing with most of the lecture today. This is the same kind of form for a covariance function as we've been dealing with for covariance matrices.

It has very similar effects. I mean before you were just talking about finite dimensional matrices, which is all simple mathematically in eigenfunctions and eigenvalues. You have eigenfunctions and eigenvalues of these things also. And so long as these are defined nicely by these L2 properties we've been talking about. In fact you can deal with these in virtually the same way that you can deal with the matrices we were dealing with before. If you just remember what the results are from matrices, you can guess what they are for these covariance functions. OK. But anyway you can find the expected value of Vi times Vj by this formula. Again we're dealing with zero-mean and therefore we don't have to worry about the mean, put that in later. And that all exists. OK.

So the next thing we want to deal with, hitting you with a lot today. But I mean the trouble is a lot of this is half familiar to most of you. People who have taken various communication courses at various places have all been exposed to random processes in some highly trivialized sense. But the major results are the same as the results we're going through here. And all we're doing here is adding a little bit of carefulness about what works and what doesn't work. Incidentally in the notes which is towards the end of lectures 14 and 15, we give three examples which let you know why in fact we want to look primarily at random processes which are defined in terms of a sum of independent Gaussian random variables time orthonormal functions.

And if you look at those three examples, some of them have problems. Because of the fact that everything you're dealing with has infinite energy. And therefore it doesn't really make any sense. And one of them I should talk about this just a little bit in class. And I think I still have a couple of minutes, is a very strange process where Z of t is IID. In fact just let it be normal. And independent for all t. OK. In other words you generate a random process by looking at an uncountably infinite collection of normal random variables. How do you deal with such a process? I don't know how to deal with it. I mean it sounds like it's simple. If I put this on a quiz, three-quarters of you would say oh that's very simple. What we're dealing with is a family of impulse functions. Spaced arbitrarily closely together. This is not impulse function. Impulse functions are even worse than this, but this is bad enough.

When we start talking about spectral density, we can explain this a little bit better by thinking this kind of process it doesn't make any sense. But this kind of process, if you look at its spectral density, it's going to have a spectral density which is zero everywhere, but whose integral over all frequencies is one. OK. In other words it's not something you want to wish on your on your worse friend. It makes a certain amount of sense as a limit of things. You can look at a very broadband process where in fact you spread the process out enormously. You can make pseudo noise which looks sort of like this. And you make the process broader and broader and lower and lower intensity everywhere. But it still has this energy of one. It still has a power of one everywhere. And it just is ugly. OK. Now if you never worried about these questions of L2, you would look at a process like that and say, gee there must be some easy way to handle that because it's probably the easiest process you can define. I mean everything is normal. If you look at any set at different times, you get a set of IID normal Gaussian variables. You try to put it together, and it doesn't mean anything. If you pass it through a filter, the filter is going to cancel it all out.

So anyway, that's one reason why we want to look at this restricted class of random processes we're looking at. OK.

What we're interested in now is we want to take a Gaussian random process really, but you can take any random process. We want to pass it through a filter, and we

want to look at the random process that comes out. OK. And that certainly is a very physical kind of operation. I mean any kind of communication system that you build is going to have noise on the channel. And one of the first things you're going to do is you're going to filter what you've received. So you have to have some way of dealing with this. OK. And the way we sort of been dealing with it all along in terms of the transmitted way forms we've been dealing with is to say OK. What we're going to do is to first look what happens when we take sample functions of this, pass them through the filter, and then what comes out is going to be some function again. And then we're back into what you studied as an undergraduate talking about functions through filters.

We're going to jazz it up a little bit by saying these functions are going to be L2 the filters are going to be L2. So that in fact you know you'd get something out that make sense. So what we're doing is looking at these sample functions V the output at time tau for sample point omega is going to be the convolution of the input at time t and sample point omega. Remember this one sample point exists for all time. That's why these sample points and sample spaces are so damn complicated. Because they have everything in them. OK. So there's one sample point which exists for all of them.

This is a sample function. You're passing the sample function through a filter, which is just normal convolution. What comes out then. If in fact we express this random process in terms of this orthonormal sum the way we've been doing before. Is you get the sum over j of this, which is a sample value of the j'th random variable coming out times the integral of pj of t, h of tau minus t, d tau. OK. For each tau that you look at, this is just a sample value of a linear functional. OK. If I want to look at this at one value of tau, I have this integral here which is a random process. A sample value of a random process at omega time a function. This is just a function of t for given tau. OK.

So this is a linear functional. As a type we've been talking about. And that linear functional is then given by this for each tau. This is a sample value of a linear functional we can talk about. OK. These things are then, if you look over the whole

sample space omega, V of tau becomes a random variable. OK. V of tau is the random variable whose sample values are V of t omega, and they're given by this. So if z of t is Gaussian process, you get jointly Gaussian linear functionals at each of any set of times tau 1, tau 2 up to tau sub k. So this just gives you a whole set of linear functionals. And if z of t is a Gaussian process, then all these linear functionals are going to be jointly Gaussian also. And bingo. What we have then is an alternate way to generate Gaussian processes. OK. In other words you can generate a Gaussian process by specifying at each set of k times, you have jointly Gaussian random variables. But once you do that, once you understand one Gaussian process, you're off and running. Because then you can pass it through any L2 filter that you want to. And you generate another Gaussian process.

So for example, if you start out with this sinc type process, I mean we'll see that that has a spectral density, which is flat over all frequencies. And we'll talk about spectral density tomorrow. But then you pass it through a linear filter, and you can give it any spectral density that you want. So at this point, we really have enough to talk about arbitrary covariance functions just by starting out with random processes, which are defined in terms of some sequence of orthonormal random variables. OK.

Now we can get to covariance function of a filter process in the same way as we got the matrix for linear functionals. OK. And this is just computation. OK. So what is it? The covariance function of this output process V evaluated at one time r another time s. One of the nasty things about notation when you start dealing with a covariance function of the input and the output to a linear filter is you suddenly need to worry about two times at the input and two other times at the output. OK. Because this thing is then expected value of V sub r times the expected value of V subs. OK. This is a random variable, which is the process V of r evaluated at time r. This is the random variable, which is the process V of t evaluated at a particular time s. This is going to be the expected value of what this random variable now is the integral of the random process z of t times this function here, which is now a function of t. Because we're looking at a fixed value of r.

So this is a linear functional, which is a random variable. This is another linear

functional. This is evaluated at some time s, which is the output of the filter at time s. Then we will throw caution to the wind interchange integrals with expectation and everything. And then in the middle we'll have expected value of z of t times z of tau, which is the covariance function of z. OK. So the covariance function of z then specifies what the covariance function of the output is. OK. So whenever you pass a random process through a filter if you know what the covariance function of input to the filter is, you can find the covariance function of the output of the filter.

That's a kind of a nasty formula, it's not very nice. But anyway the thing that it tells you is that whether this random process is Gaussian or not. The only thing that determines the covariance function of the output of the filter is the covariance function of the input to the filter plus of course the filter response, which is needed OK. And this is just the same kind of bilinear form that we were dealing with before. Next time we will talk a little bit about the fact that when you're dealing with a bilinear form like this, you can take these covariance functions and they have the same kind of eigenvalues and eigenvectors that we had before for a matrix. Namely this is again going to be positive definite as a function, and we will be able to find its eigenvectors and its eigenvalues. We can't calculate them. Computers can calculate them. People who've spend their lives doing this can calculate them. I wouldn't suggest that you spend your life doing this. Because again you would be setting yourself up as a second class computer, and you don't make any profit out of that.

But anyway, we can find this in principle from this. OK. One of the things that we haven't talked about at all yet, and which we will start talking about next time in which the next set of lecture notes, lecture 16, we'll deal with is the question of stationarity. Let me say just a little bit about that to get into it. And then we'll talk a lot more about it next time. The notes will probably be on the web some time tomorrow. I hope before noon if you want to look at them. Physically if you look at a stochastic process, and you want to model it. I mean suppose you want to model a noise process. How do you model a noise process? Well you look at it over a long period of time. You start taking statistics about it over a long period of time. And somehow you want to model it in such a way, I mean the only thing you can look at is statistics

over a long period of time. So if you're only looking at one process, you can look at it for a year and then you can model it, and then you can use that model for the next 10 years. And what that's assuming is that the noise process looks the same way this year as it does next year. You can go further than that and say, OK I'm going to manufacture cell phones or some other kind of widget. And what I'm interested in then is what these noise wave forms are to the whole collection of my widgets. Namely different people will buy my widgets. They will use them in different places. So I'm interested in modeling the noise over this whole set of widgets. But still if you're doing that you're still almost forced to deal with models which have the same statistics over at least a broad range of times. Sometimes when we're dealing with wireless communication we say, no the channel keeps changing in time. and the channel keeps changing slowly in time. And therefore you don't have the same statistics now as you have then. If you want to understands that, believe me, the only way you're going to understand it is to first understand how to deal with statistics for the channel which stay the same forever. And once you understand those statistics, you will then be in a position to start to understand what happens when these statistics change slowly. OK.

In other words, what our modeling assumption is in this course, and I believe it's the right modeling assumption for all engineers, is that you never start with some physical phenomena and say, I want to test the hell out of this until I find an appropriate statistical model for it. You do that only after you know enough about random processes. That you know how to deal with an enormous variety of different home cooked random processes. OK. So what we do in a course like this is we deal with lots of different home cooked random processes, which is in fact why we've done rather peculiar things like saying, let's look at a random process which comes from a sum of independent random variables multiplied by orthonormal functions. And you see what we've accomplished by that already. Namely by starting out that way we've been able to define Gaussian processes.

We've been able to define what happens when one of those Gaussian processes goes through a filter. And in fact, that gives a way of generating a lot more random processes. And next time what we're going to do is not to say how is it that we know

that processes are stationary. How do you test whether processes are stationary or not. But we're just going to assume that they're stationary. In other words if they had the same statistics now as they're going to have next year. And those statistics stay the same forever. And then see what we can say about it.

To give you a clue as to how to start looking at this, remember what we said quite a long time ago about Markov chains. OK. And now when you look at Markov chains, remember that what happens at one time is statistically a function of what happened at the unit of time before it. OK. But we can still model those Markov change as being stationary. Because the dependents at this time on the previous sample of time is the same now as it will be five years from now. OK. In other words you can't just look at the process at one instant of time and say this is independent of all other times. That's not what stationary means. What stationary means is the way the process depends on the past at time t is the same is the way it depends on the past at some later time tau. And that in fact is the way we're going to define stationarity. It's at these joint sample times that we're going to be looking at. Which I had a better word for that. Join sets of epochs that we'll be looking at. The joint statistics over a set of epics is going to be the same now as it will be at sometime in the future. And that's the way we're going to define stationarity.

A prelude of what we're going to find when we do that is that this covariance function, if the covariance function at time t and time tau is the same if you translate it up to t plus t1 and tau plus t1, then in fact this function is going to be a function only if the difference t minus tau. It's going to be a function of one variable instead of two variables. OK. So as soon as we get this function being a function of one variable instead of two variables, first thing we're going to do is to take the Fourier transform of this. Because then we'll be taking the Fourier transform of a function of a single variable. We're going to call that the spectral density of the process. And we're going to find that for stationary processes the spectrum densities tell you everything if they're Gaussian.

Why is that? Well the inverse Fourier transform is this covariance function. And we've now seen that the covariance function tells you everything about a Gaussian

process. So if you know the spectral density for a stationary process, it will tell you everything.

We will also have to fiddle around a little bit about how we define stationarity. But at the same time don't have this infinite energy problem. And the way we're going to do it is the way we've done it all along. We're going to take something that looked stationary, we're going to truncate it over some long period of time, and we're going to have our cake and eat it too that way. OK. So we'll do that next time.