Network-on-Chip (NoC) designs

Digital System Integration and Programming

Stefan Weiglhofer

October 27, 2020

Content

- Introduction
- Moving towards Network on Chip
- Design and Architecture
- Topology
- Bus versus Network
- Outlook

Introduction

Chip Development

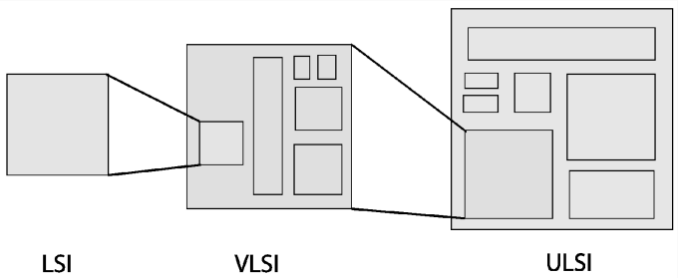


Figure 1: Paradigm shift in system scope [1]

- System scope shifts in steps
- 4 aspects: computation, memory, I/O, and communication
- Processing power and data intensity has emerged
- Number of cores and components integrated into a SoC increases

Chip Communication

- Amount of on chip communication is rising with the amount of components integrated in a chip
- Communication efficiency has become a key factor in determining system performance and cost
- Communication structures:
  - Point-to-Point (P2P)
  - Bus
  - Network

Point-to-Point

- P2P architectures can provide highest performance
  - Dedicated channel between IP pairs
  - Lack of scalability
- Example MPEG-2 Encoder:

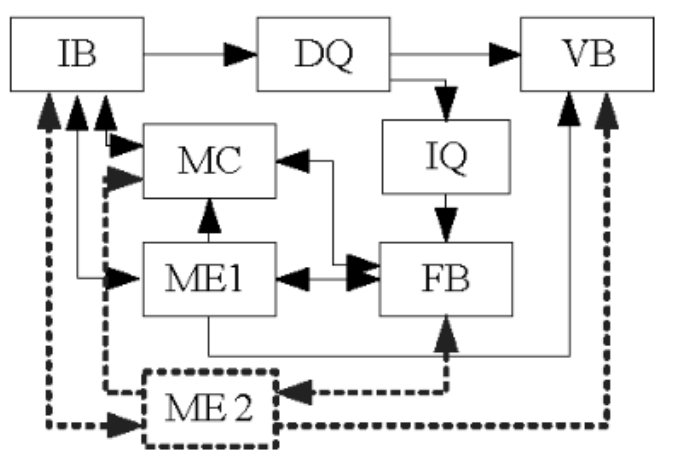
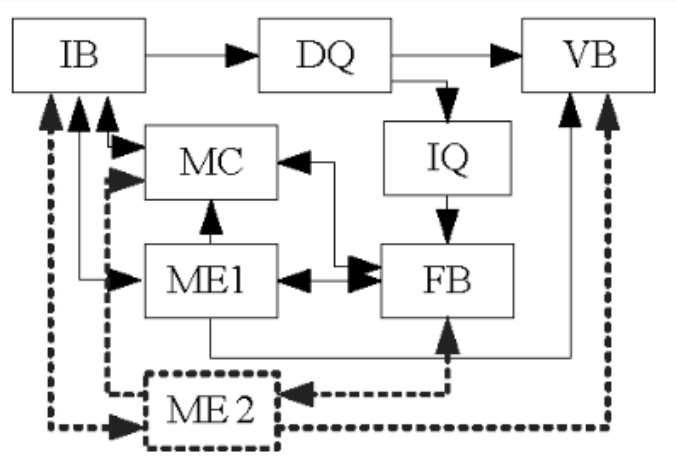


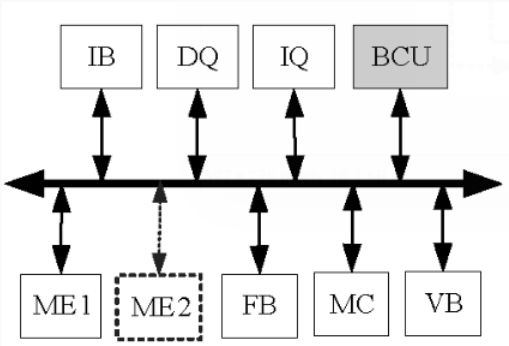
Figure 2: MPEG-2 encoding with P2P connections [2]

On-chip bus i

- IP components communicate through one or more buses interconnected by bridges.



(a)



(b)

Figure 3: MPEG-2 encoding with **(a)** P2P, **(a)** bus connections [2]

On-chip bus ii

- Topologies:
  - Shared global bus
  - Hierachical bus (bridges form hierarchy)
- Communication Concepts:
  - Static-priority (centralized arbiter grants access)
  - Time Division Multiple Access (fixed timing wheel)
  - Token passing (Allocation token circulates in a ring)
  - Code Division Multiple Access (highest prioritised signal wins)
- IP designs integrate bus interface
- Rich component libraries by various companies
  - AMBA (incl. AXI, Arm), CoreConnect (IBM), WishBone, ...

AXI - Advanced eXtensible Interface Bus i

- AMBA - Advanced Microcontroller Bus Architecture
- AXI is P2P based
- Fully specified protcol, supported by xilinx tools
- Five different channels (bidirectional data transfer)
  - Read/Write Address Channel
  - Read/Write Data Channel
  - Write Response Channel
- Different Protocols (AXI4, AXI4-Lite, AXI4-Stream)

AXI - Advanced eXtensible Interface Bus ii

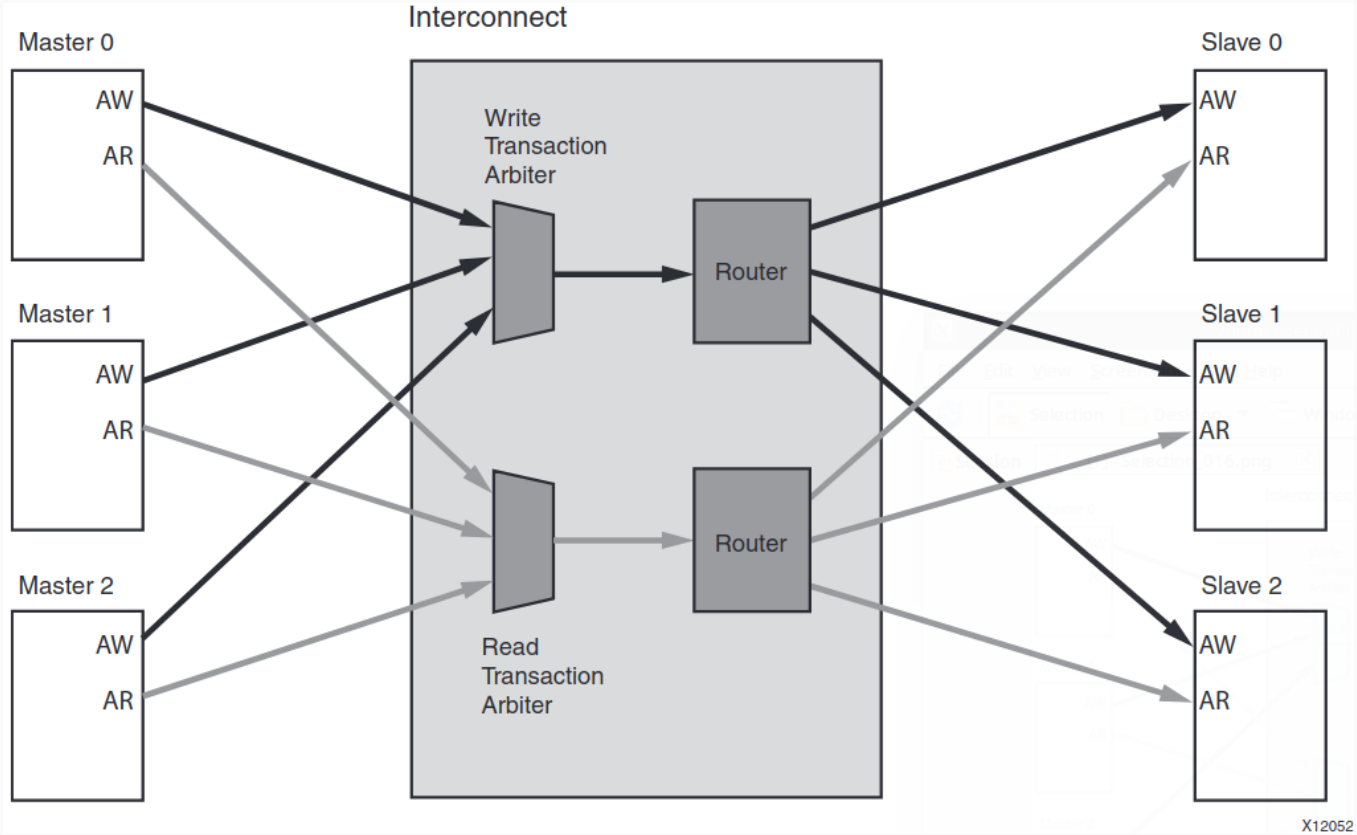


Figure 4: AXI N-to-M Interconnect [7]

Moving towards Network on Chip

Problems need Solutions i

- Wires have large propagation delays
- Synchronisation paradigm changes to globally asynchronous, locally synchronous (GALS)
- Design productivity changes - higher Abstractions needed
- Number of links needed increases exponentially as the number of cores increases.
- Solution:
  - Borrow models from network design field
  - View a SoC as *micronetwork* of components

Problems need Solutions ii

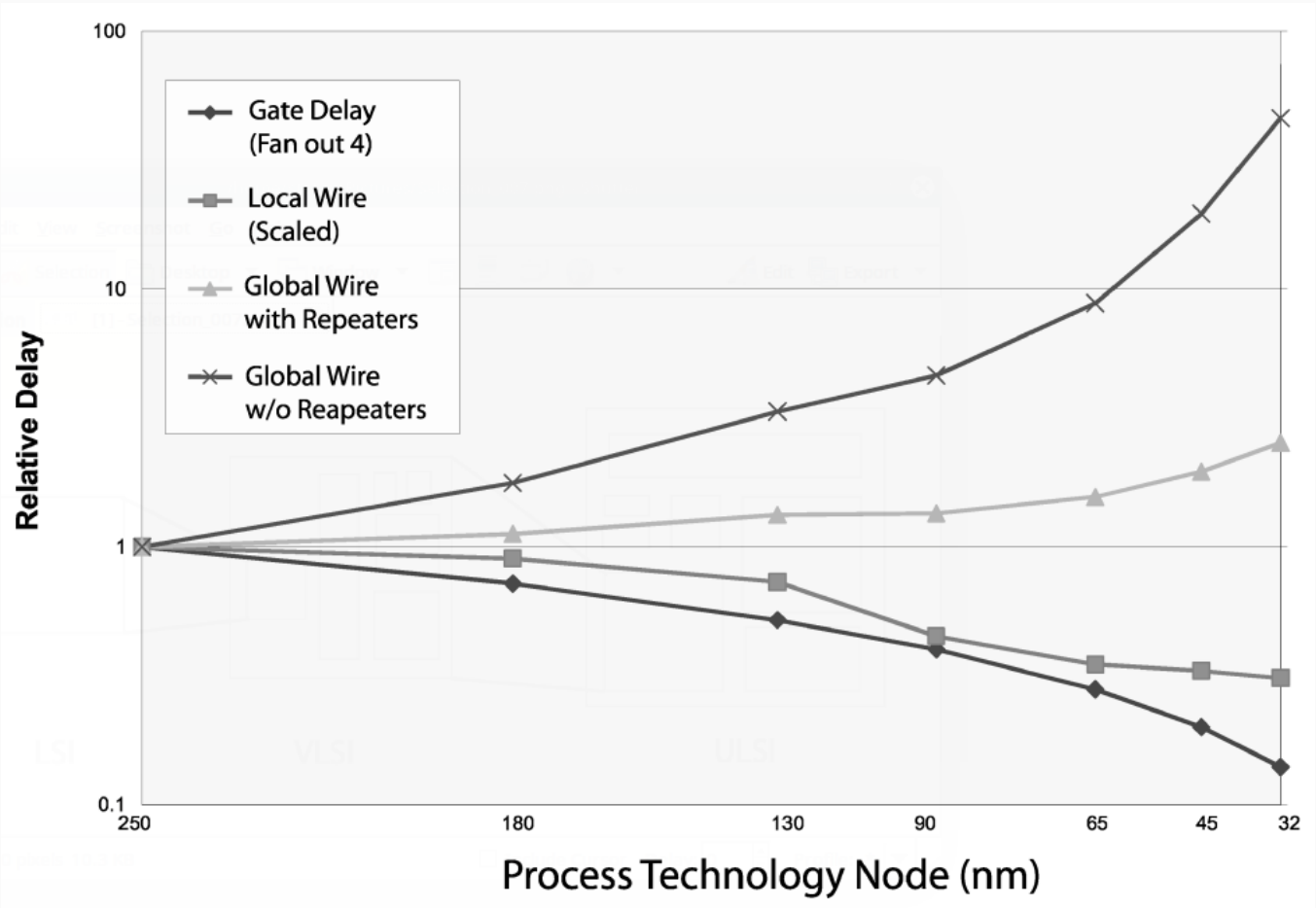
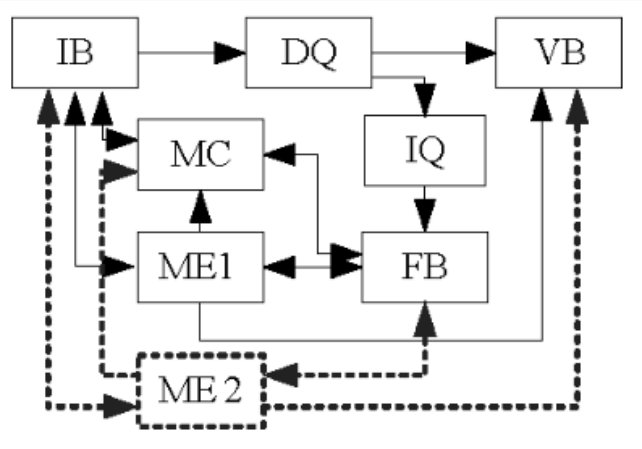
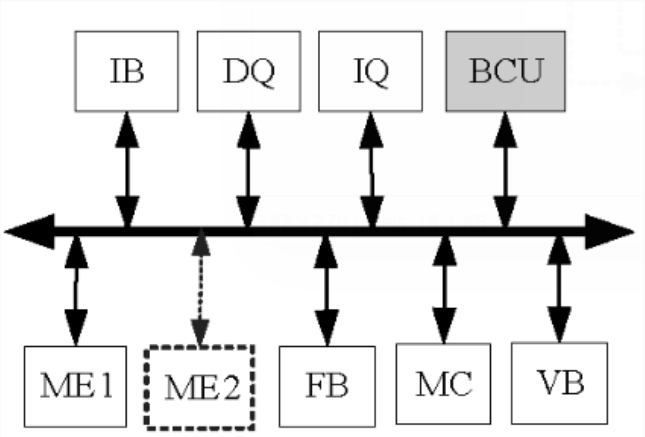


Figure 5: Projected relative delay for wires and logic gates [1]

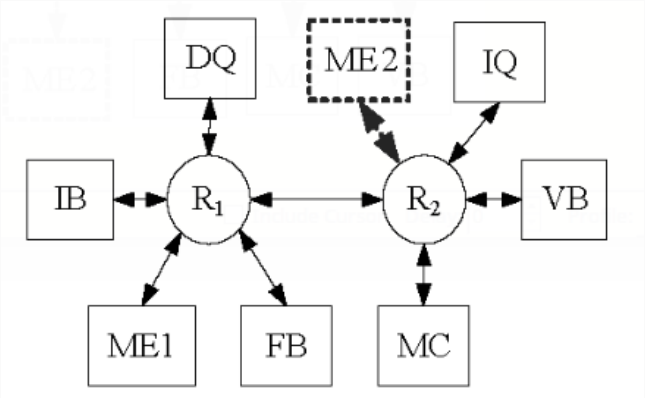
Comparison of Communications



(a)



(b)



(c)

Figure 6: MPEG2 implemented with **(a)** P2P, **(a)** bus, **(a)** NoC connections [2]

Design and Architecture

Design of a NoC i

- Use micronetwork stack paradigm (adaption of the protocol stack)
- Layers:
  - Physical: number and length of wires
  - Data Link: communication protocol
  - Network: packet transmission over network (technology dependent)
  - Transport: breaks message into network layer packets

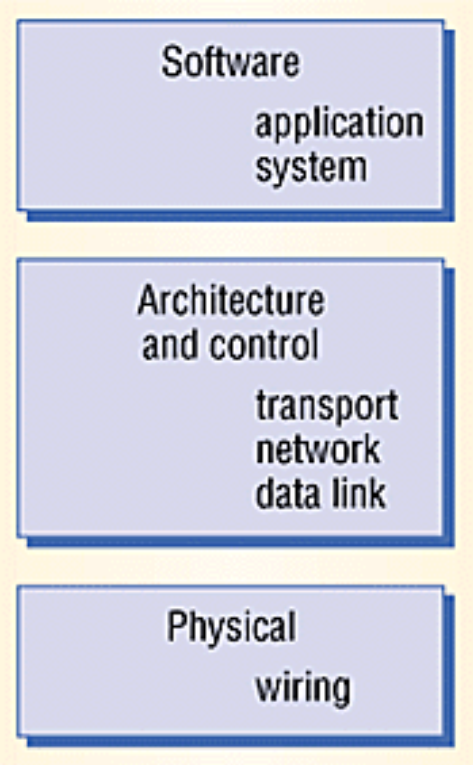


Figure 7: Micronetwork stack [3]

Design of a NoC ii

- Differences to wide-area networks:
  - Local proximity, predictability at design time (static network)
- Need to satisfy tight energy bounds
- Research issues
  - communication infrastructure, communication paradigm, evaluation framework, and application mapping
  - System composition and clustering directly affects communication infrastructure

Fundamental Components

- Network adapters
  - Interface between cores (IP blocks) and network
  - Decouple computation from communication
- Routing nodes
  - Route data according to chosen protocols
- Links
  - Connect the nodes, providing raw bandwidth
  - May consist of one or more logical or physical channels

Components Block Diagram

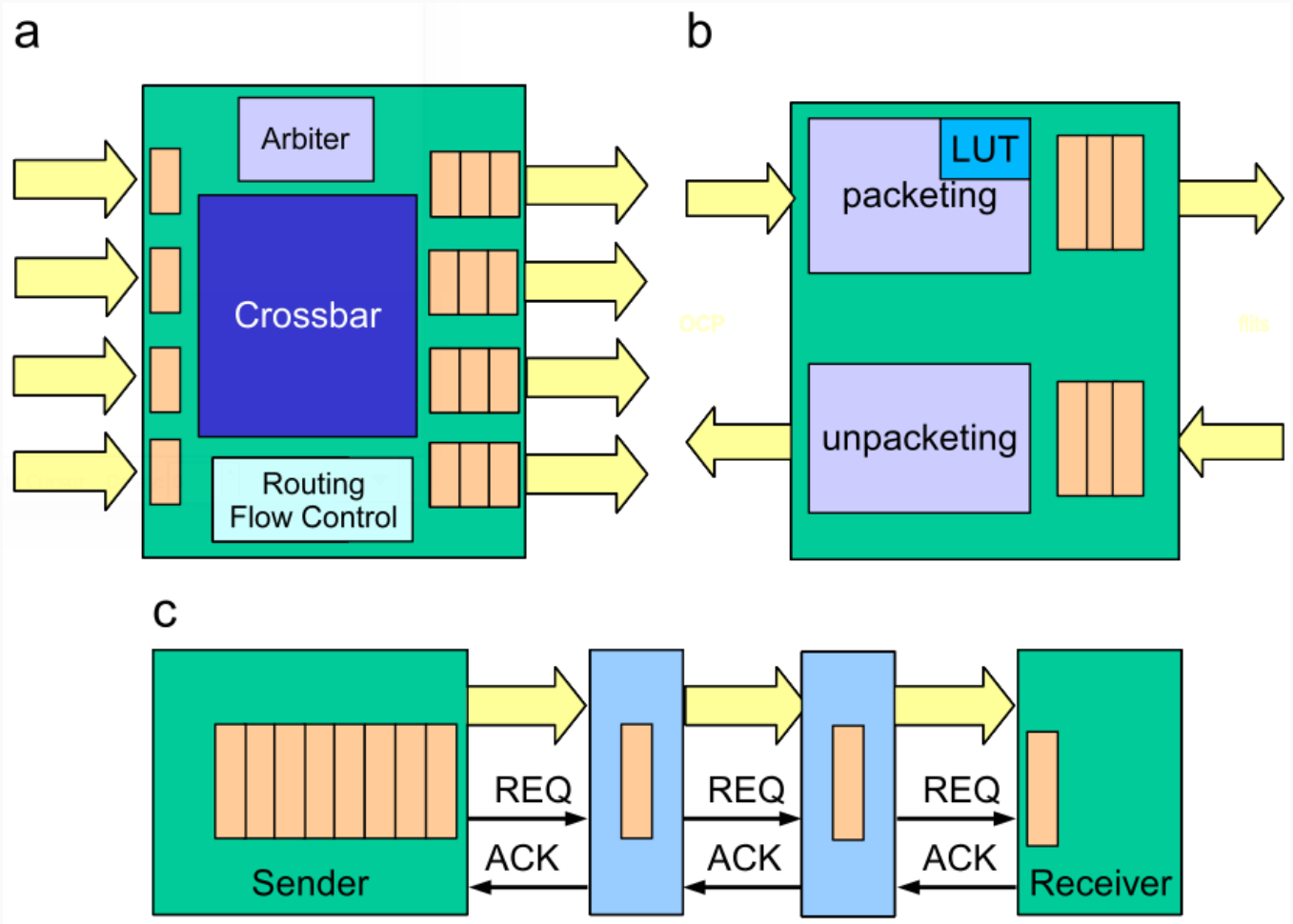


Figure 8: n pipes NoC architectureal blocks: (a) switch, (b) network adapter, (c) pipelined link [5]

Conventional Router

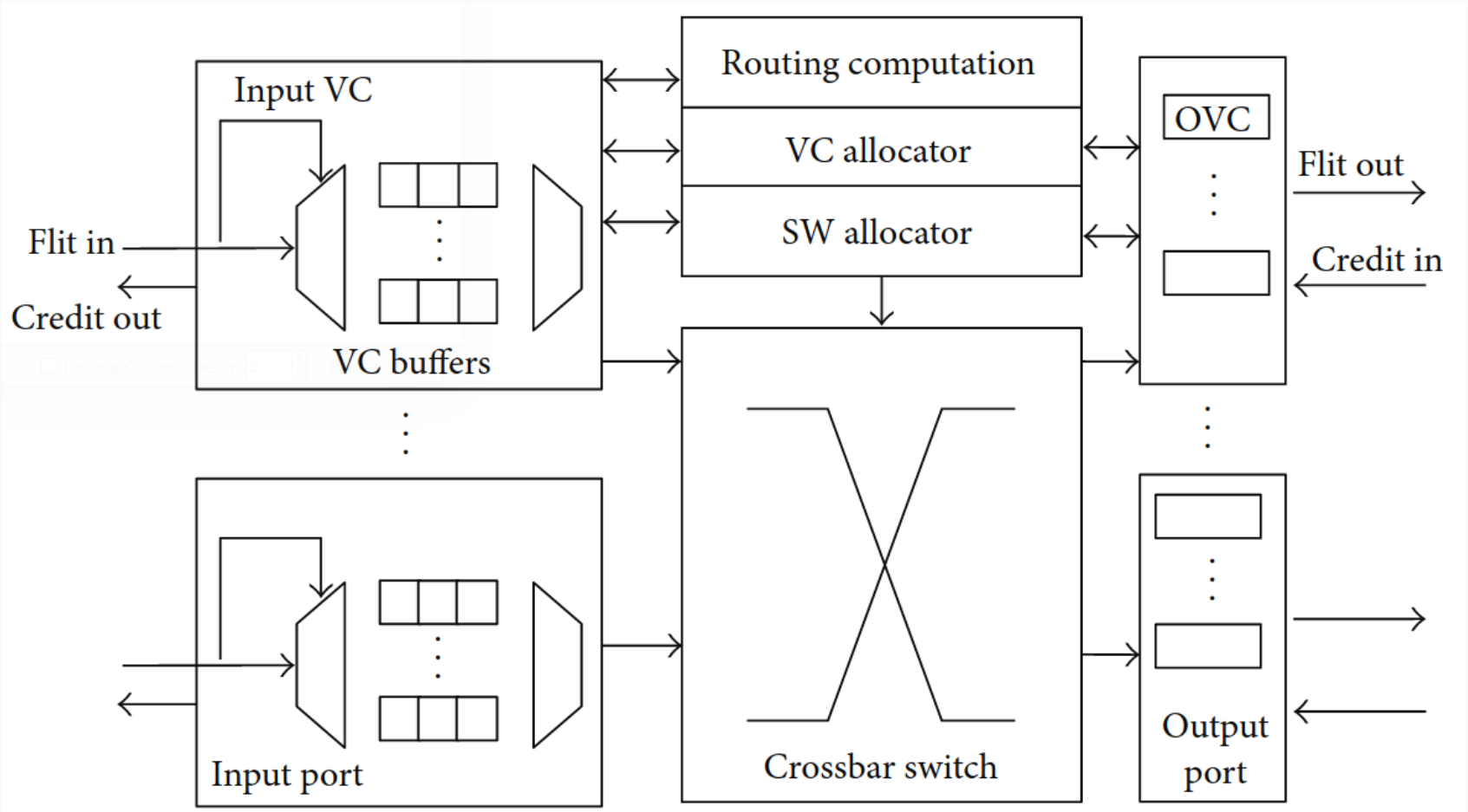


Figure 9: Conventional Router [4]

Topology

Main Topologies

- Topology defines the linking part of the network
- Most common topologies:
  - Mesh (CLICHÉ)
  - Torus
  - Folded Torus
  - Binary Tree
  - Butterfly Fat Tree

Mesh

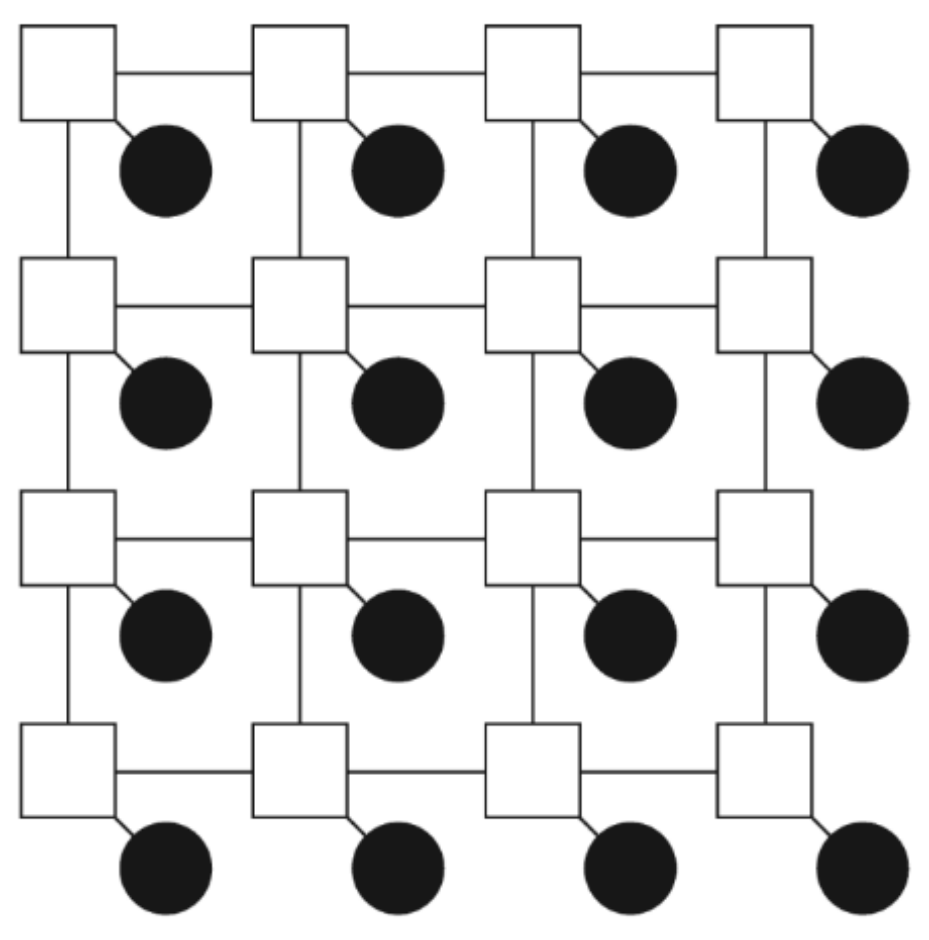


Figure 10: Mesh Topology [3]

Torus

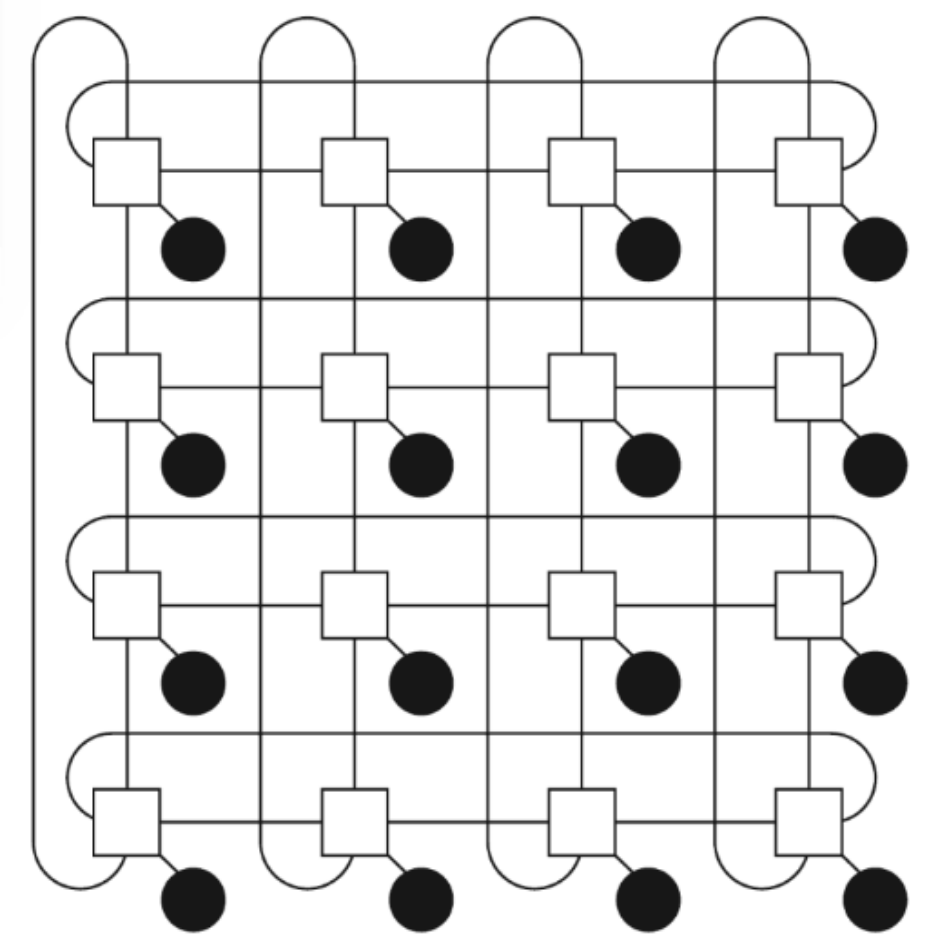


Figure 11: Torus Topology [3]

Folded Torus

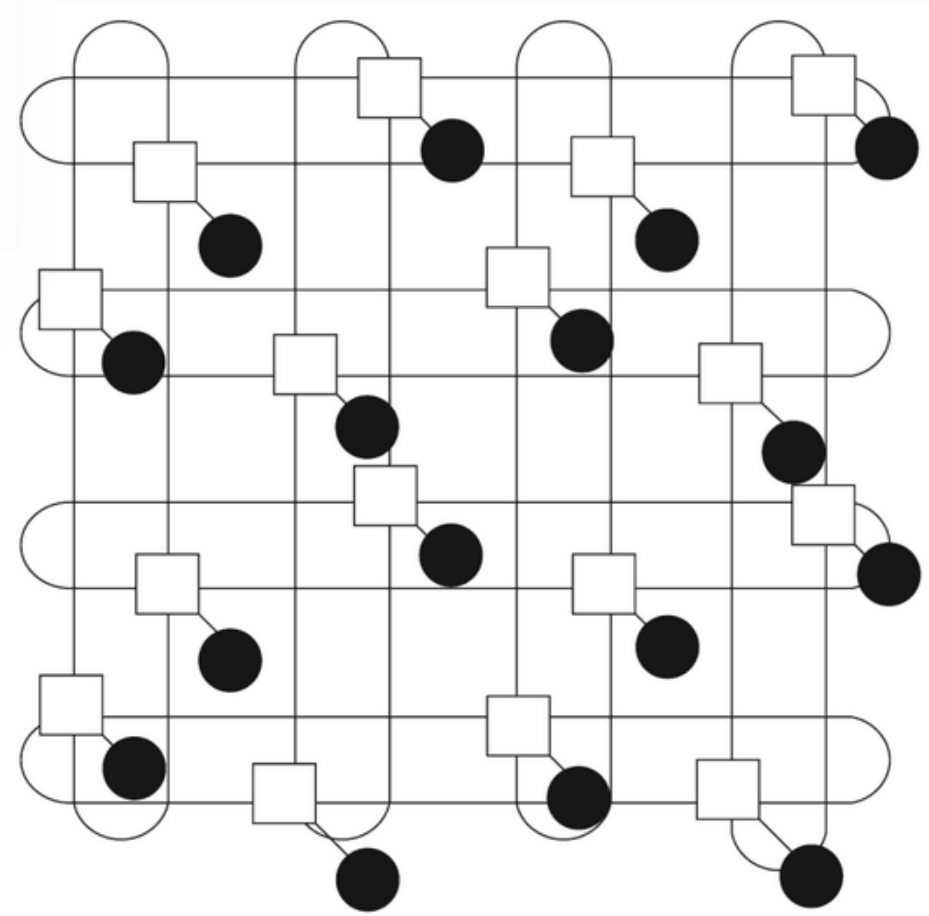


Figure 12: Folded Torus Topology [3]

Binary Tree

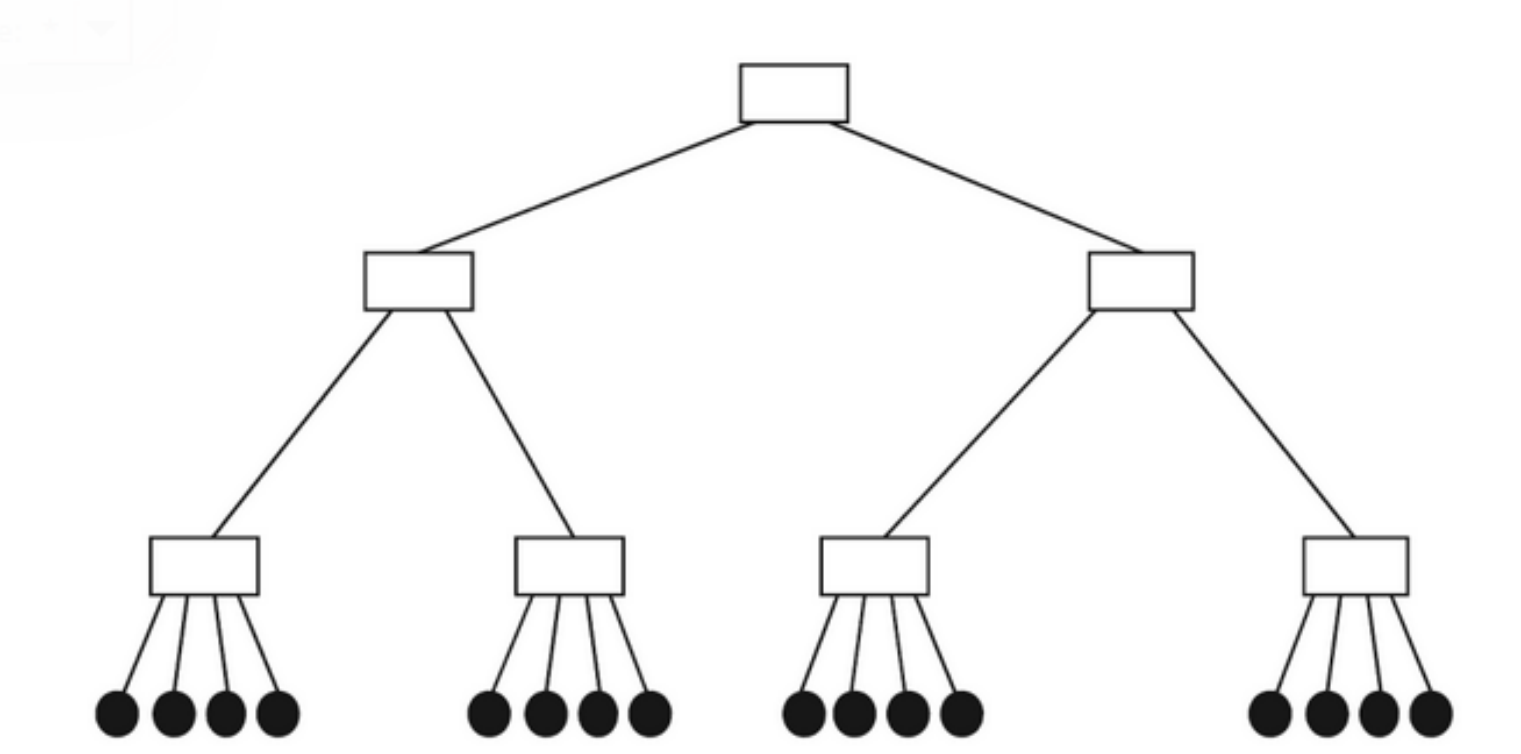


Figure 13: Binary Tree Topology [3]

Butterfly Fat Tree

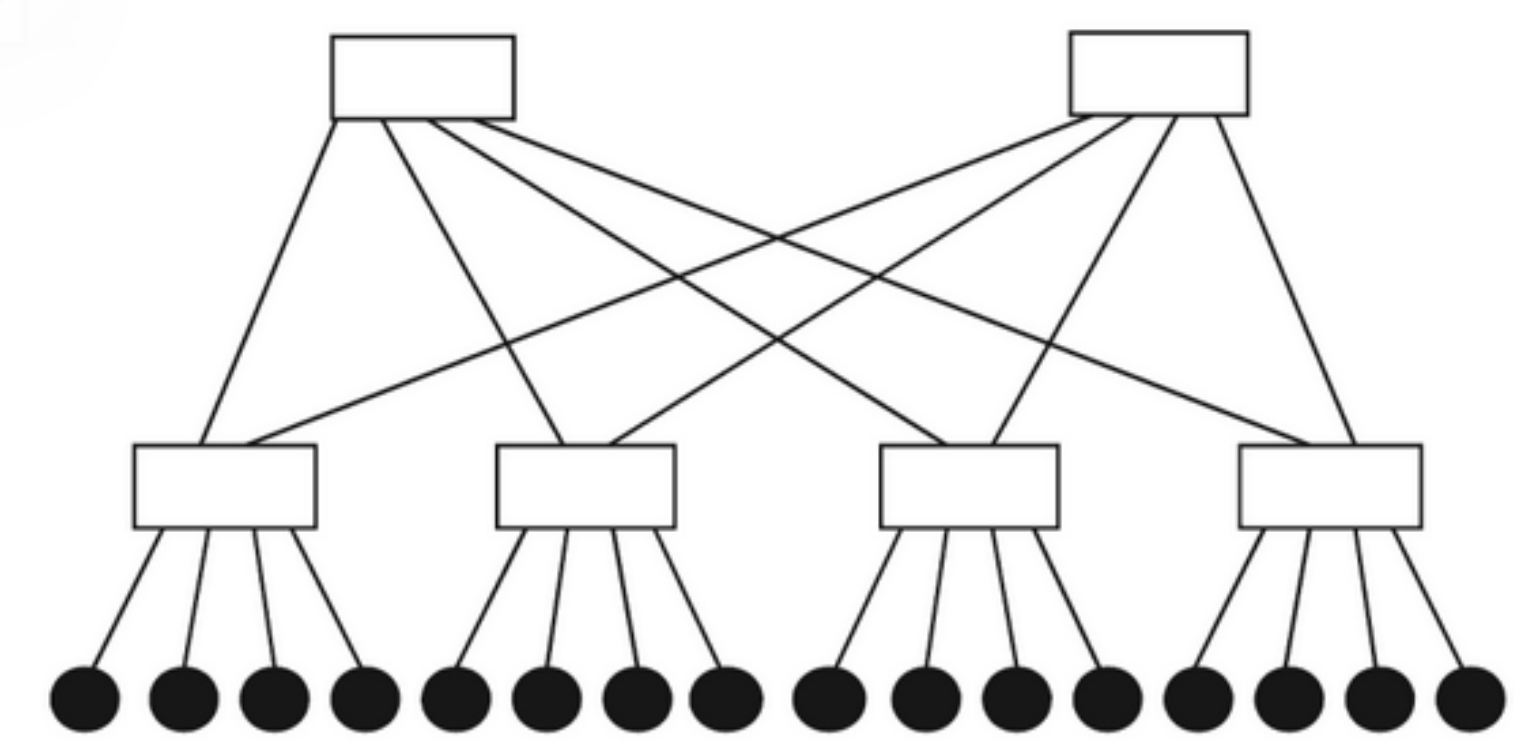


Figure 14: Butterfly Fat Tree Topology [3]

Bus versus Network

Bus Pros / Network Cons

Bus Pro	Network Con
Bus latency = wire-speed (if arbiter granted control)	Network contention may cause latency
Buses directly compatible with most IPs	Bus-oriented IPs need smart wrappers
Simple concept and well understood	More complex concepts

Table 1: Bus Pros versus Network Cons [1]

Bus Cons / Network Pros

Bus Con	Network Pro
Every added unit adds parasitics	Only point-to-point wires
Timing is difficult in a deep submicron process	Wires can be pipelined
Bus arbitration bottleneck	Distributed routing decisions
Bus arbiter instance-specific	Reuse same router
Bandwidth limited & shared by all units	Bandwidth scales with network size

Table 2: Bus Cons versus Network Pros [1]

Outlook

Outlook

- NoCs are a promising structure to overcome limitations of bus-based solutions
- Topologies have reached a level of maturity (various custom topologies to fit target application)
- 3D chip design (stacking of 2D silicons) eases implementation of complex NoCs
- Dynamic freqency/voltage scaling is still a challenge
- NoCs are already integrated in some products, but research is still needed

References i

[1] Bjerregaard, Tobias and Mahadevan, Shankar. *A survey of research and practices of network-on-chip*. ACM Computing Surveys (CSUR), 2006, 38. Jg., Nr. 1, S. 1-es.

[2] Lee, Hyung Gyu, et al. *On-chip communication architecture exploration: A quantitative evaluation of point-to-point, bus, and network-on-chip approaches*. ACM Transactions on Design Automation of Electronic Systems (TODAES), 2008, 12. Jg., Nr. 3, S. 1-20.

[3] Benini, Luca and De Micheli, Giovanni *Networks on chips: A new paradigm for component-based MPSoC design*. In: *Multiprocessor systems-on-chips*. Morgan Kaufmann, 2005. S. 49-80.

References ii

[4] Monemi, Alireza, et al. *Low latency network-on-chip router microarchitecture using request masking technique*. International Journal of Reconfigurable Computing, 2015, 2015. Jg.

[5] Atienza, David, et al. *Network-on-chip design and synthesis outlook* Integration, 2008, 41. Jg., Nr. 3, S. 340-359.

[6] Mitić, Milica and Stojčev, Mile. *An overview of on-chip buses* Facta universitatis-series: Electronics and Energetics, 2006, 19. Jg., Nr. 3, S. 405-428.

References iii

[7] Xilinx, AXI. *Reference Guide, UG761 (v13. 1)*. https://www.xilinx.com/support/documentation/ip_documentation/axi_ref_guide/v13_4/ug761_axi_reference_guide.pdf. Online; accessed 27 October 2020.