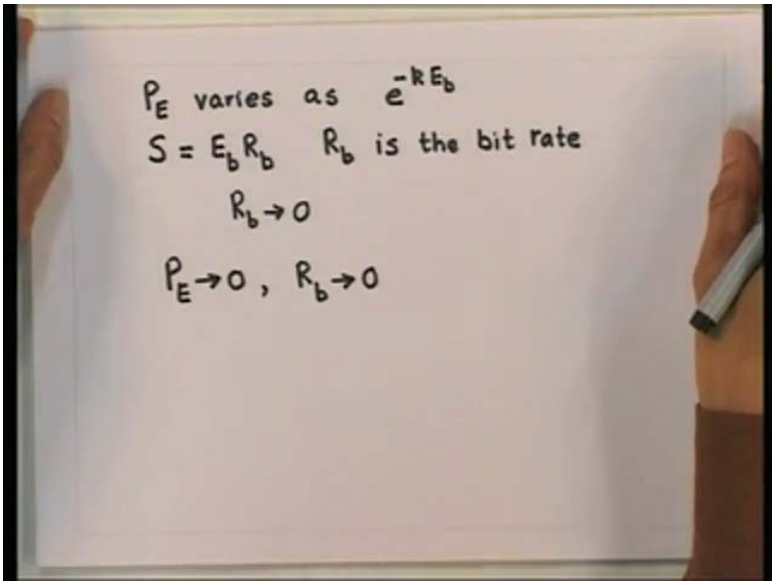
Information Theory and Coding
Prof. S. N. Merchant
Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 27
Discussion on Error Free Communication over Noisy Channel

Over a next couple of lectures, we will study Shannon's second theorem, which is the most important result in our study in information theory and coding without getting loss into the rigorous mathematical proof for the same. We will try to appreciate the importance of this theorem in the framework of a binary symmetric channel. In all modes of communication, the communication is not error free. We may be able to improve the accuracy in the transmission of digital signals by reducing the error probability. But it appears that as long as a channel noise exists, the communication be error free.

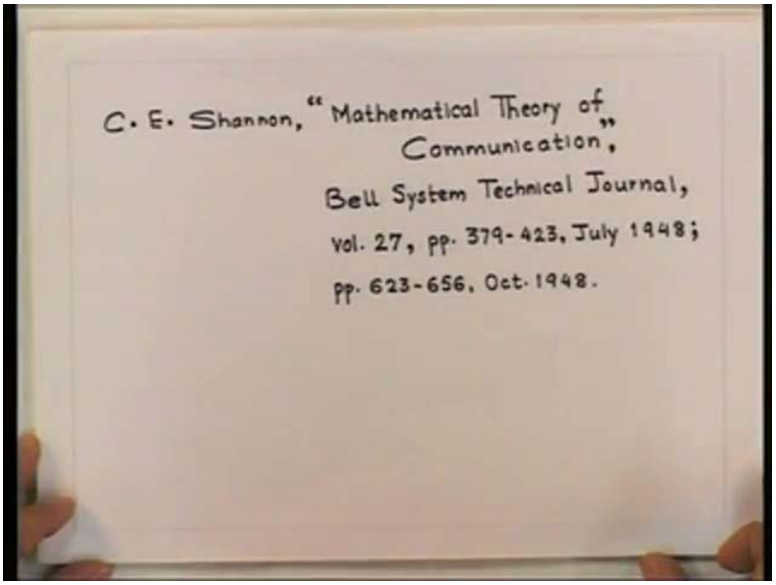
(Refer Slide Time: 01:45)



For example, it can be shown that in all digital communication system, the error probability that is P E varies as exponential minus a constant k times E b. E b is the energy per bit. By increasing E b, we can reduce the probability of error to any desired level. Now, the signal power which is indicated by S is equal to E b times R b where R b is the bit rate. Hence, increasing E b means either increasing the signal power S for a given bit rate or decreasing the bit rate R b for a given signal power S or both. Now, because of physical limitations or signal power S, it cannot be increased beyond a certain limit. Hence, to reduce P E further, we must reduce R b. What this implies that to reduce P E, we have to make R b tend to 0.

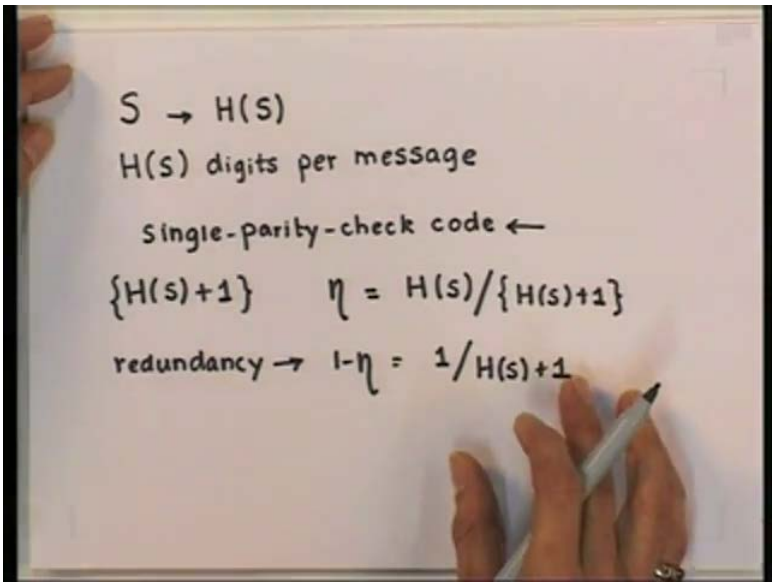
Thus, the price to be paid for reducing P E is a reduction in the transmission rate R b. So, P E tends to 0 only if R b tends to 0. Hence, it appears that in the presence of channel noise, it is impossible to achieve error free communication. Thus, that was thought by communication engineers until the publications of Shannon's classical paper.

(Refer Slide Time: 04:19)



In 1948, Shannon published his classical paper titled mathematical theory of communication in bell system technical journal in 2 parts, in July 1948 and October 1948. The gist of Shannon's paper is that the presence of random disturbances in a channel does not by itself set any limit on transmission accuracy. But instead, it sets a limit on the information rate for which arbitrarily small error probability can be achieved.

(Refer Slide Time: 05:11)



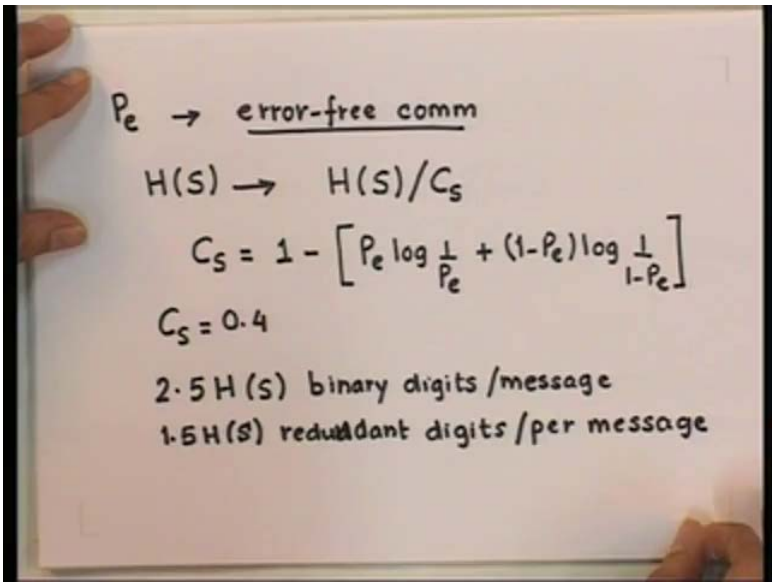
We have studied that message of source S with entropy given by H S. It can be coded by using an average of H S binary digits per message. This encoding has 0 redundancy. Hence, if we were to transmit this encoded message over a noisy channel, some of the information will be received in error. There is absolutely no possibility of error free communication over a noisy channel where messages are coded with 0 redundancy. The use of redundancy in general helped to combat noise.

This can be seen from a simple example of a single parity check code. In this coding scheme, an extra binary digit is added to each code word to ensure that the number of 1's in the resulting code; 1's code words are always either even or odd. Now, if a single error occurs in the received code word, then the parity is violated. The receiver requests for retransmission. This is a rather simple example to demonstrate the utility of redundancy. More complex coding procedures can be adopted to correct n digits error. These are discussed in later. Now, the addition of an extra digit increases average word length to H S plus 1 giving the code efficiency eta equal to H S divided by H S plus 1.

Code efficiency is defined as the entropy of the source divided by the average length of the code. The redundancy in this case is 1 minus eta. This is equal to 1 by H S plus 1. Thus, the addition of an extra check digit increases redundancy. But it also helps combat noise. So, immunity against channel noise can be increased by increasing the redundancy. Shannon has shown that it is possible to achieve error free communication by adding sufficient

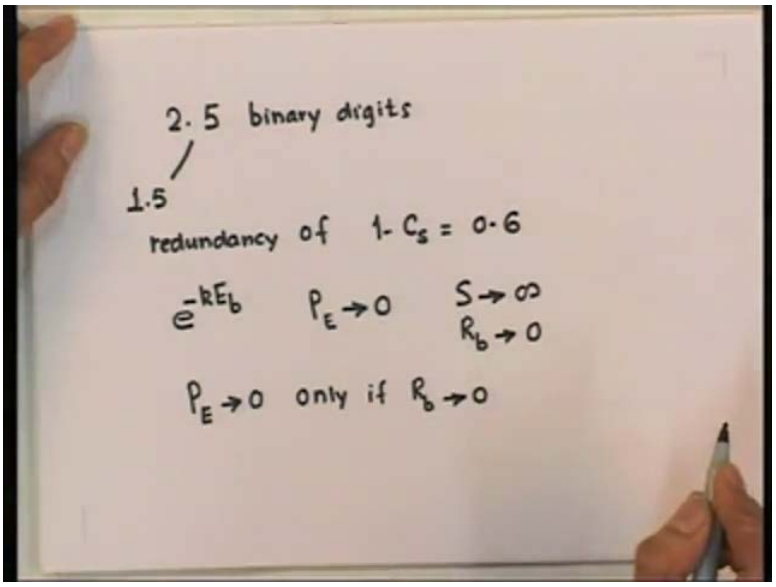
redundancy. For example, if you have a binary symmetric channel with an error probability P e.

(Refer Slide Time: 09:02)



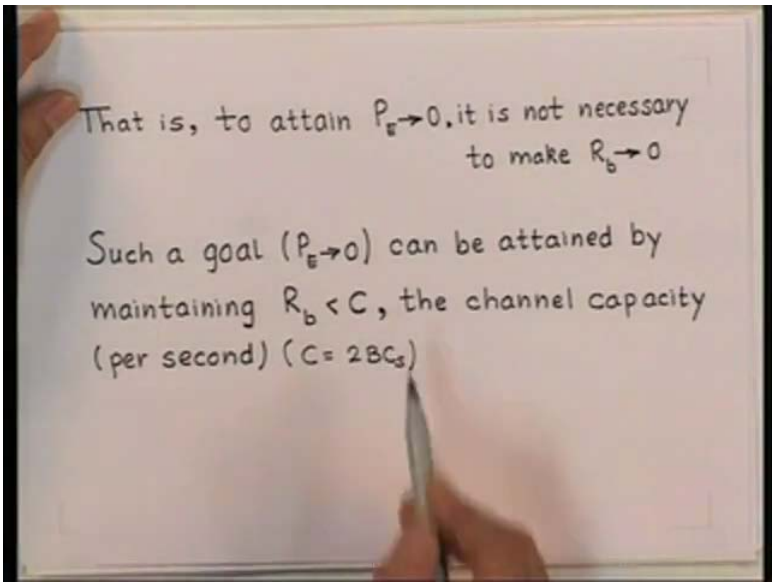
Then, for error free communication over this channel, messages from a source with entropy H S must be coded by binary codes with a word length of at least H S divided by C S. C S is the channel capacity for a single usage of a binary symmetric channel. It is equal to 1 minus P e log 1 by P e plus 1 minus P e log 1 by 1 minus P e. Now, the efficiency of this code can never be greater than C S. So, if a certain binary channel has C S equal to 0.4, what it implies? It implies that a code that can achieve error free communication must have at least 2.5 times H S binary digits per message. It is 2 and 1 half times as many digits as required for coding without redundancy. This means that there are 1.5 times H S redundant digits per message.

(Refer Slide Time: 11:32)



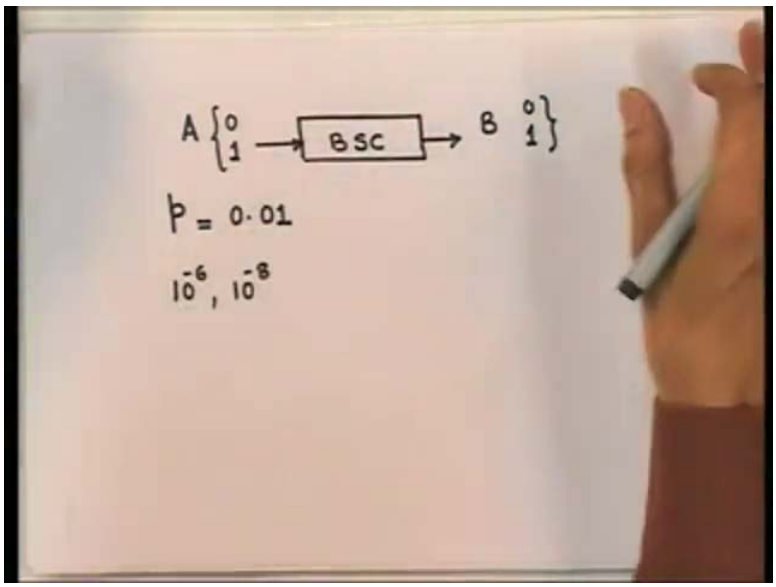
Thus, on the average for every 2.5 binary digits transmitted, 1.5 digits are redundant or check digits giving a redundancy of 1 minus C S equal to 0.6. As discussed in the beginning of this lecture, P E, the error probability of binary signalling varies as exponential minus k times E b. Hence, to make P E tend to 0, either the signal power should tend to infinity or R b should tend to 0. Now, because signal power must be finite, P E tends to 0 only if R b tends to 0. But Shannon's result shows that for a given channel as long as the rate of information digits per second to be transmitted is maintained below a certain limit, this limit is known as channel capacity. Then, it is possible to achieve error free communication.

(Refer Slide Time: 13:24)



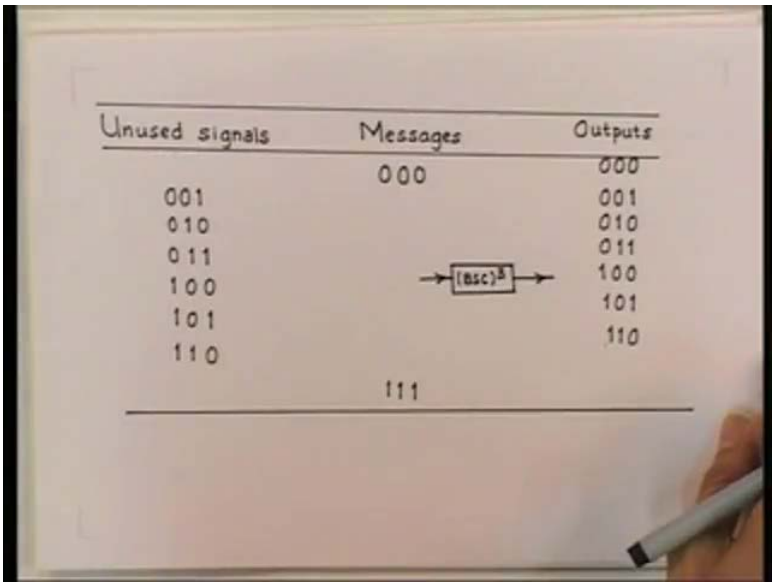
What it means is that to attain probability of P E tending to 0, it is not necessary to make R b tend to 0. So, such a goal of probability tending to 0 can be attained by maintaining the rate of information less than C. C is the channel capacity defined per second. We will derive this relation later on where C is equal to twice of the channel multiplied by C S. C S is the channel capacity per single usage of the channel. So, the question is where is the discrepancy? To answer this question, let us investigate carefully the role of redundancy in error free communication. Although, we will restrict our discussion here to a binary scheme, this discussion is quite general. It can be extended to a binary case. So, consider first the use of a binary symmetric channel to transmit reliable messages.

(Refer Slide Time: 14:47)



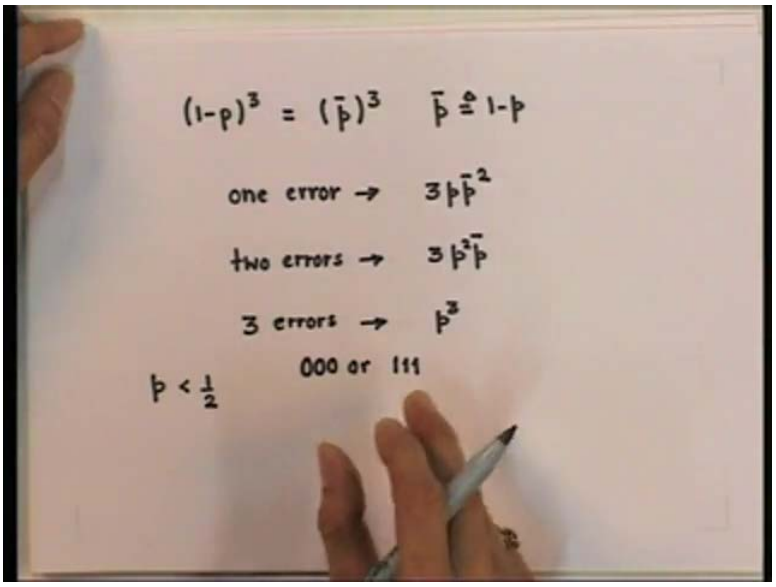
As shown in this figure, the input to this channel is alphabet A consisting of letters 0 and 1. The output of this channel is alphabet B consisting again of letters 0 and 1. To be even more specific, we assume that the probability of error of this binary symmetric channel denoted by p is equal to 0.01 means that 99 percent of the binits transmitted are received correctly. Now, for many modern data transmission systems, this level of reliability is far from adequate probability of error requirements. They are 10 is to minus 6, 10 is to minus 8 or even lower are often necessary. In order to increase the reliability of our channel, we may have to repeat our messages 0 and 1 several times. For example, suppose we decide to send each message 0 or 1 3 times. So, one method of viewing this procedure is illustrated in the figure shown here.

(Refer Slide Time: 16:35)



The output of this channel under these circumstances is an element of a third extension of a binary symmetric channel. So, the output is a binary sequence of length 3. We have used 0 0 0 and 1 1 1 sequences as message sequences. The probability that no error occurs in the transmission of our 3 digits is 1 minus p cube is equal to p bar cube.

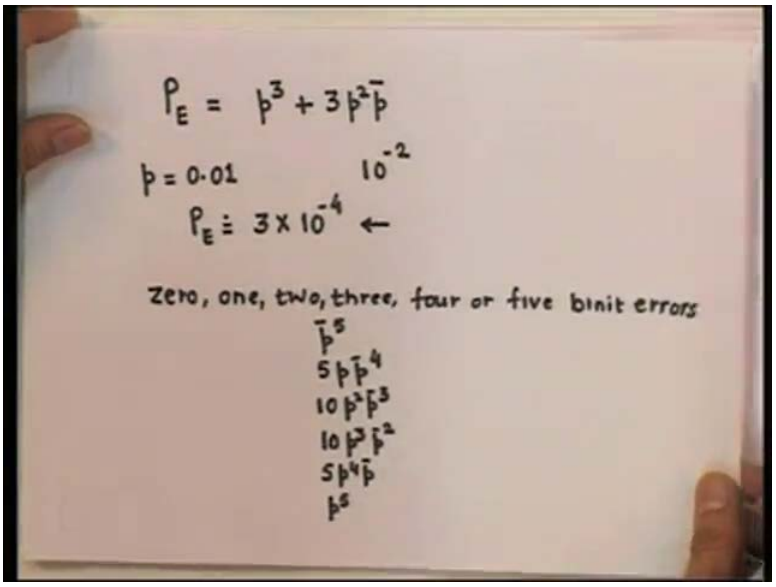
(Refer Slide Time: 17:17)



p bar is by definition 1 minus p. Now, the probability of just one error is given by 3 p p bar square. The probability of 2 errors is 3 p squared p bar. Finally, the probability that all 3 binits are in error will be given by p cube. Now, whenever p is less than half that is when the probability that a binit is received correctly is greater than the probability that it is received

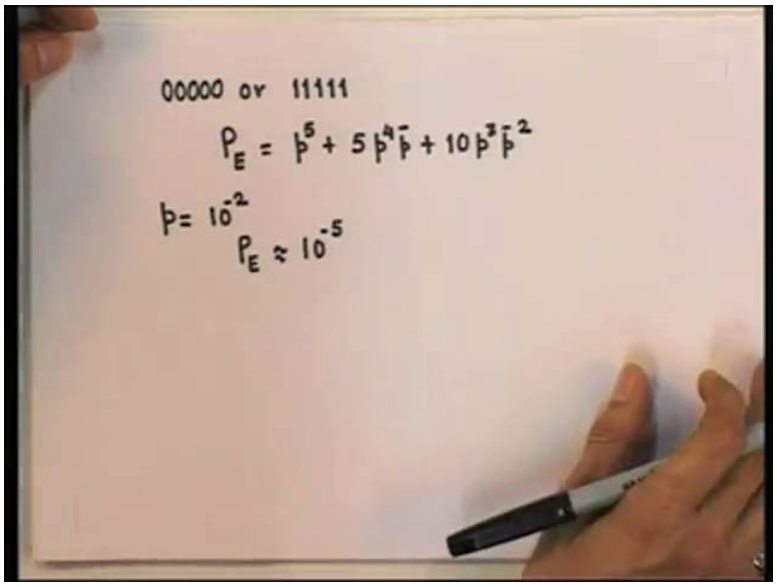
incorrectly. Then, in this case, it seems reasonable to decide that a message 0 0 or 1 1 was transmitted according to the majority vote of the 3 received binits. Now, it is easy easy to show that this decision rule is in fact the maximum likelihood decision rule. We will prove this later in this class. In any event, such a decision rule leads to a probability of interpretating the message in error.

(Refer Slide Time: 19:22)



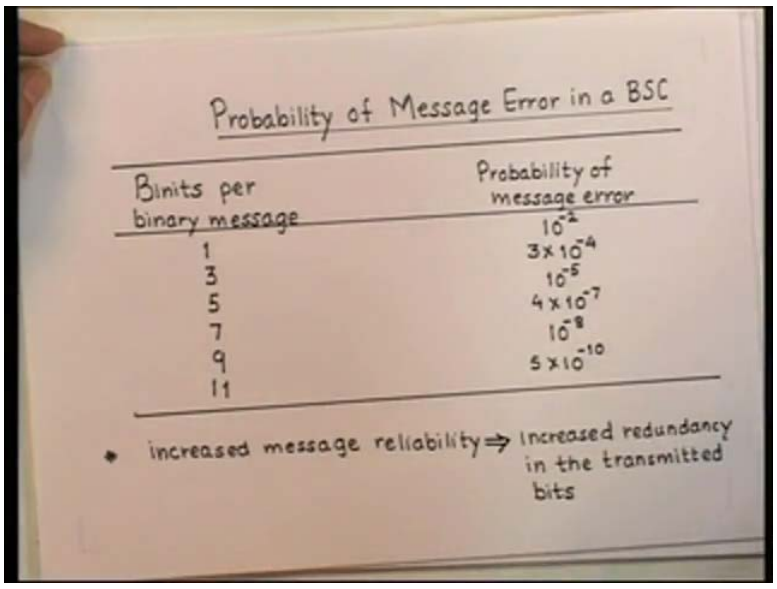
It is given by P E equal to the sum of the probability of 3 binit errors that is p cube plus 2 binit errors that is 3 p squared p bar. Now, for p equal to 0.01 this yields P E equal to approximately 3 multiplied by 10 is to minus 4. Thus, we have been able to reduce the probability of error from the value of 10 is to minus 2 to 10 is to minus 4. Having gone this far, it is not difficult to see how to increase the reliability even further. We may send five binits over the channel for each binary message we wish to transmit. This is represented in the figure as shown here. We use messages 0 0 0 0 0 to transmit 0 and the sequence 1 1 1 1 1 to transmit 1.Now, the probability of 0, one, 2, 3, four or five binit errors in transmissions is as follows. So, for the probability of 0 error is p bar 5. The probability of five binit error is p 5.

(Refer Slide Time: 22:19)



If you again use a majority rule that is maximum likelihood decision rule to decide whether 0 0 0 0 0 or 1 1 1 1 1 was transmitted, we obtain a probability of error P E equal to p 5 plus 5 times p 4 p bar plus 10 p cube p bar square that is the sum of the probabilities of five four or 3 binit errors. Now, again for p equal to 10 is to minus 2, this probability of error P E is approximately equal to 10 is to minus 5. Now, there is of course no limit to this crude method of increasing reliability.

(Refer Slide Time: 23:30)



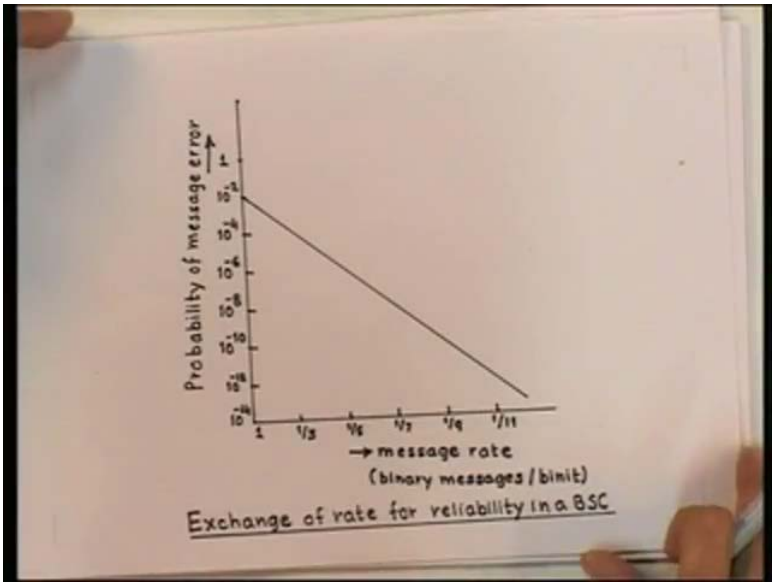
In the table shown here, we give the probability of message error when 1, 3, 5, 7, 9, 11 binits per message are used in a binary symmetric channel with single binit probability of error of p

equal to 10 is to minus 2. The improvement displayed in this table is not achieved without penalty the price we pay. For increased message, reliability is increased, redundancy in the transmitted binits. In other words, although we may decrease the probability of error from 10 is to minus 2 to 5 multiplied by 10 is to minus 10 by going from 1 binit per binary message to 11 binits per binary message, we also decrease the message rate from 1 message per binit to 1 by 11 message per binit. In general, the simple repetitive method we have described can lead to an exchange of message rate for message reliability.

(Refer Slide Time: 24:53)

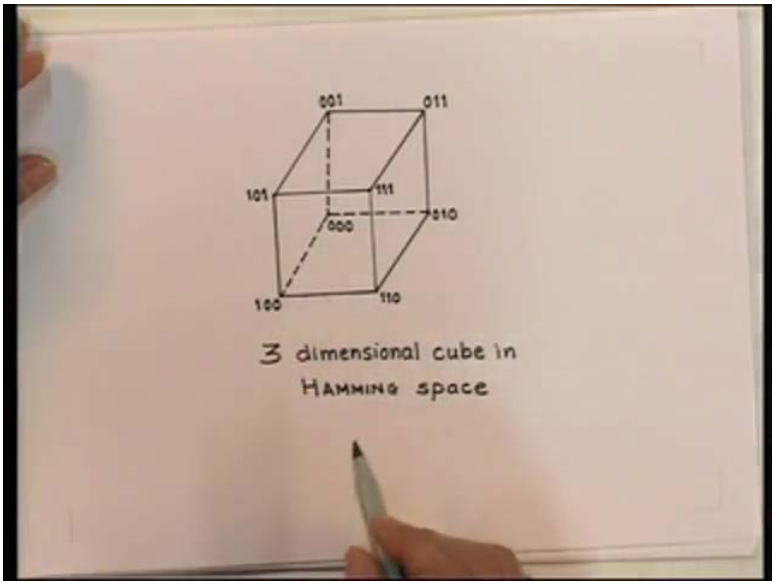
Typically, such an exchange will appear as plotted in the figure shown here.

(Refer Slide Time: 25:26)



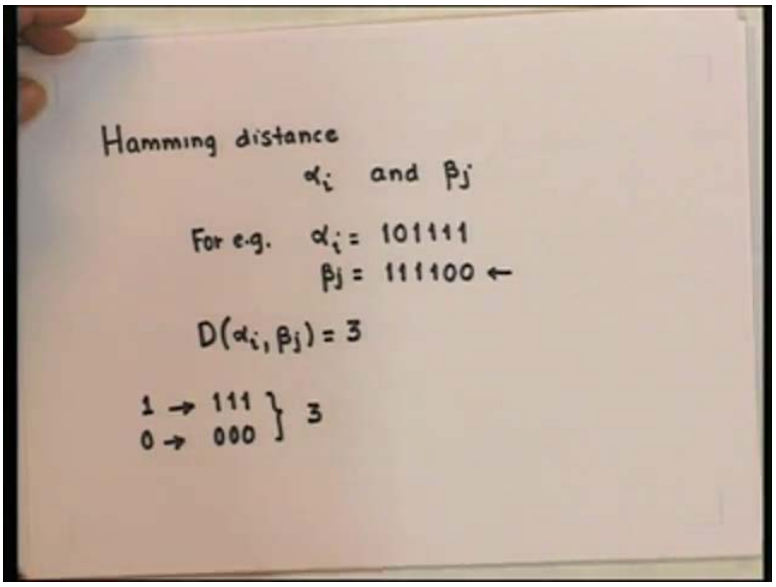
So, the probability of message error decreases with the increase, with the decrease. So, the probability of message error decreases with the decrease in the message rate. In any case, repetitions cause redundancy, but also improve the probability of error. Now, it will be instructive to understand this situation from a graphic point of view. Consider case of 3 repetitions. We can show all 8 possible sequences of 3 binary digits graphically as the vertices of a cube as shown here.

(Refer Slide Time: 26:24)



This is a 3 dimensional cube in hamming sphere. This is a three dimensional cube in hamming space. It is convenient to map binary sequences as shown in this figure and to talk in terms of what is called the hamming distance.

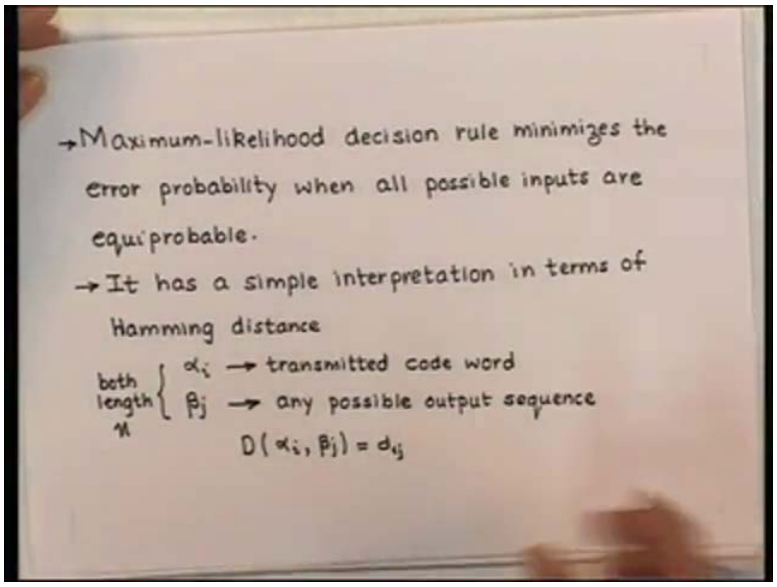
(Refer Slide Time: 26:55)



Hamming distance between any 2 binary sequences of the same length alpha i and beta j is defined as the number of places in which alpha i and beta j differ. For example, if alpha i is equal to 1 0 1 1 1 1 and beta j is equal to this sequence. Then, hamming distance denoted by D between alpha i and beta j is equal to 3 because alpha and beta j differ in 3 places. So, in the case of 3 repetitions, we transmit binary 1 by triple 1 1 1 and binary 0 by triple 0.

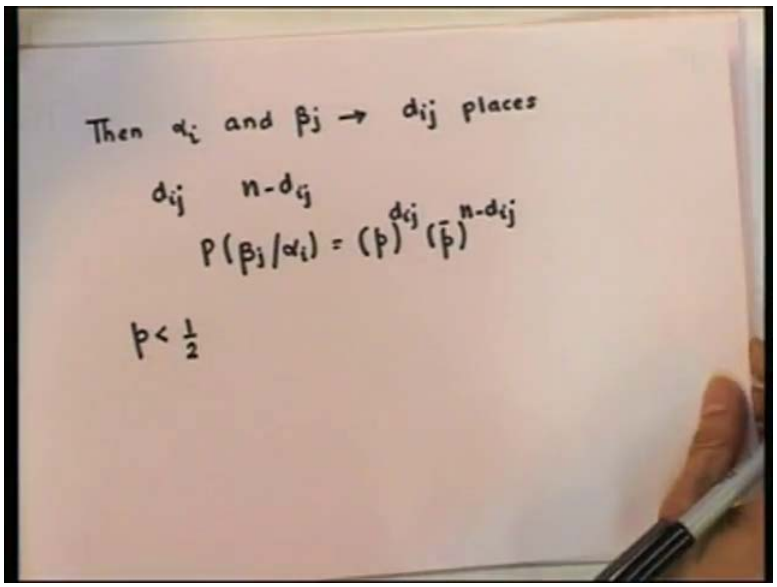
The hamming distance between these 2 sequences is 3. Observe that of the 8 possible vertices, we are occupying only 2 vertices 0 0 0 and 1 1 1 for transmitted messages at the receiver. However, because of channel noise, we are liable to receive any one of the 8 possible sequences. Now, the majority decision loss rule can be interpreted as a rule that decides in the favour of that message 0 0 0 or 1 1 1, which is at the closest hamming distance from the received sequence.

(Refer Slide Time: 29:46)



Let us provide a simple proof for the same we have studied that maximum likelihood decision rule minimizes the error probability when all possible inputs are equi probable. Now, this rule has a simple interpretation in terms of hamming distance. Let us assume that alpha i is the transmitted code word and beta j is any possible output sequence. Obviously, alpha i and beta j are of the same length n. We denote the hamming distance between alpha i and beta j as d i j.

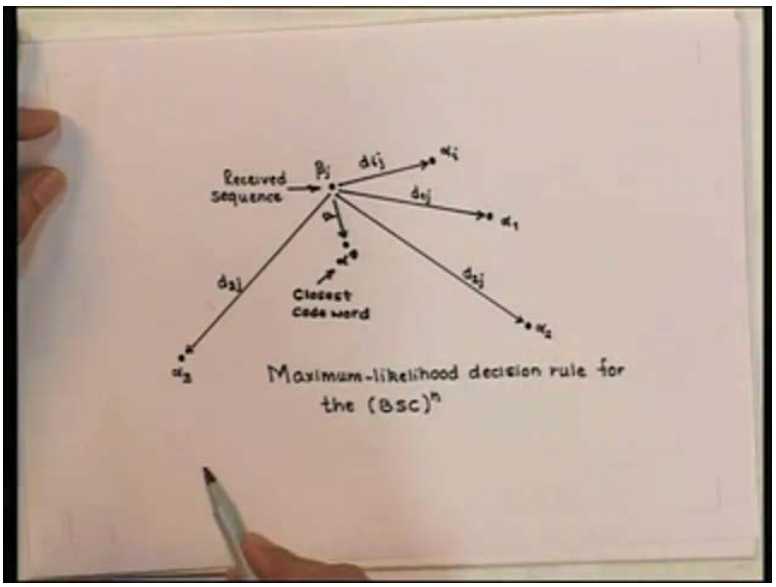
(Refer Slide Time: 30:42)



Then, alpha i and beta j differ in exactly d i j places. The probability that beta j will be received if alpha i is sent is just the probability that the errors will occur in the d i j places

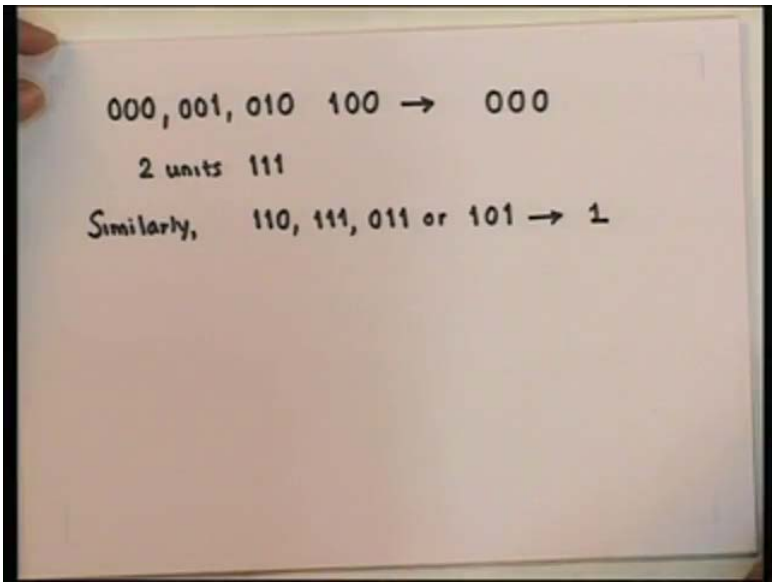
where they differ and that no errors will occur in the n minus d i j remaining places. So, probability of b j given alpha i is equal to probability p raise to d i j multiplied by p bar raise to n minus d i j for p less than half. The only case of any interest probability of b beta j given alpha i decreases with increasing d i j, further beta j is from the transmitted binary sequence. It is less likely to be received. The maximum likelihood decision rule selects that code word, which maximises probability of beta j given alpha i. Hence, for any received sequence beta j, the maximum likelihood rule selects the code word closest to beta j in the hamming distance sense. Now, this maximum likelihood decision rule can be summarised in the figure shown here.

(Refer Slide Time: 33:08)



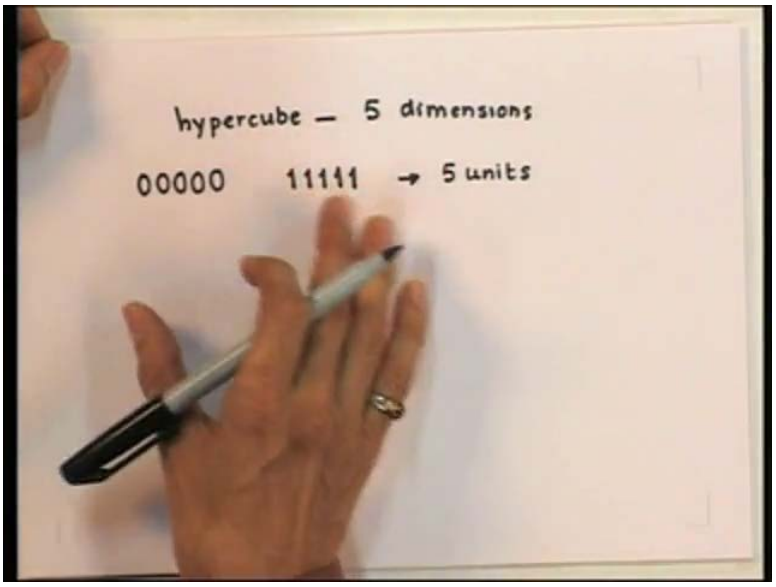
Beta j is the received sequence and alphas are the transmitted sequence. So, in this figure, beta j is closest to the code word alpha star where the distance is indicated by small d which is the smallest distance compared to all d i j's.

(Refer Slide Time: 33:50)



So, returning back to the case under discussion for 3 repetitions sequences 0 0 0, 0 0 1, 0 1 0, and 1 0 0 are all within 1 unit of hamming distance from 0 0 0. But they are at least 2 units away from triple 1 1 1. Similarly, when any of the sequences 1 1 0, 1 1 1, 0 1 1 or 1 0 1 is received, the decision is binary 1. We can now see why the error probability is reduced in this scheme of the possible 8 vertices. We have used only 2 vertices for the messages. These are separated by 3 hamming units. Now, if we draw a hamming sphere of unitary radius around each of these 2 vertices, the 2 hamming spheres are non overlapping. Now, the channel noise can cause a distance between the received sequence and the transmitted sequence. As long as the distance is equal to or less than 1 unit, we can still detect the message without error. Now, in a similar way the case of 5 repetitions can be represented by a hyper cube of 5 dimensions.

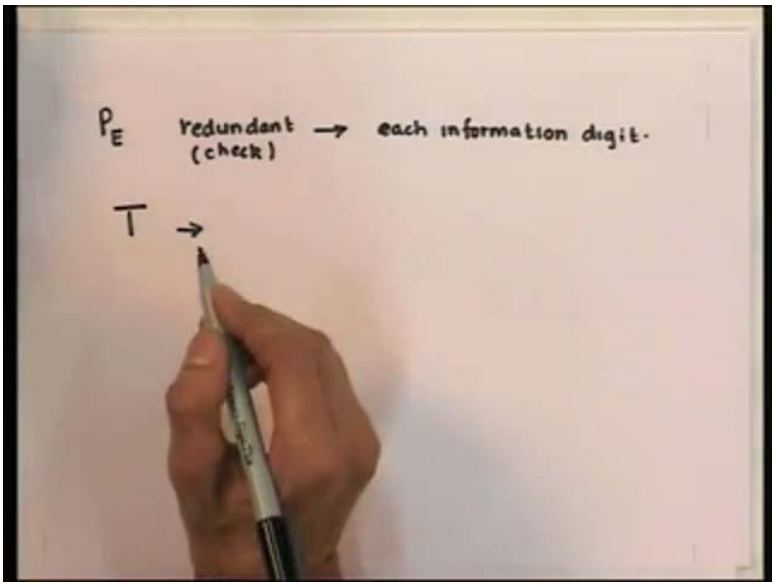
(Refer Slide Time: 35:59)



The transmitted sequences in this case are 0 0 0 0 0 and 1 1 1 1 1. These occupy 2 vertices of the hyper cube in 5 dimensions. These are separated by 5 units. Now, if we draw the hamming sphere of 2 unit radius around each of these 2 vertices, then these hamming spheres would be non overlapping. In this case, even if channel noise causes 2 errors, we can still detect the messages correctly. Hence, the reason for the reduction in error probability is that we have not used all the vertices for messages. Had we occupied all the available vertices for messages as is the case without redundancy or repetition, then if a channel noise caused an error even if it was a single error? Then, the received sequence would occupy a vertex assigned to other transmitted sequence.

We are certain to make a wrong decision precisely because we have left the neighbouring vertices of the transmitted sequences unoccupied. We are able to detect the sequences correctly despite channel errors within a certain limit. The smaller the fraction or vertices used the smaller error probability. It should also be remembered that redundancy or repetition is what makes possible to have unoccupied vertices. Now, if we continue to increase n that is the number of repetition, the probability of error will reduce. But this will also reduce the bit rate that is R b by the same factor n. But no matter how large we make n, the error probability P E never becomes 0.

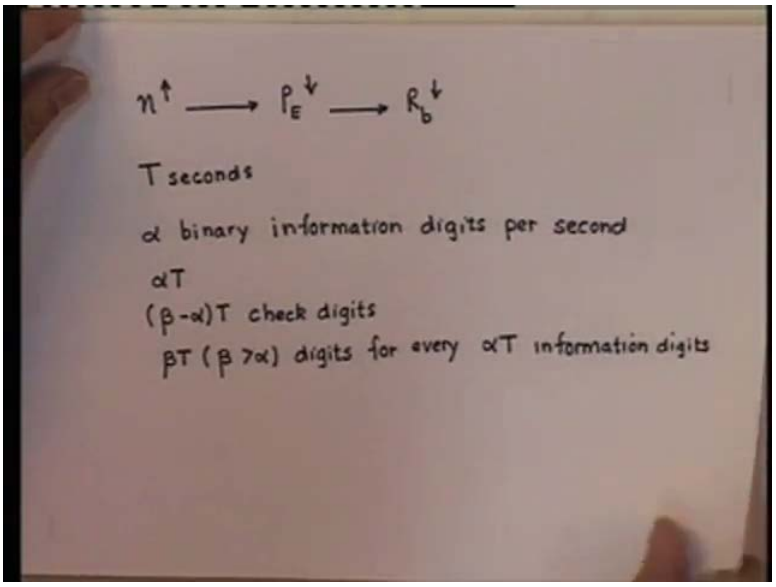
(Refer Slide Time: 38:48)



The trouble with this scheme is that it is inefficient because we are adding redundant or check digits to each information digit to give an analogy. Redundant or check digit are like guards protecting the information digit. Now, to hire guards for information digit is somewhat similar to a case of families living on a certain street hid by several burglaries. Each family panics and hires a guard. This is obviously an inexpensive and inefficient way of doing things. A better solution would be for all the families on the same street to hire a single guard. A single guard can check all the houses on the street as long as the street is not sufficiently long. If the street is sufficiently long, then we can hire more than 1 guard in using repetitions we had.

A similar analogy repetition or repeated digits is used to check only 1 information digit using the clue from the preceding analogy. It might be more efficient to use redundant for a block of information digits. This is the key to our problem. Let us consider a group of information digits over a certain interval of time of T seconds. Let us add redundant digits to check on all this digits. If we continue to increase n that is the number of repetitions, we will reduce the error probability that is P E. But we will also reduce the bit rate that is Rb by the same factor n.

(Refer Slide Time: 41:38)



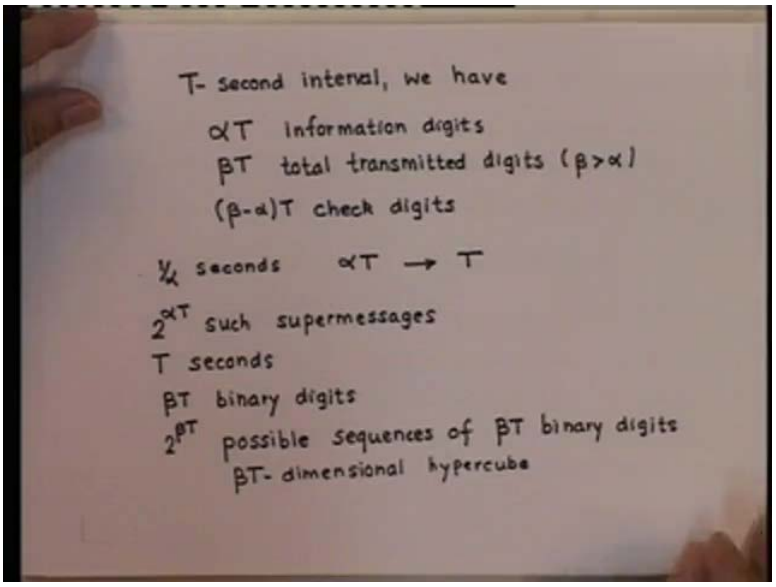
So, if n denotes the number of repetitions and if we increase this, the error probability will reduce. But this will also reduce the bit rate by the same factor n. But no matter how large we make n, the error probability never becomes 0. The trouble with this scheme is that it is inefficient because we are adding redundant or check digits for each information digit to give an analogy. Redundant or check digits are like guards protecting information digits. To hire guards for information digits is somewhat similar to a case of families living on a street, which is hid by several burglaries. Each family panics and hires a guard. This is obviously expensive and inefficient. A better solution would be for all the families on the street to hire 1 guard and share the expenses.

1 guard can check on all the houses on the street assuming that the street is not too long. If the street is too long, it might be necessary to hire more than 1 guard. But it is certainly not necessary to hire 1 guard per house. In using repetitions, we had a similar situation. Redundant or repeated digits were used to check on only 1 information digit using the clue from the preceding analogy. It might be more efficient if we use redundant digits not to check or guard any 1 information digit, but a block of information digits. Herein, lays the key to our problem. Let us consider a group of information digits over certain time interval of T seconds. Let us add some redundant digits to check on all this information digits.

Suppose, we need to transmit alpha binary information digits per second, then over a period of T seconds, we have a block of alpha T binary information digits. If this block of information digits, we add beta minus alpha times T check digits, that means beta minus

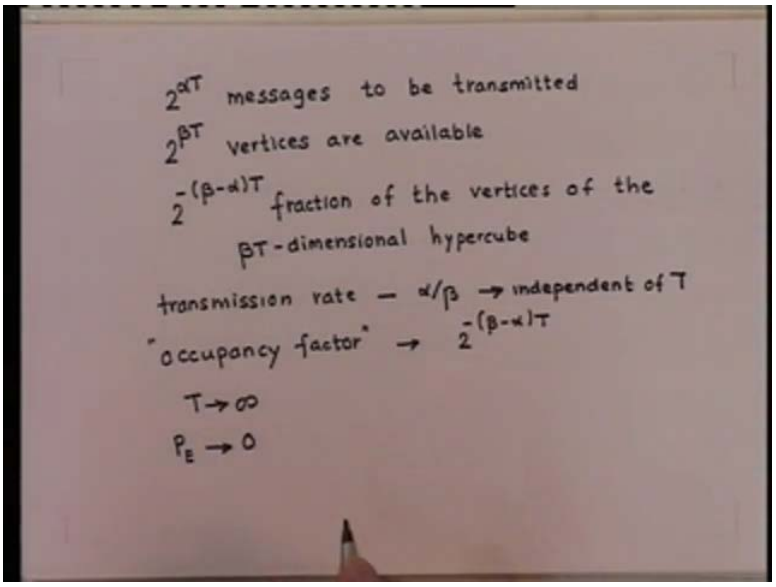
alpha check digits or redundant digits are added per second. Then, we need to transmit beta T, where beta is greater than alpha digits for every alpha T information digits.

(Refer Slide Time: 46:14)



Therefore, over T second interval, we have alpha times T information digits beta times T total transmitted. Beta is greater than alpha and this implies beta minus alpha times T are check digits or redundant digits or guard digits. Thus, instead of transmitting 1 binary digit every 1 by alpha seconds, we let alpha times T digits accumulate over T seconds. Now, consider this as a message to be transmitted. Therefore, there are total of 2 raise to alpha T such super messages. Thus, every T seconds, we need to transmit one of the 2 raise to alpha T possible super messages. These super messages are transmitted by a sequence of beta times T binary digits. What this implies that there are in all 2 raise to alpha times beta T possible sequences of beta times T binary digits. This can be represented as vertices of a beta times T dimensional hyper cube.

(Refer Slide Time: 49:30)



Now, because we have only 2 raise to alpha times T messages to be transmitted, whereas 2 raise to beta times T vertices are available, we occupy only 2 rise to minus beta minus alpha times T fraction of the vertices of the beta times T dimensional hyper cube. Observe that, we have reduced the transmission rate by a factor of alpha by beta. This rate reduction alpha by beta is independent of T. The fraction of the vertices occupied also known as occupancy factor by the transmitted message is 2 raise to minus beta minus alpha T.

This can be made as small as possible simply by increasing T in the limit as T tends to infinity. The occupancy factor approaches 0. This will make the error probability go to 0. We have the possibility of error free communication. One important question however remains to be answered. What must be the rate reduction factor that is alpha by beta for this dream to come true? We will try to seek an answer to this question in the next class. We will also look into the verification of error free communication over a binary symmetric channel.