

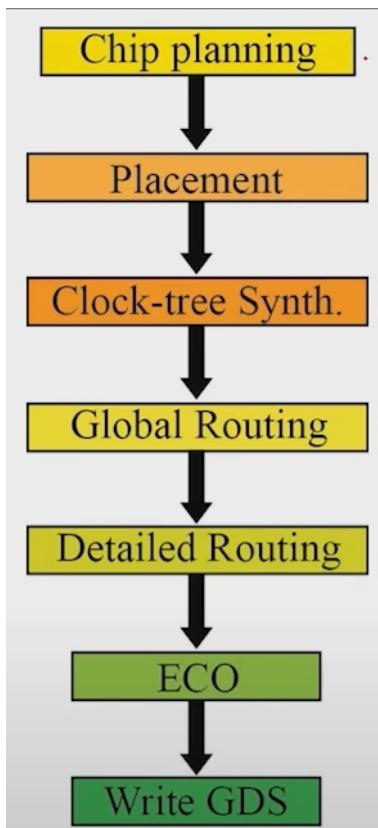
VLSI Design Flow: RTL to GDS

Dr. Sneh Saurabh

Department of Electronics and Communication Engineering
IIT-Delhi

Lecture 52 Routing

Hello everybody. Welcome to the course VLSI design flow RTL to GDS. This is the 41st lecture. In this lecture we will be discussing routing. In the earlier lectures we had seen that the physical design is broken into simpler tasks which are shown in this figure. And in earlier lectures we had looked into chip planning, placement and clock tree synthesis.



And in this lecture we will be looking into routing. Specifically we will be looking at global routing, detailed routing and then post routing optimizations. So first let us understand what routing is. So routing basically involves making physical interconnections between different components of a design.

So in earlier lectures we had seen that we have done the placement of objects. So we have placed the large objects during chip planning and the standard cells during the placement. So all the entities in our design are basically placed at appropriate places on the layout. Now once the placement is defined then we do the routing. And during routing what we do is that we create physical interconnections between different components of a design.

So these components are connected using nets in the netlist. Now using the information which is there in the given netlist the connectivity of the components is defined in the netlist that we started with. Just to recap that in physical design what we

do is that the physical design tools take input as the netlist then the sdc file in which the constraints of the design are there and also it takes the information about the libraries. And we have seen that there are two types of libraries. The first is the technology library that we saw in detail when we discussed logic synthesis and the physical libraries which are typically in the form of lef.

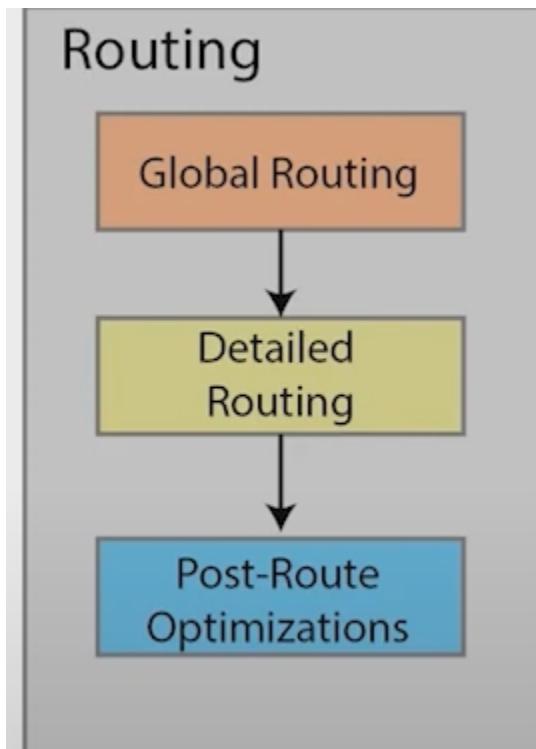
So these are three most important things that a PD tool needs. Additionally we give information about the floor plan, the die size and other things and the tool options and other things that are also the input to the physical design tool. At the end of physical design we get a layout. We get the layout which goes to the foundry for fabrication. So the netlist is given as an input to the physical design tool and whatever the connectivity is there in the netlist that connectivity is made by the router by using various layers of metals and interconnecting them and so on.

So this is the routing is a very complicated task and perhaps the most complicated task in the vlsi design flow because it involves a large number of entities. So if there are say billion instances then you can imagine that how many connections will be there it will be more than that number of instances and therefore since the number of connections or nets involved in a design is very huge making physical interconnections for them is difficult. And it is further made difficult because of various constraints that a routing tool must honor and what are these constraints? So these constraints are related to routing resources. So we discussed that in a typical IC there are say 10, 12, 15 layers of metals. But still the number of metal layers and the routing resources is fixed.

It is not that it has got infinite routing resources. So given the limited routing resource a router needs to find a solution and that is not an easy task. Further it has to honor the design rules that come from the foundry. Design rules meaning what should be the minimum spacing between two wires of a given thickness and so on. So those kinds of design rules must be honored by the router and further it must also honor the timing constraints.

So we give timing constraints using SDC file and those timing constraints must also be honored and the signal integrity issues also come up when two wires are running in parallel and because of the cross coupling capacitance one signal impacts other signal and therefore signal integrity issues comes up and the router needs to take care of those issues also. And therefore the routing is a very complicated task. And therefore we break down this routing into multiple smaller tasks. So these tasks are the first one is global routing. So in global routing we create the plan for routing for each net.

So in this when we are creating the plan we are not actually creating the physical



interconnection. We are not doing a layout for the wires. We are just saying that for this net it will cross all regions. So it is in terms of routing regions. The definition of physical interconnections that is found or that is determined during global routing is in terms of routing regions not in terms of wire layout.

And then in the next step that is when we go into detailed routing the detailed routing actually decides the actual layout of each net in the pre assigned routing regions. So the routing regions were determined during global routing right and using that information the detailed routing basically creates the layout of each individual nets in our design. And once this detailed routing is done then we do some post routing optimization so there can be some

problems in our design. We fix them or improve some QR measures like timing and so on. So those are done after detailed routing. So in this lecture we will be looking at all these design tasks.

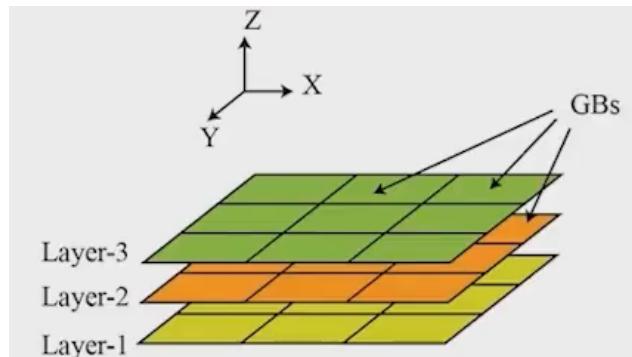
So first let us look into global routing. So in global routing what is the goal? The goal of global routing is to provide complete instructions to the detailed router on where to route each net in the design right. So the global routing is basically creating a plan right and this plan will be further used by or utilized by the detailed route right. So the plan should be such that the detailed router ideally should not fail, meaning that there should be a feasible solution which could be easily determined by the detailed router that is the primary goal for global routing. It eases the task of finding a solution or finding a feasible solution for the detailed route.

Now during this global routing what are the objectives that it must try to achieve? First is that it should maximize the probability that the detailed router can complete the routing, meaning that it should take decisions such that the routability of our design improves right or that the detailed router will not fail right, that is the first objective and the most primary objective of a global router. Additionally it should look into the total wire length and try to minimize it. The total wire length is also some kind of measure for the routability and then it should minimize the critical part delay so that maximum

operable frequency of our design can go up. Now what are the desirable characteristics or what is what is desired of a global route? So global routing is basically used many times in a design flow though we are just looking in detail at this point of time but this global routing is used at multiple design steps right. For example even during synthesis there are some flows in which we create some rough flow plan and do some rough placement and routing right and then assess the quality of logic synthesis solution and take some decisions right. So those are known as basically prototyping. We are creating a physical prototyping of our design by creating a rough layout of our design right and in those cases also this global routing is used right.

And during floor planning also we have seen that to assess the congestion of our design we can use global routing and also during placement we can use a global router to assess the quality of placement and based on that it can change the solution and so on. So since global routing is a kind of rough way of doing a routing of our design it should be very fast right that is what a desirable characteristic of global routing is. It should be very fast, a few orders of magnitude faster than the detailed route right. So that is why if it is fast then we can use in early design stages also for getting some rough information on how the interconnect will look like and based on that we can do some timing analysis or congestion analysis or make some decisions which will improve the QR of our design. Now what is the routing model that is used right?

So to make the global router fast what we do is that we use some specific type of routing model right. So what we do is that we divide the routing area into rectangular grids right. For example suppose this was our layout right this is the layout right and this suppose that there are say three metal layers in this case I am just for simplicity I am assuming that there are three metal layers metal level layer one metal level we can call it M1 this is M2 and this is M3. Then what we do is that we



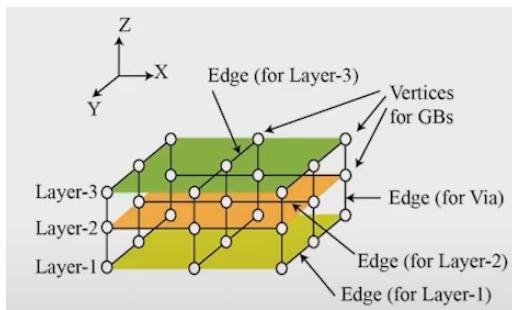
divide the layout area or routing area into rectangular grids right which is shown in as we have done here. So there is a rectangular grid for each layer right we have divided it into a rectangular grid.

Then each rectangular region is called a global bin. For example if we take this small rectangular region this is a one global bin right. Similarly this is another global bin this is another global bin right. So in this figure that we have shown here there are say one two

three four five six seven eight nine global bins right. Each individual small rectangular region is known as global bins or GBs that we have as shown in this figure right. Now various physical design tools can use various different kinds of names for the same thing right.

For example some can use the name as global tile or routing tile or say global cell or bucket right or even or we are calling it as global bin right. So just for consistency let us call it a global bin but take note that physical design tools might be using some other name for the same thing right. So now using these global bins which are built over the layout we create a grid graph right. What is a grid graph? Now the grid graph is also a similar type of graph right as a conventional graph; it contains vertices and edges. Now what are the vertices? So vertices correspond to the GBs right.

So in this case for layer three there will be three vertices and nine vertices corresponding to nine global bins right. And what are the edges? The edges represent the boundary between adjacent global bins right. For example this is the boundary between this and this so we will have an edge right. Now we have discussed earlier that the connections or the routing that is done for interconnecting various layers that follow a preferred direction for each layer right. And that is also taken into consideration while creating a grid graph right.



one vertex.

So let us take an example like what happens right. So in this case there are say nine vertices for the first layer m we are calling this as m1 m2 and m3. So for the m1 layer we have nine vertices right nine vertices others are not. And then we for the other m2 layer also will have nine vertex and for m3 also will have nine vertex right. So corresponding to each global bin will have

And then we draw the edges. Now how do we draw the edges or how do we create the edges? So we create the edges by looking into what is the preferred direction right. So we create edges between two adjacent vertices only if they lie along the preferred direction of routing. For example let us assume that the preferred direction of routing is along the y along the y axis for the layer m1 and m3 right. And for m2 let us assume that it is along the x right. So for the m3 layer for the m3 layer will have edges only along this right these are the edges.

We do not have an edge along this direction which is a non preferred direction right. Now for layer 2 or m2 layer will have edges only along the x direction right not along the y direction. Similarly for layer 1 will have only edges along the y direction that is shown right and so right. So we create edges along the preferred direction and also we create edges between two vertices that lie vertically adjacent right. For example this word this global bin and this global bin those are vertically adjacent one is lying above the other right.

And therefore we create an edge for this. Now what does this edge represent? So these edges represent the vias right. So while making interconnections we will have to go from one layer to another and for that we need to use these via layers or via edges right. Now given this grid graph how does a routing of a net is done right. So we first associate the pins with the proper global bin right.

So for example let us assume that there is an inverter which is driving another inverter right. These two inverters are driving this inverter is driving the next one and let us call that this net name is net right. So we want to create an interconnection for this net n right. So the first thing we need to locate is where the pins for this net are. So the pin is this let us call this as y pin and this as x pin right.

Now we have already done the placement right when we have come to routing we have already done the placement. Therefore we know the location of this pin y and x right. So on the on the on the on the layout we know that say for example y is at this location somewhere in the in this in this global cell. So y is here and say x is here right.

So we locate the corresponding vertex. So the corresponding vertex for this is this one and for this is this right. Now to create the layout of the wire net n what we do is that we find an optimal path in the grid graph that contains all the vertices of that net right. So in this case there are only two vertices because there are only two pins right. So we have to create a path between this point, this vertex and this vertex which are shown in red on the right. And how do we create this path? We create it using some optimality measure right.

So we need to optimize it based on some cost measure and find a routing solution or for this particular net right. So the routing solution can be something like this. Maybe it goes from this layer one to or from this vertex to the next vertex that is this one right. And then it goes along this then it goes along this and probably then it goes to a layer up then it goes through this and this right. I am just showing one of the possible paths right it need not be optimal right. So the tool will find an optimal path right and this optimal path is basically a kind of representation of what is the layout of this net right.

It is not exactly out of the net but it is the path that the layout of the net will be traversed right. So the actual layout will be created by detailed routing not during global routing and we are discussing right now global routing right. So during global routing the solution for the routing solution for this net will be as shown in this figure right. So it is not optimal but what I am just showing for the purpose of illustration is that it could be a routing solution right. Similarly it will be doing routing for other nets also and then the entire design right.

So for entire nets in our design the global router will find a solution. Now when a net crosses the boundary of 2 GBs then we say that the net utilizes the corresponding H right. For example we saw that we created this as the starting point. This one was y pin and this one was the x pin and we created the layout right. We created a lay or a path on this grid graph right. So when this happens right then if we consider for example this edge this edge was used by this net right.

So when a net crosses the boundary of 2 GBs right when for example in this case it is crossing the boundary of these 2 GBs right then the corresponding edge is this right this one which I am highlighting this one. So the corresponding edge is said to be utilized or used right. So note that one particular edge can be used by many nets right. Why can it be? Because this is a global bin that has been created it contains many routing resources for example it may contain many tracks of wires right. So one edge that I am showing here for example this edge right it is actually physically representing many tracks of wires that can be laid out right.

It is not only one track it can be many tracks and therefore each edge in our in the in the grid graph that can be used by many nets right. So the number of nets utilized by an edge is known as use of that net or we can also call it as the demand of that edge right. So we say that an edge is used by a net if the net crosses that edge right. Now if there are multiple nets which are crossing this edge then we say that this edge has got the use of that many nets right. For example suppose this edge was being used by 3 nets right it was used by 3 nets right.

Then we say that the use of this edge e let us call this edge as e is equal to 3 right because 3 nets are crossing that right. Now we also need to be careful about what is the capacity of the edge right. If one particular edge is used many times may say 1000 times that is not a good routing solution because the later on detail router will not find sufficient number of tracks in which we can lay out say 1000 nets right and therefore detailed routing will fail. Therefore while doing this global routing the router should also understand what is the capacity of the edge right. So capacity quantifies

availability of routing resources for an edge; it somehow relates to how many tracks are available for routing by the detailed route right.

So now this capacity is also called supply right. So whenever we use a net we say that the demand has been created for that edge right and when there are many resources available for an edge we say that the supply is increasing right. So the routing blockages limit the capacity right. So note that we are doing routing after doing power planning and also we have the power planning that was done during the chip chip chip planning stage right. So when we did power planning some resources say top level routing resources or routing layers that were used for power grids right.

Therefore those are not available right and there also we are doing this routing after we have done clock tree synthesis and therefore some routing resources have been used by the nets that are related to the clock nets right. And therefore when we come to this routing we are basically routing the data signals right the clock signals are already routed the power lines are already laid out right. Now we are left with data signals which are mainly targeted during the routing right. So while computing the capacity of an edge and router one should also be aware of where the routing blockages are or which are already utilized routing resources so the capacity will come down right. And it also depends on the layers and the design rules right.

For example, if we say if we have a thinner wire or thinner wire meaning that its width is small right and if we if the width is large right. Now depending on the width the design rule says how much spacing should be there and so on. And therefore if we use a thick, more wider wire then the capacity of an edge decreases right at the depending on the layer and also the design rules like what should be the spacing between one one track and the other that is decided by the design rule which comes from the foundry. Therefore the capacity computation by the when the global router is computing the capacity it should be it should also take into account the layer for which it is computing the capacity and also the design rules right. Now what should be the relationship between capacity and edge right? We said that we found that the tool found a path right, a global routing tool found a path from this point to this point something like this right.

Now while finding this path what should the constraint be on the routing on the on the on the routing path. So one of the most important constraints is that it should honor the capacity constraint meaning that the use of a or utility of an edge determines how many nets are using a particular edge that should be less than or equal to the capacity right. It should be less than or equal to capacity if it is more than it is likely that detailed router will fail because we have routed many nets through an edge and they are they are not sufficient routing resources or tracks available for detailed router to do the the layout or

create the layout for that net and therefore the capacity constraint will be violated and we measure it using a quantity which is known as overflow right. So we define overflow as use minus capacity right if use is less than capacity right so if we are using more than the capacity then we are there is an overflow right and this is a good condition it is an overflow and how much the overflow is it is defined by the difference between that. For example suppose there is a for an edge the use of that edge was say 5 and the capacity of that edge was say 2 then what is the overflow the overflow in this case will be 5 minus 2 that is 3 right.

$$\text{Capacity constraint} = \text{USE}(e) \leq \text{CAP}(e)$$

$$\text{Overflow } OF(e) = \text{USE}(e) - \text{CAP}(e)$$

$$\text{USE}(e) = 5, \text{CAP}(e) = 2, \text{so } OF(e) = 5-2 = 3$$

$$\text{Congestion } CG(e) = \text{USE}(e) / \text{CAP}(e) =$$

If the use is less than or equal to capacity then we say that then we say that the overflow is 0 because in that case the use is not more than the capacity that is a legal condition and therefore overflow is defined as 0 for those cases. And then if there is no overflow and we want to quantify which routing path is better than the other then we use a measure which is known as congestion right now how do we define congestion? We define congestion as use divided by capacity. For example if the use of a net was 4 and capacity was 6 right so the congestion will be 4 by 6. Now in another solution if say there the use was say 5 and the capacity was still 6 then the congestion measure will be more in the second situation right. So, we can quantify how much utilize what fraction of the resources are being utilized by looking at the looking at the congestion measure and typically we try to reduce the congestion not designed by by spreading out the routing or use such that it is fairly equally used or evenly used over the layer.

Now what is the global while finding the path? For example the path that is shown here will be what the routing tool will try to look into and try to optimize right. So, the first thing the tool will try to attempt is to route all the nets in our design with overflow of 0 right meaning that if the overflow is 0 then probably the routing the retail routing will succeed and that is the what primary goal of global routing is to make the retail router succeed to find a feasible solution right. Now while trying to attain it, say overflow is equal to 0 if the tool needs to sacrifice other QR measures for example, timing it can do that right because the more important thing for the tool is to find a feasible routing right. So, for example, if the tool is finding a path and this path was broken or it was having a not available routing resources all were used up by other nets right then it will probably look into say other path through this right it may take a longer path maybe it may find a

longer path or it can take a detour to between the between various various end points of a net to avoid overflow right and if it taking if it is doing a detour or taking a longer path then probably other QR measure like wire length delay those will be suffering right.

So, the routing tool can take those kinds of decisions. Now what are the challenges for global routing though so the first or the biggest challenge for a global router is to do the things or perform global routing with minimal run time and also maintain accuracy. Now what do we mean by accuracy? Accuracy of the timing measures the other other QR measures with respect to the detailed route right. So, what is global routing trying to do global routing is trying to basically simplify the problem for the detailed route right and therefore, it should be the the solution that the global router produces right it should very much be correlated with the with the with the detailed router and that is maintaining both things that is run time, lower run time and accuracy with respect to detailed router higher maintaining both them both of them is a challenging task right. So, how can we improve the run time? One of the easiest ways is to increase the size of the global bins. Now when we increase the size of the global bins what happens is the number of bins that needs to be tackled by the global router that decreases right because the layout area is fixed we are increasing the size of is the local the the global bins.

So, the number of global bins that needs to be tackled by the global router that comes down and therefore, the problem gets simplified for global router and it is run time can come down, but what did what effect it will have on detailed router. The detailed router will now see a bigger bin right and therefore, the the problem of or the complexity of the problem for the detailed router that increases right and therefore, in effect if we increase the size of global bin the problem of routing is transferred from global global routing to detailed routing it is actually not solved right. Therefore just increasing the global bin size will not help right because the accuracy will suffer and why do we want to maintain the accuracy of the global routing with respect to detailed routing. So, the reason is that see that the we are what we are trying to do is that we at the global routing stage we are trying to ah to find we we are trying to find some feasible solution for the detailed router by by finding some paths based on global bins which can be further utilized by detailed routing and also what we are trying to do is that we are trying to anticipate problems that will be seen by the detailed router right. For example, suppose a detailed router routing problem is much more difficult than the global routing problem and therefore, the run time of a detailed router is significantly higher than the global routing.

So, it may happen that when we run detailed routing it may take say 12 hours to run and at the end of 12 hours the detailed router simply says that it could not find a feasible solution right. Then that is a very bad thing because our time and design time effort and

computation time has gone waste right. So, what what we do what we typically the responsibility of global routing is that anticipate the problem that will be seen by detailed router and if there are issues for example, congestion or overflow and those things those those problems should be flagged during global routing itself and a designer can then take a corrective action right. For example, if some perturbation in placement is required then it can be done and so on. So, those kinds of design changes can be made by looking at the solution of the global routing and therefore, the global routing should try to model the problem such that it is similar to the detailed routing problem right.

So, the what what I mean to say is that the the global routing should model say the routing resources in a fashion which is similar to what is seen by the detailed router and and to do that what the global router needs to consider is the are the design rules that a foundry has basically imposed on the on the on the layout right or related to the interconnects right or if those things are considered more ah more at a at a fine grained manner then the the global router can be the then the solution found by global router will be ah will be something which will be similar to the that seen by detailed router and the design closure can be speeded right. Now, let us look into detailed routing. Now, what are the goals of detailed routing? So, the goal of retail routing is to determine the exact layout of each net including all the attributes of the wire segments such as width and location. So, now, in detail, the router will create the entire layout for a net right for it to create. So, it will break down a net into multiple wire segments and for each wire segment what is its width what is its location those will be determined during detailed routing and what constraints detailed routing must must ah consider first of all of course, the connectivity the connectivity that is defined in the in the net list that must be honored there should not be any extra open net or any short circuit net right.

So, the connectivity should be exactly the same as in our given design as in as in our given net list right and then the design rules also must be considered. So, those will be coming in from the foundry and the and the detailed router must consider the design rules and in addition to that it must look into the timing timing timing constraints the signal integrity issues and of course, the run time right because the run time of detailed routing is very very important because typically it takes very large run time and therefore, using some techniques for example, multi threading and others the detailed routing router tries to control or keep its run time lower. Now, in typical advanced process nodes or industrial designs the number of metal layers is large; it may be 10, 12, 15 and so on and therefore, there are lots of routing resources available even over the cell right. So, the cell may be using a few routing resources for the internal connections right, but the upper layers are all available and therefore, over the cell routing is done during the detailed routing right.

Now, how does the detailed routing work right? So, it works using the similar as or it creates a routing grid or what we call a detailed routing grid which is similar to the global routing grid that we just saw. However, here the granularity of those routing grids will be much finer than what we saw for the detailed route. Now, in this case the tracks with uniform spacing are right. So, in the detailed routing what the data structure or the abstraction of tracks are used. Now, what are tracks? Tracks are basically metal lines and we we so the the basically the detailed router tries to find out which all tracks to be used for making connections for a given net.

Now, these tracks are typically uniforms uniformly spaced right. So, we create these tracks with a uniform spacing, this spacing is the same for all of them right. And this spacing is known as the routing pitch and what should be the spacing of these routing tracks right. So, this will come from the foundry and why do we create these routing tracks and why the router works at the abstraction of tracks. It is because it allows us to do automation. So, it allows the use of tracks for routing or doing detailed routing simplifies the algorithm of detailed router right.

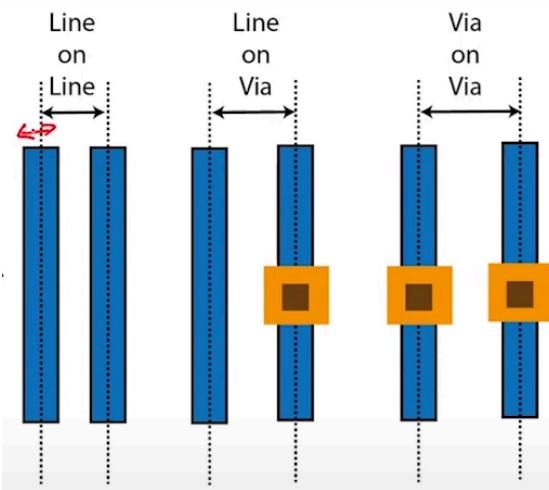
And therefore, most of the tools use this kind of detailed routing tracks as an abstraction for carrying out detailed routing. Now, what should be this routing pitch? An important thing about the routing pitch is the spacing between two tracks right. This is the spacing spacing or we call it the routing pitch, routing pitch. Now, what should be this routing pitch right? So, routing pitch is defined as the minimum spacing allowed in the technology right. Now, during routing what do we what we what do we typically want? We want to pack these wires as compactly as possible right.

We want to basically have a design or layout as compact as possible right. Then we can save on the area right die size and we can save on cost and also improve the yield right. So, typically we want to do the interconnections or create routes as packed as possible right. So, and how, and ideally if you want to pack the wires as close as possible or as tight as possible, then where do the constraints come from? The constraints come from the foundry. The foundry says that if you have this much width of the metal wire right the metal segment, then the next segment must have at least this much spacing the minimum spacing.

That kind of information will come from the foundry and the router needs to honor them right. So, these are coded in as design rules right. So, the design of the foundry gives us design rules in which the information is what should be the minimum spacing between two wires of a given width. And given this the design rules still there are many choices available right. For example, what we can do is that the routing pitch we can define as the minimum distance between two lines that is allowed by the technology.

For example, suppose this is the minimum width of the line right that can be drawn in a given technology and for that minimum width we say and the foundry says that the spacing should be say w right. Now, we can make the tracks w distant apart right all

these are $w w w$ and so on right. So, we can do this kind of or we can create tracks based on this information, but what will be the problem with this? The problem with this will be that it will be too aggressive and it will create a lot of DRC violations meaning that there will be design rule check violations. Why can there be a design rule check violation? The reason the violations can come for example, if we want to insert a via here right. For example, this line was going and then it had to change the layer either go up or go down right then we have to insert a via.



Now, if we insert the via there are different rules for via versus line right and therefore, there will be a kind of so when there is a via and there is a line the spacing between via and line will be will be required somewhat more than what will be the minimum width between two lines that is w defined by this and because of that there will be design rule violations. Now, to fix it what we can do is that we can if we want to insert a via we will leave this track and use the next track next parallel track, but then this track which we have left that is being unutilized we are not packing as efficiently as possible right. And therefore, this may not be a good choice line on line as the routing page. Another thing can be done is that we can consider saying via on line right. So, that the two the foundry will define that when there are two vias adjacent what should be the minimum spacing and based on that we create routing page right.

Now, this will save us from the DRCs, but the problem is that if we are packing all the wires as closely as possible the answer is no. Why it is no because there will be many lines where there are no vias which are in the adjacent right. There are just two lines right and in that case if there are just two lines we could have made them closer without having any violations which we are not doing. And therefore, making a track or the grid graph based on tracks with via on via spacing is also not a good solution; it is too conservative right. Typically what we can do is that there is a in the in the design rule we have that between a via and the line what should be the minimum spacing right and we can take that as the let us call let us call it as w dash and we can create the tracks

based on the width of w dash right. And therefore, whenever we add we need to add one via right we need to we need not make any we need not leave any unused space right or unused track we can have it on the same track right.

And therefore, they we have allocated space for the via right, but when we need two vias like in this situation whether we can either skew them in space or we can leave one extra line, but those situation where two vias should be exactly at the same same horizontal space right at the same space that may not be that case may not arise too often. And therefore, this line on via spacing may be more optimal or is a good trade off. Now, using the routing model that we just discussed, the tool will find a feasible solution if it exists or if it can find the right using some heuristics. So, the routing problem is an NP complete problem or a difficult problem and there are various heuristics that are used by the tool. And in this course we will not go into the algorithms of routing. If you are interested you can look into the books which are dedicated to physical design algorithms right.

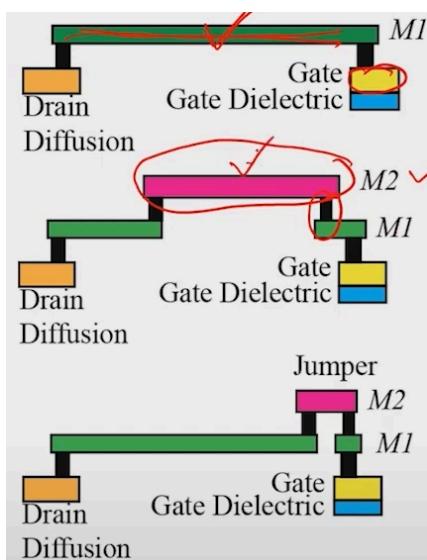
So, after the routing is done right after the detailed routing is done then some optimizations will be required or will be done to fix issues in our design right. So, till the time we have not done the routing of the signals we are not we we are just estimating that what could be the interconnect delays right based on we do not even know the layout of the interconnects. And therefore, based on some heuristics and some assumptions of how the wire will look like, the delay of interconnects or the parasitics of interconnects are computed and delays are estimated. But after we have done the routing right then we have the exact layout of the wires and then we can extract the parasitics and based on that we can do the timing analysis right which will be much more accurate than what we did during synthesis or in the earlier stages of the design flow. And once we do the timing analysis we may find some violations and then we need to fix the problem.

For example, we can improve the timing by say gate up sizing, buffering, automatic manual rerouting, automatic or manual rerouting we change some some routes to improve the the timing for a critical pass or we can do wire widening also if the routing resource is available to improve the timing. Then we can also think of improving the power dissipation in our design. For example, we can do gate downsizing for the pass where we were where the timing is not that much critical. So, there we can reduce the size of the gates and as such we can improve the power dissipation or decrease the power dissipation. But the but the the the increase in the delay will basically not impact the final QR of our design.

Also we can remove buffers on the pass where it is not, which is not required right. And

we can also do something known as we can change the cells based on its footprints right. So, now we say that in our library there will be an AND gate, two input AND gate and there will be another two input AND gate which has got the same footprint meaning that its layout is exactly the same right. And therefore, if we replace one with the other it will not matter on the layout right. But internally the transistors here may be say this way may be a low power transistor meaning that it may be having a vt of high high vt transistor or it can be high performance transistor it can contain high performance transistor where the v t or the threshold voltage of the transistor is low. And if the gate is not on the critical path we can use low power cells in there and we can save the power right.

Then we can also or after routing we should do signal integrity analysis and then fix issues for example, by increasing the distance between the wires right or maybe up sizing the victim driver right or down sizing the aggressor driver which is actually injecting noise into the other victim net. So, we can do that kind of optimization after we have done the routing right. In addition to the optimizations the two the after routing we need to also consider some other effects right or in other problems that can be there in our design.



One of them is antenna effect and we had discussed earlier that when the gate dielectric and the gate terminal of a transistor is directly connected to lower level metal M1 it can lead to antenna rule violation and this gate oxide can be prone to antenna effect induced gate damaged or what we call as also known as plasma induced gate damage right. So, this gate terminal is connected to M1 line and if this area is large or this length is large then there probably can be an antenna rule violation.

So, if there are antenna rule violations how can we fix it? So, one of the techniques is to say that we do the routing major portion of the routing in the upper layer of the metals right. For example, instead of doing the major portion of the routing in M1 layer we are using an upper M2 layer when the wire goes where most of the routing is done and we go from M1 to M2 layer very close to the gate right. So, we avoid long M1 lines directly connected to the gate terminal instead we use higher layer metal M2 for routing and the other way to fix this problem suppose this M2 layer was not available right that routing resource was already used up then probably we can use a jumper right. So, the M2 layer here is just used as a jumper. On the M1 layer it goes to the M2 layer and then to the M2 layer it goes down through a

via again to M1 layer. So, most of the routing is still done using the M1 layer, but we the there is no direct connection between this M1 layer segment this one and there is a gap right and therefore, the charge if gets if it gets accumulated it will get accumulated through on this part and the gate will not be taxed right.

This is how we can fix antenna rule violations after routing or even during routing we can make decisions for example, in this case we route it on M2 layer. So, we need not get an antenna or we avoid getting antenna rule violation. Now, after routing we can also look into some reliability issues right for example, reliability issues arising due to via defects. Now, what are these via defects right. So, we have discussed earlier that whenever there is say one metal and another metal we want to go from one metal to another metal for creating routes and we insert via right.

Now, these via are very small structures and these are prone to defects because of mask misalignment right for creating via we use multiple layers of mask and if there are some misalignment there can be defects in via right and also these via can be prone to thermal stresses and because of that via can fail right. So, if there can be open circuit defects in the via and if we are lucky then those problem will be caught during testing and then we will be having ill loss and if we are not that lucky it that problem will be found by the customers right as reliability problem may be say after few operations of the circuit there is a there comes a via defect and the circuit loses its functionality. So, that is more or that is a worse situation right. So, to avoid this kind of situation, a popular method is to add redundant via right.

So, instead of one via we had to say two via. So, even if one fails the other will provide the connection between two layers right. Now, how can we insert redundant via let us take an example. Suppose there is a metal layer M1 and M2 right and these are connected by this vias right. So, if there is one layer M1 going like this M2 in orthogonal direction and the via in between right.

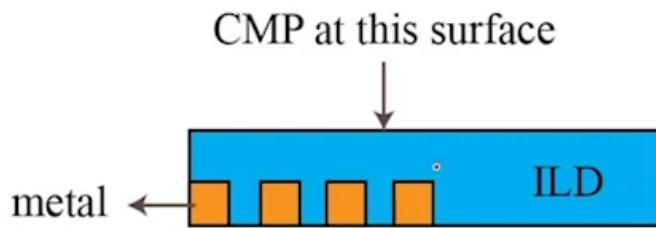
Now, if you want to insert another via where should be can we insert. So, there are two ways to do that. One is what is known as on track via on track via in this what we do is that we extend one of the lines in non preferred direction right. For example, this M2 was in horizontal direction in this direction we extend it slightly to a to a non. So, in this case we are extending this M2 layer along this direction for a short distance and making a connection right.

Similarly, we can do so for the M1 layer also. So, this is the M1 layer right this is an M1 layer. So, the M1 layer is below the M2 layer. So, it is going slightly. So, we are extending along non preferred direction right along the horizontal direction and then we

create a via right. So, these vias are on track because we are using the same track as that right. So, for short distances it is usually allowed that the routing is done in a non-preferred direction and those are basically known as jogs.

So, these for example, in this case this yellow line or M1 line is used as a jog here right in this from here to here right and in the other case this M2 line is used as jog from here to here right in this in this the preferred direction is horizontal and it is running in vertical direction M2 is running in vertical direction from here to here and it is used as jog right. We can also have off track via right we can have off track via. So, off track via what we do is that we extend this line right. So, we extend this line M1 along the preferred direction and extend the M2 line right along the non preferred direction. So, the preferred direction is this: for the M2 line it is going in this direction and here we have added a redundant via right.

Similarly, what we are doing is that we can move along this M2 in the preferred direction right and M1 along the non-preferred direction right. So, the preferred direction of M1 is vertical, but it is run for a small length along the non preferred direction and here we have this redundant right. So, this is how we can insert via to improve the or we insert redundant via to improve the reliability of our design right then we can also tackle some manufacturability issues post routing. So, what are these manufacturability issues? One of them comes because of CMP. So, remember that when we are discussing how we make interconnects right we in may we discussed dual demersing process right and in that process we need to do chemical mechanical polishing to planarize the interconnect and the in and the dielectrics right.



So, for planarizing we actually use mechanical polishing right and what happens is that if the roughness varies over the

layout then there can be irregular topography. What we mean by this is that suppose there this was our this was our side view of our chip right. So, in this region lots of metal is there right and in this region there is only inter-layer dielectric or SiO_2 right. Now, when we planarize it using chemical mechanical polishing, CMP is chemical mechanical polishing. So, when we do chemical mechanical polishing then SiO_2 being a softer material more planarization will be done in that area and will have an una more polishing will be done in this area and will have an uneven topography right.

Now, to if there are irregular topography then it can lead to yield loss and to fix this

problem what we do is that suppose in our layout there is metal layer metal layer widely used over this portion and in this region there was not enough metal layers used right and there it may create this problem of irregular topography in this region. If we do not, if we do not have an extra metal layer right and therefore, to solve this problem what we do is that we add dummy metal fills right. So, what we do is that if there is on the layout some portion of our layout is not having enough metal in that portion we fill extra metal which is not doing any functionality it is just to make the roughness of the chip evenly distributed throughout the chip right. It should not be that some areas are very soft, some areas are very hard and so on. To make the uniform hardness of a given layer we insert extra metal fills in the area where there is not enough metal.

But once we do this kind of metal filling we should be aware that the extra metals will actually change the capacitors because of the mutual coupling between the other interconnects. So, these extra metal fills that are not actually carrying any current, but they are existing in the layout layout and therefore, the potential profile of the interconnects that are in vicinity and they are carrying the signals that may change. And because of that the timing profile of our design can change right. So, we should be aware that if we are doing this dummy metal fill right dummy metal fills or these are also known as CMP fills because these are targeted for CMP or solving the problem related to CMP it is also known as CMP fills. So, whenever we have done CMP fills in our design then we should again do the extraction of the capacitance, especially the mutual capacitances and the and also carry out timing analysis again right to ensure that after dummy metal fill insertion our timing is still closed right.

Now, if you want to look into more on the topics that we discussed in this lecture you can refer to this book right. Now, to summarize what we have done in this lecture is that we have looked into global routing, detailed routing and some optimization that we can do after the route is right. Now, once we have done the routing and say we have done the timing analysis our design is clean and the implementation process is kind of finished right. Now, our design, our layout is created, all the entities are placed, the routing has been done, all the required interconnections have been made over the layout and in that sense our implementation phase has completed and the design is now ready to be taped out right. The layout has been created for our design and it is ready for tape out, but can we do the tape out right. So, we can do it, but since this is the last opportunity to check whether there is any problem in our design we should be making it or we should be doubly sure that our design is perfect or if it is behaving correctly right.

So, after we have done this implementation we are done with this implementation phase we do another round of physical verification and sign off before doing the tape out. So,

in the next lecture we will be looking into physical verification and sign off steps. Thank you very much.