

VLSI Design Flow: RTL to GDS

Dr. Sneh Saurabh

**Department of Electronics and Communication Engineering
IIIT-Delhi**

Lecture 38 Power Optimization

Hello everybody, welcome to the course VLSI Design Flow RTL to GDS. This is the 30th lecture. In this lecture, we will be discussing power optimization. Now in a VLSI design flow, we carry out power optimizations at various levels of abstraction. For example, we carry out power optimization at the system level, at the level of RTL, at the gate level or even in the physical design. So basically the power optimization techniques are important throughout the VLSI design flow.

And in this lecture, we will be discussing or having just an overview of power driven optimization techniques. So, let us first look into what are various strategies for reducing power dissipation. Now to devise strategies for reducing power dissipation, we must look into what factors the power dissipation depends on. And then we can target our techniques to those factors which are affecting the power dissipation.

Now in the last lecture, in the previous lecture, we had looked into power analysis and we saw various components of power dissipation. We understood that there are two major components of power dissipation, the dynamic power dissipation and the static power dissipation. So this is the dynamic power dissipation and this is the static power dissipation. Now dynamic power dissipation is also divided into two parts. The first one is this component, the first component is related to switching power dissipation and the second one is related to the short circuit power dissipation.

So if we want to reduce power dissipation, we must target these components of the power dissipation. Now from this equation, we can infer or deduce that on what factors the power dissipation occurs. So the most important factor that appears from this equation is the supply voltage. Why? Because supply voltage appears in all these three components and in the switching power it is square of the supply voltage. So we expect that if we reduce the supply voltage, the power dissipation will reduce substantially, not only the switching power dissipation and the dynamic power dissipation, but also the static power dissipation that can also reduce by reducing the supply voltage.

Now what are the other factors? The other factors which can or other circuit parameters which we can change to improve the or reduce the power dissipation are of the one of the things that we can infer from this equation is that if we reduce the frequency of the clock, if we reduce the frequency of the clock at which our circuit is operating, then the power dissipation can be reduced. Another thing is that if we can reduce the activity of the signal, then the switching power can reduce appreciably. And also if we can reduce the load capacitances, so these, so the power dissipation in charging or power dissipation required in charging and discharging the capacitor, those will also reduce. So power saving strategies can be devised based on this observation from what we derive or these deductions that we made from this equation and we can devise strategies to reduce the power dissipation by reducing supply voltage or clock frequency or activity or the load capacitance. Now, depending on the level of abstraction of a design and what knobs we have in controlling these parameters, these per circuit parameters, we can apply different techniques.

$$P_{tot} = C_L V_{DD}^2 \alpha f_{clk} + V_{DD} I_{SC} + V_{DD} I_{leak}$$

where,

- V_{DD} = supply voltage
- C_L = load capacitance
- f_{clk} = frequency of the clock in the circuit
- α = activity of the signal

So in this lecture, let us look into a few techniques that we can apply to our design to reduce the power dissipation. The first power reduction technique that we will see is called dynamic voltage frequency scaling. So this power savings technique basically utilizes the strategy of reducing supply voltage and clock frequency dynamically. So basing based on the requirement of the workload, the speed and the supply voltage of the circuit is varied. So this kind of power saving strategy is very popular for microprocessors.

So in microprocessors, there is a wide variation in the workload, meaning that the full speed of the microprocessor is utilized only for a few tasks and for very small time duration. What we mean is that given a microprocessor, we do not always require that

the microprocessor should run at the highest rated frequency all the time. Only for a few tasks and for a small duration, it may need to work at a high frequency. And at other periods, the deadlines can be met even at low speed. And when the speed is low, we can also decrease the supply voltage and as such we can reduce the energy consumption and the power dissipation significantly.

So to give an illustration, let us suppose that there is a processor which has been designed to work at 1.2 gigahertz and 1.2 volts. And give a task, a given task can be performed at this clock frequency. Let us assume that the given task can be completed in 10 milliseconds.

Now let us assume that this task is not very timing critical. We can afford it to be running slowly. Now if the workload requirement is such that this task can be run for a longer duration or can be completed in a longer time, then we can reduce the supply voltage and also the clock frequency. So let us assume that we can reduce the clock frequency and the supply voltage to half. So we reduce the clock frequency from 1.

2 gigahertz to 600 megahertz, that is we reduced it by half and also reduced the supply voltage from 1.2 volts to 0.6 volts. Note that when we reduce the supply voltage, suppose in a circuit we reduce the supply voltage, then the logic case for example, CMOS inverter or AND gate, NOR gate and so on for them the delay will increase, because when we reduce the supply voltage the logic gate will require more time to charge and discharge the capacitors. And therefore, the delay in general will increase.

So when we reduce the supply voltage and therefore, when we reduce the supply voltage and delay increases, then remember from the discussion on static timing analysis that if the time period should be greater than the delay of the critical path to some approximation. So if this delay increases, we need to increase the time period of our circuit otherwise we will get a setup violation. To avoid setup violation and because of increased delay of the critical path we need to increase the time period and since we want to increase the time period, we want it to reduce the frequency. Therefore when we reduce the supply voltage from 1.

2 volts to say 1.0.6 volt, it is required that we also reduce the clock frequency. So from what we say in this case we are assuming that it is reduced by half that is from 1.2 gigahertz to 600 megahertz. Now given this reduction in the clock frequency now the task will complete in double the original time because the number of clock cycles that is required to complete a job will remain the same but the time period has increased twice. And therefore it will take two times more time to complete the task.

So initially it was completing in say 10 millisecond now it will complete in 20 millisecond. Now as a result of this reduction in the frequency and the supply voltage how much power dissipation do we say. We can compute that. So we will reduce the switching power dissipation by 1 by 8. Why because from the earlier discussion we have seen that switching power is dependent on V_{DD} as V_{DD}^2 and on clock frequency.

So we have reduced the V_{DD} or supply voltage by half and half because it is V_{DD}^2 and we have reduced the clock frequency by half and therefore half. So we are getting 1 by 8th reduction in the power dissipation. Now what about the energy we save? So the energy consumption will be reduced by 1 by 4. That is not as large, it is slightly less than what we have as what we expect that gain will have in power dissipation. Now why energy dissipation is a savings in energy dissipation is less in this case.

The reason is that now the time to complete the job has increased from 10 millisecond to 20 millisecond. So earlier and remember that energy is equal to power in time. If time goes high. So in this case the job needs to run for a longer time. And therefore the energy consumption will not will not show the same kind of saving as the saving in the power dissipation.

But still we have one fourth times the original energy savings. So we are still significantly saving energy now. Now we are not discussing the architecture of the implementation of dynamic voltage frequency scaling. So at the circuit level we need to have some hooks to estimate the workload requirement or what the workload requirement of a circuit is. And based on that if the workload requirement is lower then the frequency and the voltage should be reduced and workload shoots up then the voltage and the frequency again needs to be ramped up accordingly.

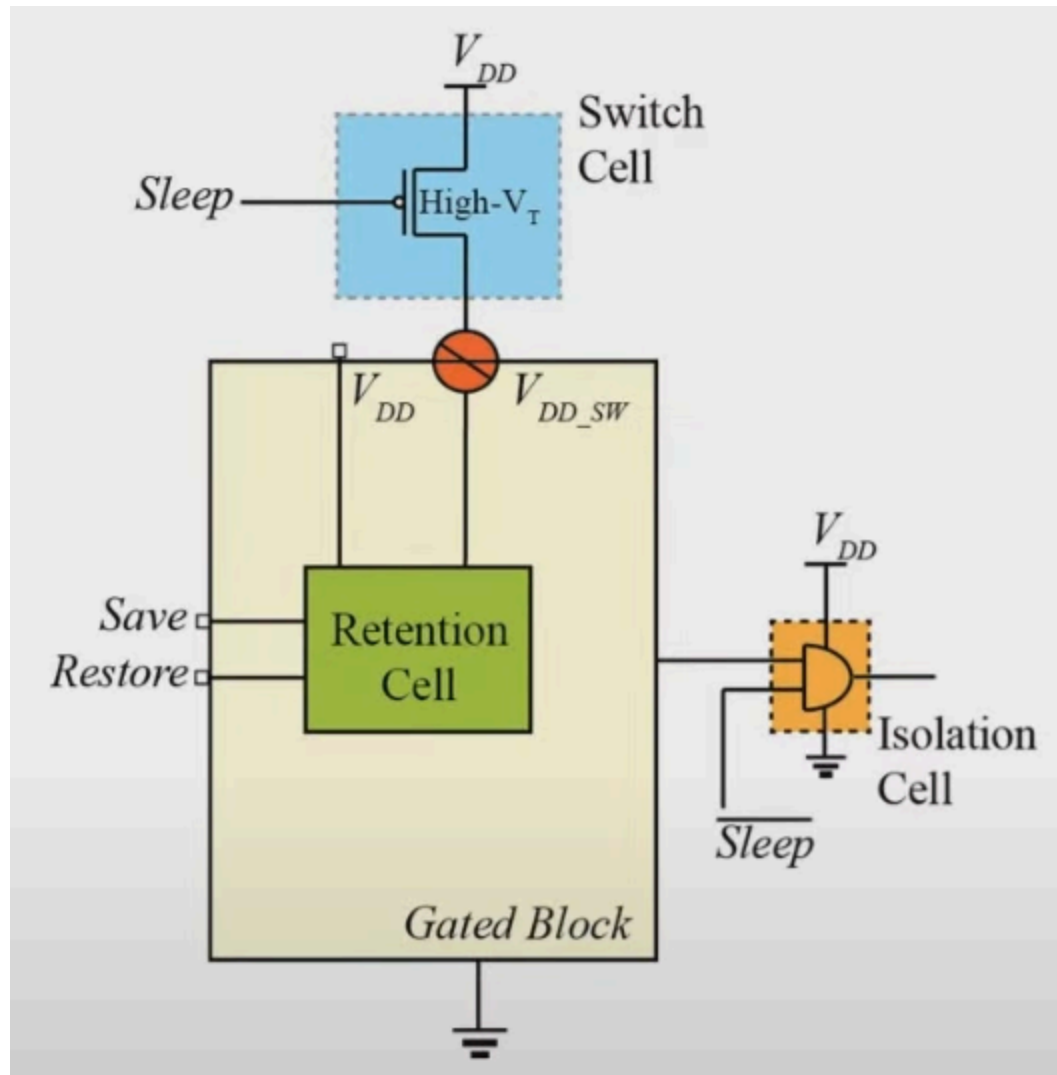
So some feedback loop has to be there in our circuit between the workload and the dynamic voltage frequency scale. So in this course we are not going into those circuit based techniques. So now let us look into another type of power optimization technique that is power gate. So we have discussed that if we reduce the voltage if we reduce the voltage of a circuit then the power dissipation decreases and switching power dissipation decreases quadratically per full time. So if we reduce the voltage we expect that the power dissipation will reduce.

Now what is the limiting case of reducing the voltage? So we can reduce from say 1 volt to 0.9, 0.8 and so on till what value we can go. So the limiting case is of course 0 if we can make the power voltage the the supply voltage as 0 meaning that we are disconnecting the power supply and that is what is being achieved by the power gate.

So in power gating utilizes the strategy of eliminating power supply voltage meaning making $V_{DD} = 0$. So in the in power power gating basically switches off the power supply for a block suppose there is a circuit in the circuit in our circuit there will be multiple blocks. Now the blocks which are not actively computing for some time we can switch it off completely. We simply switch off the disconnect the voltage source from that block and as such we will be able to save both the static component and the dynamic component. So a static component still requires V_{DD} to go for the current leakage to happen and a static power dissipation to be exhibited.

So if we simply switch off the power supply then static power dissipation will also decrease and of course the dynamic power dissipation will also reduce. So however this power gating technique requires careful circuit design we need to be very very careful about how we power down our circuit, how we power up our circuit and whether the states in our circuit are correct when it powers up when we power up again. So these kinds of complications are there which need to be handled during the circuit design and we need to also insert some special circuit elements to implement the power gating techniques. So let us look into a few such circuit elements. So the circuit elements which are required in a typical power gating implementation are switching cell, switch cell, then retention cell and isolations.

So now what are these cells? So let us take an example of the implementation and then understand what is the role of these these circuit elements.



So this is a schematic of an implementation in which this block, the gated block that is shown, is being power gated. Now to power the gate what we do is that we do not correctly connect VDD directly to this gated block. So we connect it through a switch cell. The switch cell is there which is which is a high VT cell which is a high VT cell because in the when when we want that this is this is the switching cell itself do not come consume large static power and to reduce the static power dissipation within the switching cell we keep it as a high VT.

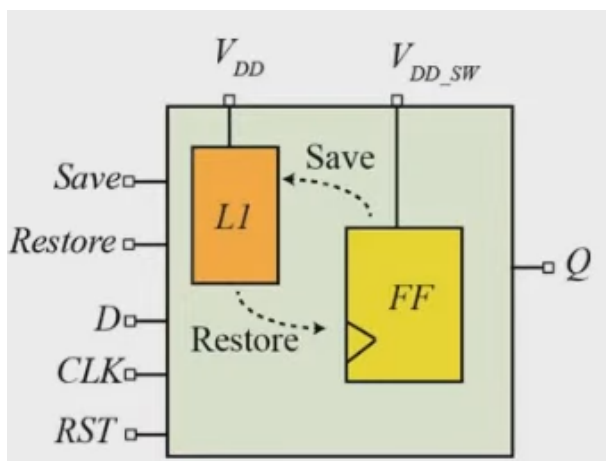
So in this case we have assumed this switching cell to be a PMOS and it receives a signal which is called sleep. Now what when sleep will be equal to 0 in the normal condition sleep will be 0. So the circuit is not that this block gated block is not sleeping meaning that it is doing some active computation. So suppose sleep is 0 in when sleep is equal to 0 then the PMOS will be switched on. So the VDD voltage that is here will be applied at this point and this again goes to all our circuit elements: the combination

circuit elements, sequential circuit elements and so on.

So when sleep is equal to 0 those circuit elements will be operating normally. So in this for designing a switch cell it should be designed such that when sleep is equal to 0 or PMOS is turned on in that case the resistance shown by this switch cell should be very very small. So that the voltage drops from VDD to the point in the gated block that is very very small. We want if we have say 1.

1 volt here we ideally want that 1.1 volt should appear at this point with minimal with minimal resistance being exhibited by the switching cell. Now this is the case for sleep is equal to 0. Now what happens when sleep is equal to 1 sleep is equal to 1. Now if the sleep is equal to 1 then what happens is that this switch cell is switched off because it is a PMOS. So this will be switched off and therefore VDD and this point will get disconnected and if these are disconnected this is basically power gated.

So the VDD voltage will not come to this gated block and the circuit elements most of the circuit elements in this gated block will be powered by VDD underscore SW and those circuit elements will be actually switched off when sleep is equal to sleep is equal to 1. Now in addition to the switch cell we have another kind of cells which are there in our gated block that are known as retention cells. So when we take our circuit to a sleep mode after we awake it or make it again active at that point we want that the values of the registers for some of the registers in the gated block restores the original value that it was that had gone before the sleep state. So to have that kind of restoration capability in the implementation we decide that in our gated block some of the cells will not be ordinary cells but will be retention cells. Now what is a retention cell? Retention cell is similar to a flip flop but the thing is that it receives 2 voltage sources.



So the important thing is that it receives 2 voltage sources, one is the switched voltage source and the other is always on voltage source that is VDD is directly applied. So the VDD that is shown coming to this gated block is not going to all the circuit elements it is going to only the retention cell there are few retention cells with the gated block which will

be powered by 2 supplies one is the always on supply that is VDD and other is the switched mode. Now before taking our circuit to a sleep mode what we do is that we give a safe signal safe signal to this retention retention cell. So this is a retention cell. In the retention cell there is a pin which is safe. Now if we assert this pin safe what happens is that whatever the value was there in the flip flop inside this retention cell that is returned to a latch that is returned to a latch and this latch is always on power.

So it will retain the old value but what will happen to the flip flops output value it can take any arbitrary value because VDD underscores SW that will be floating once we go into the sleep state. So the state of FF will be unknown but before going to the sleep state L1 has stored or saved the value which was there of at the flip flop and later on when we want to move out of the sleep state to the active state what happens is that we restore that we assert the signal restore at the at the at the retention cell and it restores the value it writes the value to this flip flop and the and the flip flop again gets the previous value which was before going into the sleep state. So this is important because when when when we switch off the power supply or when the when the sleep is asserted if the flip flop is the VDD of the flip flop is is is removed the flip flop goes into an unknown state and once we were back to from the sleep state we want that flip flop starts from the state which was earlier and that is why we need to convert the flip flops which want to retain the state to retentions. So that is the second most important in circuit elements. The third important circuit element that is required for implementing power gating strategy is the isolation cell.

Now why do we require an isolation cell because when the gated block goes into a sleep state the values will be driven by some gate for example suppose this was an inverter. Now this inverter power supply has been cut off. So the value at the output of this inverter can be hanging, it can be 0 it can be 1 or something in between. Now if we connect this output to some other circuit elements in the fan out and its input become is something between 0 and VDD not a good not a good 0 or not a good VDD or it is not deterministic then what happens is that this logic gate can draw a large amount of power because the transistors within this this logic gate that may that will see an input voltage which is somewhere between 0 and VDD it may be say VDD by 2 and if it is in VDD by 2 it can draw a large amount of short circuit current and this this this circuit can get damaged. And to protect that what we do is that once the circuit goes into a sleep state then we isolate our circuit using an isolation cell.

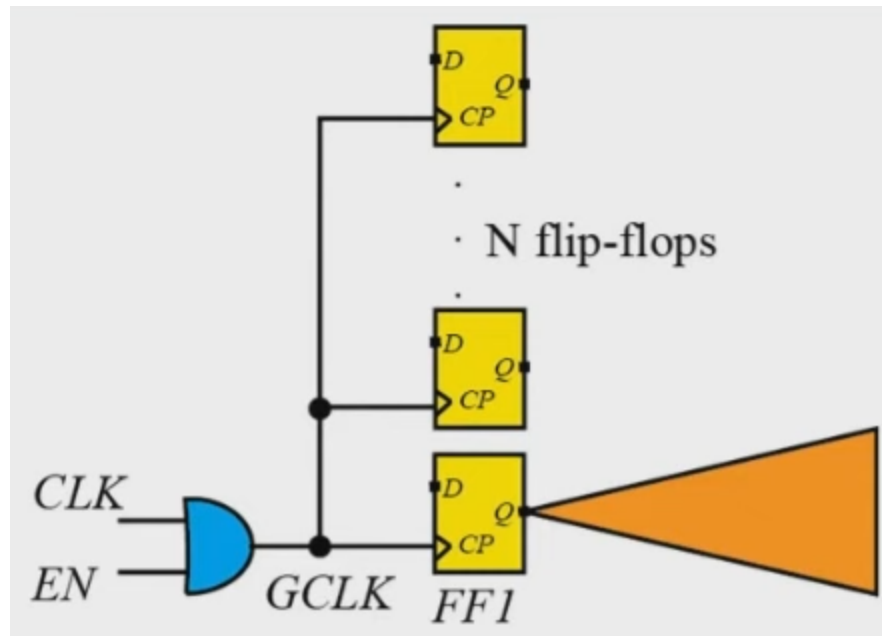
Now a simplistic isolation cell can be just an AND gate. So whenever sleep is equal to 0 this sleep bar will become 1 and whatever the value at the output that will be output of the isolation cell. But whenever sleep is equal to 1 sleep bar is equal to 0 and therefore the output of the isolation cell will be tied to a voltage of 0 not anything between 0 and 1

not say for example VDD by 2 and therefore it will it will it will help the circuits in which is in the fan out of this gated block to understand to to to not draw a large amount of current and under and have a deterministic value at their inputs. Now let us look into another power saving strategy and that is known as clock gate. So clock gating utilizes the strategy of reducing activity in the clock network .

Now assume that there is a set of n flip flops that captures new data conditionally . So if that condition is false there is a group of flip flops which are reading data at the D pin only under certain conditions . So if that condition goes false then what we can do we can say that the clock we can simply disable the clock signal. We can shut off the clock signal for the case when the condition is false . So it will save power in charging and discharging capacitors in the clock network and also the circuit elements which are inside the flip flop .

For example if we consider a flip flop and the clock signal is going to this clock clock pin of the flip flop internally there are master slave latches and there are circuit elements on the clock path of the flip flops internally . Now if we shut off power sorry shut off the clock signal the transitions within those circuit elements in the clock network or in the clock path within the clock within the flip flop those will also not charge and discharge and therefore will save power not only in the clock network but also inside the flip flops . So this is the basic motivation of using the clock or employing clock gating gating strategy. So in this case the logic synthesis tool needs to find the enabling condition of the clock and based on that it will say that okay only when this condition is true then the clock is allowed to propagate along the clock path . Now how can we implement this clock gating strategy?

So it may seem that we can simply implement it using an AND gate . So there is a clock and there is an AND enabled signal and a clock signal . Whenever enable is equal to 1 clock will propagate and this N flip flops will be will be will be capturing the value and when enable is equal to 0 then the transition in clock signal will not reach the clock pin of the flip flops and it will be disabling the transition in the in the clock network and inside the flip flops . So this looks like a simple implementation but this implementation has a flop . What is that flop? So the simple AND of enable and clock will lead to glitch and how it can lead to glitch? You can understand that.

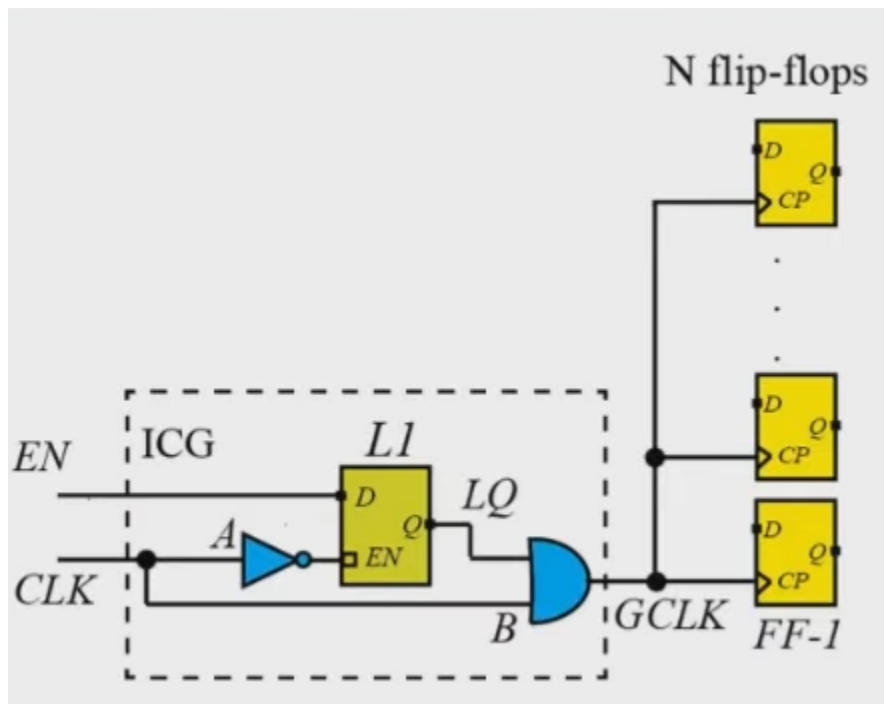


So suppose this was the clock signal, this was the clock signal and the enable signal made a transition from 0 to 1 when the clock was high. In this small duration thus the clock the enable signal made a transition from 0 to 1. Now what will happen to this GCLK? In the GCLK it is a simple AND of and of in clock and enable. So in this small duration the clock and enable signal both are 1. So we will have a 1 value but then suddenly the clock goes down and the clock goes down and therefore since the GCLK is or gated clock is the AND of the clock and enable the GCLK will also come down with the clock.

Now we have a kind of glitch: glitch means a signal which is changing sharply for a small duration of time. Now if this GCLK feeds this flip-flop what will be the problem. There can be several problems. One of the problems that can be found directly on this FF1 is that there are constraints of clock period, there are constraints of clock period which are defined for the flip-flops. Now if that is and there are constraints of minimum pulse width.

So those kinds of or kind of constraints can be violated by this glitch signal and therefore these flip-flops can fail. The other problem is that if there is a data path going from this FF1 to some other flip-flop FF2. Now how much time duration did it require or what time duration does this path have to propagate the signal from Q of this FF1 to the D of FF2. It is a very short duration because see that the rising at the edge of the clause occurs here and the next rising edge occurs here. And therefore the time duration for the data to go from FF1 to FF2 is just this much.

It is not the complete clock cycle . It is only a small duration of the clock cycle and therefore we can therefore what can happen is that even if we do, we have done static timing analysis for this path and this path passes the setup requirement because it was considering a complete clock cycle for the data to go from FF1 to FF2. Later on because of the glitch, a small duration will be there within which the data has to go from FF1 to FF2 and as a result the path will fail . It will fail the setup requirement and therefore we cannot use this kind of implementation of gated clock . So this is not a good implementation. Then how can we implement a gated clock? So typically we use an integrated clock gate or ICG .



Now what does an ICG contain? So an ICG contains a latch which is enabled by the clock signal through an inverter . So whenever the clock is 0 then this latch gets enabled and the clock signal goes to the AND gate similar to the earlier implementation but the enable signal goes through the latch to the negative triggered latch and then it reaches the AND gate . Now how does this latch help us in avoiding the glitch condition? Let us understand . Now suppose this was the clock signal and this one was the clock signal and the enable signal came very close to the falling edge of the clock . In the earlier implementation it would have led to glitch but in this case what happens is that the enable so this value *LQ* changes only when the clock is low because there is an inverter on the path .

So in this duration LQ does not change; it is held constant. Only in this whenever the clock is 0 the value of LQ can change for example it can change at this point and at this point and it can with when clock is equal to 1 the LQ is stable. So latch negative sense is negative sensitive and it allows enable to propagate only when clock is low only when the clock is low then only enable can come to LQ and another thing is that when clock is high whenever clock is high then only this output can change if the clock is low the output GCLK will be 0. Now GCLK can become 1 only when clock is high but when clock is high what happens this latch L1 became gets disabled why because it is being enabled through an inverter when clock is equal to 1 this becomes 0 and therefore this latch gets disabled meaning that it is producing LQ at the output LQ it is producing a stable value or the it is it is producing a value that was last last it is not change. So whenever the clock signal is 1 the output LQ will be stable and therefore the glitch cannot propagate to GCLK.

So these integrated clock gates are implemented and these are actually provided in technology libraries also so that synthesis tools can directly use them in the net list. Now we have looked into resizing techniques in the earlier lectures for improving the timing of our design. Now we can also use resizing for improving the power dissipation or reducing the power dissipation. So how can we use it? We can use smaller shells in the non-critical path of the circuit. So if there are some portions of our circuit which are very non-critical non-timing critical there we can use smaller cells with smaller cells. As a result what will happen is the capacitance shown by those cells will be lesser and we remember and we know that the switching power depends on the capacitance.

So if we reduce the capacitance of the cells then the driver will need to charge and discharge less capacitance and therefore the power dissipated will be or switching power dissipated will be low. In this lecture we have seen a few or looked into some of the techniques of power of reducing power dissipation. If you want to know more you can refer to this book. To summarize in this lecture we have looked into various strategies to reduce the power dissipation. Specifically we looked into the technique of a dynamic voltage and frequency scaling. We looked into the technique of power gating, clock gating and also resizing of the cells in our circuit.

Now with this we have completed the topics of logic synthesis. So in the next lecture onward we will be looking into an important topic of VLSI design flow that is design for test or DFT. Thank you very much.