VLSI Physical Design with Timing Analysis

Dr. Bishnu Prasad Das

Department of Electronics and Communication Engineering

Indian Institute of Technology, Roorkee

Week - 08

Lecture 38

Introduction and Optimization Goals – Global Routing

Welcome to the course on VLSI Physical Design with Timing Analysis. In this lecture, we are discussing about the global routing. So, we will discuss mainly about the introduction and optimization goal of the global routing. So, the content of this lecture includes introduction to the routing, then we will discuss about some of the terminology which is used in case of routing phase and then we will discuss about what are the optimization goals we need to set for the global routing. Then finally, we will go for representation of different routing regions. So, this is the typical VLSI physical design flow where we are concentrating on the signal routing which comes after the clock tree synthesis and the signal routing is one of the very important phase of VLSI physical design as it connects all the interconnects inside the chip. So, it is basically divided into two parts, one is called global routing and the detail routing. So, basically what we are doing here is that during the global routing the pins with the same electrical potential are connected with the using the wire segments. So, in case of global routing the pins with the same electric potential are connected using the wire segments. So, after the placement whenever you do complete the placement the layout area is represented using some of the routing regions.

So, we will discuss about what is routing regions in this lecture. So, this routing region is useful for how to create the routing tracks and also, connect the all the interconnects. So, if you can look into this layout and you can see that some of the blocks are placed the block A, block B, block C and block D is placed. So, the placement of the cells or the blocks are already done. Now what is the extra input we are giving to the routing problem? Basically we are giving providing the netlist.

So, netlist is basically the collection or the list of nets. So, the lists of nets are provided and your placement of the blocks are already provided. So, and one more thing is also, needed. So, these are the nets, net names and these are connecting points actually. So,

then we need additional information which is comes from the process that is related to technology.

So, we need something technology information or the DRC rules of the metals, only the metals. So, you we need the DRC rules of metal layers. So, all the metal layers not only one metal layer. So, how do you connect them? Let us say consider the net N1. So, the net N1 is connected with C4, D6 and B3.

If you can go to C4, C4 is this pin and D6 this pin and B3 is this pin, how I can connect them? So, if I go and connect them, So, I consider two types of metal, one is called the vertical metal, one is called the horizontal metal. So, the vertical metal is this, this is your vertical metal connection and this is your dashed lines are called the basically the horizontal metal segment. So, now we have some connection between them that is called the via, this is called the via. So, this is vertical metal, this is horizontal metal. So, now they are connected. So, this same strategy is applicable for all the nets N2, N3 and N4 and all the nets are routed. So, this is an example of how we can do a routing inside a VLSI chip design. So, if you can go into the full chip routing, it is divided into multiple phases. What are the main phases of a full chip routing? First phase is called your global routing. So, the global routing is basically the coarse grain assignments of the metals to the routing regions, the coarse grain assignment of the routing resources to the routing regions.

So, it is not the actual assignment, it is a coarse grain assignment of the routes to the routing regions. But in case of detailed routing, we assign the actual metal lines. We assign the fine grained assignment of the metals to the routing tracks. So, in case of global routing, it is a coarse grain assignment, in case of detailed routing, it is a fine grained assignment of the metal lines. So, now this is the second phase, the detailed routing is the second phase.

The third phase is the timing driven routing. What is this timing driven routing? The third phase is basically what we are worried about the optimization of our critical path. So, we need to look into the critical path and check if my timing is not met, we need to do net topology optimization, resource allocation on the critical nets. So, what we are doing net topology optimization and resource allocation. So, some of the things will apply to the detailed routing to basically on the critical nets to improve the speed of the overall design. So, let us look into the actual difference between a global routing and a detailed routing. In the left hand side drawing is a global routing. So, what is happening here? We have let us say this N1 is connected to these three blocks. So, one line is there, we do not know which metal will be connected connecting them. So, it is a connection.

So, there is no actual metal is assigned here. This is a one net. So, that was the information given to the tool and it does this kind of routing. Whenever I am going to the detailed routing, I am assigning the horizontal metals and the vertical metals to the net. So, if I have this N1 here, So, I have a horizontal metal here, then I have a horizontal metal here, then I have a vertical metal here. And these are connected through via metal 1 and metal 2 via. So, here let us say my metal 1 is if I assume horizontal and metal 2 is vertical, then the via will be metal 1 to metal 2 via. So, this will be this black solid line is basically the metal 1 and dotted dashed line is basically your vertical metal or the metal 2 and the via is between metal 1 and metal 2. So, now we will discuss into what is the placement and how it is different from all other points, all other phases of the VLSI physical design. So, this is the first placement of the blocks.

So, the lines whatever it is showing it is the lines internal to the blocks or the standard cells. So, then what we have we are defining here the wire tracks. The wire tracks are basically, So, if you can see here let us say in case of a typical process if you have multi-layer metal is there, each of the metal has different tracks, each of the metal has different tracks. So, based on that this track is defined in the layout area. Now this shows a global routing, it is not the actual metal lines, but all the cells which is close to each other that will be assigned to particular area and some of the routes are done. Then this is a congestion map, if there is any kind of congestion then we will get the final detail routing. So, detail routing will assign the actual metal lines and it will route all the nets in the design. Now we will discuss about the global routing in detail. So, in case of global routing, So, the wire segments are tentatively assigned within the chip layout. So, the wire segments are assigned within the chip layouts.

Chip area is represented by coarse routing grid. It has a coarse routing grid is created and this wire segments are tentatively assigned to each of the nets. And available routing resources are represented by edges within the capacity in the grid graph. So, whatever the routing resources are there that will be represented by edges with the capacity. What is a capacity means how many lines can pass through that channel or the switch matrix that will also, be available to the router. So, nets are now assigned to those routing resources. Then after you have grid, what are the things steps there? First, basically we have a coarse routing grid, then we can define the capacity of each of the edges, then we will assign the nets to those edges. So, in case of detail routing, there are basically two steps are there. One is called the net ordering and second one is called the pin ordering. So, in case of a net ordering, we will choose one of the net in the design.

In case of the pin ordering, we will use what are the pins that is using that net. In case of net ordering, we will choose one of the nets that will be routed. After you choose that net, then we will check that what are the pins that is connecting to that nets. So, then we can add priority to the nets and the pin when while you are doing the routing. So, which will make our routing process faster.

So, there are some terminology involved with the routing. So, one of them is the routing track or the column which is available either horizontal or vertical varying path actually. The tracks can be horizontal tracks or the vertical tracks. So, the routing tracks can be of two type, horizontal track or a vertical track. Then we have routing regions. What is this routing region? Which basically contains the routing tracks. So, in case of a routing region, we have multiple routing tracks are there. So, we have multiple routing tracks are there in the routing regions. So, the routing region is divided into two types. One is called uniform routing region and the second one is called non-uniform routing region.

So, what is uniform routing region is defined by basically evenly spaced horizontal and vertical grid lines that induce actually the uniform grid over the chip area. So, we have evenly spaced horizontal and vertical lines that creates a uniform grid over the chip area. So, whenever we look for standard cell based design style, this uniform routing region is used. So, which has a predefined routing area and the routing area is evenly spaced in case of a standard cell based design style. And this non-uniform routing region is formed by the horizontal and vertical boundaries that are aligned to external pin connections and macro boundaries.

So, this non-uniform routing region is basically formed whenever the routing regions are depending upon the external pin connection or the macro cell boundaries. So, based on that this non-uniform routing region is defined. And here mainly this horizontal and vertical grids are not uniformly spaced. So, in case of non-uniform routing regions, the horizontal and vertical grid lines are not evenly spaced. So, what is the example of non-uniform routing region? The custom or full custom design style usually leads to non-uniform routing regions.

We will see some examples of this in this lecture. So, what is a channel? Channel is basically if you can see here, we have this is the one row of standard shell, this is the another row of standard cells . So, if you can see here the pins can be located somewhere here, somewhere here like this for each of the cells . So, and there will be some pins here, here like this. So, what it says that it is a rectangular routing region.

So, this is a rectangular space. If I draw it in a different color, this area, this area is basically a rectangular routing region with pin on two opposite sides. So, there are pins in the two opposite sides. The pins are there in the top or bottom, but no pins is there in this direction or in this direction. So, no pins in this direction. So, the pins are there in two opposite direction, other two opposite direction there is no pins. So, that is called a channel. So, there are two types of channel is there, one is called horizontal channel and vertical channel. So, here these are horizontal channels, all are horizontal channels. So, here there is a very interesting concept is there. So, if you can see here, you have these pins are there and these are the routing tracks are there. So, this routing tracks if you can

see how many routing tracks here, this is one track and this is another track, this is another track.

So, one, two and three tracks are there. So, what is the important point here is that this middle of a metal to middle of a metal is called the pitch. So, let us say I have a metal here. This is one metal line and another metal line. So, this metal to middle of this metal is called the pitch, D pitch. So, this pitch will determine how many tracks can be possible in the routing region. In this region, how many tracks is possible? So, basically if you can see here, we can discuss this diagram in detail. So, you have a terminology called channel capacity. We discussed the channel in detail. Now we will go to the channel capacity. What is the channel capacity is basically number of available routing tracks in a region routing region.

So, number of available tracks or columns in a routing region. So, basically if you have a single layer routing, the capacity is height H of the channel divided by the D pitch. Basically if you have only one layer is routed, So, we have let us say only the metal one or two is routing. So, I have basically H this height. I already defined the D pitch. So, basically the channel capacity will be defined by H by D pitch for that layer. I have one layer, So, one channel capacity. But if I have multiple routing layers or multiple metal lines are there, then the capacity sigma is basically sum of the capacity of all the layers. So, what we are doing here the sigma of all the layers is basically summing all the capacities, sum over the capacities.

This is for single layer routing. This is the formula. This is for multi layer routing. We are finding the capacity for each layer and summing up. So, this is the formula.

$$\sigma(Layers) = \sum_{layer \, \varepsilon \, Layers} \left[\frac{h}{d_{pitch}(layer)} \right]$$

Now we discuss about the capacity which is very important. This channel capacity is a very important factor while discussing the basically routing in case of a VLSI physical design. So, now we will discuss about basically T junction which occurs when a vertical channel meets a horizontal channel in a macro placement. So, you have a vertical channel. So, this is a vertical channel and it meets a horizontal channel.

So, this one is creating a symbol T. So, that is why it is called a T junction. So, this T junction is one of the terminology used in case of a global routing. So, we have defined a channel, we have defined a T junction and in case of T junction we have a basically this vertical channel and we have a horizontal channel they meet at one point. Now we have a concept called switch box. Switch box is a type of basically area inside the VLSI chip where you have basically interconnects or the metals can route in all the directions. If you can see here the signal the routing can be possible in this direction horizontally and also,

vertically. So, this one is called a switch box. Switch box is basically we have intersection of horizontal and vertical channels. So, this is an example of a switch box. So, switch box is divided into two different category one is called 2D switch box and second one is called 3D switch box.

What is this 2D switch box and 3D switch box? In case of 2D switch box we have basically four boundaries. Four boundaries are there. We have four boundaries are there. For example top this is the top side, this is the bottom side and this is the left and this is the right. So, you have four boundaries. This is basically top, bottom, left and right. So, it is possible in whenever you are routing in two layers of metal. So, 2D switch box is related to two layer routing of the metals. So, this is related to 2D switch box.

Now we will discuss about the 3D switch box. In case of 3D switch box what is happening? It has six boundaries actually. It has six boundaries. What are the six boundaries? The same four boundaries and two additional basically the upper and the lower metals. It can go to multiple metals. So, this is let us say if you can see here the metal 5 and 4 is used in case of your 3D switch box. In case of 3D switch box we have top level metals and the below level metals are used. So, it has six boundaries. So, those boundaries are basically top, bottom, left, right, then upper and below, up and down, up and down. So, you have upper level metals and the lower level metals are used.

So, that is called 3D switch box. So, we will now discuss about this optimization goal. What is this optimization goal in case of a global routing? So, basically it determines whether a given placement is routable or not. So, whenever I enter into the routing phase my placement of the cells are already defined. So, the placement is given to me. Then I need to find that for that placement of the cell whether my router can route the design or not. So, that is my optimization goal, one of them. Then determine a course routing. Then what it does is that it will first do a course routing of all nets within available routing regions. The course routing is related to your global routing. So, this is basically a global routing.

So, we are discussing about the global routing. Then whenever we are doing course routing of all the nets in the available routing regions, our main goal is to minimize the total wire length. The first goal is to minimize the wire length. So, if I minimize the wire length what is the advantage? My interconnect delay will be lesser. If the interconnect delay will be lesser my speed will be improved.

My capacitance and the resistance in the path will be reduced. The RC in the interconnect will determine the delay in the path. So, if my interconnect length is less my speed will be improved. We are basically the delay of the signal in the net we are reducing. So, the delay in the nets is reducing means that the speed of our overall design is improving. So, we have three different types of design approaches there. One is called

a full custom and standard cell based design and gate array based design. So, we will discuss each one of them individually. So, first we will discuss about the full custom design style. So, what is happening here in case of full custom design style our routing regions are non-uniform. Why it is non-uniform? Because your design is determined by the other macro blocks and it depends upon the design of our basically the way the full custom design is done by manual method based on our requirement we define the channel.

So, the routing region is non-uniform in case of a full custom design style. Often different shapes and heights and width. As you can see here in this example you have only one vertical track and you have basically this is horizontal this is horizontal this is horizontal again this is a vertical. So, it is not uniform across the design. Across the design this routing tracks are not uniform. So, basically these are the channels. This is a vertical channel. This is a vertical channel. This is a horizontal channel. Similarly other cases. Now what we can do is that we can create a graph. So, one more thing here is that the height the routing channels are of different shapes and height and width are of different size. And they can then what we can do is that we can represent that routing regions in terms of a graph. So, we have if you can see here this is a vertical track or a vertical channel. So, this vertical channel corresponds to this one. Now this is again have a horizontal channel this one which is corresponds to this one. So, if you can see we have So, basically this one is corresponds to the channel between block A and B.

So, here you have a block A and block B and this is the channel is 1. So, we can similarly create the blocks and this is the channel ordering. This is a channel ordering. These are type of the channels and this is the ordering of the channels which is used to create the graph which represents the routing regions. So, we will discuss about standard cell based design style.

So, in this case we have two different approaches are there for routing. One is basically with case 1 with limited number of metal lines where we are using only metal 1 and metal 2 for routing. So, in older technology node we do not have enough metal lines So, we use only metal 1 and metal 2 for routing. In those cases we use a extra cell in standard cell library that is called feed through cells to route the net across multiple cell rows. So, how this feed through cell is useful is explained using this diagram. So, if you can see here we have 5 pins are there. So, this is pin 1, pin 2, pin 3, pin 4, pin 5. So, I have 5 pins are there. So, I need to connect all the 5 pins. So, if you can see here we are using this black basically this black color boxes are called the pins and this white color cells are called So, these are called the feed through cell. So, what is the use of this feed through cell? It is passing the interconnect from one channel this is the one channel let us say this is a channel 1, this is a channel 2 and this is a channel 3. So, passing the signal from row 1 to row 2 and row 3 through these feed through cells. So, this is So, for connecting these 5 pins we use 3 feed through cells. Then the second approach here is that we have

basically second approach is to reduce So, this is basically the standard tree solution with minimal wire length. So, here we have minimal wire our objective is to reduce the interconnect length. So, this is the approach we have 3 feed through cells are there. But if I go to the second solution this is the first solution to the first routing solution this is the second routing solution.

In the second routing solution my routing is different. My interconnect distance is higher if you can see here I have added this extra lines here. Now what is the point here is that we have basically reduced one feed through cell. So, here this feed through cell is not used. So, this feed through cell I am writing in short feed through cell is not used this feed through cell is not used.

So, this is a standard tree solution with fewer feed through cells. So, this is a standard tree solution with minimum wire length. So, there are two solutions to the same routing problem. So, we have multiple solution to the same problem. So, how the things changes depending upon our objective function and constraints. In the second solution our target is to reduce the number of feed through cells. In the first solution our target is to minimize the wire length. Now we will go to the gate array based design style. So, in this case of gate array based design style the sizes of the cells and the sizes of the routing region between the cells are fixed. The sizes of cells and the sizes of the routing regions of the cells are all fixed because it is done whenever you fabricate the cells fabricate the gate array.

Whenever you fabricate the gate array those routing regions are already fixed. Now what we are doing here is that So, you have a gate array with channel height of 4. Channel height of 4 means we have 4 interconnect 1, 2 this is 1, this is 2, this is 3 and this is 4. So, 4 tracks are there channel height of 4. So, if I do one possible routing is that you can do the connections like this but what is the problem where we are minimizing the wire length but what is the problem with this one is that C is not routed.

The CNET is not routed. There is no routing resources available here. How can you solve this problem? The problem can be solved if you can basically modify the placement of the blocks. So, one method to solve this problem you can do the modify the placement of the blocks such a way that all the nets can be routed. So, the way it was placed that CNET is not routed but we can solve the problem because the routing region is fixed we cannot modify the routing region in this case. So, here maybe one of the solution is to change the placement of the blocks to make the signal routable, all the signals routable. So, in this lecture we discussed about basically some of the terminologies involving routing. We discussed about global routing and detail routing.

Thank you for your attention.