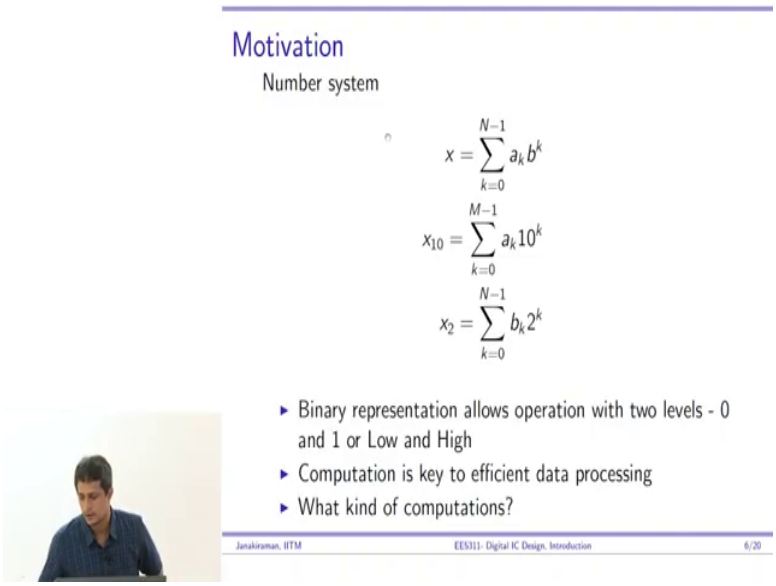
Digital IC Design
Prof. Janakiraman Viraraghavan
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture – 01
Introduction

So good morning and welcome to this course on Digital IC Design.

(Refer Slide Time: 00:20)



So, what is the you know motivation behind this course right. What leads us to study this course, what makes it possible for us to do such a course? So, mathematically we are given this you know we for example, if you want to represent a number in base 10 number system. If, you want to represent a number in base 10 right, then you know that you can write it as summation a k 10 power k right. And, all numbers can be uniquely represented by this

representation. For example; the roman numerals cannot guarantee this right. Even representing any arbitrary number in roman numerals is not possible right.

So, the number system with the invention of 0 really sort of enabled this for us. Now, there is nothing sacrosanct about 10 right, you can do this in any arbitrary base b. So, what are the values that the numbers a k can take if you represent it in arbitrary base b? 0 to b minus 1 right.

So, nothing special about this I can take this b all the way down to 2 right. And, that gives us the binary number system, where I can represent X 2, which is basically the same number X in binary number system, as b k 2 power k. Of course, the N number of bits or the number of symbols needed to represent the same value will be much more in binary compared to the decimal number system. N will be much greater than M right, because I have only since it is going in powers of 2 as opposed to powers of 10.

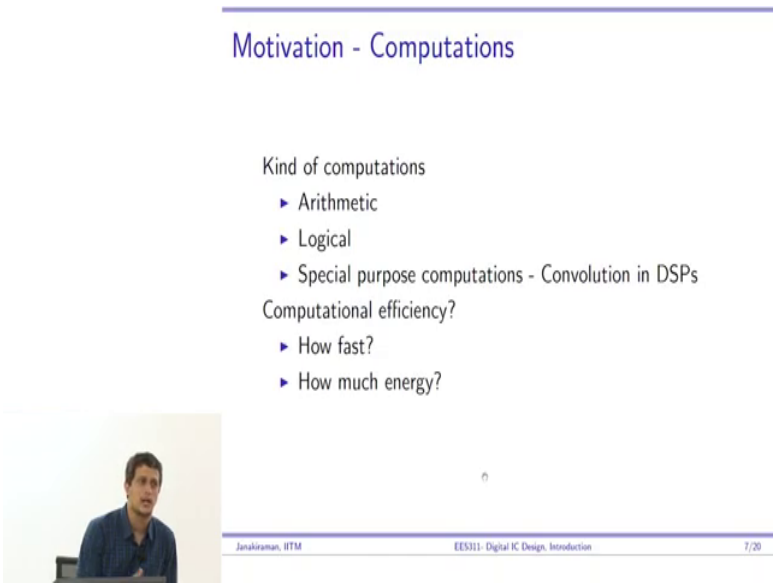
So, once you are given this guarantee mathematically, it is possible to go ahead and do computations in a consistent manner right and achieve all that we need. So, it turns out that if b k can take only 2 values, then those 2 values can be called logic 0 and logic 1. So, this is the motivation for general digital logic right. And, computation is really the key to efficient data processing right.

So, for example, their neural networks has existed for a long time right, it is been used for recognition and artificial intelligence in general right, but what is it that has made deep neural networks, you know all of a sudden very famous now. What changed from neural networks 2 decades ago to deep neural networks now?

It is basically the computation power right, you have computers that can evaluate at monstrous speeds now right. And, therefore, you are able to train really large networks right, 150 layers with arbitrary number of hidden units in each layer, fully connected convolution whatever it is you are able to train these networks right.

Of course, they are going to consume a lot of power, but you are at least able to do it because it is computation speed has gone up significantly in the last 2 decades right. And, you have also been able to pack so, many more transistors in a single microprocessor today right.

(Refer Slide Time: 03:45)



So, what are these kind of computations? These computations are essentially not surprisingly arithmetic and logical right. Everything you can just boil it down to a simple addition, multiplication, subtraction or logical maybe you can do ands, bitwise ands, ors, xors and so on right.

So, if you look at your cryptography it is heavily using the concept of an xor right, you can use you parity and stuff like that. So, these things are ultimately they just boil down to some

logical operation, which happens bitwise or arithmetic where you do something bitwise and then start propagating the carry and so on across bits right.

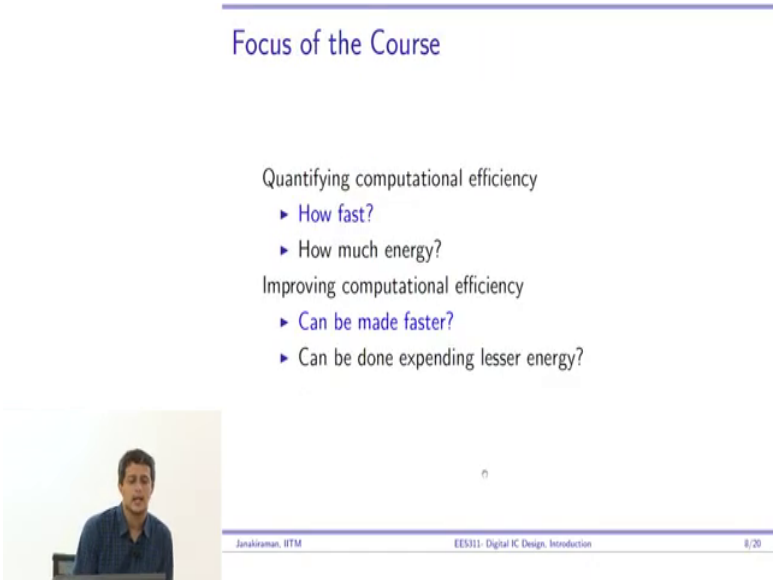
In some cases like DSPs you can also do some special instructions like convolution that is the heart of a DSP. So, in that case you can make that a single instruction right. So, what I mean by single instruction is in a microprocessor, you can say con of something right, which means in 1 cycle once the convolution is done it will proceed to the next instruction.

So, those atomic instructions is what I am talking about here right. Those are the building blocks for your microprocessor and naturally the question to ask is how good is the computational efficiency? Ok. And, that is what we want to improve, but then what is computational efficiency? So, it turns out that this can be quantified in 2 ways; 1 is how fast, second is how much energy? Right.

So, it is important to just keep in mind that energy is different from power right. If, I have a server that is connected to a power outlet, I do not care too much for efficient energy processing because I have sufficient energy coming from the power outlet, I just have to keep the power under control right. So, that it does not go out of control in and heat up the processor that is all.

But on the other hand if you are doing something on your mobile phone, where you are charging on the outlet once and then you are roaming around for the rest of the day, then you have to ensure that you do your operations in an energy efficient manner right. Because, ultimately every coulomb of charge that is or nanocoulomb of charge that is coming out of your battery is going to drain the battery out little by little by little right.

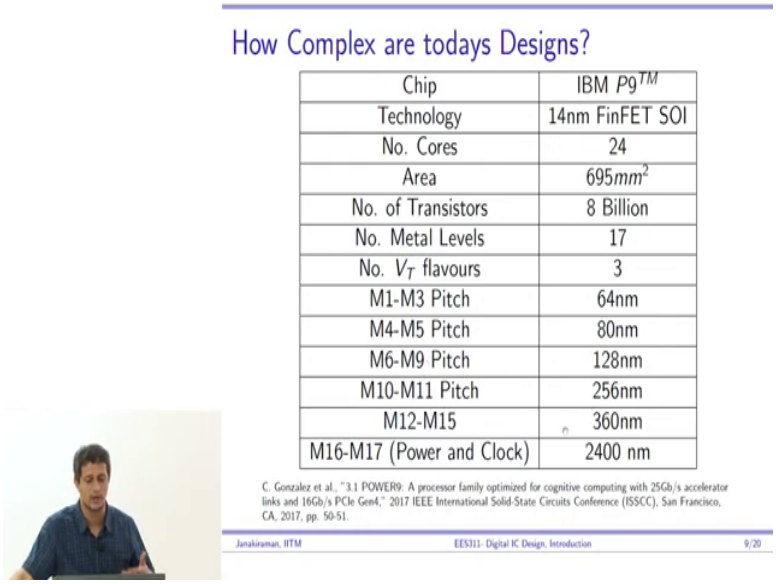
(Refer Slide Time: 06:18)



So, how much energy is also very important, but unfortunately in this course we will not have time to go into all of those details right. There is lot to do just with delays right. How fast and just answering the question can it be made faster itself is going to consume us in this course.

So, therefore, I will not have time to touch upon how much energy and stuff like that I will focus only on delay, how slow or how fast is a particular circuit, how can it be made faster? Ok. The can it be done expending lesser energy, I will allude to this a little bit in when I talk about energy and power, but that will be about it in this course.

(Refer Slide Time: 07:02)



So, before we jump into the details of this course, I think it is important to get a feel for what a microprocessor chip looks like in today's technology right. So, today's technology by the way in production is 14 nanometer in 2019 right, 7 nanometer is in technology development stage there is really no product that has come out in mass production right. Whereas, 14 nanometer has come out in a market, what I show here is basically the IBMs power 9 processor ok, this is a server chip from IBM. So, the focus of this chip is entirely on performance it has to do it very fast ok.

So, the here we are not too worried about constraining power or energy of this particular microprocessor, it is you are just tuning things to do it really fast ok. What is the technology it is the 14 nanometer FinFET SOI technology, SOI is Silicon on Insulator. If, time permits I will allude to that at some point. FinFET again is was introduced only in 14 nanometer up to 22 nanometer the transistors where just scale down regularly ok. And, they were able to

achieve their targets without much problem, but beyond a point beyond 22 nanometer it became difficult to contain the leakage current right, keeping the on current this thing.

So, just a order of magnitude question if the on current is let us say 1 unit. What will be the off current? Like order of magnitude, it is about a thousandth of the power 10 power minus 3, that is what a technology can guarantee beyond that it is very difficult right. So, they found it very difficult to achieve this beyond 22 nanometer with the old transistor technology, they had to move to FinFET.

And, that is basically some 3 D structure I will try to just tell you at least what that structure looks like at some point when I do the transistor, but let us just take it for now in 14 nanometer to handle leakage current, you had to go to a radically new technology ok, which also makes technology development very very very expensive by the way ok. The number of FABs have come down to a very small number, because it is not profitable for E anyone and everyone to just run a fab these days.

The investment you have to make in technology to get that small gain of you know performance or power is extremely high, you know it was not that high about a couple of generations ago, coming from 90 to 65 or 65 to 45 not much of a difference, but now it is very very very expensive ok. What is the area of this microprocessor? It's about 695 mm square, it is about 25 mm by 25 mm right. The size of a 10 rupee coin almost, that is how small it is and what does this contain? 8 billion transistors ok.

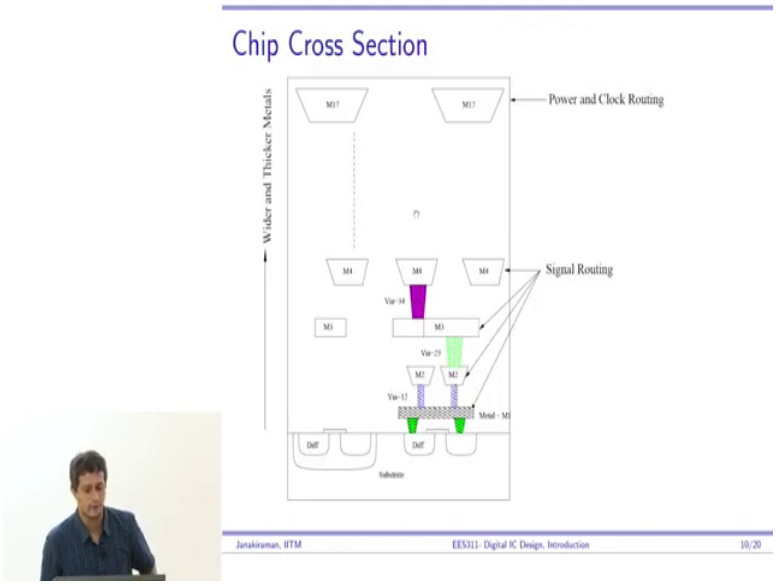
So, now if the on current of a transistor is a micro amp and the off current as I told you is a 1000 right, nano amp why is the off current even a cause for concern, it is a nano amp right yeah.

Student: 8.

Yeah, you have 8 billion of these transistors leaking a nanoamp each which means very quickly we are just leaking amperes of current right. And, what does leaking mean, it means that my device is not in use it is not doing anything right. And, in spite of that I am losing

amperes of current from my power supply, my battery is getting drained that fast right or it is drawing that much power and therefore, it is a big problem.

(Refer Slide Time: 11:08)



So, the number of metals, what is a metal and let me just give you a quick illustration here of what the manufacturing process looks like right, below you have this substrate, then you have the transistors ok. This Diff stands for diffusion ok, we will we will come to what those diffusions are (Refer Time: 11:24) in some. So, this is an N MOS transistor and this is the P MOS transistor right, we will define all that more formally in some time right. And, then you have a set of metals.

So, what is this transistor? That is called the front end of line that is the first thing that you manufacture to do something, then you start building metals over and above that. This metal region right the region of manufacturing that is basically called the back end of line. So, in the

back end of line they have 17 levels of metal in that microprocessor. What does that mean metal M 1? For example, is shown here, it is connected to this green via, that is called a via, which is the contact between a metal and a lower level metal or the diffusion ok.

This is a contact to the diffusion the green color right. And, M 2 is the metal level 2 which basically is connecting metal is connected to metal 1 through this blue via. What is a via it is either a cylindrical structure or a frustum that of metal, which connects the higher level of metal to a lower level of metal.

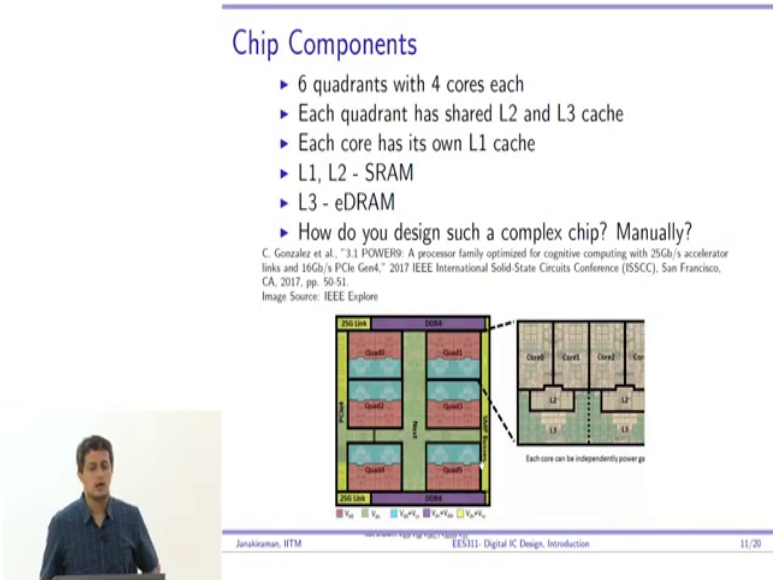
Similarly, you have M 3 M 4 all the way to M 17. So, with this you have about 17 levels of metal right. And, what you notice here is the pitch basically which is talking about the spacing between metals right is 64 nanometer for M 1 to M 3 right, which means a spacing the width of the metal also has to be less only then the spacing can also be less right, then you have M 4 to M 5 it is 80 nanometer, M 6 128 nanometer and so on.

So, what happens is the lower level metals right, because you are closest to the transistor will be used to connect transistors sitting near each other right. So, what happens is; what happens is you have 1 transistor sitting here, something sitting nearby I do not want to go all the way to the top level of the metal and come down I need to go as quickly as possible. So, that has to happen through the lowest level of metal. And, therefore, since these are short distance communication I do not have to worry about the resistance so much. So, I can make the width lesser right and pack more wires also at the lower levels.

But, as I go higher I would have to drive for example, the power lines or the clock lines, these actually drive a lot of current it can drive amperes of current right. Now, in that case I need to ensure that the IR drop on that wire is not very high. And, therefore, you make those wires very wide, but of course, you cannot pack as many of them right.

So, lower level of metals for nearby communication as you go higher it is basically for far communication right. And, the power and clock are basically routed at the topmost levels, which is the most the widest metal and also very very low resistance.

(Refer Slide Time: 14:34)



So, what does this look like right? People figured out that just scaling a single microprocessor right, a single core and improving it is frequency forever is not giving enough returns. So, you had to make it multi core, you know you have to distribute your computing and all that stuffs. So, that is that is what is happened over many years in the microprocessor technology.

And, therefore, in this chip they have 6 quadrants with 4 cores each right, by the way all the details are in this paper if you want to read it; 6 quadrants with 4 cores each and each quadrant has a shared L 2, L 3 cache right.

So, people figured out that going outside to the dram and fetching data is extremely slow, you cannot keep doing that therefore, you needed to have some faster access, but of course, you cannot have all the access to I mean access to all the data at 1 shot you start doing something known as caching right. And, that is why you have a hierarchy of caches called L 1, L 2, L 3,

L 4 and then the memory main memory. This basically grows in size right, but becomes slower and slower as I go the higher level of cache.

So, with that they have shared L 2 and L 3. So, L 2 is shared between core 0, core 1 and another L 2 between core 2 and core 3, and then you have a shared L 3 cache, which is you know again shared between the cores as shown here ok, these are memories. So, it is interesting to note that I have equal amount of area dedicated to memory as I do for computational power today right. So, what does this tell you? It does not it says that just improving computational speed is not enough, I need to have access to the data right. And, especially in today's world I have this concept of big data and data analytics where the volume of data is so; huge that there is no way I can get all of it into my processor.

So, you need fast memory access and then good computational speed. So, that is why you see half the chip actually dedicated just to the caches. And, what is not shown here is in each core you have an L 1 cache. So, it is actually more than 50 percent of the area that is allocated to memories in terms of caches, in today's microprocessors.

So, a small portion of this is actually the compute element which is the ALU very small portion right, there are other things like you know the serial links then the DDR 4 interfaces and so on. There are a lot of other things that needed that you need to communicate to the outside world right, but a small area in this is actually your computational part which is doing the a ALU job or some other computation as needed ok.

Now, the question is how do I design such a chip? It has 8 billion transistors, it has lot of interconnect right. In the interest of time I will not do that calculation, the if I just lay out all the interconnects the metals in 1 straight line right. It turns out that the length of that interconnect can anyone guess who is not there in the last class, what is the order of magnitude of this number, millimeters, nanometers, what 25 mm by 25 mm chip ok.

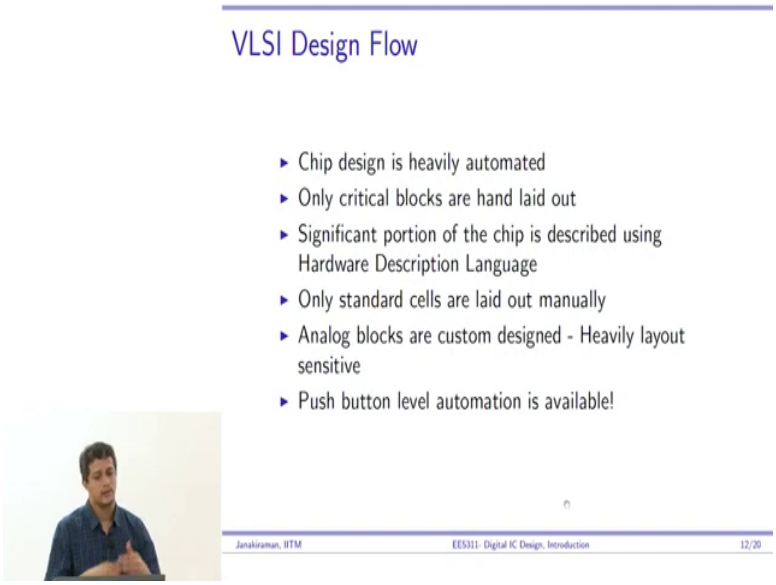
And, I have you know this kind of a thing let us say 2500 nanometer take M 7 M 70 2500 nanometer which ok, let us assume that the width also is 2500 nanometer. What do you think

this net distance would be just order of magnitude I do not care whether it is a 11.5 or 12.3, I need the unit there.

Some guess I need a guess otherwise I am not going forward, like do you think it will be nanometers it will be like meters or what yeah micrometers ok. So, I urge you to go do this calculation just figure out if I have a wire which is 2500 nanometer, spaced apart by again 2500 nanometer right; that means, every 5000 nanometers, I have 1 wire ok, running across the chip, which means the length of the wire is again 2500 nanometer.

Just calculate the total length, it is an astonishing number the total length is 11 kilometers, in an earlier chip. So, this chip is actually higher it is in the order of tens of kilometers. So, how do you even design such a chip, which has 8 billion transistors, which has 11 kilometers of interconnect you cannot possibly do this manually right.

(Refer Slide Time: 19:53)

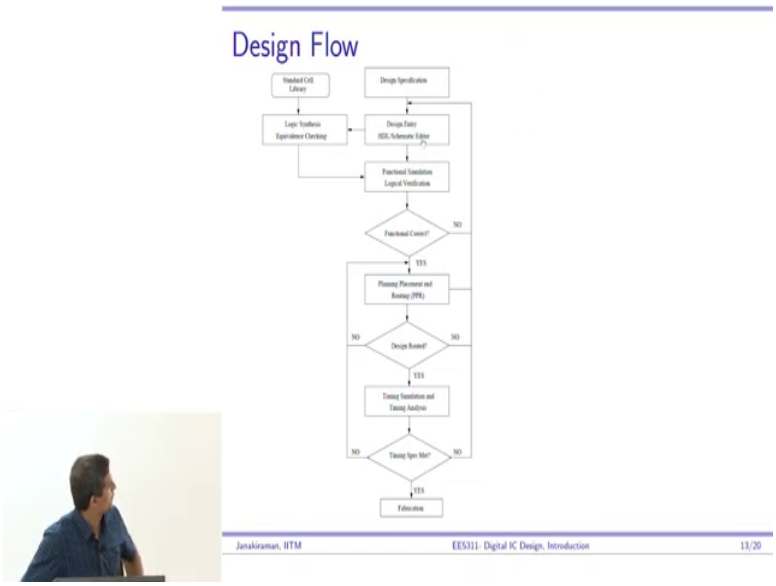


And, therefore, what do I do it turns out that the VLSI design flow is heavily automated right, lot of this is just you write something known as Verilog code, then you can synthesize and get your layout right. Layout is basically the which you learn in this course right. The eventual thing that you can shift to the foundry and say manufacture this right, it is a set of masks that you will make.

Only the critical blocks are hand laid out, which in this course you will do only hand layout you will not do synthesis. So, significant portion of the chip is described using hardware description languages like Verilog or so on right. Only some select things called standard cells I tell you what they are laid out manually analog blocks are usually custom designed right.

So, it is like a push button now, I give all my settings say this is my speed of the microprocessor that I need 2 Gigahertz I need to finish this operation in this time, go ahead and synthesize my circuit, I will just give it Verilog code it will do it will do it right.

(Refer Slide Time: 21:00)



In fact, this is the design flow I know it may not be visible to the last bench, but all I want you to notice here you can go see the details later, you give a design specification you do some other things right, but at every point there is a decision box. What is the decision box it says is the functionality correct? If, yes then you proceed to the next state otherwise you go back and redo something.

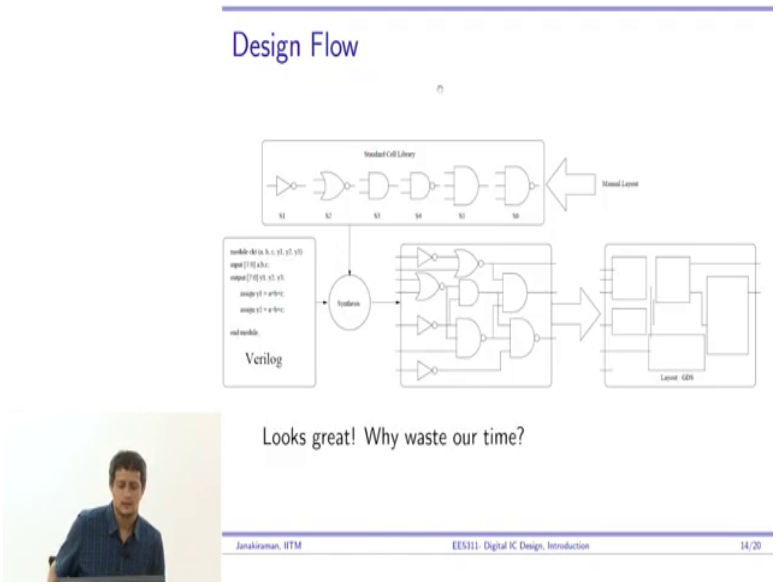
Then you check is the design routed correctly if yes you do something, otherwise you go back and do something else right timing spec is it met if yes then you complete the design otherwise you go back and do something. So, the point is this is not a straight you know one unidirectional flow chart, where I give something and I am guaranteed to get an output this loop may never converge. And, that is the reason you are even doing this course, we are going

into the nuts and bolts of building this microprocessor elements right, specifically the ALU and we are trying to figure out how much time it takes, how we can make it faster? Right.

So, the point is if the design was so, automated is it not a waste of time to even study this course, because everything you push a button and then you are done right. It turns out that this loop may not converge and when it does not converge as a designer you need to figure out why and you need to be able to fix it ultimately. So, to do that to understand the synthesis report saying look, I have thrown my hands up I cannot synthesize the tool will just tell you that, you cannot just give up saying you know I cannot get this circuit working right. If that is the case robots would replace your job right, why would I need a human being there right if it is an automated decision like that.

So, then your job is to go in and figure out what went wrong, which is the guy who's actually causing a violation, which block inside it is actually consuming all the time can it be made faster. Now, to answer all those questions you need to do this course. Digital IC design is a very fundamental digital circuit design course, where we are going to deal only with transistors and that is what I am going to think.

(Refer Slide Time: 23:19)

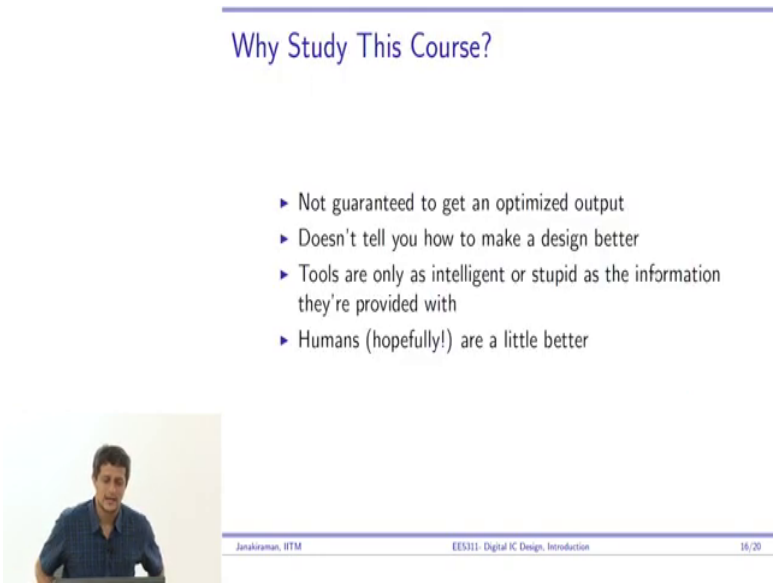


So, if you look at the design flow you have something called a Verilog code here right, then you have something called a standard cell. A standard cell is basically a set of gates, which are designed have been hand laid out and given to you. And, say that look these are my universal gates. Now, you have to synthesize your circuit just with this.

For example, I could have just a Nand gate here. And, say synthesize my circuit it is possible right. It is a universal gate I can synthesize any logic circuit with that right, but of course, it is not efficient. So, you give a few more gates and say that ok, I have inverter Nor, Nand and so on. And, then you say that I want to make my circuit. So, the synthesis block will take this Verilog and generate a circuit like this.

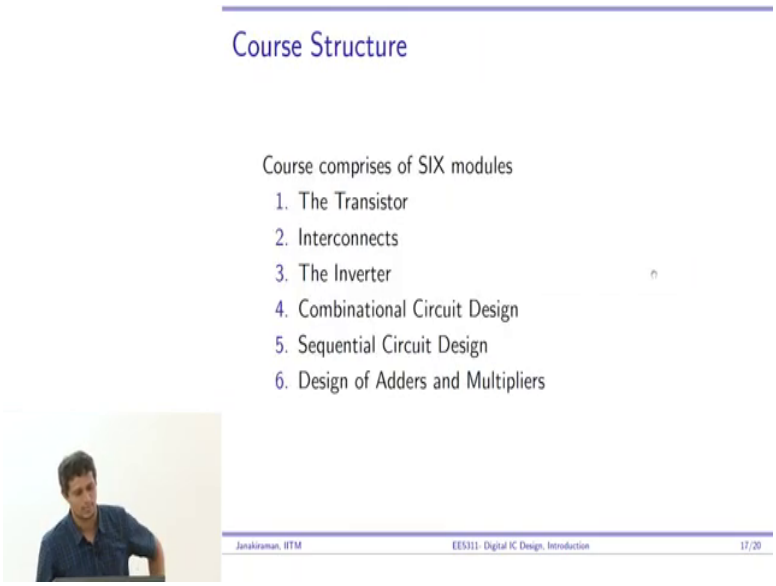
In schematic right, in schematic basically it is just a netlist it is a text which just tells you how to connect the transistors and the gates together, then you do a place and route and then you get the layout that is what happens ok.

(Refer Slide Time: 24:21)



So, as I told you the loop is not guaranteed to converge you need to know what went wrong right. And, humans are; obviously, more equipped to make informed decisions and that is the reason your jobs also will be secured at least for some time until AI really comes and sort of takes that over.

(Refer Slide Time: 24:44)

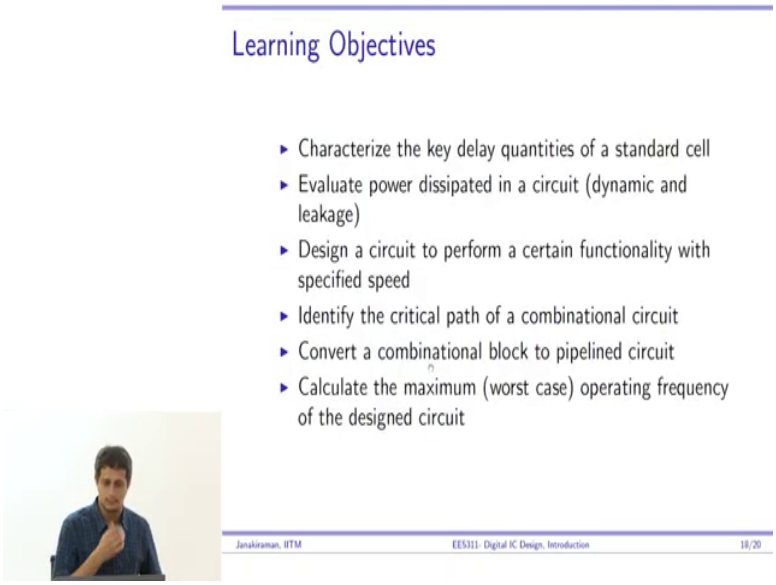


So, what is the course structure the course structure is essentially this comprises of 6 modules ok. The transistor, interconnects, the inverter, combinational circuit, design sequential circuit design and design of adders and multipliers, this last block is basically the thing which is going to deal with the ALU actually how do I design my multiplier, how do I design my multiply accumulator right, how do I design my adders and so on right, but I need to do all of this before I can get there ok.

So, I have divided into 6 modules, this will take you know just this is more or less an introduction if I unfortunately I cannot make it a prerequisite, because lot of students come here first semester they have to take this course right. Otherwise, I would like to spend more time on the other modules rather than the transistor, but I will spend enough time to give you the necessary abstraction to move ahead in this course ok.

Then interconnects will be very short it mainly I want to talk about the Elmore delay model ok, then the inverter the combinational circuit design, this module 4 and 5 will actually be bulk of the course ok. Combinational circuit design is I give you any arbitrary logic and I want you to design a transistor level circuit that will implement that logic in a given time, that is what we do there ok. Sequential circuit is basically how do I design flip flops, how do I design latches and so on.

(Refer Slide Time: 26:19)

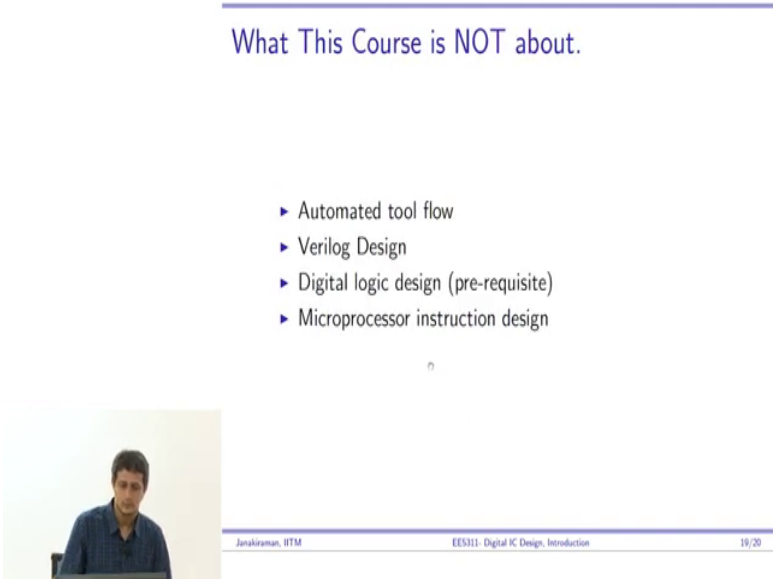


These are the learning objectives after you are done with this course, you should be able to characterize the key delay quantities of a standard cell, you should be able to evaluate the power dissipated in the circuit both dynamic and leakage that is only, what I that is the only thing that I will have time for with respect to power and energy how to evaluate these numbers? Ok.

Then, design a circuit to perform a certain functionality with specified speed ok. Identify the critical path of the combinational circuit, then convert a combinational block to a pipeline circuit. Pipelining is another fantastic concept where you can improve the throughput of your system. How do you do that right, where do you go and put your flops, how do you make sure that the circuit throughput can double triple and so on right. Then, you should be able to calculate the maximum worst case operating frequency of a designed circuit ok.

These are the learning objectives after we are done with this course we will come back and revisit this to make sure that you know you are you are comfortable answering all these questions.

(Refer Slide Time: 27:25)



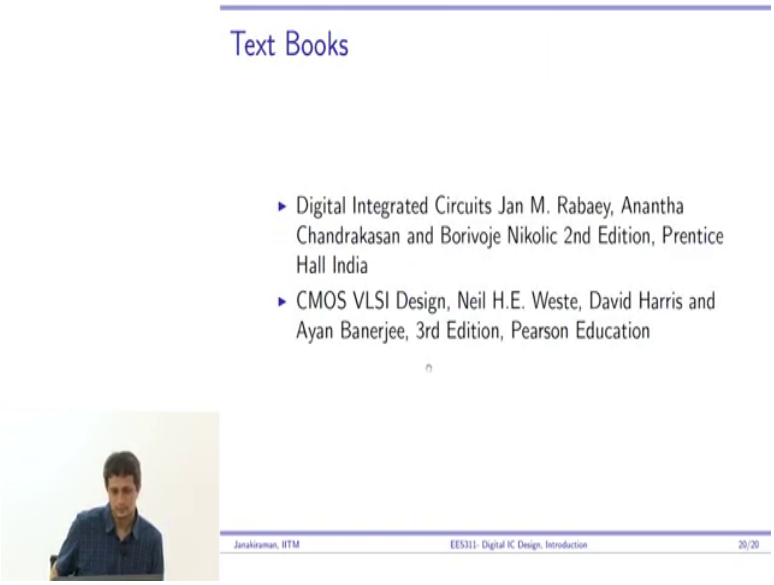
It is also good to know upfront what this course is not about. This course is not about automated tool flow, I am not going to take you through that synthesis flow that is another

course, which Professor Nagendra is offering, but I do not think that is open to too many students, because we have limitations on the licenses and so on.

Verilog design it is definitely not about Verilog design, you will not write a single line of Verilog code in this course ok. It is not about digital design in the conventional industry sense, this is digital circuit design. Digital logic design is a prerequisite ok, you should know what a multiplexer is what a decoder is all those thing how to solve, simplify a given Boolean expression all that you should know de Morgans Laws, these are prerequisites.

So, I am not going to cover any of that in this course. And, it is not about microprocessor instruction design, where I try to find out what best new instruction I should introduce in the microprocessor to make it better it is not about that right. We are just looking at certain addition, multiple, multiplication, operations and at the nuts and bolts level of the circuit ok.

(Refer Slide Time: 28:43)



These are your textbooks, Jan Rabaey and Anantha Chandrakasan you have a book called digital integrated circuits all of these on is on the website by the way do not worry about it. CMOS VLSI Design right, Neil, Weste and David Harris, so, some modules I have taken from you know the Rabaey, some I have taken from Weste and Harris.

So, I will be providing whatever slides I use right for to you through the website, but do not bank on it please do take notes, because it is it you will not be able to follow that ok, that is more for my reference rather than your reference.