A 51mW 1.6GHz On-Chip Network for Low Power Heterogeneous SoC Platform

Kangmin Lee, Se-Joong Lee, Sung-Eun Kim, Hye-Mi Choi, Donghyun Kim, Sunyoung Kim, Min-Wuk Lee, Hoi-Jun Yoo

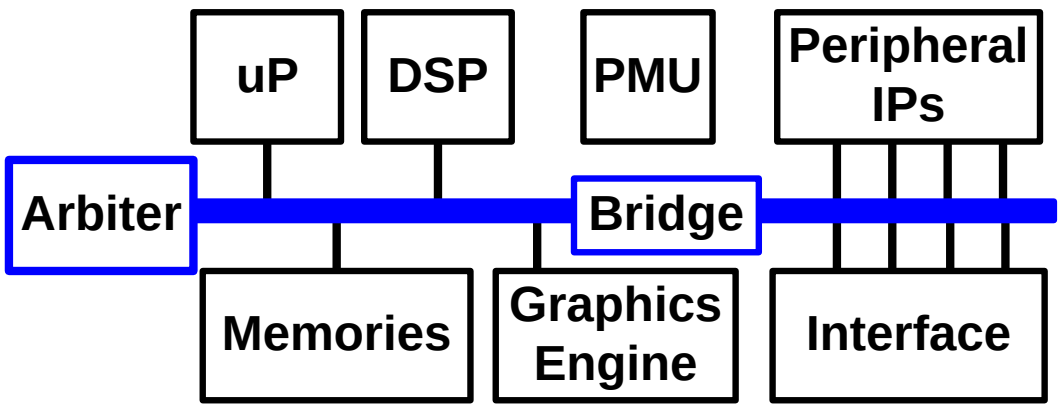
Semiconductor System Lab.
Korea Advanced Institute of Science and Technology

Outline

- Motivation
- On-Chip Network Architecture
  - Overall architecture
  - Features & protocol
- Low-Power Techniques
  - Small-swing global link
  - Crossbar partial activation
  - Serial-link encoding
  - Frequency scaling
- Implementation Results
- Conclusion

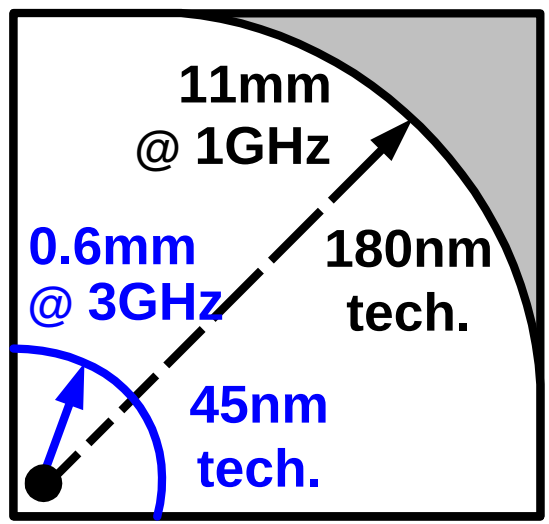
Motivation (1/3)

- **Communications Problems in SoC**
  - Bandwidth bottleneck on a shared-bus
  - Complicated arbiter and global interconnections
  - Difficulties in global clock synchronization



Shared Bus Structure

Shrinking Clock Domain

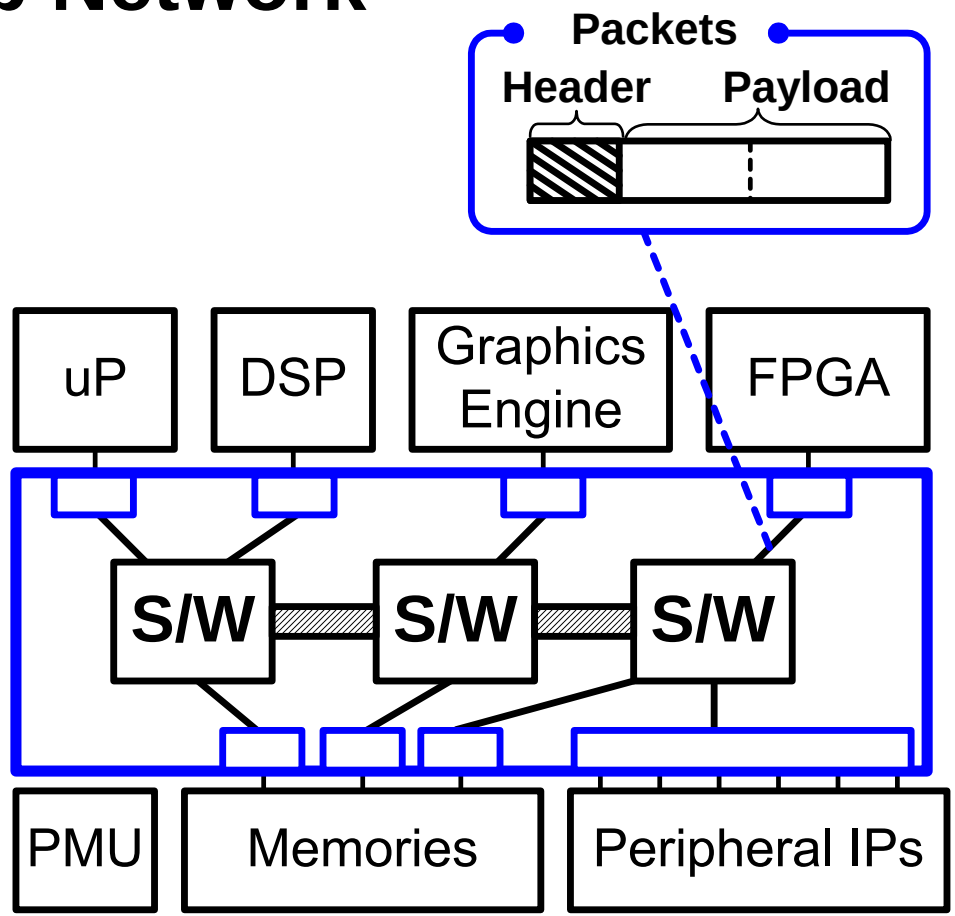


Motivation (2/3)

- **Solution: On-Chip Network**

❖ **Performance**
- Large bandwidth
- QoS control
- **Low-power design**

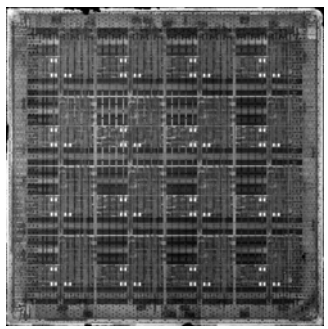
❖ **Easy Design**
- Design reuse
- Scalability
- Synchronization



Motivation (3/3)

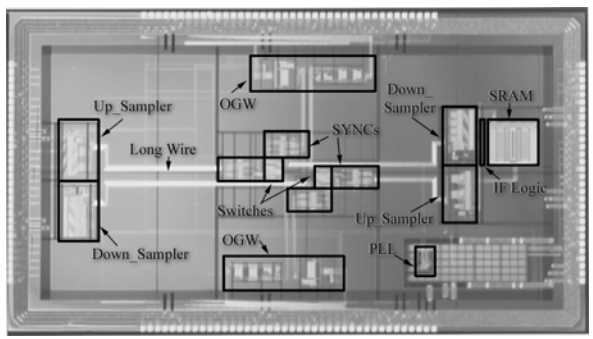
- **Previous On-Chip Networks**
  - Designs not optimized for power, only for performance

MIT-RAW *



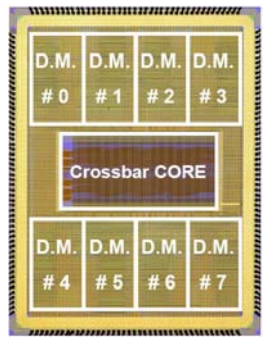
25000mW @ 29GB/s

KAIST-BONE *



264mW @ 6.4GB/s

STM *



N/A @ 0.8GB/s

*[ISSCC 2003]

- **This Work**

→ **Low-Power On-Chip Network Implementation**

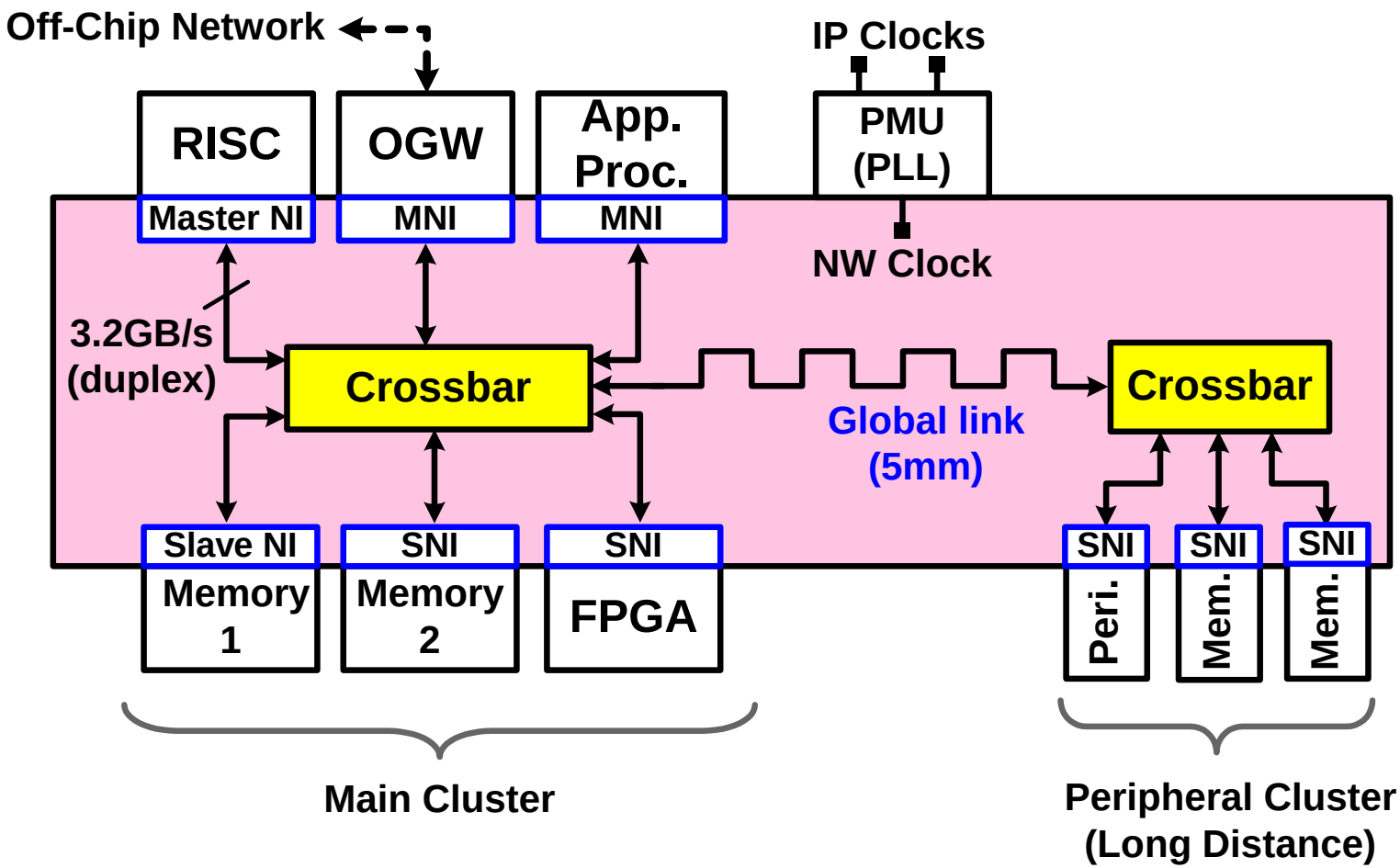
→ **51mW @ 11.2GB/s**

Comparison of Chip Implementation

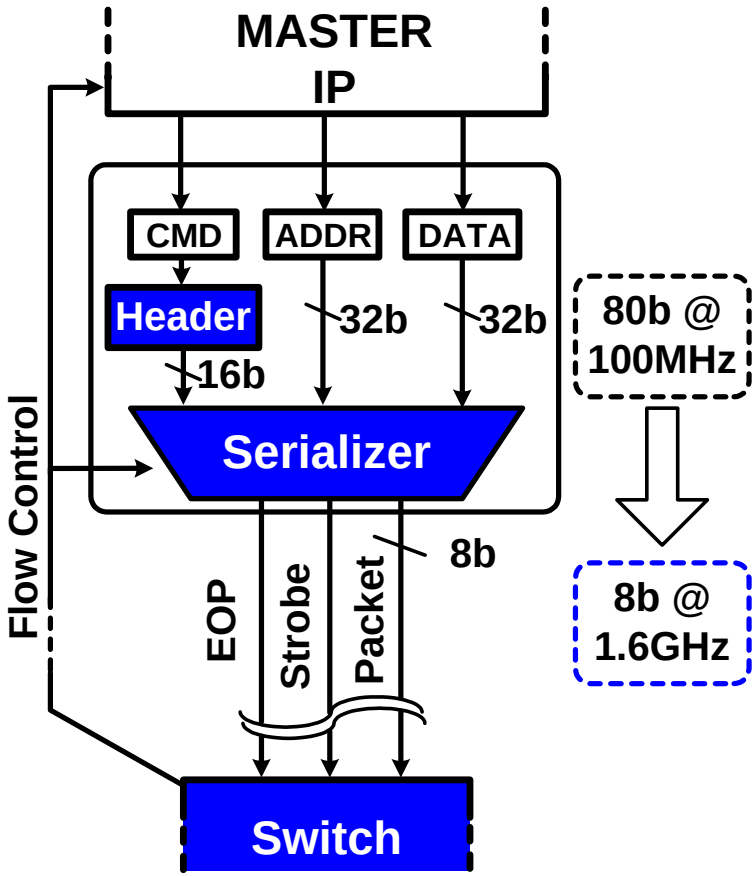
	ISSCC 2003	**This work**
Generation	1st	**2nd**
Process Technology	0.35μm CMOS	**0.18μm CMOS**
Topology	1-level of Star	**2-levels of Star**
Aggregate Bandwidth	6.4GB/s	**11.2GB/s**
Power Consumption	264mW	**51mW (@full BW)**
Power/Bandwidth	41 mW/GB/s	**4.6 mW/GB/s**
Integrated IPs	1kB SRAM Off-chip Gateway	**32bit RISC x 2** **FPGA (64-LE)** **8kB SRAM x 2** **Off-chip Gateway**
Low-Power Techniques	X	**4 major techniques**

Architecture Overview

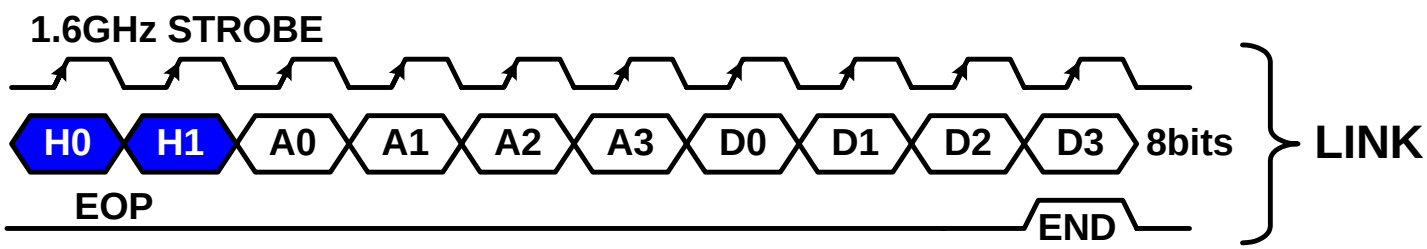
- Prototype for Heterogeneous SoC Platform



Architectural Features (1/2)



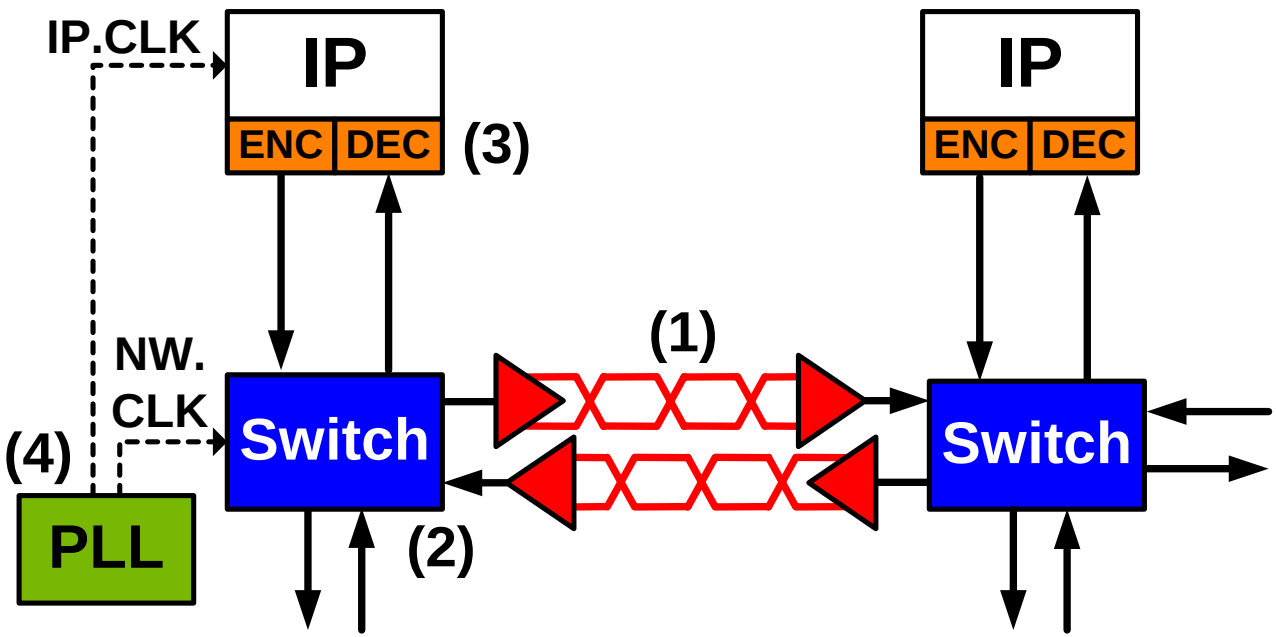
- **x10 Serialization & Speed-Up**
  → Reduces network area
  → Increases network bandwidth
- **Plesiochronous Comm.**
  by source synchronous scheme
- **Flow Control**
  by 1-bit back-pressure



Architectural Features (2/2)

- **Protocol**
  - Burst packet operation (BL=1,2,4,8)
  - Prioritized packet → 2-level of QoS
- **Crossbar Switch**
  - Cut-through switching @ intra-cluster packets
    → Reduces Latency
  - Store & Forward switching @ inter-cluster packets
    → Increases Throughput
  - Deterministic Source Routing
    → Simple Lookup H/W Implementation
  - Round-Robin Scheduler → Fair arbitration

Low-Power Techniques: Preview



(1) Low-swing Differential Signaling on Global Links

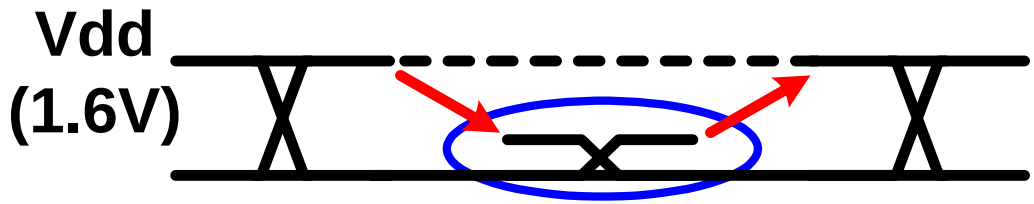
(2) Crossbar Partial Activation Technique

(3) Serial-Link Encoding Technique

(4) Operating Frequency Scaling

Low-Swing Global Link

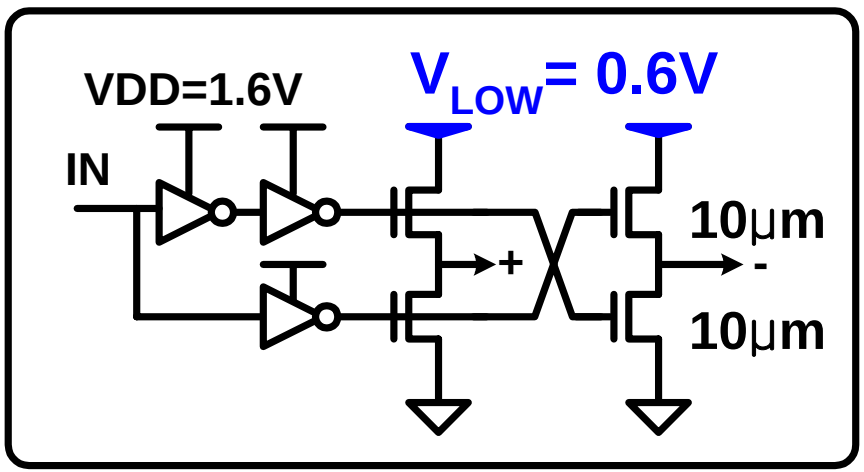
: Low Power Technique (1)



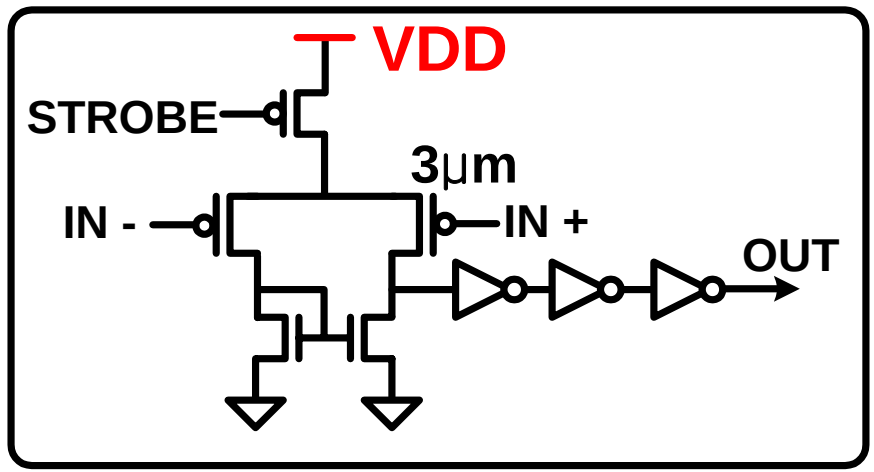
- 5mm long link
- Differential wires

Power reduciton ~ V^2

Transmitter



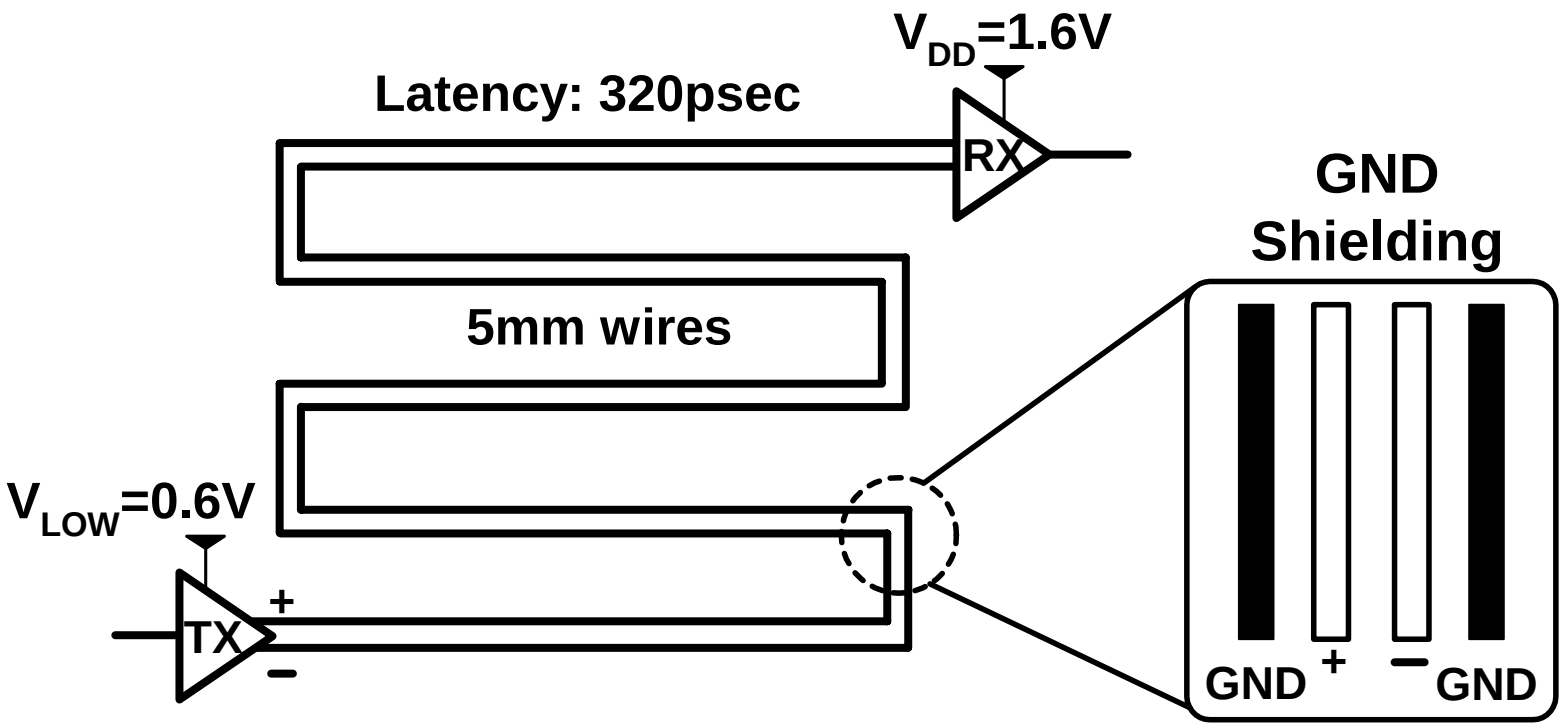
Receiver



- Overdriving with 0.6V supply
- PMOS inputs for low-voltage
- Clocked with Strobe

Low-Swing Global Link

: Low Power Technique (1)

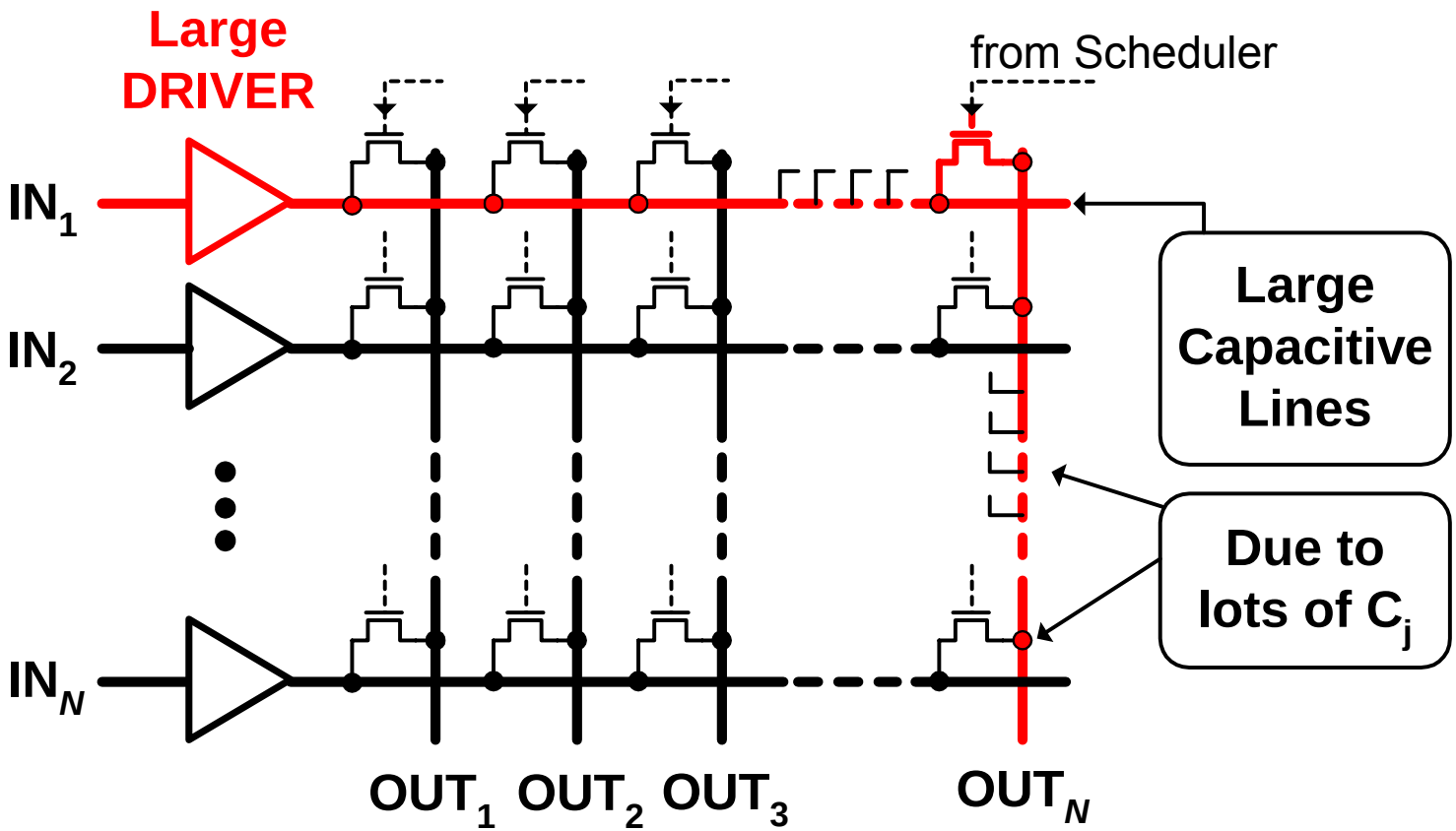


- No area-consuming Repeaters
- Differential signaling increases Noise Immunity
- **Energy ~ 0.35pJ/bit**
  : **1/3 the power of a full-swing repeated link**

Crossbar Partial Activation

: Low Power Technique (2)

- **Conventional Crossbar Fabric**



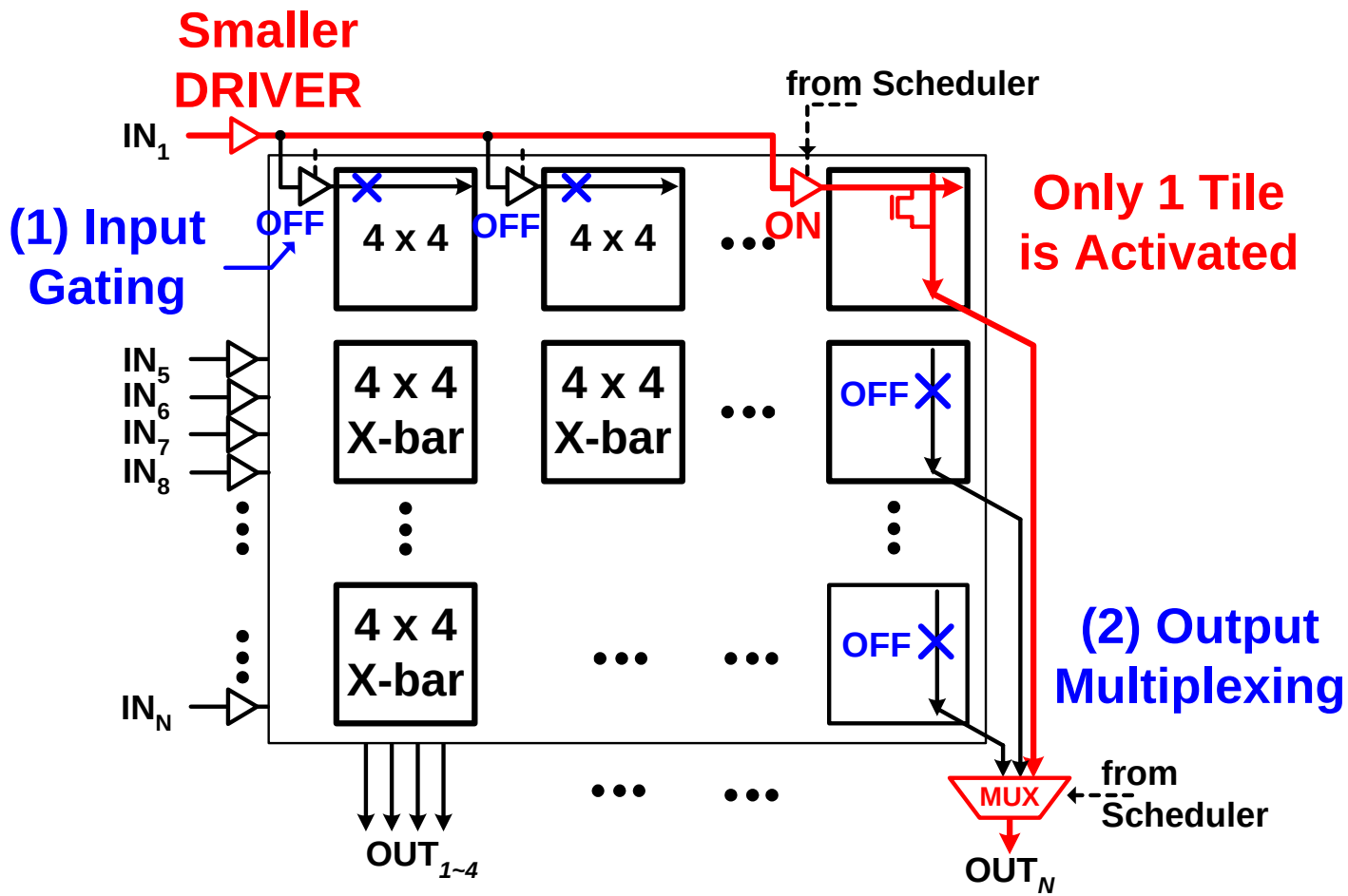
- **Waste of Power** due to Large Load on the lines
  → **Avoid the unnecessary load!**

Crossbar Partial Activation

: Low Power Technique (2)

(2/3)

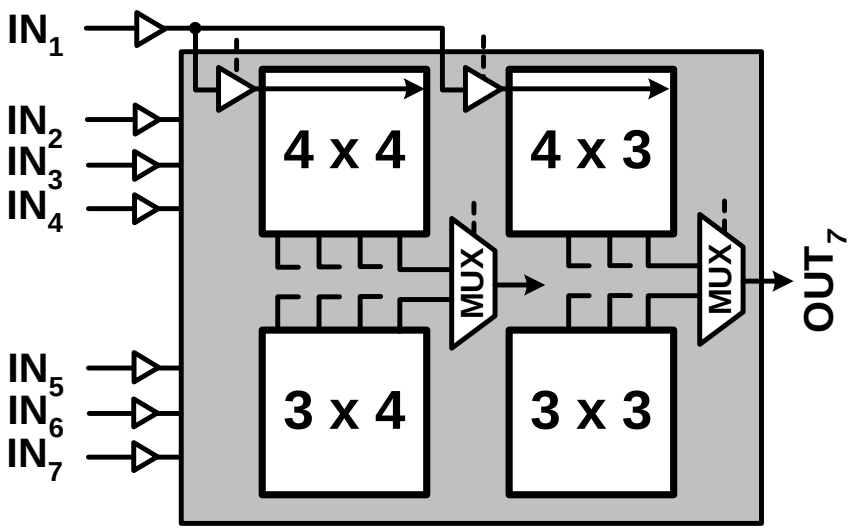
- **Proposed Technique: Split into Smaller Tiles**

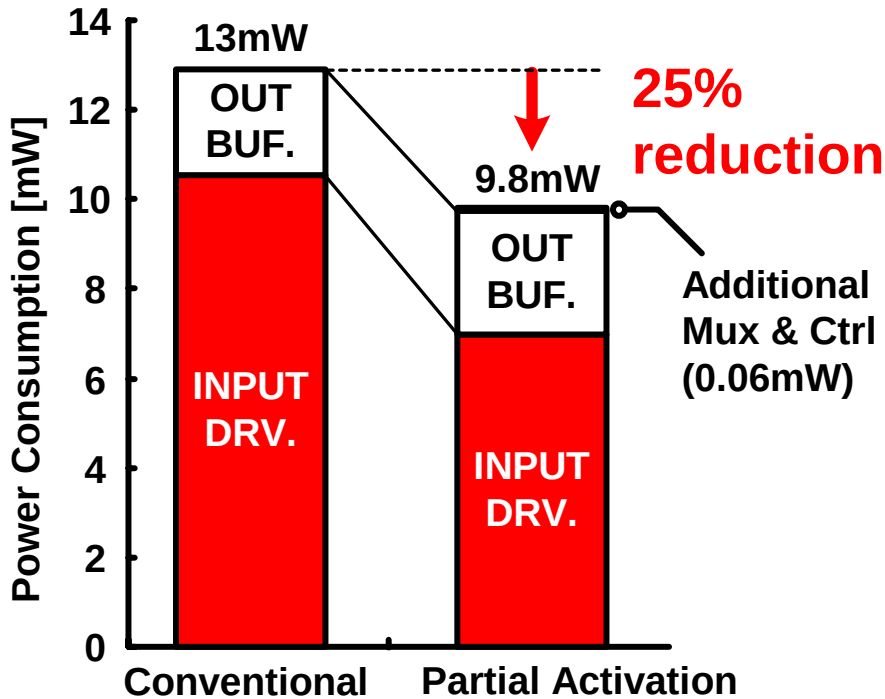


Crossbar Partial Activation

: Low Power Technique (2)

- **Power Saving @ 7x7 Crossbar**





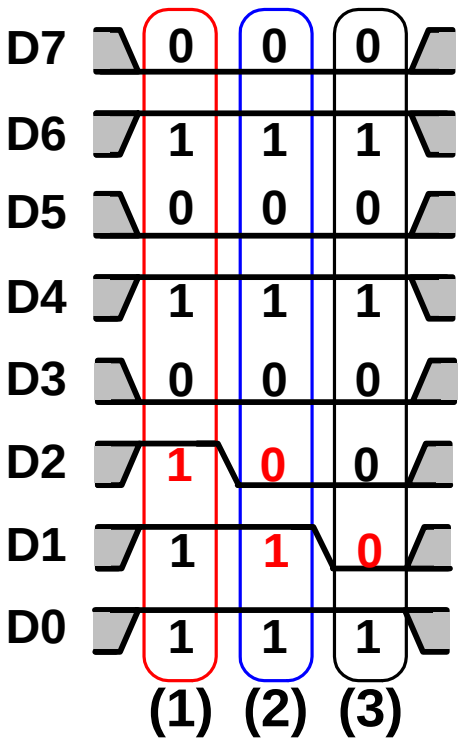
(1) Input Gating
(2) Output Multiplexing ⇨ Wire Load is reduced ⇨ Low-Power Short-Latency

Serial Link Encoding

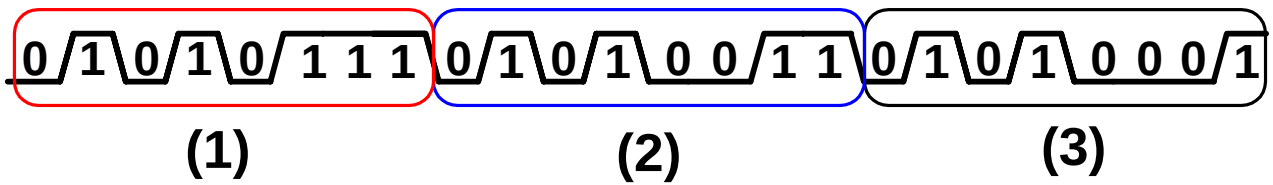
: Low Power Technique (3)

- **Serialization increases transitions on wires!**

Parallel Bus



Serial Link



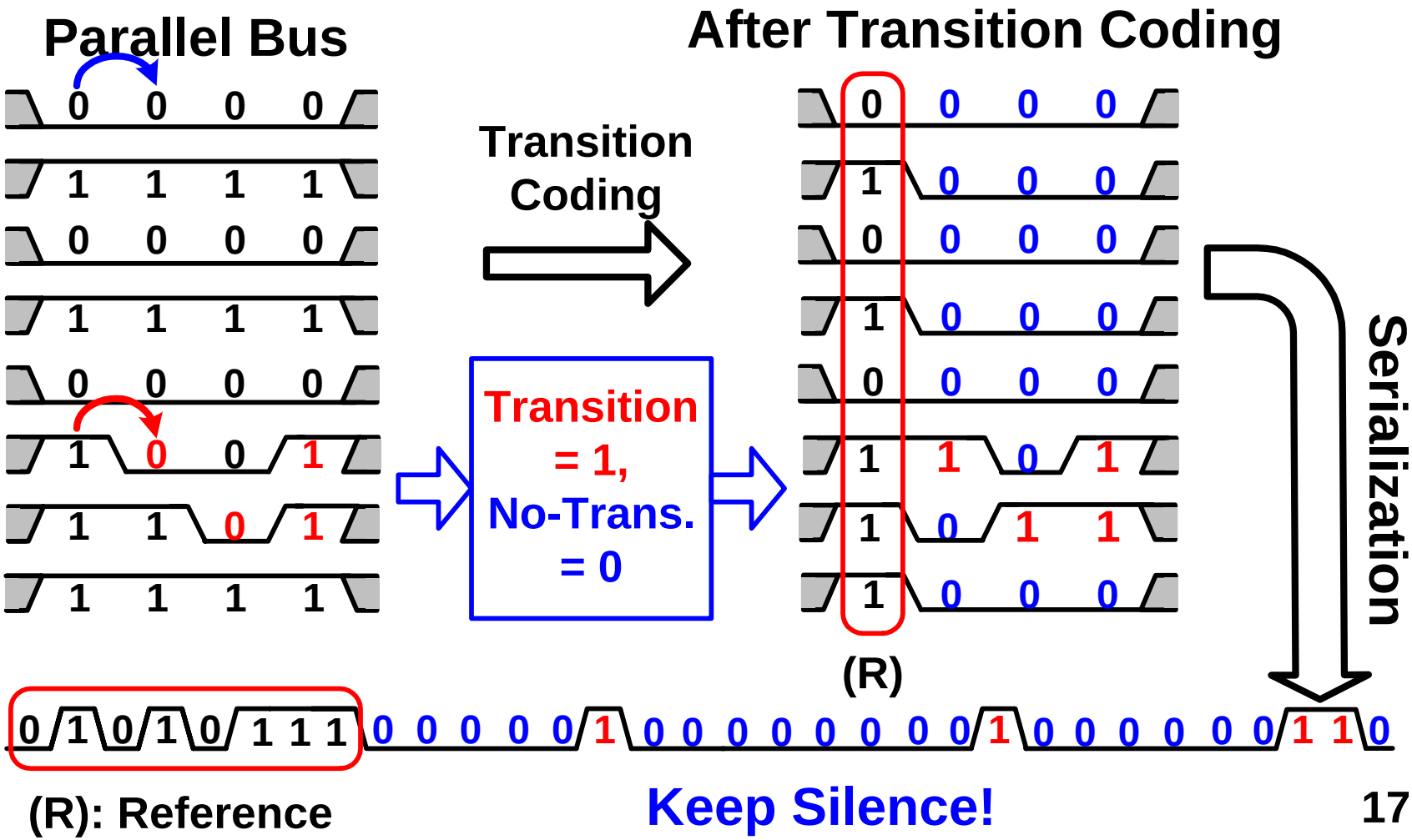
More transitions on a serial wire

- **Goal: Reduce transitions on serial link**
  → **Low-Power Serial Link Encoding**

Serial Link Encoding

: Low Power Technique (3)

- **SILENT: Serial Link Encoding Technique**

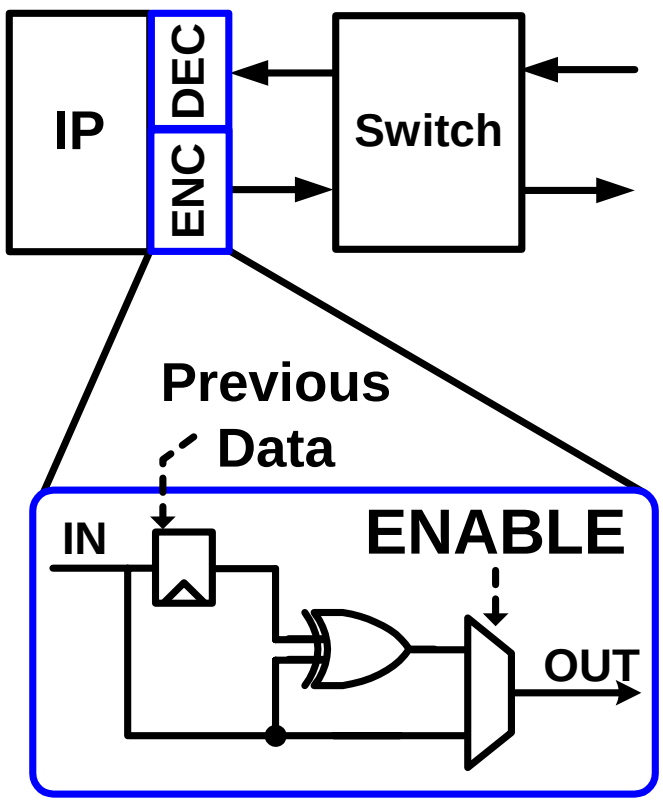


Serial Link Encoding

: Low Power Technique (3)

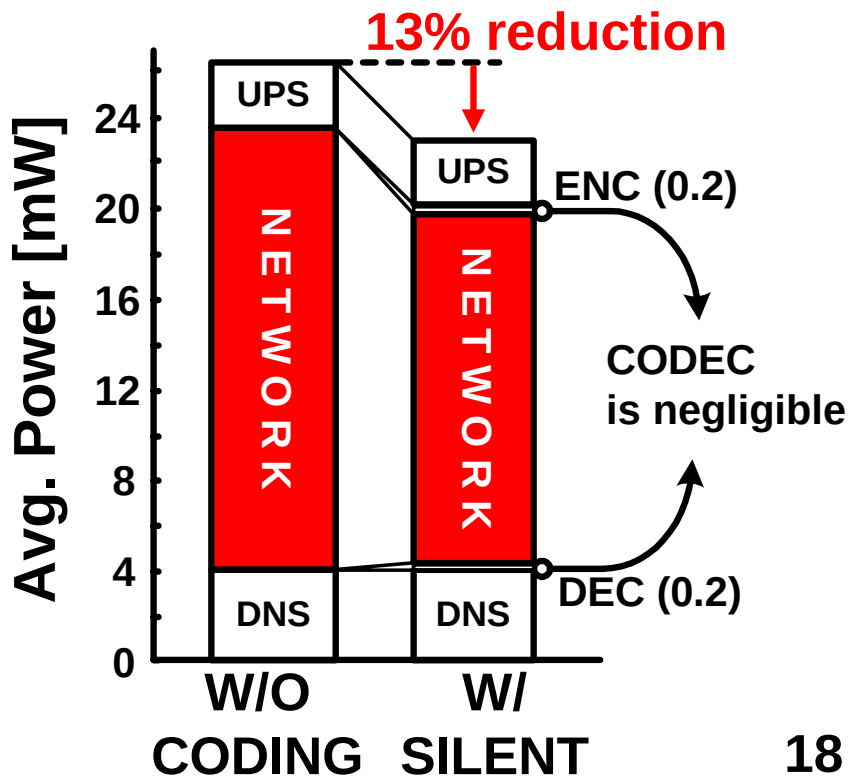
(3/3)

▪ CODEC Circuits



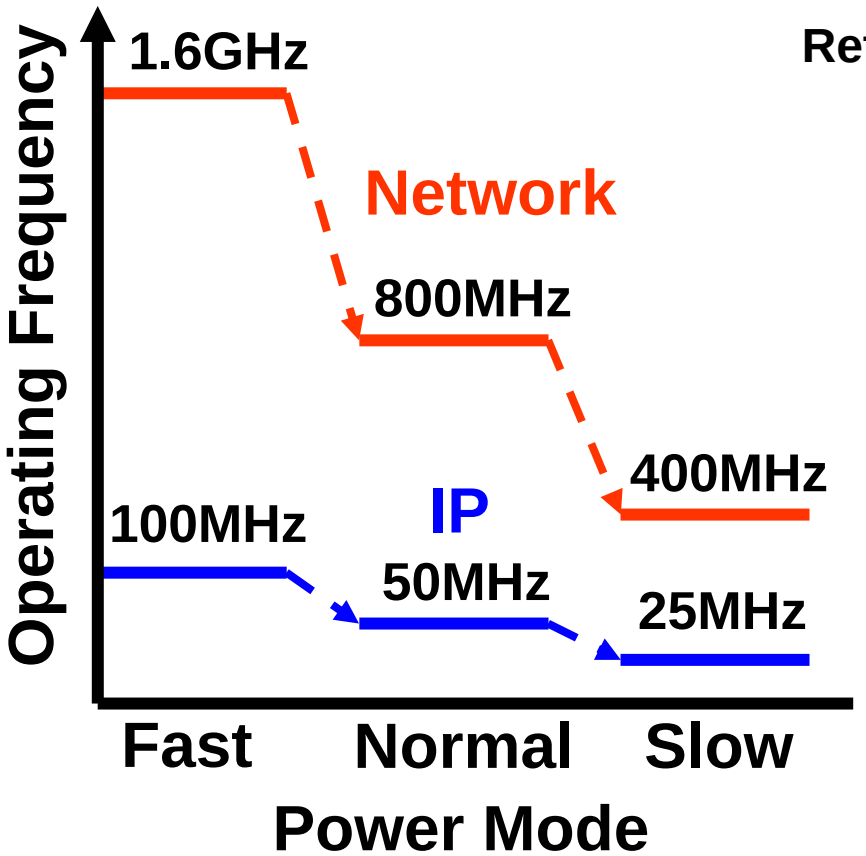
• On & Off by software

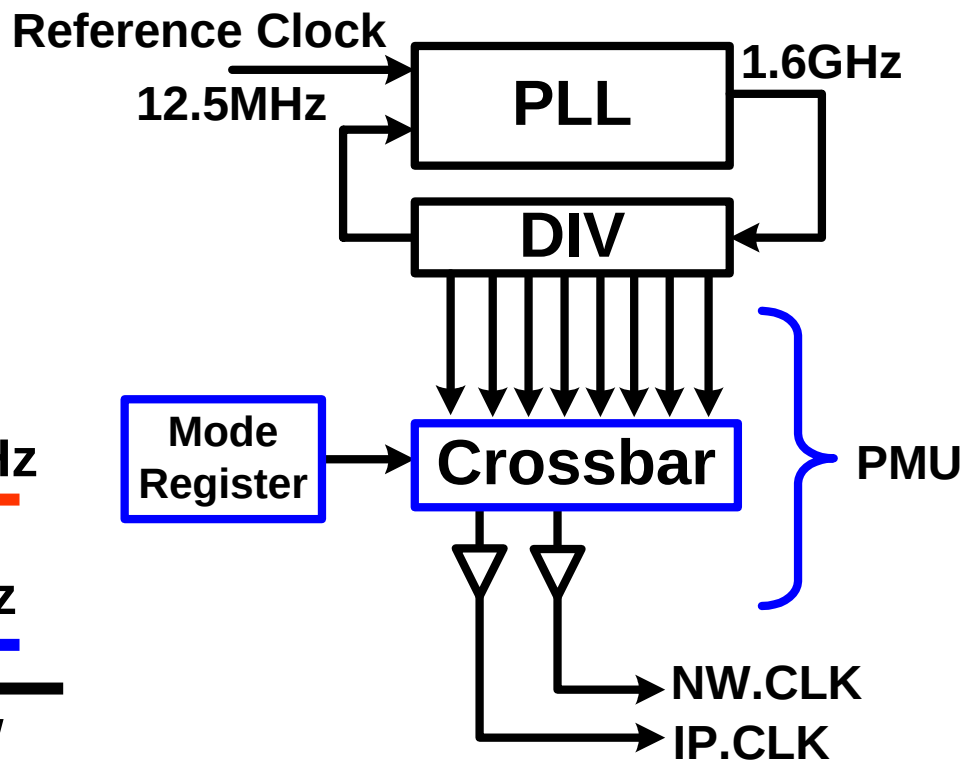
@ 3D Graphics Applications



Frequency Scaling

: Low Power Technique (4)

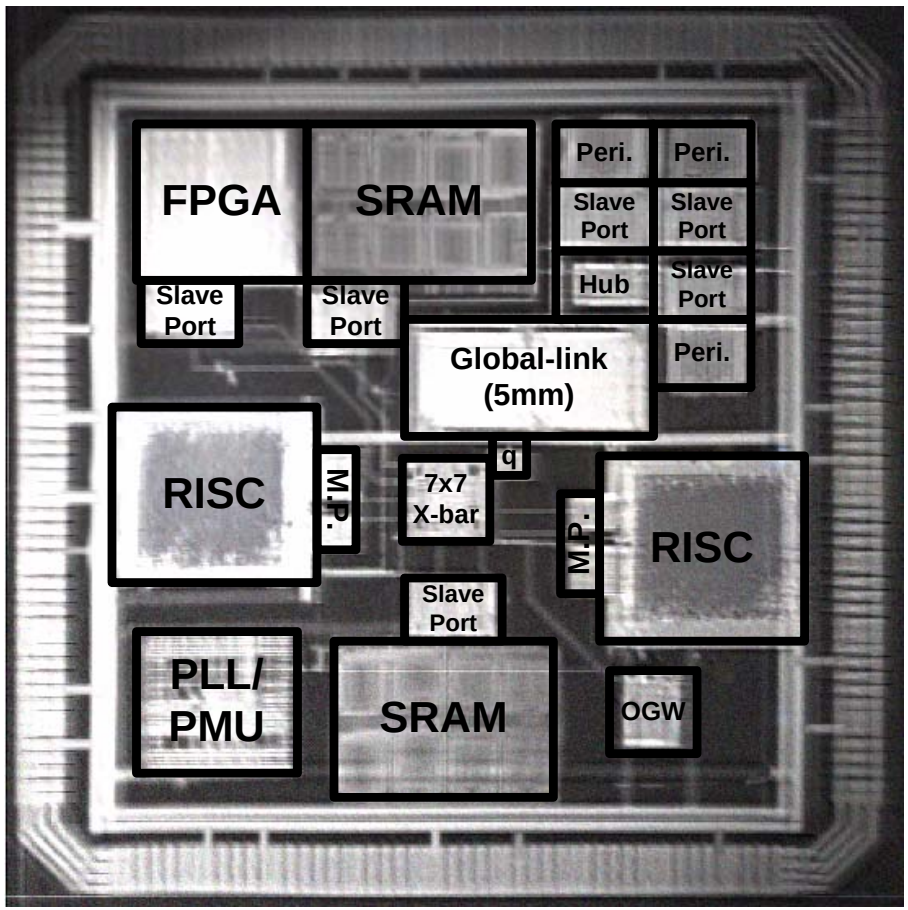




- Operation Frequencies are scalable according to applications and power modes

Implementation Results

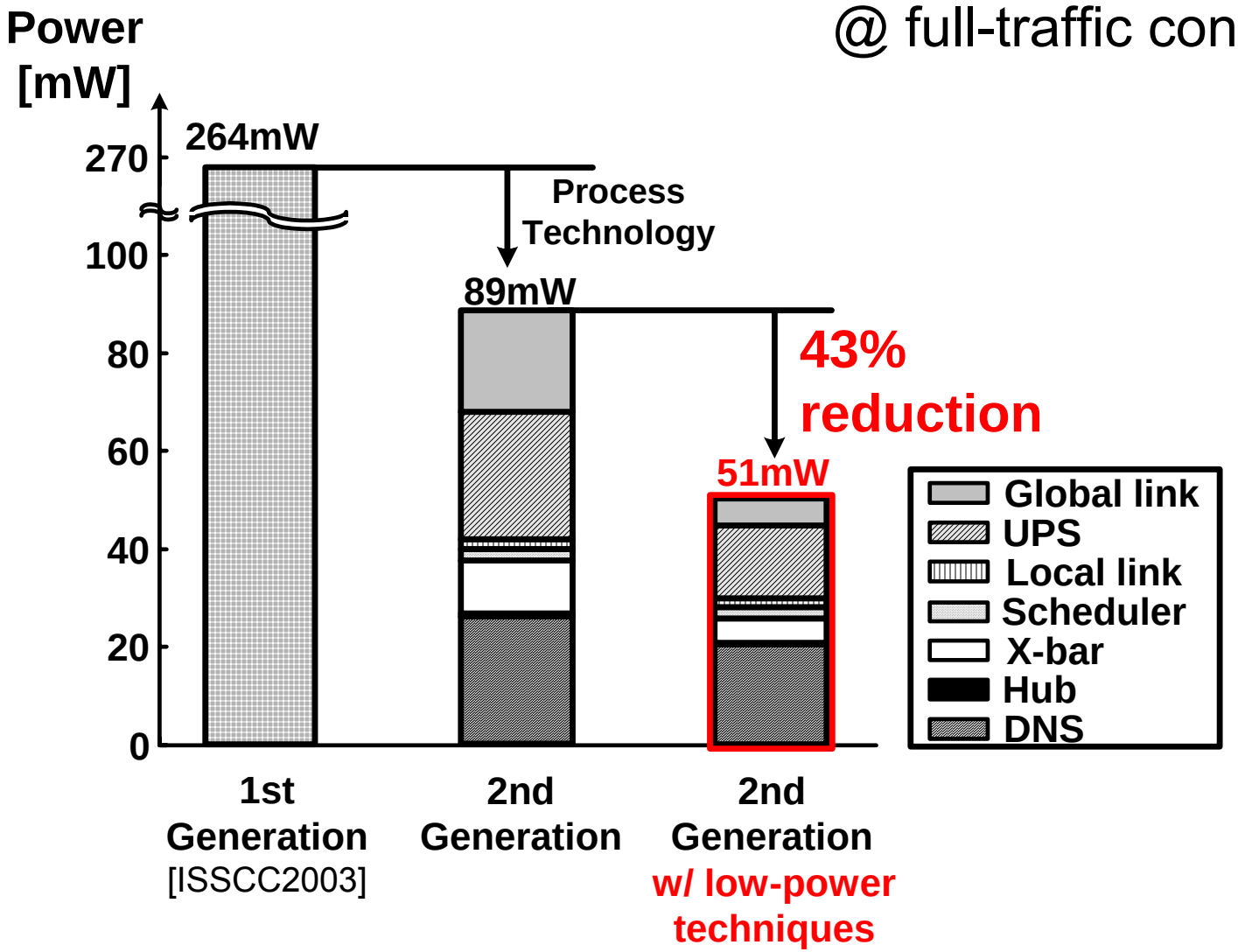
- **The low-power OCN-based SoC platform**



- □ **0.18μm 6M CMOS Tech.**
- □ **5mm x 5mm**
- □ **Power Supply**
  - **1.6V: Logic/Analog**
  - **3.3V: I/O**
- □ **OCN Power Consumption**
  - **Less than 51mW**
- □ **Aggregate Bandwidth**
  - **11.2GB/s**
- □ **Various IPs for Multimedia App.**
  - **32b μP x 2 (@ 100MHz)**
  - **FPGA (64LE)**
  - **64kb SRAM x 2**
  - **Off-chip Gateway**

Overall OCN Power Reduction

@ full-traffic condition



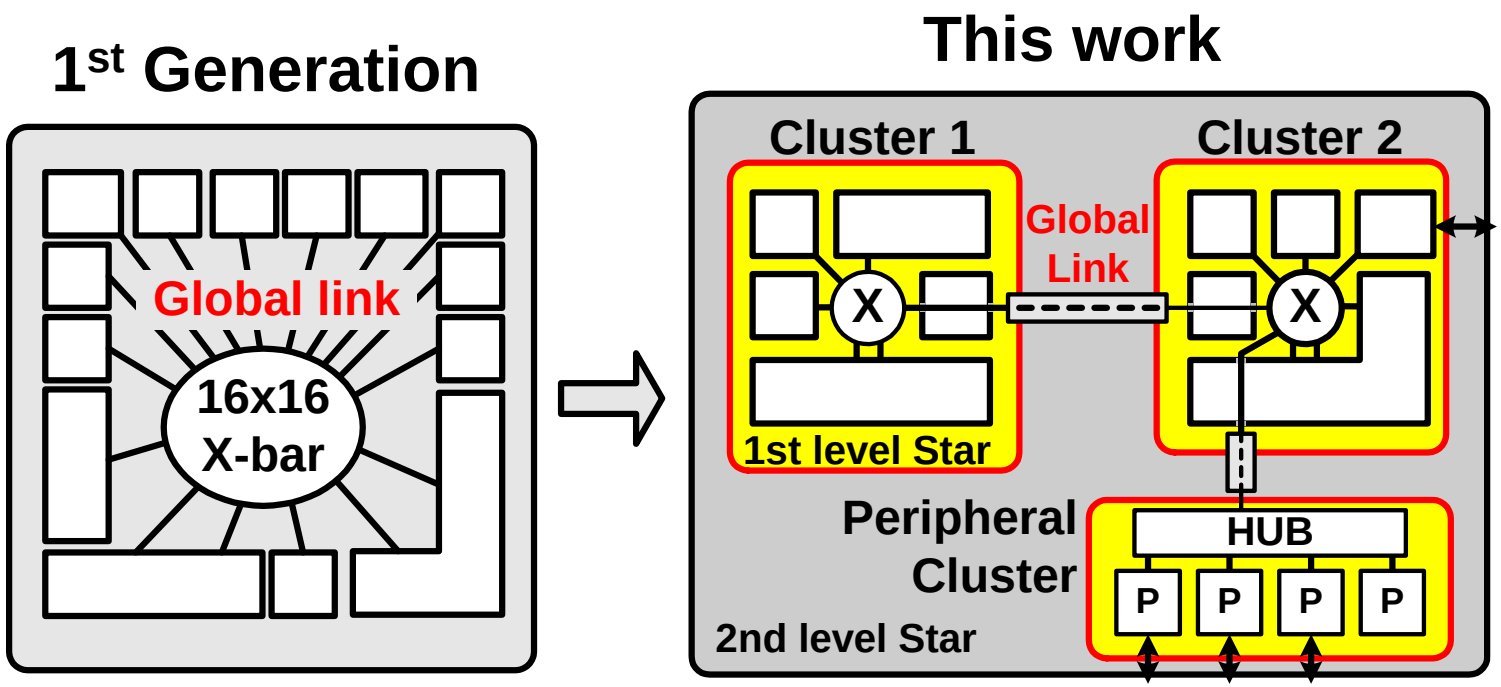
Conclusion

- **A Low-Power On-Chip Network for Heterogeneous SoC Platform**
  - Hierarchical Star Topology
  - Low-Power Techniques save **43%** overall network power.
    - Low-swing global link
    - Crossbar partial activation
    - Serial-link encoding
    - Frequency scaling
  - 51mW, 11.2GB/s Bandwidth
- **Various IPs for Multimedia Applications**
  - μP x 2 + FPGA + Memory + Off-chip gateway

Topology

Supplementary for Q&A

- Cluster based Hierarchy reduces # of global links
  → Power Saving by Half

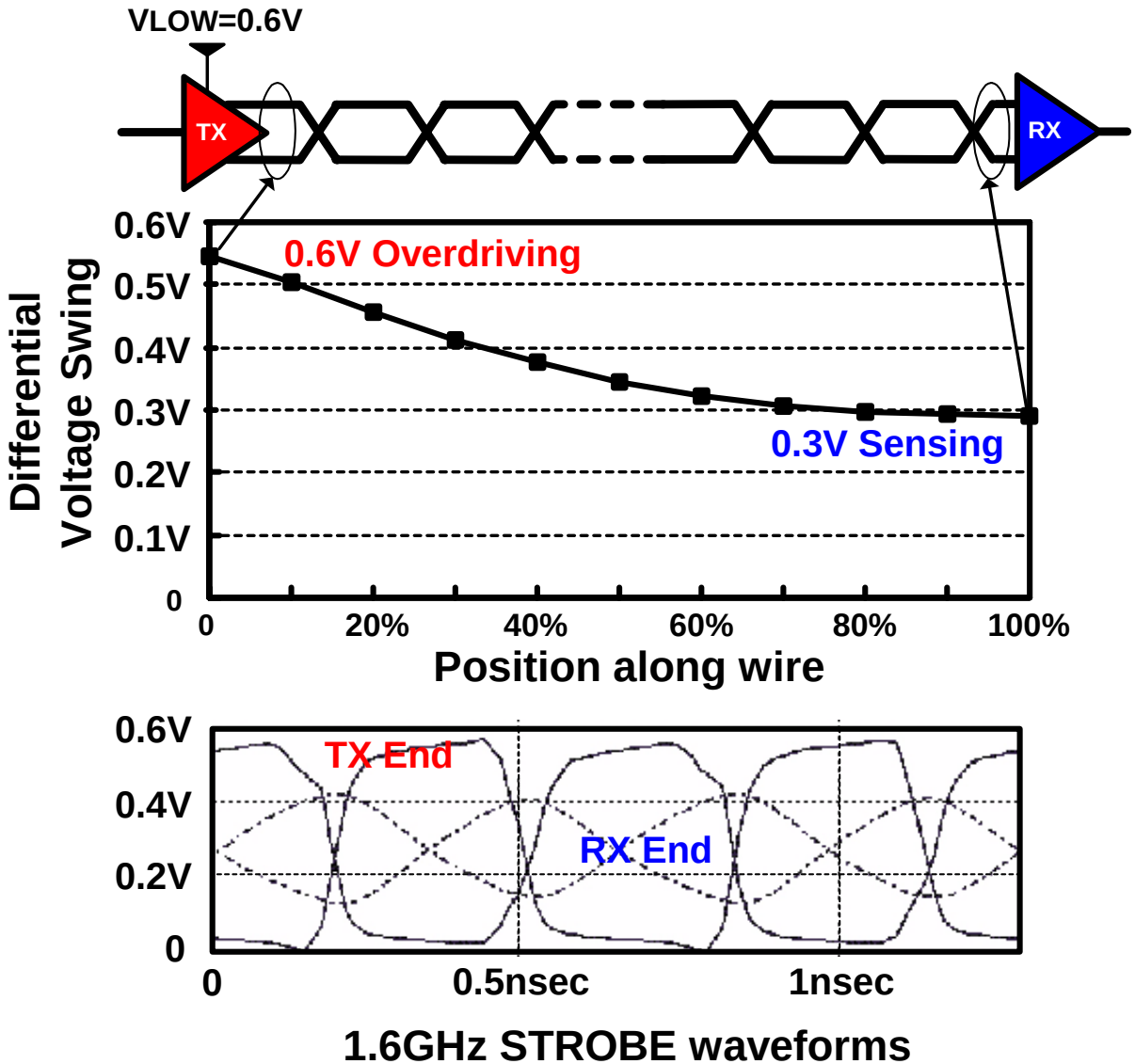


Flat Star-topology

Hierarchical Cluster-based Star-topology

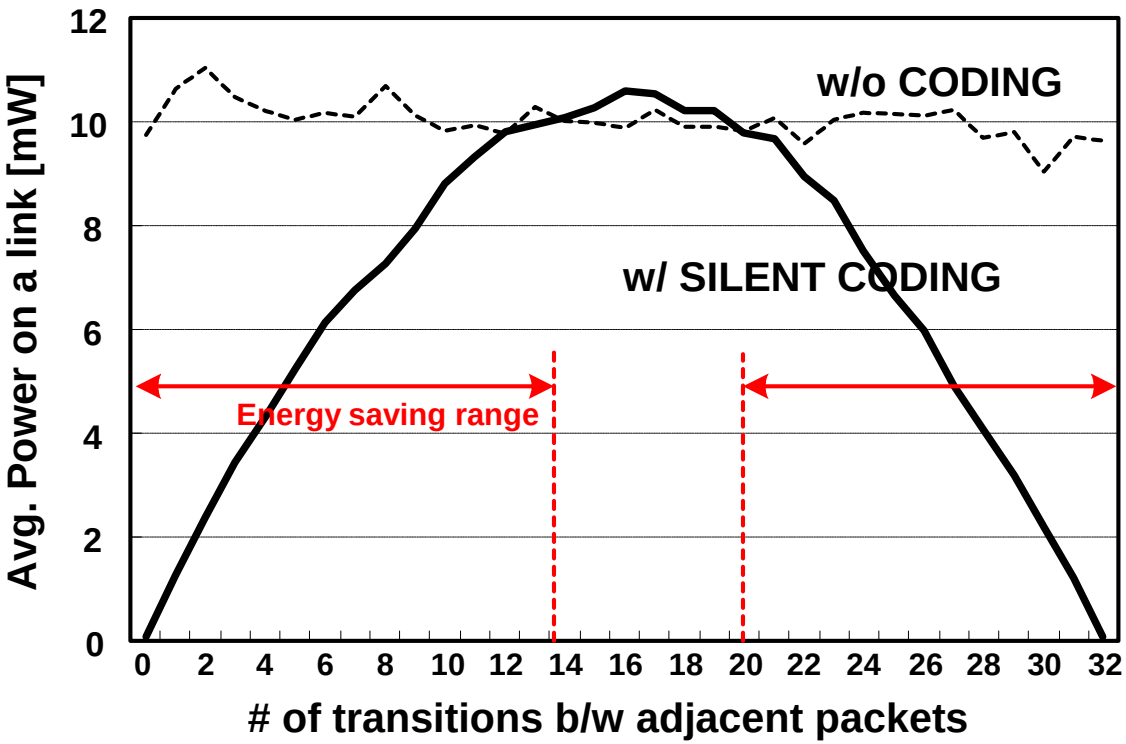
Low-Swing Global Link

Supplementary for Q&A



SILENT: Serial Link Encoding

Supplementary for Q&A



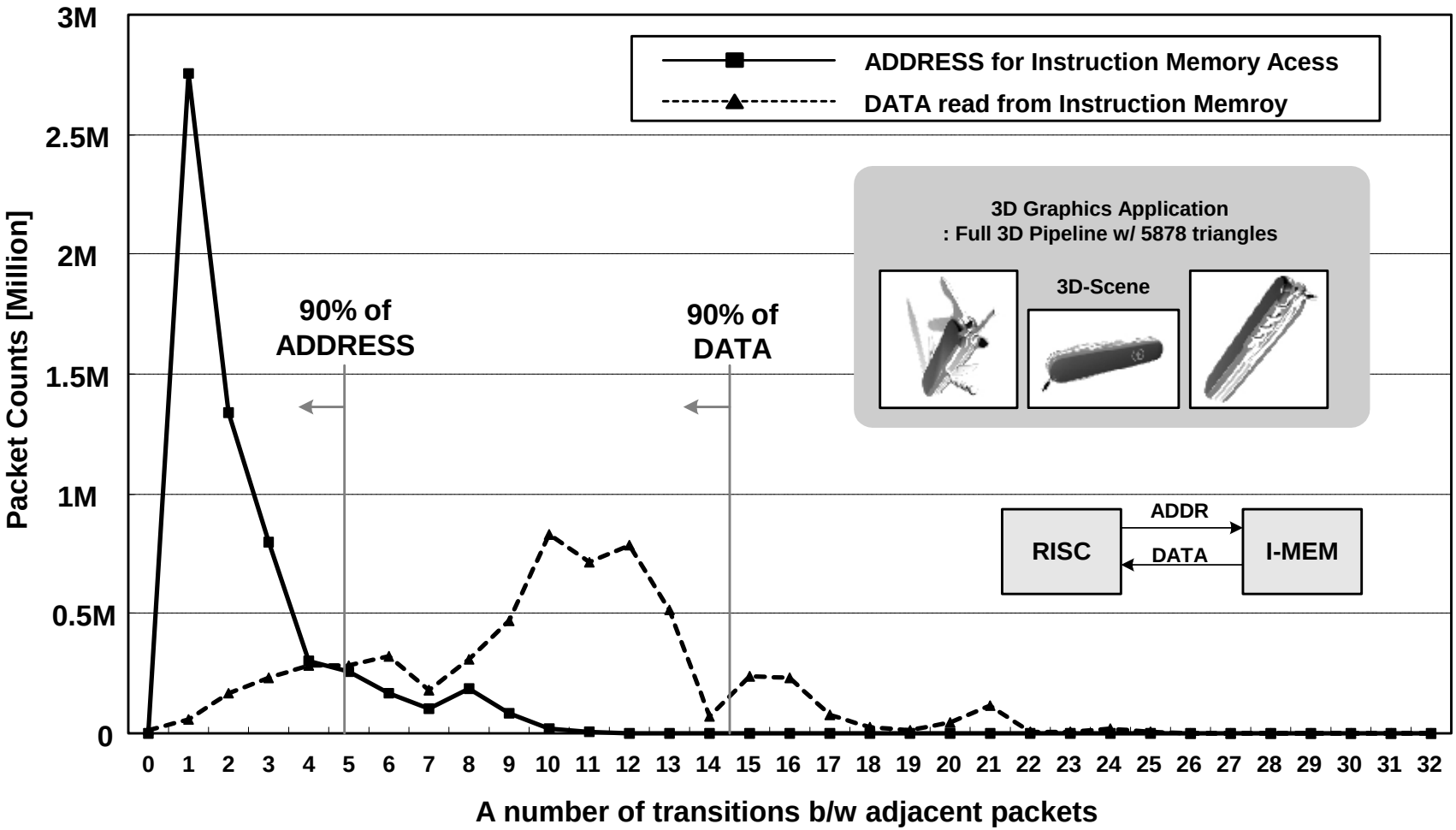
An example

00010001_010.... 1st Packet
00011101_010.... 2nd Packet

: 2 transitions

SILENT: Serial Link Encoding

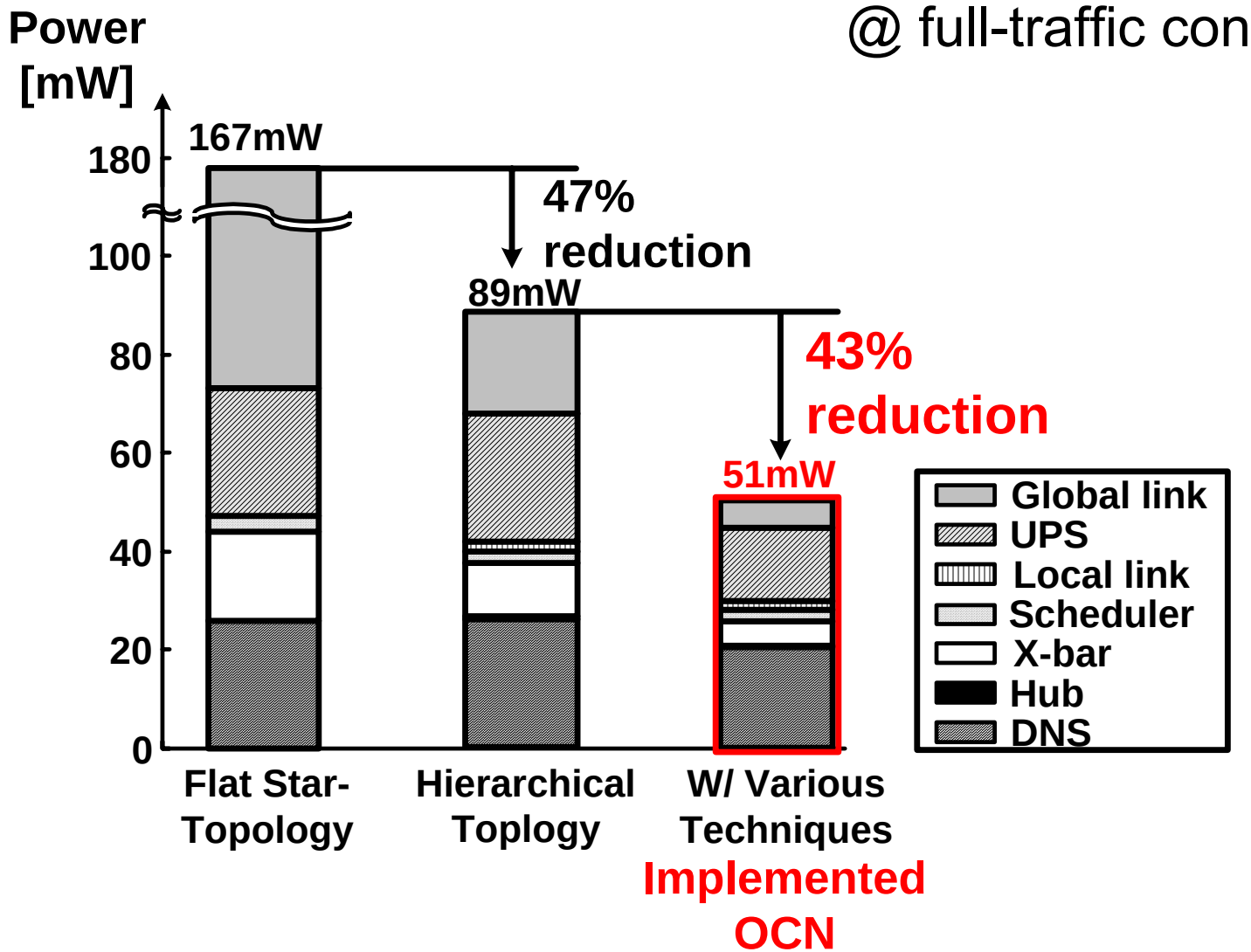
Supplementary for Q&A



Overall Power Reduction

Supplementary for Q&A

@ full-traffic condition



Overall Power Consumption

Supplementary for Q&A