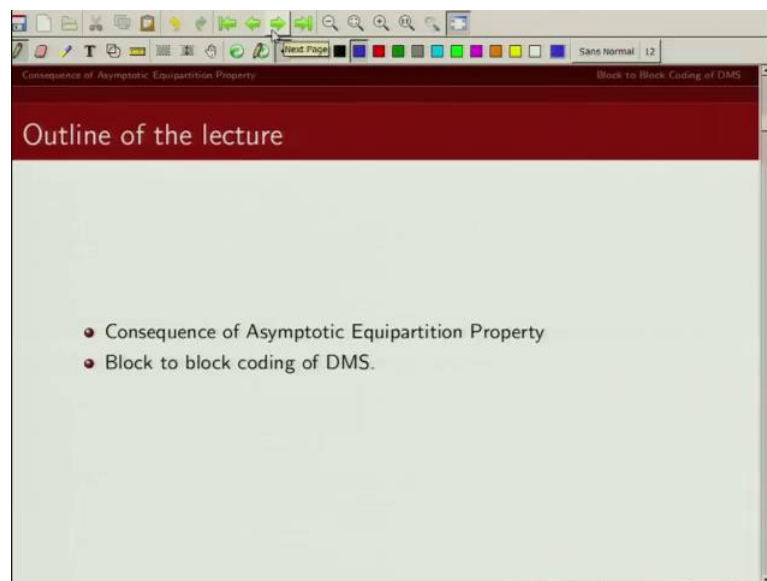


**An Introduction to Information Theory**  
**Prof. Adrish Banerjee**  
**Department of Electronics and Communication Engineering**  
**Indian Institute of Technology, Kanpur**

**Lecture – 06B**  
**Block to Block Coding of DMS**

Welcome to the course on An Introduction to Information theory. I am Adrish Banerjee. So, in this lecture, we will continue our discussion on block to block length coding of a discrete memory less source. In the last lecture, we described what we mean by typical sequence and we showed some properties of typical sequence which are collectively known as asymptotic equipartition property. In this lecture, we are going to see the consequence of those asymptotic equipartition properties what is the consequence of them for block-to-block length coding.

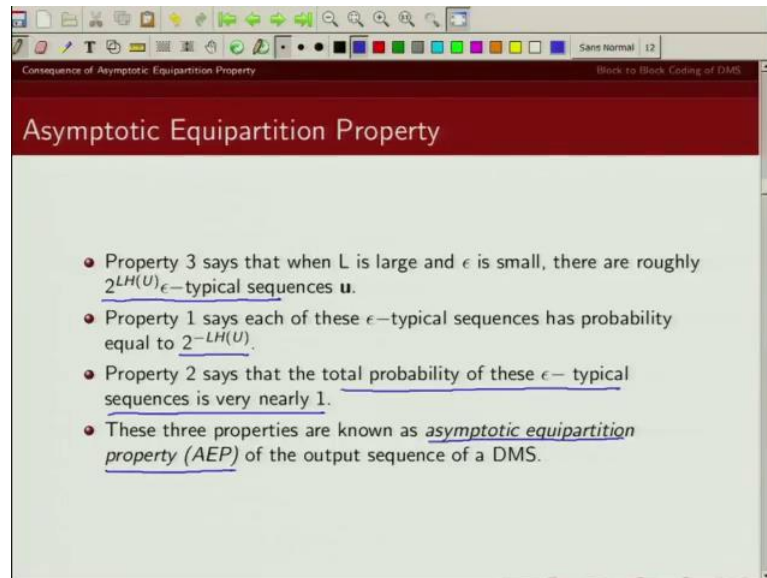
(Refer Slide Time: 01:00)



So, we will first quickly refresh our asymptotic equipartition property and its consequence and then we will talk about block-to-block length coding. Now, remember when we are talking about block-to-block length coding is a lossy compression because we have large blocks of data which we are mapping into small blocks of data. However, since given a source distribution we know not all sources are equally likely to drop in some sources are which we showed typical sequence which are likely to come out of a particular source we will try to encode those sources using unique codewords. Whereas,

the sequences which are not likely to happen, they non-typical sequence; we are not going to sign a unique codeword to them and hence we will get some lossy compression.

(Refer Slide Time: 01:58)



Consequence of Asymptotic Equipartition Property

### Asymptotic Equipartition Property

- Property 3 says that when  $L$  is large and  $\epsilon$  is small, there are roughly  $2^{LH(U)}$   $\epsilon$ -typical sequences  $\mathbf{u}$ .
- Property 1 says each of these  $\epsilon$ -typical sequences has probability equal to  $2^{-LH(U)}$ .
- Property 2 says that the total probability of these  $\epsilon$ -typical sequences is very nearly 1.
- These three properties are known as asymptotic equipartition property (AEP) of the output sequence of a DMS.

So, please quickly go over the three properties of asymptotic equipartition property. So, the third property says that number of typical asymptotic sequences given by 2 raise to the power  $L$  into  $H$  of  $U$ . And property 1 says that each of these typical sequences happens with probability 2 raise power minus  $LH$  of  $U$ . And the third property says the total probability of these typical sequences nearly 1, which means that if you take large  $n$  of  $L$ , most of time the sequences that will come out of this source are likely to be typical sequence. And as we said these three typical these three properties are collectively known as asymptotic equipartition property.

(Refer Slide Time: 02:57)

Consequence of Asymptotic Equipartition Property

Block to Block Coding of DMS

### Consequences of Asymptotic Equipartition Property

- Let  $X_1, X_2, \dots, X_n$  be independent identically distributed random variables drawn from the probability mass function  $p(x)$ .
- We are interested in short description of these sequences.
- Let us divide the set of sequences in  $X^n$  into typical set,  $A_\epsilon^{(n)}$  and its complement.
- We will require  $n(H + \epsilon) + 1$  bits to represent the typical set and not more than  $n \log |X| + 1$  bits to represent its complement set.
- We can prefix the typical set by 0 and its complement by 1.
- This code is one-to-one and easily decodable.
- Typical sequence requires short description of length  $\approx nH$ .

So, let us see now how we can use these properties to design our block-to-block length and codeword. So, let  $X_1, X_2, \dots, X_n$  be independent identically distributed random variables, they got drawn according to some probability mass function  $p$  of  $x$ . Now, we are interested in short description of these exercises right. So, let us divide this set of sequences into two sequences one, which is typical sequence; and other one, which is a complement of the typical sequence. Now, how many typical sequences exist that is block length  $2$  raise to power block length  $H$  of  $U$ . So, to represent uniquely each of the sequences in the typical set, we would require  $nH$  plus epsilon plus 1 bits right, because we know the number of typical sequences we have block length of  $L$  it is  $2$  raise power  $L$  into  $H$  of  $u$ .

So, to uniquely define each of these typical sequence using a codeword we would require these many number of bits  $n$  of (Refer Time: 04:32), the non-typical set cannot be more than  $n \log$  cardinality of  $X$  plus 1. Now, we can prefix the typical set by 0 and non typical set by one. So, essentially, we are increasing the block length by one. So, we are going to use  $nH$  plus epsilon plus 2 bits to describe codewords which belong to typical sequences and we are going to use  $n \log$  cardinality of  $X$  plus 2 bits to represent non sequences which are coming out of non typical set. Now, clearly this code, which I designed so far is one-to-one and is easily decodable. Why, because each of the typical sets I have a side unique codeword so far I have done this same thing for non typical set

as well because I am using this many number of bits and as we said typical sequence which typically require block line into H of U number of bits to represent it.

(Refer Slide Time: 05:53)

Consequence of Asymptotic Equipartition Property

Expected codeword length is given by

$$\begin{aligned}
 \underline{E(I(X^n))} &= \sum_{x^n} p(x^n) I(x^n) \\
 &= \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) I(x^n) + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n) I(x^n) \\
 &\leq \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) [n(H + \epsilon) + 2] + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n) [n \log |X| + 2] \\
 &= \underline{Pr(A_\epsilon^{(n)}) [n(H + \epsilon) + 2]} + \underline{Pr(A_\epsilon^{(n)c}) [n \log |X| + 2]} \\
 &\leq n(H + \epsilon) + \epsilon [n \log |X|] + 2 \\
 &= n(H + \epsilon')
 \end{aligned}$$

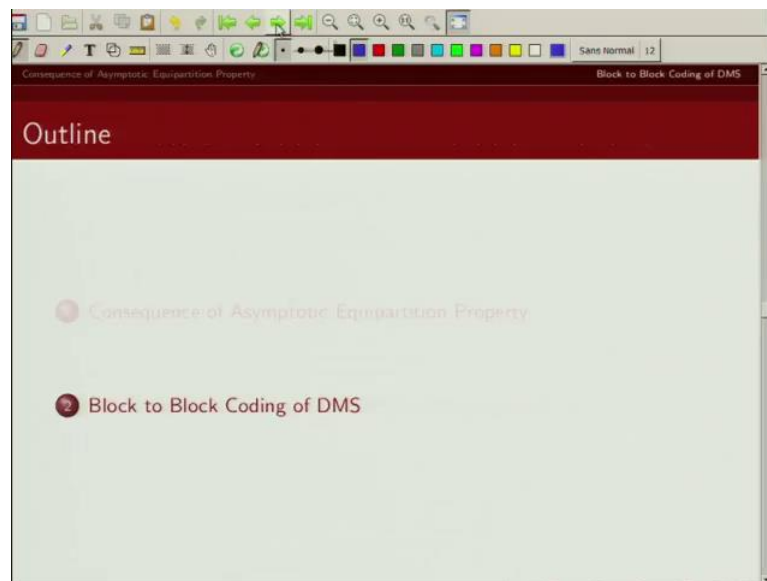
where  $\epsilon' = \epsilon + \epsilon \log |X| + \frac{2}{n}$

So, let us compute the expected codeword length. Now, expected codeword length can be written as sum over all the sequences of length n probability of occurrences of those sequences multiplied by length of those sequences. Now, we have partitioned our set of sequences into two, one corresponding to typical set, another to complement of typical set. So, this particular term that you see here corresponds to sequences are belongs to typical set and this summation correspondence to sequences that belongs to non typical set. Now, what is the length required to describe the typical set? Now I have was using n times H plus epsilon plus 1 bits to represent a typical set. Remember I was prefixing it by a 0. So, I am totally using n into H plus epsilon plus 2 bits to describe these codewords correspondence to typical set. Similarly, for the non typical set, I was using n log of cardinality of X plus 1 and I am prefixing it by 1 to denote that it belongs to non typical set. So, the length of these codewords belonging to non typical sequences is given by n log of cardinality of X plus 2.

Now, what is this probability, this probability is basically probability of occurrence of typical set and this is the probability occurrence of sequence being not a typical set. Now, we know from asymptotic equipartition property that this probability is very small some epsilon something like that and this probability is close to 1. So, I can then this

probability then can be upper bound if I consider this to be one and I consider this probability to be epsilon, I can then upper bound this probability by  $n$  times  $H$  by epsilon plus 2. And this is epsilon times  $n \log$  of cordiality of 2, this much larger than 2; so swiping at this way. Now, this can be written as  $n$  times  $H$  plus epsilon dash where epsilon dash given by this quantity this is expected codeword length if I use this scheme which as I just described. Now, epsilon is small, so you can see and  $n$  is large. So, this epsilon dash is very small, so roughly I am using close to  $n$  times entropy of this source.

(Refer Slide Time: 09:06)



Now, let us take the block-to-block length coding for theorem for the discrete memoryless source.

(Refer Slide Time: 09:13)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Given a K-ary DMS with output entropy  $H(U)$  and given any positive numbers  $\epsilon_1$  and  $\epsilon_2$ , there exists, for all sufficiently large  $L$ , a D-ary block code of blocklength  $N$  for block message sets of blocklength  $L$  such that
$$\frac{N}{L} \leq \frac{H(U)}{\log D} + \epsilon_1$$
$$N < L$$
and such that the probability,  $P(F)$ , that the codeword will not uniquely specify the message satisfies
$$P(F) < \epsilon_2$$

Proof:

So, we are given of theory discrete memoryless source, whose output entropy is given by  $H$  of  $U$ ; and we are given two positive number epsilon 1 and epsilon 2 they are small. Then for large  $L$  that exist at  $D$ -ary block length of code  $N$  for all those messages of block  $L$  such that  $N$  by  $L$  is less than equal to entropy of  $U$  by  $\log D$  plus epsilon 1, such that probability that the codeword will not be uniquely decodable that probability the codeword will not uniquely specify the message of probability of error basically that is also upper bounded by epsilon 2.

So, what are we talking about here. So, we have an input of block length  $L$  and we are coding it into output sequence of block  $N$ . So, clearly  $N$  is less than  $L$ , and then only we will get compression. Now, what should be  $N$  by  $L$ ? So, this says the relation between  $N$  by  $L$ . If you take  $N$  by  $L$  to be less than equal to  $H(U)$  by  $\log D$  plus small epsilon then probability of basically being error probability of failure is upper bounded by some epsilon. Now, we will show the converse also that this ratio  $N$  by  $L$  cannot be very, very small because if we if this ratio is very, very small then we will show this probability of error basically a failure that increases that will show subsequently. So, let us show this result.

(Refer Slide Time: 11:29)

Block to Block Coding of DMS

- Given a K-ary DMS with output entropy  $H(U)$  and given any positive numbers  $\epsilon_1$  and  $\epsilon_2$ , there exists, for all sufficiently large  $L$ , a D-ary block code of blocklength  $N$  for block message sets of blocklength  $L$  such that
$$\frac{N}{L} \leq \frac{H(U)}{\log D} + \epsilon_1$$
and such that the probability,  $P(F)$ , that the codeword will not uniquely specify the message satisfies
$$P(F) < \epsilon_2$$

Proof:

- Suppose we assign a unique D-ary codeword of length  $N$  to each of the  $M$   $\epsilon$ -typical source output sequences  $\mathbf{u}$ , but use a single additional codeword to code all the non-typical source output sequences.

So, suppose we are assigning a unique D-ary codeword of length  $N$  to each of the  $M$  typical sequences, but we use only a single additional codeword to code all non-typical sequences. So, there are total  $M$  typical sequences we are uniquely defined in those  $M$  typical sequences by a codeword. However, for all non-typical sequences, we are using only one codeword. In other words, if we transmit a non-typical sequence then the receiver will not be able to make out what was the sequence received; but if we transmit a typical sequence, the decoder will be able to exactly find out what sequence was transmitted. And remember from the AEP property we know that if  $L$  is large most of the sequences generated by this source are going to be typical sequence. So, for typical sequence, we are assigning a unique codeword, whereas, for all the set of non-typical sequence we are just using one codeword.

(Refer Slide Time: 12:47)

Block to Block Coding of DMS

- The smallest  $N$  that satisfies this condition is given by
$$D^{N-1} < M + 1 \leq D^N$$
- Thus,
$$M \geq D^{N-1}$$
or  $(N-1) \log D \leq \log M$ 

$$(N-1) \log D \leq (1+\epsilon)LH(U)$$
or  $\frac{N}{L} \leq \frac{H(U)}{\log D} + \frac{\epsilon H(U)}{\log D} + \frac{1}{L}$ 

$$\frac{N}{L} \leq \frac{H(U)}{\log D} + \epsilon$$

$\epsilon \approx 0$   
 $L \uparrow$

So, then number of codewords we require is  $M$  plus 1;  $M$  for  $m$  typical sequence and 1 for all non-typical sequence. So, this smallest  $N$  that satisfies this condition because we are using a  $D$ -ary codeword, so smallest  $N$  that will satisfy this condition given by this. So, the  $N$  should be such that that as  $M$  plus 1 is greater than  $D$  raise power  $N$  minus 1 and it is less than equal to  $D$  raise power  $N$ . So, further, I mean simplifying it I am just writing  $M$  is greater than  $D$  raise power  $N$  minus 1 or  $D$  minus take a log  $D$  minus  $n$  log of  $D$  is less than equal to log of  $M$ . And what is  $M$ ,  $M$  is upper bounded by to raise power 1 plus epsilon  $L$  times  $H$  of  $u$ . So, if I take log of that, I get this expression right.

Now, simplifying further I can write this, this relationship in this particular fashion. So,  $N$  by  $L$  is less than equal to  $H(U)$  by log  $D$  plus epsilon  $H(U)$  by log  $D$  plus 1 by  $L$ . Remember we are talking about epsilon when epsilon is very small to close to 0 we are talking about  $L$  which is very, very large. So, when epsilon is small this term will be close to 0. When  $L$  is large, this term is close to 0. So, we can write this as  $N$  by  $L$  less than equal to  $H(U)$  log of  $D$  plus epsilon 1. And remember since most of the sequences are typical sequence of probability of error is going to be small is less than epsilon from the property two of typical sequence we know the total probability of epsilon typical sequences close to one so that will then prove this result.



(Refer Slide Time: 15:20)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Given a K-ary DMS with output entropy  $H(U)$  and given any positive numbers  $\epsilon_1$  and  $\epsilon_2$ , there exists, for all sufficiently large  $L$ , a D-ary block code of blocklength  $N$  for block message sets of blocklength  $L$  such that

$$\frac{N}{L} \leq \frac{H(U)}{\log D} + \epsilon_1$$

and such that the probability,  $P(F)$ , that the codeword will not uniquely specify the message satisfies

$$P(F) < \epsilon_2$$

Proof:

- Suppose we assign a unique D-ary codeword of length  $N$  to each of the  $M$   $\epsilon$ -typical source output sequences  $\mathbf{u}$ , but use a single additional codeword to code all the non-typical source output sequences.

That if you have a K-ary discrete memoryless source, whose output entropy is given by  $H$  of  $U$  then there exist a D-ary block code of length  $L$  for message sets of length  $L$  such that  $N$  by  $L$  is less than equal to  $H$   $U$  by  $\log D$  plus epsilon 1, and this while ensuring that probability of basically making a mistake is also bounded. Now, the next question to ask is can we make this  $N$  by  $L$  ratio very small. What is the consequence of that because this says  $N$  by  $L$  is less than equal to  $H$   $U$   $\log D$  plus epsilon, can I make  $N$  by  $L$  very small. To answer to this, we will come in the next slide where we will show what happens. So, we cannot make  $N$  by  $L$  very small without increasing our probability of failure. So, we are going to show that next.

(Refer Slide Time: 16:28)

Block to Block Coding of DMS

- The block-to-block (lossy) source coding theorem for a DMS states that given any positive numbers  $\epsilon_1$  and  $\epsilon_2$ , there exists a source coding scheme as shown in Figure 1 for which

$$\frac{N}{L} < \frac{H(U)}{\log D} + \epsilon_1$$

and

$$P(F) < \epsilon_2$$

where  $P(F)$  is the probability that  $[\hat{U}_1, \dots, \hat{U}_L] \neq [U_1, \dots, U_L]$ .

So, we are going to show for a discrete memoryless source. So, given positive number epsilon 1, epsilon 2, there exist a source coding scheme such that  $N$  by  $L$  is less than this and probability of failure basically where we do not estimate our bits source bits correctly is bounded by epsilon. So, this is the kind of block diagram that we are looking at. We have a discrete memoryless source, which is fitting out this  $U$  i's. Now, this  $U$  i's are  $K$ -ary alphabet. So, each of this  $U$  i's can take a different values this is been said to and my input log length is  $L$ . This is been said to a block source and coder which take this  $\log L$  bits and transforms into block of length  $n$  and these codewords are  $D$ -ary codewords. Now, once I have this codeword at the receiver essentially I need to deconstruct and get back my original sequence. So,  $\hat{U}_1 \hat{U}_2 \hat{U}_L$  is estimate of the source bits which were encoded using this block-to-block length encoder.

(Refer Slide Time: 18:08)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Defining the average error probability over the segment of L digits by
$$P_s = \frac{1}{L} \sum_{i=1}^L P_{ei}$$
and noting that  $P_{ei} = P(\hat{U}_i \neq U_i) \leq P(F)$ , we see that
$$P_s \leq \underline{P(F)}$$
- Hence, we see that the lossy source coding theorem implies
$$\underline{P_s \leq \epsilon_2}$$
so that  $P_s$  can be made arbitrarily small.

Now, as I said we cannot make this ratio N by L very small. So, we are going to now quantify in terms of probability of error. So, we define this average probability of error in this particular fashion, where probability of error at the ith bit is defined like this. So, if  $U_i$  is not same as  $\hat{U}_i$  that is defined as probability of error at the ith bit. And if we take the average of error over all L bits divide by L that is going to give us our average probability of error. Now, this average probability of error is less than this probability of failure. So, probability of error is upper bounded by probability of F and this we have derived in the previous lecture that this depends on K, this depends on block length L, this depends on epsilon square, and this depends on p min probability of U. And since this is small average probability of error is also bounded.

(Refer Slide Time: 19:32)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Show that  $P_s$  cannot be made arbitrarily small when  $N/L$  is smaller than  $H(U)/\log D$ . More precisely, if
 
$$\frac{N}{L} \leq \frac{H(U)}{\log D},$$
 then
 
$$h(P_s) + P_s \log(K-1) \geq \left[ \frac{H(U)}{\log D} - \frac{N}{L} \right] \log D$$

Proof:

- We know that
 
$$I(U_1, \dots, U_L; X_1 \dots X_N) = \frac{H(X_1 \dots X_N)}{L} - \frac{H(X_1 \dots X_N / U_1 \dots U_L)}{L} \leq \frac{H(X_1 \dots X_N)}{L}$$

Now as I said we are going to ask the converse. So, in other words, we are going to ask can we make probability of average probability error very, very small if we make this ratio also very small. So, the answer lies here. So, here we cannot make ratio N by L much, much smaller. You can see from this relation, if we make this N by L ratio smaller and smaller then probability of error is bounded from below by larger quantity. So, we cannot make this N by L ratio much, much smaller without increasing the average probability of error. So, we are going to prove now that the average probability of error is related to number of bits that we use to encode our block of data of length L. So, binary entropy function is using capital H notation, by binary entropy function of this average probability of error plus probability of error log of K minus 1 is upper bounded by uncertainty in U divided by log D minus N by L whole multiplied by log of D.

Let us prove this. So, before you prove this I just want you to again go back and look at the block diagram - this block diagram. So, we are using  $U_i$ 's to denote our output of a discrete memoryless source. We are using  $\hat{U}_i$  to denote the estimate of this  $U_i$ 's and we are using  $X_i$  to denote our source encoded bits. So, you can clearly say  $U, X$  and  $\hat{U}$  hat they form a Markov chain right. So, we can make use of data processing lemma. So, let us proceed with the proof of this. So, this result relates the average probability of error to number of bits used to encode in the case of block-to-block length coding. Now, from the definition of mutual information, you can write the mutual information between  $U_1, U_2, \dots, U_n$  and  $X_1, X_2, \dots, X_n$  as uncertainty in  $X_1, X_2, X_n$  minus as uncertainty  $X_1$ ,

$X_2, X_n$  given  $U_1, U_2, U_n$ . And we know for discrete kind of variables this quantity is greater than equal to 0. So, I can upper bound this mutual information by uncertainties in this  $X_i$ . So, mutual information between  $U_1, U_2, U_n$  and  $X_1, X_2, X_n$  can be upper bounded by entropy of  $X_1, X_2, X_n$ .

(Refer Slide Time: 22:57)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Using data processing lemma we have  $U_i \rightarrow X_i \rightarrow \hat{U}_i$

$$\begin{aligned} I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) &\leq I(U_1 \cdots U_L; X_1 \cdots X_N) \\ &\leq H(X_1 \cdots X_N) \\ &\leq N \log D \end{aligned}$$

- Since  $U_1 \cdots U_L$  are i.i.d.

$$H(U_1 \cdots U_L) = LH(U)$$

Next, we are going to use data processing lemma. We just said that  $U_1, U_2$  basically  $U_i$ 's  $X_i$ 's and  $\hat{U}_i$  they form a Markov chain. So, then mutual information between  $U_i$  and  $\hat{U}_i$  has to be less than mutual information between  $U_i$  and  $X_i$  this is the result which follows from data processing lemma. So, we can write the mutual information between  $U_1, U_2, U_L$  and  $\hat{U}_1, \hat{U}_2, \hat{U}_L$  this from the data processing lemma follows that this is less than equal to mutual information between  $U_1, U_2, U_n$  and  $X_1, X_2, X_n$ . Now, this quantity we are shown in the previous slide that this quantity is upper bounded by uncertainty in  $X_1, X_2, X_3, X_n$ .

(Refer Slide Time: 24:08)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Using data processing lemma we have

$$\begin{aligned}
 I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) &\leq I(U_1 \cdots U_L; X_1 \cdots X_N) \\
 &\leq H(X_1 \cdots X_N) \\
 &\leq N \log D
 \end{aligned}$$

$D^N \log D^N$

So, we plug that in, so what we can write then is this, this is upper bounded by this quantity correct. And each of this exercise is  $D$ -ary random variable and this is a block of length  $N$ . So, they are total  $D$  raise power  $n$  possible values of  $X_1, X_2, X_3, \dots, X_n$  and we know from the property of entropy that this is greater than equal to 0 and less than equal to log of number of possibilities of  $X$ . So, in this case, number of possibilities of  $X_1, X_2, \dots, X_n$  is  $D$  raise power  $n$ . So, this entropy is upper bounded by log of  $D$  raise power  $N$  which is nothing but  $N \log$  of  $D$ .

(Refer Slide Time: 25:24)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Using data processing lemma we have

$$\begin{aligned}
 I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) &\leq I(U_1 \cdots U_L; X_1 \cdots X_N) \\
 &\leq H(X_1 \cdots X_N) \\
 &\leq N \log D
 \end{aligned}$$

- Since  $U_1 \cdots U_L$  are i.i.d.

$$H(U_1 \cdots U_L) = LH(U)$$

- We can write

$$\begin{aligned}
 H(U_1 \cdots U_L | \hat{U}_1 \cdots \hat{U}_L) &= H(U_1 \cdots U_L) - I(U_1 \cdots U_L; \hat{U}_1 \cdots \hat{U}_L) \\
 &\geq LH(U) - N \log D
 \end{aligned}$$

So, what we have shown so far is mutual information between  $U_1, U_2, \dots, U_L$  and  $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_L$  is less than equal to  $N \log D$ . Now, since we are considering a discrete memoryless source. So, this  $U_i$ 's are identically distributed and they are independent. So, if they are identically distributed and independent I can write this  $H$  of  $U_1, U_2, \dots, U_L$  as  $H$  of  $U_1$  plus  $H$  of  $U_2$  plus  $H$  of  $U_L$ ; and since  $U_1, U_2, \dots, U_L$  are identically distributed this will be same as  $H$  of  $u$ . So, this  $H$  of  $U_1, U_2, \dots, U_L$  can be written as  $H$  of  $U$ ,  $L$  times. So, then this will be  $L$  times  $H$  of  $U$ .

Now, from the definition of mutual information, I can write the uncertainty in  $U_i$ 's given  $\hat{U}_i$  to be equal to uncertainty in  $U_1, U_2, \dots, U_L$  minus mutual information between  $U_1, U_2, \dots, U_L$  and  $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_L$ . Now, this quantity we know from this relation that this is equal to  $L$  of  $H$  of  $U$ . So, we plug this value to here now this quantity we have upper bounded by  $N \log D$ . So, if you are subtracting larger quantity, so then this would be greater than equal to  $L H U$  minus  $N \log D$ .

(Refer Slide Time: 27:19)

Block to Block Coding of DMS

- We know that
 
$$H(U_1 \dots U_L | \hat{U}_1 \dots \hat{U}_L) \leq \sum_{i=1}^L H(U_i | \hat{U}_i)$$
- Using Fano's lemma we have
 
$$\sum_{i=1}^L H(U_i | \hat{U}_i) \leq \sum_{i=1}^L \{H_2(P_{ei}) + P_{ei} \log_2(K-1)\}$$

$$= \left\{ \sum_{i=1}^L H_2(P_{ei}) \right\} + LP_s \log_2(K-1)$$

$$\frac{1}{L} \sum_{i=1}^L H(U_i | \hat{U}_i) \leq \left\{ \frac{1}{L} \sum_{i=1}^L H_2(P_{ei}) \right\} + P_s \log_2(K-1)$$

*Handwritten notes:*  $U_i, \hat{U}_i \rightarrow L\text{-ary random variable}$ ,  $L=K$

Next, so using chain rule you can write this  $H$  of  $U_1, U_2, \dots, U_L$  given  $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_L$  we can write them as  $U_1$  given  $\hat{U}_1$  plus  $U_2$  given  $\hat{U}_1, \hat{U}_2$  like that we can write. And then we can use the property that conditioning cannot increase entropy. So, this results comes from two fact first we are applying chain rule second we use the fact that conditioning cannot increase entropy. So, if we combine those two



results we can write uncertainty in  $U_1, U_2, \dots, U_L$  given  $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_L$  this is upper bounded by  $H(U_i | \hat{U}_i)$  sum over all  $i$ 's.

Now, from Fano's lemma, we know the uncertainty in  $U_i$  given  $\hat{U}_i$ , this is upper bounded by binary entropy function of probability of error  $\log$  of number of possibilities of this random variable  $U$  and  $\hat{U}_i$  and that. This  $U$  is the theory random variable and we have shown this is given by  $\log(L - 1)$ , where  $L$  in this case is this is a theory random variable. So, in this case, number of random variables. This is cardinality  $U_i$  is  $K$ . So, from Fano's lemma, if you recall we can write  $U$  given  $\hat{U}$  is less than equal to binary entropy function of this probability of error plus probability of error  $\log(L - 1)$ , where  $U$  and  $\hat{U}$  are  $L$ -ary random variable. In this case,  $L$  is in this case  $L$  is  $K$  in our  $U$  is a  $K$  ary random variable, so that is why I am writing here this as  $K$ .

Now, next thing what I am doing is I am summing of this over all  $i$ 's. So, I am summing this over all  $i$ 's. So, I am sum this, this over all  $i$ 's, I sum this right hand side over all  $i$ 's, if I do that I get this expression. Next, I can divide this whole thing by  $1/L$ . Now, if I divide this whole thing by  $1/L$  what I get here is this term is going to be  $1/L$  summation  $H(U_i | \hat{U}_i)$ . This will become  $1/L$  summation binary entropy function of this error and this  $P_{ei} \log(K - 1)$  this was  $L$  times we know  $P_{ei}$  is given by  $1/L$  summation if you look at what is  $P_i$  just go back to definition of  $P_i$ .

(Refer Slide Time: 31:12)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Defining the average error probability over the segment of  $L$  digits by
 
$$P_s = \frac{1}{L} \sum_{i=1}^L P_{ei}$$

$$\sum P_{ei} = L P_s$$

and noting that  $P_{ei} = P(\hat{U}_i \neq U_i) \leq P(F)$ , we see that

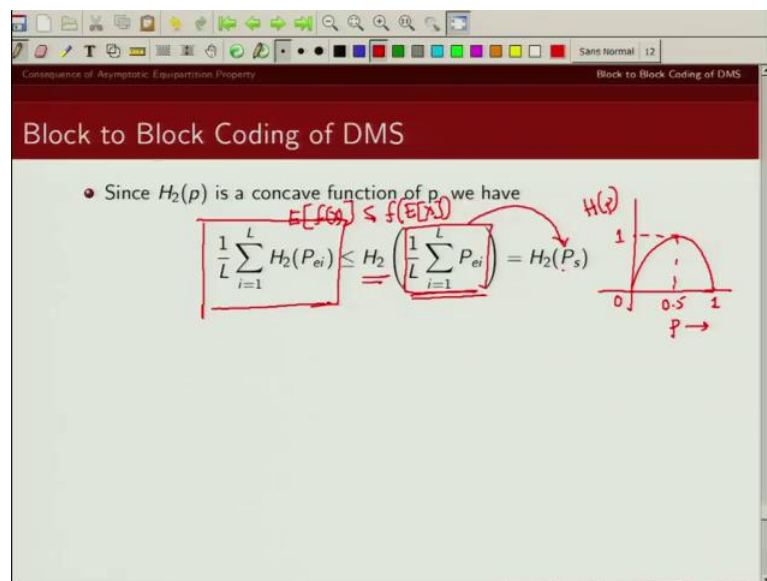
$$P_s \leq \underline{P(F)}$$

- Hence, we see that the lossy source coding theorem implies
 
$$\underline{P_s \leq \epsilon_2}$$
 so that  $P_s$  can be made arbitrarily small.



This summation  $\sum_{i=1}^L P_i$  is  $L$  times  $P_s$ , this summation  $\sum_{i=1}^L P_i$  is  $L$  times  $P_s$ . So, if we plug that value in here what we get is so when we do summation this is  $L$  times  $P_s \log$  of  $K$  minus 1 and now we are dividing by  $1$  by  $L$ . So, this will be  $1$  by  $L$  times this and this will be  $P_s \log$  of  $K$  minus 1. Now, look at this function, this is like we are calculating the expected value of a binary entropy function. Now, what do we know about binary entropy function the binary entropy function is a concave function.

(Refer Slide Time: 31:59)



A binary entropy function, if you recall looks like this, this will be 1 corresponding to  $p$  equal to 0.5 and this is 0 and 1. So, binary entropy function looks like this. So, this is the concave function. And from Jensen's inequality what do we know if the function is the concave function then expected value of the function expected value of the function is going to be less than function of expected value of  $x$ . So, since binary entropy function is the concave function of  $P$  we invoke Jensen's inequality which says expected value of the function is less than function evaluated as expected value. So, what is the expected value of the binary entropy function that is this quantity and what is the function, function is the binary function evaluated at expected value of  $S$ , here it was  $P_i$ . So, this is the expected value of this  $P_i$ . And what is this term from the definition of average probability of error this is nothing but binary entropy function of the average error.

(Refer Slide Time: 33:31)

Consequence of Asymptotic Equipartition Property

### Block to Block Coding of DMS

- Since  $H_2(p)$  is a concave function of  $p$ , we have
 
$$\frac{1}{L} \sum_{i=1}^L H_2(P_{ei}) \leq H_2\left(\frac{1}{L} \sum_{i=1}^L P_{ei}\right) = H_2(P_s)$$
- Combining above equations we get
 
$$\frac{1}{L} H(U_1 \cdots U_L | \hat{U}_1 \cdots \hat{U}_L) \leq H_2(P_s) + P_s \log_2(K-1)$$
- From above equations, we get
 
$$H_2(P_s) + P_s \log_2(K-1) \geq \frac{1}{L} \{LH(U) - N \log D\}$$

$\geq \left\{ \frac{H(U)}{\log D} \right\} \left\{ \frac{N}{L} \right\} \log D$

So, then combining this expression, and this expression what we get is of course, this expression as well we combine this expressions. So, this is says uncertainty in  $U_1, U_2, \dots, U_L$  given  $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_L$  is upper bounded by this and from here you get another upper bound. So, combine these two, we get this expression. Now, what is this term? This also we have proved earlier. This is given by go back and see this term is given lower bounded by  $LH(U) - N \log D$ . If we use this result, what we are going to get is an expression like this; and further simplifying, we get the desired expression that we wanted to prove.

So, what we have shown here is this the binary entropy function of this average error plus average power  $\log$  of  $L$ , this is lower bounded by  $H(U)$  by  $\log D$  minus  $N$  by  $L$  whole multiplied by  $\log$  of  $D$ . So, if you try to make this much, much smaller compare to this, you are going to pay a penalty in terms of your probability of error average probability of error. If this becomes much smaller than this quantity, this lower bound is going to increase, so this will increase your probability of error. So, you cannot make  $N$  by  $L$  arbitrary small, and still achieve very low average probability of error.

So, then to summarize, if you the case of block-to-block length coding is the case of lossy source compression, you take large blocks of data and we are basically coding it into small blocks of data. The trick is you should try to encode all these  $M$  typical sequence using unique codewords and assign one codeword for all non-typical

sequences. Now, since the probability of occurrence of this typical sequence is close to one, then we are considering large block size, we will be able to correctly decode at the receiver most of the time.

So, with this, I will conclude my discussion on block-to-block length coding.

Thank you.