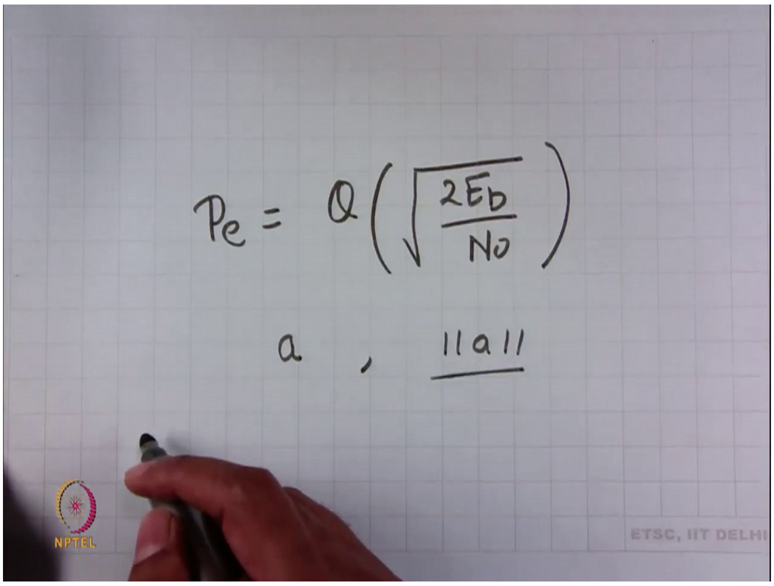
Principles of Digital communication
Prof. Abhishek Dixit
Department of Electrical Engineering
Indian Institute of Technology, Delhi

Lecture – 32
Detection
Theorem of Irrelevance & Waveform Detection

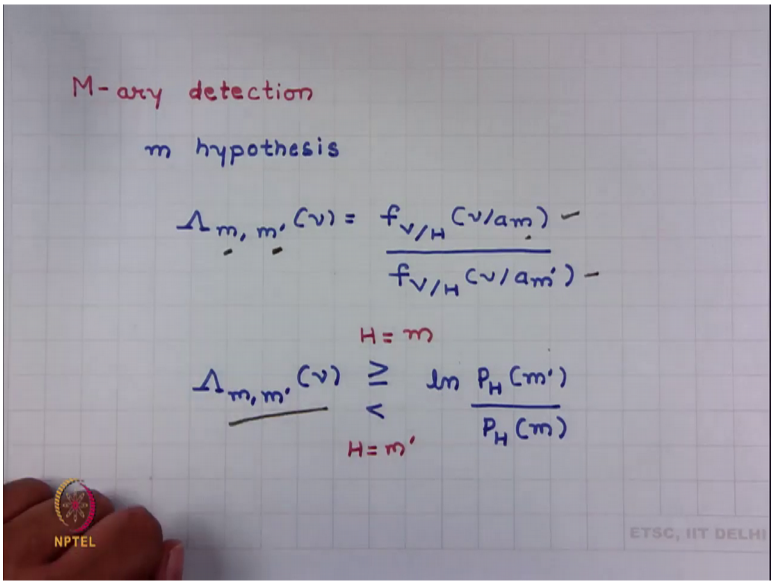
Good morning. Welcome to a new lecture in Detection. In this lecture we will basically talk about Theorem of Irrelevance and Waveform Detection. In last lecture what we did is, we took a simple case of binary antipodal PAM and we derived the bit error rate for binary antipodal PAM.

(Refer Slide Time: 00:45)



And we have seen that for binary antipodal PAM, the bit error rate is given by Q of square root of 2 E b by N naught where E b is debate energy and n naught denotes the one sided power spectral density of noise and then we started looking into vector detection. And we have said that for the case of vectors, everything remains same except that. Now instead of a, where a denotes the numerical value of the signal in binary antipodal PAM we have to work with norm of a; that is it and there was other differences which we have looked into.

(Refer Slide Time: 01:40)

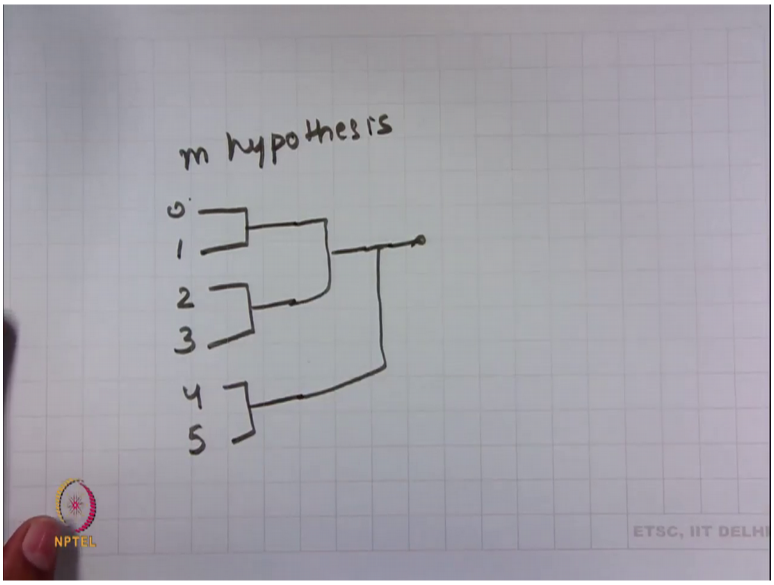


Today we will start with first M-ary detection and in M-ary case we have m hypothesis, so m possibilities. And we can define the likelihood ratio in the same way as we have done previously. The only difference is, because there are m hypothesis now we need to have more number of likelihood ratios.

So, you can compare for example, m with m dash and let us say for the hypothesis m, if we transmit a signal a m and for hypothesis m dash we transmit a signal a m dash and so this likelihood ratio of m and m dash can be given by this. This was the ratio of likelihood when a m is transmitted divided by the likelihood when a m dash is transmitted. So, we get this likelihood ratio. And again if you have to decide on the hypothesis when comparing m and m dash, you need to see that if this likelihood ratio is greater than this threshold then we select the hypothesis as m; otherwise we select the hypothesis as m dash.

So, what we do in this a M-ary case, is we take two hypothesis at a time. We compare these two hypotheses and then we decide on one hypothesis which should be considered one later evaluations.

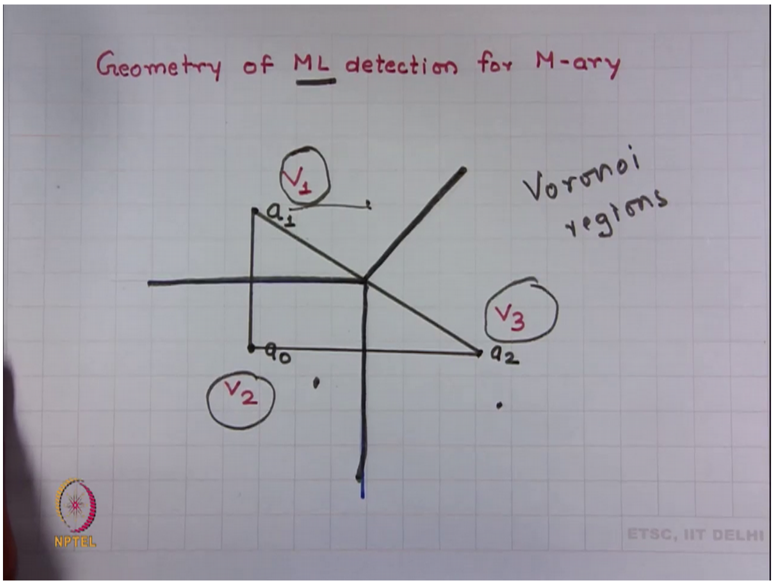
(Refer Slide Time: 03:04)



For example, if we have m hypothesis and we can consider 2 hypotheses at a time. For example, let us consider the case of 6 hypotheses, so we have to compare this with this, we decide on a winner, we compare this with this, we decide on a winner, we compare this with this, we decide on a winner and then we have to compare this with this decide on a winner and finally, you can compare this winner with this and finally, you can get the right hypothesis ok.

So, you can do decoding in this way. So, this is only difference in the case of M-ary, everything else remain same, likelihood ratios remain same, the likelihoods remain same, everything remains same, its just now you have to have more number of comparisons.

(Refer Slide Time: 04:05)



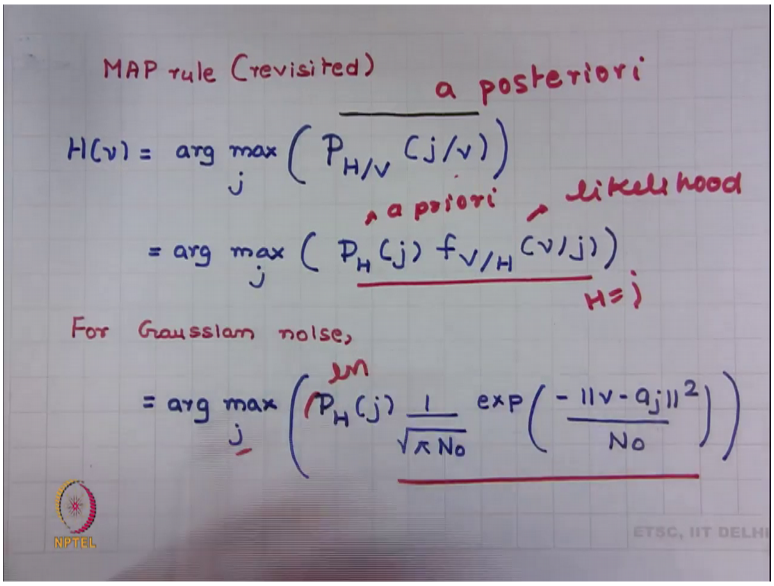
What is the geometry in the case of M-ary. So, in the case of M-ary as before we are focusing on the geometry in the case of maximum likelihood detector. So, in maximum likelihood we assume that the symbols are equiprobable. So, if we have this M L detector for M-ary you can find the Voronoi regions. So, Voronoi regions are the regions which are formed by drawing a perpendicular bisector between the 2 signals.

For example, if I consider signal a 1 and signal a naught, these two points, I draw a perpendicular bisector between this signals. So, I get this perpendicular bisector, I can consider signal a naught and a 2, I get this perpendicular bisector. I can consider signal a 1 and a 2 and I get this perpendicular bisector ok.

So, I get now 3 Voronoi regions V 1 V 2 and V 3. So, Voronoi regions are the regions which are formed by drawing perpendicular bisectors between signals. And now the detection is simple, if the received symbol falls in this Voronoi region then I go with the hypothesis 1. If received signal falls in this Voronoi region then I go with hypothesis 0. If received signal falls in this Voronoi region then I go with the hypothesis 2.

So, in terms of geometry its pretty simple, you have various signal points, you draw perpendicular bisectors between the signal points and you get some Voronoi regions. And whenever a received signal lies in a Voronoi region you assume that this signal is corresponding to the hypothesis in, whose Voronoi region that signal is line. And again you can see here as well, this M L detector is simply calculating the distance of the received signal from all possible signals and then it is deciding based on the minimum distance and this will become clear mathematically.

(Refer Slide Time: 06:36)

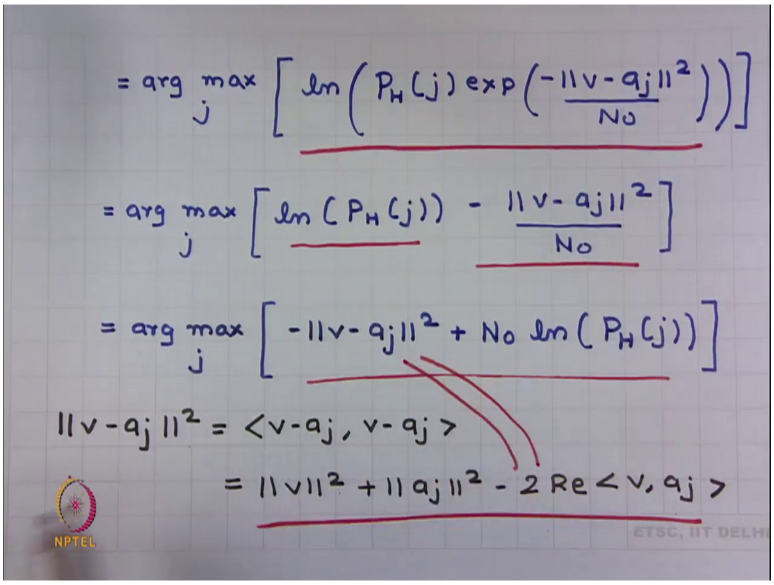


So, now I am really looking at this map detector again and by real looking at it I would derive some other ways in which you can think about this MAP rule. So, what was MAP rule , let us look at this. So, MAP says that I should choose the hypothesis as j and this hypothesis should maximize my a posteriori probability. So, this is a posterior probability. So, I should choose a hypothesis as the hypothesis which maximizes this a posterior probability. And we have also seen in the last lecture that this a posteriori probability maximization is same as maximizing this product. So, this is a priori probability and this is likelihood.

So, when we want to maximize this a posteriori probability, it is same thing as trying to maximize the product of a priori probability with likelihood. And for Gaussian noise, we also know what is this likelihood. So, I can simply substitute the probability density function of this random variable V given a hypothesis H is transmitted and let us assume that when hypothesis H is j, then the symbol value is a j

So, now we know that this is the probability density function of random variable V, given that hypothesis H j. And what MAP detector would do, is it will try to find out j for which this quantity is maximum. And to simplify stuff I just take l n here, natural log N here. So, when I try to maximize this quantity it is same as trying to maximize this.

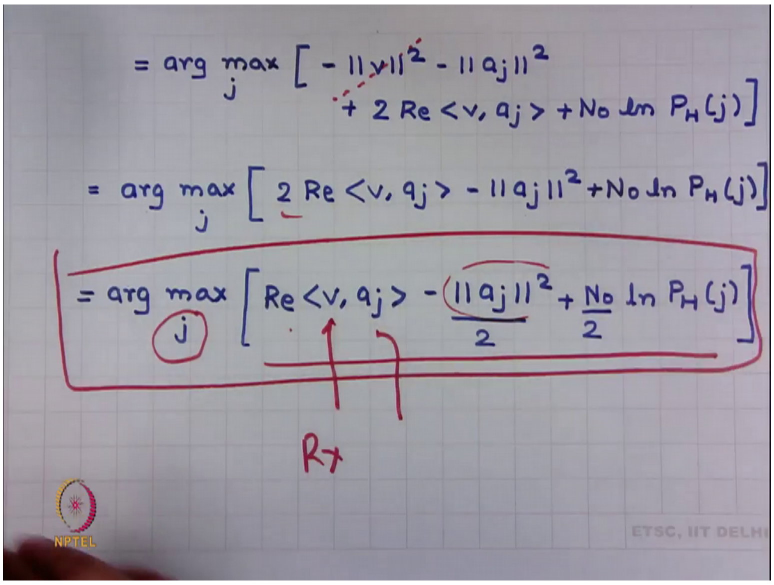
(Refer Slide Time: 08:47)



So, I am considering now this quantity with l n here. Trying to maximize this thing is same as trying to maximize this thing ok, its having a natural log in here just simplifies this little bit. So, I know that log of a times b is simply log of a plus log of p. So, from this I get this. And then what I do is, I shift this N naught to this side. So, from here I can simply get this. So, what is map detector doing, is it trying to find out j for which this quantity is maximum.

And then I use the property that norm of v minus a j square can be simply calculated by finding the inner product of v minus a j with v minus a j. And we have solved this identity before and this is a good idea that you now memorize this that is norm square of v minus a j is simply norm square of v plus norm square of a j minus 2 times real part of inner product of v with a j.

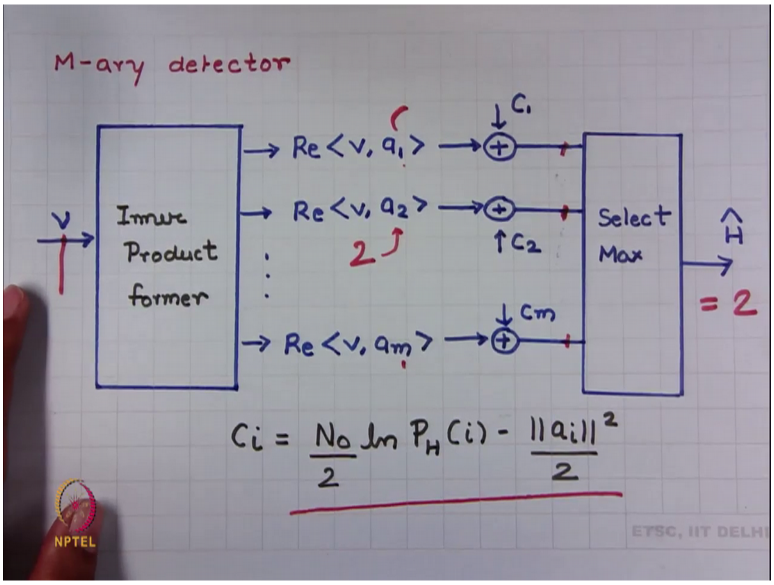
(Refer Slide Time: 10:09)



So, now substituting this thing in place of this, what I get is, that map detector should find the j for which this quantity is maximum. And as you can see that this is independent of hypothesis j, so, we can get rid of this, its not a function of j. So, I do not want to think about this, so what I have remained with is this. Now shifting two to these products I get this. So, now, what I get is that a map detector should find the hypothesis j for which this quantity is maximum and this is how you should design an M-ary detector.

So, first let us see what products we have. First we are having real part of inner product of v with a j, v is the received vector, a j is one of the possible transmitted vector. This corresponds to the energy of a j by 2. And this terms deals with the probabilities of hypothesis; namely a priori probabilities of hypothesis. So, if you have to design an M-ary detector, you have to look at this equation. And based on this equation you can design an M-ary detector. For example like this.

(Refer Slide Time: 11:50)



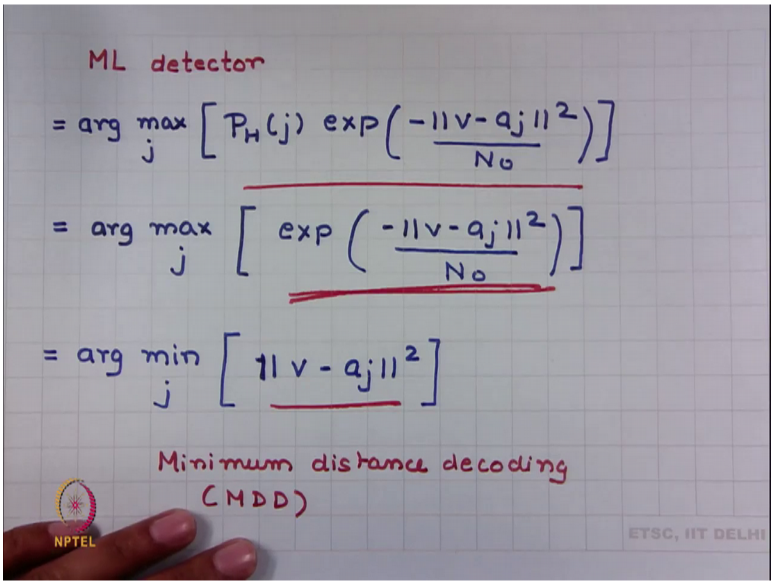
So, what M-ary detector would be doing, is its taking a received vector, it will form the inner product of this received vector with signal set. So, it would take the inner product of v with a 1, where a 1 let us say corresponds to the signal corresponding to hypothesis 1 and so on and so forth. And then it takes the real part of this thing.

So, when it does it, it takes care of this term and to this it has to subtract or add this quantity. So, it adds C 1 where C i in general is this. And based on the outcome at this point it can decide for the possible transmitted hypothesis. A possible transmitted hypothesis is the one which has the maximum resultant here.

For example, if for the second hypothesis I am transmitting a signal a 2 and this branch has the highest outcome, then my predicted value of hypothesis will be 2. So, here I am assuming that I am having m hypothesis, where this corresponds to hypothesis 1, this corresponds to hypothesis 2, this corresponds to hypothesis m. And so m at a detector is a simple detector.

And this actually we have discussed way back in lecture 4, where we have said that detector is taking decision based on inner product of the received vector with the signal set and it adds some bias to this inner product and it takes the decision. And at the time I pointed out that we will figure out what is this bias when we have covered detection. It will be a very good idea that if you have this picture in mind; if you have this picture in mind then making detectors for digital communication system is trivial all right.

(Refer Slide Time: 14:11)



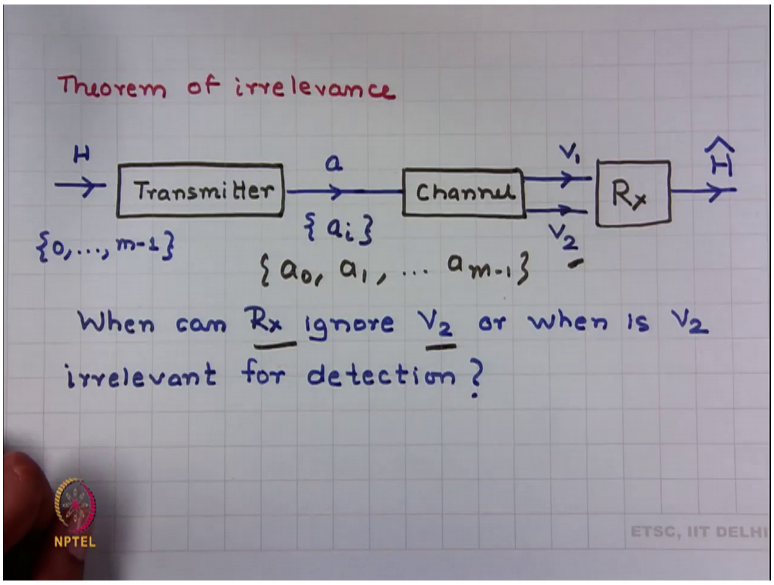
So, let us now move on to the case of M L detector. In the case of M L detector, I will be trying to maximize this quantity, but in case of M L detector these a priori probabilities are same for all hypotheses and thus we do not have to worry about this. M L detector simply tries to find out j for which this quantity is maximum. And trying to find out j for which this quantity is maximum is same thing as trying to find out j for which this quantity is minimum, because when this quantity is minimum, this quantity is going to be maximum.

And what is this? This corresponds to the square of distance of the received vector with the symbol corresponding to the hypothesis. So, actually M L detector is simply trying to find out or to select the hypothesis which minimizes the distance from the received vector.

So, M L detector is also known as minimum distance decoder, because what is trying to do, is to find out the hypothesis for which the distance or the square of the distance is minimum. If the square of the distance is minimum the distance will also be minimum, and hence M L detector is also known as minimum distance decoder. We are done with vector detection.

So, in case of vector detection we have seen how to handle this complex vectors reception for the case of binary antipodal vectors. And also we have looked into the case of M-ary detection of vectors. And we have looked into the design of a map detector and we have looked into the ideas behind M L detector and we have seen that M L detector is nothing, but it is simply a minimum distance decoder. Now, we have to go to this waveform detection and to go to waveform detection we have to look into what is known as theorem of irrelevance.

(Refer Slide Time: 16:29)

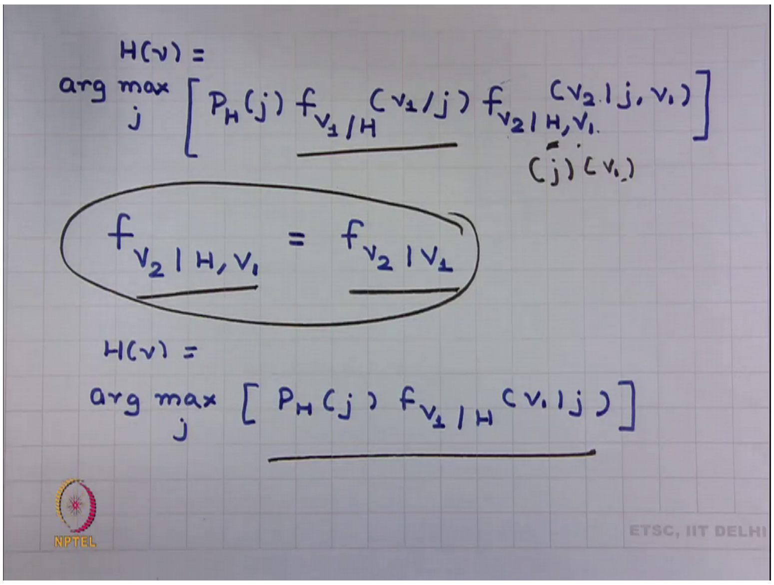


So, this will be useful theorem, if we try to understand how does the detector takes care of waveform. So, let us start, let us assume that I am having m hypothesis. Sometimes I choose hypothesis to go from 0 to m minus 1 and sometimes I choose the hypothesis to go from 1 to m, but it does not matter conceptually.

What it simply matters is that I have m hypothesis and corresponding to each hypothesis I am transmitting a signal, where signal belongs to a set; like for example, here to this set. So, I am having m quantized numbers. And whenever I try to transmit a hypothesis, I may transmit one of the numbers or one of the vectors, it does not matter dealing with numbers or dealing with vectors is not at all different.

Then I pass this through a channel and channel offers to this receiver two random variables; V 1 and V 2. And based on looking at these two random variables, receiver has to predict the hypotheses. And the question that we are asking is, when can receiver ignore this V 2 or when is this V 2 irrelevant for detection. When receiver does not have to worry about this extra random variable V 2, when can receiver simply decides optimally based on one received random variable; that is the question we are asking.

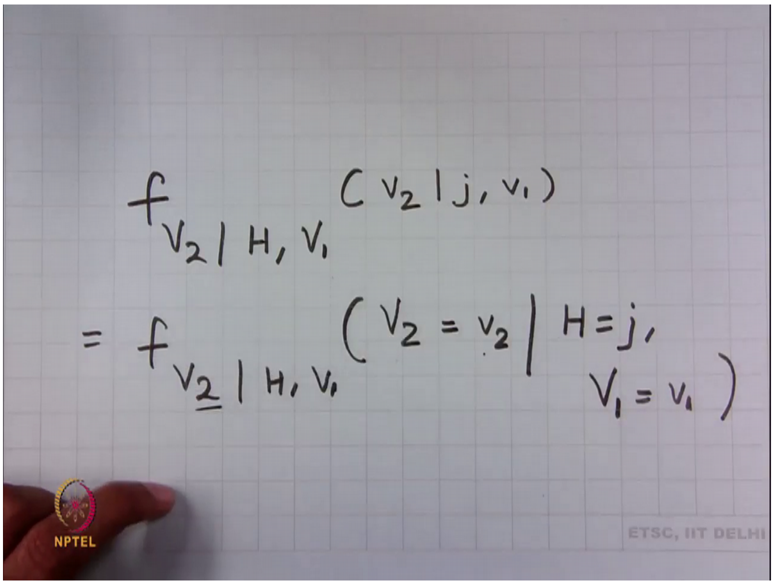
(Refer Slide Time: 18:27)



So, let us look at the map case, what would be the a posteriori probability here. So, map rule remains same; that means, you are trying to find out the hypothesis for which this a posteriori probability is maximum. And this a posteriori probability is simply given by the a priori probability of hypothesis j to be transmitted, multiplied by the conditional probability density function of random variable V 1 taking in a numerical value small v 1, given that hypothesis H s j.

So, multiply it by the conditional probability density function, that hypothesis H has taken a value j, random variable V 1 has taken a numerical value small v 1. Conditioned on this event what is the PDF that random variable V 2 takes in a numerical value small V 2. I can also write this in this way so that it is more clear.

(Refer Slide Time: 19:35)

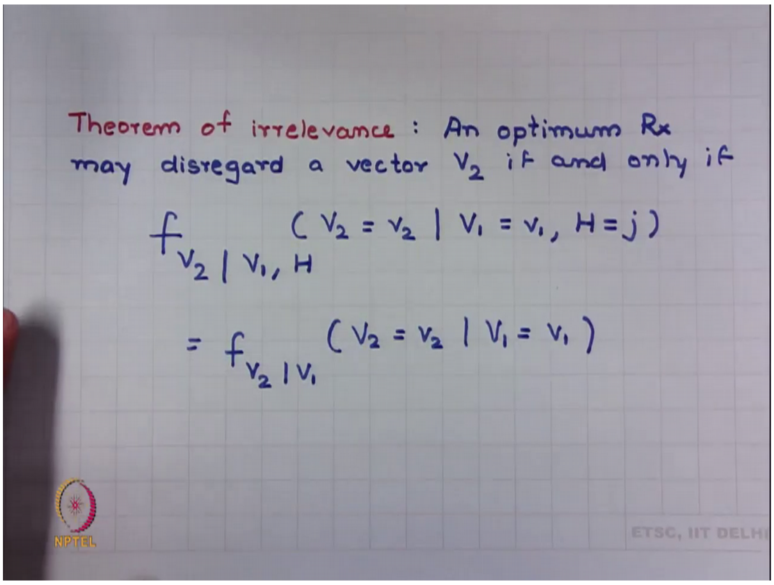


So, this can be expanded like this. So, I am saying what is the PDF for which random variable V 2 take seed name numerical value V 2, given that hypothesis is j and random variable V 1 has taken in a numerical value small v 1.

So, when we have discussed about this random variables and random processes, we have said that we usually denote random variables with capital letters and numerical values of those random variables with small letters. So, this is the conditional PDF of V 2 given H and V 1, all right. So, if this conditional probability density function of V 2 given H and V 1 does not depend upon the hypothesis; that means, this conditional PDF is same as the conditional PDF of V 2 given V 1. Then what you can see is, this term will become independent of hypothesis and thus the choice of j does not influence this term. So, this term could be ignored.

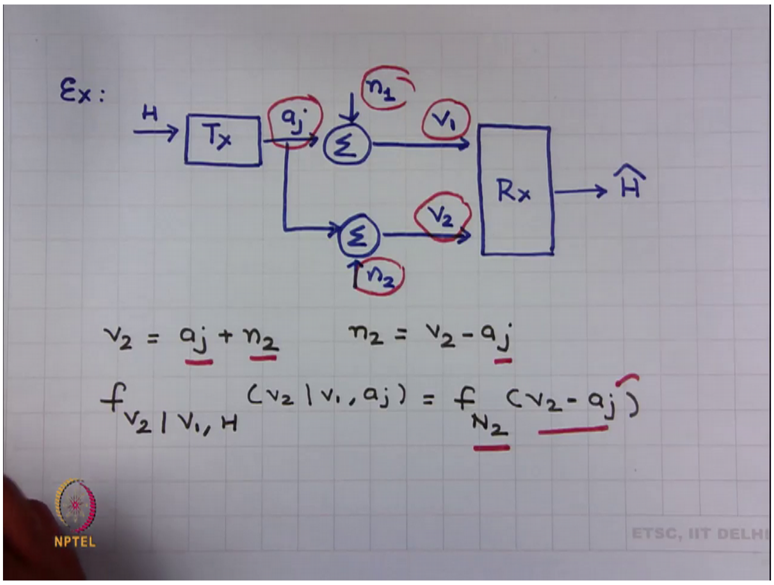
Then in that case map rule simply reduces through this rule. So, map detector does not have to worry about this term, because this term will become independent of hypothesis, if the conditional PDF of V 2 given H and V 1 is simply conditional PDF of V 2 given V 1, then this term will be independent of hypothesis and the map detector does not have to worry about this, because its constant with respect to j, its constant with respect to hypothesis does the map rule simply reduces to this. And thus if this condition is satisfied then V 2 becomes irrelevant for detector and that is what theorem of irrelevance says.

(Refer Slide Time: 21:53)



So, an optimum detector or receiver may disregard a vector V 2, if and only if the conditional PDF of V 2 given V 1 and H is same as the conditional PDF of V 2 given V 1; that means, this was independent of H and then the receiver may disregard this vector V 2 and this is theorem of irrelevance. This theorem of irrelevance will help us to prove things precisely and that is why we are using it. So, to understand this theorem of irrelevance, let us look at 3 examples.

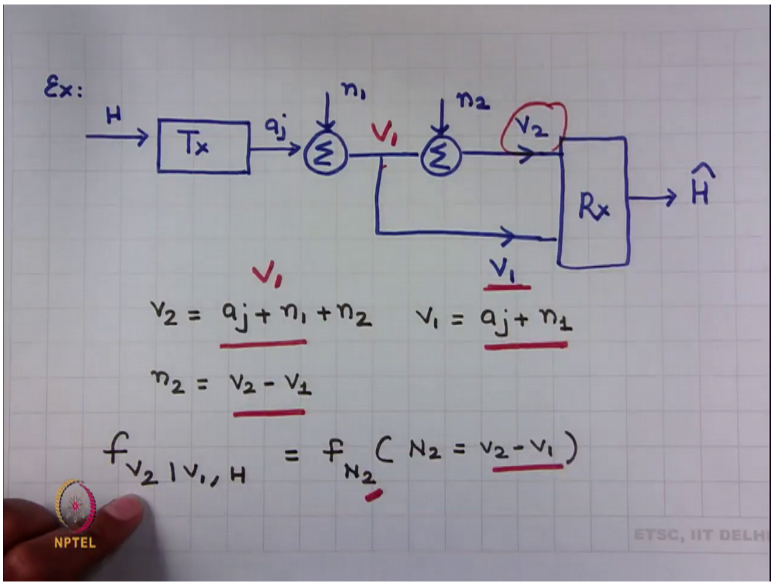
(Refer Slide Time: 22:32)



So, in this example I am transmitting hypothesis, this is the numerical value corresponding to this hypothesis, then noise sample n 1 adds to this transmitted symbol and I get a random variable V 1. Here noise sample n 2 adds to this transmitted symbol and I get a random variable V 2 and the question is V 2 irrelevant for the receiver. And how do I think about this, what is this V 2? V 2 is a j plus n 2 and what is this n 2? N 2 thus is V 2 minus a j. Thus the conditional probability density of V 2 given V 1 and H, if this has been given to us is simply given by the PDF of the noise random variable, to take in a numerical value of n 2.

And in that case numerical value of n 2 that it has to take is V 2 minus H a. And hence conditional probability of V 2 given V 1 in h is a function of a j, is a function of hypothesis and hence this V 2 is not irrelevant for receiver. And its quite obvious, because this V 2 is a function of a j and is also a function of N 2. So, by looking at this V 2, your information might improve. For example, if I have a very noisy sample in here, and this sample is less noisy, then you have improved information about this transmit symbol if you look at V 2. And hence we do is not irrelevant for receiver.

(Refer Slide Time: 24:38)

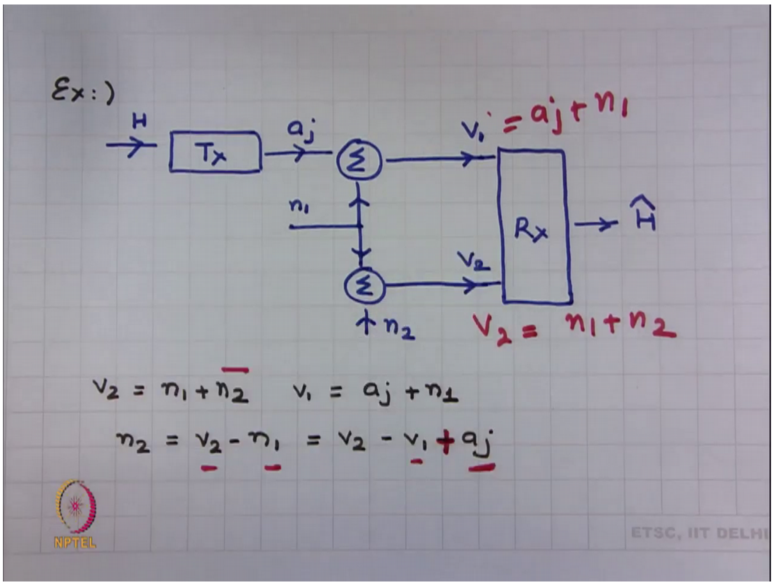


So, let us look at another case. So, here I am transmitting a j, noise sample n 1 adds. So, this leads to random variable V 1, V 2 is V 1 plus n 2. So, V 2 is V 1; V 2 is V 1, this is V 1 plus n 2. And what is V 1? V 1 is a j plus n 1. So, if we look at n 2, what is n 2? N 2 is

V 2 minus V 1. And hence conditional PDF of V 2 given V 1 an edge, is simply conditional PDF of noise random variable, taking in a numerical value V 2 minus V 1.

And this is not a function of hypothesis and hence this V 2 is irrelevant for receiver. So, this is a test that we have to do, where the channel is offering any useful information for the receiver. Our channel is offering a redundant information to the receiver

(Refer Slide Time: 25:57)



Let us look at the third case. Here I am having this V 1 as a j plus n 1. And this V 2 is n 1 plus n 2. So, V 2 is n 1 plus n 2, V 1 is a j plus n 1, what is n 2? N 2 from here is V 2 minus n 1. And what is n 1? n 1 is V 1 minus a j. So, this has to be plus anyways. So, n 2 is now a function of a j and hence this random variable is not an irrelevant random variable, this is relevant.

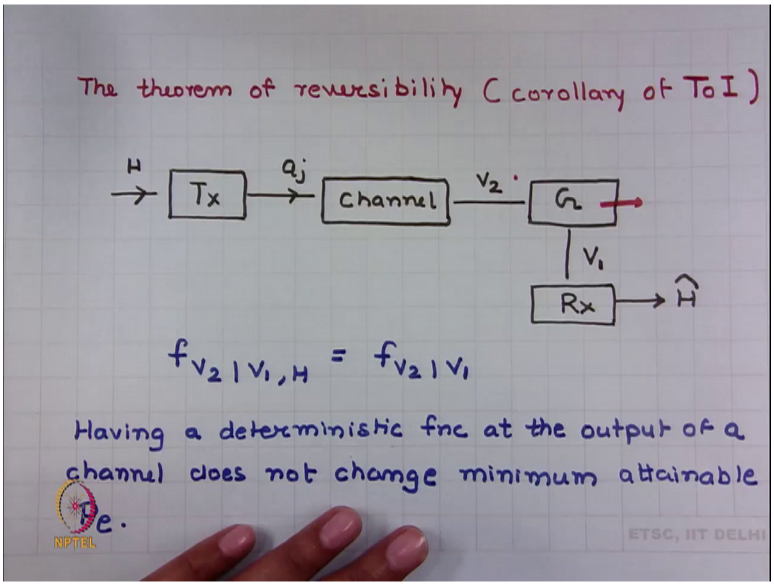
Hence what is theorem of irrelevance is saying is, if you have to find out whether a random variable is irrelevant or relevant for the detector, you have to just see whether its conditional PDF is a function of hypothesis or not. If its not a function of hypothesis then this random variable is irrelevant for the receiver. Here in this case even though it looks that this random variable is not directly a function of hypothesis, but this will help us in learning about the noise n 1. And this noise n 1 is also present here.

And thus intuitively you may feel that because by looking at this random variable I can learn about this n 1, because this V 2 is n 1 plus n 2. And this V 1 also has n 1 so, I have a better chance of trying to find out about the hypothesis. So, intuitively also you can understand about these things.

Let us look at the other examples, whether intuitively can we find about these things. So, V 2 in this case is simply V 1 plus n 2; that means, V 2 is a noisy version of V 1. If V 1 is already present and if V 2 is just noise plus V 1, then you are not helping in learning about V 1 at all and hence this is an irrelevant statistic. If I look at the first example, here this hypothesis is present in both V 1 and V 2. And as I have said it is possible that for example, here the noise amplitudes might be very small and because you are having hypotheses in both branches this is of course, not an irrelevant statistic.

So, what idea I am trying to convey is sometimes you can work this out through maths as we have done and sometimes also you can intuitively think whether random variable is relevant or irrelevant for the detector.

(Refer Slide Time: 29:05)

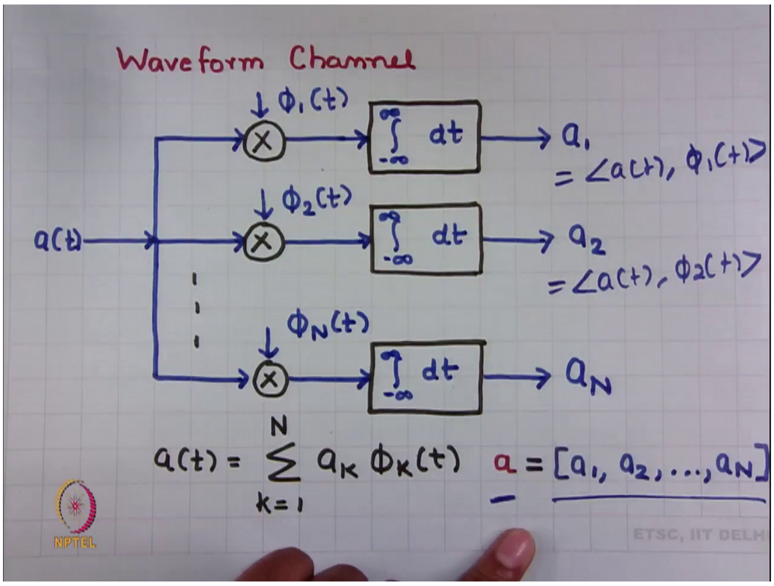


There is corollary of theorem of irrelevance and that goes by the name of theorem of reversibility. This theorem says that if at the output of a channel you are having a deterministic function. For example, this g is a deterministic function. Now the conditional PDF of V 2 given V 1 and hypothesis, is simply conditional PDF of V 2 given V 1, because this is a deterministic function. As soon as you are given this V 2, you know V 1 is not it.

And this transformation, because its a deterministic function does not at all depend upon hypothesis and hence whether you observe V 1 or you observe V 2 you get the same probability of error. Probability of error does not change by introducing any deterministic function at the output of a channel.

So, sometimes you might be given a very complicated deterministic function. And if you have to worry about what is the probability of error of this receiver, you can simply estimate the probability of error at this point and you will get the same probability of error. So, that is clear from this theorem of reversibility. Now after having understood this theorem of irrelevance, now we can safely move to this waveform channel, because we will require this theorem of irrelevance to understand this waveform channel ok.

(Refer Slide Time: 30: 42)



So, here we are having this waveform channel and this picture we have seen before as well, I am trying to revise these things. So, when we have talked about signal spaces, we have several times have shown this picture. So, we have a deterministic signal a t built using different orthonormal function. So, a t is built using different orthonormal functions or its a weighted sum of different orthonormal functions.

And though this equation looks very naive and simple, but I guarantee you that you can think about any modulation scheme, whatsoever it is, in terms of this equation. So, any modulation scheme, whether its PAM COM FSK, PPM, CDMA whatsoever it is, can be thought in terms of this basic equation. So, this equation holds for both linear modulation, non-linear modulation. Whatever it may be the kind of modulation scheme you can convert any modulation scheme to this simple basic equation, and thus this equation is very generic.

So, you can think about any modulation scheme. Any modulation scheme creates a signal as a weighted sum of orthonormal functions. In PAM there is only basic orthonormal function and all other orthonormal functions are time shifted version of the same basic orthonormal function. In com you have two basic orthonormal functions. In FSK, in M-ary FSK you have m orthonormal functions. So, whatever is the modulation scheme, all these modulation schemes can be thought in terms of this basic equation.

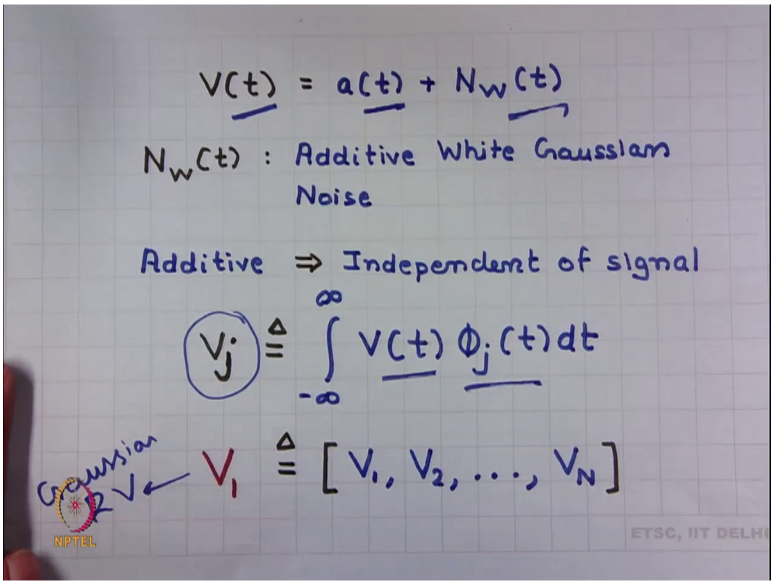
So, any waveform, you can pass through various correlators. So, we know that this is a correlator tuned to this phi 1 t. So, this correlator will extract the component of a t along this phi 1 t. So, a 1 as we have seen is nothing, but is the inner product of a t with phi 1 t. a 2 is inner product of a t with phi 2 t and so on so forth.

So, basically what you can get is corresponding to this a t you can have these coefficients; a 1 a 2 up to a N. And the collection of this element we are saying as a vector a. So, up to now we have been thinking about what happens to this vector a, we have understood clearly the vector reception; that is a. So, vector a is the collection of some coefficients; a 1 a 2 and a N.

So, a case are the coefficients of this vector a. So, when I have a waveform I pass it through a bunch of correlators, I get to the vector corresponding to this waveform. And I can generate this waveform also like this by multiplying these various orthonormal functions with these coefficients and these are the elements of that vector a. And what I have said is you can think about any modulation scheme using the simple basic idea, any waveform is weighted sum of orthonormal functions all right.

So, here my receiver would be receiving a t and it would be giving me a vector corresponding to this a t, because I know how to handle this vector, I know how to detect this vector. So, once I have this vector then it looks like my problem is half solved. And only thing that now I have to prove is that when I go from this step to this step, I do not lose out on information ok or nothing bad happens that I cannot hand it.

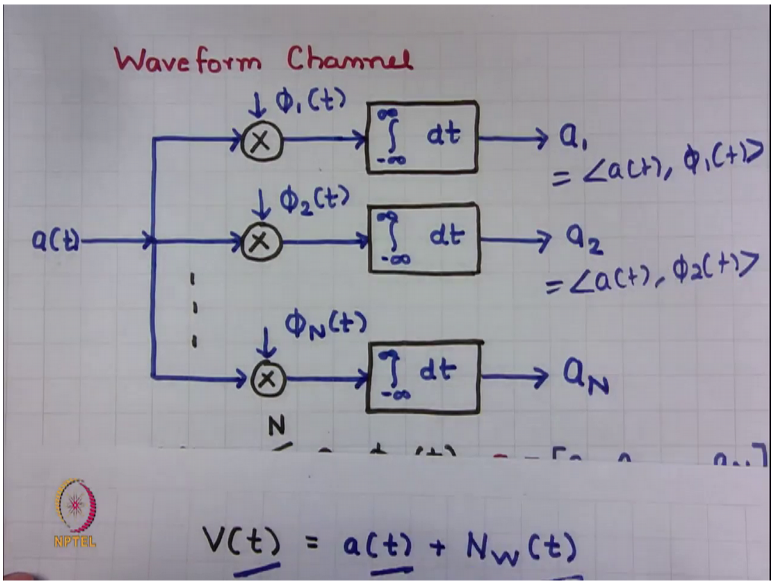
(Refer Slide Time: 34:47)



So, at the receiver I am receiving a waveform which is a t the transmitted one plus noise; that is what we have said we are assuming AWGN channel. So, noise adds to the signal, the underlying noise is assumed to be white Gaussian. So, N w t is nothing, but it is additive white Gaussian noise. And we have also said sometime back that additive means that it is independent of signal.

Let us see what happens if this random process, this is a random process, because this is a random process. So, V t is a random process and hence we are using this capital letter V of t, whereas this is a deterministic function and hence we are using a small letter a alright. So, I am having a random process V t, I am passing this random process through those bunch of orthonormal functions that I have. And I extract an element of vector in the same way as I was extracting the element of the vector a.

(Refer Slide Time: 35:47)



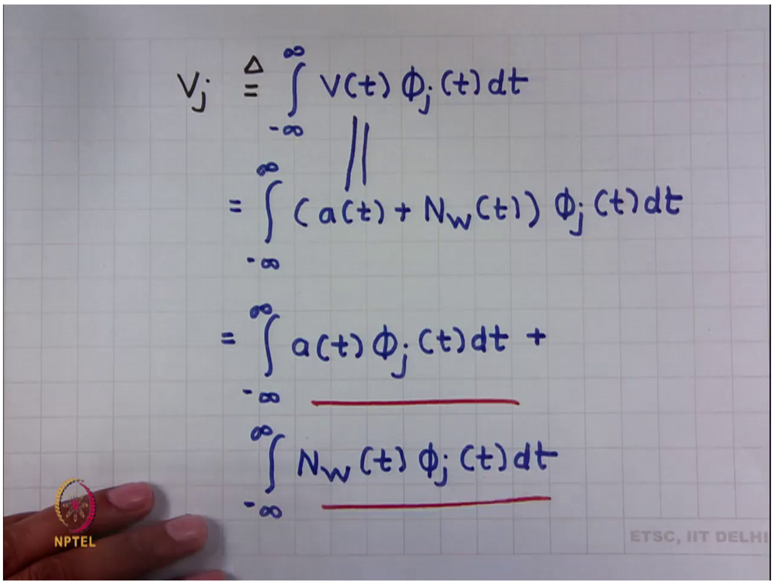
So, I pass this random process through a deterministic function. I already know what is this? This is a linear functional of the random process a linear functional of the random process, is a random variable. So, I am getting the random variable, corresponding to this orthonormal function phi j t and hence I am using a subscript j.

Similarly, I can get random variables corresponding to each orthonormal function and I can construct as a random vector. So, this is a Gaussian random vector. So, vectors in this lecture I am denoting using red color, because I am having variables and vectors at the same time. So, I thought it is a good idea to use a different color for vectors. So, I am using red color for vectors ok.

So, random vector V 1, is simply the collection of n random variables and these random variables are obtained by the projection of this random process along these orthonormal functions that I am using to generate signal set, generate or to receive both ways, it cuts both ways all right.

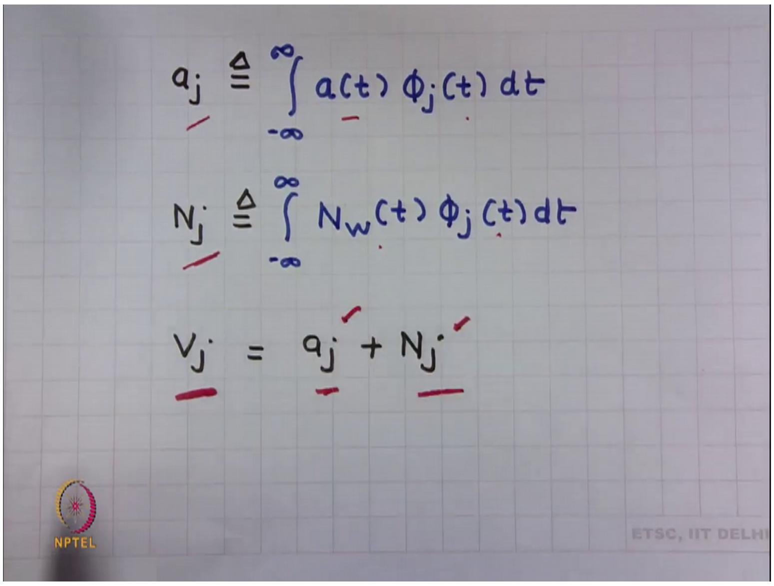
So, I have said that I am having a Gaussian random vector which I form by collecting random variables which are obtained by passing this random process through various orthonormal functions.

(Refer Slide Time: 37:27)



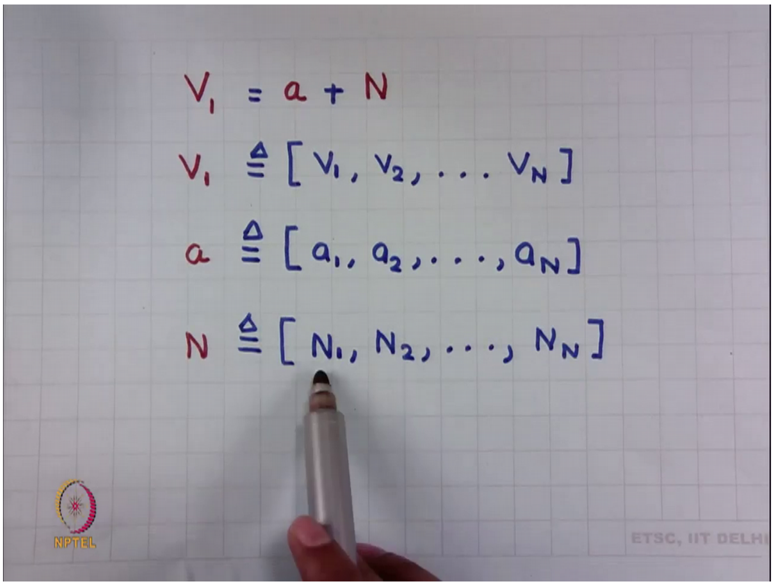
So, I am saying one random variable V j is simply obtained by projection of this random process along one orthonormal function. And I know what is this V of t? V of t is nothing, but a t plus discussion noise and I can break this down into 2 components; this plus the projection of white Guassian noise on these orthonormal functions.

(Refer Slide Time: 37:58)



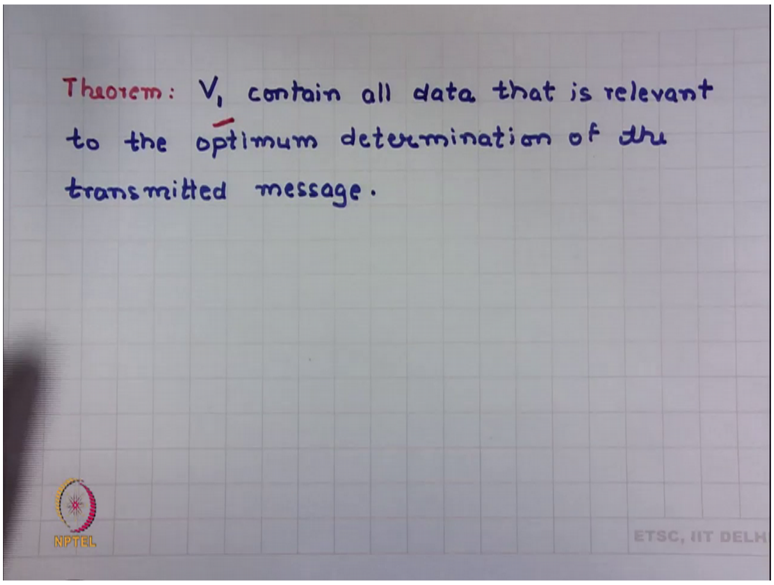
And I have already defined a js are the coefficients of projections of a t on this phi j t, and I can define n j as the coefficients of projections of this white noise on this phi j t. Once I choose this definition my V j is simply a j plus N j, where N j is the projection of white noise on these orthonormal functions, a j corresponds to the projection of signal on these orthonormal functions and V j is sum of these two coefficients ok. Its a random variable because N j is a random variable.

(Refer Slide Time: 38:40)



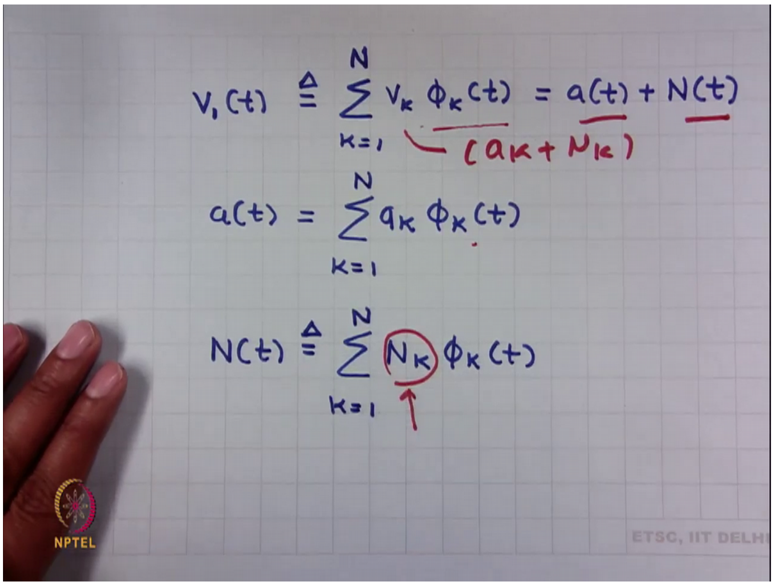
Now, if you know this then you can also know that random vector V 1, is simply a sum of 2 vectors a and N. Where the elements of a are the coefficients of projections of signal on orthonormal functions and the coefficients of N are the projections of white Gaussian noise on these orthonormal functions. So, what I am doing is, simply translating this. So, this relates that the coefficients of this random vector V 1 is sum of coefficient of vector a plus coefficient of random vector N. If V j is a j plus N j then we know that V 1 is also a plus N all right.

(Refer Slide Time: 39:37)



Why we are doing this, because we are trying to worry about stating and proving a theorem that V 1 contain all data that is relevant to the optimum determination of the transmitted message and this is a deep statement. Because it says that this random vector V 1 is sufficient and optimum for the detection of the transmitted message and this is what we are going to prove.

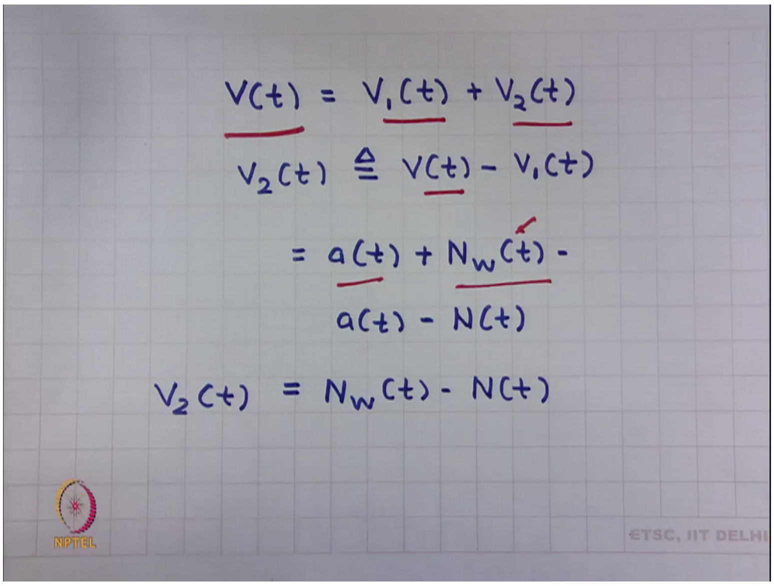
(Refer Slide Time: 40:17)



So, V 1 t, as I have said is simply a weighted sum of these orthonormal functions and the weights are decided by these random variables V k. And these random variables V k is simply a k plus N k. And we know that this is a t and this is N t, so you can also say that V 1 t is simply a t plus N t. It is the deterministic signal and N t is the white Gaussian noise projected on orthonormal functions.

So, we project white Gaussian noise on orthonormal functions, you get these random variables and you form weighted sum of these orthonormal function, using these random variables you get a random process and this random process is referred to as N t.

(Refer Slide Time: 41:17)



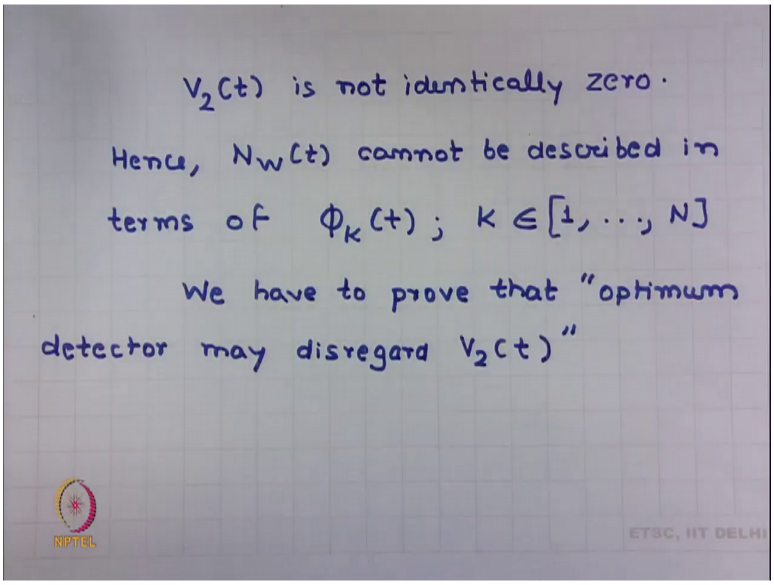
So, this is V 1 t which is a t plus N t V t; V t is the total signal that was incident on the receiver. And this total signal we say is sum of V 1 t plus V 2 t, where I define this V 2 t as V t minus V 1 t, so obvious. And what is this V t? V t is a t plus white Gaussian noise, its complete white Gaussian noise, its unprojected white Gaussian noise, it its a noise that lies in infinite dimension minus V 1 t and V 1 t is as I have said a t plus N t, N t is a finite dimensional noise. Its a noise that is lying only on the signal space. What is a signal space? Signal space is formed by those bunch of correlators.

So, if signal is expressed over N orthonormal functions, these N orthonormal functions are deciding that signal space. N t is the projection of white Gaussian noise on that signal space and w t is the total white Gaussian noise. This white Gaussian noise runs over infinite dimension. And hence V 2 t is simply this infinite dimensional noise minus the noise that is projected and the signal space. And what we are saying is, this V 1 t is only sufficient for the receiver and this V 2 t is actually irrelevant for the detection, we do not care about this V 2 t. And thus when this noise is projected only over finite number of dimension we do not lose anything on optimality, that is an argument we are trying to put forward.

So, let me restate the argument, because its kind of an important and tricky argument to understand. What we are seeing is, at the input of the receiver, we were having a signal plus infinite dimensional white Gaussian noise. Now when this signal and infinite dimensional white question noise is projected on these finite orthonormal functions. At the output of those finite orthonormal functions I am getting the signal completely plus finite dimensional noise.

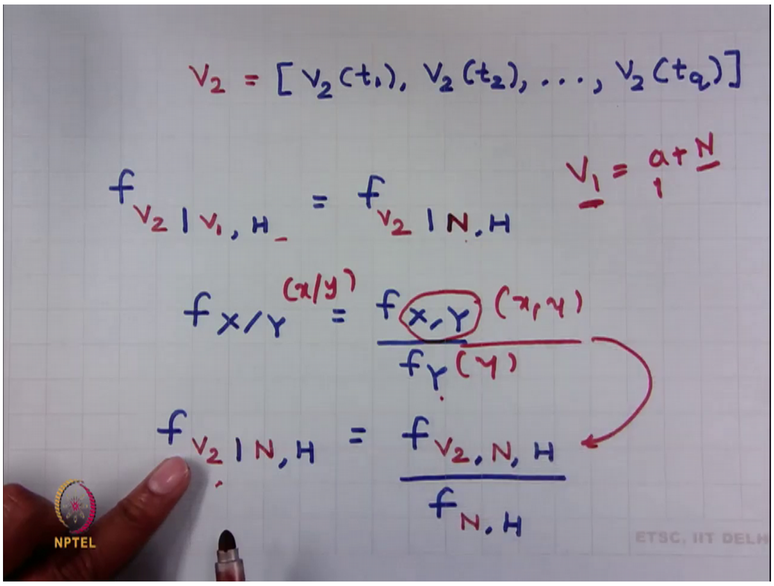
And there is certain noise that is lost out in that process. I am saying that that noise, which disappeared, is actually V 2 t and that disappeared noise is not relevant for our receiver. That noise will not improve our understanding about the hypothesis and this is what we are going to prove. V 2 t will not be identically zero and because a white Gaussian noise cannot be described in terms of finite dimensions, phi k t is a finite dimensional space, because k is a finite number. It will go from 1 to N or from zero to N minus 1 and this is what we have to prove that optimum detector may disregard this V 2 t.

(Refer Slide Time: 44:09)



And once V 2 t can be safely disregarded, we can safely say that when you convert these waveforms to vectors you are not losing out on optimality right. Whatever is lost out is not relevant for the detection. What is relevant for the detection is, the vectors that you have obtained and the output of the correlator; that is it right.

(Refer Slide Time: 45:05)



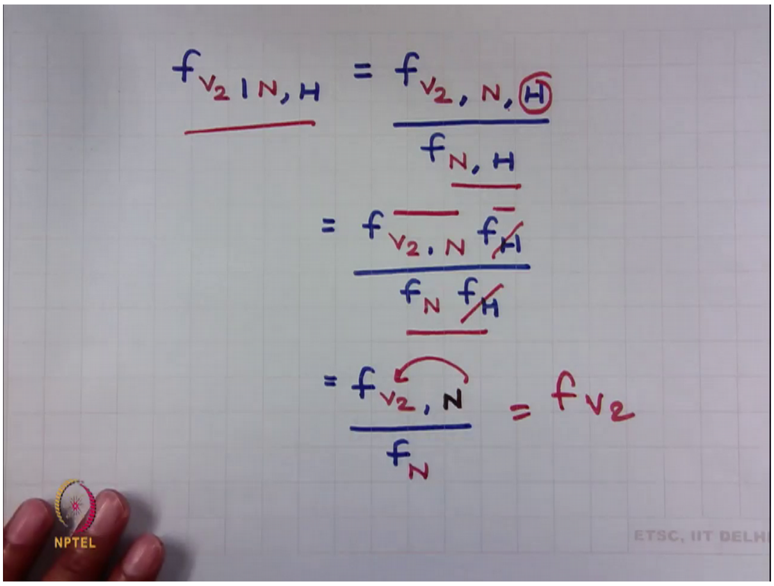
Let us see the proof. So, V 2 we are forming a random vector, red color tells us about the vector, as a collection of various random variables and I obtain these random variables by sampling the process V 2 t at various time instances. So, we have seen that if you have a random process, you sample this random process at various time instances you get random variables and this random vector is nothing, but its a collection of random variables. So, what we have to prove is, the conditional PDF of V 2 given V 1 in H is simply independent of hypothesis; that is it and then we know that this V 2 is irrelevant for deduction from theorem of irrelevance.

Let us see what is this conditional PDF of V 2 given V 1 and H, is same as the conditional PDF of V 2 given that N has taken some particular values. For example, what is this V 1? Let us say its a plus N. So, if I want to find probability density function of V 1. This is same as the probability density function of this random variable taking the numerical values N, because this is a deterministic function.

So, that is what we are saying. So, the conditional PDF of V 2 given V 1 and H, is simply conditional PDF of V 2 given N and H. And from one of the lectures on this conditional PDF we know that, the conditional PDF of X given Y is simply the joint PDF of X and Y divided by the PDF of Y. We are losing out on arguments, because this is not central here ok. I could have written this as, this as well does not matter. So, I am losing out on the arguments, just to keep things concise, but you know how you have to fail in this.

So, we know that the conditional PDF of X given Y is joint PDF of X and Y divided by PDF of Y and thus conditional PDF of V 2 given N and H is joint PDF of V 2 N and H divided by joint PDF of N and H. Using this property we can get to this.

(Refer Slide Time: 47:51)

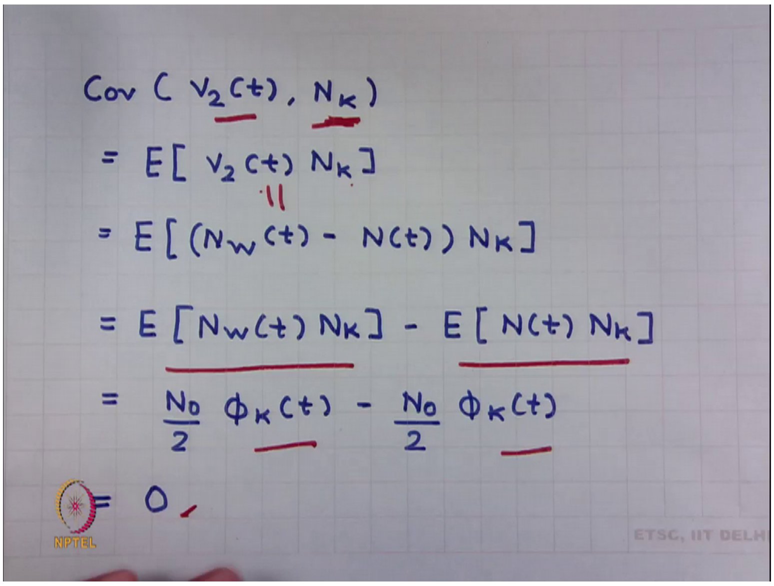


So, from this conditional PDF of V 2 given N and H, we have already said that this is the case. And this hypothesis is independent of noise and as well as on V 2, because this is a residual noise. So, hypothesis H is independent of both V 2 N H and thus the joint PDF is simply the multiplication of these PDFs. So, joint PDF of V 2 N and H, is simply the joint PDF of V 2 and N multiplied by marginal PDF of H.

And because H is again independent of N, signal is independent of noise, because its additive white Gaussian noise. So, this I can write as multiplication of these two PDF, and from here I can cancel this with this. So, from this what we can get is that this, is simply joint PDF of V 2 and N divided by PDF of N.

And now what remains is to identify that this N is also independent of this V 2. And then you can easily prove that this is nothing, but PDF of V 2. So, V 2 does not depend upon N, it does not depend upon hypothesis. That means, if you observe this V 2, neither you can learn about N noise, nor you can learn about hypothesis, you cannot learn about anything. And once you are neither learning about noise that is present in that signal space. If you are not learning about the hypothesis then of course, its a irrelevant statistic.

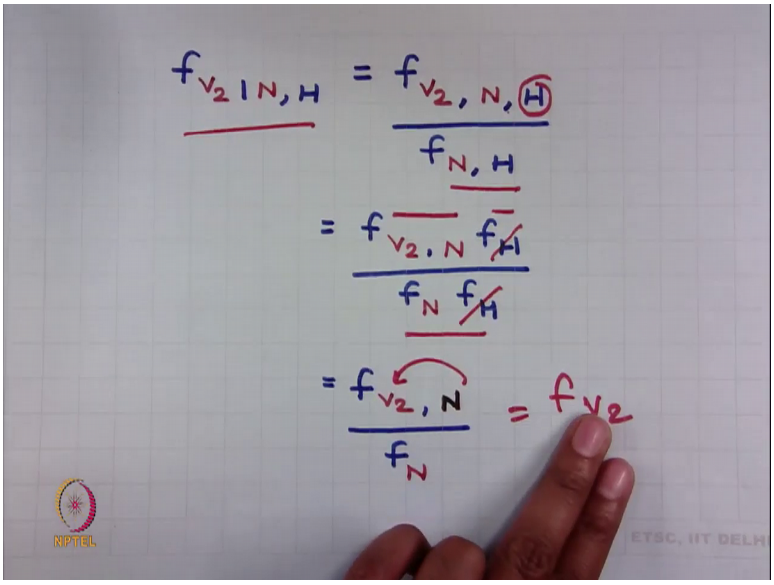
(Refer Slide Time: 49:54)



So, what we have to prove is this n is also independent of this V 2. So, to do that let us look at this covariance of V 2 t with this N case and because both are zero mean. I can simply think about this covariance as expected value of V 2 t times N k, then I can expand on this V 2 t. V 2 t is nothing, but this total white Gaussian noise minus the white Gaussian noise present in the finite dimension and we know what is N k. So, from here I can get to these two expectation.

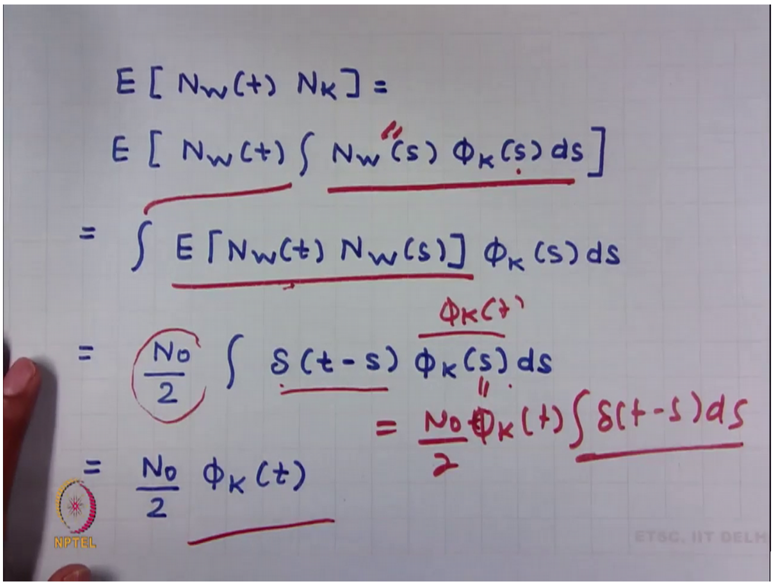
And we can prove that this is nothing, but this and this is also same as this. And from here I can get the covariance is zero. And if covariance is zero; that means, V 2 t is uncorrelated with N k. And if this V 2 t and N case are uncorrelated, we know that they are also statistically independent and this will prove that N ks are statistically independent of these V 2 ts.

(Refer Slide Time: 51:12)



And thus this will simply reduce to probability density of V 2. And thus looking at V 2 will not improve your knowledge about hypothesis. So, we have very quickly gone to this step to this step and from here to here, let us prove this as well all right.

(Refer Slide Time: 51:37)

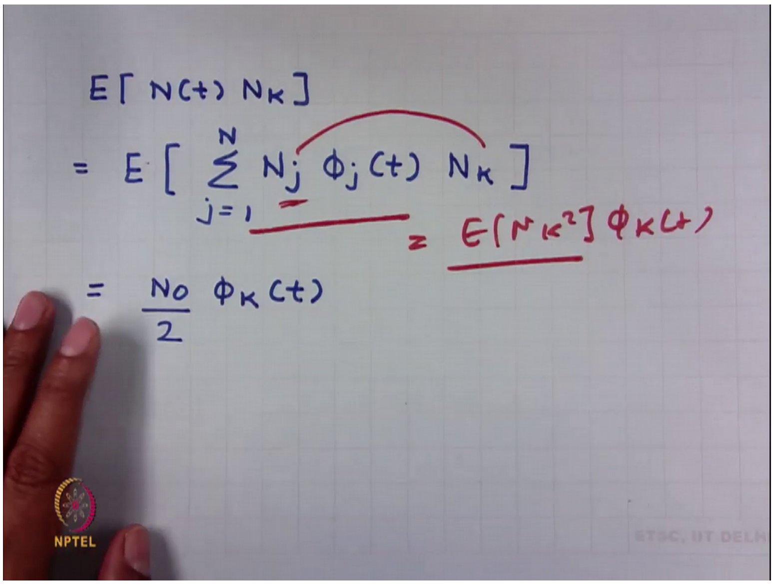


So, let us prove, what is expected value of N w t times N k. N k I can write like this, because N k is our the projection of this white Gaussian noise on these orthonormal functions. I can interchange this expectation and integration operation, pulling integration through this side, getting an expectation. Here we have done several such examples while running through week 3 4 and 5. So, I am doing this variant hast, because we have already done several such operations. In fact, we have also solved this before ok, just revising and that is why we are doing it a little bit quickly.

So, we get expected value of this. And we know what is expected value of this thing, because its a white Gaussian noise, is power spectral density is n naught by 2 and its uncorrelated. And thus we can demonstrate this through an impulse function. So, if t is different from s then you should get a flat 0, because any two samples of a white Gaussian noise are uncorrelated right. So, you can only get an expectation when s is same as t; otherwise you get a flat 0 and this effect is created by having an impulse at t minus s, because if s is not t then you get flat 0. If as the same as t then you get N naught by 2 which corresponds to the power spectral density of the noise.

So, this is a naught by 2 times integration of this thing. And this creates an effect that this s has to be t otherwise this is going to be 0. So, s is t, this is phi k t. phi k t is not a function of s, it could be pulled out. So, this becomes N naught by 2 phi k t delta t minus s t s and this is one, integration of an impulse is one, so we get this. So, there complete the first proof that expected value of N w t x N k is this thing. Now, let us check on this, what is this.

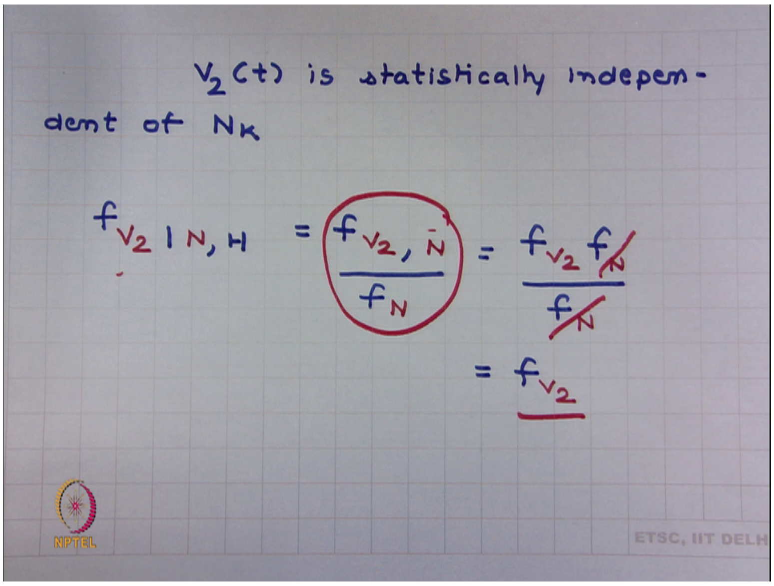
(Refer Slide Time: 54:02)



So, we have to worry about what is expected value of N t times N k. N t know I am writing in terms of orthonormal functions. Again these random variables will be uncorrelated and their expectation would be zero, except when j is same as k. So, multiplying N k with all other terms will make it flat 0 and hence in the summation there is only N k that we have to worry about. So, in that case this will be expected value of N k square times phi k t. An expected value of N k square is N naught by 2. And hence this proof is completed.

So, now you can believe me that the covariance of these two quantities is 0. Hence these two things are statistically independent. Once these are statistically independent this is same as this. So, you can write the joint PDF in terms multiplication of marginal PDFs and this cancels with this and you get conditional PDF of V 2 given N and H is simply PDF of V 2.

(Refer Slide Time: 55:03)



So, with this we have proven that in the case of waveforms, we pass this waveform through orthonormal functions, you get the signal plus the projection of infinite dimensional white Gaussian noise on the signal space and you get bunch of vectors. You get a vector corresponding to a signal; you get a vector which is obtained by the projection of that infinite dimensional white Gaussian noise onto the signal space. You add these two things up and you get a vector which we call as V 1 and we are saying that this vector V 1 is optimal for the receiver.

Whatever noise is left out is not required for detection, its irrelevant for detection, because that noise is the statistically independent of the noise present in the signal space, and its also statistically independent of hypothesis. And thus by looking at that noise you would not learn anything, either about the noise present in the signal space or about the hypothesis. And hence when we go from waveform to, vectors we lose out something, but whatever we lose out its not important for detection.

And hence when we are defining the optimum rule for waveform detectors, these rules are exactly same as the rules that we developed of a vectors. And this we will complete in the next lecture, where we will see that there is complete equivalence between the rules for waveform detection and the rules for vector detection ok. So, with this we will move into next lecture.

Thank you.