

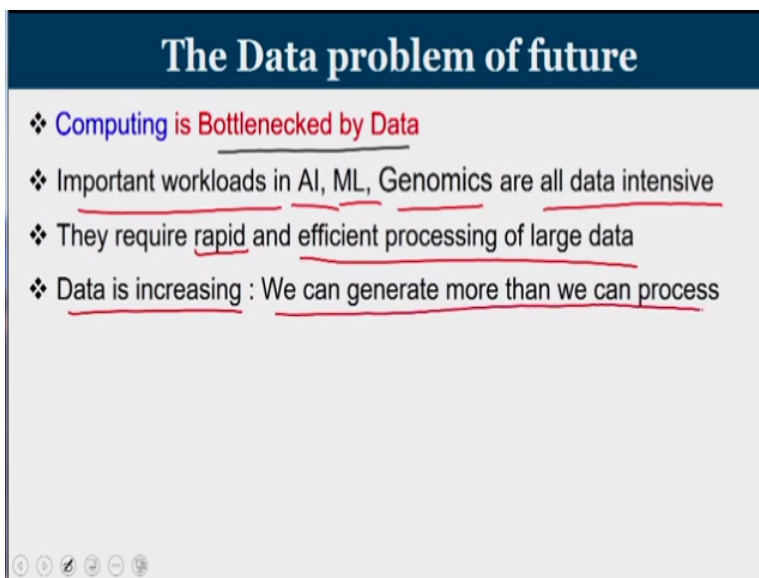
Advanced Computer Architecture
Prof. Dr. John Jose
Assistant Professor
Department of Computer Science and Engineering
Indian Institute of Technology-Guwahati

Lecture-32
How to Explore Computer Architecture

Welcome all of you to the final lecture the concluding lecture of this 8 weeks NPTEL course on advance computer architecture. And this course will give you an executing summary of what we have learned and what are the way ahead in exploring computer architecture further. We are in a world where we are having data explosion lot of applications that we work on our computers and our mobile gadgets work on data audio data sometimes video data, sometimes text data, so huge amount of data is been processed.

Over the years the computer architecture was not develop considering these data patterns only in the recent few years this tremendous data explosion happened. So our architectures need to be intelligent enough, need to be supporting enough to facilitate operation on these kind of data. So what we see is specific problems pertain into data management and at the operating system level at the system software level and even at the hardware level we need to have intelligent mechanisms to handle these data.

(Refer Slide Time: 01:49)



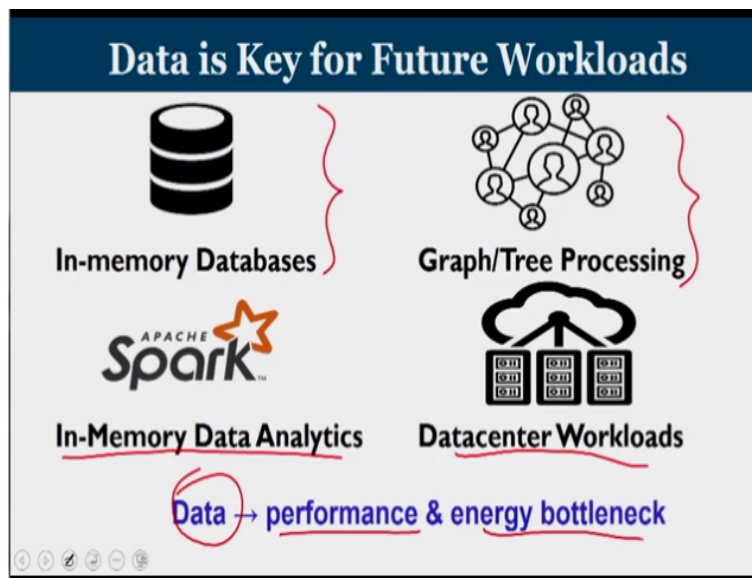
The Data problem of future

- ❖ Computing is Bottlenecked by Data
- ❖ Important workloads in AI, ML, Genomics are all data intensive
- ❖ They require rapid and efficient processing of large data
- ❖ Data is increasing : We can generate more than we can process

Navigation icons: back, forward, search, etc.

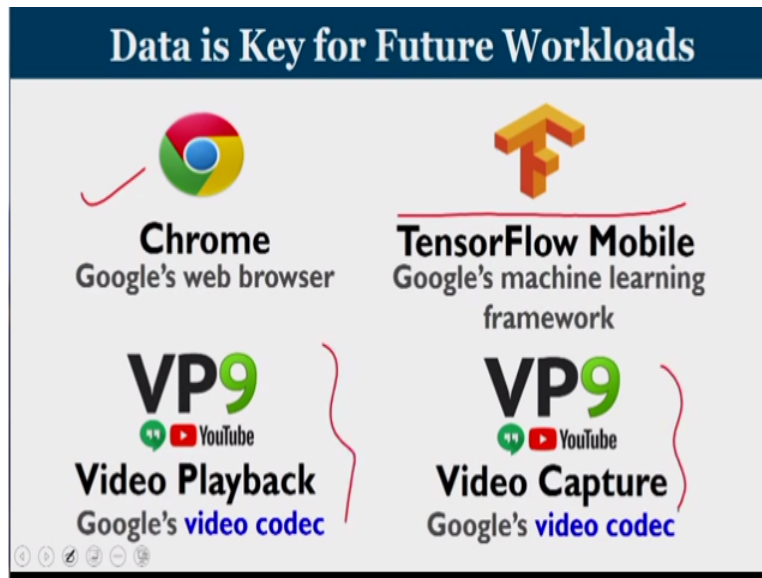
So computing is bottlenecked by the data that is the challenge that you are going to face in the near future. Import and workloads in artificial intelligence, in machine learning, in genomics all are data intensive. And majority of the application require rapid and efficient processing of large data and we know that these processing happen in the hardware. So data is increasing and we are actually generating more data than we can process.

(Refer Slide Time: 02:23)

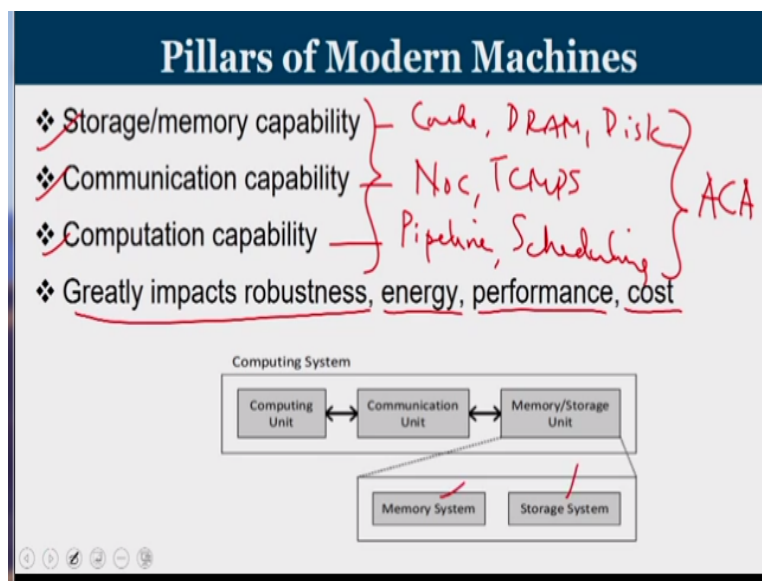


We are in the era where the future applications, the future workloads that are going to run on our computers are going to be massively data intensive. We are going to handle with databases inside memory sometimes we may have to handle with graph and preprocessing applications. We may have to do analytics on memory and then we may have to deal with datacenter workloads. So once you move into data centric approach performance and energy is a bottleneck as far as processor designs been concerned.

(Refer Slide Time: 02:56)



Even the web browser that we are using that is going to handle huge amount of data and essentially when you breakdown it in architecture concept they are even a web browsing itself is an execution of a program it is involves fetching, decoding, execution of instructions. Similarly the tensor flow is the machine learning framework that google use and then we have the google's video codec, video capturing mechanisms all these are going to handle with huge amount of data.
(Refer Slide Time: 03:27)



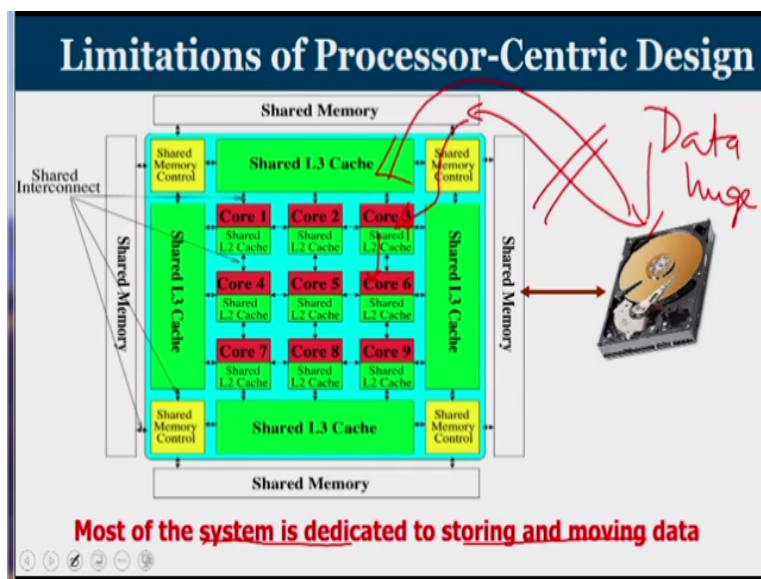
And we know that as far machine design is been concerned, we need to have 3 important pillars that we have covered in this course. We started with computation capability of the machine in terms of pipeline, in terms of scheduling these are the concept that we learned. And then we

learned about storage and memory capacity in terms of cache then we have DRAM and then we have disk.

And then we are talking about the communication capability as well in terms of NOC and TCMPS. So our entire 8 week course on advance computer architecture was giving you a flavor of how should be the computation capability and how should be the storage and memory capability and how should be the communication capability, so overall we use the design approach.

So these 3 aspects how the computation unit is been designed, how the memory is been organized, what are the capabilities of communication system are going to greatly impact the robustness, the energy consumption, performance and cost associated with execution of applications. So in total our computer system consist of computing communication and memory concepts and memory further subdivided into memory which consists of RAM and cache and the permanent storage.

(Refer Slide Time: 04:59)



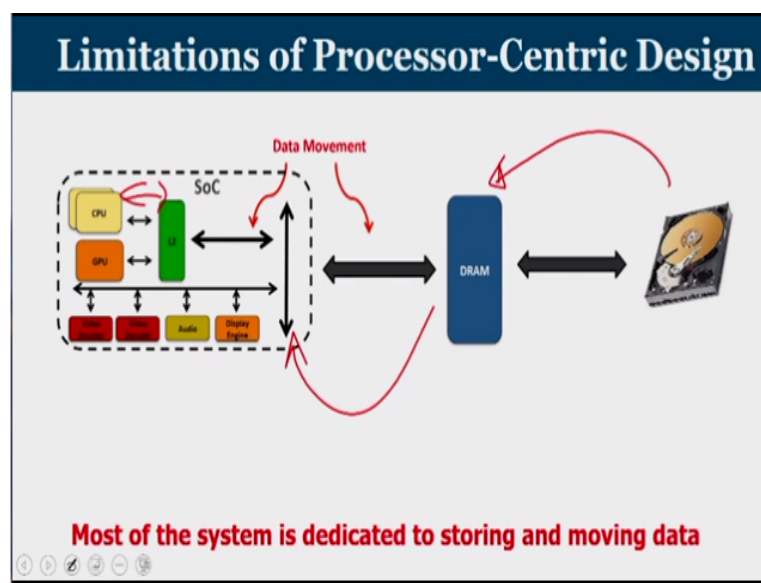
So these are the traditional TCMPS that we have seen in this week, it have some inherent so limitations when you talk about this data intensive applications. So most of the systems that we have seen in this course is dedicated to storing and moving data, you keep data in hard disk and

then you move to memory controllers and then from there you move to L2 caches and then you move to the course.

So it is all about your processor is going to be there in one place, the task is executed in the processor as and when required instruction and data are been moved. And when this data is going to be huge then it involves lot of movement of data from off-chip to on-chip and from different components in off-chip. These kind of architectures we are working really well as far as the kind of application that we have seen up to the last decade.

In the decade since a data is a dominating factor we should think about alternate designs that can process these data fast.

(Refer Slide Time: 06:03)



So the limitations of processor-centric design is lot of movement is been involved from hard disk to DRAM from DRAM to the chip and then from L2 to L1 and many other operations that we have seen, your mispenalty is there. So when you have lot of data I cannot keep everything in cache, so caches will encounter cache misses and then we try to come up with optimization design. So most of the system is dedicated to storing and moving data to facilitate processing.

(Refer Slide Time: 06:30)

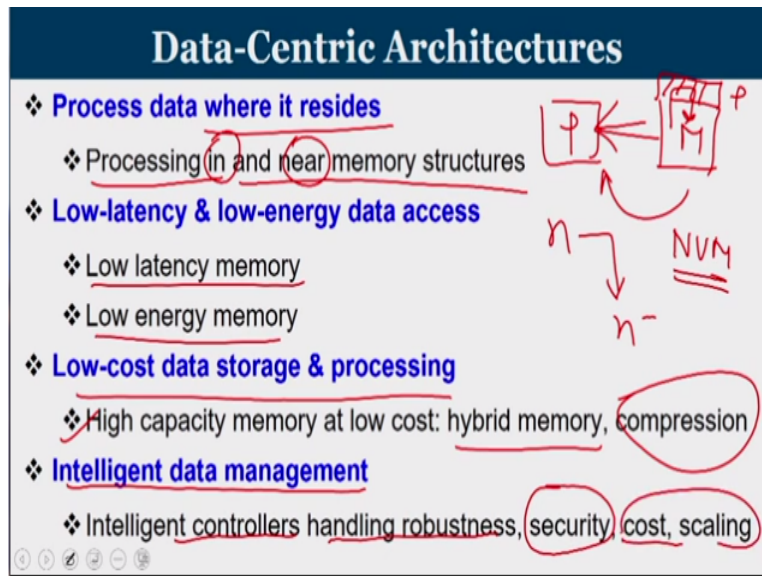
Handle Data Well

- ❖ Ensure data does not overwhelm the components
 - ❖ via intelligent algorithms
 - ❖ via intelligent architectures
 - ❖ via whole system designs: algorithm-architecture-devices
- ❖ Take advantage of vast amounts of data and metadata
 - ❖ to improve architectural & system-level decisions
- ❖ Understand and exploit properties of (different) data
 - ❖ to improve algorithms & architectures in various metrics

So what computer architecture of the future is heading onto is can you handle data well, for that ensure data does not overwhelm the components, how can it be done. We can use intelligent algorithms, we can use intelligent architectures or we can use the whole system design by clubbing intelligent architecture and algorithm devices. And take advantage of vast amount of data and the metadata on it, this will help us to improve architectural and system level decisions.

And we need to understand and exploit properties of different data as well and this will help us to improve algorithms and architecture in various metrics. So with the concepts that we have learned in this course like advance computer architecture, we have to understand the context in which our processors are going to work. Essentially our processors are been used for execution of task and when task are becoming more data-centric in modern day, optimizations at the architectural levels are required to facilitate this.

(Refer Slide Time: 07:46)



So the future is all about data-centric computer architectures where you process the data where it resides. And how it is been possible there is something called in memory computing or near memory computing, processing in and near the memory system. So traditionally we have a processor and then we have memory systems and through the BUS we are going to bring it, can I add a small processing module near the memory.

Such that some of the operations rather than moving the data can I process it and then store it back that is called processing in and near the memory system. And then we should focus on low latency and low energy data access, can we have mechanisms by which I could store data in places where the latency is low. And can you have or can we have techniques by which the energy associated with reading and writing from memories are going to be less.

In this context non volatile memories NVMs are also gaining significance and then can you design high capacity memory at low cost. Low cost data storage and processing, something like hybrid memory and let us say that is a I have n bit of data. Now can I represent this n bit of data into lower than n bits, can I compress my data such that data would not take enough space but still I can get back the data as and when required.

So we need higher storage if you cannot go for higher storage can you compact the data, so compression friendly architectures, hybrid memory architectures, low latency memory

architectures and low energy memory architectures are going to be some of the important areas in which architecture community is going to work on. Then at the end we should have intelligent data management it is done with a help of intelligent controllers.

We have seen that DRAM controllers, we have seen hard disk controllers, so can these controllers be put more intelligent such that we can manage the data well by handling robustness. And security of hardware is yet another important aspect, can I take care of the security, can I take care of Trojans that is been coming, keeping in mind the cost and the scaling aspect.

(Refer Slide Time: 10:10)



So what is the way forward, we need to design data-centric system and intelligence spread around it, do not center everything around traditional computation units. We have to rethink our computational process from a different angle, we need to have better cooperation across layers of a system. That is something called co-design rather than designing different components and putting it together let us try to understand how one component is trying to interact with other, can we have processors which will talk more frequently with caches and try to know what cache want.

So that is called careful co-design of components and layers at system level, at architectural level and at the device level. We need to come up with better, richer and more expressive and flexible interfaces. We know that there are different interfaces that we have between processor and

memory, can the interface be more expressive, can the interface be more flexible, can interface be more richer and better as far as efficiency is concerned as far as cooperation is concerned.

And better than worst-case design, we always design things for worst-case, this is the worst traffic that you are going get. So we design it for handling such kind of worst case traffic, we should know that these kind of worst-case happens rarely. So try to focus our designs on common case and then, so heterogeneity in the design. Rather than keeping everything homogeneous we should think about design which is a co-design of various components like we discussed.

At the same time some components are used for certain category of applications whereas some other components in the same chip will take a dominant role when it be handling some other class of application. So can we have specialized hardware that the hardware be asymmetric it enables more efficient design. So no one size is going to fit all, so in this way the future architectures should be more intelligent in nature.

(Refer Slide Time: 12:16)

How to explore computer architecture ?

- ❖ Refer to IEEE/ACM transactions & journals
 - ❖ IEEE TCAD, IEEE-TVLSI, IEEE-TOC,
 - ❖ ACM-TODAES, ACM-TECS, ACM-TACO
 - ❖ JPDC, JSC, JSA, CAL, ESL
- ❖ Refer to top tier conferences
 - ❖ ISCA, HPCA, MICRO, ASPLOS, PACT, DATE, DAC, ICCAD
 - ❖ ICCD, ISVLSI, ASPDAC, VLSI-SoC, GLSVLSI, NOCS, NoCArc
 - ❖ HiPC, VLSID, VDAT, ISED

Now some of you may be thinking after learning a course like advance computer architecture over this last 8 weeks, what is a way forward. Some of you might have found some interest in studying this topics further, some of you maybe planning to do higher studies in this domain

where are lot of opportunities that is available for you. There are waiting for you, if it is in terms of higher studies premier institutes in the country we have vibrant architecture research group.

And there are lot of R and D sector also in this domain, lot of corporate sectors also work in hardware manufacturing. So efficient skill set is been needed, the concept that we have learned in this course will help you in building a base which will be acting as a gateway entry point either to higher studies or to a scientific or an engineering carrier in this domain. To know further what is happening with a base that you got from this course reading research materials will help.

There are research articles that are published, so I will give you a quick overview of where can you find this research articles. Generally in computer architecture domain we have peer reviewed journals that are published regularly by professional societies. And then we have peer reviewed conferences as well which are conducted on an annual basis across different locations of the globe and this conference proceedings are also available for learners.

If you wanted to explore computer architecture further refer to IEEE or ACM they are professional societies and we call them transactions, special category of journals. And we have journals from other societies as well, so IEEE transactions on computer aided design IEEE transactions on VLSI, IEEE transactions on computers. ACM transactions on design automation of electronic systems, ACM transactions on embedded computing system, ACM transactions on architecture and code optimizations.

These are some of the top tire journals where you get access to research articles published in architecture domain. And then we have journal of parallel and distributed computing, journal of supercomputing, journal of systems architecture, computer architecture letters, embedded system letters. Here also you find very good research articles and then we have premiere conferences international symposium on computer architecture, high performance computer architecture, micro architecture, architectural support for programming languages and all.

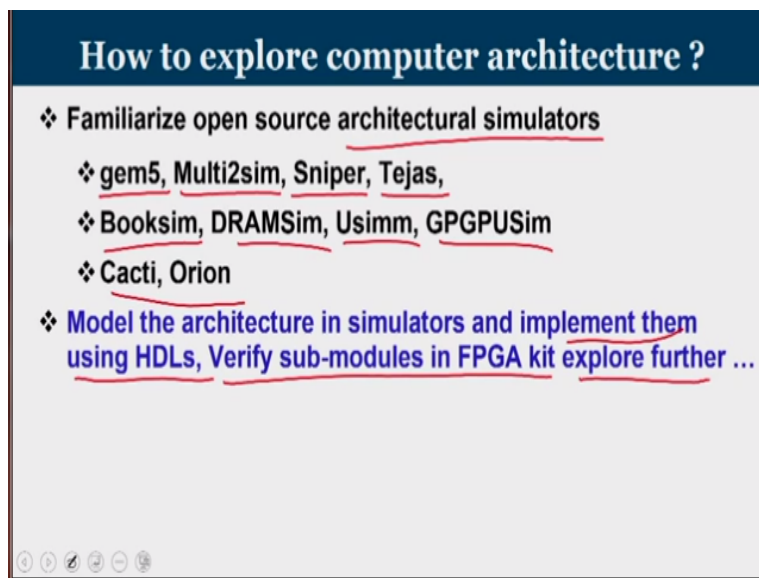
And OS parallel architecture and compilation techniques, design automation and test in Europe conference, design automation conference, international conference on computer aided design.

These are all the top conferences where you get extremely high quality research material from this domain. We have the next tier level conference, conference on computer design, the international symposium on VLSI, Asia south Pacific design automation conference, VLSI system on-chip conference.

Grade like symposium on VLSI, network on-chip symposium, network on-chip architecture workshop, these are all the next level conference. Here also you get reasonably good research material which through lights into the future of computer architecture. And for the Indian community we have these conferences which are located in Indian cities happening annually, high performance computing VLSI design, VLISA design and test, the international symposium on electronic design.

These are all international conferences located in India, so reading articles from these conference proceedings. And if possible try to attend these conference it will give you a bigger exposure and you will be able to network with researchers who work in this domain. These conferences are having very good keynotes there are many student contests, PhD forums. So that will get you a chance to be in-connection with the architecture community.

(Refer Slide Time: 16:43)



How to explore computer architecture ?

- ❖ Familiarize open source architectural simulators
 - ❖ gem5, Multi2sim, Sniper, Tejas,
 - ❖ Booksim, DRAMSim, Usimm, GPGPUSim
 - ❖ Cacti, Orion
- ❖ Model the architecture in simulators and implement them using HDLs, Verify sub-modules in FPGA kit explore further ...

Now apart from reading these material what is further you have to familiarize with similar task especially when you work in architectural systems, open source architectural simulators will help

you a lot in implementing this ideas. So one such tool was introduced to you in this course gem5 similarly you have multi2sim, sniper, Tejas many tools are there which will give you full system simulation.

If you are focusing on 1 micro architectural feature like cache, like GPUS, like DRAM then booksim, DRAMsim, Usimm, GPGPUsim all will help and these are all power tools cacti and Orion. So if at all you wanted to know more about things that is been happening I request to you go and read this research materials and then implement some of them within this architectural tools.

So and then try to see are you able to get the same kind of research what is been claimed in this papers. So once you implement things in this simulators then you will get to know more statistics and you may find some interesting observations. And that is a basis for doing computer architecture research and projects with a research background in computer architecture domain.

And people from the electrical science background especially from electronics and the electrical engineering hardware description language which you might have learned along with your courses like Verilog, VHDL and Bluespec. If you have proficiency in that or if you can learn them you can actually model you can write hardware codes, hardware description language codes for the techniques proposed architectures or algorithms that are proposed in these papers.

And try to implement them using hardware description language, so model the architecture in the simulators and implement them using HDLs and verify this sub-modules in FPGA kits or SOC kits. And then you will be able to explore further tools like synopsis, cadence (()) (18:45) will give you more of the critical path, the timing, the delay, the area overhead analysis etc.

(Refer Slide Time: 18:53)

Summary

- ❖ Multicore processors and on-chip clouds are going to become an integral part of future digital technologies.
- ❖ Understanding the hardware of such system will help us to design with conceptual clarity.
- ❖ Our country need good computer architects and processor design engineers with hands on exposure to VLSI design flow to cater the growing demand of skilled personnel in this domain.

So the summary is multicore processors and on-chip clouds are going to become an integral part of future digital technologies. Understanding the hardware of such system will help you to design with conceptual clarity. Our country need good computer architects and processor design engineers with hands on exposure to VLSI design flow to cater the growing demand of skilled personnel in this domain.

(Refer Slide Time: 19:21)

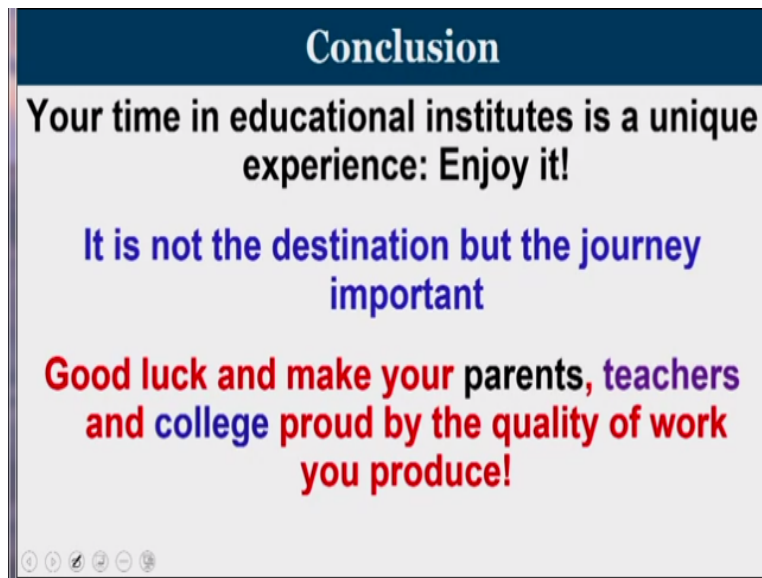
Our role as educated citizens

Let us make ourselves **up-to-date** in our respective subjects with latest technology enabled learning and **practice healthy, sound learning and research practices, academic teamwork to mutually inspire** each one of us such that we get **transformed as potential technocrats, engineers, scientists, teachers and researchers of next generation.**

So being educated citizens who are in the part of learning more, let us make ourselves up to date in our respective subjects with latest technology enabled learning and practice healthy sound learning and research practices, academic team work to mutually inspire each one of us. Such

that we get transformed as potential technocrats, engineers, scientist, teachers and researchers of next generation.

(Refer Slide Time: 19:49)



So to conclude your time in educational institutes is a unique experience you should enjoy it. It is not the destination that the degree that you get is not that important but the journey that is important. The way in which you learned, the concept that you learned is more important and good luck to all of you and may you make your parents, teachers and college proud by the quality of the work that you produce.

So we have come to the end of this course and lot of inputs are been given in this final video lecture which will help you in exploring further in this domain, think of pursuing higher studies in the systems field in the architecture field. And the entire national system the higher education institute like IITs, NITs, IIITs and central CFTIs all willing to take aspiring candidates to explore computer architecture, it was a pleasure working with you all in this course, I wish you good luck, thank you.