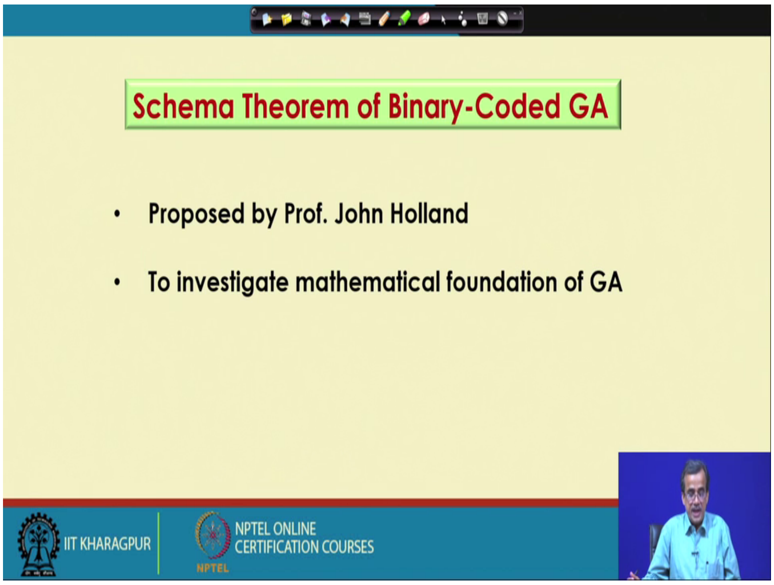
Traditional and Non-Traditional Optimization Tools
Prof. D. K. Pratihar
Department of Mechanical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 09
Schema Theorem of BCGA

Let me start with the schema theorem of binary coded GA and this particular schema theorem is going to indirectly proof the convergence of the binary coded GA. Now, let us see the principle of this particular the schema theorem.

(Refer Slide Time: 00:35)

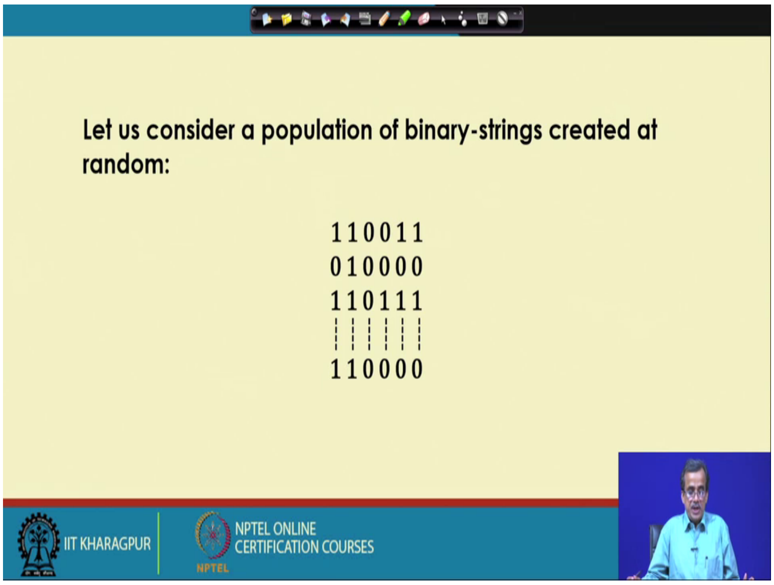


Schema Theorem of Binary-Coded GA

- Proposed by Prof. John Holland
- To investigate mathematical foundation of GA

Now, the concept actually it was given by Professor John Holland who actually introduced to the concept of the binary coded GA. Now, here as I told the purpose is to investigate the mathematical foundation of this particular the binary coded GA.

(Refer Slide Time: 00:58)

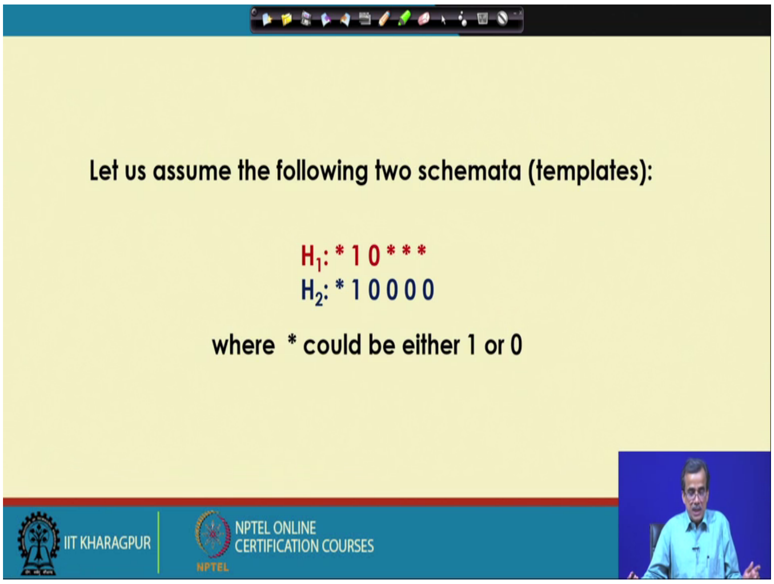

Let us consider a population of binary-strings created at random:

110011
010000
110111
¦ ¦ ¦ ¦ ¦ ¦
110000

Now, to discuss this particular theory let us start with a population of binary strings created at random. Now, supposing that to start with the binary coded GA, so we have created this particular the binary strings at random using the random number generator.

Now, actually what I do is we try to find out the similarity among this particular the binary strings.

(Refer Slide Time: 01:30)


Let us assume the following two schemata (templates):

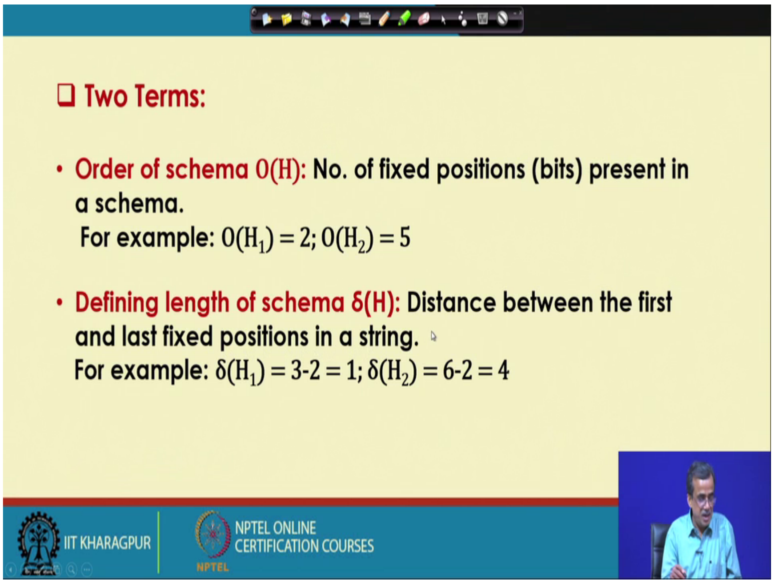
H_1: * 1 0 * * *
H_2: * 1 0 0 0 0

where * could be either 1 or 0

Now, to define or to explain this particular schema theorem we take the help of some schema or the template and schemata is actually the plural of this schema, so schemata means templates. So, here for simplicity I am just going to consider only 2 templates or only 2 schemata H 1 is star 1 0 star star star. Now, star means it could be either 1 or 0 and there are some fix numbers like 1 and 0. So, star as I told it can take either 1 or 0. So, this is 1 schema or 1 template. Similarly there is another schema I am going to consider that is H 2 and that is nothing, but star 1 0 0 0 0. Now, here there is only 1 star. Now, this star could be either 0 or 1 and we have got the remaining 5 fixed bits consisting of ones and 0s.

Now, if I just try to find out that how many strings which are present in this particular population are going to follow a particular template that is H 1 I can find out for example, H 1 says star 1 0 star star star. Now, if I see the population the first string is going to follow 1 0 1 0 1 0 1 0. So, all the strings which I am showing here of course, there are some hidden binary strings here, but all the 4 strings which I am showing here are going to follow, so this particular the first template on the first schema. Similarly that H 2 is nothing, but star 1 1 0 0 0 0 0. So, if I see this particular population. So, it is star 1 0 0 0 0, so if I see this particular think. So, 1 1 0 0 0 0 the second is going to follow third is not going to follow, but the last one is going to follow this particular H 2 just to repeat show all those strings shown here are going to follow that template H 1, but only string second and the last one are going to follow that second schema that is your H 2.

(Refer Slide Time: 04:13)



☐ **Two Terms:**

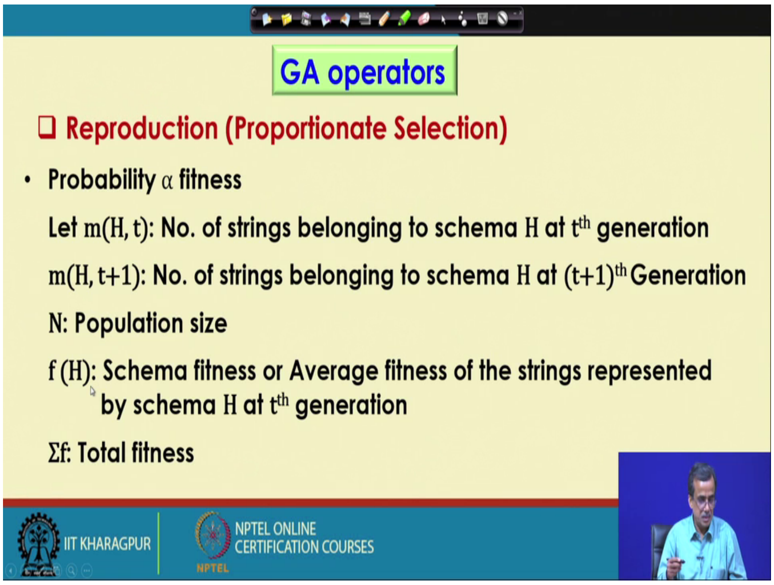
- **Order of schema 0(H):** No. of fixed positions (bits) present in a schema.
  For example: $0(H_1) = 2; 0(H_2) = 5$
- **Defining length of schema δ(H):** Distance between the first and last fixed positions in a string.
  For example: $\delta(H_1) = 3\text{-}2 = 1; \delta(H_2) = 6\text{-}2 = 4$

Now, we define 2 terms one is called the order of is schema that is denoted by O H and that is nothing but number of fixed position or bits present in a schema. Now, if you see this particular H 1, H 1 we have got only 2 fixed bits. So, order of this particular H 1 that is O H 1 is equals to 2 next is your order of H 2 H 2 is 1 2 3 4 5, so there are 5 fixed bits. So, order of this particular the H 2 is nothing, but 5.

And we try to define another term that is called the defining length of a schema that is your delta H and delta H is defined as the distance between the first and the last fixed bit position in a string. Now, if you just see H 1. So, H 1, this is the last fixed bit position and this is the first fixed bit position. So, the defining length is nothing but the 3 minus 2, so 3 minus 2 will be the defining length for one that is nothing, but 3 minus 2 is equals to 1.

Similarly, if I want to find out the defining length for H 2, so the last fixed position is first, second, third, fourth, fifth, sixth position is the last one last fixed bit and the first fixed bit is the second position; that means, 6 minus 2 that is 4 will be the defining length for the second template. So, delta H 1 is equals to 3 minus 2 equals to 1 and delta H 2 is 6 minus 2 that is equals to 4. So, I have define this particular 2 terms order of schema that is O H and defining length of a schema that is nothing, but delta H. And I will carry out this analysis with the help of these two terms that is order of schema H and defining length of schema H.

(Refer Slide Time: 06:34)



GA operators

❑ Reproduction (Proportionate Selection)

- Probability α fitness

Let $m(H, t)$: No. of strings belonging to schema H at t^{th} generation

$m(H, t+1)$: No. of strings belonging to schema H at $(t+1)^{th}$ Generation

N: Population size

f (H): Schema fitness or Average fitness of the strings represented by schema H at t^{th} generation

Σf: Total fitness

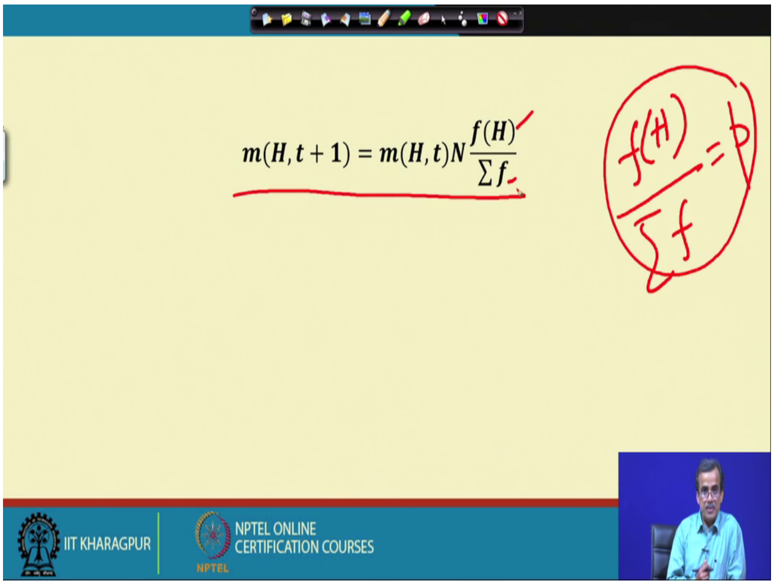
Now, I am just going to concentrate on the GA operators and let me start with the proportionate selection that is nothing but the roulette wheel selection where the probability of selection in the mating pool of a particular GA string is proportional to the fitness. The higher the fitness the more will be the probability of being selected.

Now, supposing that m H t is going to represent number of strings belonging to schema H at t th generation; that means, how many strings are going to follow schema H at t th iteration that is nothing, but m H comma t. Now, our m is to find out what is m H t plus one; that means, how many strings are going to follow the same schema H at t plus 1 th iteration. That means let me take a very simple example supposing that the population size of the GA is say 100 and a particular schema H is followed by say only 5 GA string out of 100 in first iteration. Now, there is a possibility in the second iteration this number 5 may increase to 8 or it may also decrease to 2 then how to increase or how to decrease a number. So, that principle I am going to discuss.

So, N is the population size and. Now, I am just going to define another term that is called f H. So, f H is nothing, but the schema fitness or average fitness of the string. Now, supposing that say 5 strings are going to follow a particular template say schema H.

Now, I can find out the fitness value of all the 5 strings which are going to follow a particular schema we add their fitness values and divide by 5 so that will be the average fitness of the strings which are going to follow the schema H and that is nothing, but the schema fitness denoted by f H and this particular is the total fitness summation f is nothing but the total fitness. Now, I can find out what will happen to the schema H at t plus 1th iteration if I know what is m H t.

(Refer Slide Time: 09:18)



$$m(H,t+1) = m(H,t)N\frac{f(H)}{\sum f}$$

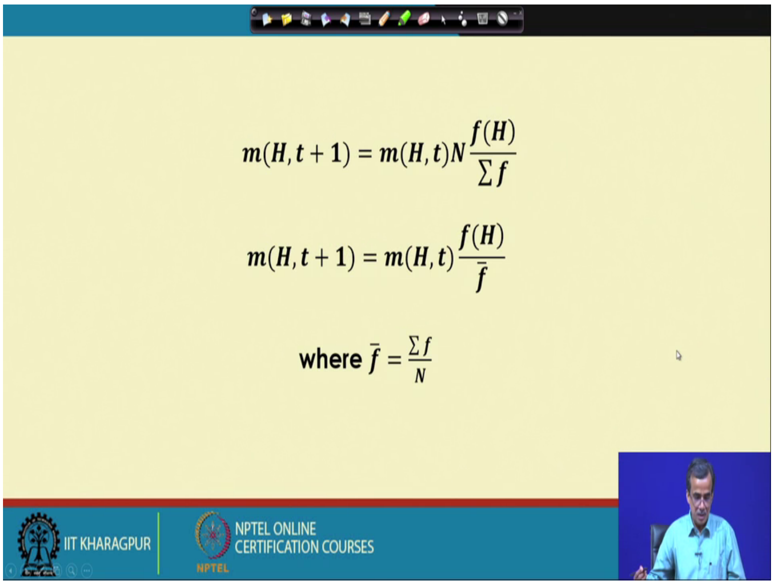
$$\frac{f(H)}{\sum f} = p$$

So, if I know that how many strings are going to follow a particular template at tth iteration and what will happen to this particular schema H in the next iteration that I can find out using this particular expression.

So, m H t plus 1 is nothing, but m H t multiplied by N multiplied by f H divided by summation f. Now, how to arrived at this particular expression? As I told that we are using the principle of proportionate selection. Now, proportionate selection is the probability is proportional to the fitness. Now, we have already calculate calculated what is f H that is the schema fitness and I also know what is the total fitness there is summation f. Then what is the probability according to the proof proportionate selection that this particular schema H will be selected in the mating pool that probability is nothing, but f H divided by the sum of all the fitness values. So, this is the probability that this particular schema H will be selected through this particular the proportionate selection. Now, if this is the probability value. Now, m H t plus 1 is nothing but m H t multiplied by N is the population size multiplied by this particular the probability.

So, using this reproduction scheme, I can find out this particular the expression; that means, what will happen to schema H at t plus 1th iteration if I consider only proportionate selection. Now, let us see what happens if I use the other parameters other GA parameters.

(Refer Slide Time: 11:26)



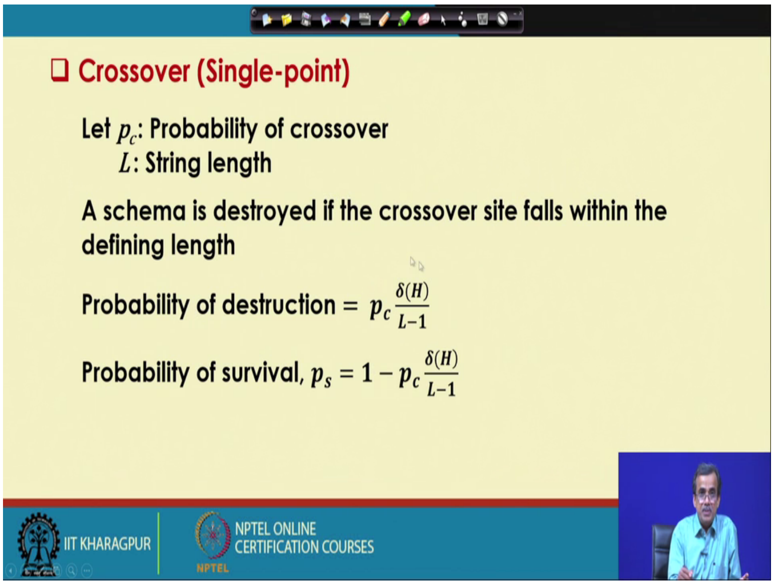
$$m(H,t+1) = m(H,t)N\frac{f(H)}{\sum f}$$

$$m(H,t+1) = m(H,t)\frac{f(H)}{\bar{f}}$$

where $\bar{f} = \frac{\sum f}{N}$

Now, this can be written as like your m H t plus 1 is m H t multiplied by f H divided by f bar because this average fitness f bar is nothing, but the sum of the fitness values divided by the population size. So, from this particular expression I can find out this expression.

(Refer Slide Time: 12:00)



❑ Crossover (Single-point)

Let p_c: Probability of crossover
L: String length

A schema is destroyed if the crossover site falls within the defining length

Probability of destruction = $p_c\frac{\delta(H)}{L-1}$

Probability of survival, $p_s = 1 - p_c\frac{\delta(H)}{L-1}$

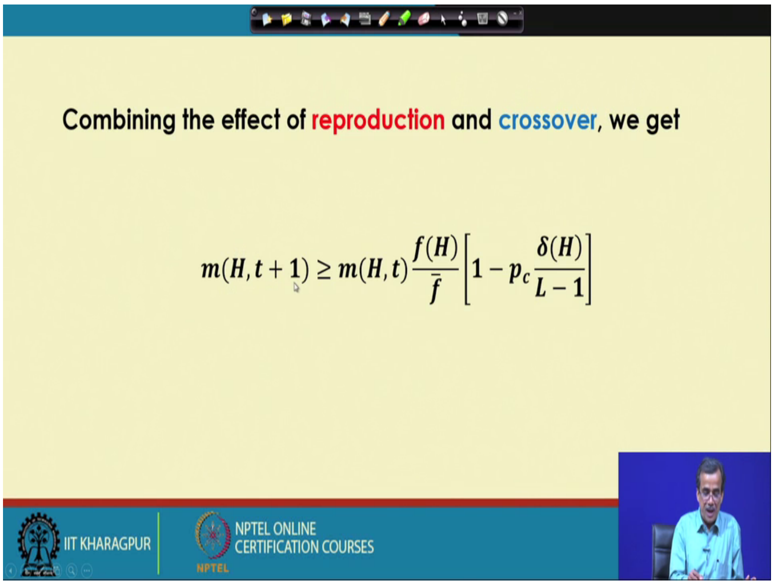
Now, I am just going to concentrate on the crossover and for simplicity let me consider the single point crossover. Now, supposing that p c is the probability of crossover and capital L is the string length. Now, a scheme will be destroyed if the crossover site falls within the defining length.

Now, if you remember the defining length is the difference between the last and the first the fixed bit position. Now, if the crossover site falls within this defining length there is a possibility that this particular template will be lost and if you want to survive, so we will have to take the precaution. Now, what is the probability that this particular schema whose defining length is nothing, but your delta H that is defining length will be selected inside that particular the danger zone that particular probability is nothing, but delta H divided by L minus 1. Why is it so? Now, delta H is a defining length and if the string length is L minus 1. So, I have got L number L minus 1 places for the selection of crossover site. So, the probability that the crossover site will fall within the defining length is nothing, but delta H divided by L minus 1. So, that particular expression I am just going to write here.

So, the probability of destruction is nothing, but delta H divided by L minus 1 multiplied by p c because p c that is the probability of crossover is the another probability that is also a probability of destruction. Now, here there are 2 probability values what is the p c another is the delta H divided by L minus 1 both are probability of destruction, but they are independent and that is why we multiply these 2 probability values just to find out what should be the probability of destruction. Now, if this is the probability of destruction. So, very easily you can find out the probability of survival that is p s is nothing, but 1 minus p c delta H divided by L minus 1.

So, probability of survival is nothing, but 1 minus the probability of destruction. Now, if I just combine this thing along with whatever we got through the reproduction.

(Refer Slide Time: 15:00)

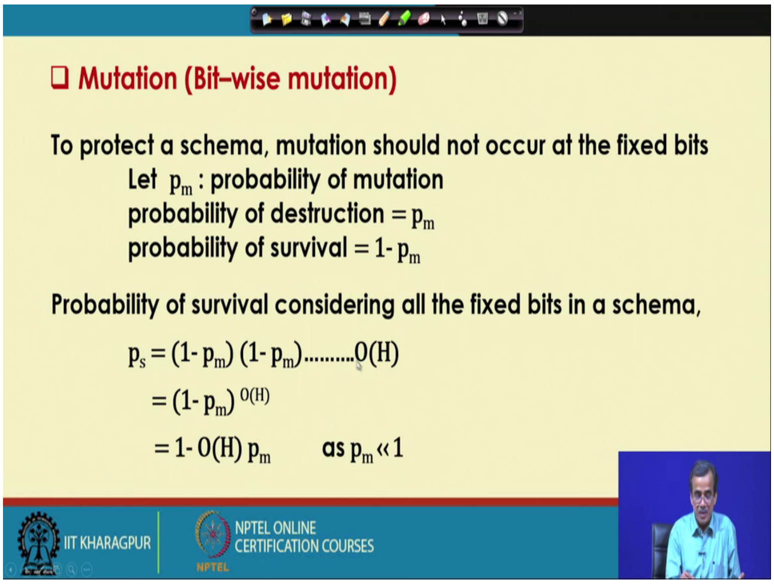


Combining the effect of reproduction and crossover, we get

$$m(H,t+1) \geq m(H,t)\frac{f(H)}{\bar{f}}\left[1 - p_c\frac{\delta(H)}{L-1}\right]$$

That means, if we combine the effect of reproduction and crossover. So, will be getting the updating rule that is m H t plus 1 is greater than equals to m H t multiplied by f H divided by f bar. This part we have already seen multiplied by the probability of survival that is 1 minus p c multiplied by delta H divided by L minus 1. Now, this m H t plus 1 is greater than equals to this particular expression if I combine reproduction and crossover.

(Refer Slide Time: 15:50)



❑ **Mutation (Bit-wise mutation)**

To protect a schema, mutation should not occur at the fixed bits

Let p_m : probability of mutation
probability of destruction = p_m
probability of survival = $1 - p_m$

Probability of survival considering all the fixed bits in a schema,

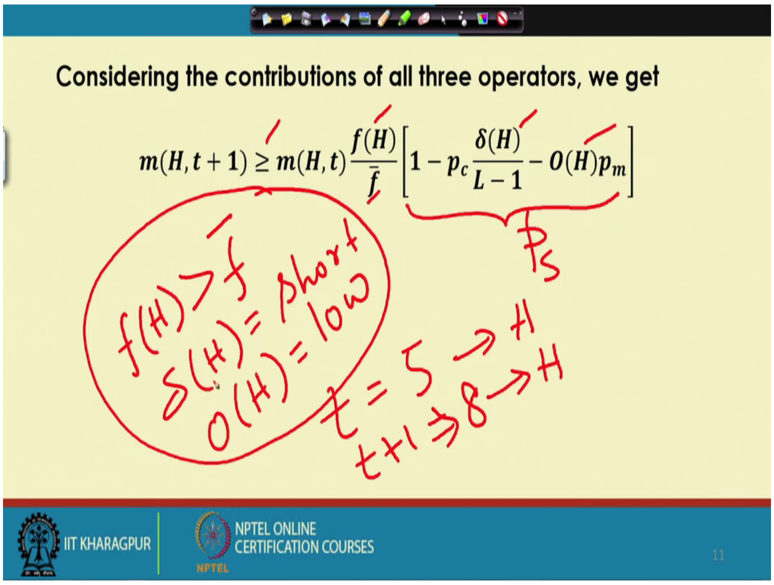
$$p_s = (1-p_m)\,(1-p_m)\ldots\ldots 0(H)$$
$$= (1-p_m)^{0(H)}$$
$$= 1 - 0(H)\,p_m \qquad \text{as } p_m \ll 1$$

Now, let us see if I consider the mutation and here I am just going to consider bitwise mutation. Now, if I consider the bit wise mutation and we can say that to protect the schema mutation should not occur at the fixed bits. Now, the number of fixed bits are denoted by the order of a particular schema; that means, on the fixed bits there should not be any mutation just to preserve that particular the schema.

Let pm be the probability of mutation and this is nothing, but the probability of destruction. Now, if this is the probability of destruction then the probability of survival is nothing, but 1 minus pm. Now, what is the probability of survival? If I consider all the fixed bit position in a schema that is denoted by p s, p s is nothing but 1 minus p m. So, 1 minus pm is actually the probability of survival for the first fixed bit 1 minus pm is the probability of survival for the second fixed bit and similarly we have got order of H. So, many fixed bits. So, this becomes actually 1 minus pm raise to the power order of H.

Now, as pm is very small compared to 1, approximately we can write down it is equal to 1 minus order of H multiplied by pm and this is nothing, but the probability of survival.

(Refer Slide Time: 17:37)



Considering the contributions of all three operators, we get

$$m(H,t+1) \geq m(H,t)\frac{f(H)}{\bar{f}}\underbrace{\left[1 - p_c\frac{\delta(H)}{L-1} - O(H)p_m\right]}_{p_s}$$

$f(H) > \bar{f}$

$\delta(H) =$ short

$O(H) =$ low

$t = 5 \rightarrow H$

$t+1 \Rightarrow 8 \rightarrow H$

Now, if I combine the probability if I combine the effect of reproduction then single point crossover and bitwise mutation, I will be getting this is the rule for updating of this particular the schema. So, the rule for updating is m H t plus 1 is greater than equals to m H t multiplied by f H divided by f bar multiplied by 1 minus p c delta H divided by L minus 1 minus order of H multiplied by pm. Now, this is nothing, but the probability of survival if I consider this crossover as well as mutation. So, this is actually the probability of survival denoted by p s and f H is the schema fitness f bar is the average fitness m H t is the number of GA string going to follow schema H at tth generation and m H t plus 1 is the number of GA string which is going to follow the schema H at t plus 1th iteration.

Now, here you can see that I have put 1 sine greater than equals to not only equal there is a reason behind that there is a reason behind putting this particular the greater than size the reason is as follows. Now, during this GA operation there is also a possibility that a new GA string may also participate or may also follow a particular schema that is schema H, so just to keep that particular possibility intact here I have put greater than equals to sign.

Now, let us see what does it mean what is the conclusion from this particular the expression we will be take very hypothetical situation. Supposing that f H is greater than f bar that is the schema fitness is more than the average fitness say delta H that is the defining length is short defining length then comes O H that is the order of schema H that is say low let us see what happens. If I consider the schema fitness is more than the average fitness then the defining length that is this is short and order of H is low so what will happen, the probability of survival is going to be more. And moreover f H is found to be greater than f bar so there is a possibility we will be getting a higher value of this particular m H t plus 1 compared to your m H t; that means, at tth iteration if I find out that only 5 bits are going to follow schema H there is every possibility at t plus 1 th iteration there is a possibility that might be 8 strings are going to follow scheme H, provided so f H is found to be greater than f bar delta H is short and order is low.

Now, in sentence we can say that if I can find out a schema whose average fitness is more than the average fitness of the population and if it is having short defining length and low order, that particular schema is a very good schema and that is going to receive more and more number of copies and more and more number of GA strings are going to follow that particular schema.

Now, in other words we can we can say that the schema which is having this particular property is a very good schema and which is going to make this particular search which will be that work that particular very good schema will be the building block of this particular the GA search. This is almost similar to the situation supposing that we want to construct a very strong building and to construct a strong building the quality of the bricks is very important and if you get very good quality bricks, so good quality bricks are nothing, but the good quality templates or the schema. And if I can find out good quality schema that will receive more and more attention and its number is going to increase. And on the other hand if I get a bad schema there is a possibility the number of GA string going to follow schema H is going to be reduced with the number of iteration and ultimately that particular bad schema may be deleted from this particular the population.

Now that means the fact which I am going to reveal that if we get a very good schema, we will be getting more and more copies; that means, as the GA iteration proceeds there is a possibility that a majority portion of this particular population will be occupied by the very good schema and there is a possibility that if GA is running properly, so he will be getting the multiple optimal solution. Now, there actually the fitness values will be the same, but that particular fitness the value we can reach with different combination of the design variables.

Now, this is another very big advantage of the GA that will be getting the multiple optimal solutions. Now, out of this multiple optimal solution the user can select any one. Now, just to summarize, I can say that, this particular schema theorem is an indirect proof, mathematical proof for the convergence of this particular the binary coded GA.

Thank you.