Introduction to Semiconductor Devices

Prof. Dr. Naresh Kumar Emani

Indian Institute of Technology - Hyderabad

Module No # 13

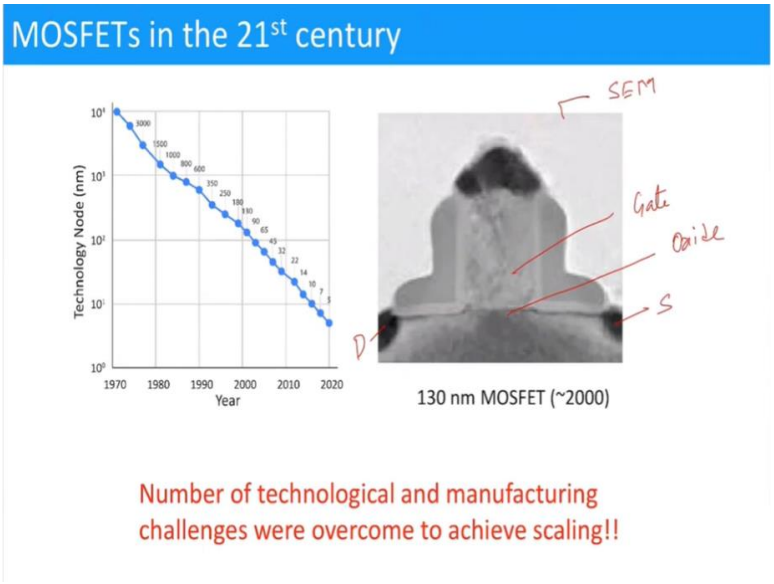
Lecture No # 67

MOSFET's in the 21st Century

This document is intended to accompany the lecture videos of the course "Introduction to Semiconductor Devices" offered by Dr. Naresh Emani on the NPTEL platform. It has been our effort to remove ambiguities and make the document readable. However, there may be some inadvertent errors. The reader is advised to refer to the original lecture video if he/she needs any clarification.

Hello everyone welcome back in the last few weeks we have understood how MOSFET works we have analyzed various situations and then we have developed I believe in good intutive insight into how MOSFET works? Today I would like to show a glimpse of what there is in the most modern semiconductor devices that are being manufactured currently.

(Refer Slide Time: 00:39)



So we have seen this graph already we have seen how the technology notes have been evolving. And whatever MOSFET theory we have developed that is primary applicable to the long channel MOSFET and we have discussed the short channel corrections in the last lecture. So this is how a MOSFET at the end of the twentieth century would look like 130 nanometers is roughly where in 2000 we were there.

So the MOSFET looks like this where in this is a gate and then there is a oxide in between here and there is a channel I mean this is how physically MOSFET like this is what is known as the SEM picture scanning electron Micrograph SEM. So you can use electron to take this image because it is a nano scale device you cannot use a normal regular microscope we need a special type of microscope called as scanning electron microscope.

So with that you take an image you can see the various structures and there is a great amount of detail and these edges at you are seeing on the source in the drain. So this is how a MOSFET have look like at the end of the twentieth century. And you did of course earlier it good shot. And you did of course short channel effects and then whatever we discuss is all applicable to this MOSFET.

So, what happen in the last 20 years and this is usually not discussed in any text books and I believe the textbooks are currently being rewritten so to account for this there been tremendous amount of progress and it is my hope that I can give you some glimpse to that. So the material that I am going to show you today is taken from various presentations mainly by scientist at Intel at various conferences.

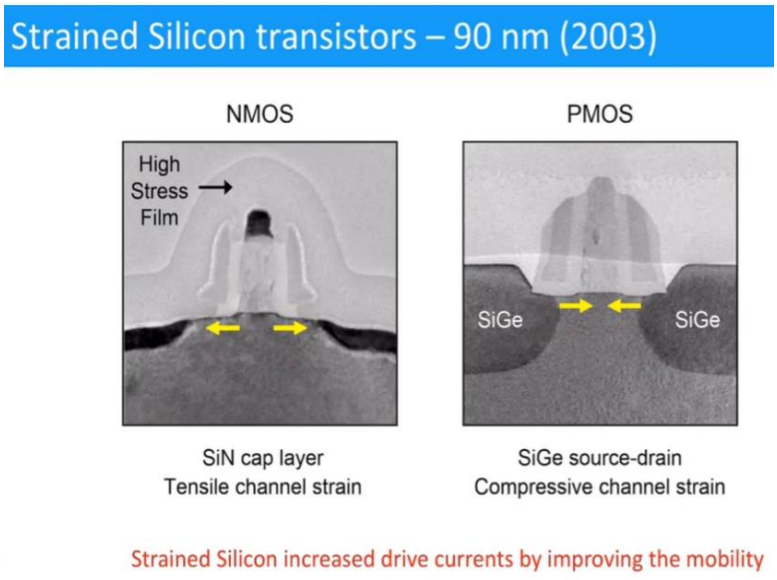
I will, give the reference to the conferences in the bottom so if you are interested it could always go back to the presentations. And I believe with a amount of knowledge that you are gained over semiconductor devices in this course. You will be able to understand what we are talking about.

So if you want to know the just development you should go back and look at presented at conferences.

So, in the last 20 years at great number of technical challenges to overcome and that has enabled us to do continue the scaling process and what are these? I will only give you the glimpse of it is not my aim to discuss these in great detail. It is probably you know it should make you feel that I did not know that is the reaction I am looking for. And I am not really trying to explain the in depth what is happening that would require an entirely different course of depth.

So, this was 130 nanometers you know this is 2000 right at the turn of the century and scientist where already realizing that. You know the reason being the scaling of oxide and you know scaling the transistor improve the drive currents and of course pack in and more transistors. But then at the turn into century already scientist realize that it is running into some problems because we could not scale supply voltages and mobility for also getting affected. So, they wanted to improve the drive current that is on current right.

(Refer Slide Time: 03:52)



And one of the great inventions that was required was to introduce what is known as strained transistors was introduced in 2003. And what we mean by strain? Essentially we have the regular MOSFET and now on top of it I mean you understood that when we studied the basics of lattices mobility and all that. We saw that the mobility of a semi- conductor is dependent on the bank structure and band structure is infulenced by the lattice spacing

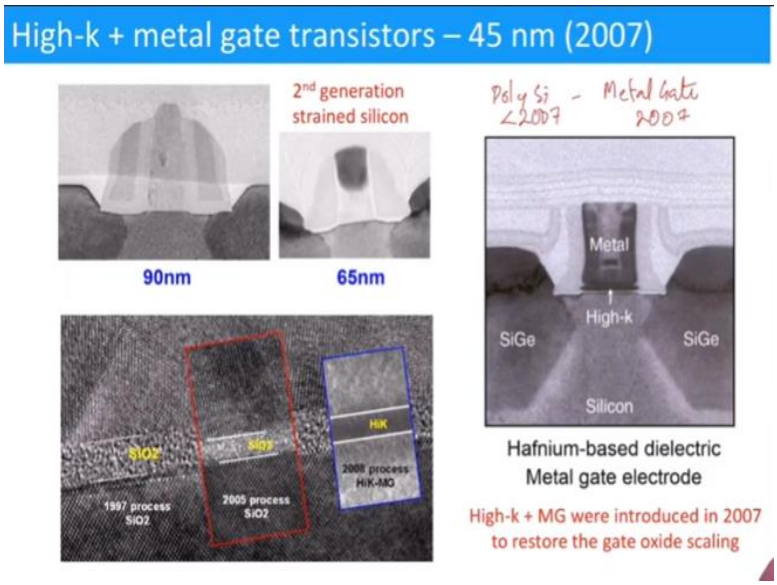
So, scientist want to increase the drive current in a transistor and we know that drive current in the transistor is simply given by now it is proportional to mobility right we have seen that expression for it. And can we play around mobility? You can of course play around with the vg and all that but in addition somehow you can improve the mobility of the silicon itself right. That will be great and a way do that is actually to apply strain.

Strain essentially means either you know I applied tensile strain which mean I stress the latical little bit or apply compressive strain by means I compress a lattice little bit. Let us say the lattice constitute is 5.4 angstroms I can stress it and make 5.5 angstorms . If I do that I am applying tensile strain and if I compress it then I make my lattice constant 5.3 then applying compressive strain. So this can be done by incorporating by external materials so in 2003 Intel introduces idea of you know strain silicon of course this were research even before that.

But 2003 was when they made it into fabrication you know the actual devices. Before that of course a lot of research has done by various researches universities and so on. So now how did they incorporate this tensile strain? They essentially deposited a you know silicon nitride capping film on top of the transistor which cost the transistor to stretch a little bit and that is why we show that you know NMOS transistor is actually have a strain which is stretched mobility was increased by that.

And similarly in PMOS transistor they introduced silicon germanium source drains which were actually costing the channel to get compressed and that would increase the mobility again. So why and all that of course you know you have to take a lot of detail that you will do if you are doing a research. But for now I will just want to introduce this is the concept of strain that means your lattice is manipulated to improve the drain current and this was done in 2003 and in worked you know sorry couple of years.

(Refer Slide Time: 06:20)



So by 2005 already they make the transistors smaller so this was the transistor which was looking in the 2003. And then by 2005 they had the second generation of strained technologies and they managed to improve the drained currents further, But then a big challenge came up when we introduced a MOS capacitor we talked about oxide thickness and we said that oxide thickness is actually coming down to nanometer scale and that was a major problem.

And one of the ways to work over is that we have to introduce a High-k metal gate or rather High-k dielectric. So this is what was done in 2007 so roughly 13 years back so this is basically 1997 the silicon dioxide is basically substantially thick and this is what is known as TEM picture transmission electron microscope wherein you are shining electrons and then into the lattice and

then seeing the transmission and what you see here the regular structure is all silicon and this is silicon dioxide which is a amorphous and then poly-silicon or you know on the top.

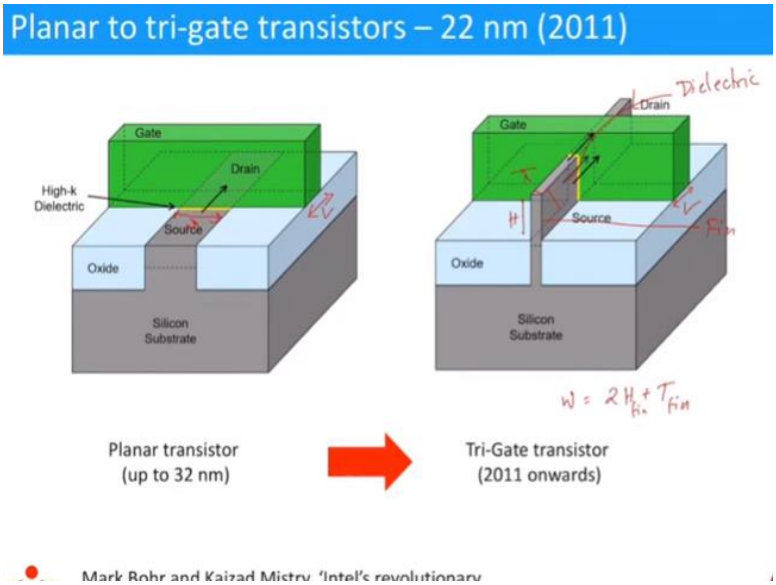
This is in 1997 the oxide was substantially thick but by the time we came to 2005 it was like 0.8 nanometers and because of that there was lot of quantum tunneling happening which was increasing the leakage current and that can impossible to you know the live with. Then of course people knew that we introduce High-k dielectrics we can increase the physical thickness of the oxide and thereby reduce the tunneling current and finally this was introduced in year 2007 when Intel introduces High-k metal gate.

And they introduces this High-k oxide by then the if you introduces a high-K dielectric it was not enough to introduce High-k because they were lot of other electrical issue and because of that even the gate have to replaced. So before 2007 the gate was poly silicon but in 2007 it became metal gate it replace both dielectric and the gate. So that is why this is called as high k metal gate process in 2007 and this is less than 2007.

Remember originate in 1960 and 70 it was you know metal gate but then we change over to policy metal gate for a adjustment of the VT and so on. But and also some lithography related issues and finally in 2007 we are back to the metal gate. And this was introduced to essentially use to restore that you know scaling of oxide so you could continue to reduce the thickness of the dielectric and then increase a drive current this was done in 2007.

So we call these transistors as High-k metal gate transistors and this is 45 nanometer technology. So if you bought a computers year on 2008 or so you probably would have bought a computer with high-k metal gate. So this was okay it worked or again one more generation they managed to work.

(Refer Slide Time: 09:07)



But then it eventually became you know difficult to improve the drive currents any further. The reason was that if you look at the traditional technology you know the one that was discussed so far. We call them as planar transistors the reason that called as planner transistors is that now we are familiar with the structure of transistor is just rotated. So basically you have the source you have the drain and you have the substrate and this is the field oxide that we talked about or you know shallow oxide solution whatever it is.

It is the isolation between transistors so where is your channel length? You know channel length is here this is the length of the channel and this is the width of the transistor. So you see that this is a gate and under the channel there is inversion channel that formed with the dielectric you have shown in yellow. And below the dielectric you have just you know square of inversion level right or a rectangular form just under the dielectric you will find the inversion layer.

And essentially all the current conduction happening in the plane that is why it called as a planar transistor and this is what was used till upto 2011. But the challenge was because increasingly difficult to control the inversion charge with the gate voltage. So Intel came with a really

revolutionary technology of course this was also studied or sometime but it was finally introduced in 2011.

So the idea was that we will not have a planar technology anymore but will have a out of plane silicon so now the silicon substrate is actually cominga up like this as a thin slab. We call it a fin and now a days now it is more popularly known as the fin is thin strip of silicon just coming up, and then you surround that with dielectric here this is yellow again right this is dielectric then surrounded by the gate.

So now where will be the channel from? The channel forms in the 3 dimensions it will be in this line here this is a channel under the oxide there will be a piece of channel here and then the piece of channel here right. So this is a 3 dimensional transistor or we call it tri-gate transistor it effectively is like a 3 gates you know so what will the width of this device if you think about it usually related to the traditional width of the transistor what will it be?

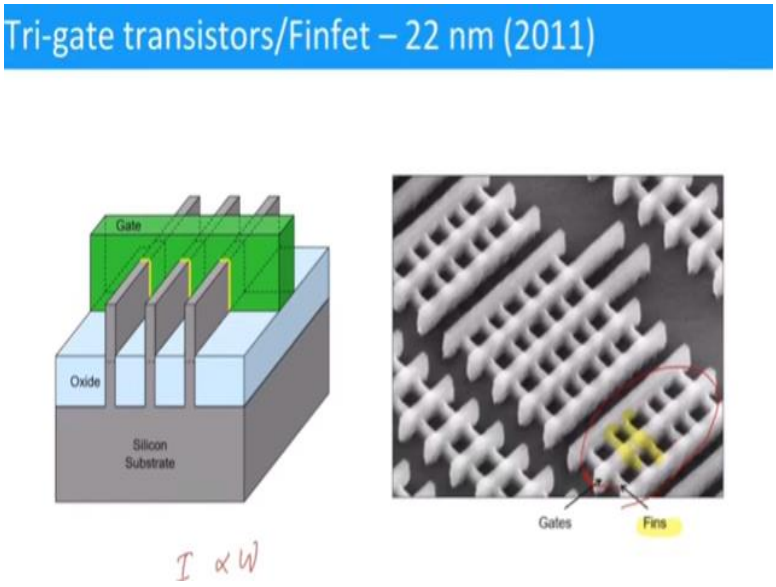
So we can already we can clearly see that this is the length because the distance between the source and the drain that the silicon source is here silicon drain is here of course it is doped it is not going to be the same silicon. And under the channel there is no doping that is why it is there so now and then it is opposite dope P type dope and then N + source and drains. So this is the length what will be the width of this transistor? The width will be let us say this is height of the fin is H let us say this is height of the fin pitch and this is a width of this is thickness of the film T.

The total width would be will be 2 times height + the thickness of the film T fin and H fin right so this is the width of the transistor. So you see what happen we are able to make the dimensions smaller by introducing these fins because if you try to go smaller and smaller then lot of practical challenges. So they could not make the transistors smaller and smaller so the finally they went out

of the plane in the third dimension and that is known as the finFET or tri gate transistor technology that Intel introduce in year 2011.

And of course there is lot of detail and if you can go back and just look for this presentation you will find some details and with your background in semiconductor devices now you should be understand some of that. So now listen what it is and the beautiful thing is you do not need to have simple one fin you can have more fins.

(Refer Slide Time: 12:52)

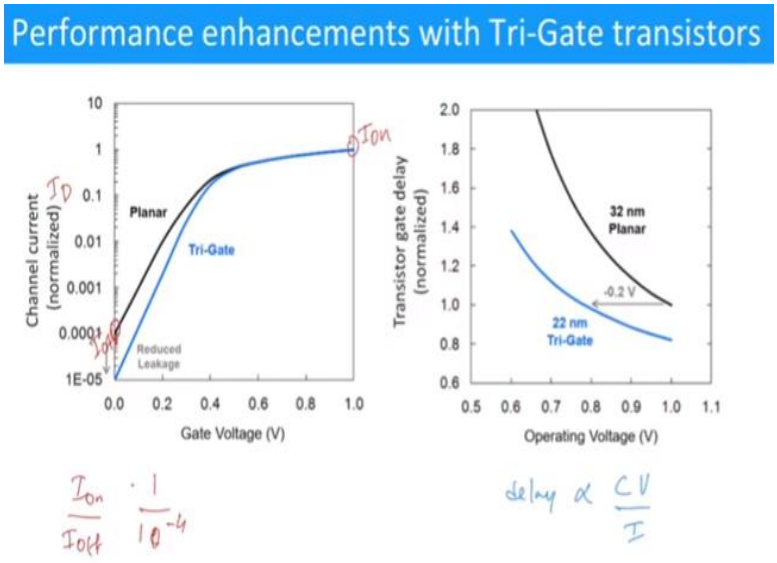


So you can have multiple fins like this and you can have oxide like this so that way you are able to increase the width of your transistor. The physical area lateral area is thick very you know small but you can have much larger width so you know already what happens when we increase a width I is proportional to width of the transistor. So if you increase the width we are increasing the drive current on current of your transistor and that is why you are actually able to you know charge your capacity faster and discharge the capacity faster.

So you are actually improving the speed of a transistor so this is what schematically looks like and practically if you look at it this is how the transistors look like just you know gate till upto the gate there is going to be lot of things on the top of it. But here you see these are the fins these things are the fins of silicon and across that you have a gate. So if you look at this as one transistor you can count how many fins are there 1, 2, 3, 4, 5, 6 , 6 fins are there, there is a certain height certain thickness.

We can estimate what is width and of course gate length will be nearly whatever you know this is a length of the transistor. So this is how Intel managed to do in 2011 and we have the finFET technology it looks beautiful but what is the exact benefit in term of performance.

(Refer Slide Time: 14:14)



That you can see in this graphs here this is again taken from the same presentation by Intel now this is 22 Nanometer technology. So in that planar gate technology we have seen this on off ratios we have talked about sub-threshold current so this is let us say on current I_D versus V_G. So we know the on current to off current I_{OFF} this so you should look at the planar transistor this is basically comparison of 2 technologies planar and tri-gate technologies.

So they are normalized so on current in this case is 1 I_{ON} is 1 off current is let us say you know 1×10^{-4} So you have 4 orders of magnitude change in the planar technology but we said that for ,you know good transistor for digital application especially you have to have higher I_{ON} to I_{OFF} ratio . So you need to have high I_{ON} to I_{OFF} ratio and that was a achieved by the tri-gate transistor or the finFET they call it tri-gate of it now we call it finFET.

You see there is clearly you know this here the I_{ON} to I_{OFF} ratio is 10 power 5 yes right so that is it 10 times better transistor than the traditional planar gate. And why was this possible? Well think about it you are actually in traditional MOSFET have the oxide and below that you had the channel. So you are only using the gate to continue the channel but in a fin side they are actually having a gate which is surrounding the channel.

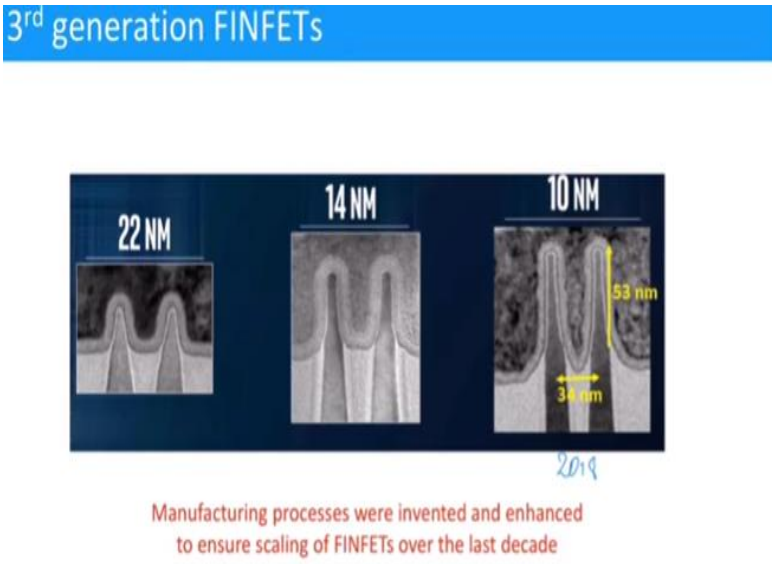
So you have better electrostatic control on the channel and that is why if you want to switch it on of course you turn on. But when you want to switch it off because you are electrostatic controlling better you have a lower leakage current. So this is terms of the leakage current and we also studied this you know really short time. We said that time delay the delay this side is proportional to C time V by I current.

$$delay \propto \frac{CV}{I}$$

So if you increase the drive current we are going to reduce the delay if you reduce the voltage you can also reduce the delay and also power dissipation right and this is a very qualitative picture I am giving you. So what happened is we said that because of the threshold voltage you know we could not reduce it further it sort of saturated and we would like to reduce it further but there was technological issues as well.

But with Intel tri-gate technology it turned out that they were actually if you compare the delay verses the operating voltage. In a traditional technology it was like this whereas if you go to tri-gate technology we are able to achieve much lower delays for a similar operating voltage. So in principle you could also reduce operating voltage and still get a same amount of delay. So that is how Intel also manage to improve the performance of a transistor this is in year 2011 so 1 decade back.

(Refer Slide Time: 17:17)



So all of these technologies where developed you know within our life time and of course after that what happens in the 10 years well the technology improved even further you know we started out with the fin. If you look at the fin that was you know very nice picture right you know because it is technical low magnification. But if you go back and look at it closely in 2011 the fins where of this shape the height of the fin was not much.

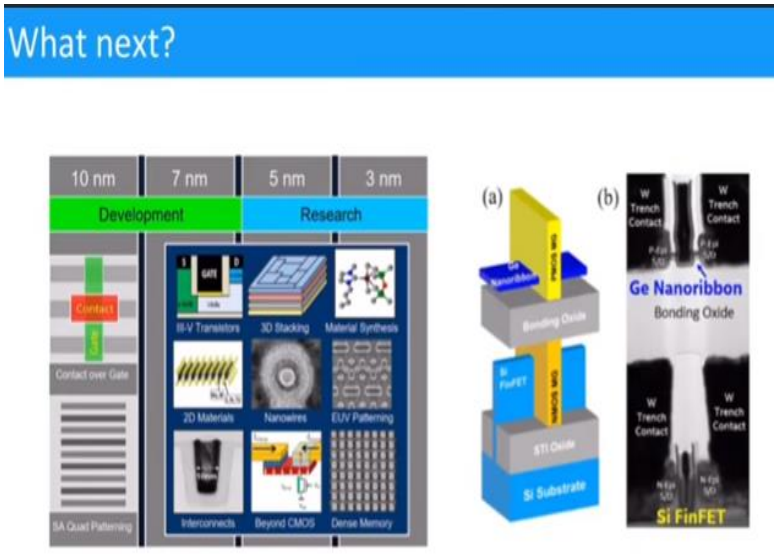
And then there is a surrounding gate and over the next 10 years you know this is 2018 I believe 10 nanometers about 2018 they managed to improve the height of the fin. And also surrounding you know gate. So basically they managed to bring the fins closer and also increase the height of the

fins. And by doing that we are able to get much higher performance of a transistor and it might seem I mean it seems easy enough right just make it.

But this is actually pushing the limits of technology it took 10 years to reach from 22 nanometers to 10 nanometers or 8 years. The reason is they have invent a lot of processes you know these are done using various expensive equipment if you look at the state of the semi-conductors nano meters the fabs that are rewired to fabricate cost up enough 5 billion dollars or 8 billion dollars or so.

So it is lot of expensive equipment that is required to actually to make this so but you know Intel as managed to do all of this equipment's of course after Intel lot of other companies have also done AMD, Global founders ,TSMC and so on. Lot of companies have done these things so these are the improvements that were done.

(Refer Slide Time: 18:56)



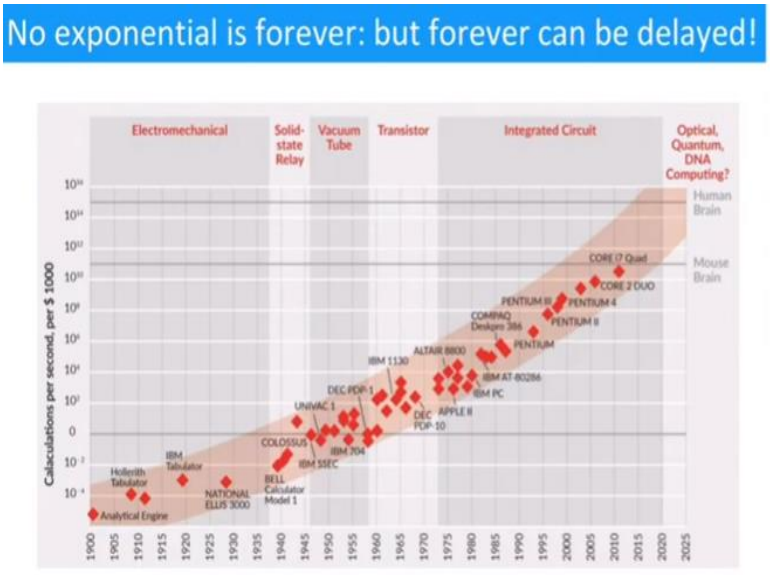
And finally what next it is okay right we have understood it high level position what current technology. But now the race is on to develop a next generation of technologies and what are they?

Well lot of things being discussed because all of these channels is all of this technologies have some basic limitations. So many thing are been discussed for example we can introduce III-V semi-conductors so that we can improve the performance or you can use Nano-wires as a transistor.

Or you know we have various other technology are being explored we do not know what is going to work but it is really the cutting edge of resource right now. And I just wanted to show you the latest last year in 2019 there was paper of Intel where they show. We have now see the fin has you gone up the transistor has go out of plane now. But in addition what they have done is an NMOS was fabricated like a FinFET but then on top it you fabricated PMOS structure not next to it but on top of it.

That means you are actually further compressing to the transistor right this is kind of a 3D stacking of transistors which is what was done by Intel again. They demonstrated this in 2019 so there is a lot of progress that is happening and my effort was that basically we want to give you a glimpse of it beyond the textbook.

(Refer Slide Time: 20:35)



So and what else and so in the end I would like to stop with this slide this is what I have shown in the introduction and this is the exponential increase in the computational capacity in last century right. So and especially in last 50 70 years is starting with year 1948 we had about 1 computation per dollar per second I say now the increase in this various processes we are managed to increase this and now we are come close to 2020 somewhere here or the other what next?

There are whole range of ideas that are been discussed no one really knows what? They believe that you know optical technologies might play a crucial role because optical technologies are inherently have higher bandwidth and they actually can go faster. That could be one potential direction being an optical interconnect to chip or we bring in quantum computing which is completely a different paradigm altogether.

And it if it really works out in field it will change technology you cannot even image what we can do right now. Just like now somebody there in 1950's they could not have imagined what MOSFET would have done in the last 70 years. Similarly if quantum technologies really ease practical we manage the implementation and it actually at a cost effective then it will changes a world. So and or it could be you know based on bio systems using a something called as Neuromoprhic similar to neuromorphic bio-system of neuro-computing thing like that.

A various things that are being tested in the scientific community and it is being my effort to show you glimpse of that the reason is if you want to work at the cutting edge we need our youngsters should be educated in these things so that is been the effort and you are welcome to explore some of these things work and I am sure you will find it fascinating with that I will stop my discussion on the MOSFET's part of the course.

In the next 2 weeks I will discuss opto-electronic device which is another fascinating dimension alright so with that I will stop thank you for your attention and will see you next week have a great week bye.