**NPTEL**

**NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING**

**IIT BOMBAY**

**CDEEP IIT BOMBAY**

# Quantum Information and Computing


**Prof. D.K.Ghosh**
**Department of Physics IIT Bombay**

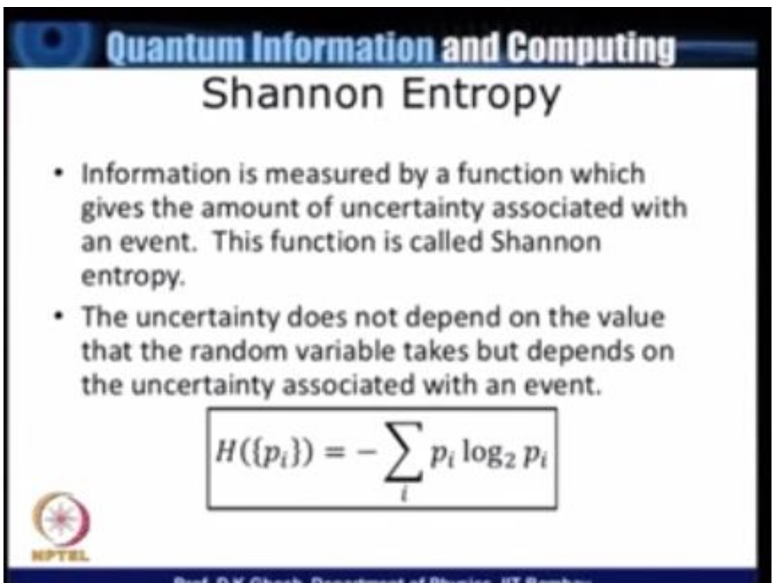

**Modul No.07**

**Lecture No.36**

## Shannon's Noiseless Coding Theorem

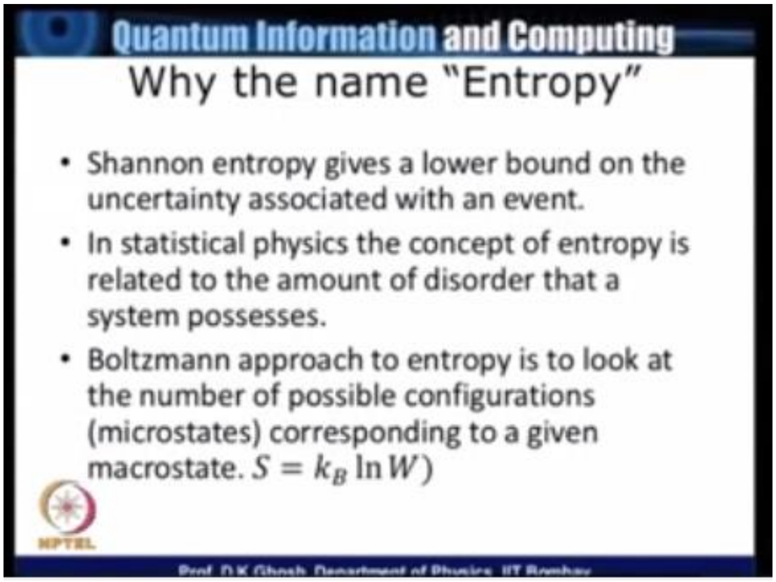
In the last lecture we have introduced a quantity which we call as the Shannon entropy which essentially measured the degree of uncertainty.

(Refer Slide Time: 00:29)



Which is associated with an event why I called it entropy we will see as we go along today. The point to be noticed regarding the quantity which I defined as the entropy is that the uncertainty does not depend up on the values that the random variable takes, but it depends up on the probabilities that is the uncertainty associated with an event and that is why we said that supposing I have an event which has various possibilities of happening and if the $i^{th}$ $^{h}$as probability $P_i$ I define the uncertainty function H as equal to $-\Sigma_i P_i$ log of $P_i$.

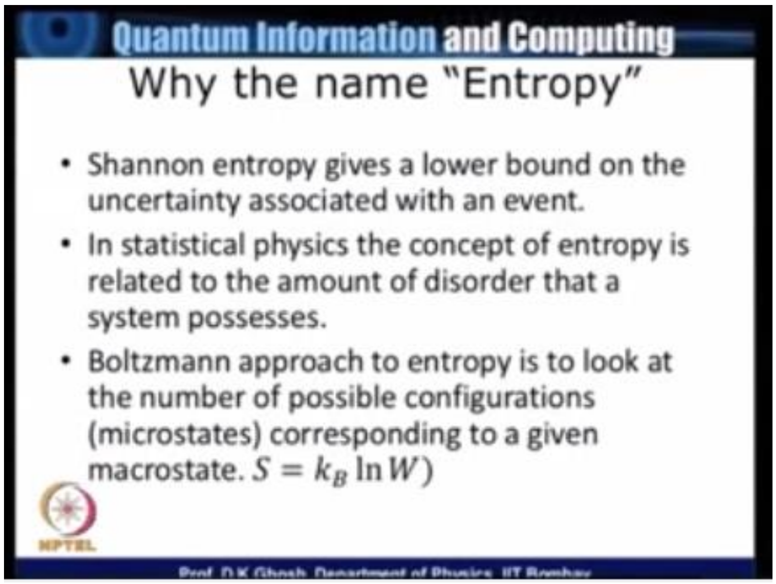
(Refer Slide Time: 01:21)




And I have pointed out that it is traditional to take the base of the logarithmic to be 2. So what is the Shannon entropy, this Shannon entropy it gives as we have discussed the lower bound that is associated with an event we had seen there are other ways of calculating information particularly in terms of asking questions out of a set of possibilities and we have shown at least illustrated that the bound that is given by Shannon entropy is realized in other words we have said that we you cannot code a message with the compression being given by a quantity less than this Shannon entropy.

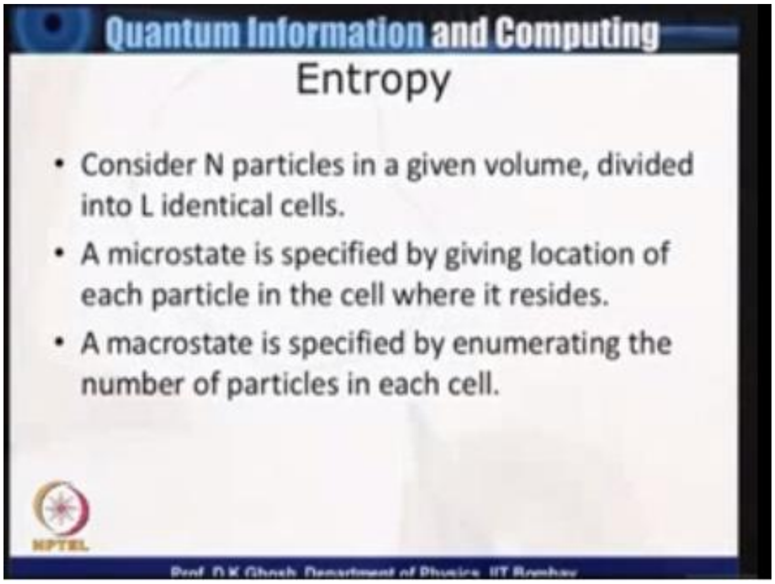
Now the next question is why do you what entropy have you recall that you have been associated with this word entropy from your knowledge of the second law of logarithms and in fact the person who introduced entropy was Boltzmann. But to understand what is the similarity of the entropy that I am talking about which is also occasionally called as the information entropy.

(Refer Slide Time: 02:50)



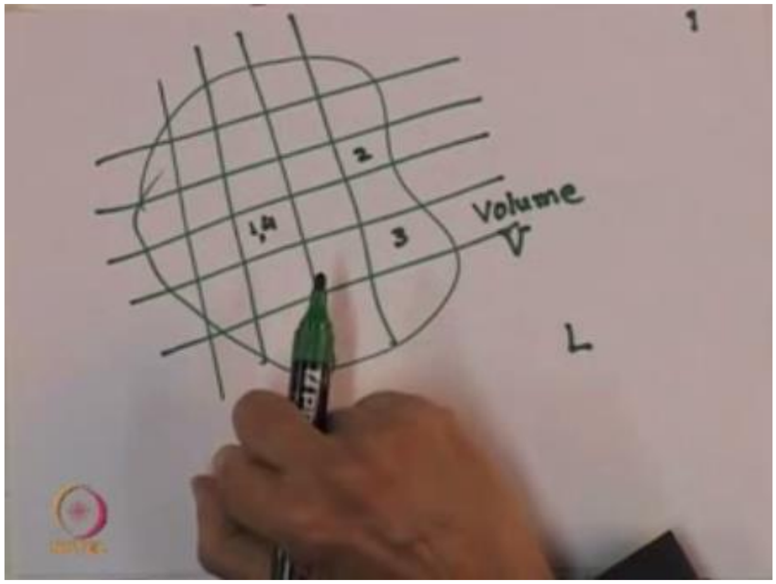

I would have to take you little bit on the concept of entropy as it is understood in statistical physics. Now what Boltzmann did was the following that supposing you have a system where the macroscopic state is decided by, now when I talk about a macroscopic state it means the states which have characteristics which are decided by the gross picture of this ensemble, but corresponding to every macroscopic state I may have several microscopic state and what Boltzmann did is to show that the entropy that we talk about in this case is given by k log w where w is the number of configurations corresponding to a macroscopic state and KD is usual Boltzmann's constant now just to be specific.

(Refer Slide Time: 03:58)




Let us consider n number of particles in a given value and let us suppose that I distribute this volume.

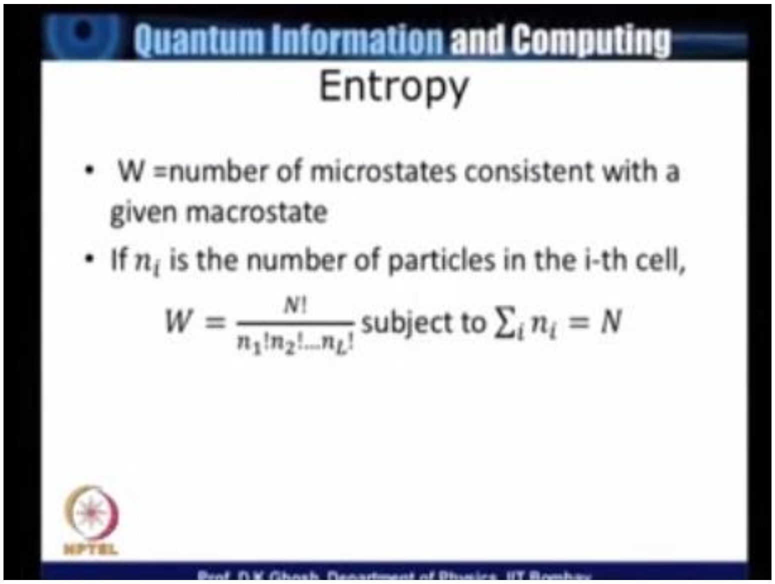
(Refer Slide Time:  04:10)




Into several cells this is the volume that you have and this volume I distribute to several identical cells so I could do this for example.  So supposing I have L number of identical cells and I have N number of particles to be located in each sides. Now when I am talking about a microscopic state or a microscope I need to supposing I number the particles from 1 to n then I need to tell you where does practice number 1 go, where does particle number 2 go so it could be for example particle number goes here particle number 2 goes there particle number 3 goes there particle number 4 goes here again like this.

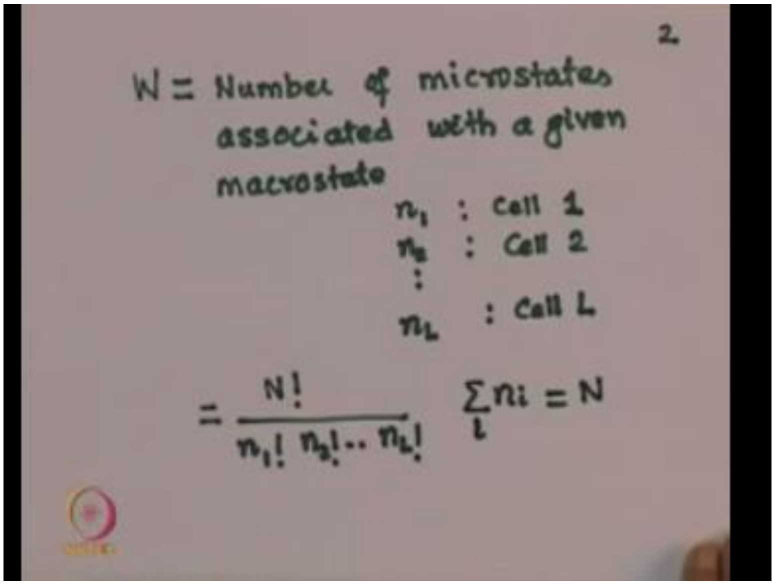
So by giving details of where the particle is I would specify a microscopic of this is. Now what is the macroscopic now while define an macro state I do not care about the individual identities of the particle but I will say there are two particles in this state there might be 4 particle in this state etc… now these 3 particles here for example in this particular cell that it is particle number 1 and 4 or whether it is particle number 7 and 20 it does not make any difference for description of a macro state so let us talk about the following.

(Refer Slide Time: 05:47)




Let us say.

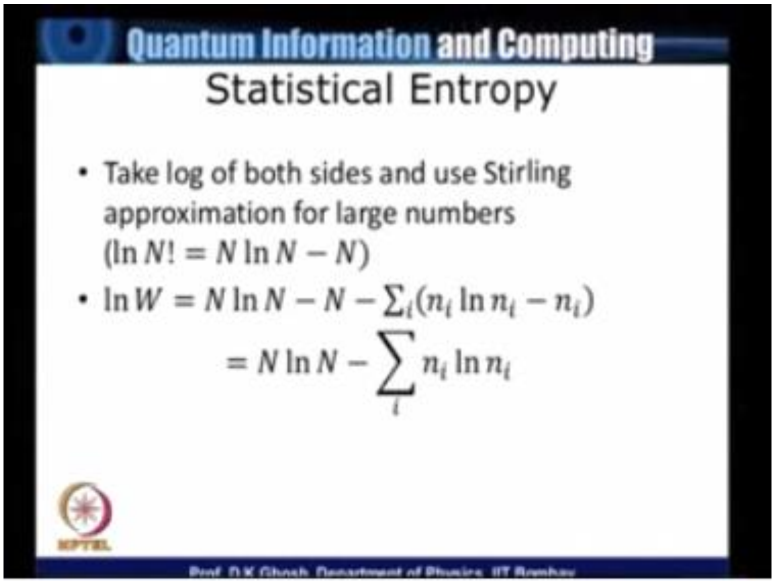
(Refer Slide Time: 05:48)




W is a number of micro states associated with a given macro state. So what does it mean? Supposing I have told you already that we have L number set, so macro state is specified by saying there are N1 number of particles in cell number 1, n2 number particles in cell 2 likewise nL number of particles in cell number L, so since I do not care about which particles are there, so the number of configurations that I have.

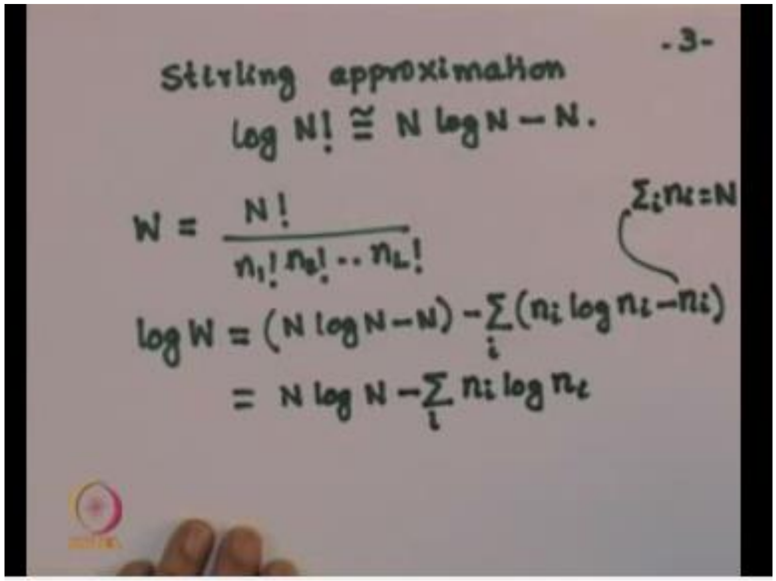
Is essentially obtained from the elementary query of permutation and combination which is given by factorial N where capital N is the number of particles that I have divided by n1! n2! …..nL! Now clearly sum over i ni = N, now what I do is this, I take logarithm on both sides.

(Refer Slide Time: 07:24)



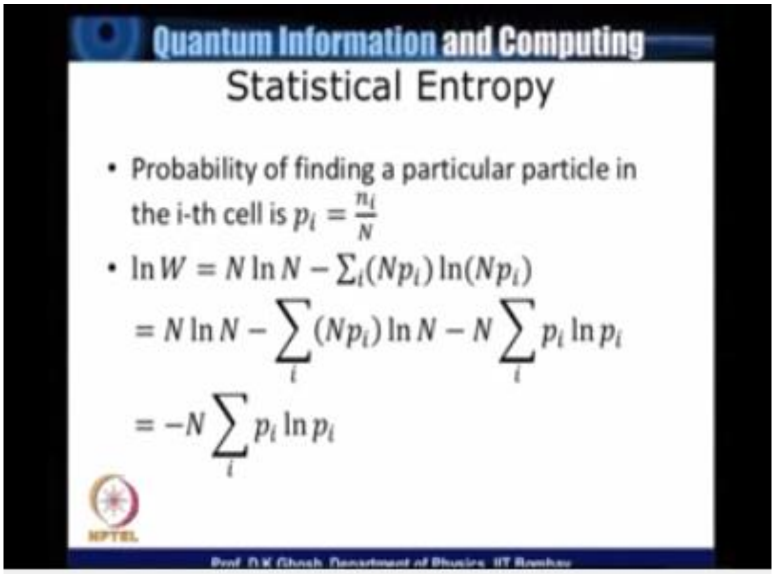

Now when I take the logarithm on both sides I use a formula for large number I assume all these numbers that I have are large because I am talking about a statistical ensemble in a molecular assembly so this Stirling approximation tells me.

(Refer Slide Time: 07:41)



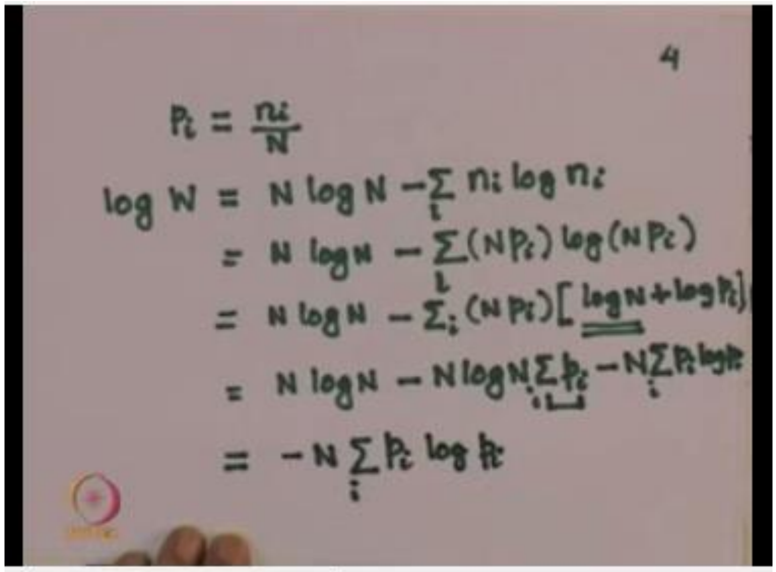

That if you have a large number log of and you take the logarithm of the factorial then you can approximate this as N log N – N. So let us look at what that it give me regarding W so which was !N/ n1! n2! …nL! So I get log W = log of!N so this term is N log N – N – log of all these things and so therefore - sum over I (ni log ni – ni) you can see that this –n and +n here which comes because of the fact that sum over i ni = N they cancel out and I am left with N log N – sum over i ni log n.

(Refer Slide Time: 09:20)




Now I would like to write it in a slightly different way by realizing that what is the probability of finding a particle, particular particle in this cell i, so since I said the number of particles in the cell i is ni.

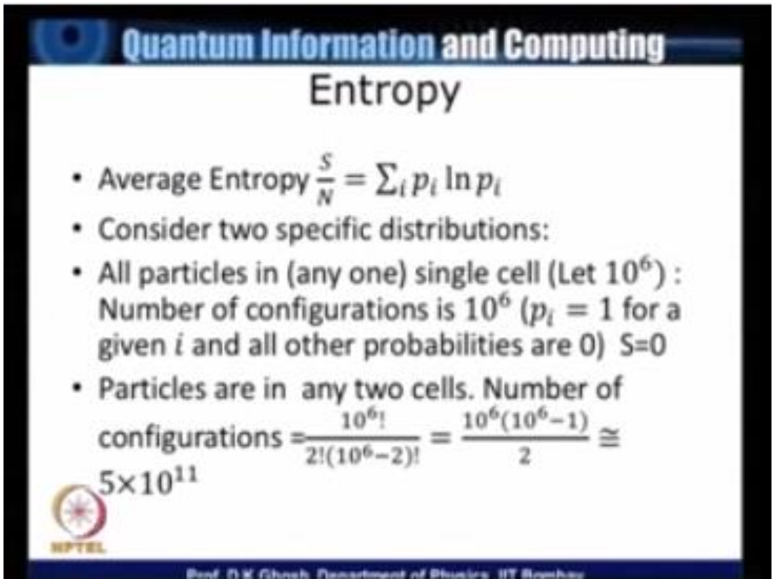
(Refer Slide Time: 09:46)




The probability pi of finding a particular particle in the cell i is given by ni/N, so I will re-write this log W which we had just now seen is given by N log N – sum over i ni log ni in the following way, let us keep this as N log N – sum over i this ni have you realize, this is nothing but n times Pi log (N Pi ) open up the terms there so I have N log N – sum over i (NPi )[log N + log Pi] so notice here log N does not depend upon i and N also comes out of it.

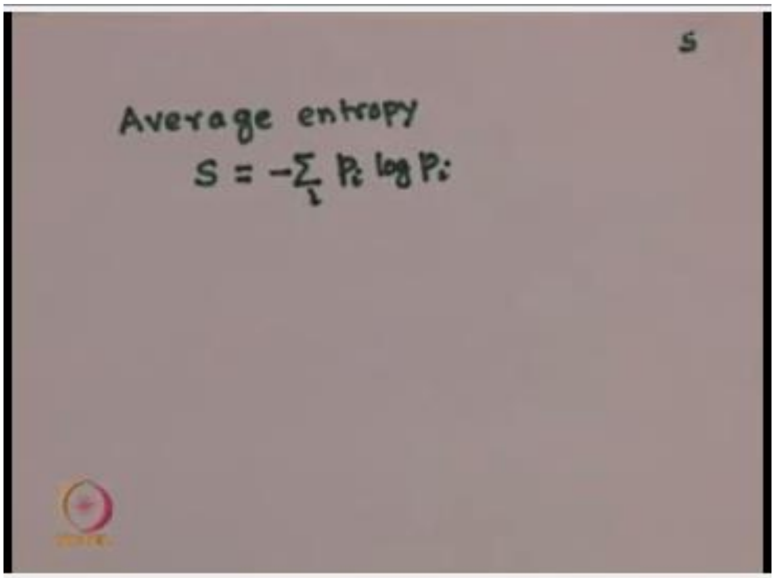
So the first term is –N log N if I take it out sum over i Pi but then that must be equal to 1 because is just a sum of probability minus N sum over i Pi log Pi. So this quantity since sum over i Pi = 1 I can write it as –N sum over i Pi log P.

(Refer Slide Time: 11:40)



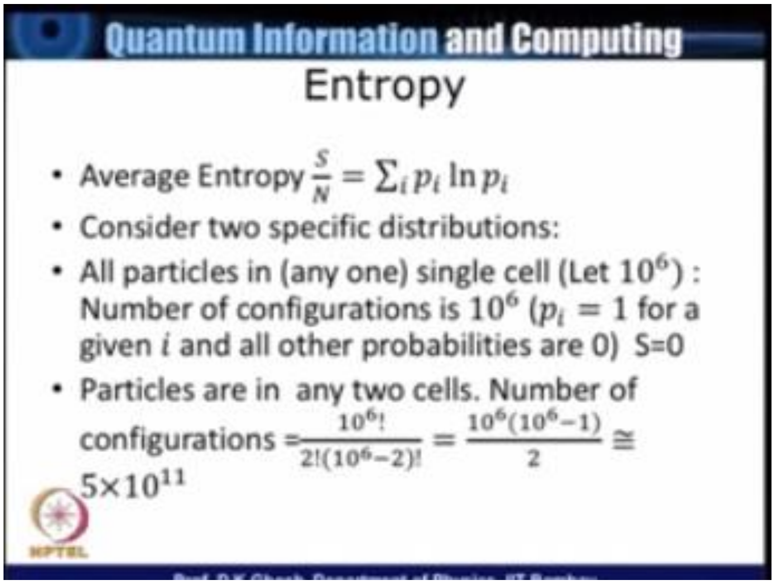

So therefore the average entropy is just obtained by dividing this quantity by N and so average entropy is written as.

(Refer Slide Time: 11:53)



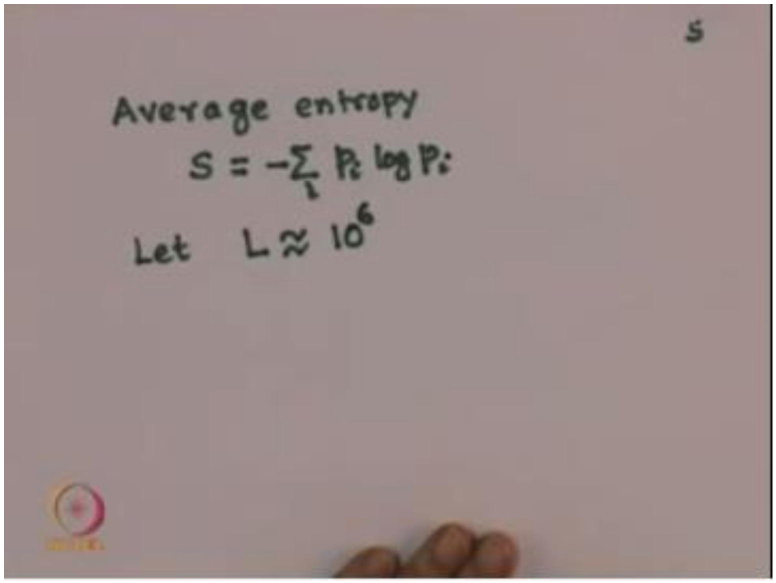

Which is simply obtained by dividing this by N so which is equal to -∑i Pi logPi.

(Refer Slide Time: 12:15)



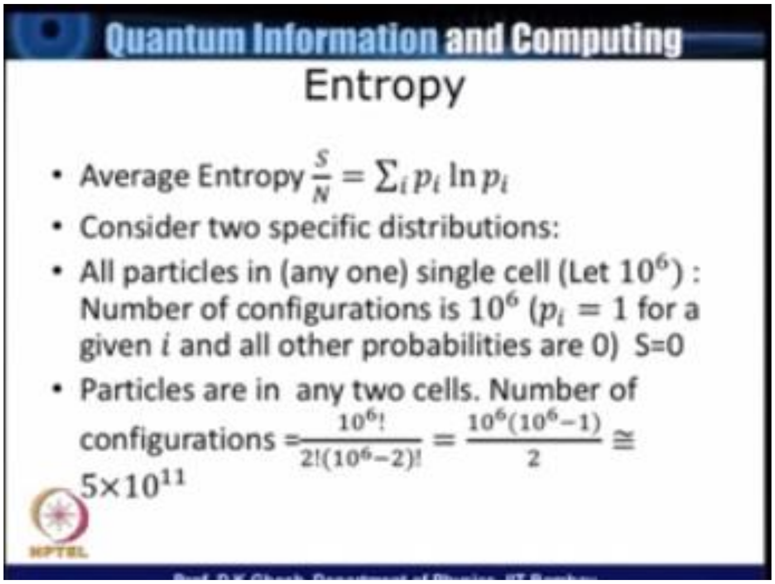

Now to understand the name entropy look at the following situation, let us consider two specific distribution, in the first case let me put all the particles in a single cell for specific calculation let us assume.

(Refer Slide Time: 12:36)



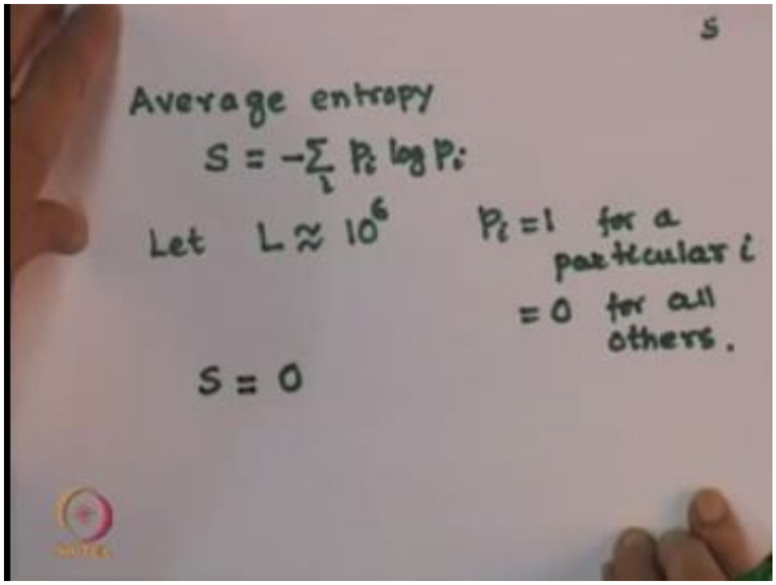

L be of the order of $10^6$ as shown, that I divided the volume into $10^6$ identical surface, so when I say.

(Refer Slide Time: 12:49)



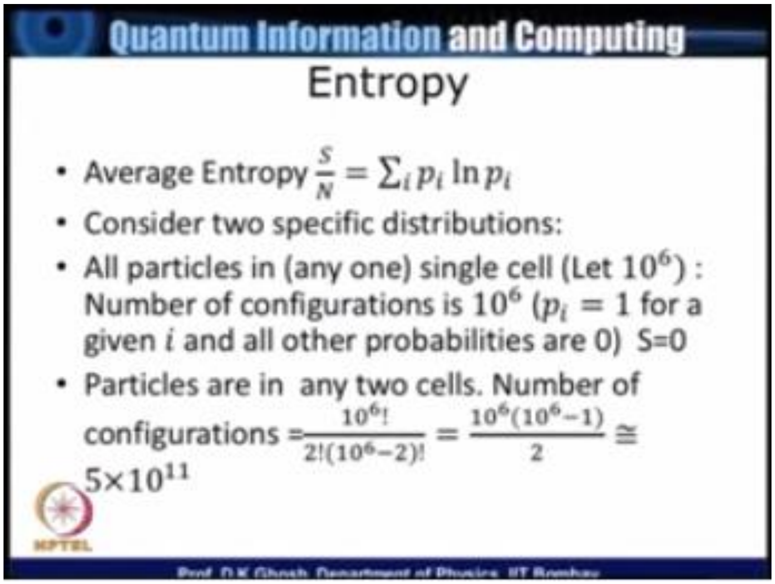

All the particles S and N or R in a given cell what I mean is.

(Refer Slide Time: 12:54)



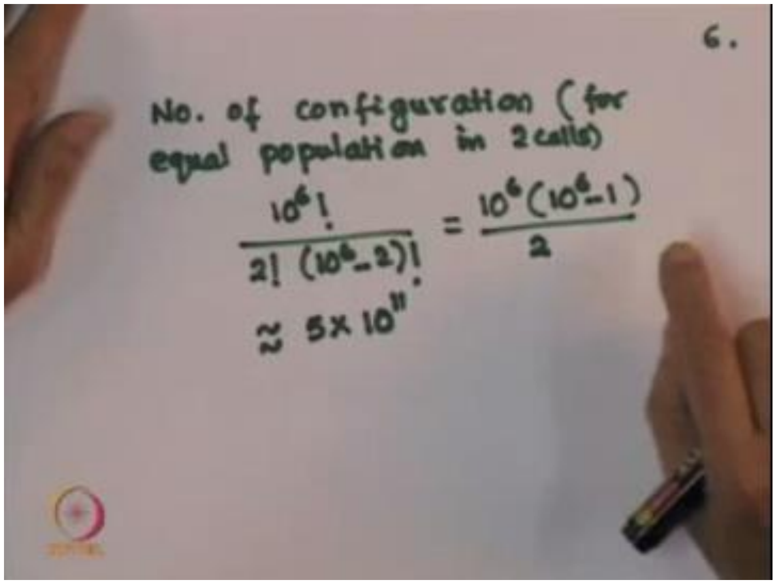

Pi=1 for a given i and is equal to 0 for all others. Now if that happens then my entropy because whether it is 0 log0 which as you know limit of x logx goes to 0 so this would be given by 0 whether as log of 1 is also 0.

(Refer Slide Time: 13:35)



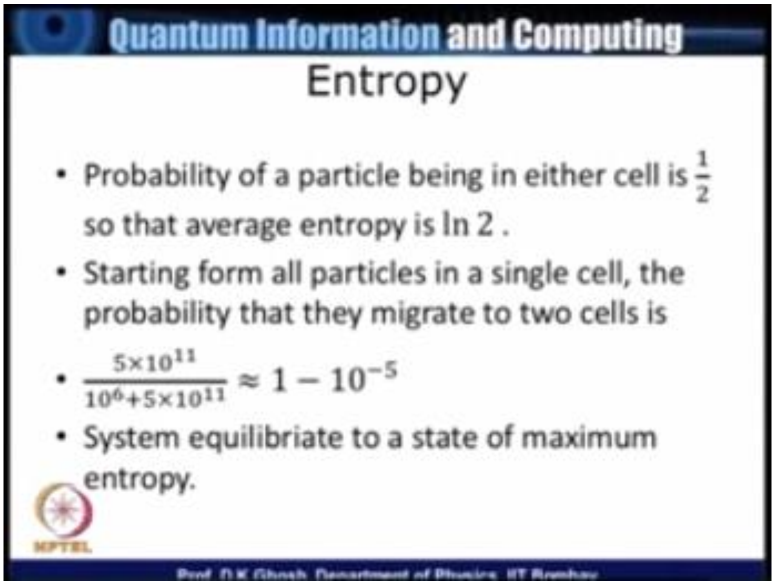

Now let me now talk about a slight a different situation, let us suppose instead of all the particles being in one cell let me say that there are in two different cells equally distributed in two different cells. Now if they are distributed in two different cells out of a possible $10^6$ number of cells the number of configurations.

(Refer Slide Time: 14:03)



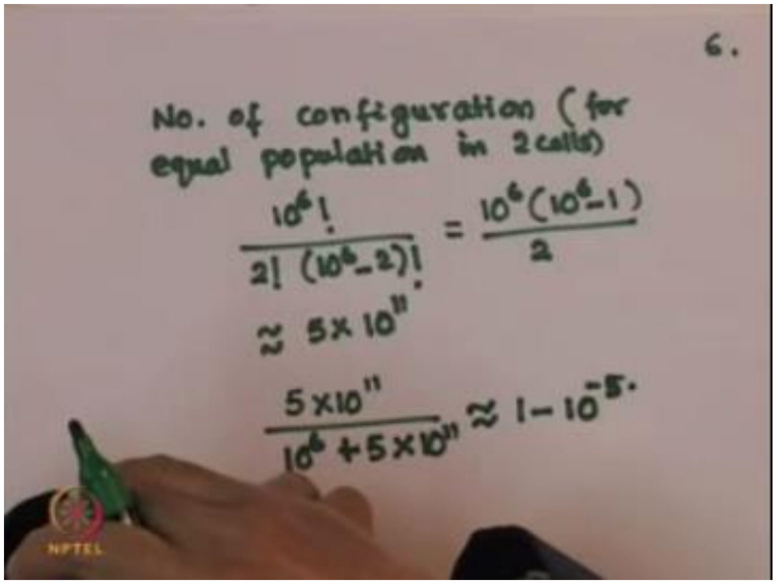

For equal populations in two cells, now that is clearly given by $10^6$ we do which is $10^6!/\ 2!(10^6-2)!$ So we will just multiply these things and you find this nothing but $10^6(10^6-1)/2$, now which is approximately equal to $5\text{x}10^{11}$ this is comes only from this term $10^6(10^6)$ which is $10^{12}/2$ the other term in much smaller compare to $10^{12}$.

(Refer Slide Time: 15:11)



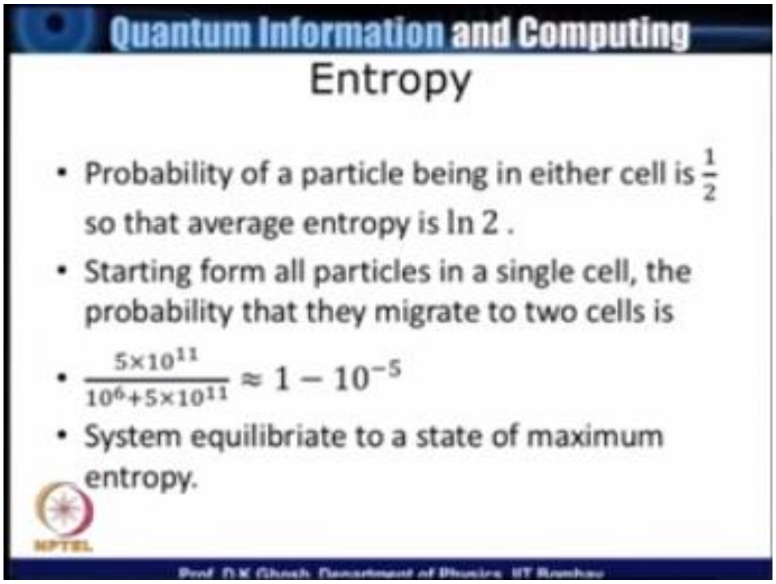

So look at this situation, now in this case the probability of the particle being in either cells if 1/2, so therefore the average entropy is just log2. Now so if I started with all particles in a single cell where my entropy was 0 suppose I migrate to two cells then the probability with which they migrate because of.

(Refer Slide Time: 15:46)




This is given by $5x10^{11}$ divided by the number of configurations in single cell $+5x10^{11}$ again because that is the number there, which is approximately $1\text{-}10^{-5}$. Now look at this the, if I had just two alternatives since I had $10^6$ cells.

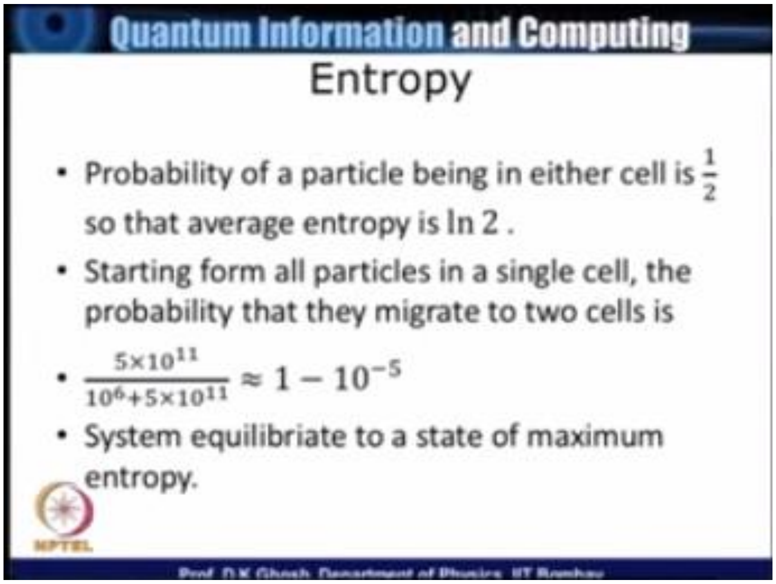
(Refer Slide Time: 16:19)



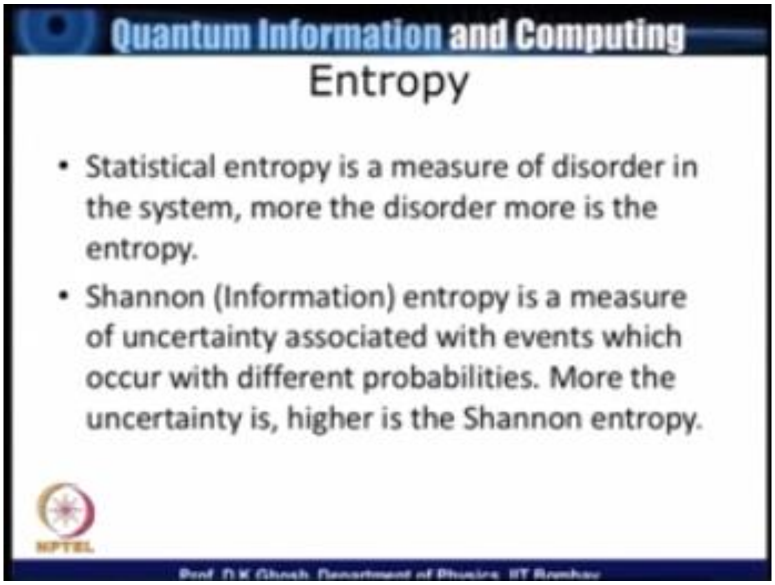

In the first case, so I had essentially the probability to be $10^{-6}$ because I am saying all the particles must go to one particular cell out of the $10^6$. When I relax the condition and said let half of them go to one cell and the other half go to another cell I found that the relative probability is very close to 1. So in other words, the systems tend to equilibriate to a state of maximum entropy.

.

(Refer Slide Time: 16:55)



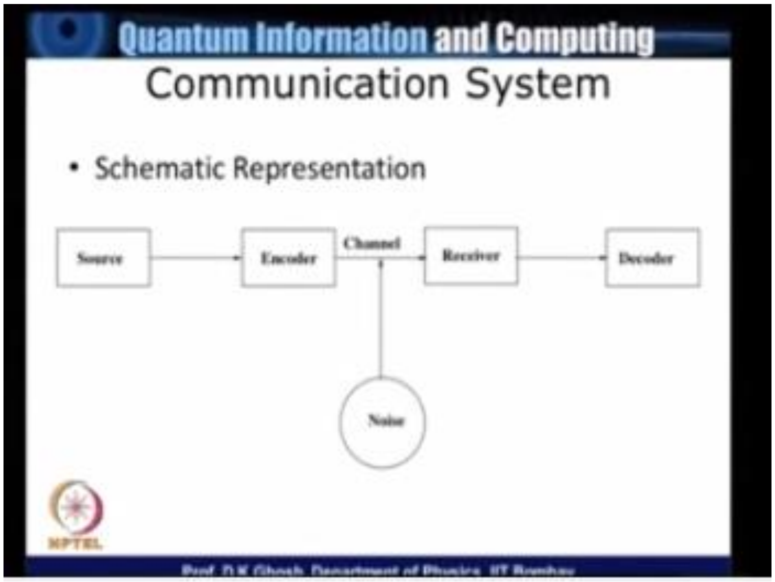
(Refer Slide Time: 16:58)




So therefore, the statistical entropy is a measure of disorder in the system more the disorder more is the entropy. Now let us come back to Shannon entropy, the Shannon entropy is a measure of uncertainty associated with events which occur with different probabilities. More the uncertainty is higher is the Shannon entropy. So you see the one to one corresponds both in terms of physical interpretation and in terms of the expression that we are obtained in statistical physics and for the information theory.

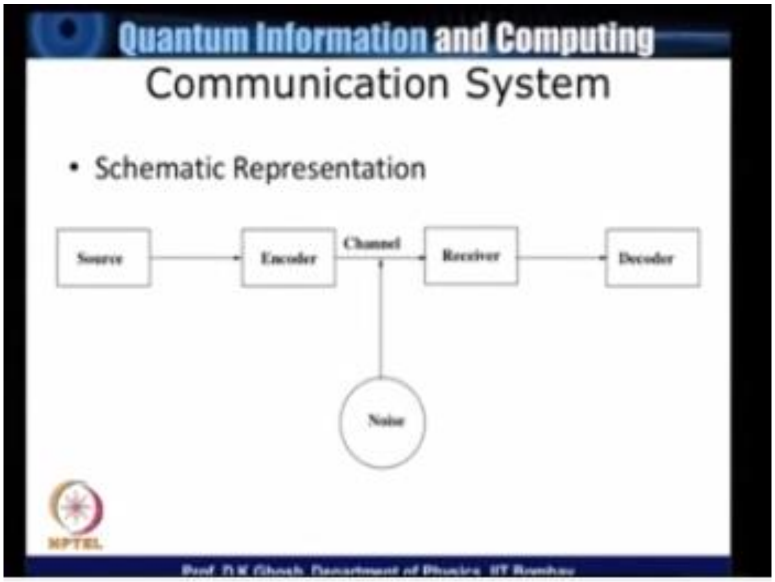
It was Boltzmann who were actually connected the thermionic entropy which was connected with the heat Q and the temperature by suggesting that the constant which comes in front of the logw should be identified with a constant with later on was given the name of which the name of Boltzmann constant so therefore the entropy in statistical physics becomes k log. In this case I do not need to contribute to the statistics or statistical physics or come back with. So therefore, we take the definition of Shannon entropy to be – sum over I pi log pi which it what I indicated by this functions f. With this let me look at what is a typical communications this is schematic diagram.

(Refer Slide Time: 19:00)



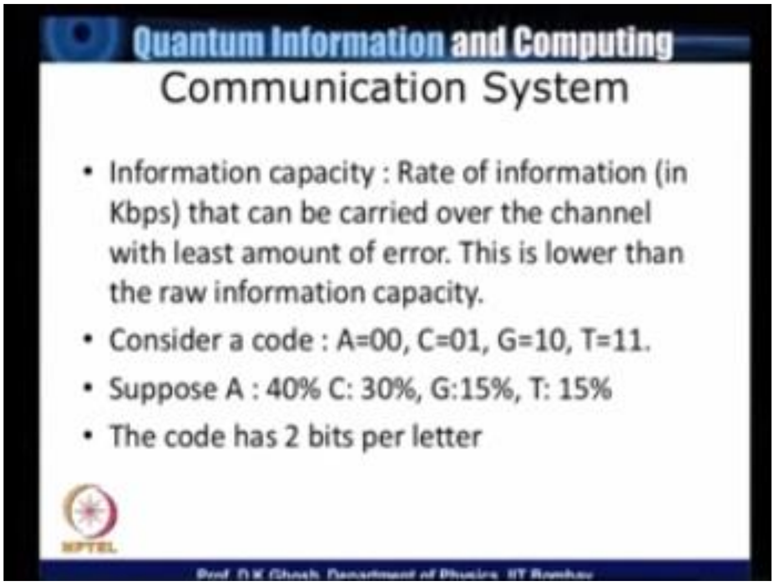

Of the communication system is shown in this slide, so what happen for the communication system is are the source which generates the message that I am trying to communicate but I have to encode it in terms of binary digits so which I do and then over a channel which could be any type of channel which is old style telegraph there was one type of channel now there are fiber optic channel or whatever is your channel, micro channels it would go and receive in a receiver. But then receiver we will receive it in a coded form so he will now have to apply a decoder to get back the origin but on the way what happens is.

(Refer Slide Time: 20:00)



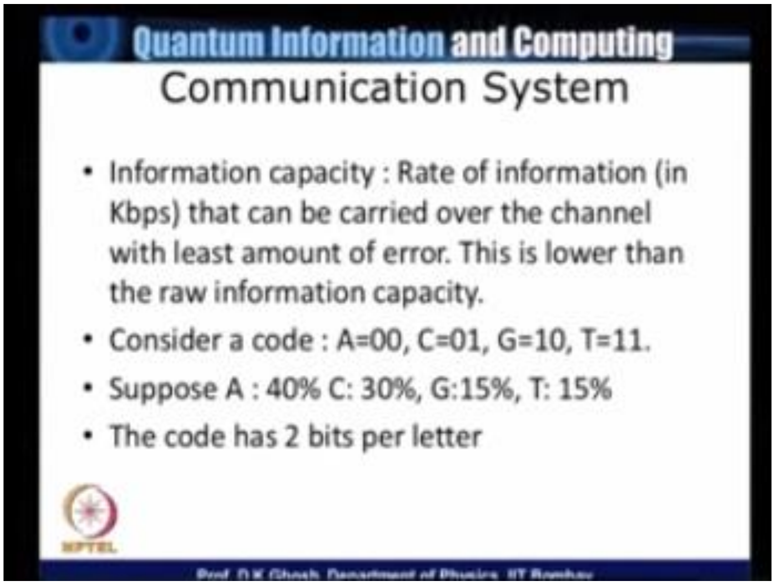

The system picks up, the channel picks up external and this is what we discussed by talking even about the quantum communication and so therefore the surroundings, the super code, the noise been that is being sent with random disturbances and so therefore the job of a communication systems is to sum over there eliminate or minimize such one.

(Refer Slide Time: 20:39)



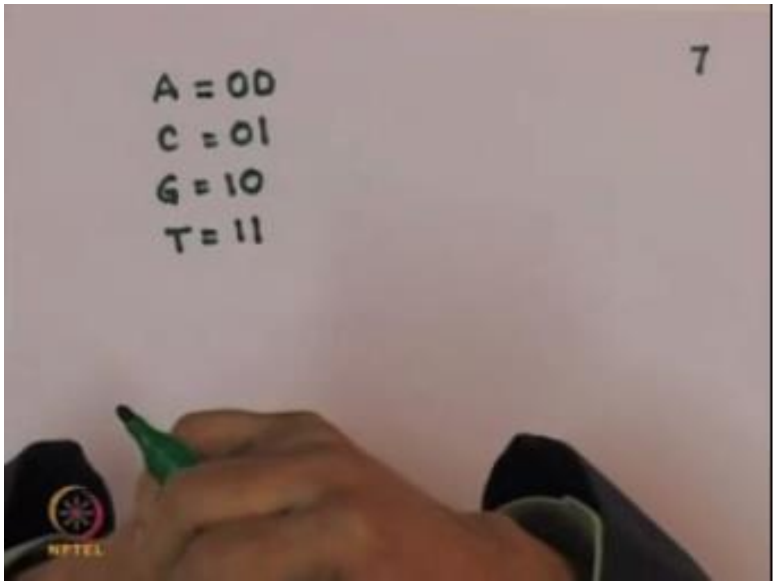

We define the information capacity of a communication system as the rate of information usually measured in kilo bits per second that can be carried over the channel with least amount. I assume for the purpose of this discussion that I am talking about the raw information capacity in practice the real information capacity is lower than the raw information capacity because of the presence of life. Now what type of a code do I have.

(Refer Slide Time: 21:27)



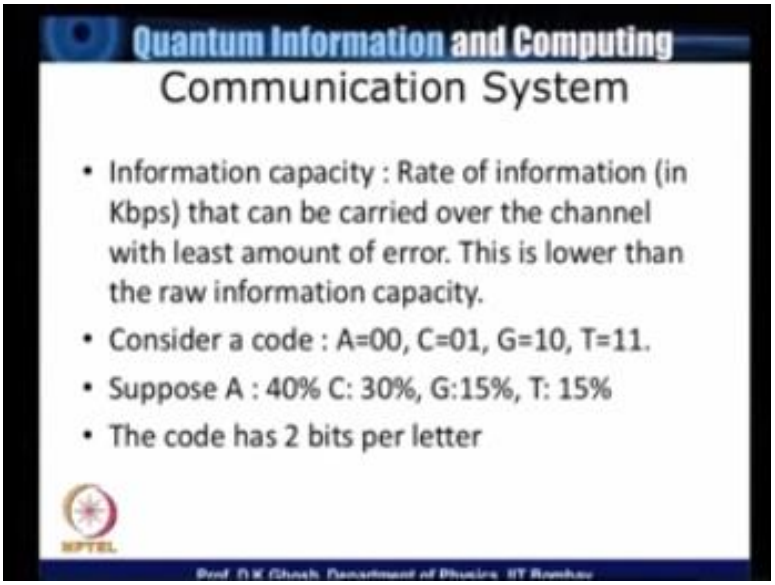

So let us look at a code in which I take these letters.

(Refer Slide Time: 21:31)



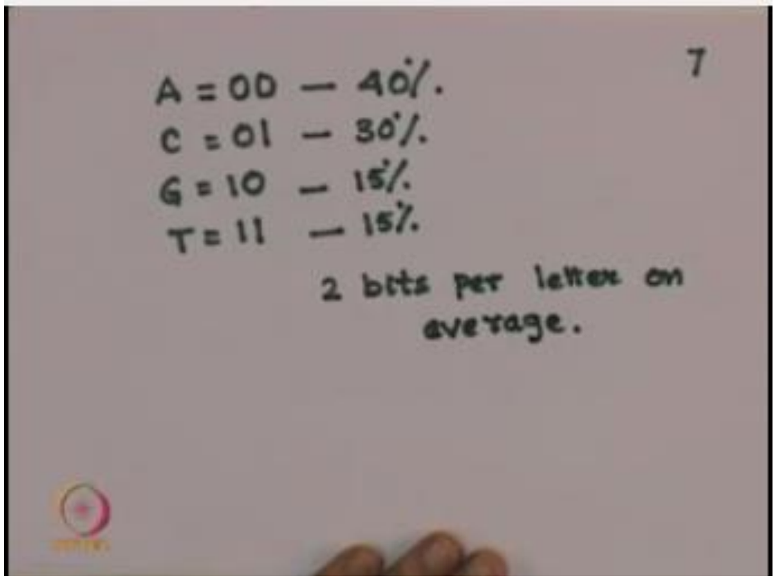

You might identify this letters I said A = 00, C= 01, G = 10, and T = 11 what I am trying to do here, I am trying to send a DNA sequence, I am trying to send a DNA sequence by trying to indicate what is the sequenced with ACGT occur and since I have to send it in a binary form I use this code.

(Refer Slide Time: 22:15)



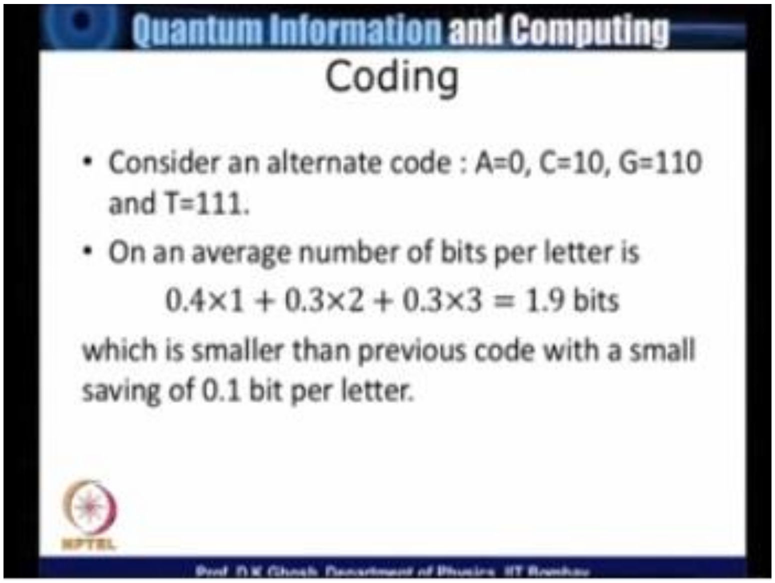

A = 00, C = 01, G = 10, T = 11 now let us suppose that the letter A appears with.

(Refer Slide Time: 22:24)



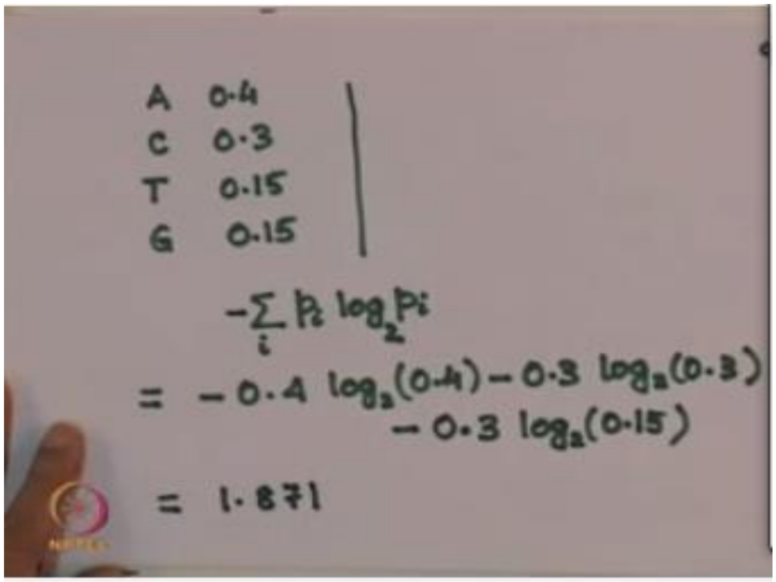

With 40% letter C 30% and each of G and T occurs with 15%. Now since each one of this letter is coded by two bits by average letter or each letter has an average length of two bits so there are two bits for letter on an average.

(Refer Slide Time: 23:11)



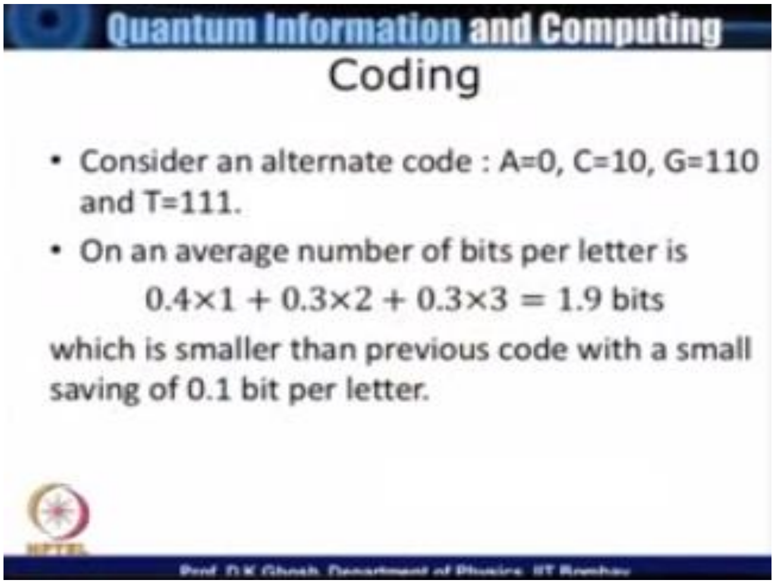

Let us consider alternative model in which what I do is this I take the same probabilities of occurrence but I code A = 0.

(Refer Slide Time: 23:23)



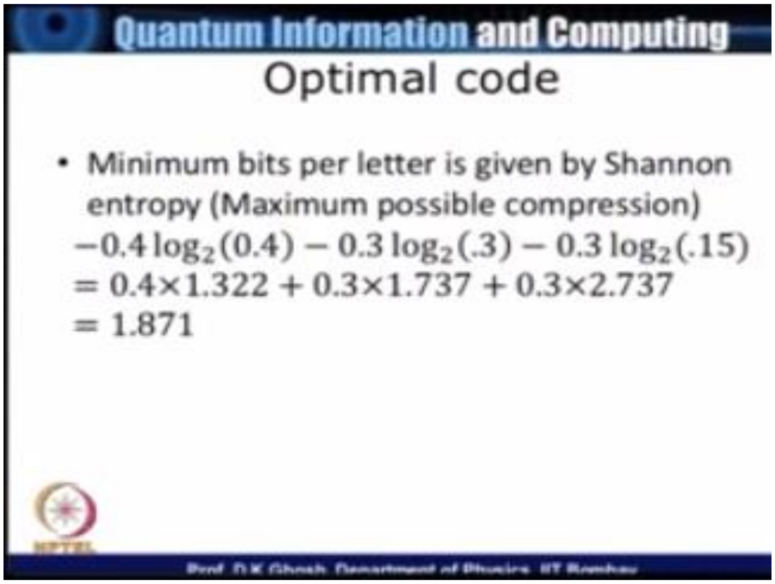

Single bit C = 10, G = 110, and T = 111 and notice I have just taken a abrupt scheme which I have looked up this you recall 40%, this is 30%, 15% and 15%. So if I am to calculate the average number of bits I have it I have 0.4 x 1 because there is just 1 letter then +0.3 x2 letters +0.15 x 3 letters and 0. 15 x 4 letters add it up you find this works out to 1.9 bits per letter. The previous code I have two bits per letter this code I 1.9 bits per letter so I have a same. So in these two examples that we have given we have seen.

(Refer Slide Time: 24:44)




In the first case I had 2 bits per letter in the second case there is 1.9 bits per letter which is the smaller very small at one time you respect the previous one but an advantage never being. The question is what is the optimal code, can I give a limit on what is the maximum compression possible and that is what is given by the Shannon entropy.

(Refer Slide Time: 25:16)

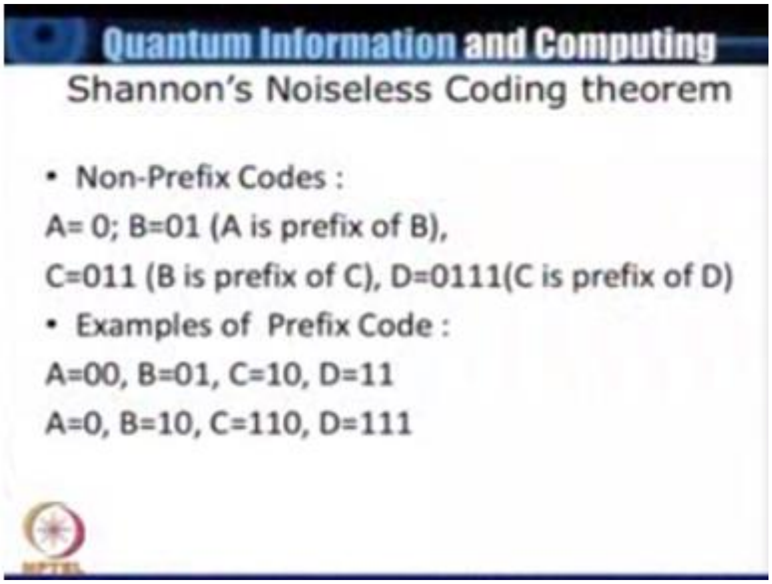



Let us take this example again remember in Shannon entropy I do not have to worry about this specific coding that I am take but I am going to talk about what are the probabilities with which the various events in this case appearance of A C T or G here. So we have said A had a probability of 0.4, C had the probability of 0.3, T had a probability of 0.15 and G had a probability of 0.15 well whether it is CT or TC immaterial for our calculation.

So if I calculate the sum enter of people here is I get $\sum$ I $-p_i$ log $p_i$ and this is logarithm to the base to this is -0.4 log 0.4 -0.3 log 0.3 – there are two terms 0.15 each so I put add them up and write is as 0. 3 again, $\log_2$ 0.15. You can take a calculator and work this out just be a little careful most standard calculator do not have logarithm calculate to the base two that is related to calculator you will find this is equal to 1.871.

This is the maximum compression that is possible for any of the codes that you care to write down that is called Shannon's noiseless coding theorem and only thing is that is applicable for what is known as.

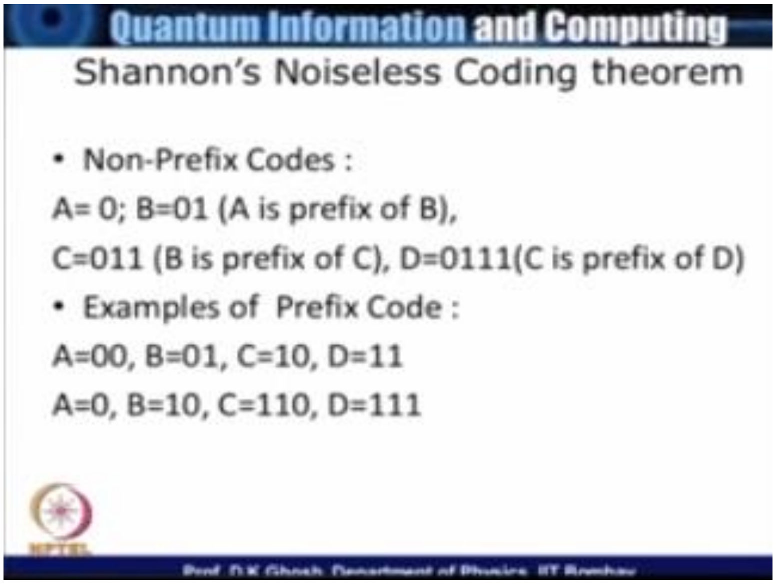
(Refer Slide Time: 27:23)




A non prefix codes this slide shows what is meant by a non prefix code by definition in non prefix code is a code in which the code for given letter is a not a prefix for a second, so the previous case that you considered we said A is 0 B is 01 but look at these 01 the code for A is the prefix to B so that tells me that code and talking about is not a prefix code, the same example for C which is 011 but then B which is 01 is a prefix for C and like as for D.

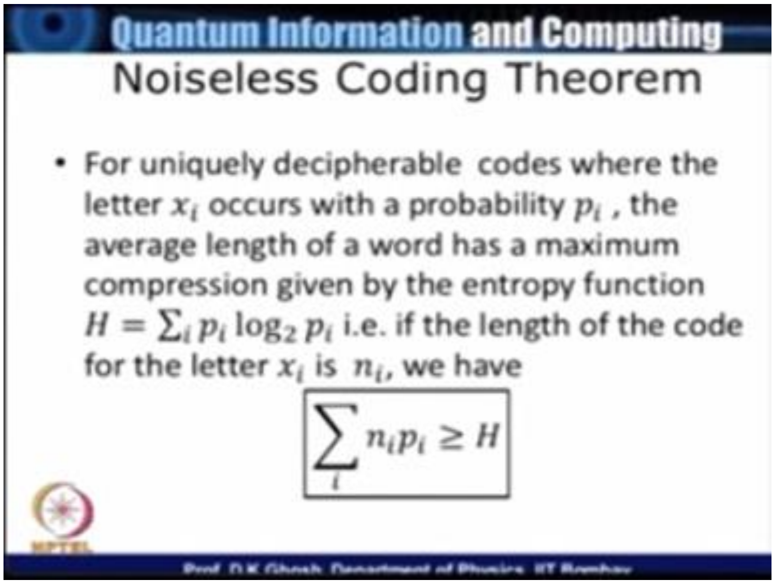
Example the prefix code would be for is to you take A = 00, B = 01, C = 10, D = 11, you notice that none of these is the prefix for the other one or take another one A = 0, which means in none of the letters 0 should come as the first one D = 10.

(Refer Slide Time: 28:45)




So again what I want is 10 should not appear as the first two letters of the reaming letters. So I cannot use C= 101 or 100, C = 110, D = 111, these are examples of what I known as it is called.

(Refer Slide Time: 29:03)




What Shannon's noiseless coding theorem said is the could I that if you consider uniquely decipherable codes where the letter $x_i$ occurs with the probability $P_i$ the average length of a word which word consist of several letters has a maximum compression which is given by the entropy compression. And that is $H = -i -\sum_i P_i \log P_i$ there is a –i missing that slide. And so, therefore suppose I use for the letter $X_i$ a code of length $n_i$ then I must have $\sum_i n_i P_i$ which is the length of my average length of my letters must be greater than the entropy function that we talked about.

And with this I conclude my discussion of classical with this I conclude my discussion of classical information theory. In the next lecture I would go to discuss another entropy known as Ven Neumann Entropy which is a direct extension of the classical Shannon entropy to the case were we consider instead of classical distribution classical ensamples we considered quantum ensamples.

**NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING (NPTEL)**

**NPTEL**
**Principal Investigator**

**IIT Bombay**

Prof. R.K. Shevgaonkar

**Head CDEEP**

Prof. V.M. Gadre

**Producer**

Arun kalwankar

**Online Editor**
**& Digital Video Editor**

Tushar Deshpande

**Digital Video Cameraman**
**& Graphic Designer**

Amin B Shaikh

**Jr. Technical Assistant**

Vijay Kedare

**Teaching Assistants**

Pratik Sathe
Bhargav Sri Venkatesh M.

**Sr. Web Designer**

Bharati Sakpal

**Research Assistant**

Riya Surange

**Sr. Web Designer**

Bharati M. Sarang

**Web Designer**

Nisha Thakur

**Project Attendant**

Ravi Paswan
Vinayak Raut

**NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING (NPTEL)**