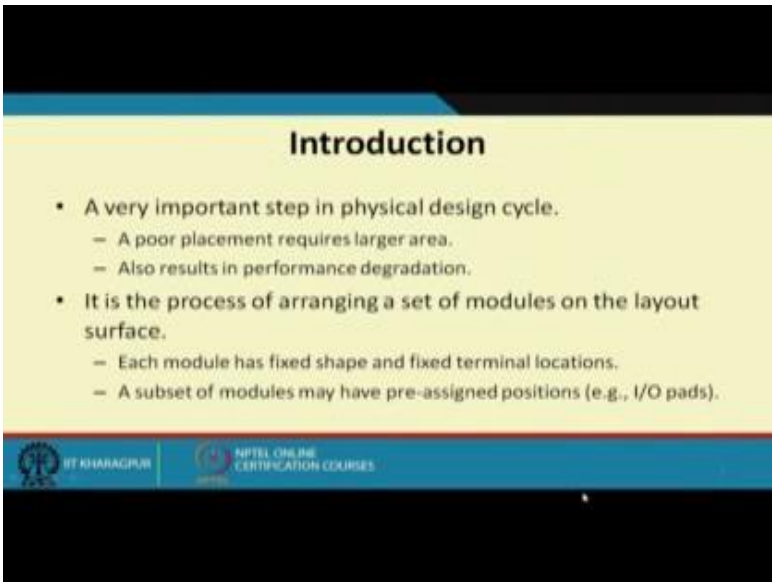
VLSI Physical Design
Prof. Indranil Sengupta
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture -11
Placement (Part I)

So, we now consider the problem of placement; now in the sequence of steps in the physical design automation cycle partitioning is followed by floor planning, with pin assignment and then comes placement. Well, means although we said that was certain design styles likewise standard cells or gate arise, floor planning and placement are not much different they can be merged together, but for full custom style we have to follow on this steps. So, we start our discussion on the placement problem in this lecture.

(Refer Slide Time: 01:08)



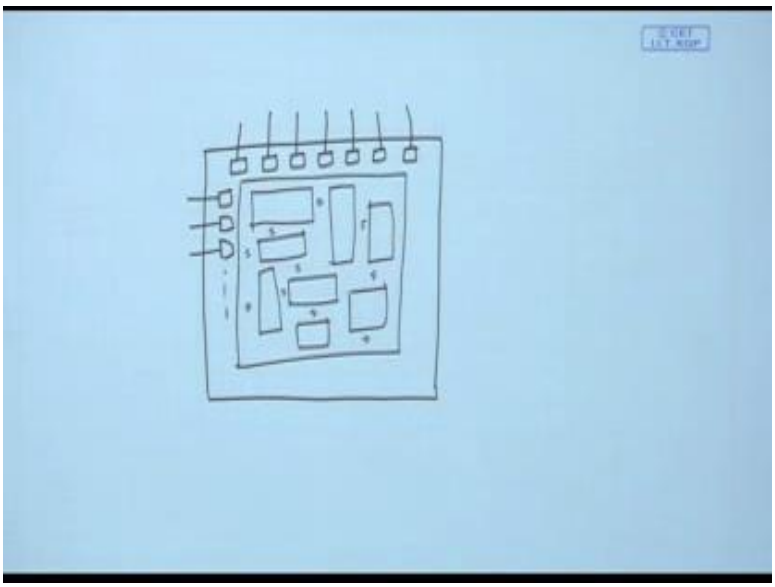
So, as I said this is one of the most importance steps, because it incurs a very large you can say impact on the overall performance and cost of the chip, because you see in today's chip you know that the density of the components, the number of components you can pack in a chip is increasing day by day, today you can pack billions of transistors in a chip.

But if you look at this standard layout of these circuits inside a chip you will find that interconnections occupy sizable percentage of the total area, because of this memory

chips are often very much more dense as compared to standard logic or process a chips; because in memory cells the layout is very regular interconnections are very crisp. So, if your placement is not good, your interconnection problem we interconnection cost can go up, that might increase to a I mean increase the cost of the interconnections, the cost of the chip as a result. So, placement problem is important, if your placement is poor the area requirement can be large and if the area is large some of the critical parts can also increase in the lengths, which may result in degradation of the performance.

So, essentially placement says; how to arrange set of modules on the layout surface? At this stage we assume that the modules has fixed shapes and fixed pin locations. So, the steps of floor planning and pin assignments of already been done, now you have modules that are having fixed shape and fixed locations will however, there are some modules which have pre assigned locations like I O pads.

(Refer Slide Time: 03:35)



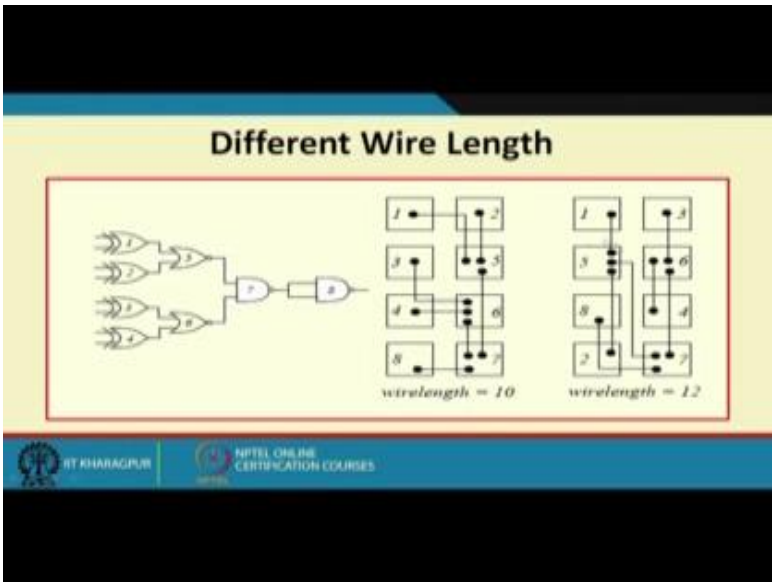
Like you see I am just showing you suppose this is the typical layout of a chip. So, for the pins are usually nowadays taken from all 4 sides, so the input output pads are dedicated circuitry that are used to drive the input and output signal lines.

Now, this I O pads are located adjacent to the external pins. Now because the pins of this chip will be fixed, the location of the I O pads will also be fixed. So, these are some blocks whose positions are pre assigned, but within the layout area that is

available to it in between here within, here you can place all your blocks. Well in the full custom size the blocks can be of any shapes and sizes all possible shapes and sizes you can have. So, you can have something like this also.

So, this is what the placement problem talks about. Now in addition as I said you can also add some space in between for interconnections, which means you need to know how much space is required or what is the complexity of the interconnections.

(Refer Slide Time: 05:01)

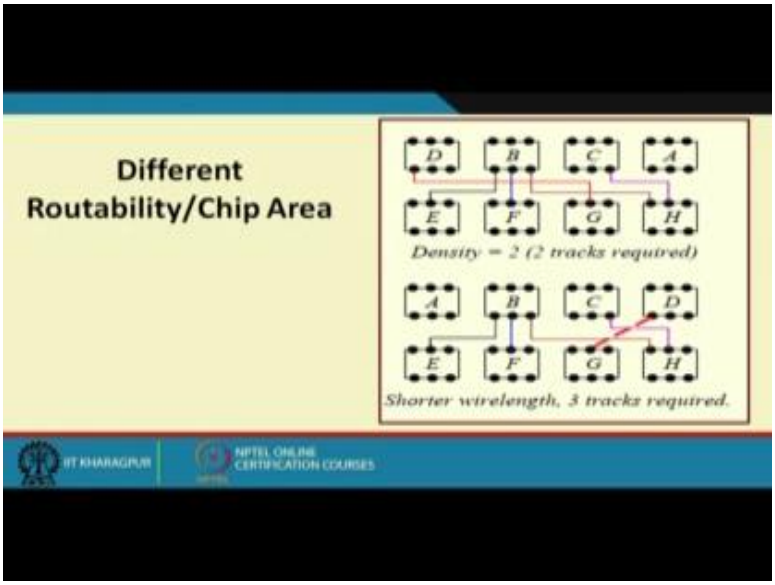


Let us take an example here, which shows you the impact of wire length with respect to 2 possible placements. While these are very simple example which shows 8 gates and let say each gate represents or module. So, there are 8 modules, suppose if I place the modules like this, this numbers indicate the gate numbers. So, you see 1 is connected to 5. So, 1 has to be connect this, 5 has 3 connections.

So, 1 is connected to here, 2 is connected to 5 and 5 is connected to 7. So, for interconnections there are 3 pins; so these are shown by solid dots, the interconnections. So, similarly say 3 is connected to 6, 4 is connected to 6, and 6 is connected to 7 and here 7 is connected to 8 this one. So, if you just calculate the wire length here, it comes to 10. So, how did I calculate? You see I am assuming that adjacent cells the cost is 1. So, the cost of this metal will be 1 plus 1 2 3 4 5 6 7 8 9 and 10 cause becomes 10.

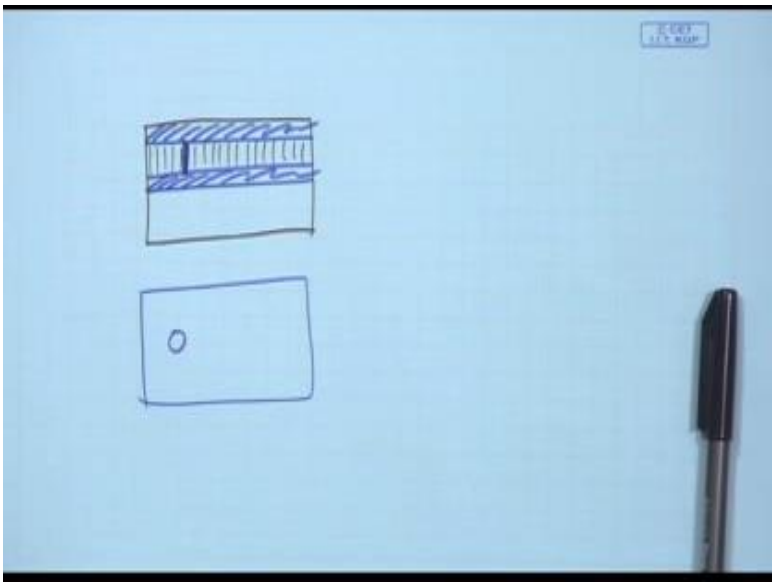
Suppose my placement was like this where block 5 is here, 6 is here, 8 is here, 4 2 7 like this. So, the interconnections corresponding to my net list will be like this. So, now, if you calculate the wire length it will be 1 2 3 4 5 6 7 8 9 10 11 and I think 12. So, it becomes 12. So, depending on the placement your wire length might increase. So, one of the main objectives of placement may be to reduce the overall wire length, because one intuitive argument you can give that if I minimize the overall wire length it is expected that the performance of my circuit will also improve well. It is true most of the time that, but there can be exception situations, your wire length might be reducing, but your critical length might be becoming longer that may be an extreme case all right.

(Refer Slide Time: 07:48)



Let us take with respect to routability. So, here we are considering standard cell kind of a placement, but we are just counting how many tracks we are needing for interconnection. So, again the same example there 8 blocks, some interconnections are there, suppose my connections are like this from D block I have a connection to the G block from B to E, B to F, B to H, C to H. So, if I have a connection like this then as you can see all my connections well here one assumption I am making and I am consisting with that assumption that when in a channel this is like channel routing. So, whenever I am routing the connections, I am assuming that there will be 2 layers of metal like let me explain.

(Refer Slide Time: 08:50)



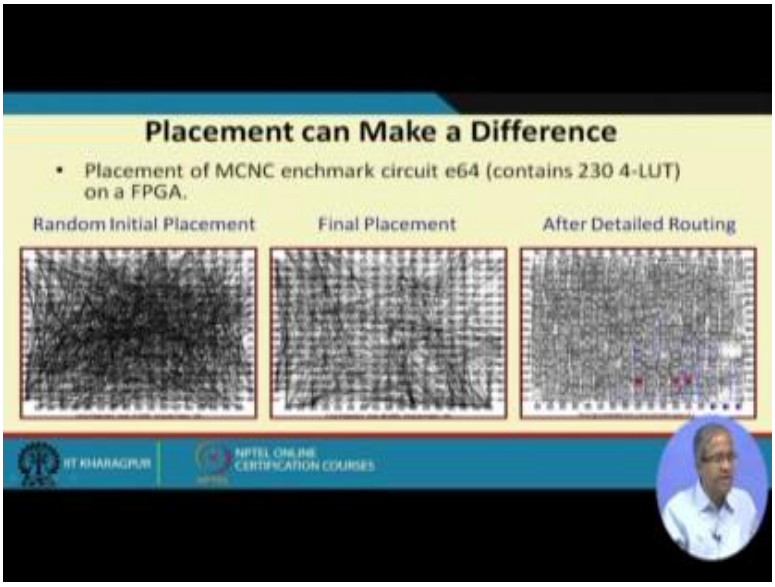
So, if I have a cross sectional view. So, in a cross sectional view this can be one metal surface, this can be another metal surface and there will be some kind of a insulating material in between. So, they are not touching each other right they are on 2 different layers. So, the 2 metal layers or an 2 different list. So, as a matter of convention I can run all horizontal lines on one of the layers and all vertical lines on the other layer and whenever I need to make a connection, I do something called a wire connection wire contact; wire contact to look like this whenever I need from a this layer to this layer, I drill of also now I am showing a top view, in a top view so whenever I need I drill a whole, I put metal between these 2 less. So, that there is a connection like this, right.

So, in this diagram if you see a follow the same convention, vertical and horizontal lines are in 2 different layers. So, that even if some lines are crossing there is no short circuit, because the vertical lines are on one layer, horizontal line is on a different layer the connections are only at the junction, so wherever the horizontal and vertical segments are meeting each other. Now in this solution you can see that I need only 2 tracks, for routing the total routing area you can measure in terms of tracks I need 2 tracks, but if I move them around in a slightly different way, so you can see that now my wire length is obviously shorter, here my wire length so longer, some of the wires is pity longer, but here my wire length is; obviously, shorter, but I already laid out these on 2 layers, but to connect this third line I need third layer, I cannot do it on 2

layers otherwise they will be a short circuit here right. I need another third layer to connect them.

So, this wire length and the number of tracks these are actually some kind of a conflicting requirement, wire length does not always necessary mean that your total area of routing shorter. Trucks do mean that, reducing tracks means shorter area for layout, but increasing wires wire length may mean and increase in delay. So, you have to make a balance between the 2, all right.

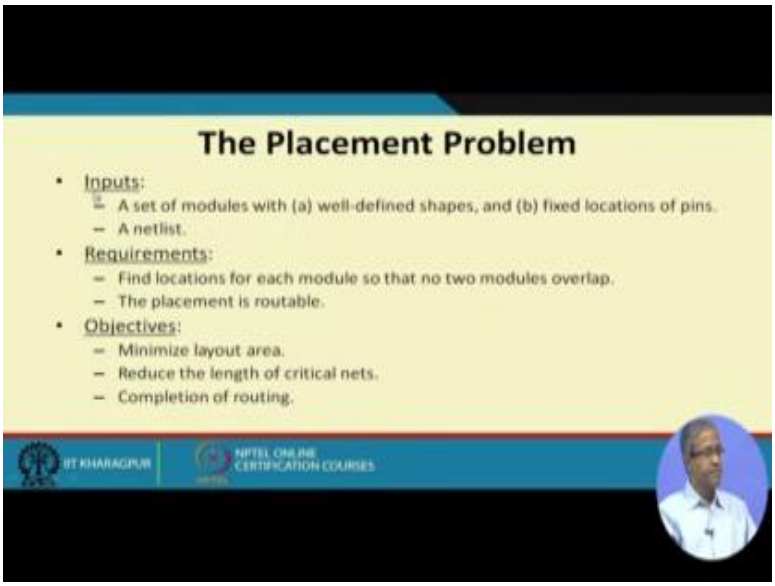
(Refer Slide Time: 11:43)



Now, this is a interesting diagram; this is set of diagrams which are for one example circuit, this is actually bench mark circuit this b is missing, this is called e 64. This e 64 was map to an FPGA where 230 look up tables where used.

So, after mapping the connections where drawn a straight lines and this is that mesh diagrams zigzag diagram which looks really ugly; this was with respect to random initial placement, but after you have done a good placement if you again draw it you see that it looks much neater, most of the connections are local not very not from one corner of the chip to the other and after detail routing. So, this is only by drawing straight lines for after detail routing, the connections look like this it is much cleaner; just an example that a good placement can lead to a much neater network, much shorter wire lengths.

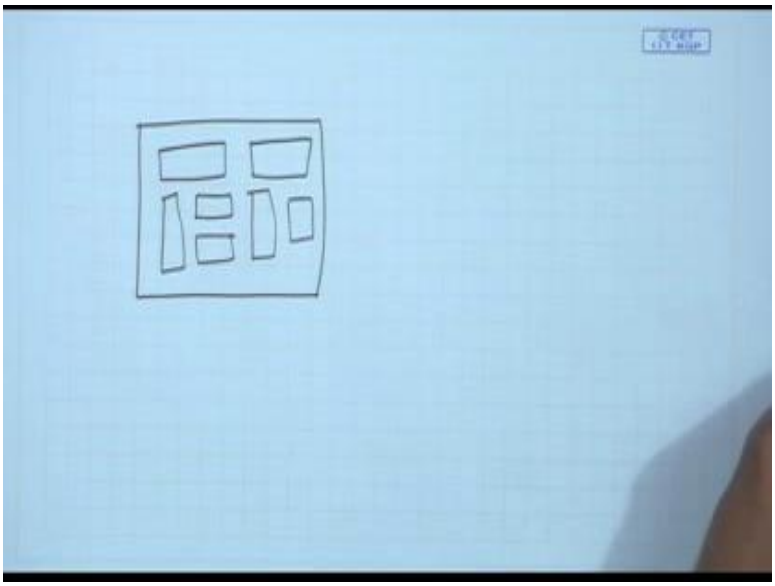
(Refer Slide Time: 12:58)



Now coming to the placement problem, so the inputs is a set of modules with well define shape. So, here none of them are flexible like in floor planning we had said that some of the blocks are flexible in terms of their height and width, but the area was fixed, but here it is nothing like that.

Every block has fixed height and fixed width everything has been finalized, and also the modules have fixed location of pins; when addition you have a net list which tells you that which pin is getting connected to which other pin that is the net list. Requirement first one is obvious, you have to place all the modules on the silicon floor determine locations; obviously, the 2 modules must not overlap. Secondly, the placement is routable.

(Refer Slide Time: 14:08)

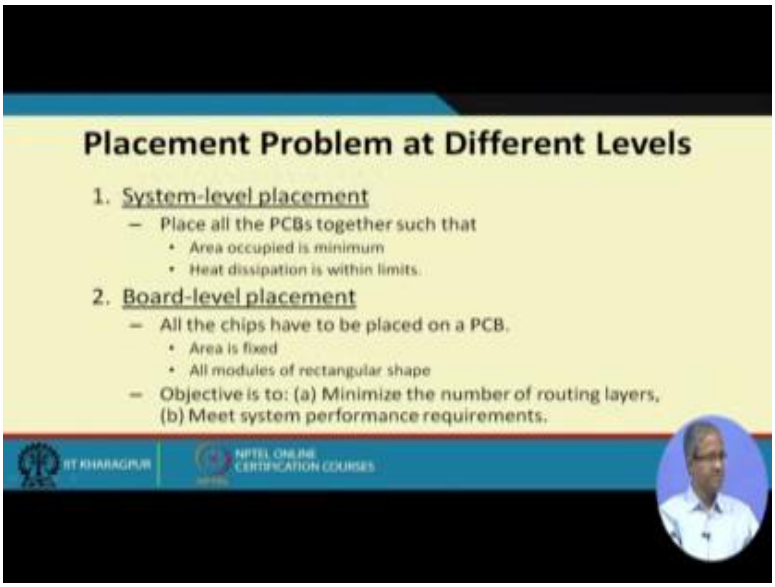


So, what is meant by routable? It means that when you place the blocks, you should keep sufficient space in between. You should keep sufficient space in between so that you will interconnections can be made. So, there has to be space between the blocks is unlike floor planning where you are showing that the blocks where touching each other, but in placement we are not showing the blocks is touching each other were also keeping some adequate space in between, through which the interconnections can be laid out.

So, the objectives can be multiple, minimizing layout area is of course one, for high performance circuits reducing the delay; that means, the length of the critical nets is important. In the third one is very important problem, you see you can have a very good placement, but during this subsequent step of routing you may find that some of the nets you are unable to rout, because you do not have adequate space to draw the lines. So, the placement problem also should ensure that routing completion is guaranteed, but we will see may at this stage this completion of routing guarantee of an cannot be given all the time, but you can always make a good guess, you can keep sufficient space so that I means you if you feel that routing can be done. But what might happen later on is that, you may find that you may have to move some blocks around.

So, after routing you will see that some of the nets you are not able to rout, you can again come back to placement. Move some cells around make more space go back and then complete the routing; those are some techniques which people do use.

(Refer Slide Time: 16:14)



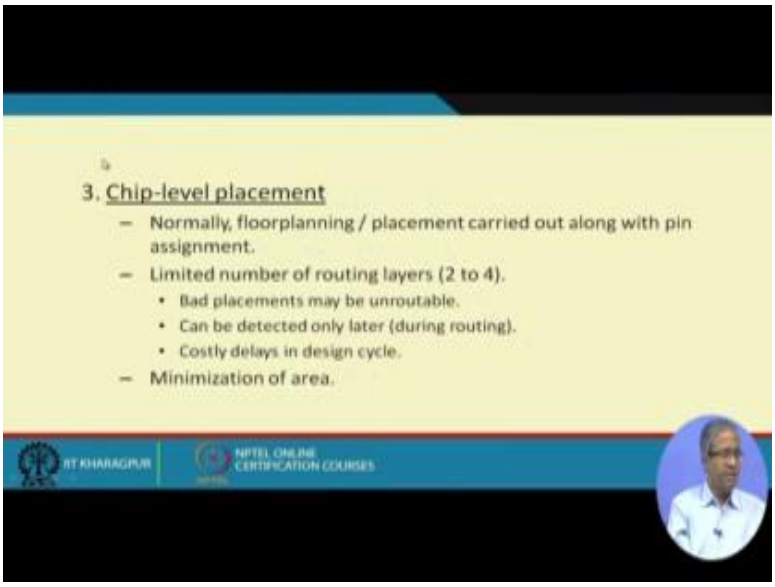
Now this placement problem can be considered at different levels of abstraction; like it system level where your system consist of a number of boards or printed circuit boards PCBs, like say whenever you manufacture seem is earlier we are talking about desktops, how to make our desktops smaller. Now you think of our laptops, how to make our laptops smaller and sleeker, it is becoming thinner and thinner and sma; but when you open the laptop, you will really be able to appreciate the amount and the kind of engineering that goes into the planning of the design.

How they have made is smaller you can really see that. So, system level placement concerns placing the boards together, in such a way that the total area occupied is minimum and also the heat dissipation does not go beyond bounds these both of these are very important; because systems will be having some cooling mechanisms, some cooling fans and may be some more sophisticated ways. So, my air cooled or water cooled, liquid cooled that the number of different ways in which you can cool the systems to dissipate the heat. Similarly within a board you can think of the chips, how to place the chips? See for a PCB in normal your area is fixed, because you are

designing a PCB for a say particular desktop, you know what is the total size of the box your PCB cannot be larger than that.

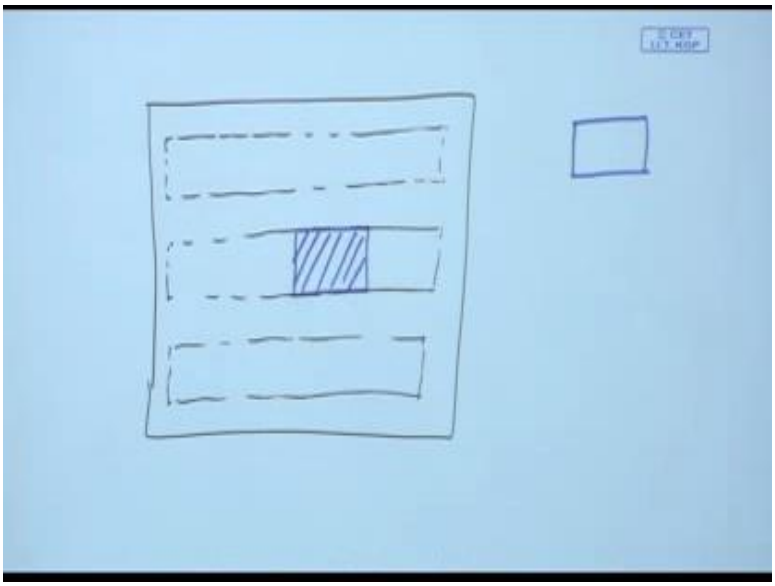
So, your area is typically fixed known before hand and all the chips (Refer Time: 18:10) the modules or a rectangular shape. So, the objective will be 2 minimize the number of routing layers in PCB is also there a typically multi layer routing multiple layers on which the metal layers are running to compute they routing. Now less the number of layers less will be the cost of the PCB, more will be the reliability of operation. So, it is important to reduce the number of layers and of course, to meet system performance requirements in terms of the delays.

(Refer Slide Time: 18:44)



Well now if you go inside the chip, chip level placement you talked about the modules and the blocks. So, floor planning and placement are carried out with pin assignment for most of the chip level placement, because as I said in the beginning most of the chips that you manufacture today they use this standard cell design style. In a standard cell design style floor planning and placement need not be 2 separate steps, because whenever do a floor planning.

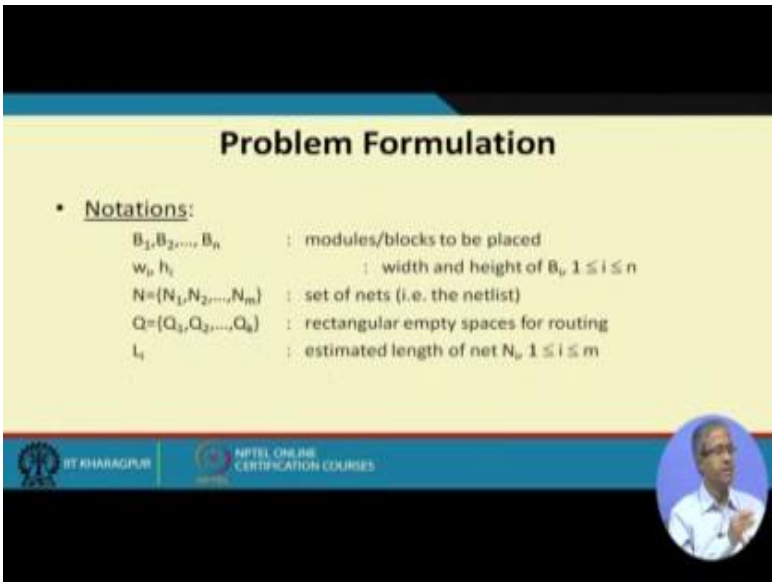
(Refer Slide Time: 19:28)



It means you are doing placement, because your floor is restricted means you see your chip area, what is the floor? For full custom design your entire silicon surface is a floor, but for standard cell design you are floor definition is restricted.

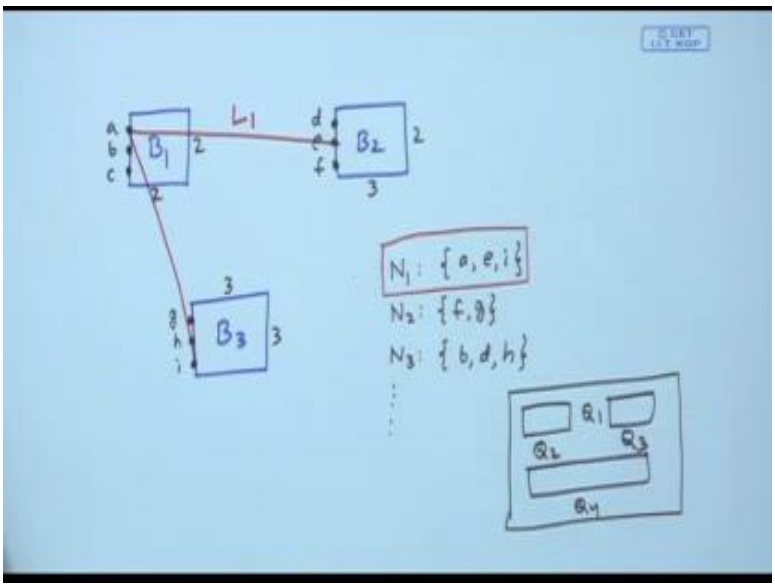
So, you have the rows of standard cells, these are the position for the rows. So, now, your floors are restricted to one of these rows, now suppose we have a block of shape like this you want to place, you can place it either here or here or here suppose it decides you put it here, this is your floor planning as well as placement both. So, you cannot distinguish between floor planning and placement in a standard cell design right. And the routing layers are often limited of course, with the different means say in fabrication technology you can have much larger number of layer 7, but less the better and if you have a bad placement; obviously, I have mention it earlier, you may not be able to complete the routing. So, you may have to come back again iterate on the design make changes, but these might you might introduce some costly delays in the design cycle, which you may not want of course minimization of area is important right.

(Refer Slide Time: 21:01)



So, these are some of the issues, now let us look into the problem formulation from a slightly different angle, now here we will saying is that this blocks B 1 to B n at the modules or blocks to be placed, with each block is having some particular width a particular height, you have a set of nets see N 1, N 2, N m are the net list what I mean is something like this.

(Refer Slide Time: 21:30)



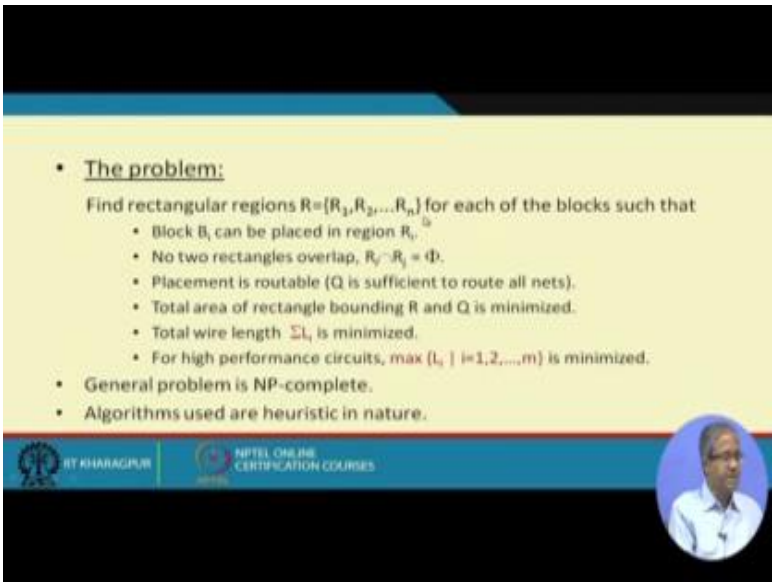
Let me try to illustrate let say I have a block B 1 here. Let me explain with 3 blocks B 1, B 2, B 3. So, I have 3 blocks whose width and heights are known suppose this is 2 and 2, this is 3 and 2, this is a 3 and 3.

So, width and height are known; now there are some pins let say this block 1 have some pins let us give some names. Let say 3 pins each let us taken example let us give some names to each of this pins a b c d e f g h and i then the nets let us say N 1 tells you that the pins a, e, and i they have to be connected which means a, e and i belong to one net. This N 2 may say this f and g have to be connected; this N 3 may say b, d and h of to be connected and so on. The nets are specified in this way nets not necessarily may mean 2 terminal nets, they can be multi terminal nets they wire has to connect a to e, e to i all together, right.

So again coming back; so, I have set of blocks, I have there width and heights, I have a set of nets not only that I have a set of empty spaces for routing. So, empty spaces for routing and I have variable l i which indicates the estimated length of routing of net N i; like when I say let say let say the first net a, e, i. So, a is here, e is here, i is here. So, let say so a has to be connected to e, and i, I can connect to, a I can connect to, e I can connect to any of the middle point let say I am connecting from a to i. So, this is one way of connecting. So, once we find out a strategy of connecting them I have to estimate the length of this net 1 L 1. So, we shall see later that at this stage we can make a good estimate of the length of every net.

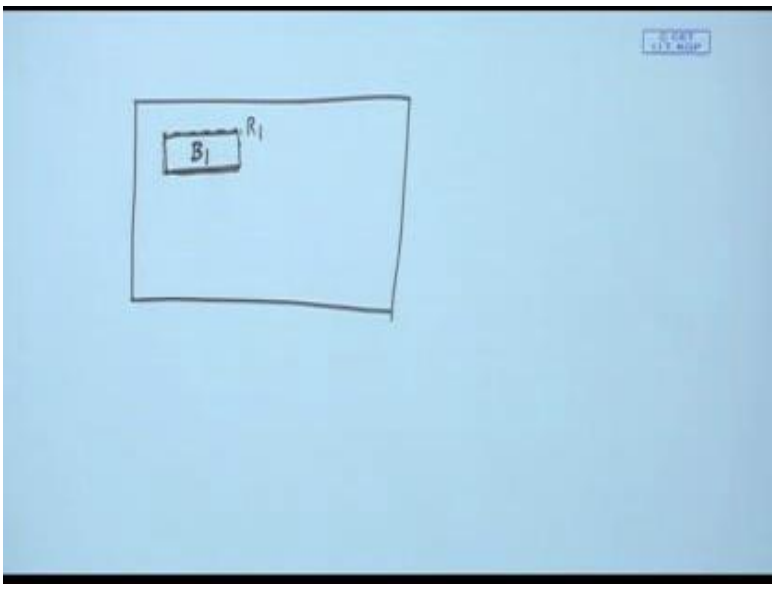
In addition we have this Q I which means the rectangle empty spaces that are left for routing, say means I earlier mentioned that whenever you place the blocks to place them like this. So, this can be your empty space Q 1 this can be another space Q 2, this can be Q 3 and so on. Similarly on this side this can be Q 4. So, these spaces are already kept in for routing for interconnecting minutes, fine.

(Refer Slide Time: 25:07)



So, now when we I am talk about the problems, the rectangular region we have to find out some rectangular regions for each of the blocks; such that block B i can be placed in region R i; see here what we are saying is that we have to identify some rectangular regions on the floor on which I have block B i can be replaced.

(Refer Slide Time: 25:37)

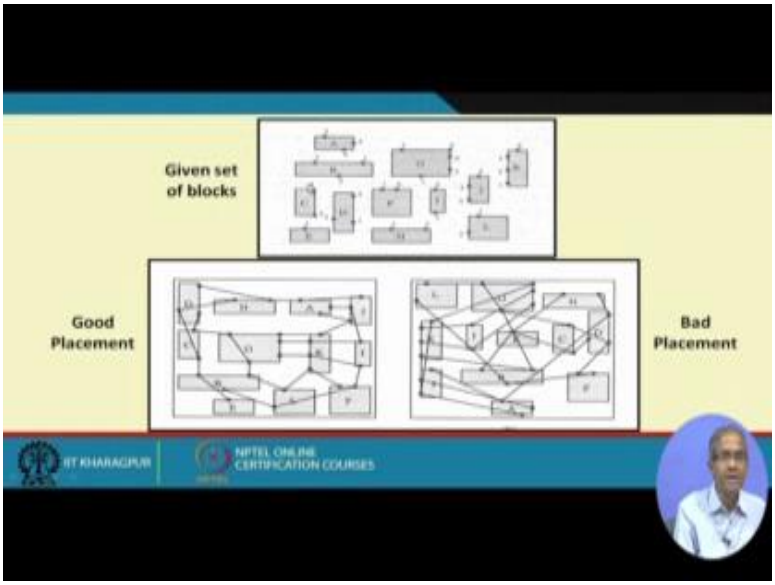


See this is a mathematical o f saying it, but actually what it means is that this R i is this space holder, this is your R 1 on which you place block 1; one which you place block B 1 this means same thing. So obviously, know to rectangular; so will overall

over lap this placement will be routable, placement routable means the amount of space that you are kept Q 1 Q 2 Q 3 that is sufficient, the total area of the rectangular bounding R and Q see R is the area of a replacing the blocks, Q is the area your kept for routing. So, R and Q combined will correspond to the total are of the chip; so this total area of the rectangular bounding R and Q has to minimize. The total wire length has said we have to make an estimate that has to minimized, not only that for high performance circuit you have to limit the delays, the maximum wire length may also have to be kept within the limit.

Now, the general problem of course, is very difficult computational complex and many heuristics are available which work pretty well for very complex circuits.

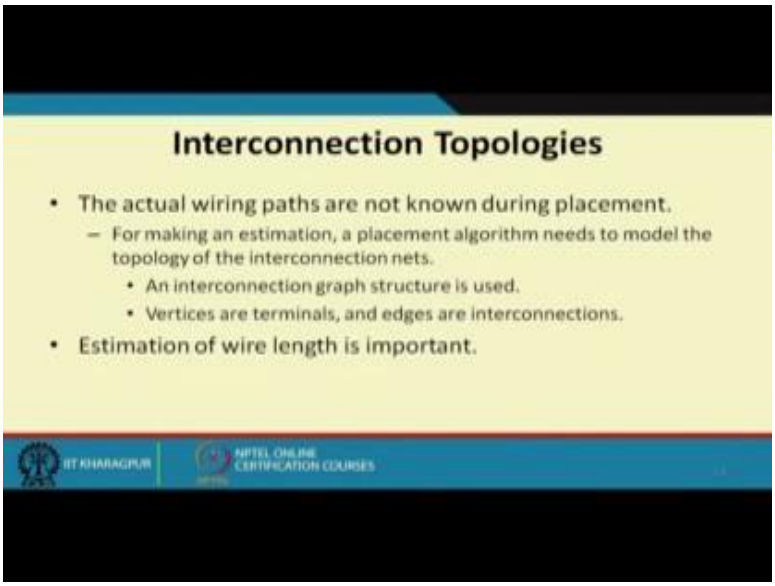
(Refer Slide Time: 27:04)



So, here there is another example that I am showing; here I show a diagram where there are many blocks with some pins, here depends switch a number same 1 1 and 1 means they all belong to the same net. 2 2 and 2 means they belong to the same net. The 2 alternate placements I am showing what they pin connections are shown by straight lines, just to see that how complexities this corresponds to a good placement and these sequence lot of zigzag and long connections, this corresponds to bad placement.

So obviously, you can feel that we have a good placement; your subsequence step of routing will be easier.

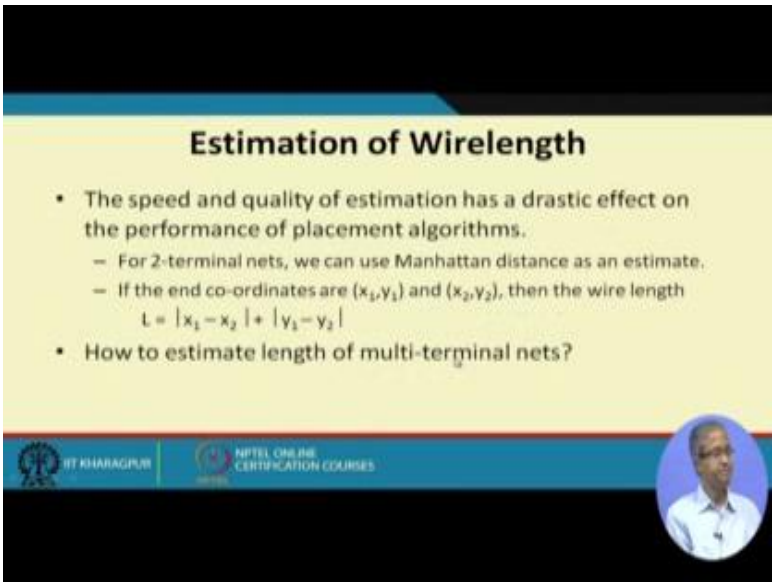
(Refer Slide Time: 27:49)



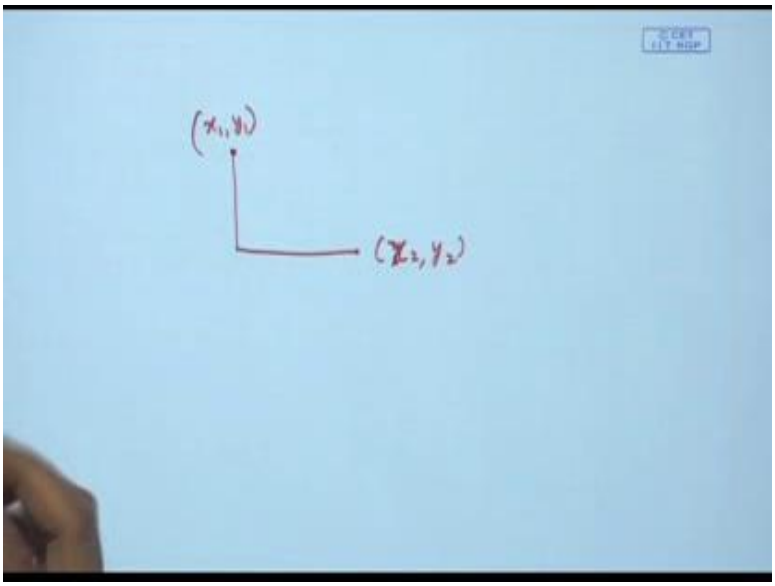
Now, one important step as I said is that during placement once you place the blocks or you decide to place the blocks you have to make an estimate of the length of the interconnection, because this l I, l I is very important minimization of l I, minimizing the longest l I both are equally important. So, you have to make a good estimate on the length of the interconnections, for which some interconnection topologies may have to be assumed, they actual varying paths are not known during placement.

So, for making estimation some placement algorithm needs to model the topology of the interconnection nets. So, we shall see several such interconnection graph structures. So, where the vertices indicate terminals and edges indicate the connections. So, estimation of wire length is used using that graph structure which as I said is important.

(Refer Slide Time: 28:57)



(Refer Slide Time: 29:08)

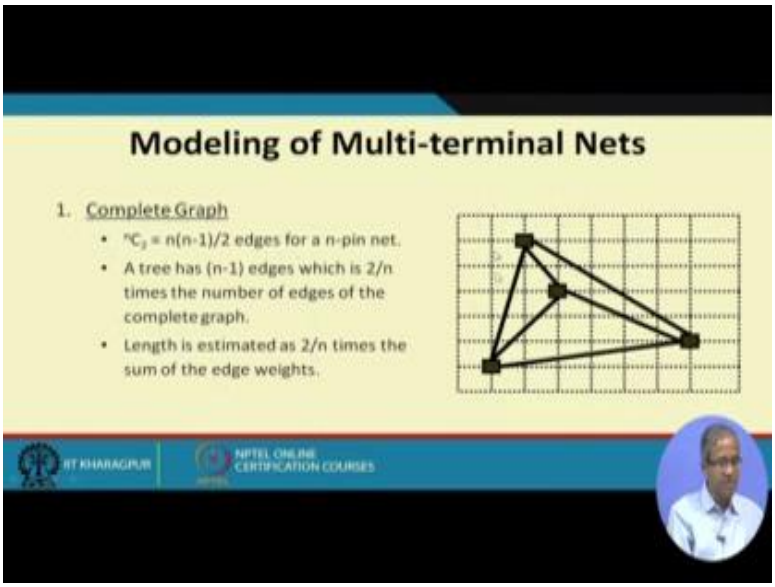


Let see. So, one thing for 2 terminal nets simple manhattan distance is use as a estimate. So, as I said what is manhattan distance? If you have one point with coordinate x 1 and y 1; another point with coordinate y say x 2 and y 2, you can directly take this. So, what do we total length; 1 minus y 2, x 1 minus x 2 some of that. So, it is mention that.

So, for 2 terminal nets, if you take the manhattan distance, the wire length will be simple this, but for a general multi terminal nets how to estimate and that is the

important point here. Now the mechanism to make this estimation should be fast, and also quality should be good it should not be a very bad estimate it should be a reasonable good estimate.

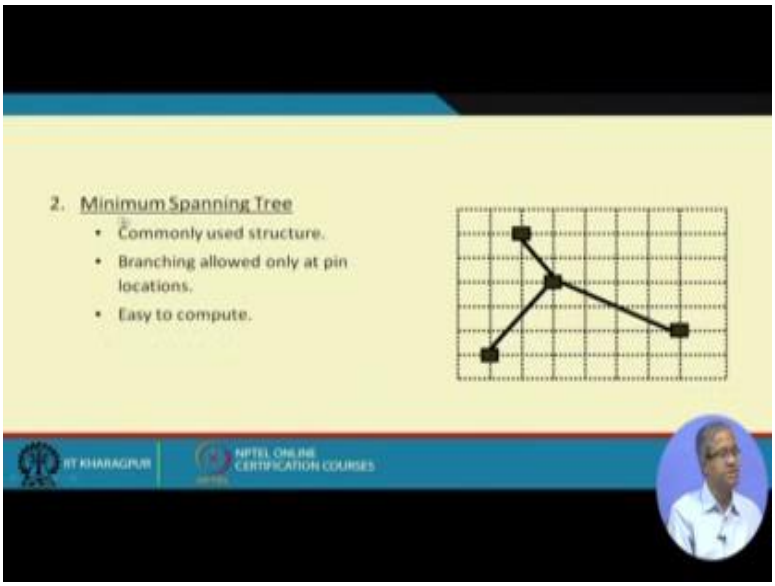
(Refer Slide Time: 30:02)



So, I am showing several alternatives, first is a complete graph well here I am showing on a grid 4 points, complete graph means every vertexes connected to every other vertex. So, for n nodes the number of edges will be n C 2 for n into n minus 1 by 2. So, for 4 it will be 4 into 3 by 2, or 6 there are 6 edges.

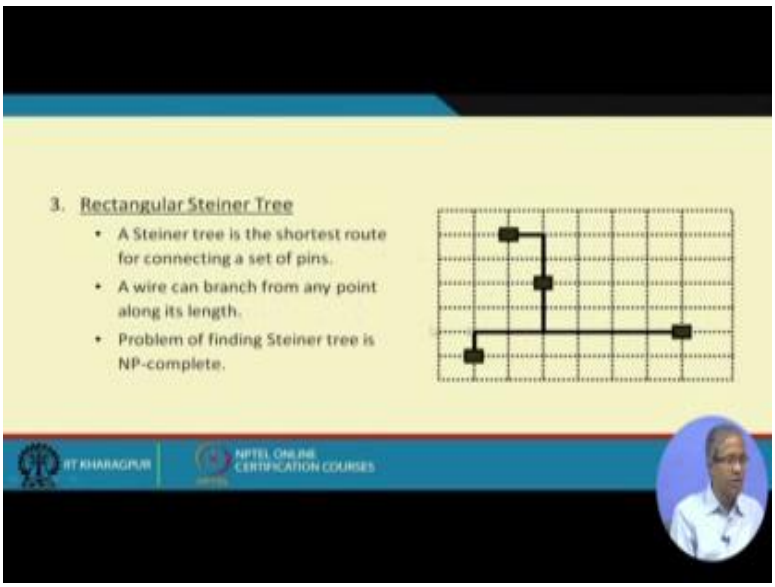
So, if you have a complete graph structure like this. So, you can make a calculation of the manhattan distance of all this 6 edges say between these 2 it will be 1 2 3 4 5 6 between these to it will be 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 in this. So, you can do a sum total. But for an n terminal net you do not need so many n C 2 edges, like n minus 1 edges are sufficient to make the connection; for 4 points 3 lines are sufficient. So, you are using means n by 2 mode number of edge n by 2 times mode edges. So, in this estimate after you take this sum total, you multiply that is sum total by 2 by n as a normalization factor and you get an estimate now here the problem is that you have to make a calculation for all this n C 2 edges. So, the computation complexity is pretty high in this method; this is one approach.

(Refer Slide Time: 31:54)



Second approach is minimum spanning tree; this is whatever saying that for connecting 4 points I need 3 edges; this is one possible spanning tree other may be this, this, this or this is minimum, this some of the edge widths is minimum that is called the minimum spanning tree. Now in a spanning tree you can only use the pin locations has the vertices for drawing the edges. So, here what will be the cost 1 2 and 1 3, here 1 2 3 4 5 6 and 3 and 6 9 10 11 12 13 14. So, total cost is 14 for minimum spanning tree.

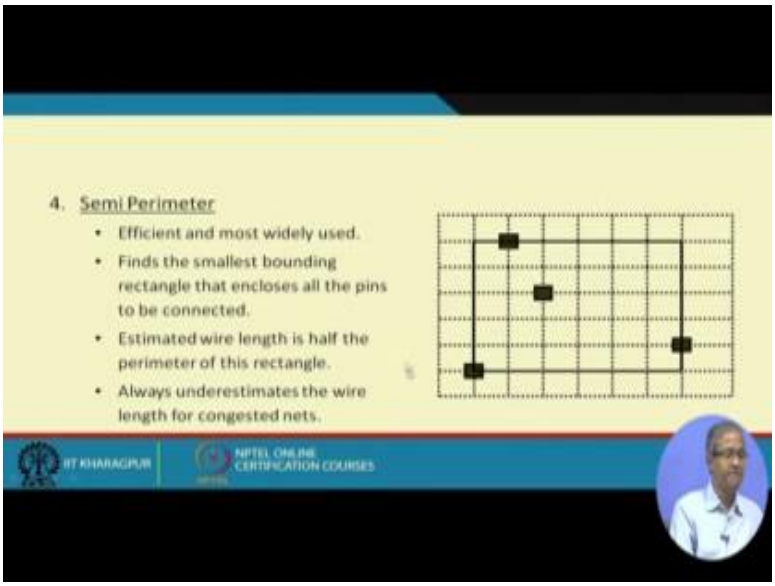
(Refer Slide Time: 32:43)



But the point is that for actual layout you layout like this, like here your allowing only the vertices to make the connections, but this kind of Steiner tree says here a wire can branch from any point along the length like here we have connected these 2 points, for connecting other points you need not have to start from here you can start from any point in the middle.

So, you see this will have a much shorter length 1 2 3 4 5 6 7 8 9 10 11 12 less, but the trouble is that getting the minimum length Steiner tree is computational difficult, but there are a number of adopt estimates or heuristics available, is in which you can get a good Steiner tree.

(Refer Slide Time: 33:31)



But the other very fast method which is quite popular is that you imagine this smallest bounding rectangle, that encloses all the pins and take the semi parameter height plus width that is estimated as the wire length. Of course, the actual wire length will always be larger than this, this always underestimates the wire length, but it is the very fast estimate and it gives a reasonable estimate also not very means widely deviating. So, I think with this we start we just come to the end of this lecture, in the next lecture we shall be continuing with the placement problem we shall be looking at some of the data structures and some of the actual algorithms that are used for placement.

Thank you.