

Multicore Computer Architecture – Storage and Interconnects

Dr. John Jose

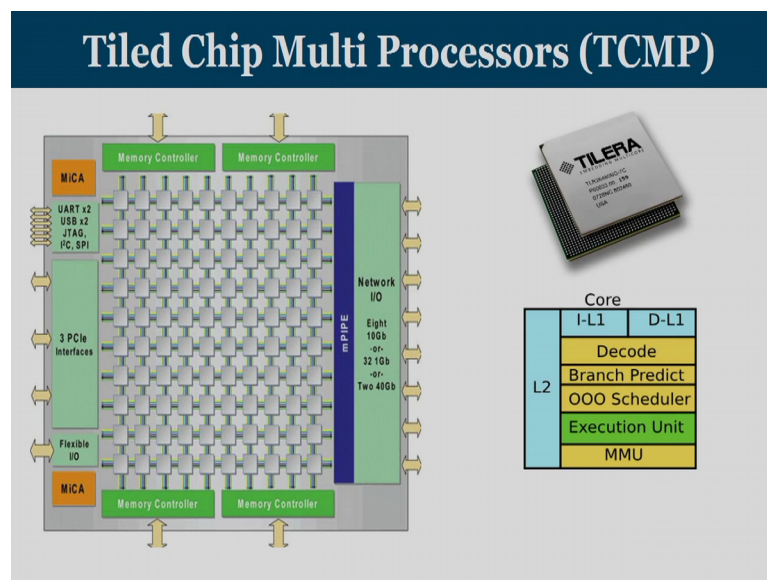
Department of Computer Science and Engineering
Indian Institute of Technology, Guwahati, Assam

Lecture –18

Emerging Trends in Network on Chips

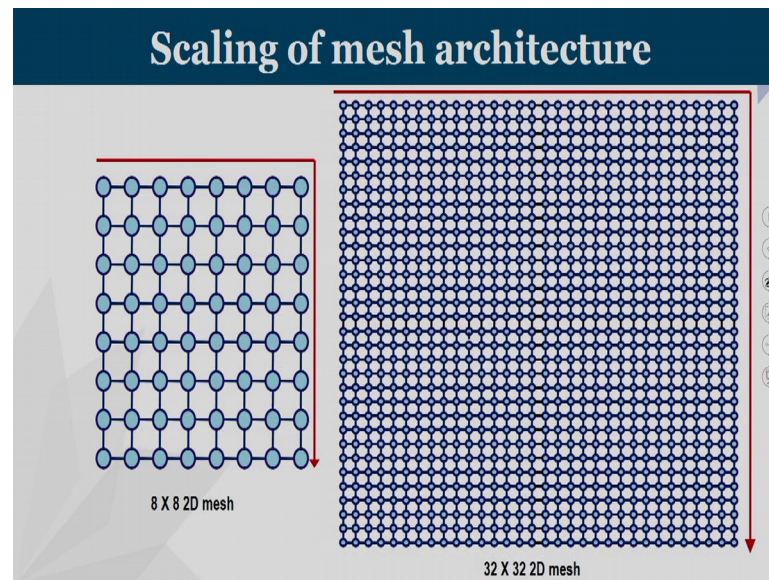
Hi, hello everyone. Welcome to lecture number 18 of this course. This is going to be the last lecture of this course. We will be having some advanced trends in network on chip that is going to be discussed today. There is one more tutorial from TCMP system and with this we are going to wind up this course. Today in this lecture we are going to work on discuss on Emerging Trends in Network on chip.

(Refer Slide Time: 00:54)



So, we have seen that tiled chip multi core processors TCMP was our point of discussion for the last few lectures we have seen what is the internals of the router micro architecture and how TCMP system work, optimizations on it and that of the cores and the cache are going to be there inside your TCMP system.

(Refer Slide Time: 01:16)



Now, look at this scenario. So, far we were discussing about the 8 by 8 2-dimensional mesh system. Now, in this case if we really wonder to travel from one end of the chip to another you are going to take roughly 14 hops that is going to be the diameter of the system. Now, consider larger systems like this; this is organized as 32 by 32 2-dimensional mesh system. So, it is going to give us 1024 cores, it is really huge our future systems are going to be of this order. We have 1024 cores on a single chip.

So, if such kind of magnitude is there our traditional routing algorithms our mesh structure will it support we have to find out. In this case, if you wonder to travel from one end of the chip to another it is taking so, many number of hops and you know that your router will consume one or two cycles and link is going to take another one cycle. So, it is going to be a quite a large count to on the edging here.

(Refer Slide Time: 02:24)

Emerging Trends in NoC

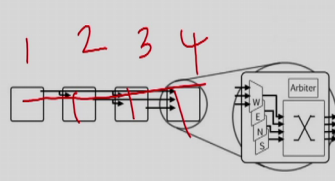
- ❖ Multidrop links ✓
- ❖ Three-dimensional NoC ✓
- ❖ RF interconnect ✓
- ❖ Nanophotonic NoC ✓
- ❖ Wireless NoC ✓

So, let us try to find out to reduce the diameter of the network how can I travel from one end of the chip to another when it is a huge multi core system. Some techniques we will discuss today. The first one is multidrop links then we have three dimensional NoC, we have RF interconnects, we have nanophotonic and then we have wireless NoC. Our attention will be more on wireless NoC which is going to be the future trend, but a couple of minutes I will spend on discussing on multidrop links 3D NoC, RF and nanophotonic NoC.

(Refer Slide Time: 02:55)

Multi-drop Links

- ❖ Links where signals reach multiple routers in same clock cycle

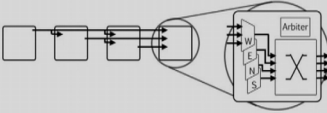


So, what is the idea of multi drop links? Can we design links where signals reach multiple routers in the same clock cycle something like this? Here we have the first router, we have the second router we have the third and the forth router and we have specially designed links which are known as multi drop links. So, a signal that is starting from 1 router whether it can get down a 2 or 3 or at the 4. So, that is going to be called the multidrop links, it is something like an express highway channel.

(Refer Slide Time: 03:27)

Multi-drop Links

- ❖ Links where signals reach multiple routers in same clock cycle
- ❖ Reduces the diameter of the network without the need of very high radix routers
- ❖ These links are built up and broken down by sending bypass signals towards the destination



- B. Grot et al., "Kilo-NOC: A Heterogeneous Network-on-Chip Architecture for Scalability and Service Guarantees," ISCA-38, 2011

- T. Krishna, *et al.* "Single-Cycle Multihop Traversals over a Shared Network on Chip," IEEE Micro, 34(3), 2014

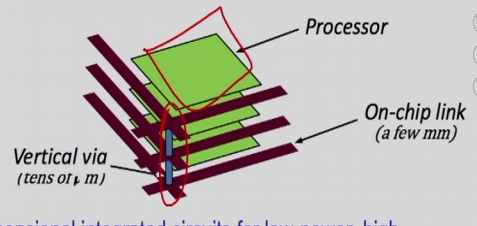
So, it reduces the diameter of the network without the need of very high radix routers and these links are built up and broken down by sending bypass signals towards the destination.

A couple of research work is already done in this and today in this lecture I am going to give the links for certain research papers also. So, that the interested students can work on this can go through it for finding out some good interesting topics.

(Refer Slide Time: 03:54)

3D - NoC

- ❖ Benefits
 - ❖ Higher packing density
 - ❖ Improved noise immunity
 - ❖ Overall superior performance



The diagram illustrates a 3D NoC architecture. It shows multiple layers of processors (green squares) stacked vertically. Vertical vias (red lines) connect the layers, with a label indicating they are 'tens of μm ' long. On-chip links (red lines) connect processors within a layer, with a label indicating they are 'a few mm' long. The processors are labeled 'Processor' and the links are labeled 'On-chip link (a few mm)'. The vias are labeled 'Vertical via (tens of μm)'.

- J. Burns *et al.*, "Three-dimensional integrated circuits for low-power, high-bandwidth systems on a chip," in IEEE ISSCC 2001

The second type of a network that we are going to discuss is about 3D NoC. So, far we have seen about two dimensional mesh or (Refer Time: 04:05) NoC. Can we start multiple such layers one over other and provide vertical interconnects and that is known as a three dimensional NoC. So, we have vertical interconnects that is there. So, we have tiles. now this multiple tiles are start one over other and then we have links which are there on the chip and we have vertical links also. What are the benefits? We can have higher packaging density and better noise immunity and we get superior performance as well. So, this paper will give you more detailed discussion on 3D NoCs.

(Refer Slide Time: 04:43)

Vertical Interconnects

- ❖ **Near-field coupling schemes**
 - Eliminate the need for a physical connection between layers
 - Energy per transmission higher than state-of-the-art vias
- ❖ **Inductive coupling schemes** : Longer transmission ranges than capacitive ones for similar data rates
 - N. Miura, *et al.* "A High-Speed Inductive-Coupling Link With Burst Transmission," IEEE Journal of Solid-State Circuits, 44(3), 2009
- ❖ **Capacitive coupling schemes** : Area overhead approximately one order of magnitude lower than the inductive options
 - ❖ Q. Gu, *et al.* "Two 10Gb/s/pin Low-Power Interconnect Methods for 3D ICs," Solid-State Circuits Conference, 2007

The next type is the in the 3D NoC we have the vertical interconnects and we can have interconnects which are not by physical connections. So, we can use near field coupling schemes which eliminate the need for physical connection between layers. So, we do not have really a physical connection, we can have two different mechanism; one is called the inductive coupling and other one is called capacitive coupling. So, for longer transmission ranges we have used the inductive coupling and the overheads of such kind of scheme can be eliminated by capacity coupling.

So, both by inductive and capacitive coupling, it is possible for us to communicate to upper layers without having through silicon wire. So, one approach is have vertical links physical connection second approach is have inductive coupling. So, whenever there is a charge that comes in one layer you come to know what is going to be the data that is to be recorded. So, by closer vicinity of a inductive and capacitive coupling also we can transfer data from one layer to another in a 3D NoC.

(Refer Slide Time: 05:55)

RF Interconnects

- ❖ Alternative to traditional VI signalling through metallic wires
- ❖ Transmission of electromagnetic waves over micro-strip transmission lines within the metal layers of the chip
- ❖ Signals are modulated using carrier waves at frequencies up to several GHz and then guided through the transmission lines
- ❖ Signals propagate at the speed of light instead of at the charging and discharging speed of RC wires, and need to be taken back at the baseband frequency and demodulated at the receiving end

– M.-C. F. Chang, V. Roychowdhury, L. Zhang, H. Shin, and Y. Qian, "RF/wireless interconnect for inter and intra-chip communications," *Proceedings of the IEEE*, vol. 89, no. 4, pp. 456–466, 2001

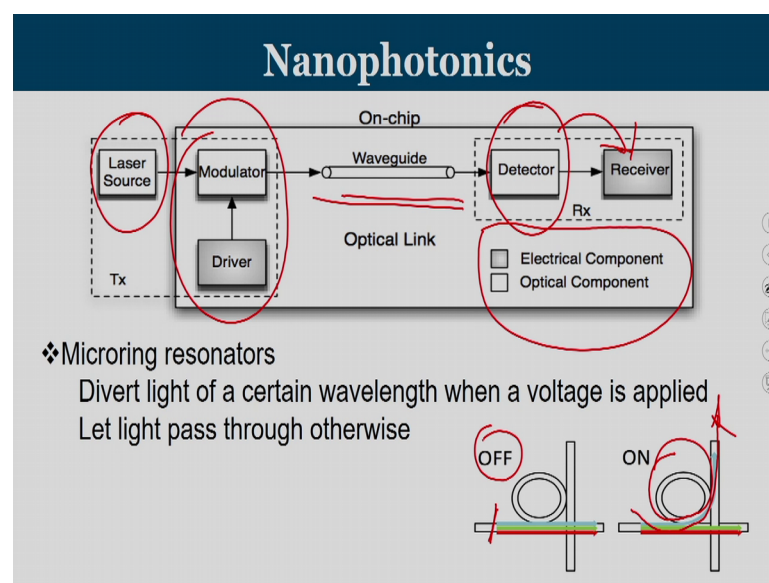
The next technique is going to be RF interconnects. So, it is alternative mechanism to traditional voltage and current signally. So, normal signaling is voltage and current signaling through metallic wires and the.

Replacement technique is going to be can you transmit the electromagnetic waves over micro strip transmission lines within this metal layers. So, we have your metallic layer and on top of the metallic layer rather than going for voltage current signally can we

send electromagnetic waves over this micro strip and that technique is called RF interconnects in mechanism. So what we do: signals are modulated using carrier waves at frequencies of the order gigahertz and then you send it through guided transmission lines and signals propagated the speed of light instead of charging and discharging of RC wires and we need to take it back to the baseband frequency and demodulate at the receiving. So, you have a modulation process at this ending side and then we have a demodulation process at the receiving side.

So, signals are going to be modulated with career waves of the order of the gigahertz and that is going to help you to transmit data at the speed of light than that of the charging of the RC wires. So, we get a bit of a more explanations in this paper which is recently published.

(Refer Slide Time: 07:25)



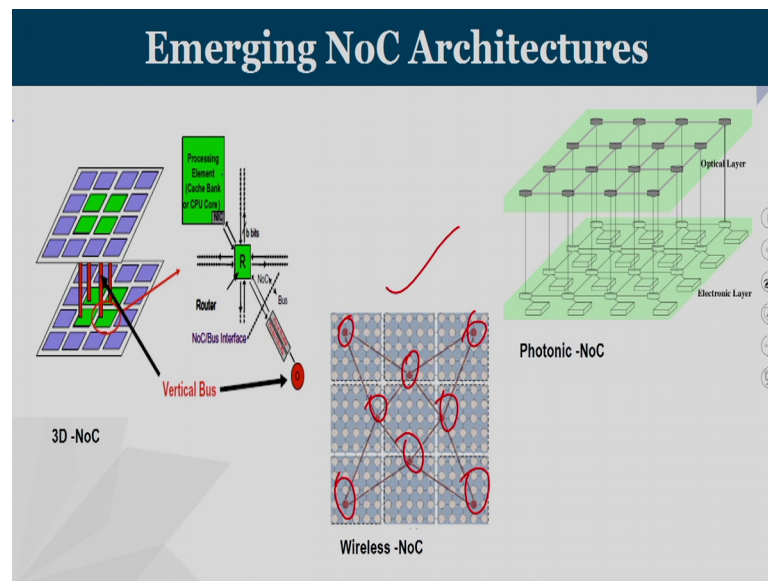
The next technique is all about fiber optic NoC or it is also known as nanophotonics. We use micro ring resonators which will divert light to a certain wavelength when a voltage is applied. So, only the other light passes through. So, we have a laser source and then we have a modulation through the optical wave guide, we are going to send the signal, we have to detect whether light is present or not and then we are going to convert into electrical.

So, you have at electrical component as well as an optical component in your NoC and this is a emerging field, lot of research work is happening in this field about the

nanophotonics. We can understand that when you have a light that is going to pass through, when you apply some voltage on this microring your light is going to bend across that is one way or if it is in off position then there is no bending.

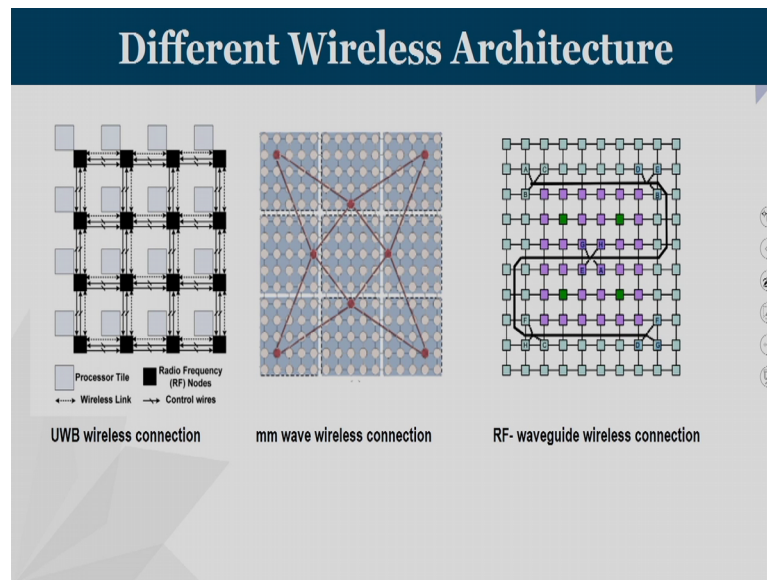
So, bending can be realized by applying proper potential on the light signal that are passing through the wave guide.

(Refer Slide Time: 08:53)



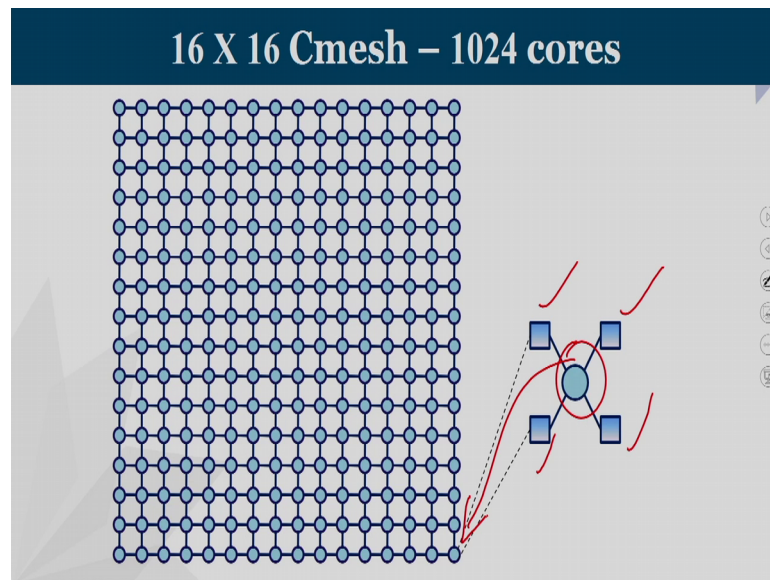
And so, we have seen the most important one of them are the 3D NoC with vertical interconnects. We have photonic NoC and the next one is called the wireless NoC which we will spend little bit more time today. So, if you look at a this structure, we can see that in each of this region each region is marked with blue coloring, each of the region we have identified certain nodes which have a wireless transmission capacity, wireless transmission and reception. So, you have normal communication that happens, on top of this normal communication we have this wireless access points. So, long distance communication can be realized over this wireless links.

(Refer Slide Time: 09:21)



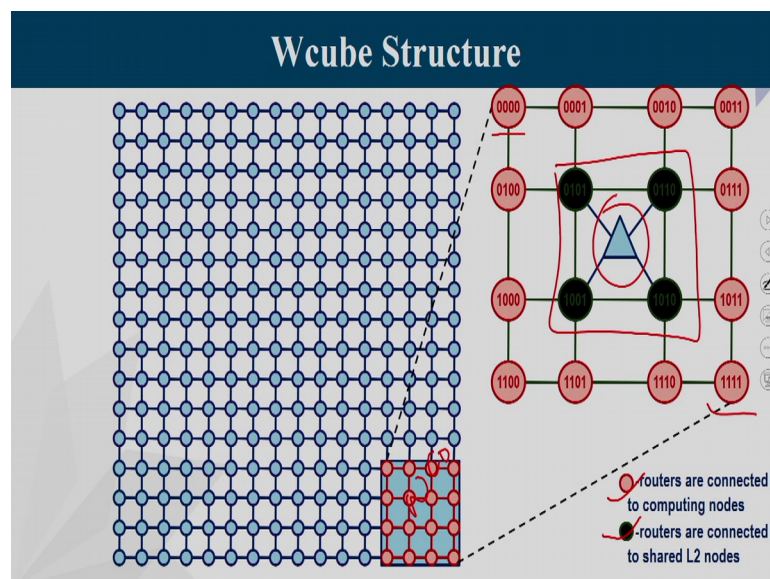
Now, we have different types of wireless architectures. So, first one is called a ultra wide band wireless communication where radio frequency is used. We have waveguide wireless communication there the wave guide is passing through channels and then we have the millimeter wave. So, that is what we are going to focus millimeter wave wireless connection. So, in millimeter wave the wave length is going to be the order of millimeters which can only reach only up to very small distance restricted to a chip. So, long distance communication you can send your data or packet in to this wireless hubs and this wireless hubs are going to send millimeter waves to adjacent an wireless hubs there by your signal reaches the required destination.

(Refer Slide Time:10:22)

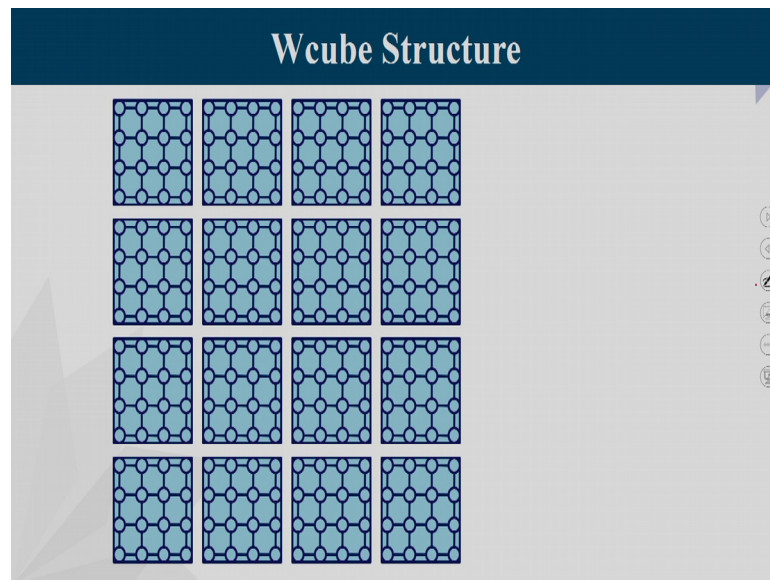


So, consider a 16 by 16 concentrated mesh; it is called on a Cmesh. Now, what do you mean by concentrated mesh? You have a 16 by 16 organization, 16 rows and 16 columns that will give you 256 points. Now, each of these node is having a concentrated of 4; that means, each of them is connected to another 4 nodes like this. So, this is the node what we have seen and we have 4 of the processing units connected to this router. So, we have a 256 routers are there, but total of 1024 cores that is called a concentrated mesh connection.

(Refer Slide Time: 10:40)

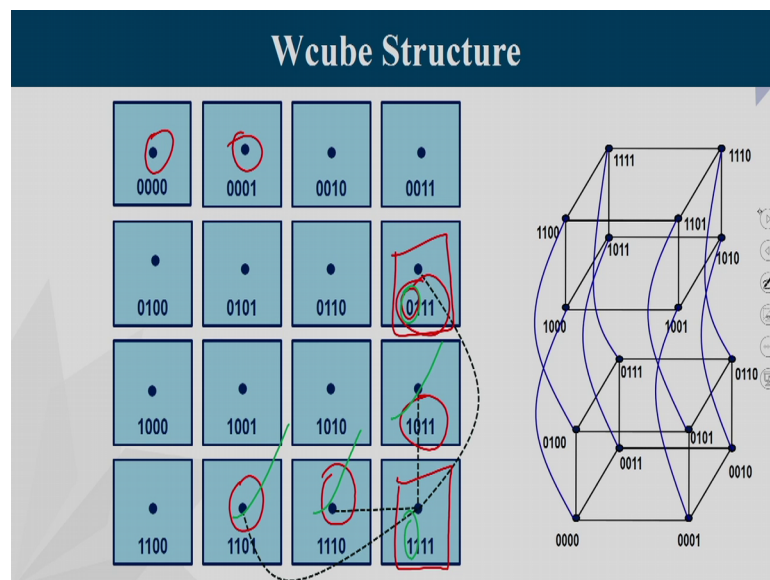


(Refer Slide Time: 12:23)



So, we can see that like that the entire 16 by 16 mesh can be divided into 16 W cubes.
So, this is the structure of a W cube.

(Refer Slide Time: 12:36)

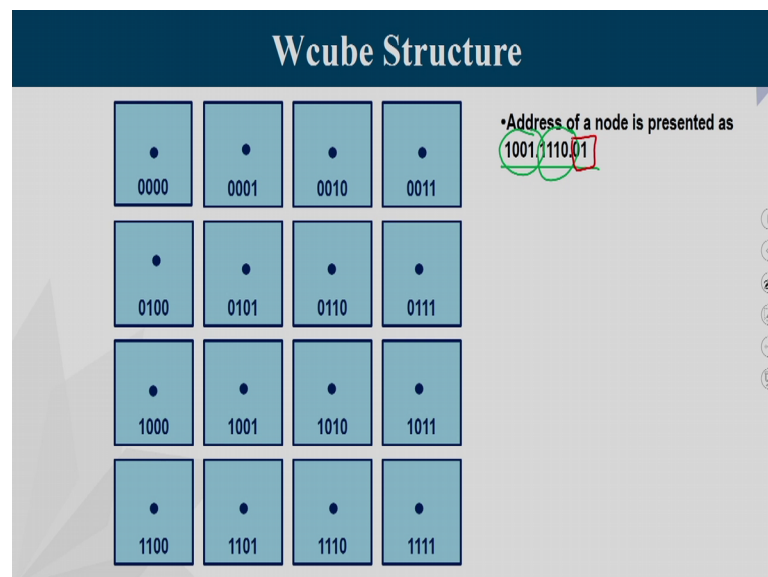


Now, each of the W cube can be now represented as W cube 0 0 W cube 0 1 1 0 like that.
So, that is the numbering scheme given to this w cube and what you see here in the each of the W cube is the transmitter. So, each of the W cube has it is own transmitter and receivers.

Now, let us see which all W cube can communicate to others. In the bottom right corner, we have a W cube 1 1 1 1 and these W cube can communicate to 4 of its neighbors. The peculiarity of 4 of its neighbors are these are the neighbors which are at a hamming distance of 1. So, 1 1 1 1 can communicate with 0 1 1 1, it can also communicate with 1 0 1 1, it can also communicate 1 1 1 0 and 1 1 0 1. So, we have 4 of the W cubes to which the node 1 1 1 1 can communicate. So, how are you going to obtain this?

What is the peculiarity of this? 0 1 1 1; that is this is at a hamming distance of 1 with 1 1 1 1. The numbering shows that these both the W cubes differ only at the most significant bit. So, the most significant bit is one that is varying. So, this most significant bit is varying all the remaining three bits in the W cube numbering remain same that means, they are at a hamming distance 1. Similarly, these W cubes also are at a hamming distance 1. They differ only in one bit position. So, a W cube can transmit to all other W cubes which are at a hamming distance 1 in its numbering; that means, this is going to be the logical connection, it is a hypercube structure. These are all which are connected to the other nodes.

(Refer Slide Time: 14:29)

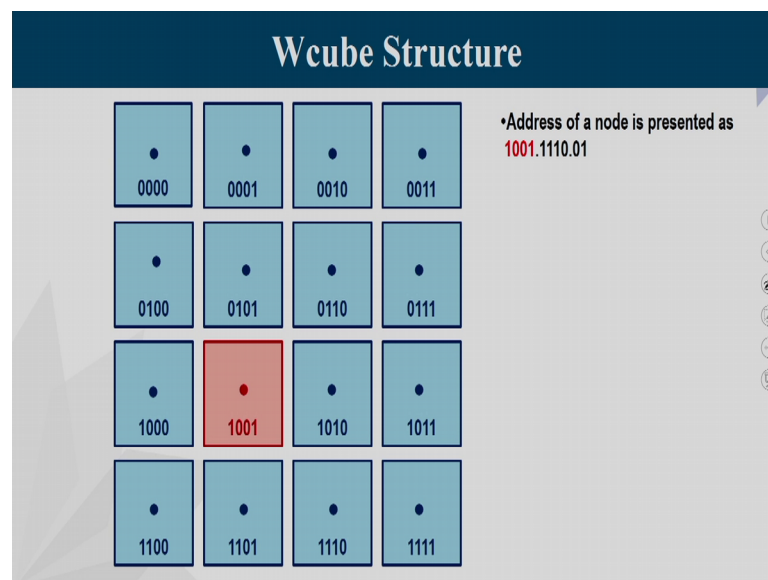


Now, let us try to find out the addressing scheme. We have mentioned that there are 1024 cores that are going to be connected with the help of this structure in a concentrated mesh. So, how you will tell the numbering? If it is 1024 nodes that are available which

are connected by an 256 grid structure; 16 by 16 grid structure. This 1024 nodes have to be represented using 10 bits. These 10 bits, the address of a node is presented like this.

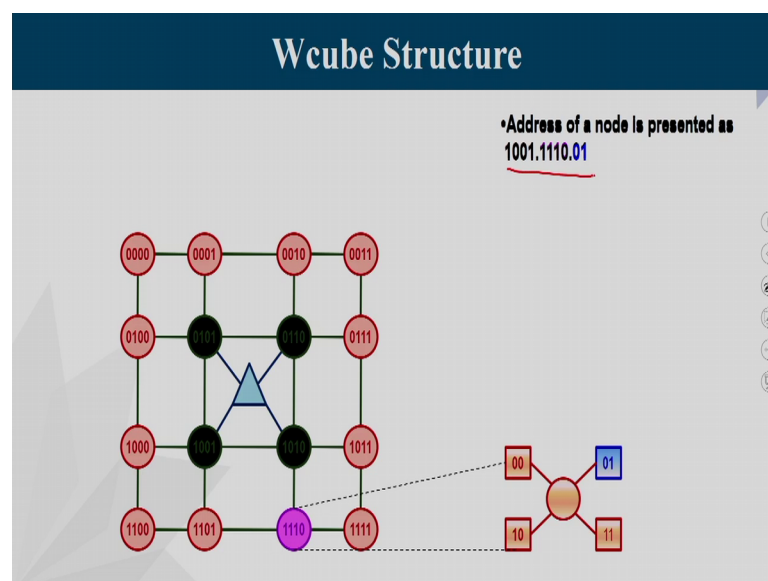
The first portion will tell you the W cube number, second portion will tell you the node within the W cube and the last portion will tell you out of that which is going to be the node under discussion.

(Refer Slide Time: 15:17)



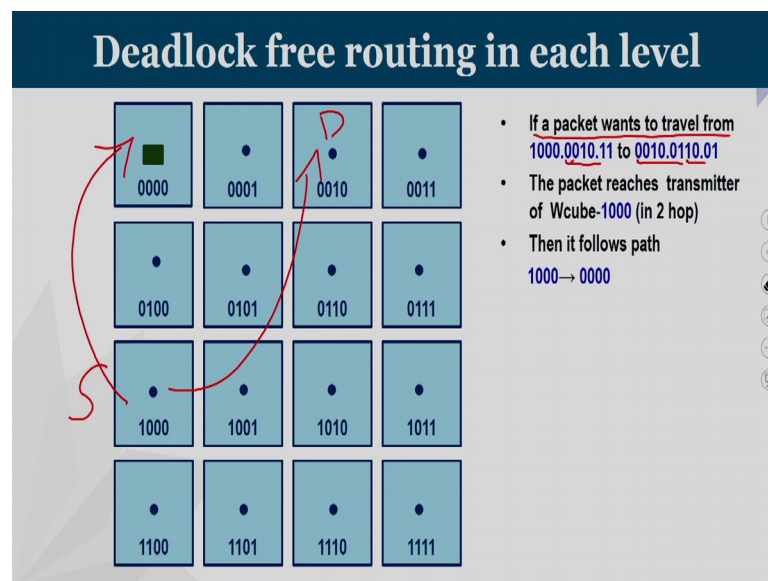
So, these means, 1 0 0 1; so this is the W cube that is been chosen with this number.

(Refer Slide Time: 15:25)



Now, within that W cube there is zoomed version of 1 0 0 1, we are going to 1 1 1 0. So, 1 1 1 0 is been shown and within 1 1 1 0 we have to go to 0 1. So, this is the way how the addressing mechanism works. The first 4 bits will tell you which is the W cube, next 4 bits will tell you which is the node within the W cube that is been chosen by the best 4 bits and the last 2 bits will tell you, you have reached already the router. Now, what is the core number that is connected to the router? And since the concentration is 4 meaning 4 cores are connected to a single router.

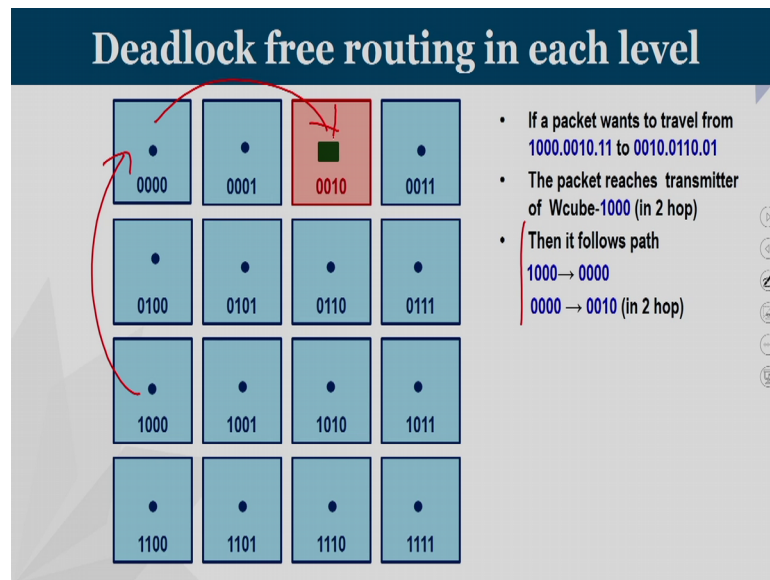
(Refer Slide Time: 16:05)



Going further, let us find out let us say if a packet wanted to travel from 1 0 0 0 0 0 1 0 to 0 0 1 0 to 0 1 1 0 0. So, the meaning is this is the source of the packet; it is travelling from that W cube to a W cube this. So, this is the communication that we have to take. In the normal wired communication by applying x y routing, you can reach there. Now, let us see how a deadlock free routing is possible. So, we will try to understand how a deadlock routing happens. First it goes to a hamming distance 1.

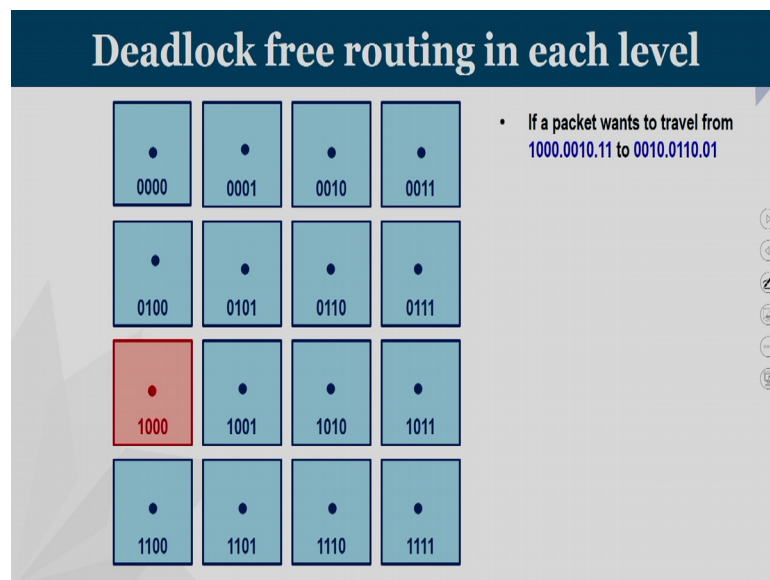
So, 1 0 0 is differing from 0 0 0 0 only by 1 bit position. So, that can be directly communicated. So, you may be having a query why cannot the source and destination directly communicate? The reason is the source and destination is not having a hamming distance of 1. They differ in more number of bits. So, it is not possible. So, what we do is we are trying to jump into an intermediate wireless hub which is at hamming distance. So, this is the one that is being chosen.

(Refer Slide Time: 17:30)

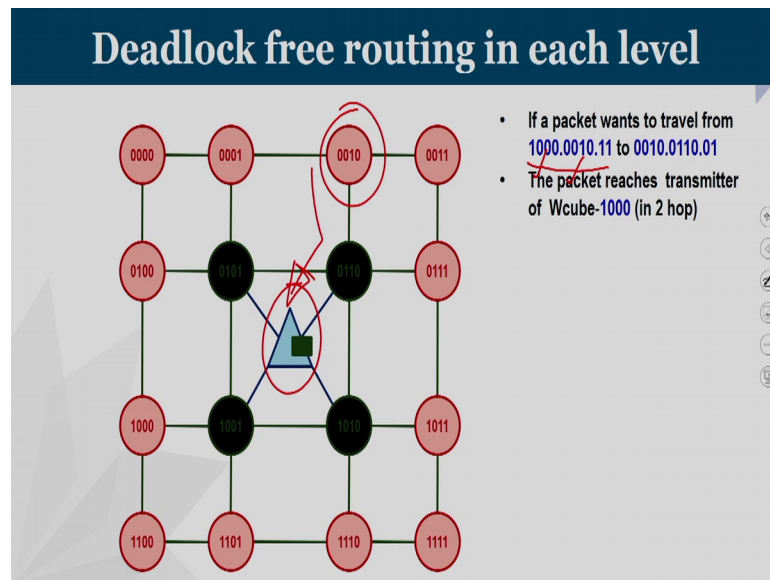


And then from 0 0 0 0, we are moving in to 0 0 1 0. So, in 2 hops, I reach the W cube of my destination.

(Refer Slide Time: 17:40)



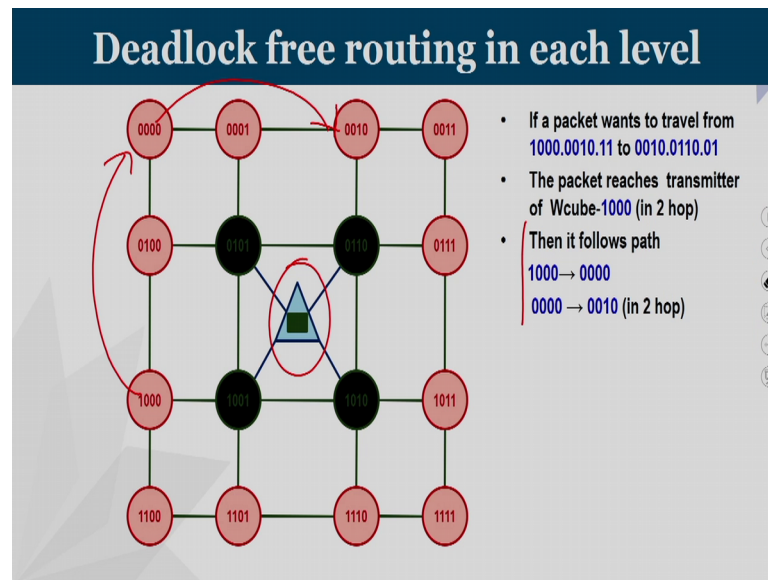
(Refer Slide Time: 17:42)



Now, from 1 0 0 0, I have reached to that particular point. So, from 1 0 0, it is 0 0 1 0. So, we have to identify which is 0 0 1 0; this is the sub node in W cube 1 0 0. So, it has to reach the transmitter.

So, the packet reach the transmitter of W cube into 2 hops; 1 and 2, this is the way how the packet reaches. And then this is the way how the packet reaches. Then it follows the path 1 0 0 0 to 0 0 0 0 that is the path, first path; hop number 1 and hop number 2. So, the packet is now at the destination, at the destination W cube. Now, we will see what happens in destination W cube.

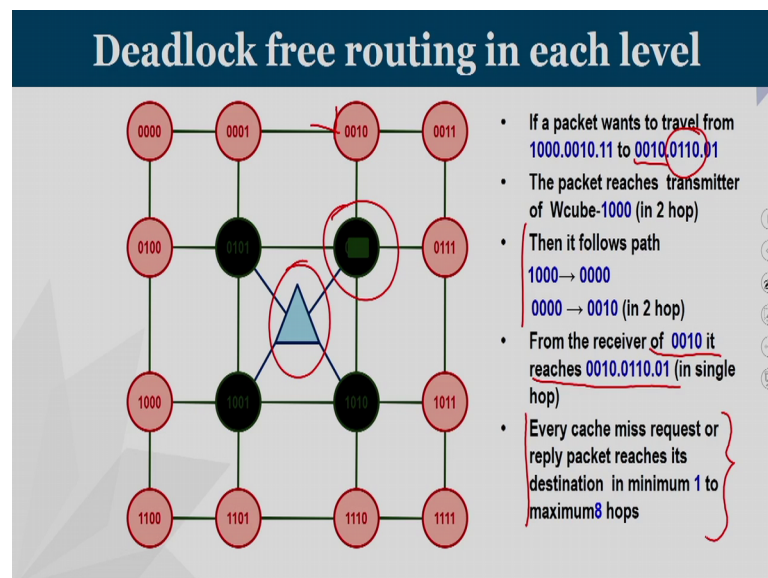
(Refer Slide Time: 18:25)



Packet has reached the transmitter of the destination W cube. Now, from the transmitter depending upon the destination node, so, we have already reached W cube, 0 0 1 0.

Now, from the transmitter I have to reach to 0 1 1 0.

(Refer Slide Time: 18:45)

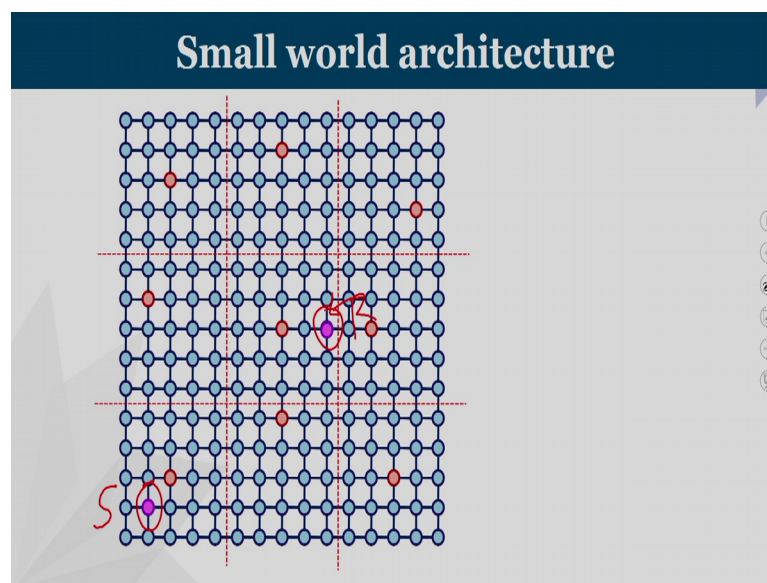


So, 0 1 1 0 is this. So, packet is going to move from the receiver of 0 1 0 0 1 0, the packet is moving in to 0 1 1 0 in a single hop because the destination is a cache node. So, the packet moves there. Now, once the packet is going to move you are going to get the

data. So, every cache miss request or reply packet reaches its destination in minimum 1 to maximum of 8 hops.

So, because of this peculiar structure, whatever be the transition that happened from one end of the chip to another, it can happen at most with 8 hops because in 2 hops you can reach this transmitter and then you may take maximum 3 or 4 hops to reach your W cube of the destination. Once you reach the receiver of the W cube of the destination, you can shift it to the other locations which are there within that W cube.

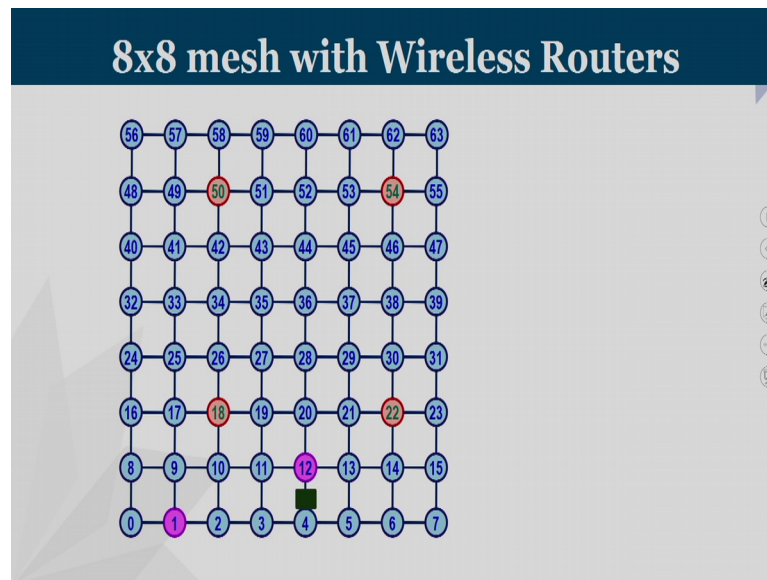
(Refer Slide Time: 19:43)



Now, let us try to see at different kind of an architecture. So, far we have seen a concentrated mesh of W cube architecture. Now, we will see small world architecture given this large NoC, we are going to subdivide this NoC into 9 independent regions and each of this region will have its own wireless transmitter. So, when you look at the slide, we can see all the nodes are marked as blue, but one node per region is marked with orange color and the peculiarity of them is these are nodes which are having wireless points. They can transmit and receive data in the wireless manner.

So, these are heterogeneous nodes these nodes can send data in wired to its neighbors as well as it can send data wirelessly from to its adjacent wireless routers. Now, consider this violet color you have a packet that is from source and it is moving to this node as the destination. So, from this point to this point my data has to travel. Some animations are being prepared.

(Refer Slide Time: 20:54)



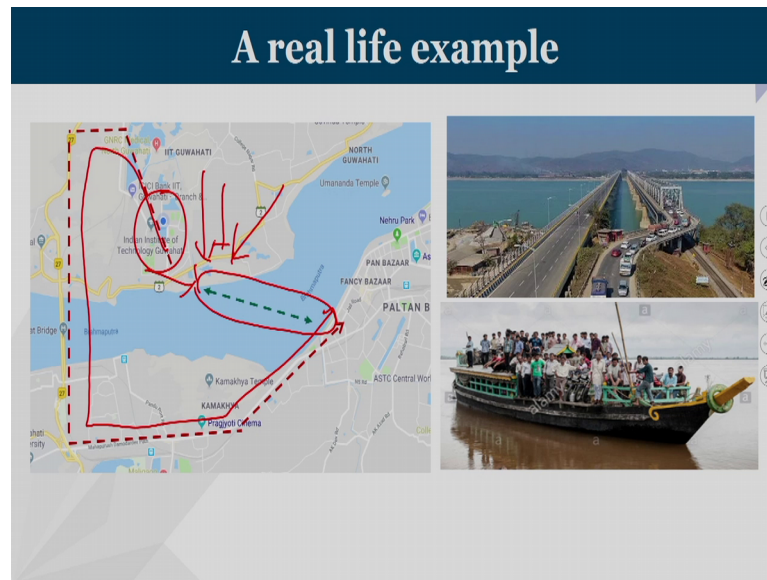
So, if you look at the animation you can see that first the data which is there is going to travel in x y routing to the nearest wireless router. From the wireless router of the source region there is a single hop jumping to the wireless router of the destination region. So, your packet moves like this and from there the packet is moving to the corresponding destination; this is how in small world architecture data flows. So, you have an initial x y transmission to the nearest wireless hub. From the wireless hub of the source region to the destination, it is a wireless transmission.

And once it reaches destination wireless hub again you apply xy routing. Now, it is an 8 by 8 mesh with a wireless router. This can be a scenario where routing number 18, 22, 50 and 54 are marked as wireless routers. So, these routers have normally 4 of it is wired neighbors plus you have few wireless neighbors also. Now, consider the case that is a packet wanted to travel from 1 to 62 in an 8 by 8 mesh NoC with 4 wireless hubs. So, packets travel from 1 to 18 in xy routing from 18 to 54 by wireless routing and from 54 to 62 again by xy routing. Now, consider the case that your source is 1 and destination is 12. In this case, both the source and destination is relatively closer, there is no need to go through a wireless point. So, the package will travel from 1 to 12 directly using xy routing.

So, it is kind of a hybrid routing, whether should I take a wireless routing or not. So, that is a adoptive that is being embedded into the routers intelligence. The router should be

intelligent enough to understand should I stick on to wire wired routing that we have seen from 1 to 12 or should we stick on to a wireless routing that we have seen from 1 to 62.

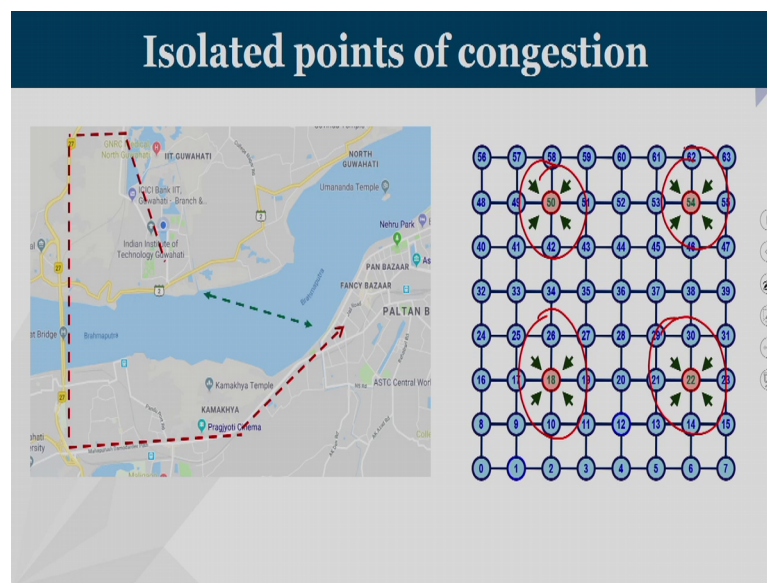
(Refer Slide Time: 23:02)



So, I would like to draw your attention to a real life example. So, IIT Guwahati is located on the banks of river Brahmaputra. So, this is the campus and it is adjacent to the Brahmaputra river and Guwahati city is on the other side of the river and it is being connected through the Sirat bridge which is been shown. So, this is the path through which we have to go through the city that is a path through which we travel via road and here we have the bridge.

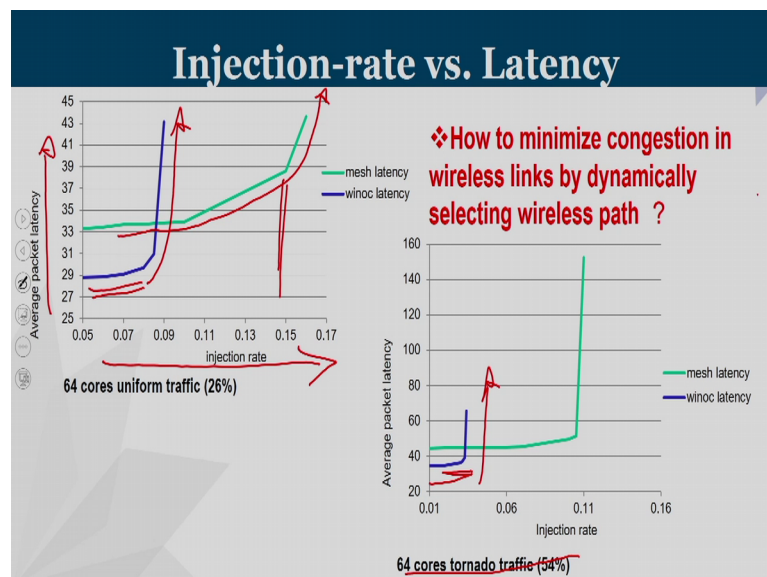
So, it is a pretty long path. We do have free services also where we can take a ferry such that you can cross Brahmaputra river and that is much more time consuming. This is a way how the ferry looks like, but when too many people come to use the ferry like say- we have people coming from different regions of this north bank of Brahmaputra river then the ferry is going to be over crowded.

(Refer Slide Time: 24:05)



We had similar scenario like this when in a wireless router when many of them are trying to send it through wireless, the wireless hubs are going to be highly crowded. So, we get isolated points of congestion and we need to write adaptive routing such that these isolated points of congestion are been eliminated.

(Refer Slide Time: 24:23)



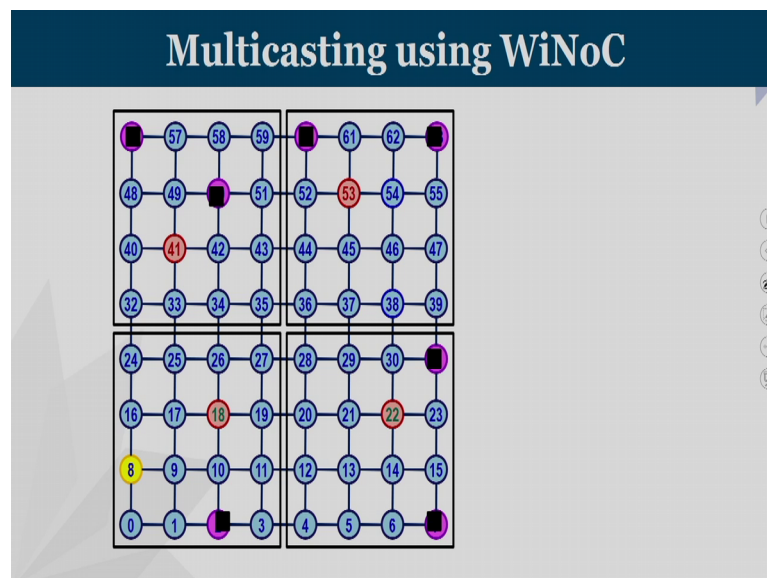
So, some experiential result shows that this is a result which we showing packet latency on the y axis average packet latency and x axis loads shows the load in the network.

So, the x axis is injection rate and the y axis is average packet latency. This is a load versus latency graph. What you see in green color is the latency when it is a normal mesh NoC. We can that as the load increases, the latency is also going to increase and it goes to an exponential fashion this is called a saturation point.

Now, when you look at the wireless NoC, an NoC with a wireless structure that is a blue graph that is called winoc NoC. We can see that since we take up the shortcut path through wireless, the latency is very low initially for very load. But as the load increases more number of packet try to send it through the wireless node leading to early saturation. So, the take away is using wireless NoC will save you latency if the load is very less.

As the load increases there will be more number of packets which will combit to reach the wireless NoC point and there will be a queuing delay because already the wireless transmitter is sending packets which has reached there early. This scenario is visible in the transpose. So, in the case of a tornado traffic also where you can see the initial latency is less but it is saturated very early. Adaptive routing techniques are been explored in this context.

(Refer Slide Time: 24:23)



Now, let us try to see one important application of wireless NoC; one of the application is called multicasting. What do you mean by multicasting? Multicasting is the scenario in which there is one source and I have to send data to multiple destinations. So, one data

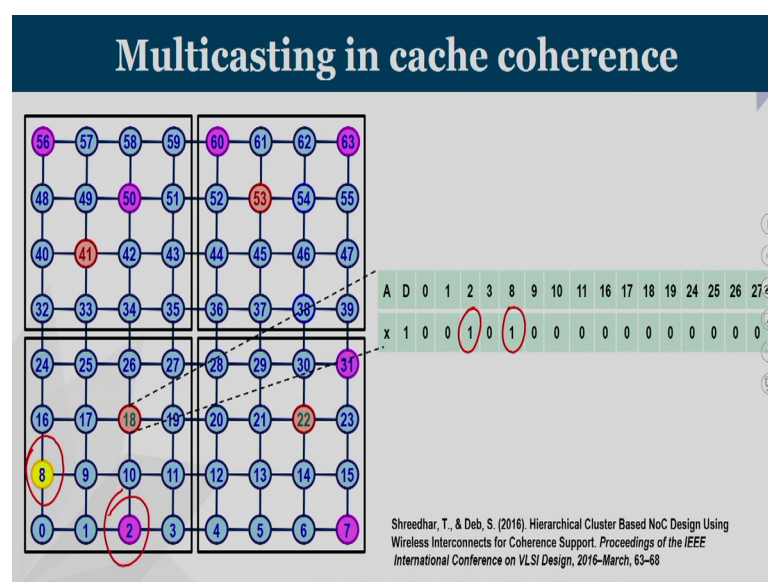
has to reach multiple destination. Can I make use of the available wireless NoC infrastructure to facilitate multicasting forces a bit more faster?.

So, consider the scenario. Let us say we have 4 wireless access points at 18, 22, 41 and 53, it is a small world architecture. Now, there are the nodes where you see with violet colour 1, 7, 31, 50, 56, 60 and 63 are the destination. A data which is starting from 8 that is a source has to reach all these nodes which are marked with violet color. And how I am going to use? So, a packet from 8 will go into 18 and at 18 you are going to broad cast it. So, three messages will be created from 18.

So, these three messages are going to reach into the appropriate wireless nodes and you create duplicate packets from these wireless hubs; such that all the packets are going to reach into the appropriate destinations. So, this wireless infrastructure helps you to broadcast and then selectively multicast once it reaches each of these quadrants.

So, in applications we need certain scenarios where a same data has to reach to multiple points and this is very much needed in cache coherence. So, where one node will validate or invalidate the data that is located in many other nodes together. In that case, having a wireless NoC infrastructure is going to help us.

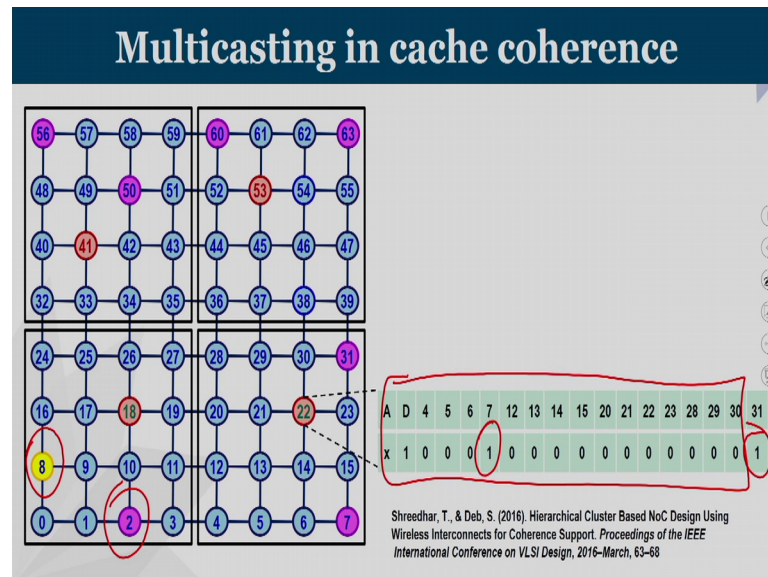
(Refer Slide Time: 28:01)



So, how it works? The multicasting in cache coherence; let us say we have a packet from 8 that goes all the way to 18. The packet starts from 8, it reaches 18 and it is having some

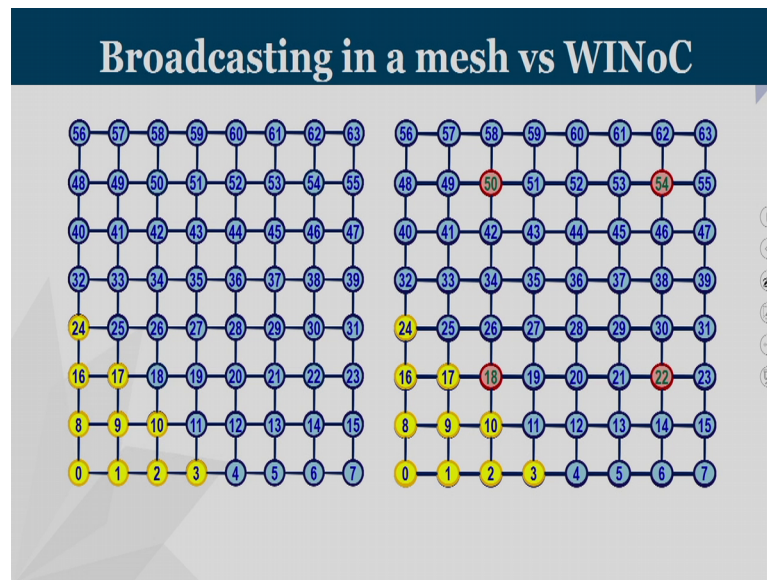
information which will tell where it should be going. So, this meaning is the packet has to go to 2, packet has to go to 8. So, 2 and 8 are the nodes which should get these packets in this quadrant. So, 18 we will come to know where all this data has to be sent.

(Refer Slide Time: 28:33)



Let us say a 18 is going to send the packet 22. Once the packet reach 22 it knows, which are the quadrants in 22 that are to get the data? So, 7 and 31 is going to get the data. So, once the packet reaches 22, you have to create appropriate forwarding mechanisms such that the data reaches 7 with one packet and data reaches 31 with the other packet. So, this kind of a specialized header structure will help the wireless routers to take appropriate actions in forwarding a multicast packet.

(Refer Slide Time: 29:12)

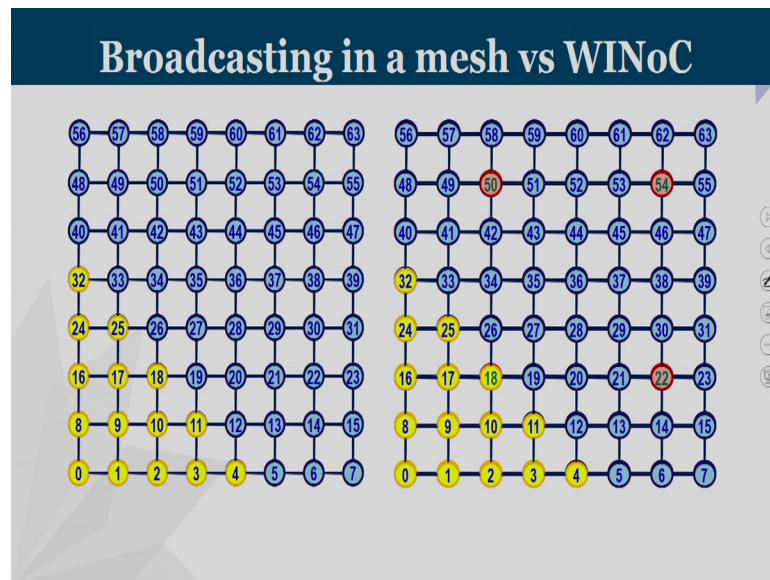


Now, coming into the broadcasting scenario, what do you mean by broadcasting? Let us say a packet from one node has to reach all other nodes. Sending clock signals is an important requirement of broadcasting. Powering up the system, booting; you have to send a special command to all the router such that they should start, it is booting process power up reset. So, these kinds of signals which are known as broadcasting which should reach all the nodes and that can be better improved with the help of wireless infrastructure.

The left side will show how a broadcasting message from 0 will reach all other nodes in a conventional two dimensional mesh NoC. The right side is a wireless NoC with the 4 nodes 18, 22, 50, and 54 having wireless access point. So, how the same broadcast message starting from 0 is going to reach all other nodes? So, that is being shown with a help of an animation. Let us say in a data which start from 0 will reach node number 1 and 8 in one hop. So, these are nodes which are already received the data. So, nodes that have received the data it is marked in yellow color and nodes that are yet to receive the data is marked in blue colour.

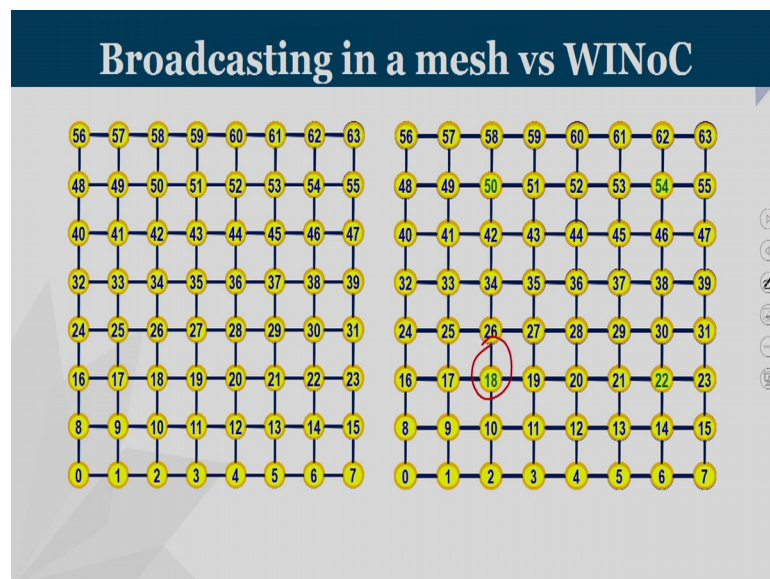
Let us see how this yellow colour is going to progress. In the next 2 clock cycle, the neighbors of those who are marked will become yellow, the next clock cycle, the next set of neighbors are going to get the data. So, this is same in the case of left side wired NoC and right side the hybrid NoC.

(Refer Slide Time: 30:52)



Next they are going to reach the same point.

(Refer Slide Time: 30:55)

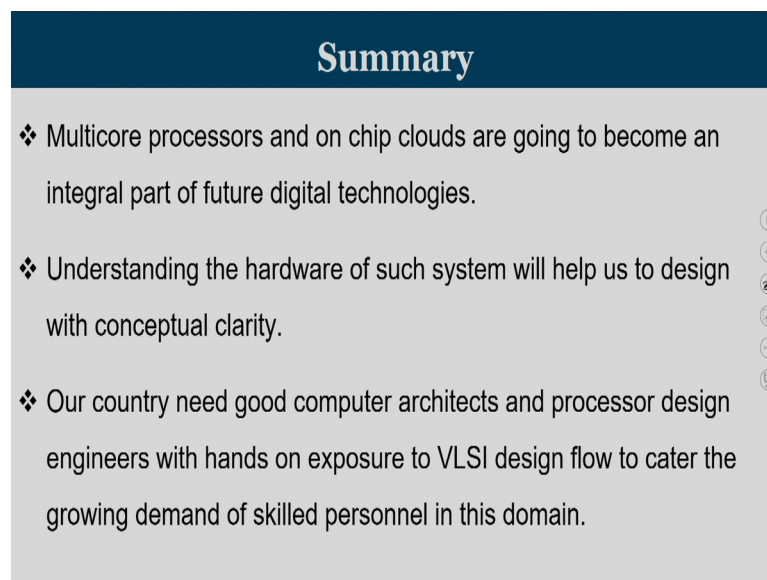


Now, what we can see that it has already reached the wireless NoC router. So, in the next hop, 50, 54 and 22 also will get the data and then they are going to send it to their neighbors. So, with this the wireless NoC has transmitted the data or rather broadcast at the data to all its nodes, whereas in the case of wired still more nodes are pending. So, it takes a few more cycles for the data to reach all the nodes. In this way by having a wireless NoC infrastructure, your data will reach all the nodes rather very fast. So, wireless NoC

can be used for broadcasting can be used for multicasting and can be used for better communication from end to end.

So, we have given a brief overview about what is the infrastructure of a wireless NoC by W cube structure by small world architecture. So, the summary of this is multi core processors and on chip clouds we have seen.

(Refer Slide Time: 32:05)



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.

Lot of computers in a single chip that is what is known as on chip clouds are going to become an integral part of future digital technologies and understanding the hardware of such system will help us to design with conceptual clarity. So, our country needs 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.

If we want our country to have our own processors, the digital technology has to be empowered. It is not about writing software's alone. We can design our own chips. We spend lot of money in importing chips. So, why cannot we develop our own processors? Initiatives already taken and the people with the background of computer architecture are the need of the hour.

So, I hope that whatever discussions we had in this course will really help you to go deeper into the topic and further read about architecture related research materials. Students can work for projects in this domain, faculty can mender students to do take up

small computer architecture related projects. So, I will give you a couple of advices how to work with this domain those who wanted to do research, those who wanted to do projects in this domain, where can you get material from.

(Refer Slide Time: 33:28)

How to explore computer architecture ?

- ❖ Refer to IEEE/ACM transactions & journals
 - ❖ IEEE TCAD, IEEE-TVLSI, IEEE-TOC, JPDC, JSC
 - ❖ ACM-TODAES, ACM-TECS, ACM-TACO
- ❖ Refer to top tier conferences
 - ❖ ISCA, HPCA, MICRO, ASPLOS, PACT, DATE, DAC, ICCAD
 - ❖ NOCS, NoCArc, ISVLSI, ICCD, ASPDAC, VLSI-SoC, GLSVLSI,
 - ❖ HiPC, VLSID, VDAT, ISED
- ❖ Familiarize open source architectural simulators
 - ❖ GEM5, Multi2sim, Sniper, Tejas,
 - ❖ BooSim, DRAMSim, Usimm
 - ❖ Cacti, Orion
- ❖ Model the architecture in simulators and implement them using HDLs, Verify sub-modules in FPGA kit explore further ...

So, how to explore computer architecture further? So, of course, the good research is always found in good transactions. We have IEEE and ACM transactions and good journals. IEEE Transactions on Computer Aided Design, Transactions on VLSI, Transactions on Computers, Journal of Parallel and Distributed Computing, Journal of Supercomputing, then ACM Transactions on Design Automation of Electronic Systems, ACM Transactions on Embedded Computing Systems, Transactions on Architecture and Code Optimizations; these are the top journals in this domain.

So, if you wanted to know what is the kind of research happen in this domain, I request you to go and read the research materials available on this and learned for multi core computing, caches, network on chips, storage on chips, etcetera. We have very good peer reviewed conferences also in this domain, the Symposium for Computer Architecture a very good highly rated conference, High Performance Computer Architecture, the Micro Architecture Conference.

Then we have the ASPLOS; Architectural Support for Programming Languages and Operating System, we have Parallel Architecture and Compilation Techniques, Design Automation and Testing Euro, Design Automation Conferences, International

Conference on Computer Aided Design; these are all well reputed conferences where research on computer architecture multi core computer architecture that interconnects caches, processors are all presented.

Then specifically to network on chip, we have network on chip symposium, network on chip architecture workshop; these are exclusively for NoC domain. Then we have International Symposium on VLSI, International Conference on Computer Design, Asia South Pacific Design Automation Conference, VLSI System on Chip Conference, Great Lakes Symposium on VLSI Design. So, these are all good conferences which will give you a good material. These are considered to be tier 2 conferences and the first one is tier 1, this is tier 2.

And then these conferences happened in India; High Performance Computing Conference, the VLSI Design, VDAT, International Symposium on Electronic Design; these are all conferences within India itself. And then we need to go through the material that is been published and remodel existing work. This is a basic step to work with the architecture related projects whatever is the existing work that is already mentioned in these papers.

Try to redo them and that is being done with the help of simulator and some of the graphs that we have discussed in this course is been obtained the values are obtained by working with the simulators. So, we have full system simulators like GEM5, Multi2sims, Sniper, Tejas and all. We have micro architectural simulators which will deal with only certain aspects of the architectural like Booksim, DRAMSim, Usimm and all. So, full system simulator we will model your processor, your OS, your memory hierarchy, your interconnects; everything will be there, but if you wanted your study to be focused on certain topics, then the micro architecture simulators will help.

And then we have power tools which will model this. Cacti is a tool which will help you to estimate the power associated with a storage systems. Orion will help you to find out power associated with interconnection systems. So, read research material try to work with the open source simulators, model the existing work on the stimulators, try to see the numbers what they claim in this research papers are really obtaining in your experimental modelling also.

And once you come up with good designs this system simulators will give you; what is approximate time, throughput, latency, speed up and all and then whatever is architectural change that you have proposed we have to model in hardware description language. So, model the architecture in simulators and implement them using HDLs and verify the sub modules in FPGA kits to test whatever design that you have proposed is indeed really working or not.

So, with this I complete this lecture. So, we have started from the fundamentals of micro processors, we learnt a bit of pipelining, the advance features of pipelining, what are super scalar processors. Then we looked into on chip storages, we understood what our cache memories and then the interconnection mechanisms. We had in between a quick overview of DRAM systems. So, with this course I have tried my level best to give an overall picture about what is multi core computer architecture all about.

So, I requests majority of you to explore this domain further. There were a good number of candidates already done this registration and on a day to day basis working with the assignments. So, prepare well for the end exam course that is scheduled on Aug on October 7th. So, lot of questions we get. If you are very much thorough with the assignments that will really take you up to 30 to 40 percent of these questions, go through the slides, go through the study materials.

So, attending all this video lectures surely it will be you can score more than 70 to 80 percent of marks. There will be couple of questions which will check your deeper understanding. So, overall those who have put in systematic effort, I am sure that this course is going to help you in understanding and appreciating the hardware of these machines and students who really do well at the end of this course I can consider a couple of them to join our research group for internships and we welcome faculty members also to have active collaboration with the computer architecture group at IIT Guwahati. So, we look for really motivated students and faculty members to have collaborations in this field.

I hope you enjoyed this course. We were planning to have more number of courses related to architecture domain in these coming semesters. So, wish you good luck.

Thank you all the best.