



先进封装与集成芯片

Advanced Package and Integrated Chips



Lecture 9 : SoC/Chiplet Interconnect
Instructor: Chixiao Chen, Ph. D



Course Project

- Option I: Presentation (I guess most students would choose)
 - Pick up one paper from the pool (<https://365.kdocs.cn/l/cpbY4B34cUnj> first come first served), send a email or told me in wechat group, I will updated the paper selection in time.
 - If you want to pick a paper not listed, please send me an email to get granted.
 - Slides and Presentation: Prepare a 15min slides to introduce the work and 3-5 min Q&A. Send the slides to the homework mail after presentation.
- Option II: Project
 - Complete a component (AXI-streaming controller+PCS, PHY-TX/RX pair, Clock Generation,) design according to UClle standard (not in a group)
 - Report and Presentation: Prepare a 5 min introduction and send the design report (<10 pages A4) to the homework mail, NO need for slides, show your draft report during presentation is OK.
- 1st Deadline: Tell me your choice by 5.6 on class / mail / wechat.

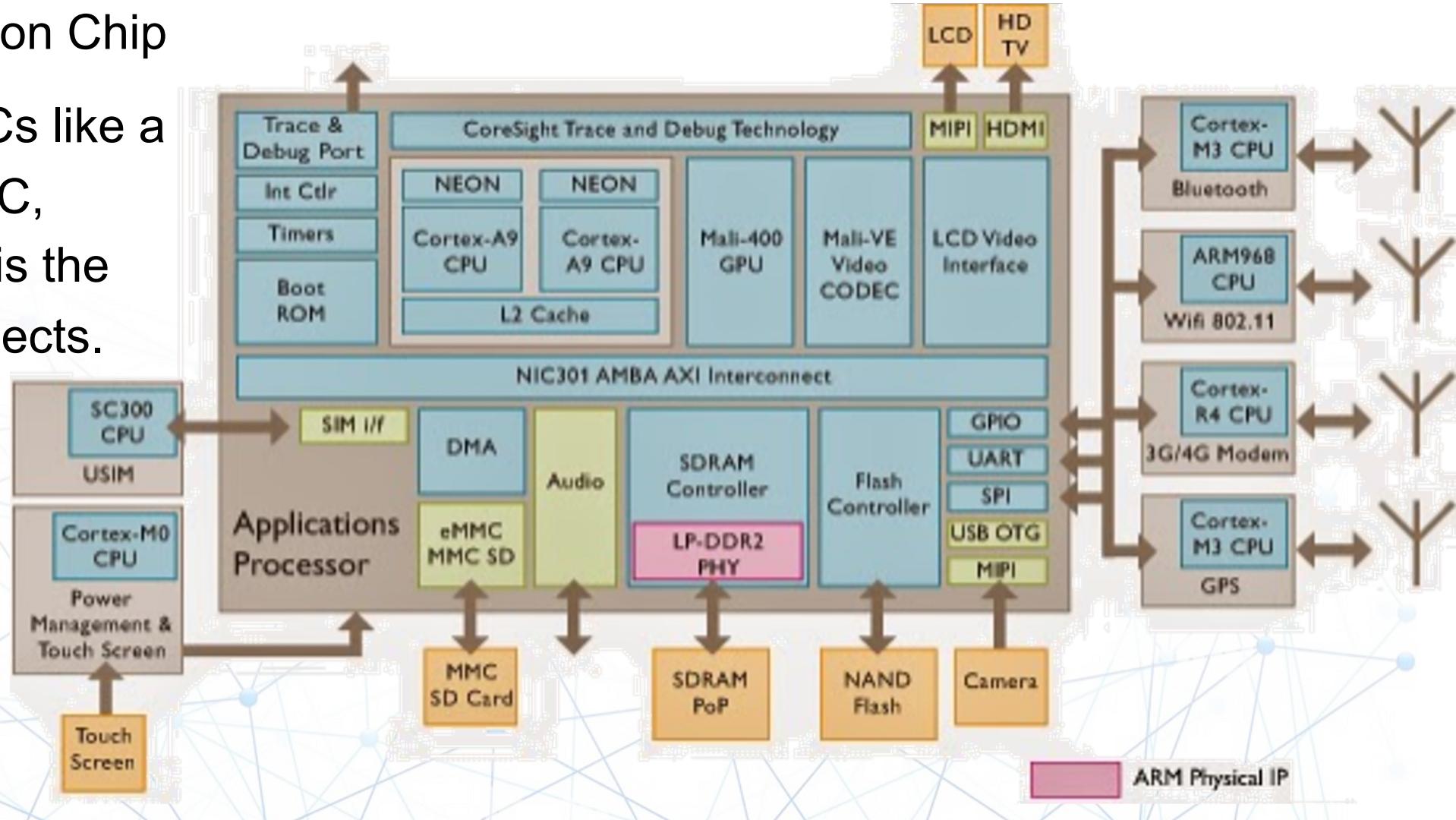
Overview



- Review on SoC Interconnect
 - Bus based on-chip communication
 - Network-on-Chips (NoCs)
- From SoC Peripherals to Chiplet Interconnect
 - Case Studies
 - How advanced packaging affects System Performance

Communication-centric Design for SoC

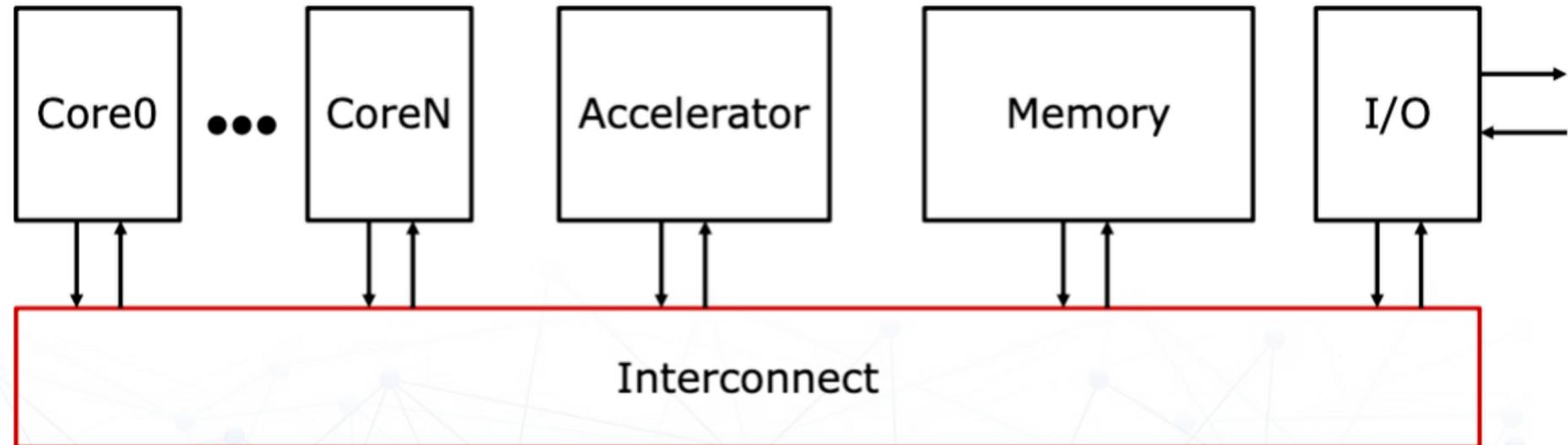
- SoC = System on Chip
- For modern SoCs like a smart phone SoC, communication is the most critical aspects.
- It affects performance, power, area (PPA) & time-to-market.



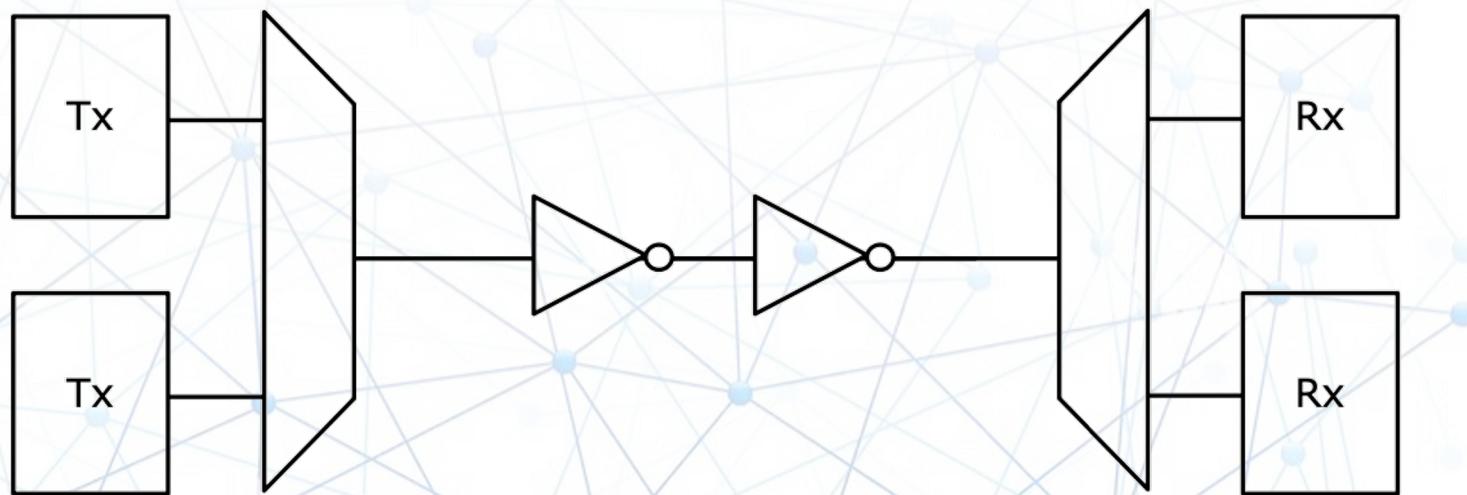
On-Chip Interconnect: Physical and System View



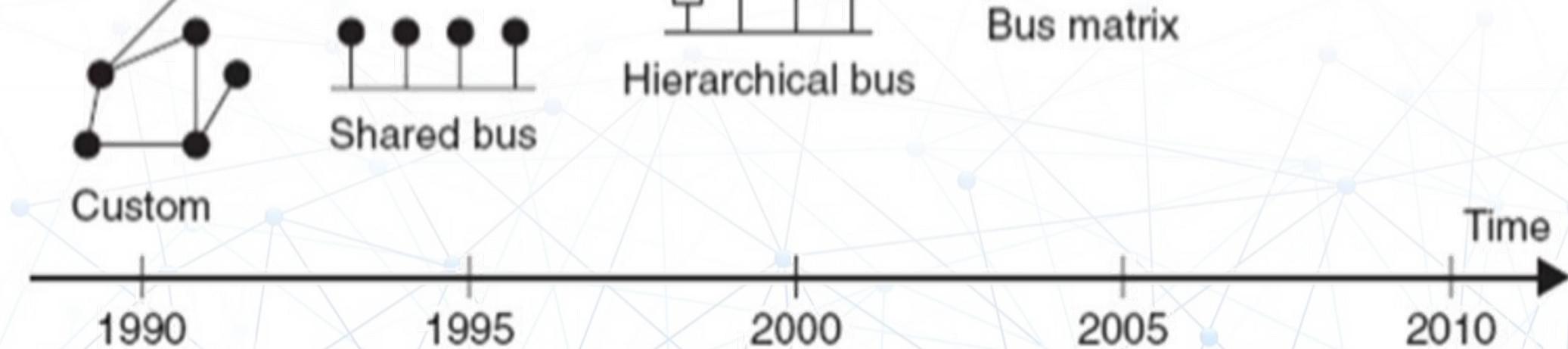
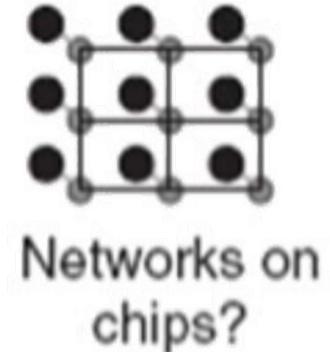
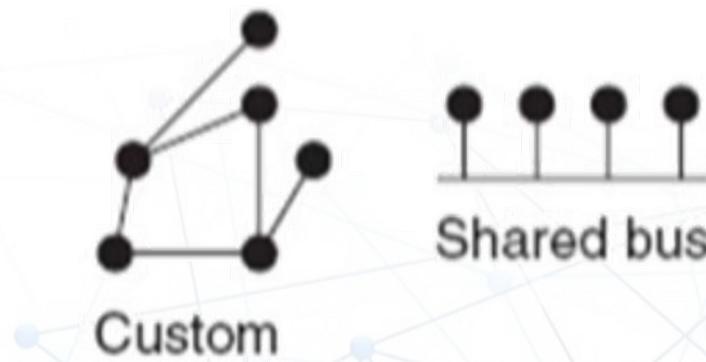
- Interconnect: communication infrastructure connecting all IPs together



- Physical Implementation
Interconnect: to avoid too many wires, we need to multiplex data over a group of shared wires



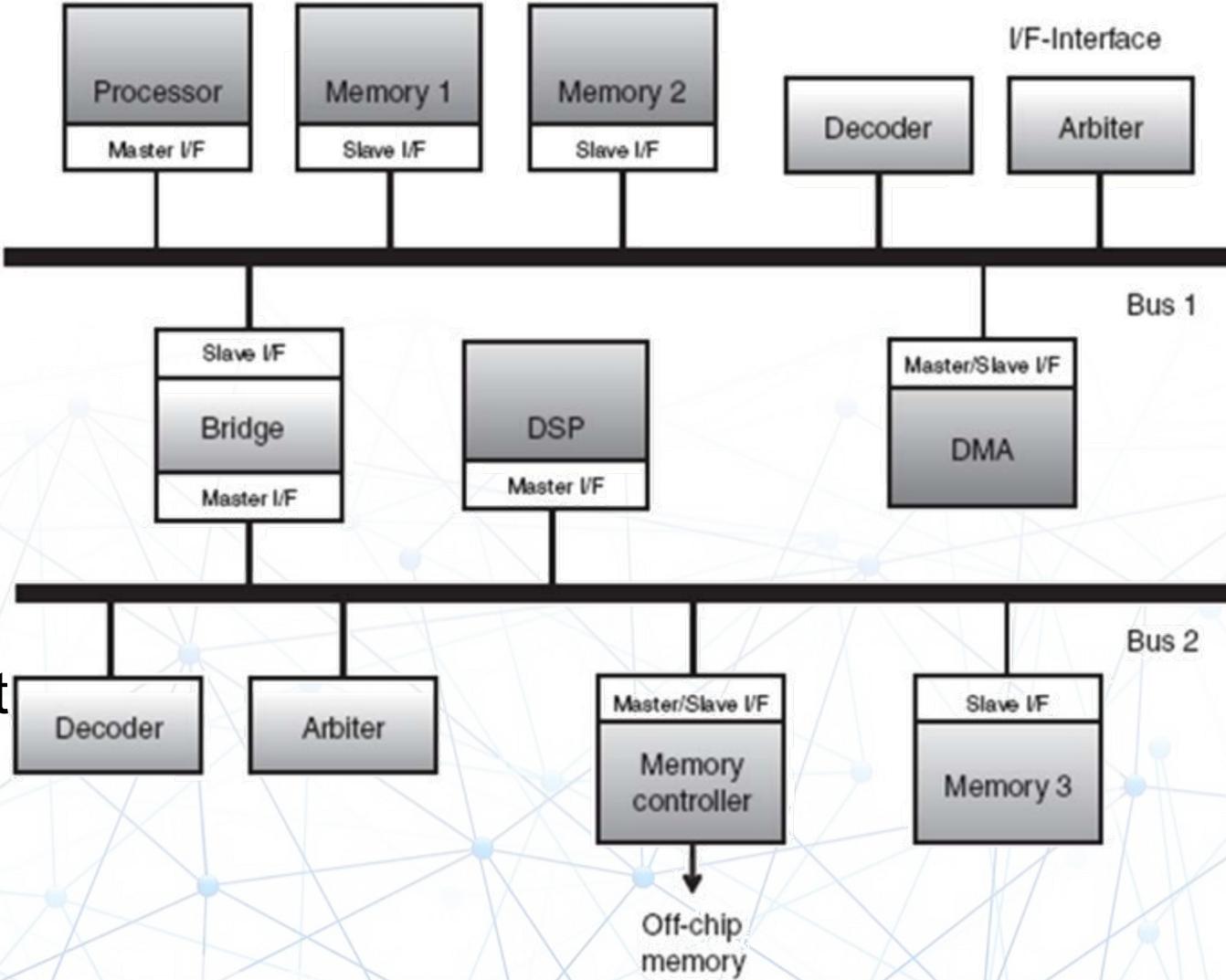
Evolution of on-chip Interconnect



- Bus is the most simplest and widely used SoC Interconnect.
- Bus definition: A collection of signals to which multiple IPs are connected

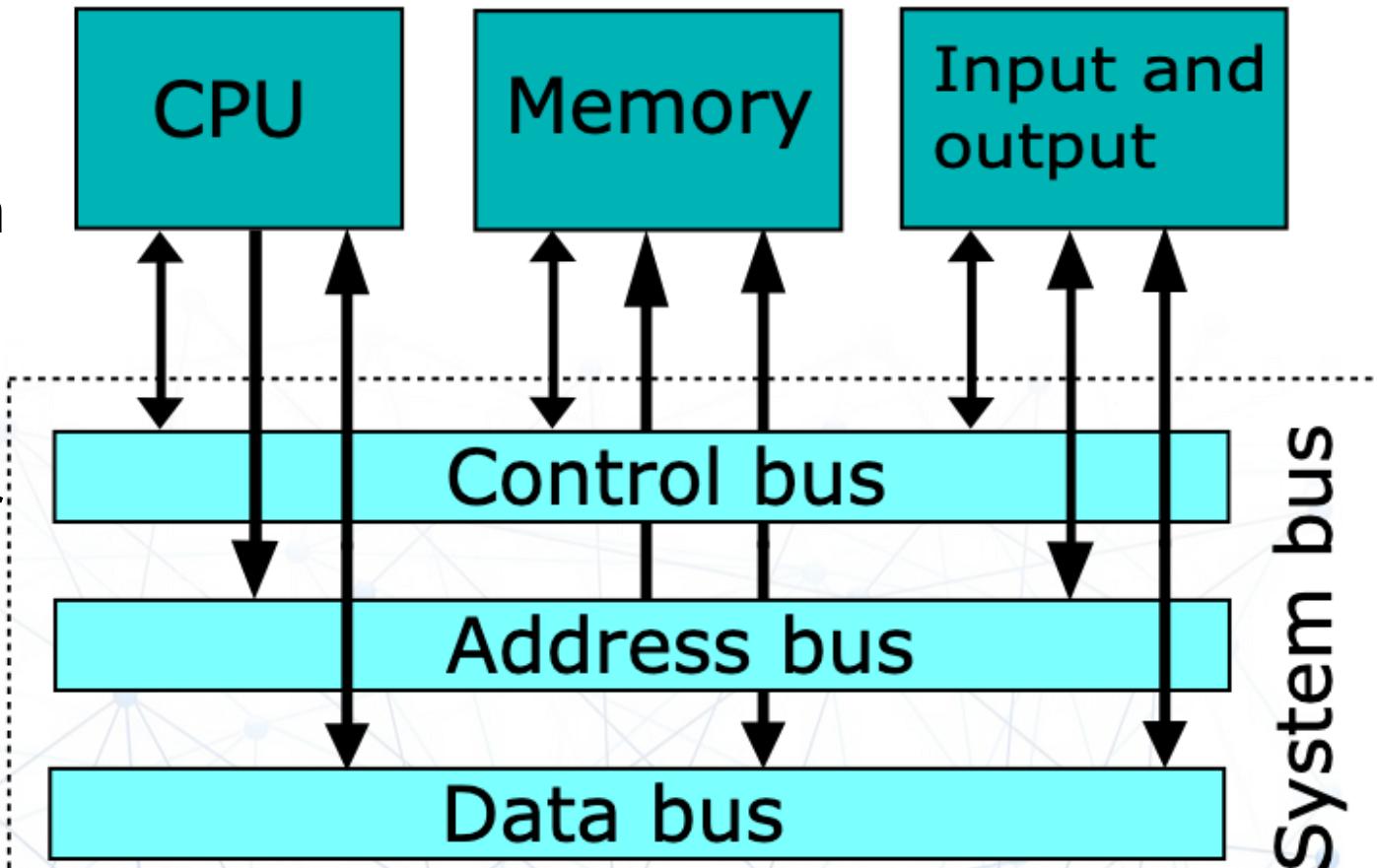
Bus Terminology

- Master: IPs that initiates a read/write data transfer.
- Slave: IPs that only responds to incoming transfer requests.
- Arbiters: Control bus operation by selecting master to grant data transfers.
- Bridge: Connects with different bus, acting as slave one side and master on other side.

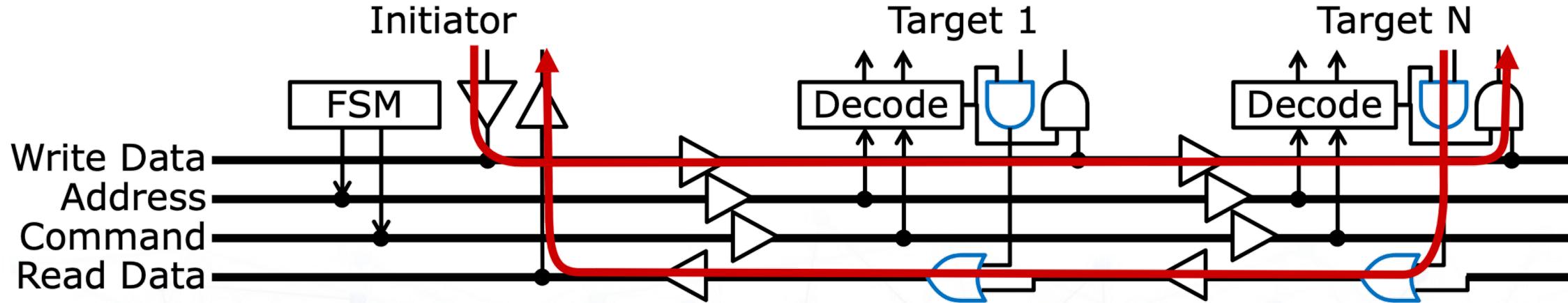


Bus signals

- Address: transfer data's source and destination, uniformed encoded for all on-chip IPs, driven by masters only
- Data: real information sent and received by bus, can be shared or separated for read and write
- Control: includes request and acknowledgements, specify different type of data transfer (R/W, burst, cacheable, byte mask, ...)



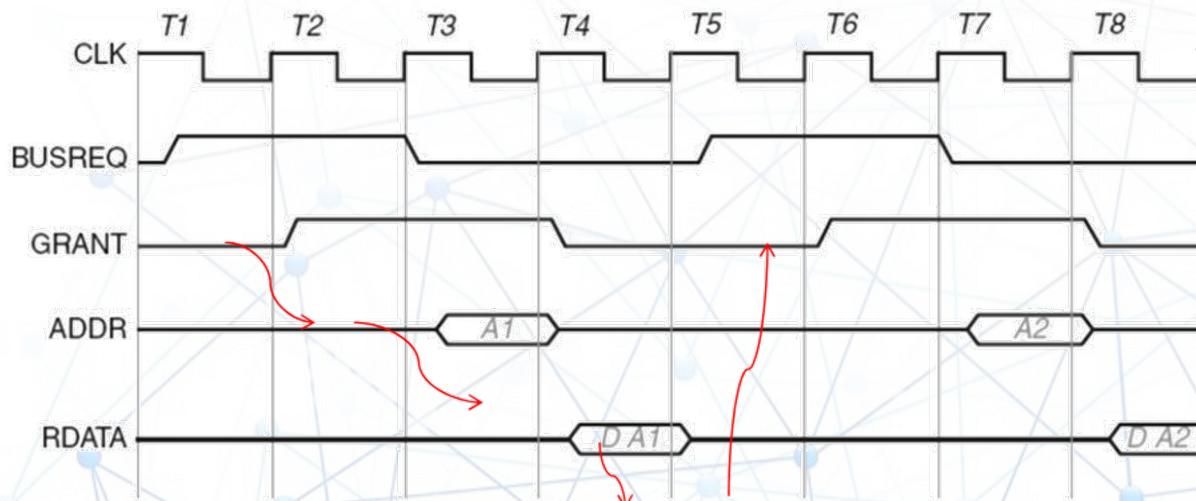
Basic Bus Circuit Implementation (Digital)



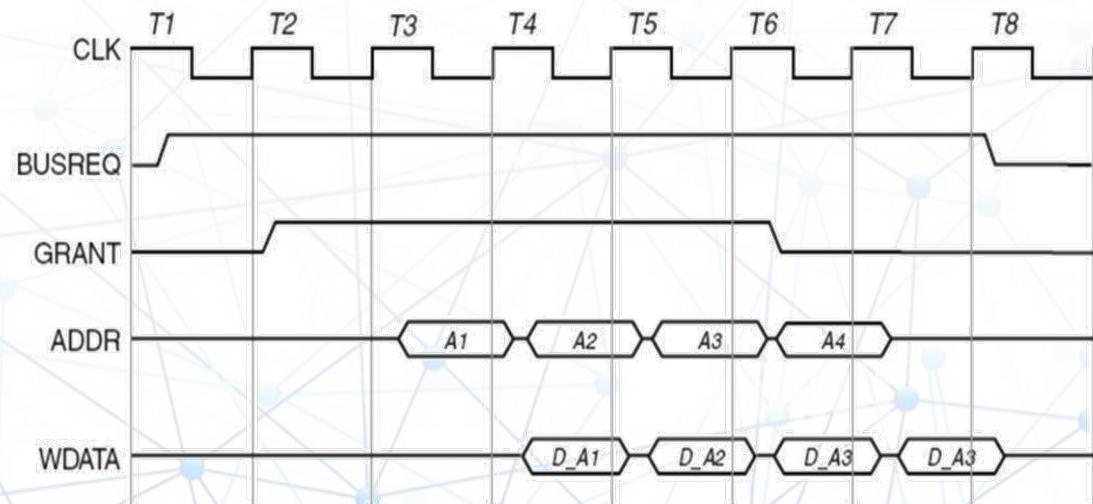
- Historically tri-state drivers (high impedance to disconnect) is not friendly in recent CMOS digital circuit design.
- Current bus implementation adopts differentiate Read & Write Data channels to replace tri-state drivers
- Pipelining stage can be inserted to prevent long-distance transition and latency.

Bus Transfer Modes

- Single data transfer (w/o pipeline)
 - first request to access bus
 - access granted/acknowledged
 - sent address and control signals
 - send/receive data in subsequent cycles



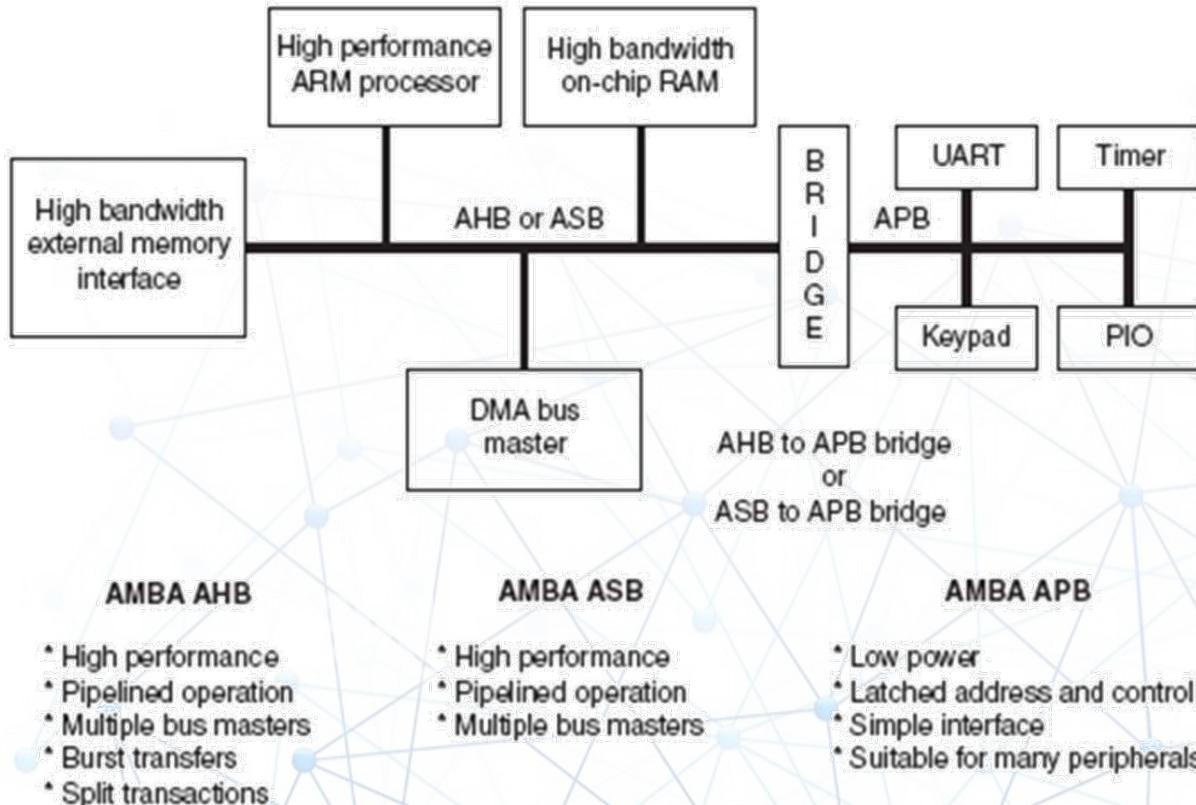
- Burst data transfer (w/i pipeline)
 - send multiple data with only one cycle control (save time for arbitration)
 - Continuous data transfer for recent AI applications



- There should be **a protocol or standard** for bus communication.

AMBA Bus Protocol

- Advanced Microcontroller Bus Architecture, open standard but owned by ARM
- AMBA 2: AHB Advanced High-performance Bus, AMB 3&4: AXI Advanced Extensible Interface



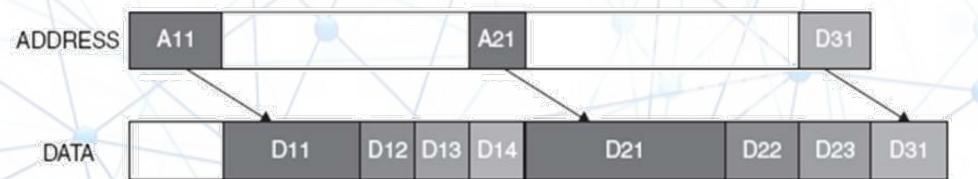
- **AHB Burst**

- Address and Data are locked together (single pipeline stage)
- HREADY controls intervals of address and data



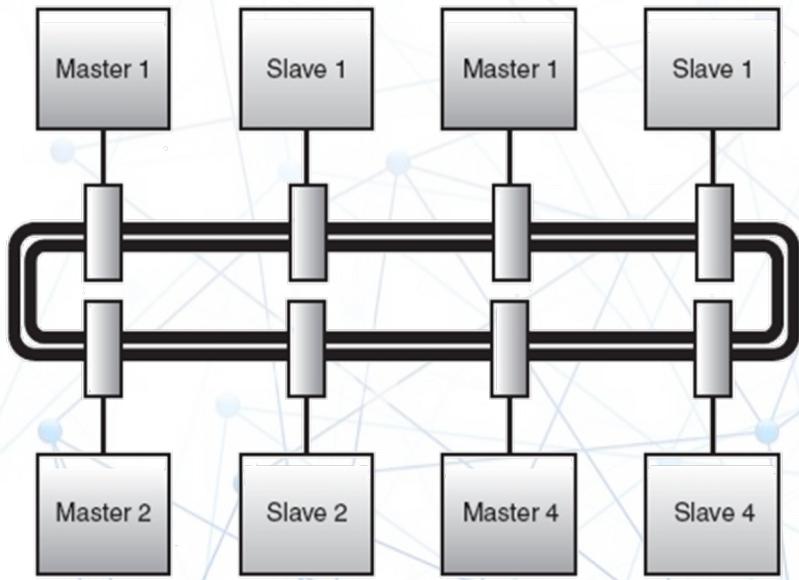
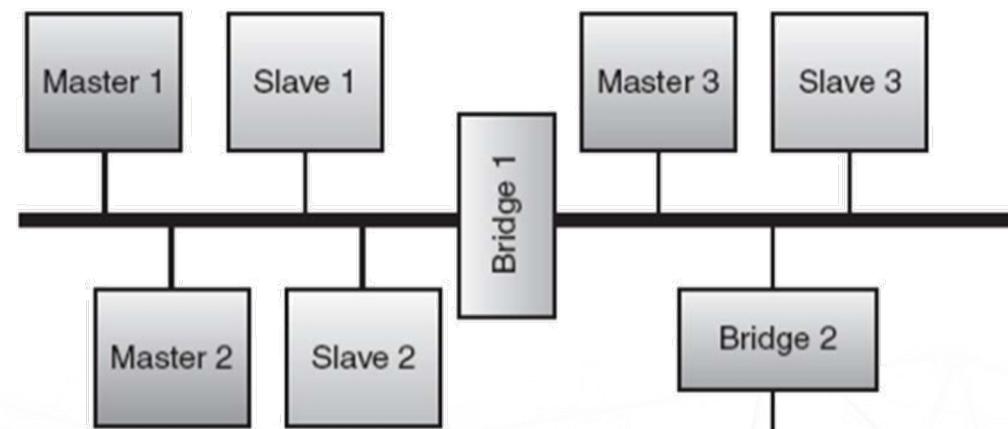
- **AXI Burst**

- One Address for entire burst

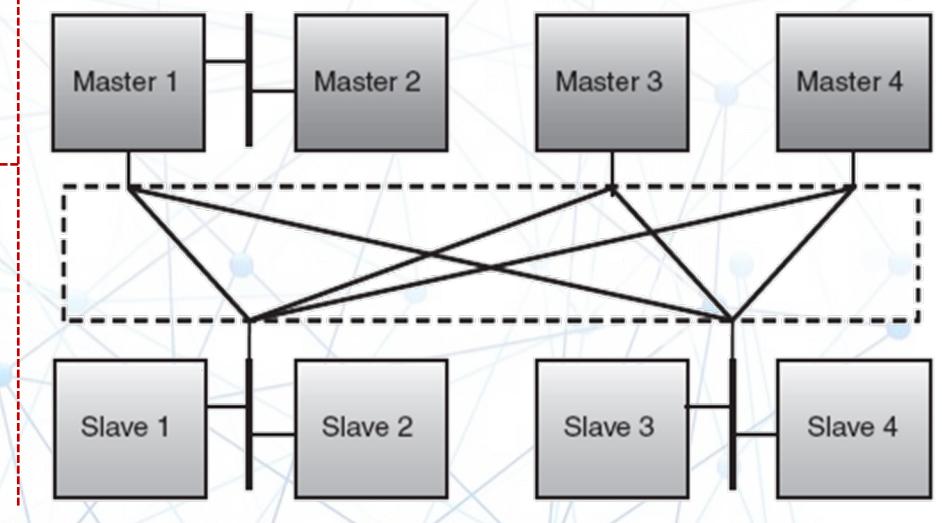
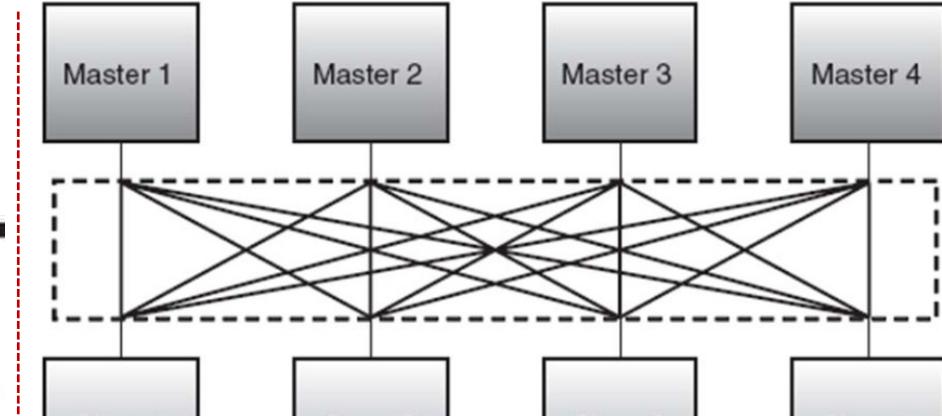


Bus Topologies

- Hierarchical shared bus for different clock SoCs

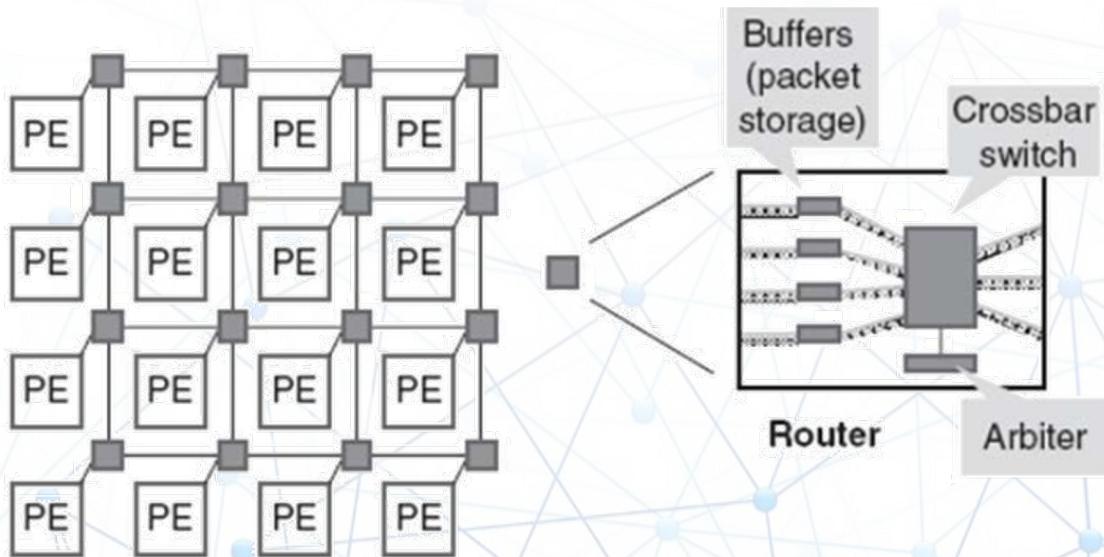


- Fully/partially Crossbar matrix
- Ring bus



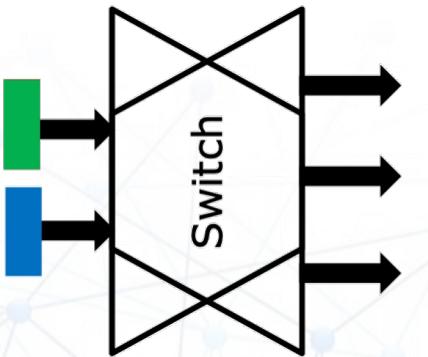
Networks-on-Chips (NoCs)

- Network-on-chips is a packet switch based on-chip interconnection schemes designed by a layered methodology. “route-packets, not wires”
- NoCs use packets to route data from the source to destination PE via a network fabric that consists of routers, as well as links.



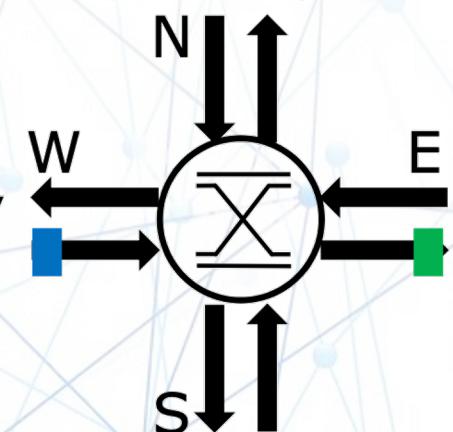
Switch:

little or no queuing,
ports not directional



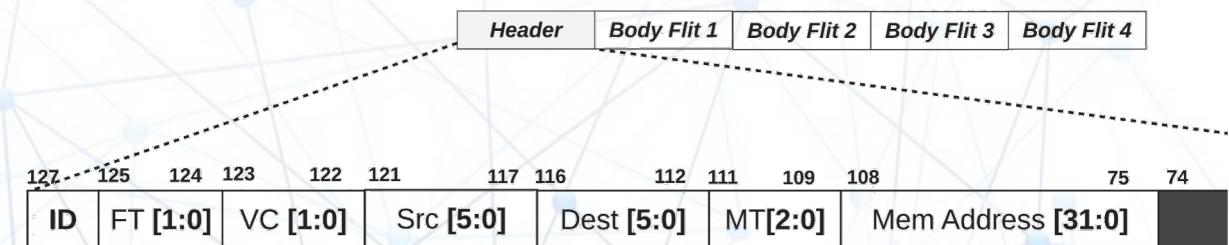
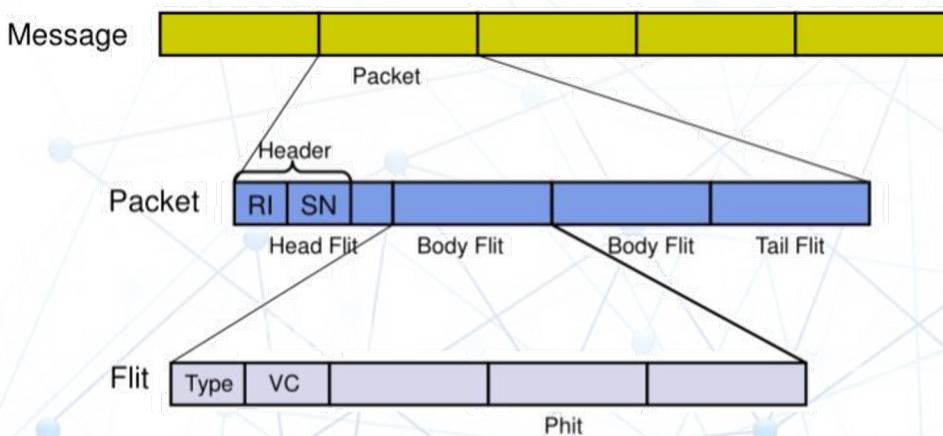
Router:

often more queuing,
directional ports



Packet and Flit

- Each core sent messages/packets including information how data flows through routers in the NoC.
- Phit (Physical control digit) is a unit of data that is transferred on a link.
- Flit (flow control digit) is unit of switching.

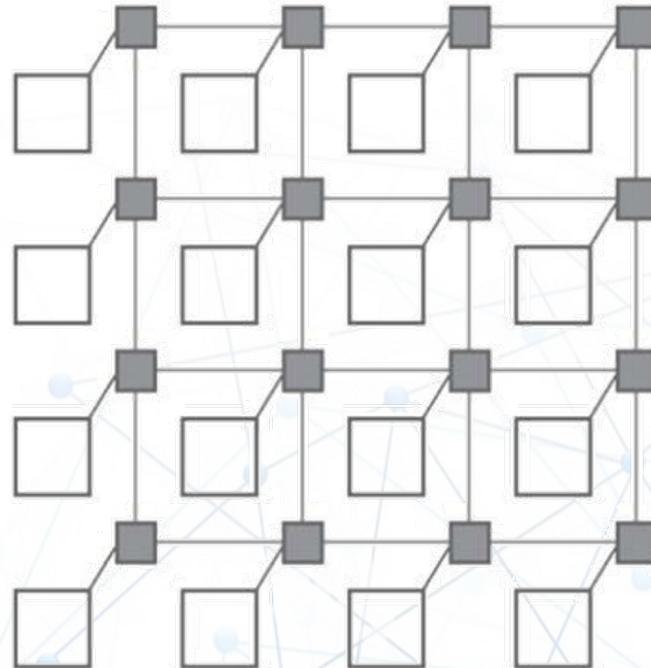


Messages, Packets, Flits and Phits are handled in different layers of the network protocol

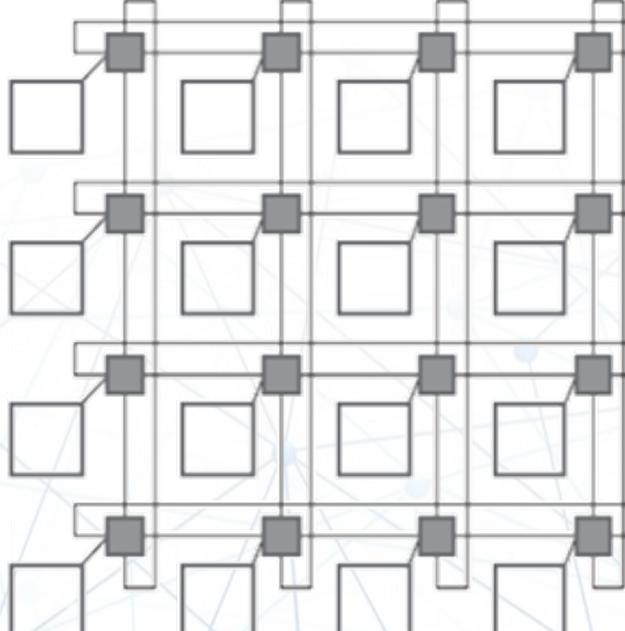
NoC Topologies



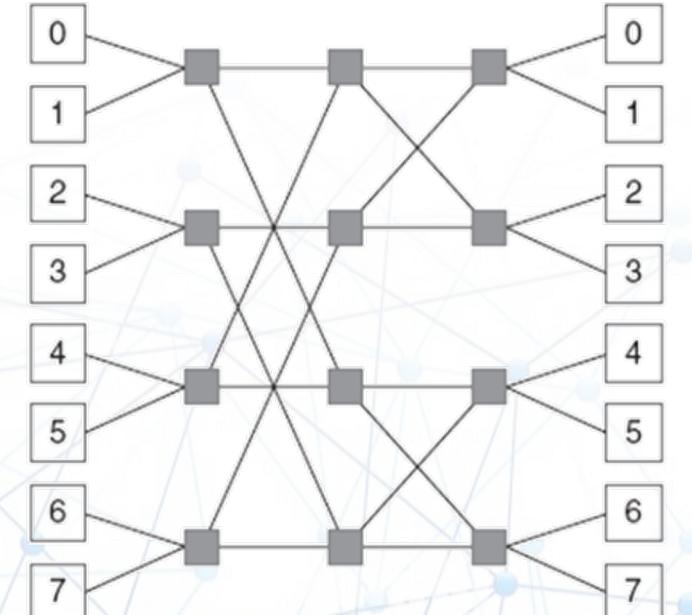
- NoC is kind of an advanced bus, which is more friendly on scalable architectures.
- Many NoC Topologies is used: 2D mesh, Torus, butterfly, fat-tree,...



2D Mesh: equivalent link length



Torus : Ring connection for nodes
at the edges, long end-around link



Butterfly

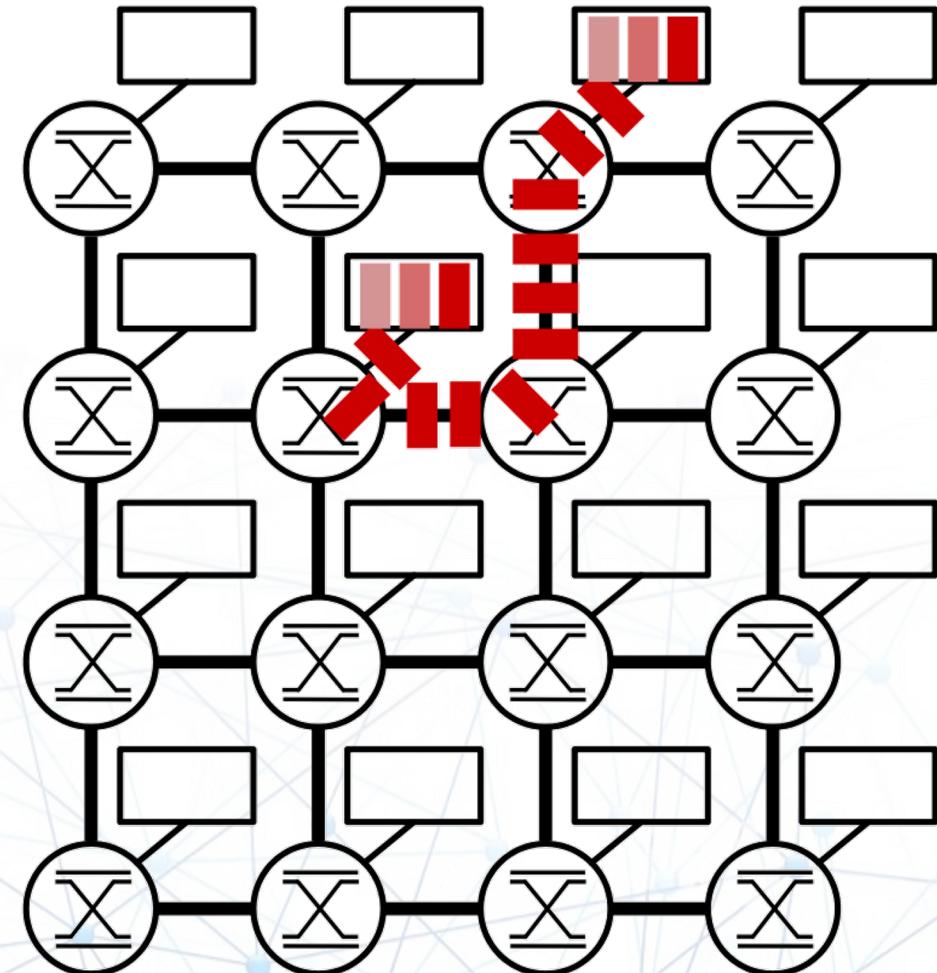
Routing: Packet-Switched Based Interconnect



- Data grouped in packets
- 1 packet : 1 or more data words
 - One word is a "Flit" (Flow-control unit)
 - e.g.: 1st flit = base address & command
2nd flit & next = data burst



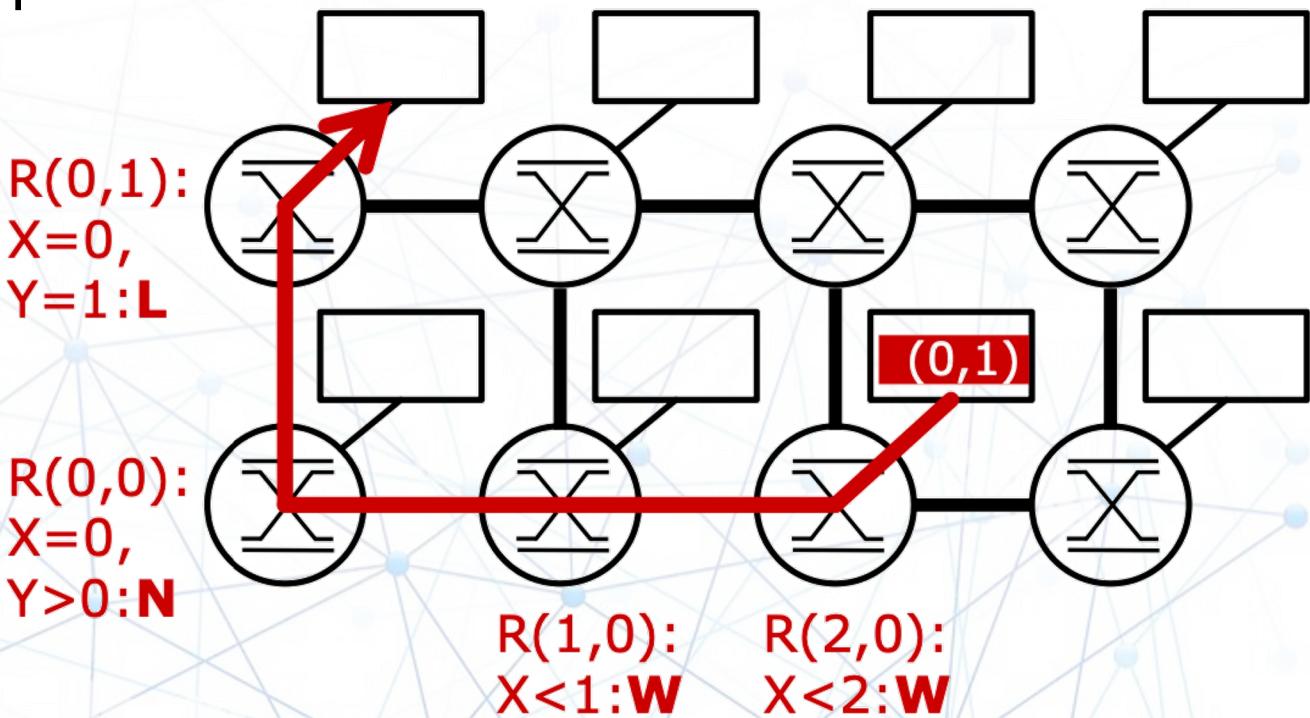
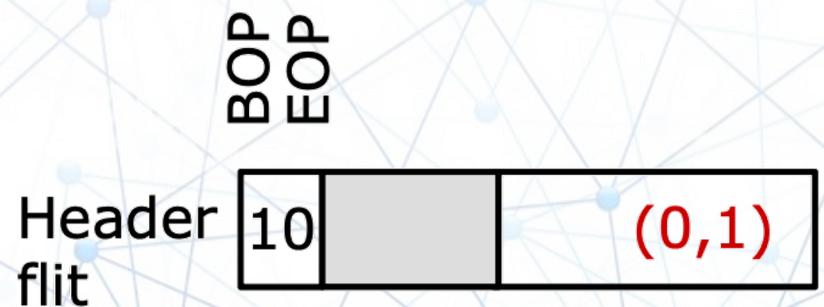
- Each packet contains routing information in the header flit
- Packet routing is atomic
 - No flit interleaving with other packets
 - Can span multiple blocks



Courtesy by Y. Thonnart, ISSCC 2021 Tutorial 8

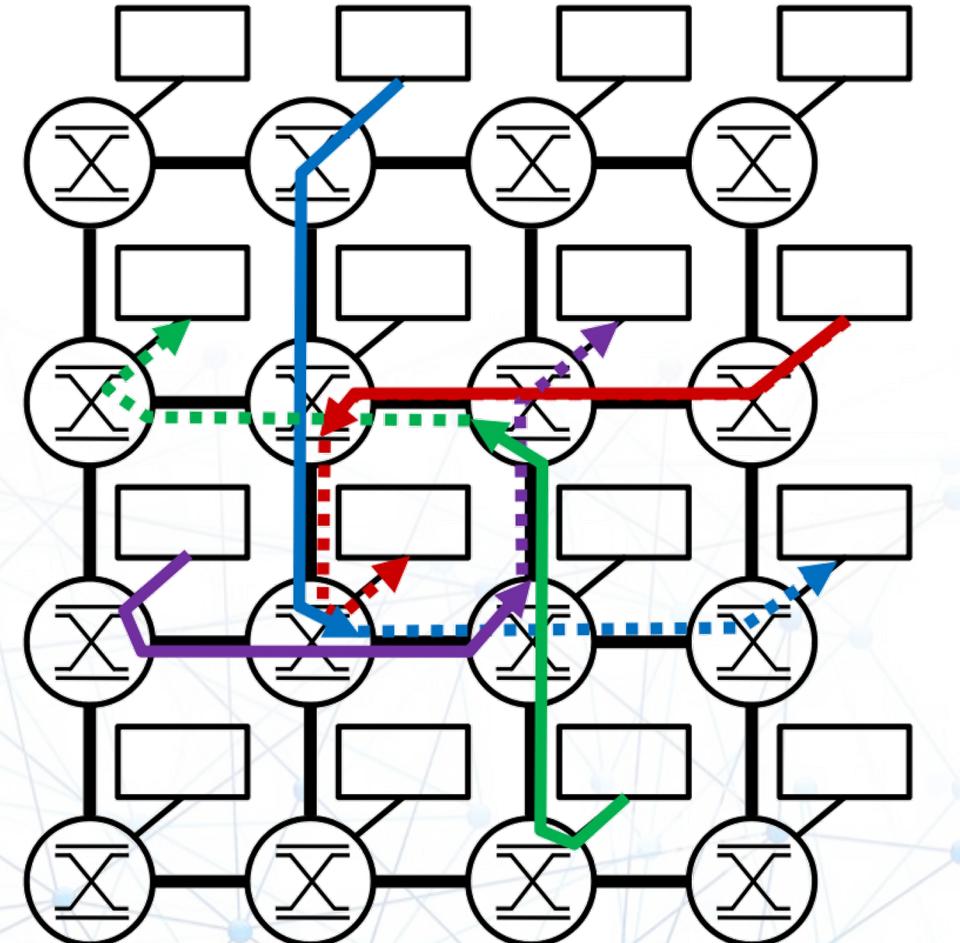
Routing & Packet Format

- In 2D mesh NoC, Coordinates-based Routing is most commonly used
 - Destination coordinates is located in header
 - Comparison to Router coordinates for X-Y routing
- Other methods includes indicate sequence of turning encoded in header flit
 - “East East North Local” ...



Routing: Traffic and deadlock

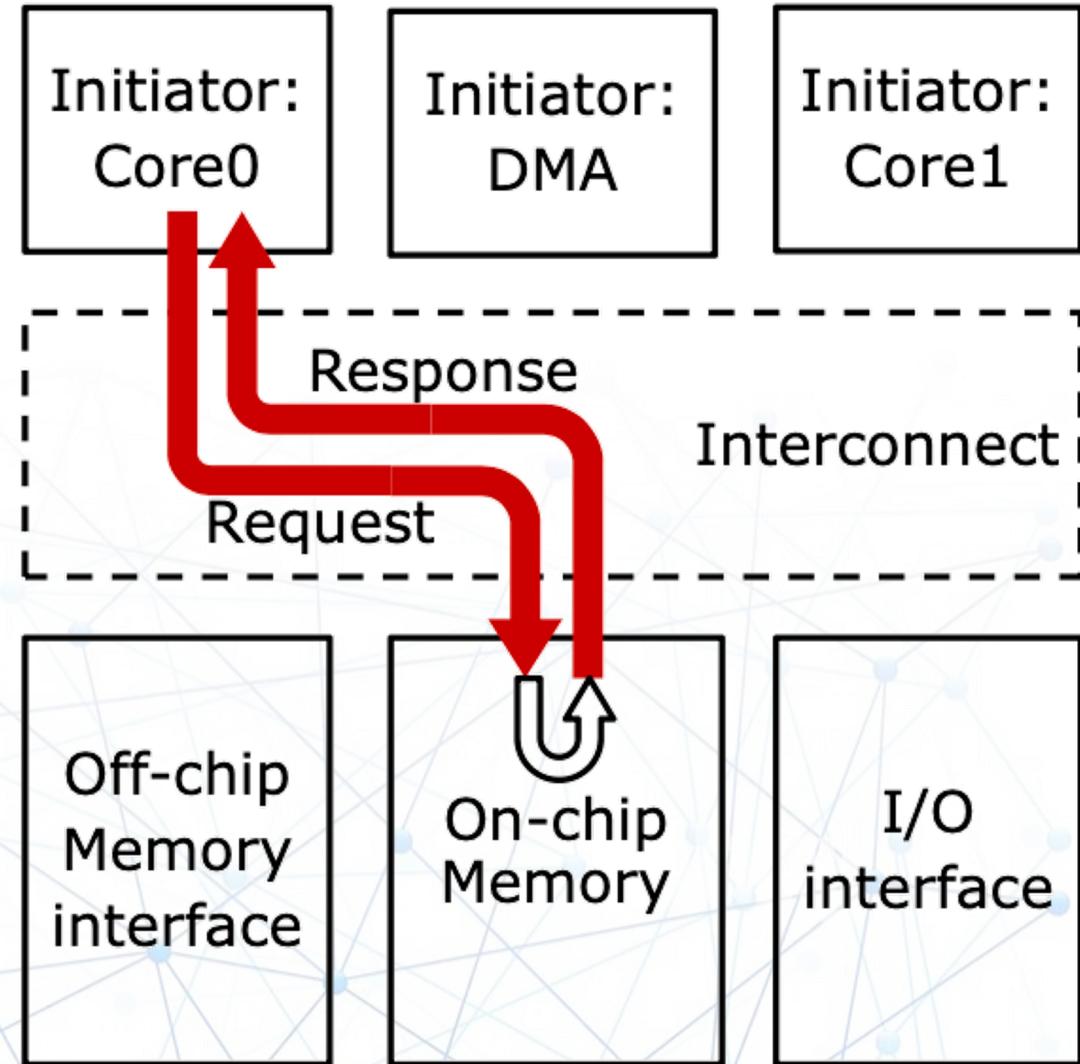
- Queuing behind a stalled packet waiting for an output
 - Potential trail accumulating
- Invalid routing algorithms may create cycles of stalled packets
- Potential deadlock
 - No packet can make progress to destination
- Solved by forbidding some turns
- E.g. X-Y routing: always X first



Courtesy by Y. Thonnart, ISSCC 2021 Tutorial 8

Transaction-based Interconnect

- Memory access is another common type interconnect, which normally use transaction based interconnect.
- Normally memory have specific protocols
- Memory Request/Response have different/independent channels, therefore multiple request are allowed.
- Memory Interconnect issues: coherence in multi-core architecture
- AMBA 5: CHI (Coherent Hub Interface)



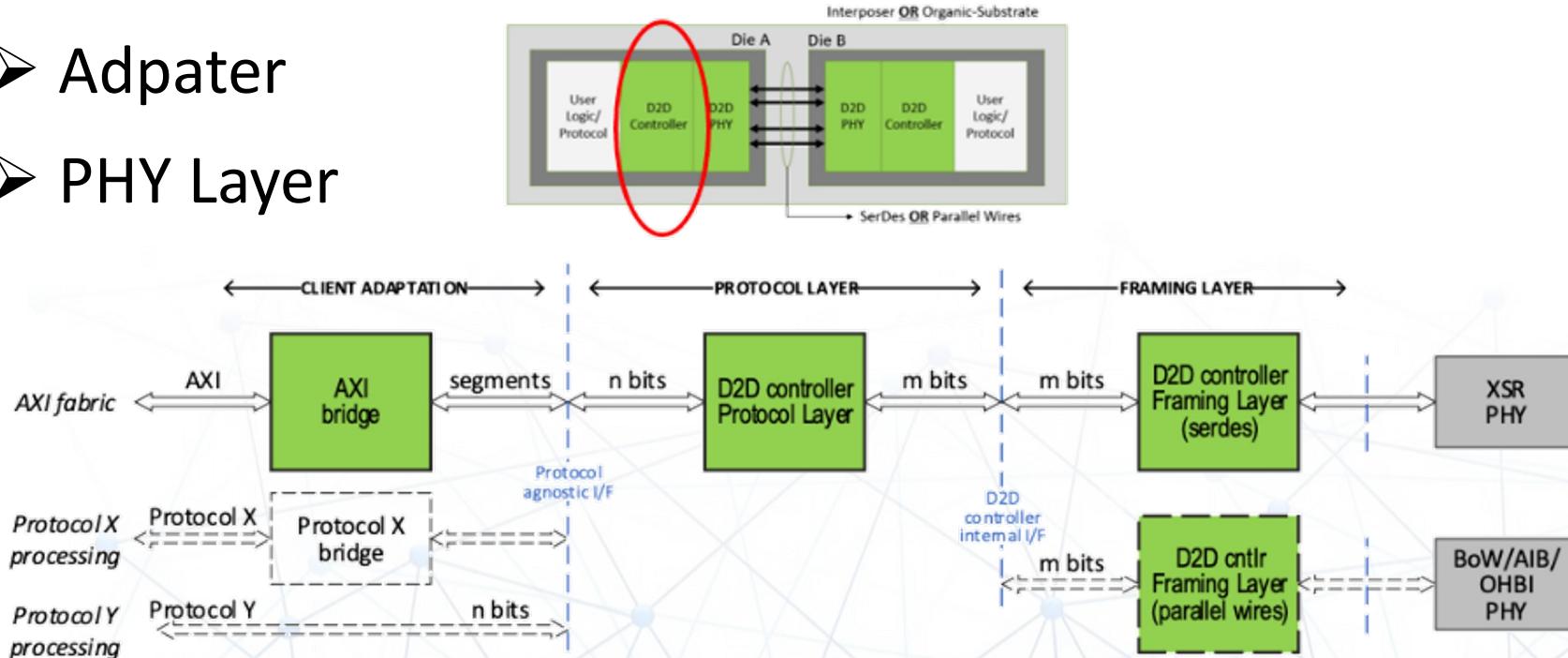
From Bus/NoC to Off-Chip Interconnect

➤ Protocol Layer

D2D Controller Features

➤ Adapter

➤ PHY Layer



Client Adaption

- Adaption of Client signal to Controller Interface
- Support for AXI4, TL, xGMII etc.

Protocol Layer

- End to End error free delivery
- Optional Flow Control
- Optional Re-transmission

Framing Layer

- Function depends on Parallel wires or SerDes
- Lane alignment/De-skew
- Optional FEC Engine

UCle Protocol Stack

Protocol Layer

PCIe 6.0 (256B Flit), CXL 3.0 (256B Flit)
CXL 2.0 (68B Flit), Streaming (raw)

Streaming CXL PCIe

FDI

Die-to-Die Adapter

Protocol arbitration/muxing, CRC/Retry
Link management, Parameter negotiation

RDI

Physical Layer

Link initialization, training, Power states
Lane mapping and remapping,
Scrambling, Lane reversal

Sideband

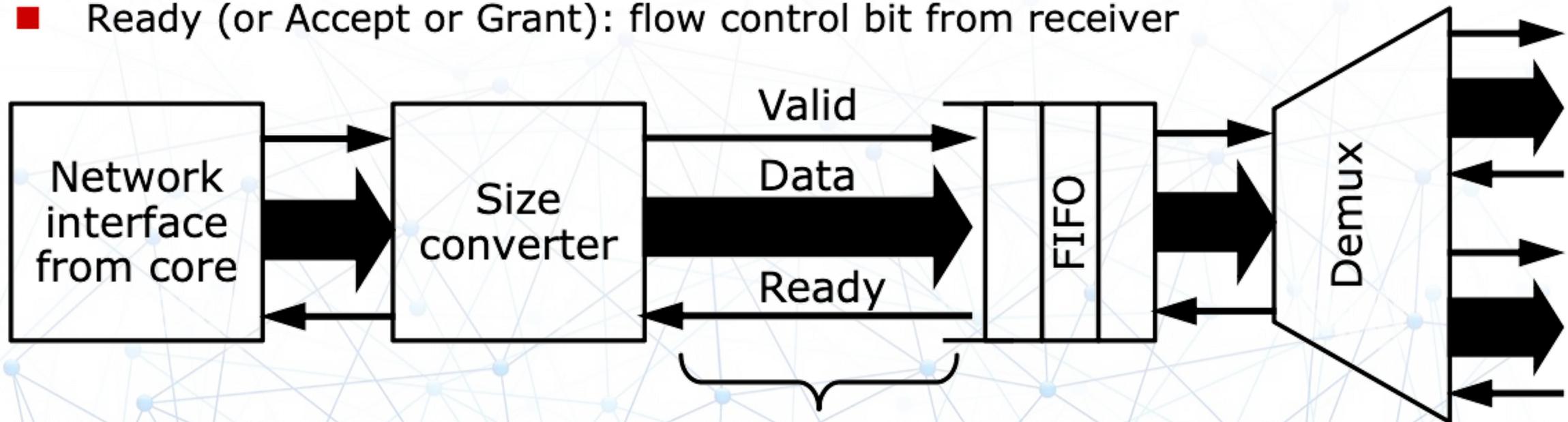
Electrical / AFE

Transaction/Packet Based Interconnect

- are naturally good protocol supporting inter-chiplet interconnect.

- Minimal flow control required

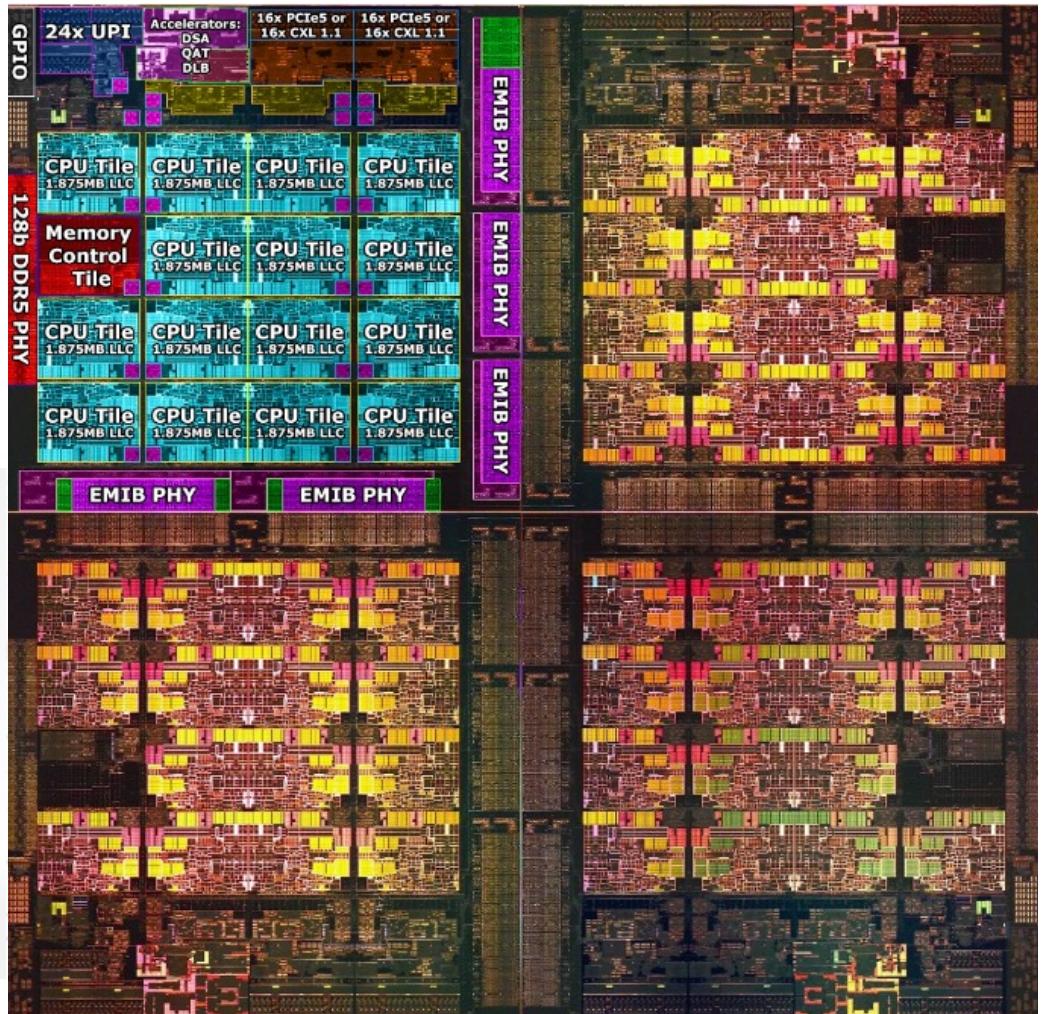
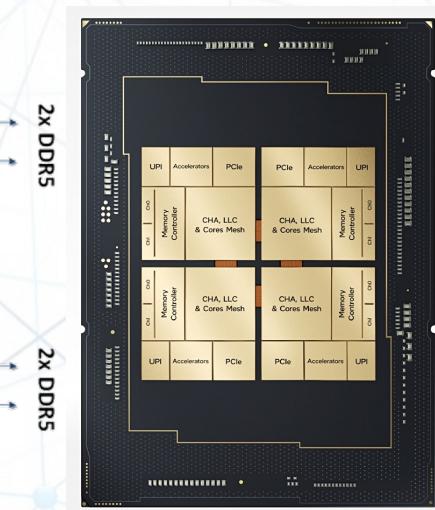
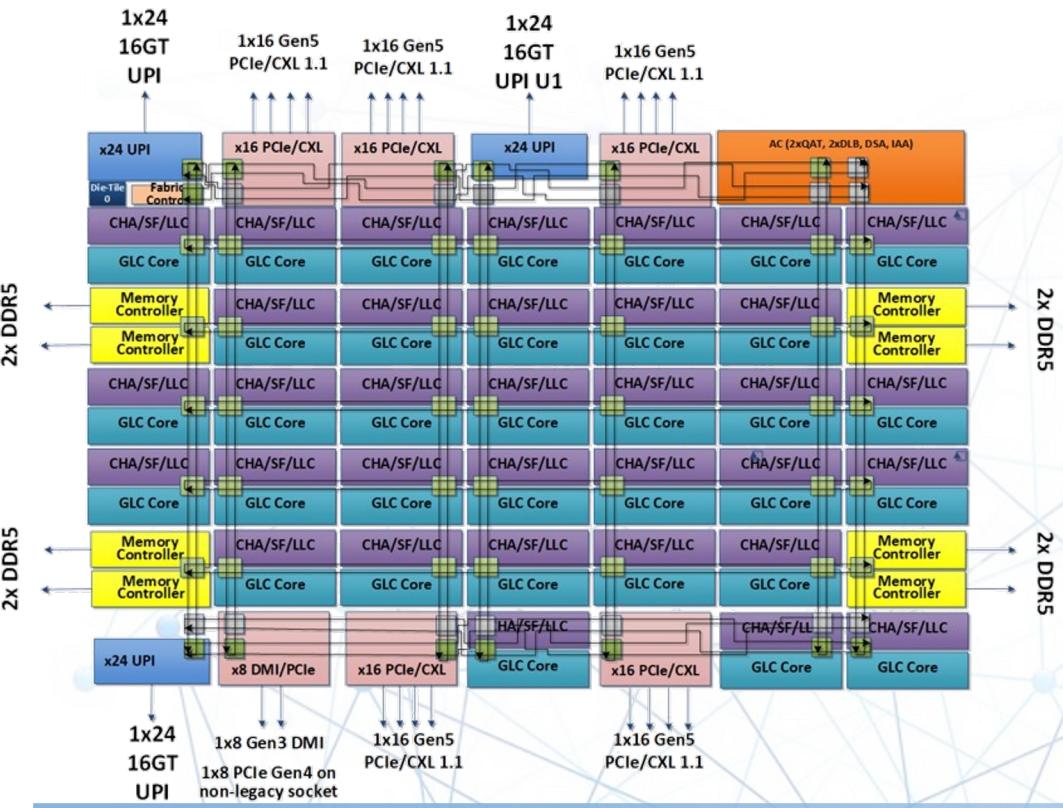
- Data: payload word
- Valid (or Send or Request): flow control bit from sender
- Ready (or Accept or Grant): flow control bit from receiver



Case Study: NoC Partition with Multiple Chiplets



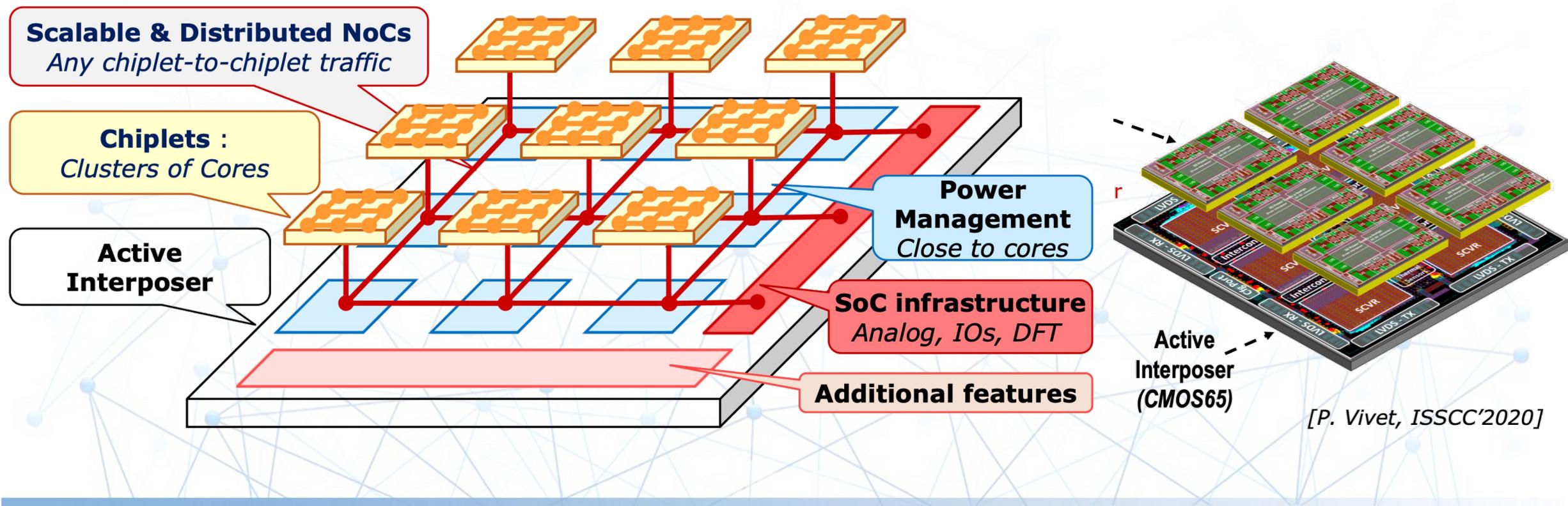
- Intel Sapphire Rapids
- 4xCPU die(15 core each die) + 2D Mesh NoC
- MCM Direct Connection with EMIB



Case Study: NoC Partition with Multiple Chiplets



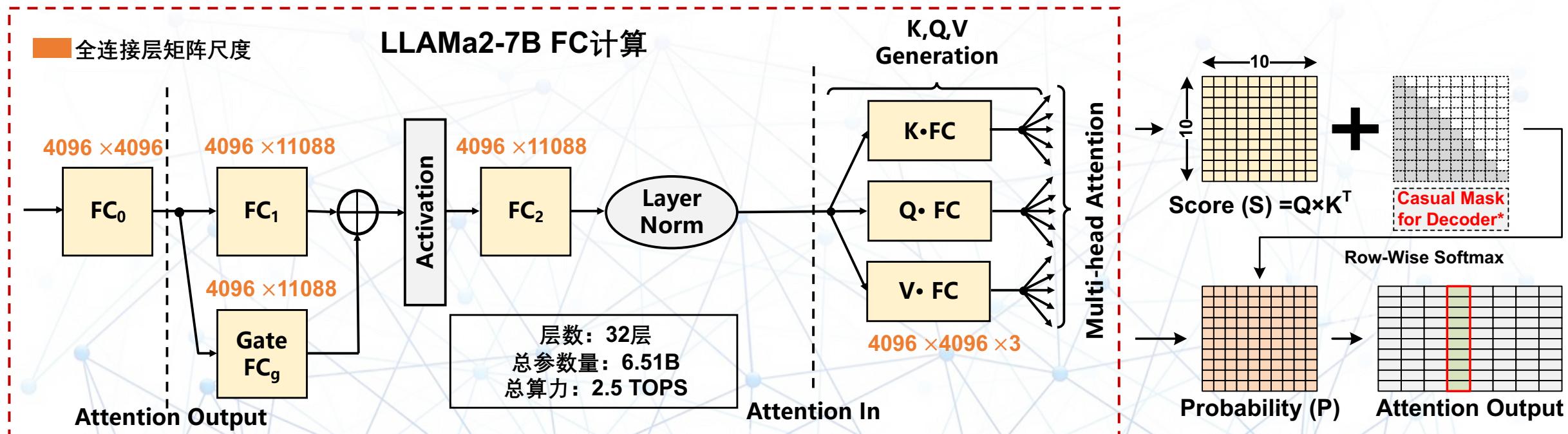
- Cea-Leti: IntACT-Active Interposer with 6 chiplets and 96 cores
- 3-layer of distributed interconnect: On-chiplet 2D Mesh, short distance (Synchronous NoC) and long distance (Asynchronous NoC).



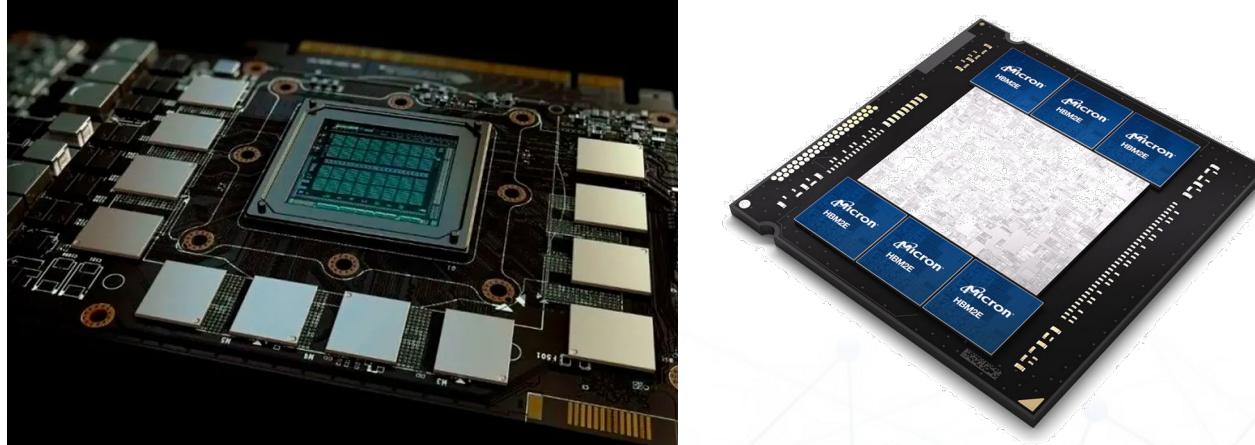
How advanced Packaging affects Interconnect?



- Given a 1466 TFLOPS @ FP8 AI processor with 64MB on-chip weight buffer storage (similar to Nvidia L40S), Please estimate the overall token rate when it is deployed to complete a 7-billion LLAMA-2 model inference (FP8). Assumed the overall model is buffered in (a) a 48GB GDDR6 external DRAM, (b) 80GB HBM2e .



How advanced Packaging affects Interconnect?



	GDDR 6 – 48GB	HBM 2E – 80GB
Per Pin Data Rate	16 Gbps	3.2Gbps
Data Pin Count Per Bank	32	1024
Per bank bandwidth	64 GBps	409.6 GBps
Bank Number	12	6
Overall Memory Bandwidth	864 GBps	~2040 GBps (with little loss)