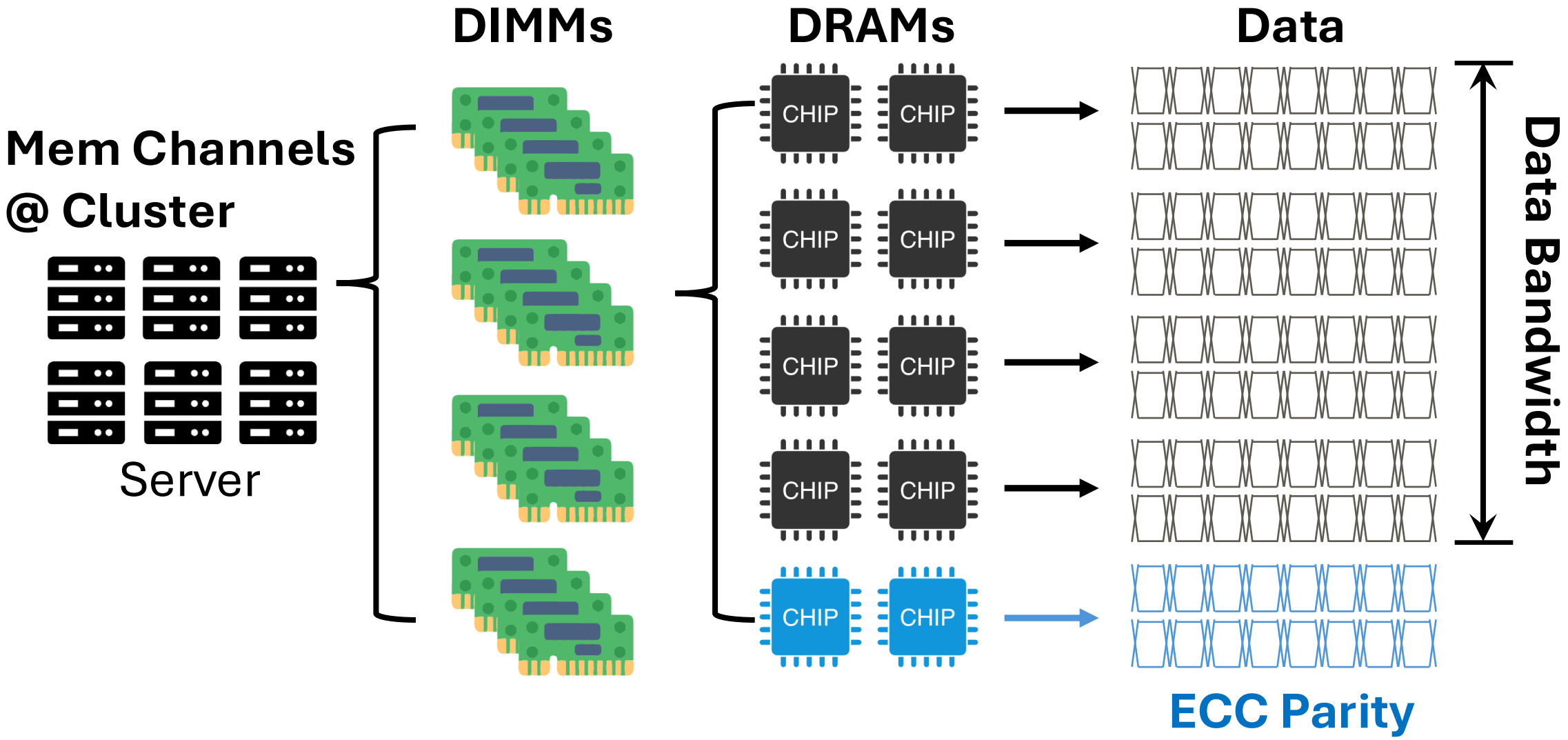
32nd IEEE International Symposium on High-Performance Computer Architecture (HPCA-32)

ASPA: Reassigning DDR5 Parity Bandwidth

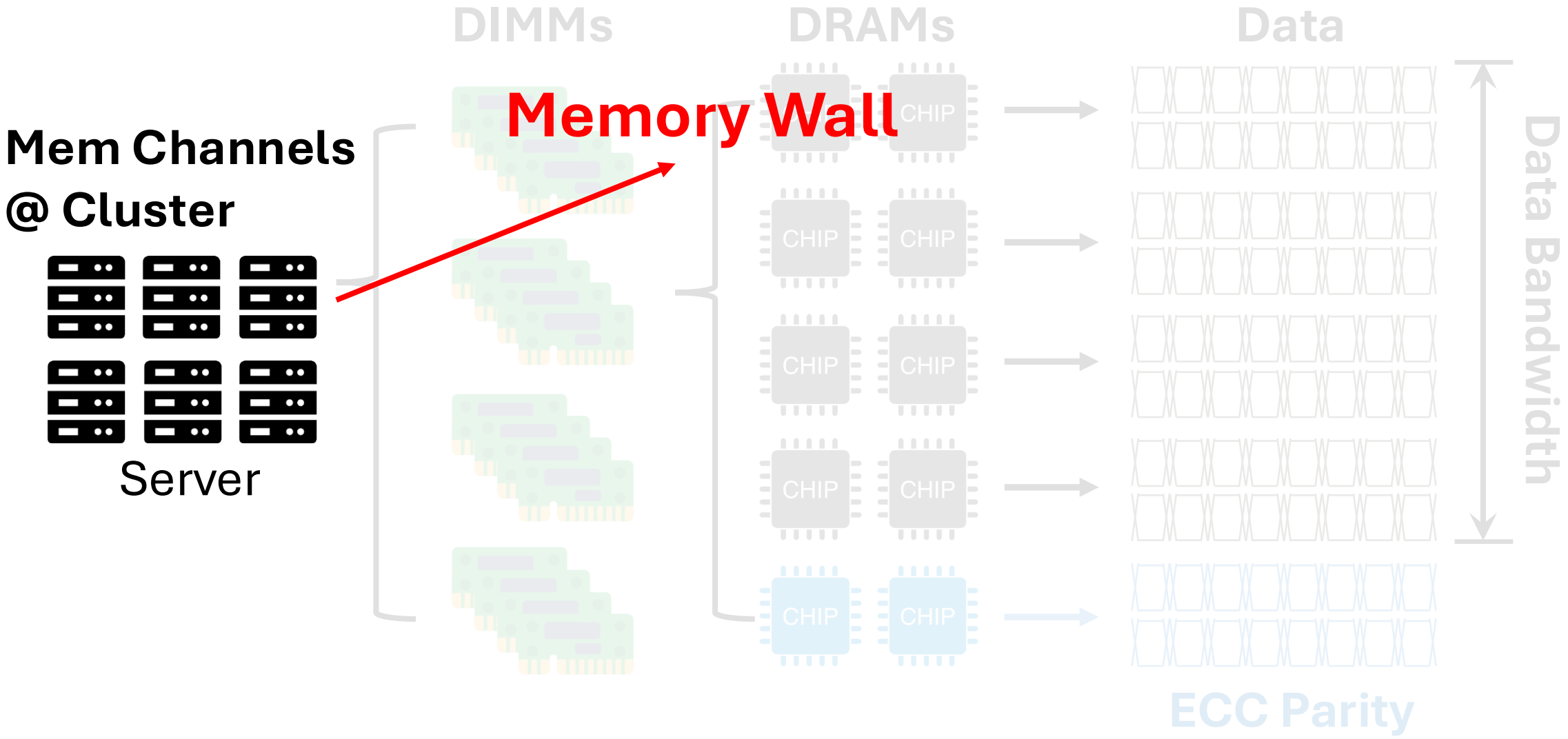
Fan Li†, Qiufeng Li‡ , *YananGuo*§, Weidong Cao‡ , Xin Xin†

†University of Central Florida, USA
‡ George Washington University, USA
§University of Rochester, USA

DDR5 Memory Systems



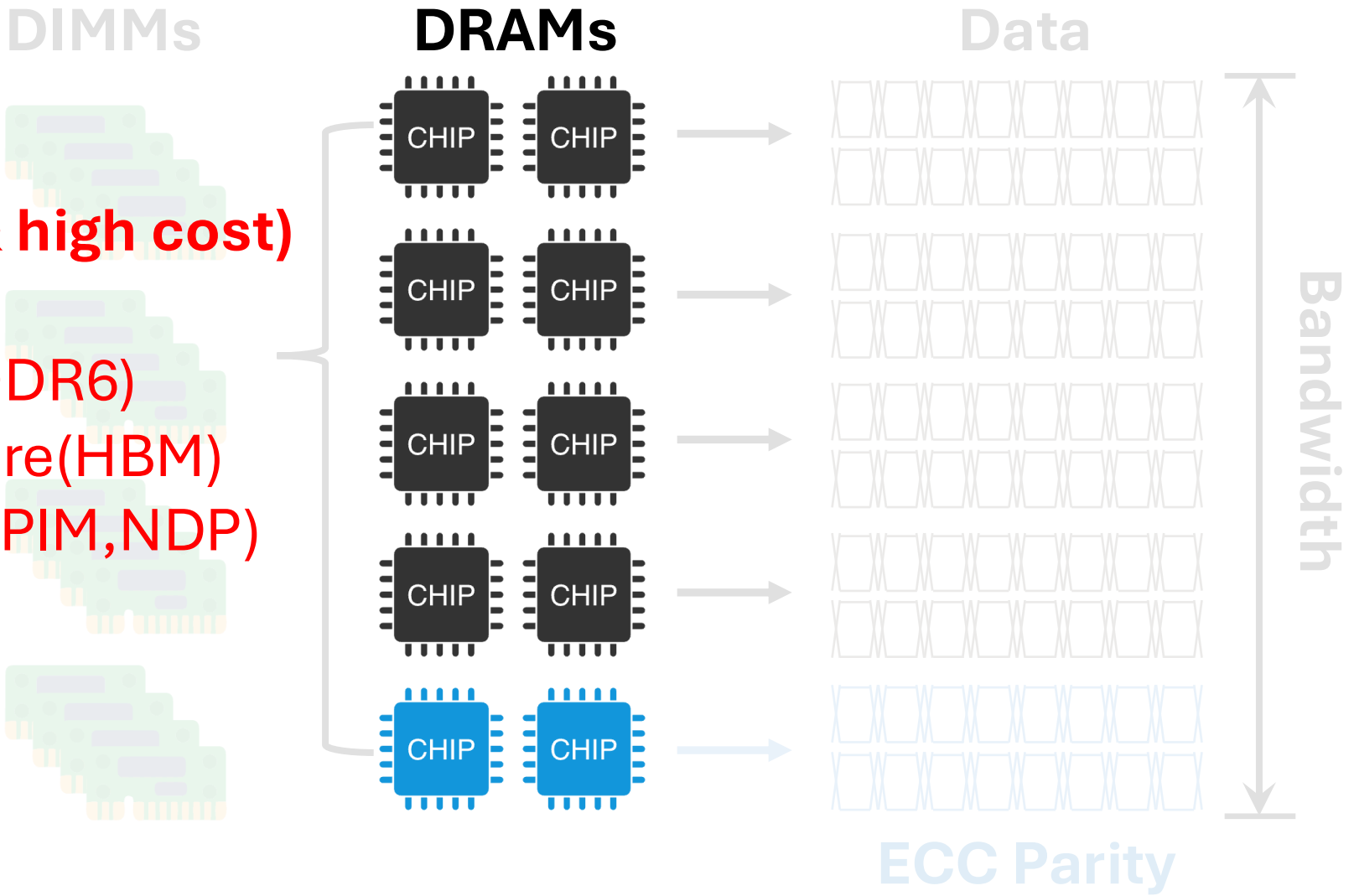
DDR5 Memory Systems



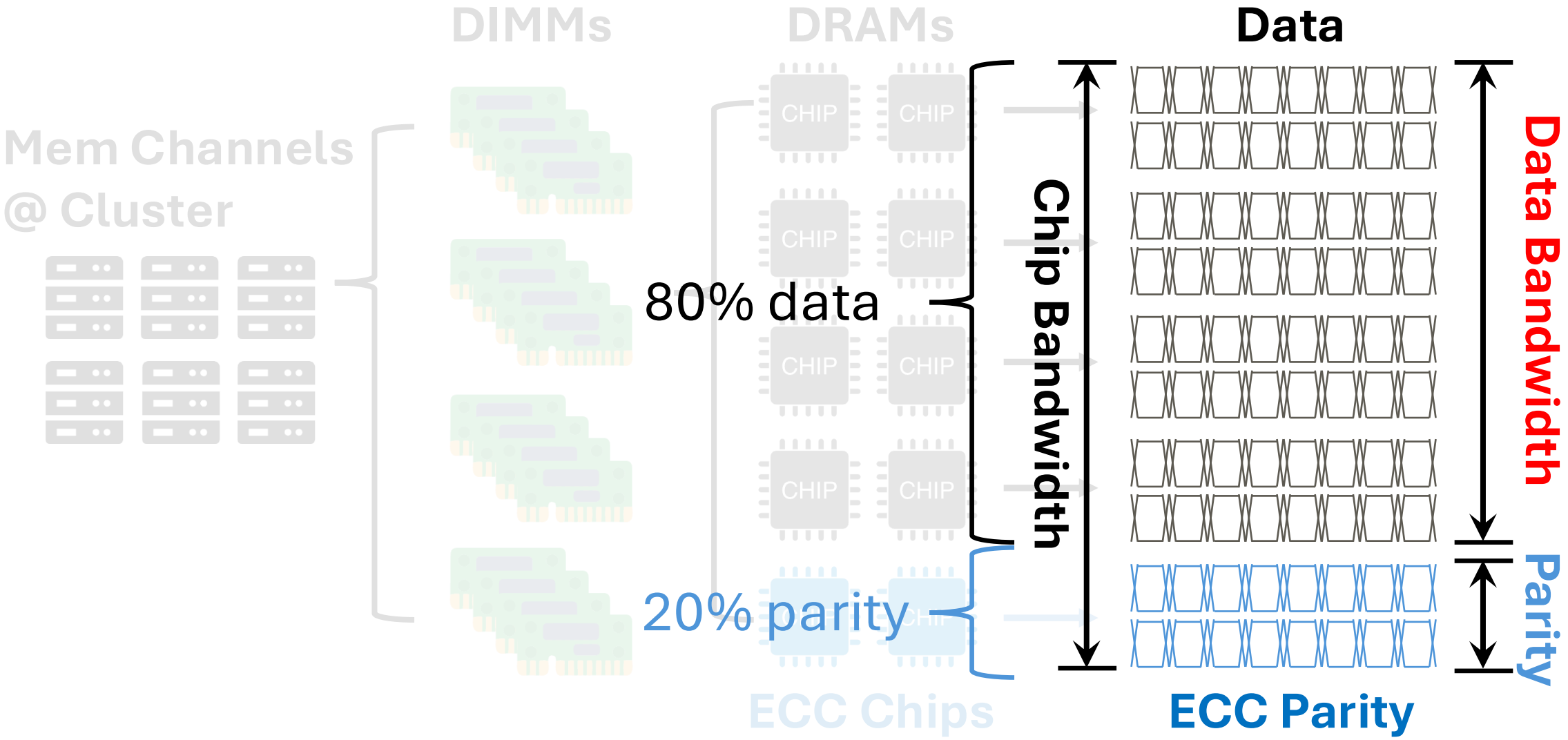
DDR5 Memory Systems

Popular solutions
(bandwidth extension & high cost)

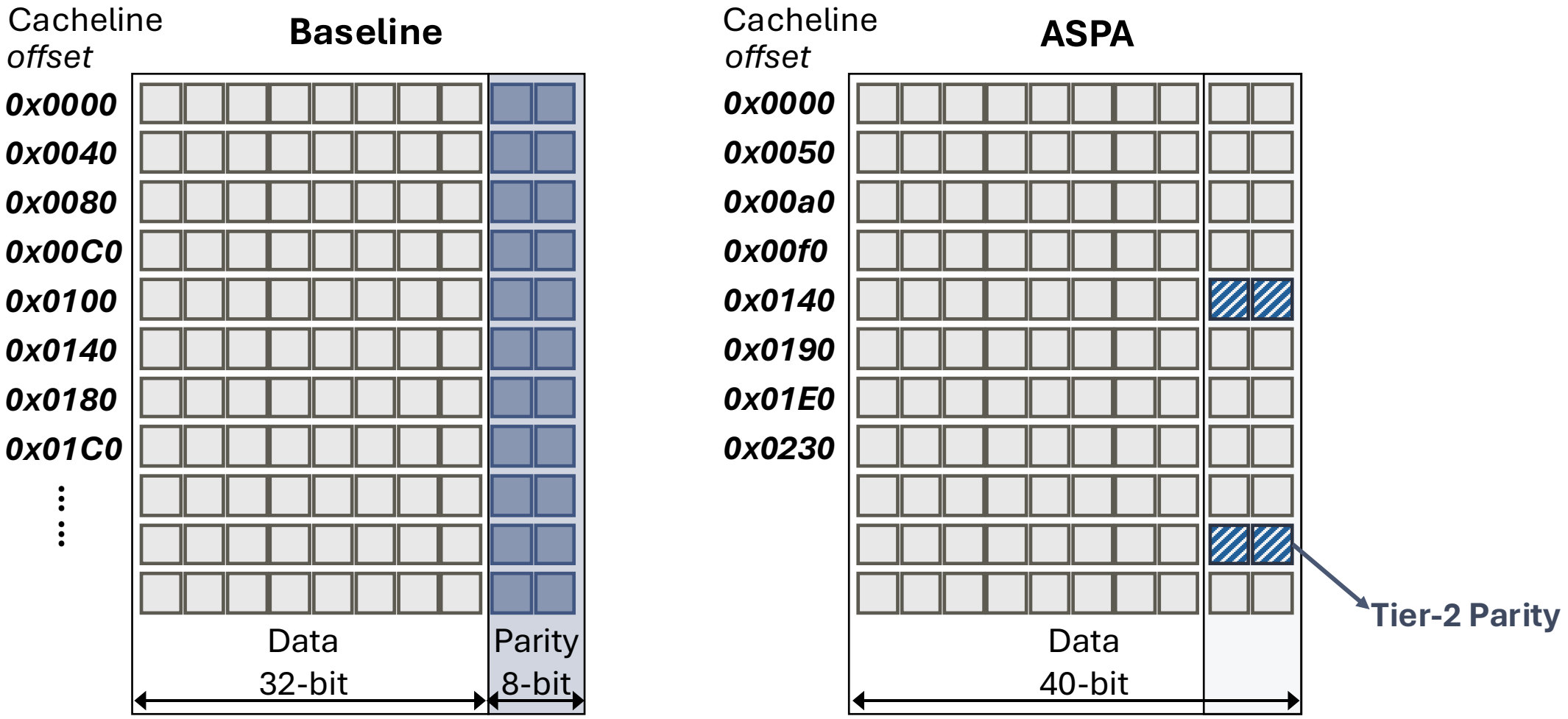
-New DDR generations(DDR6)
-New memory architecture(HBM)
-New memory paradigm(PIM,NDP)
…



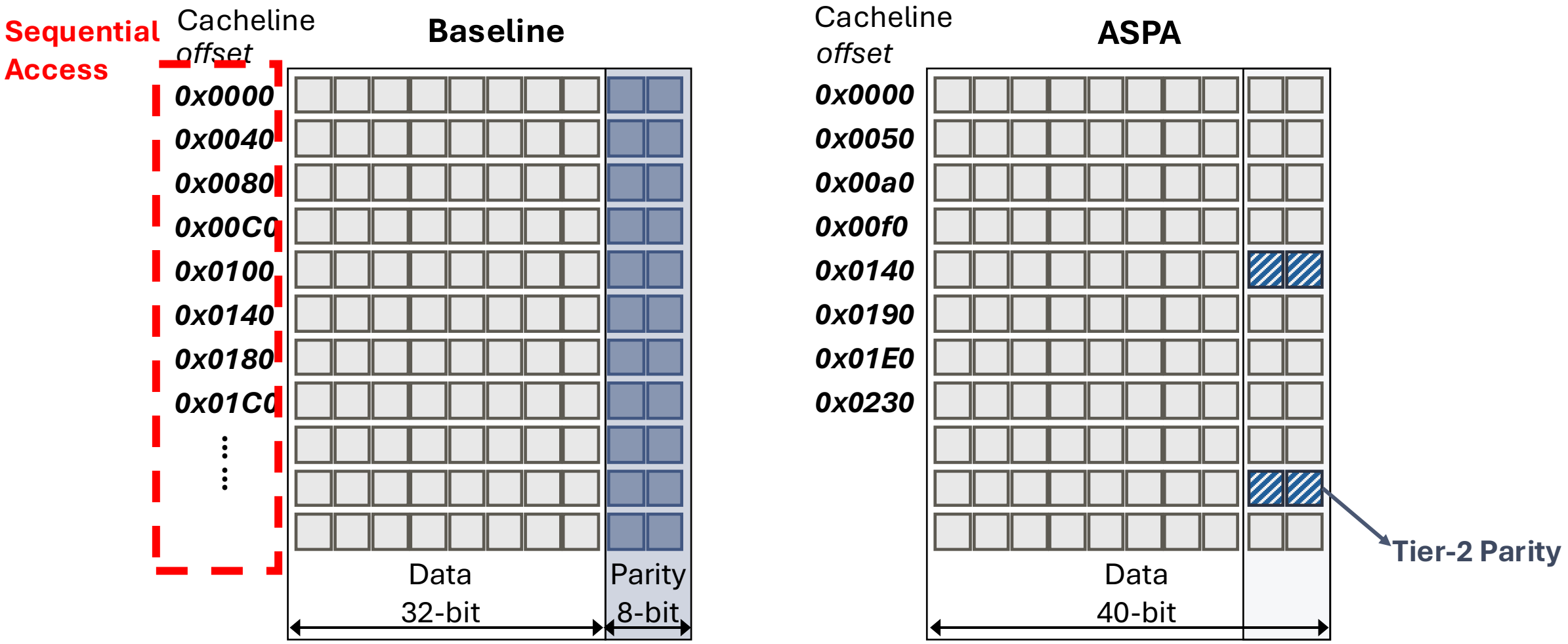
DDR5 Memory Systems



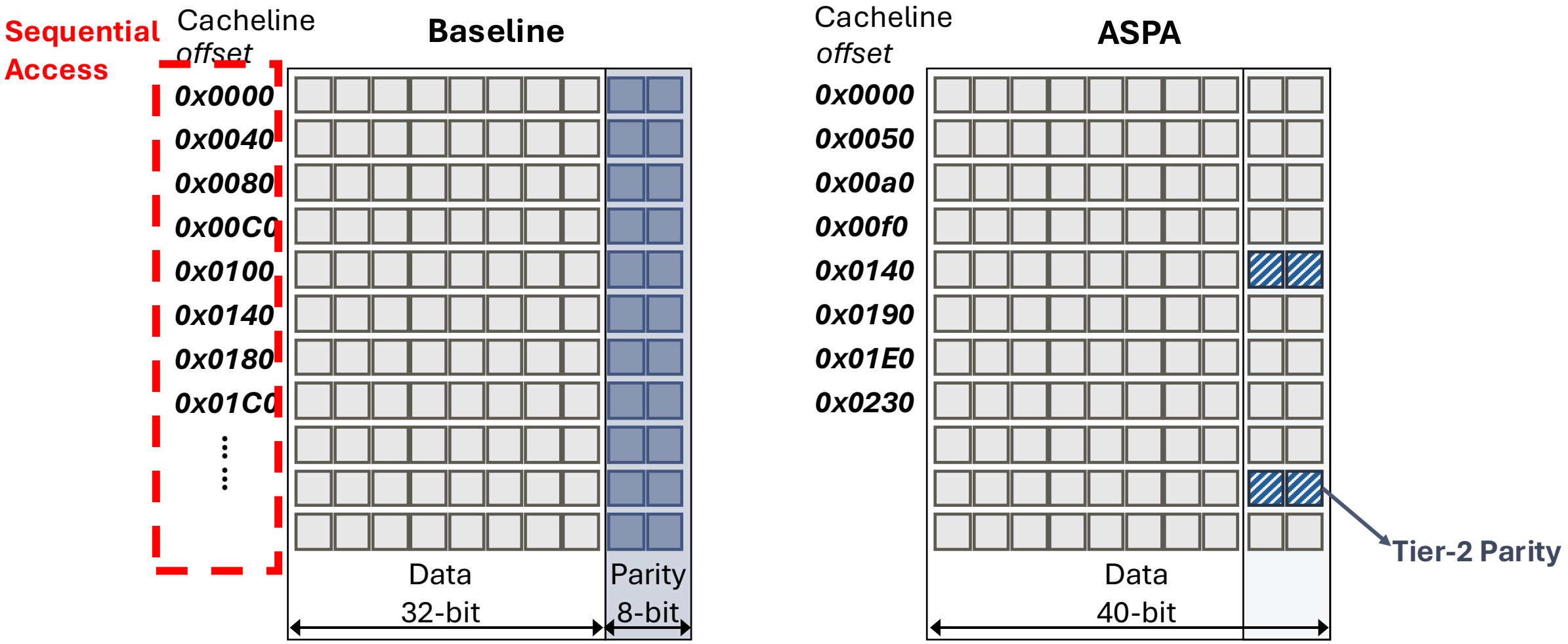
Design Overview



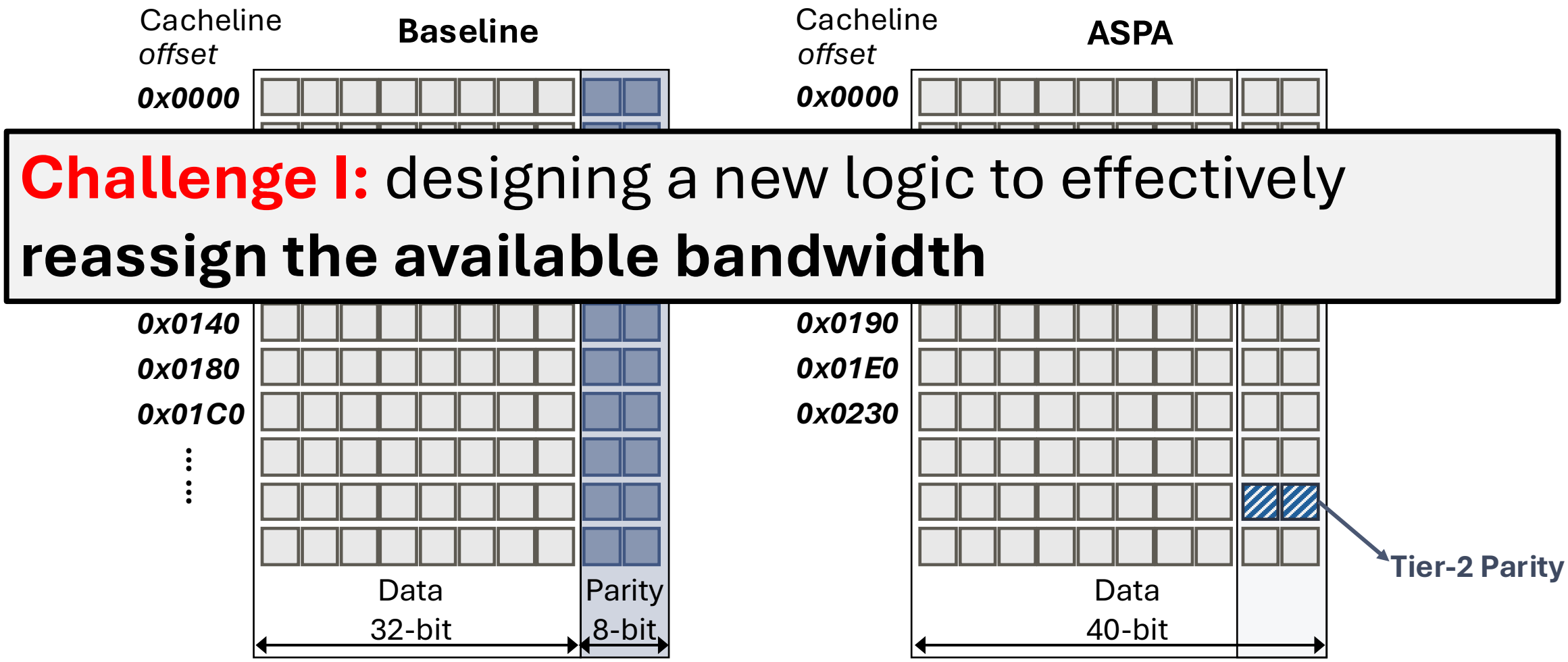
Design Overview



Design Overview

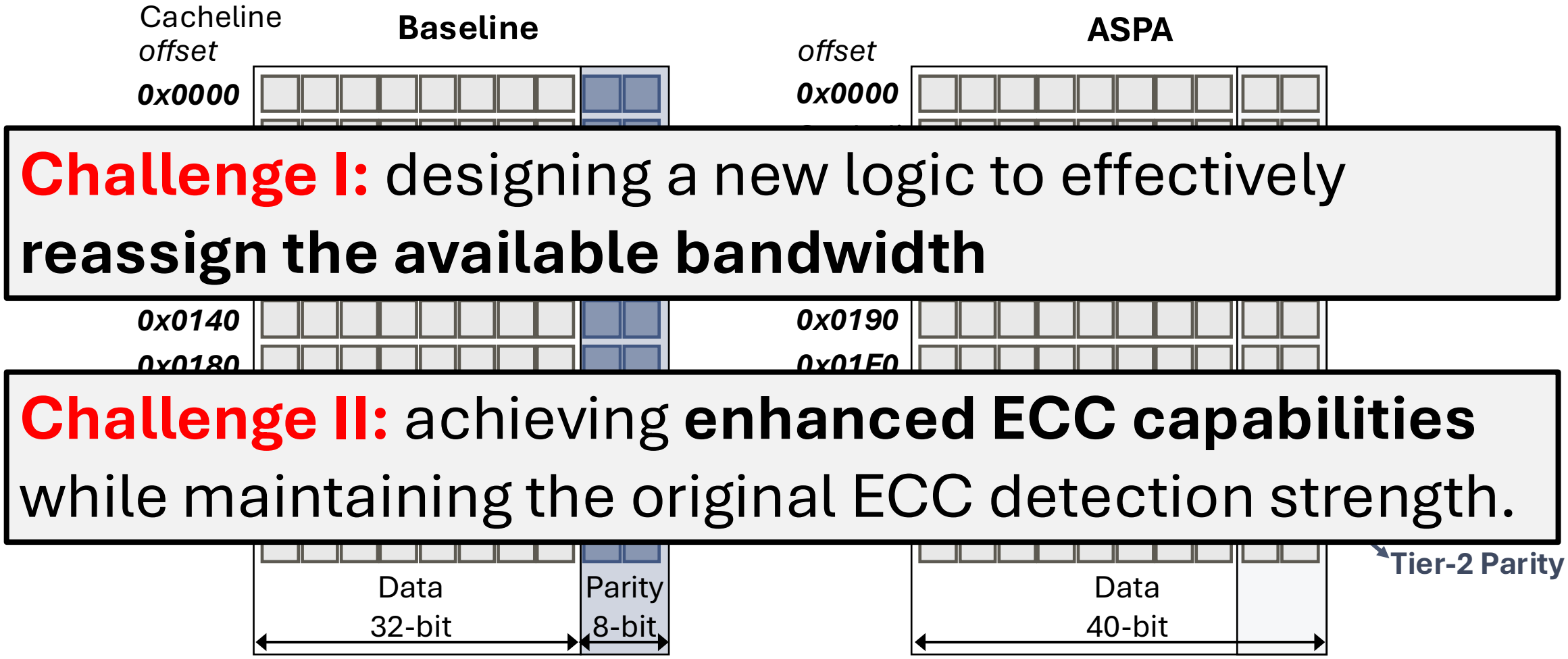


Design Overview



Challenge I: designing a new logic to effectively **reassign the available bandwidth**

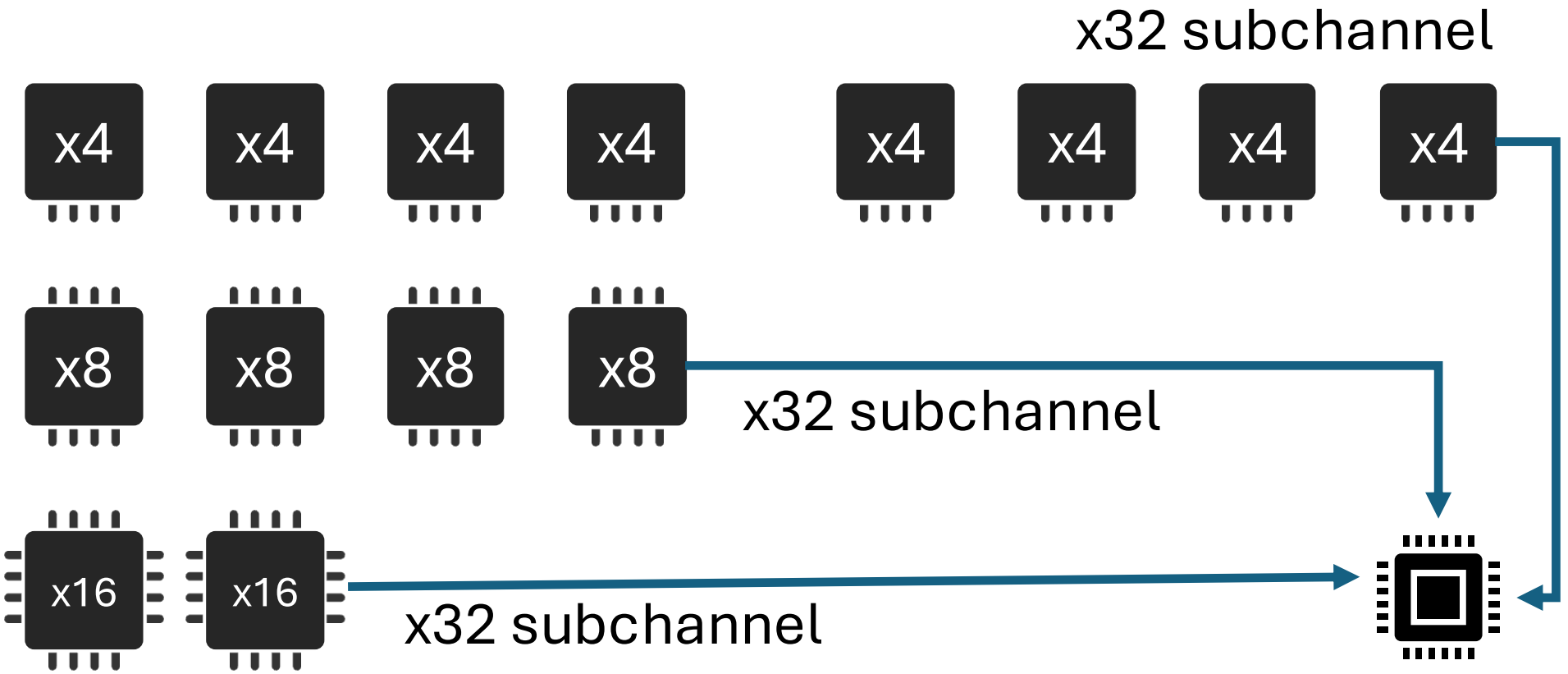
Design Overview



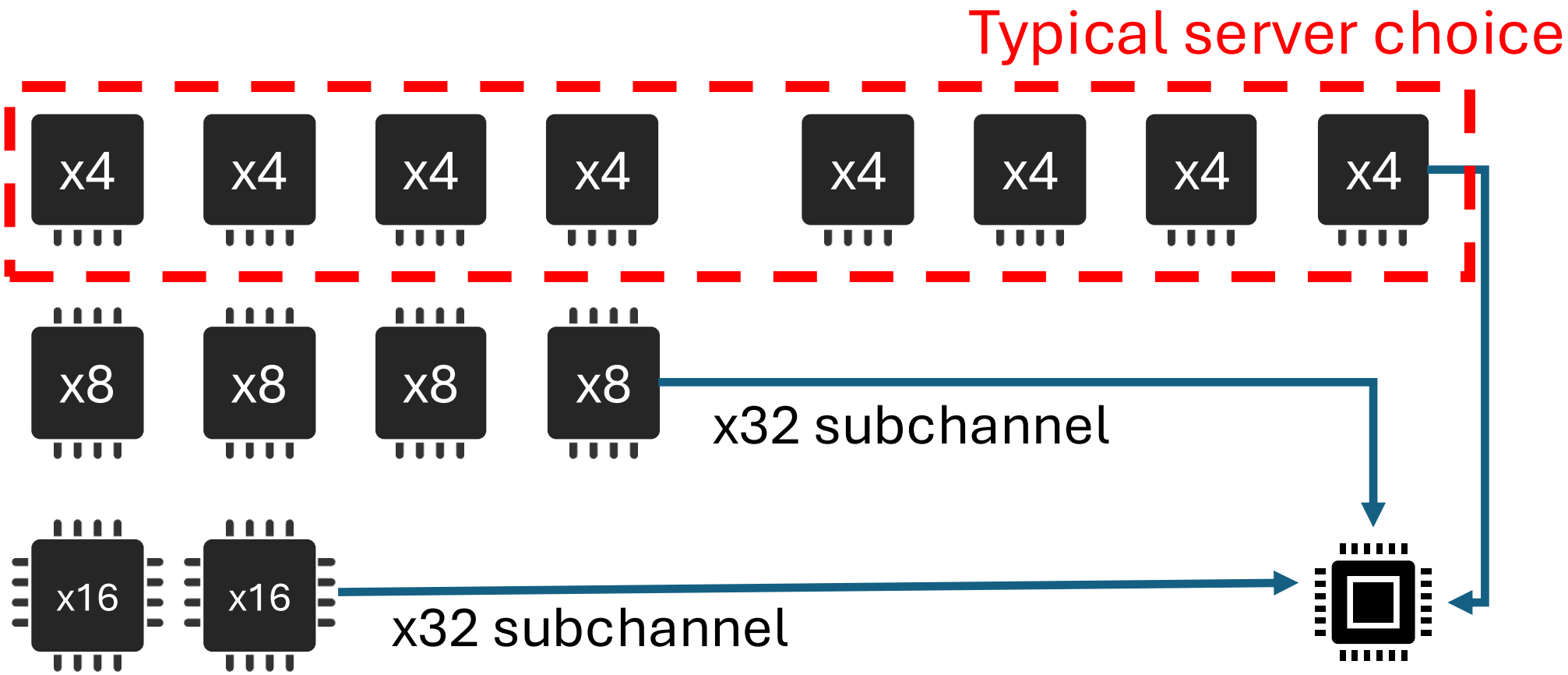
Challenge I: designing a new logic to effectively **reassign the available bandwidth**

Challenge II: achieving **enhanced ECC capabilities** while maintaining the original ECC detection strength.

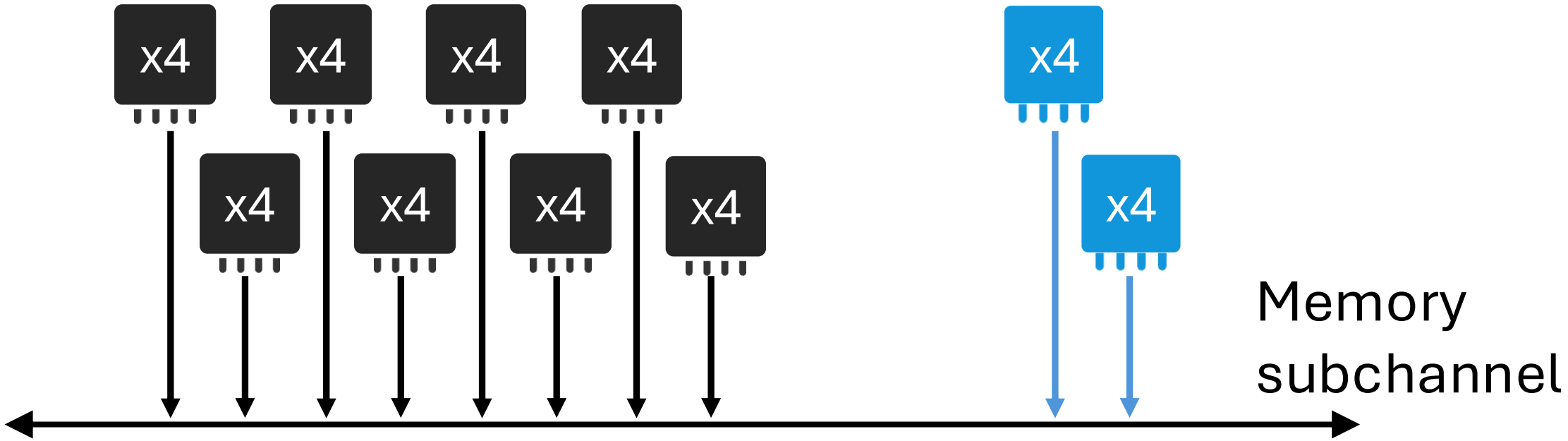
DRAM I/O Configurations



DRAM I/O Configurations



Bandwidth Distribution

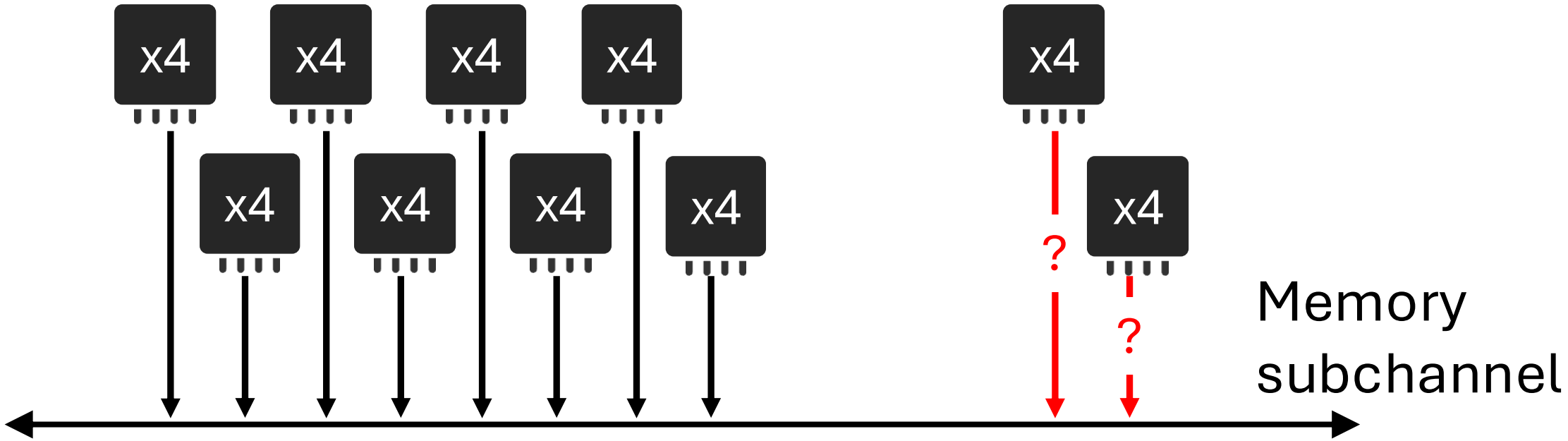


4b*BL*8chips=64B data
(Cacheline)

4b*BL*2chips=16B parity
(ECC codeword)

BL(Burst Length)=16

Bandwidth Distribution

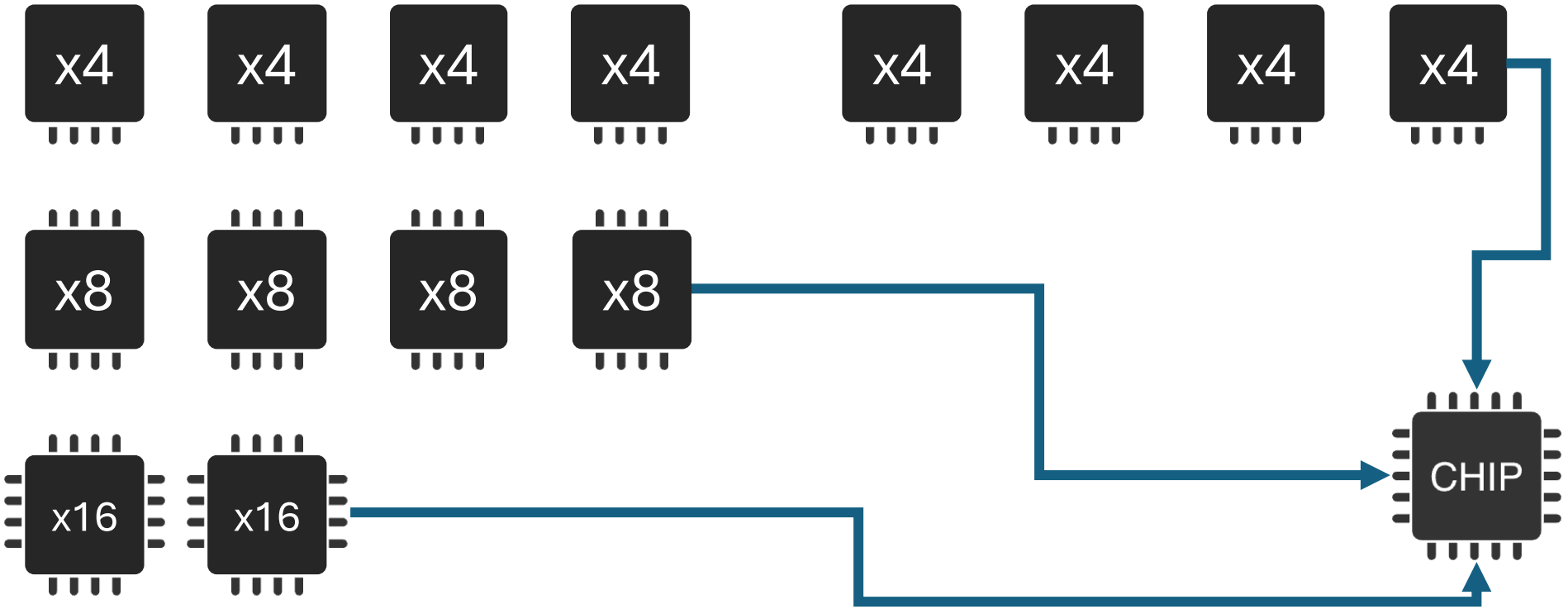


4b*BL*8chips=64B data
(Cacheline)

4b*BL*2chips=16B **data**
(1/4 Cacheline)

BL(Burst Length)=16

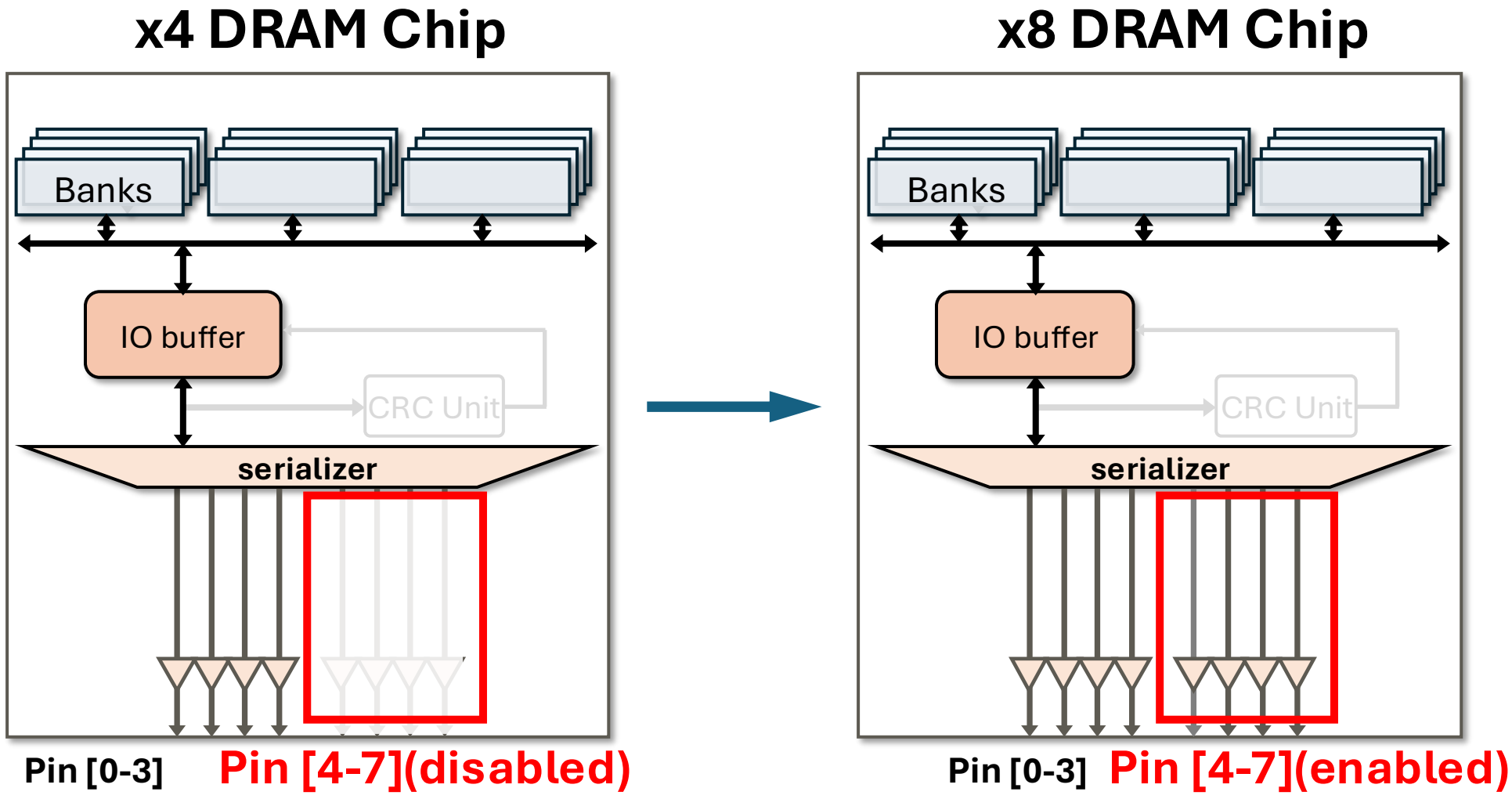
Common Die



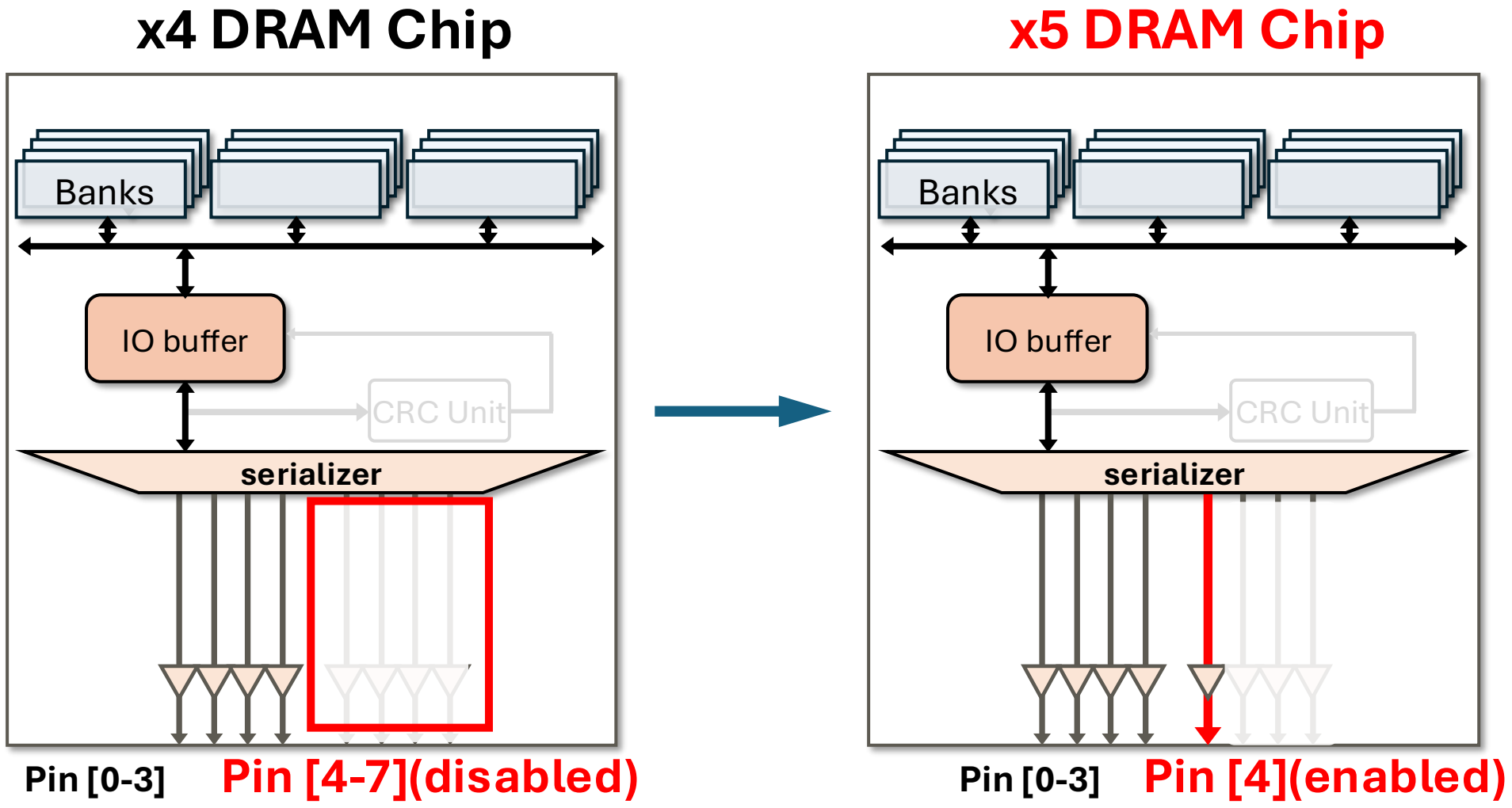
Common die

Same operation logic

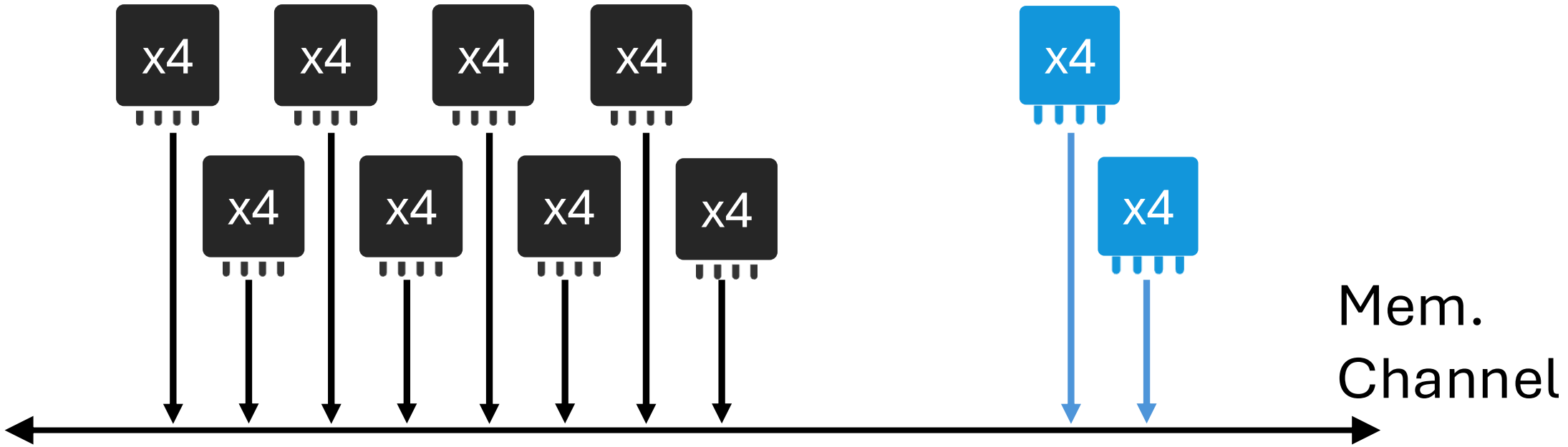
Challenge I: Reassign Bandwidth



Challenge I: Reassign Bandwidth



Challenge I: Reassign Bandwidth

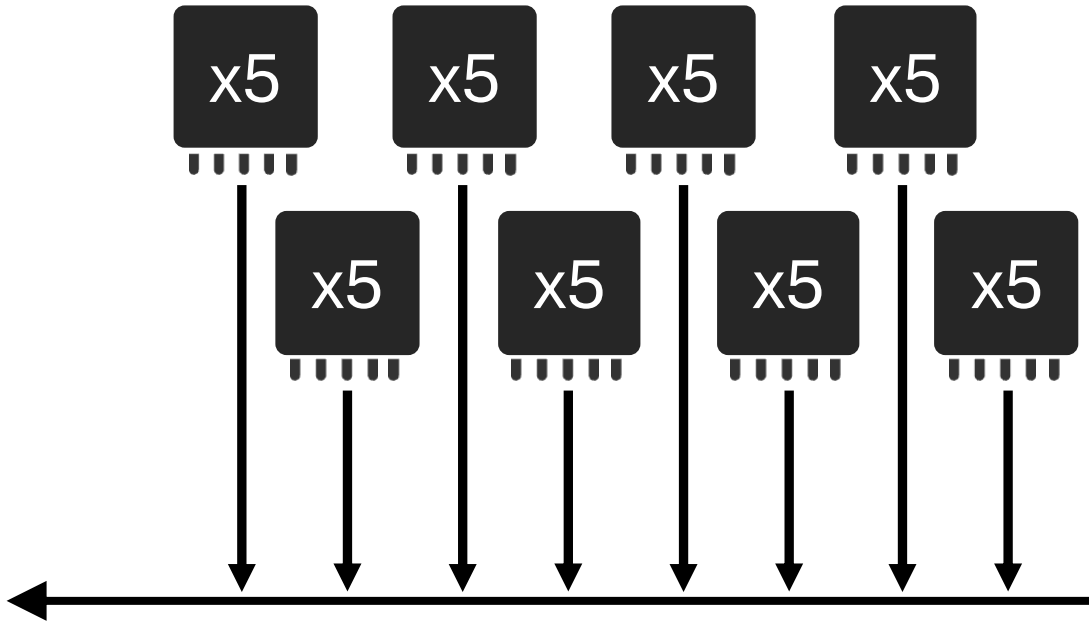


4b*BL*8chips=**64B data (Cacheline)**

4b*BL*2chips=128b parity

BL(Burst Length)=16

Challenge I: Reassign Bandwidth



Baseline

- **16 transfers = 1 Cacheline**

ASPA

- **26 transfers = 2 Cachelines**

Mem. Channel

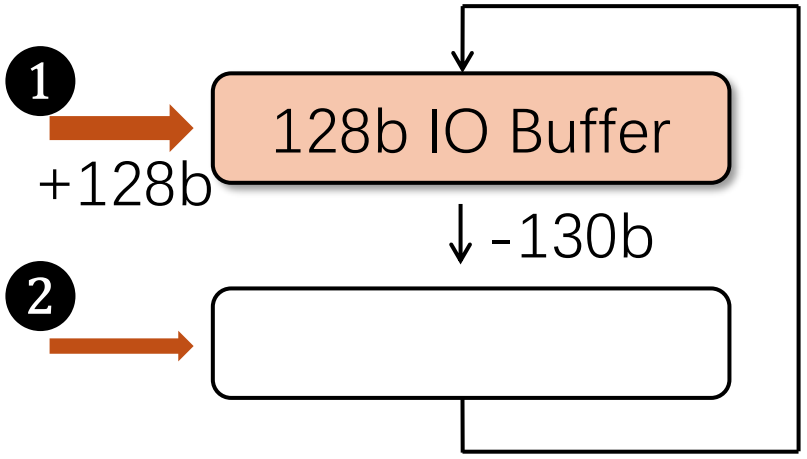
5b*BL*8chips=**130B data**

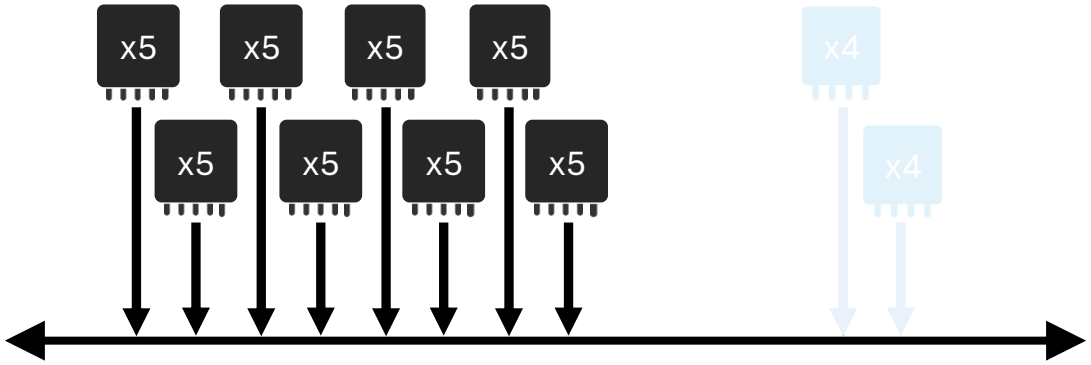
(2 Cachelines)

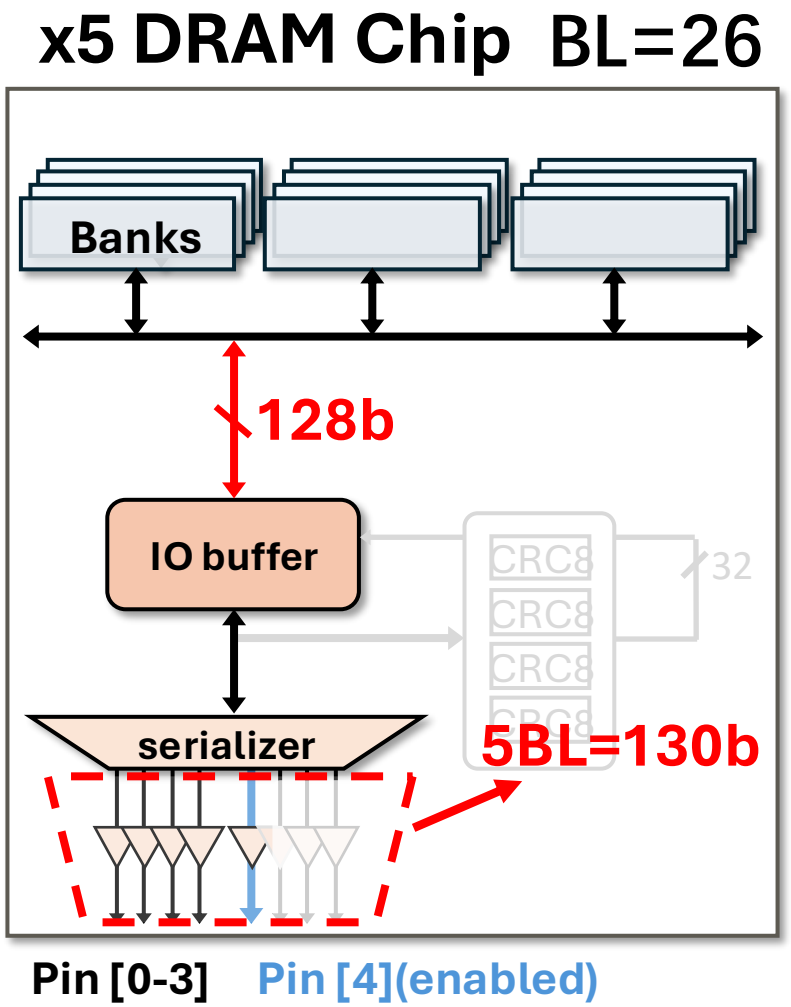
4b*BL*2chips=128b parity

BL(Burst Length)=26

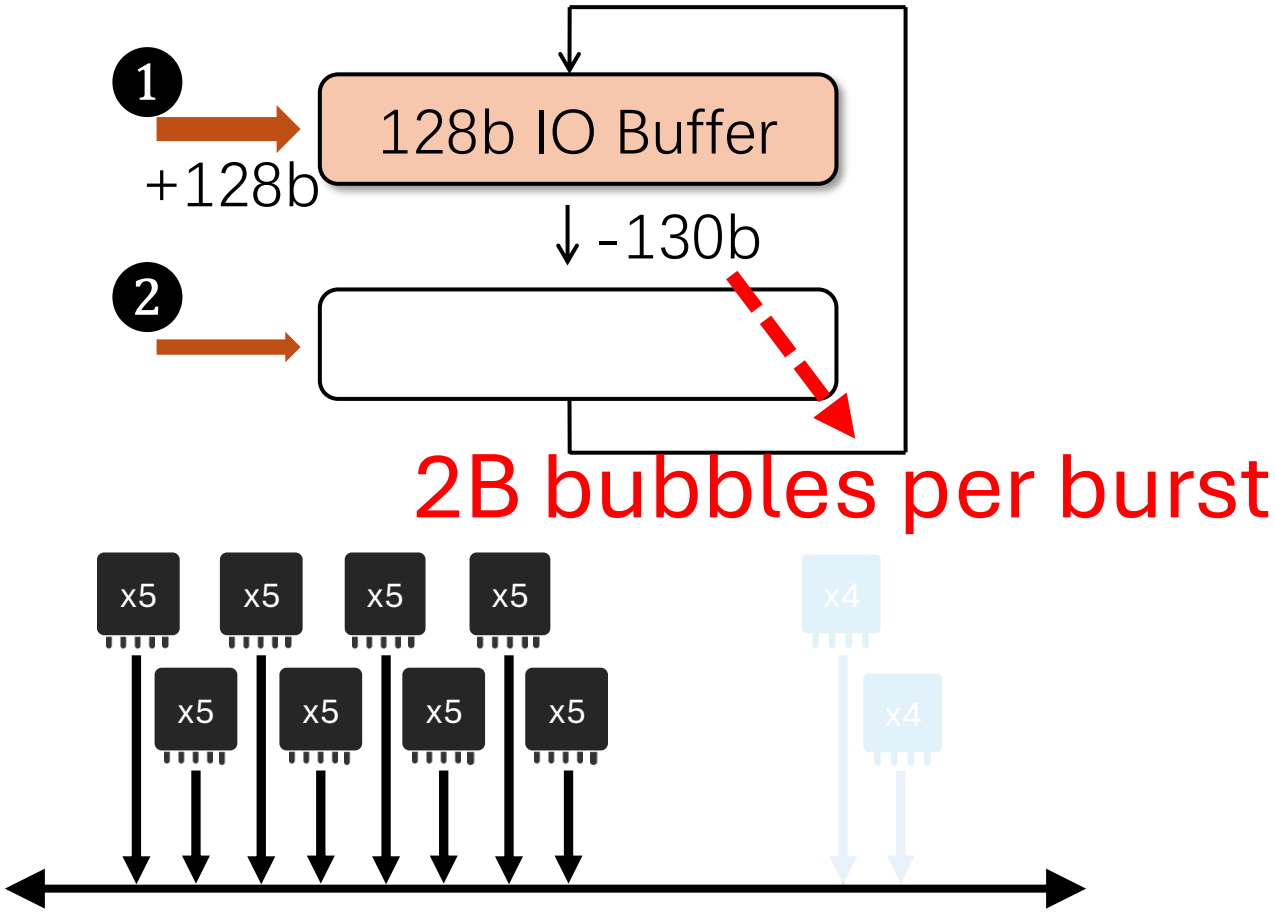
ASPA -- Fixed Buffering

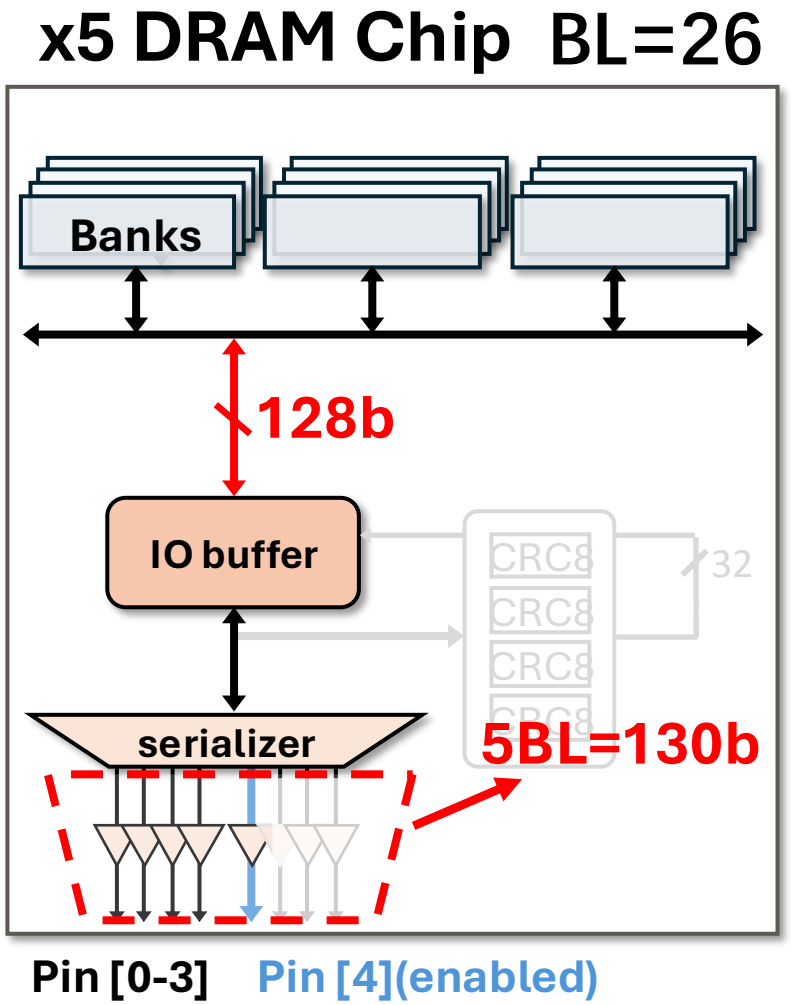




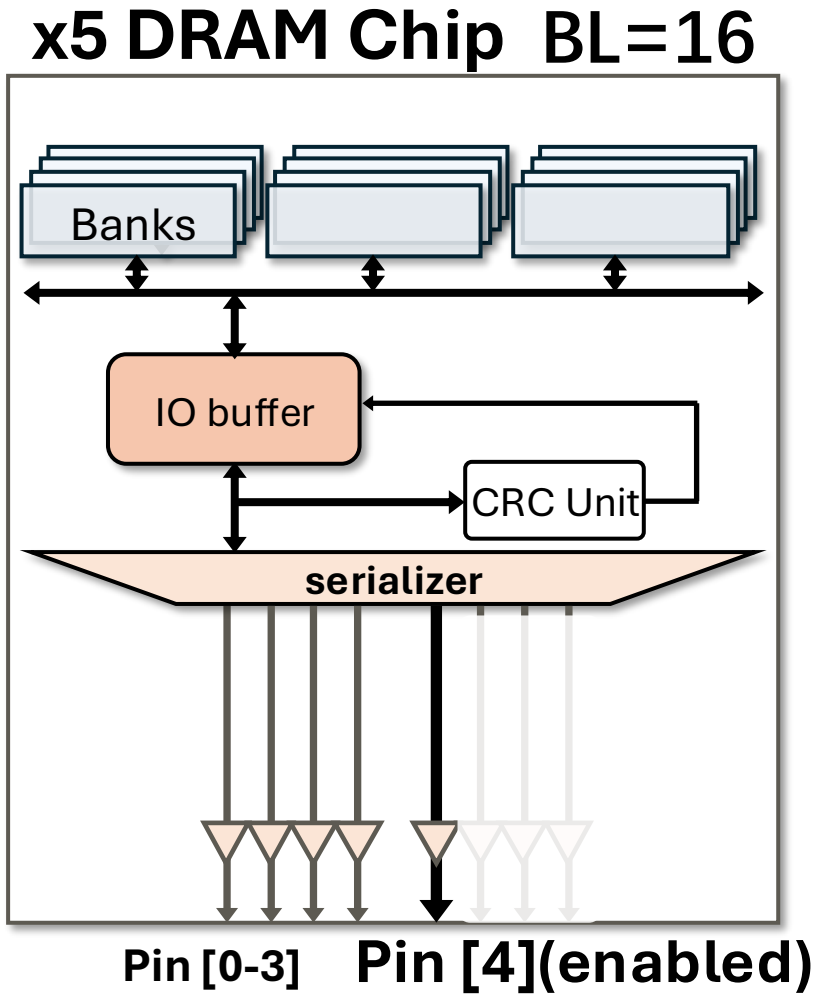


ASPA -- Fixed Buffering

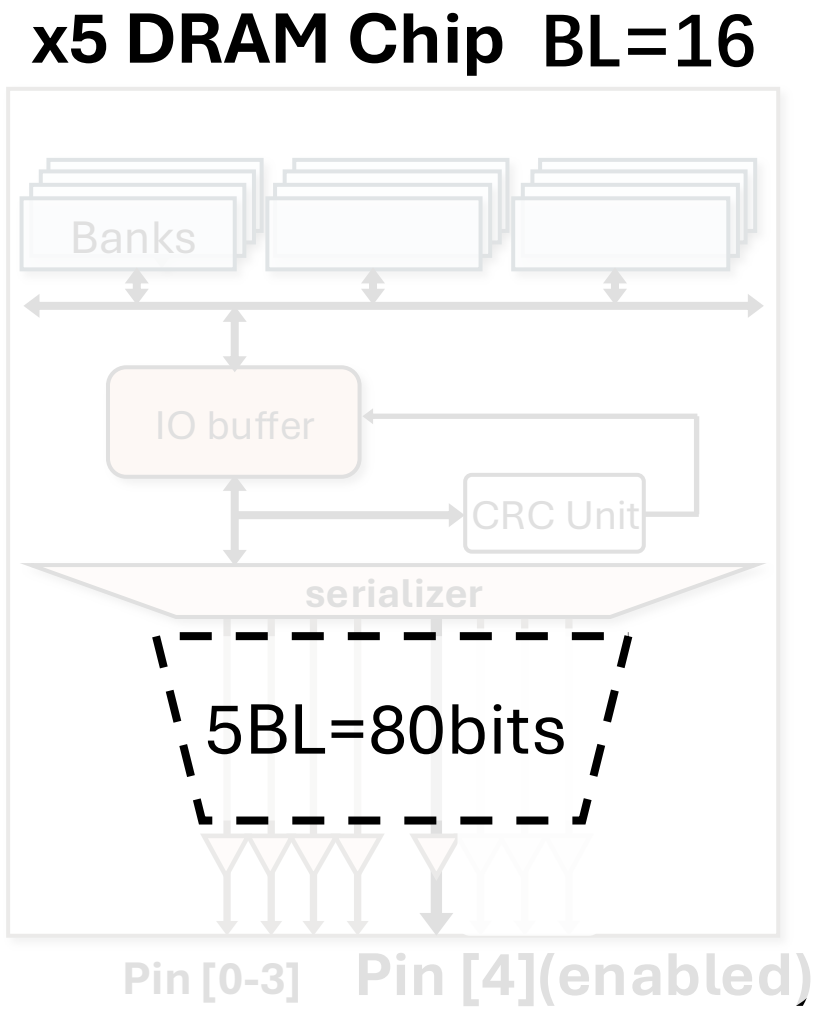




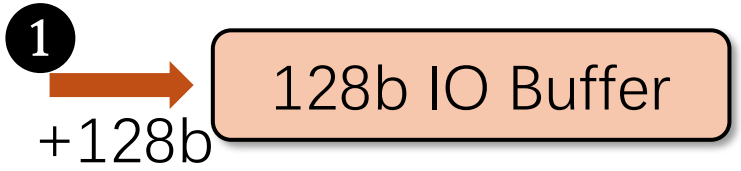
ASPA -- Adaptive Buffering

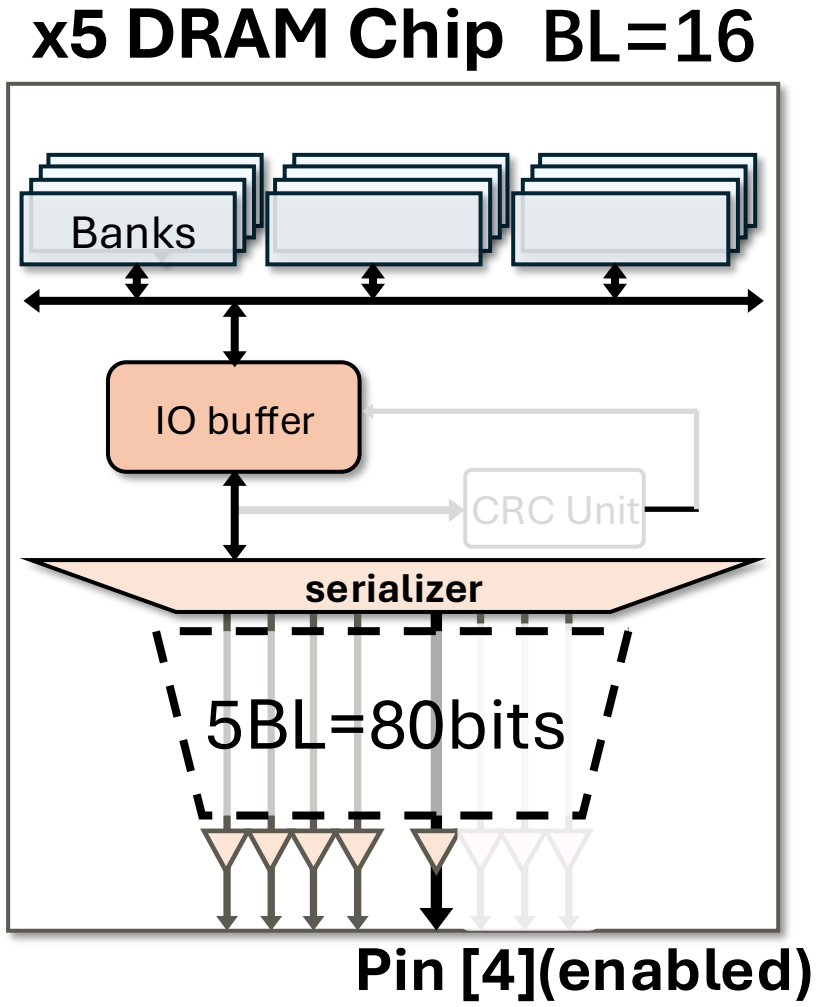


ASPA -- Adaptive Buffering

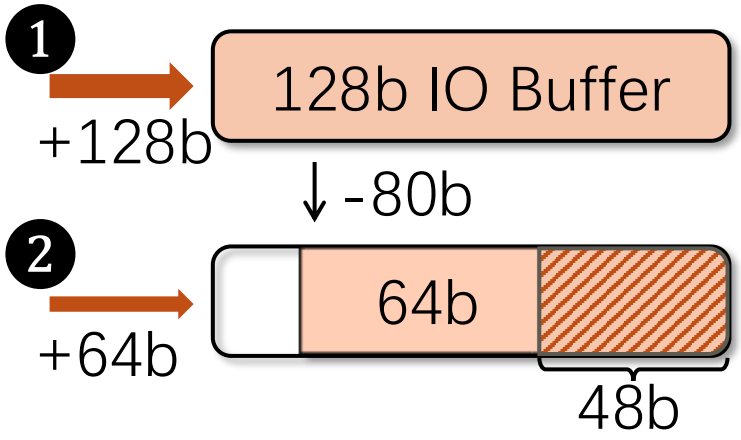


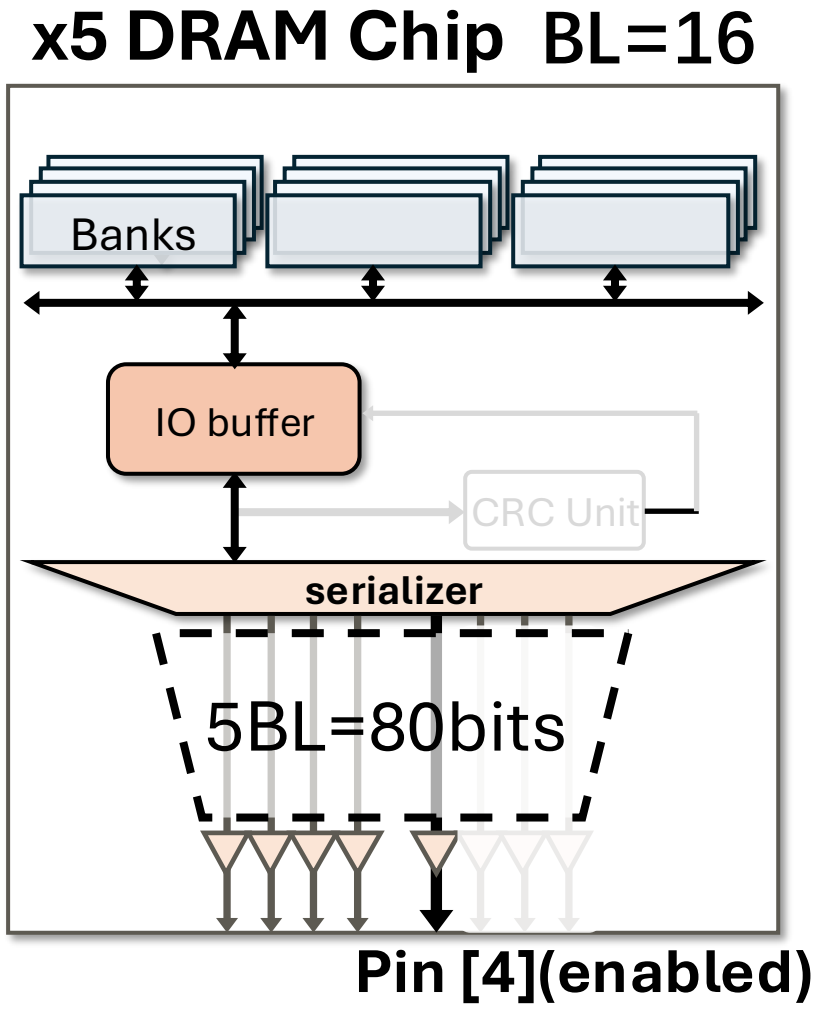
ASPA -- Adaptive Buffering



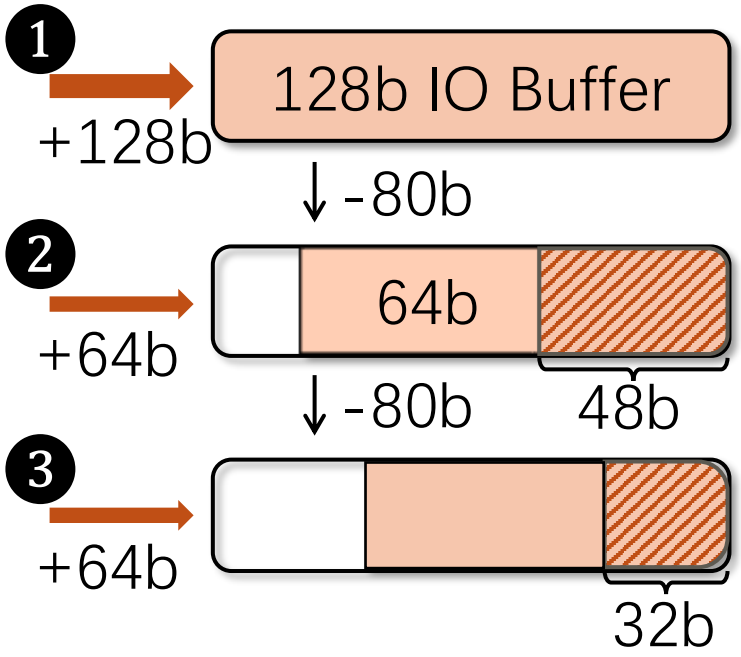


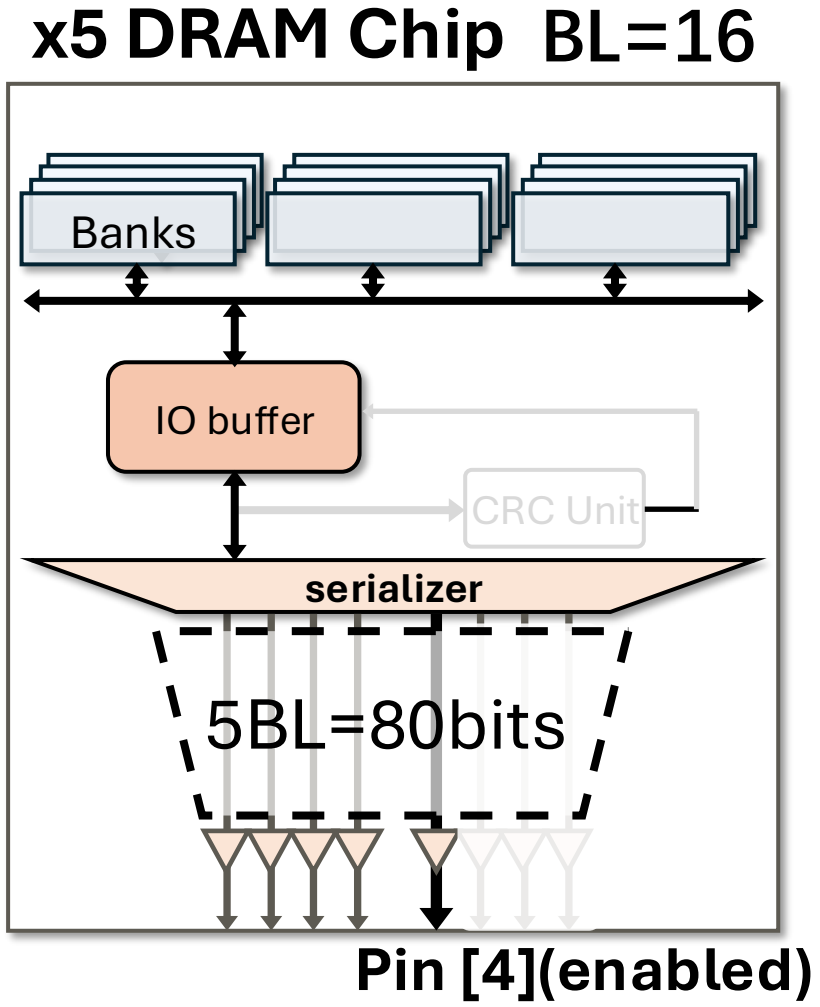
ASPA -- Adaptive Buffering



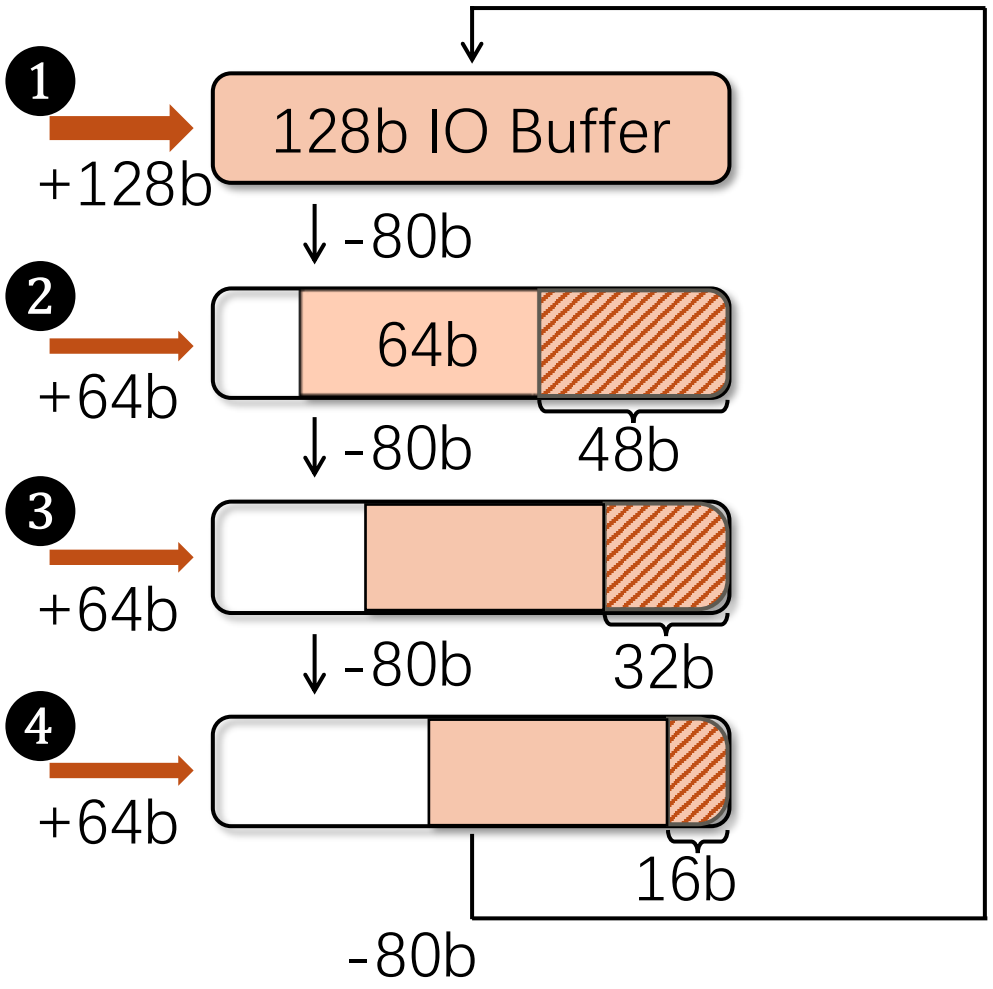


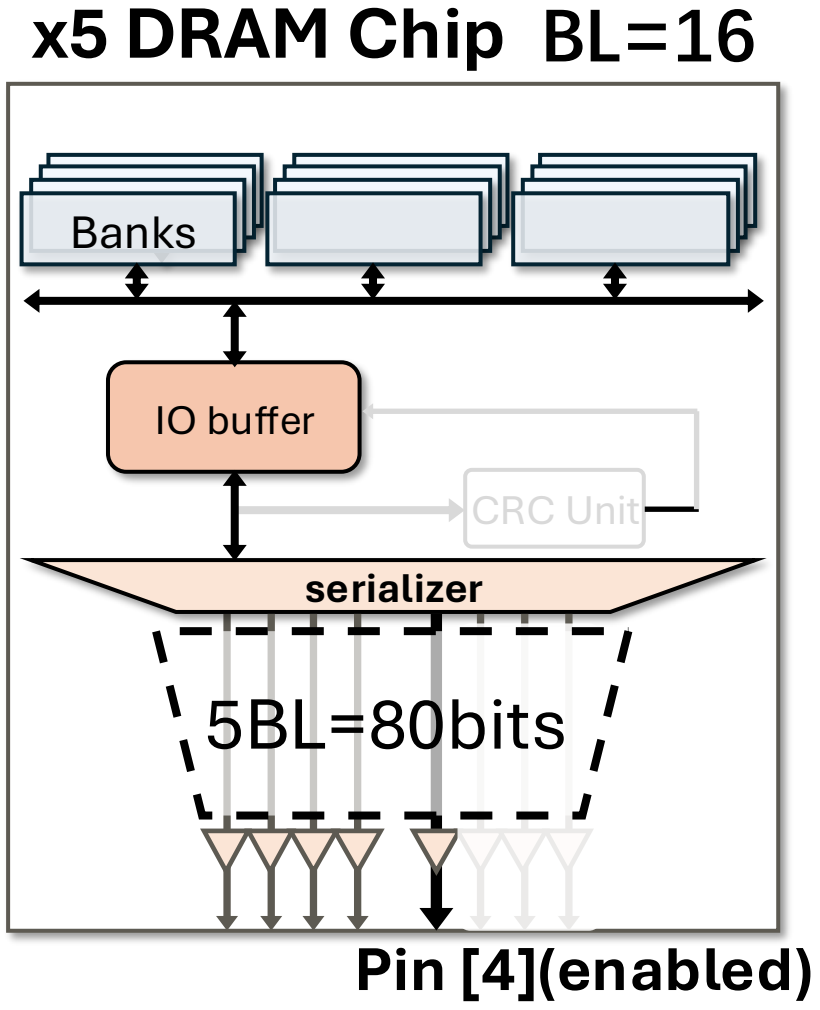
ASPA -- Adaptive Buffering



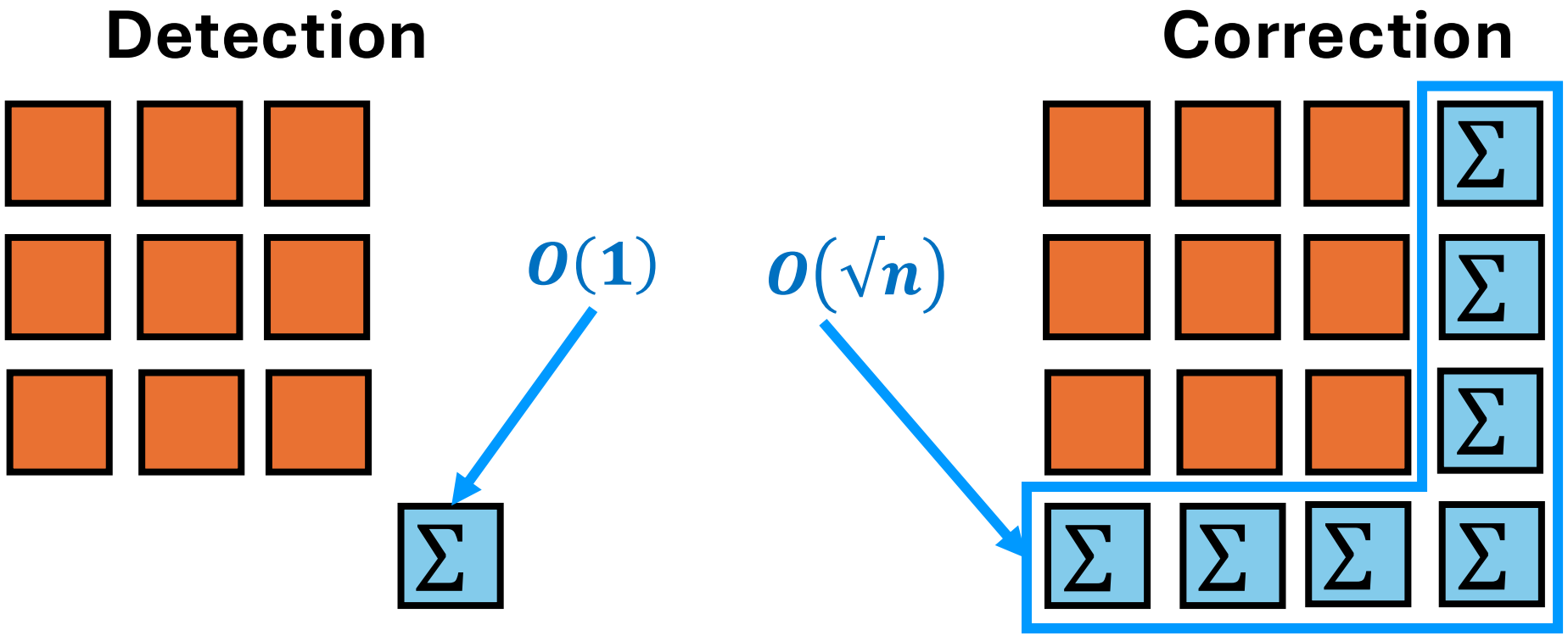


ASPA -- Adaptive Buffering

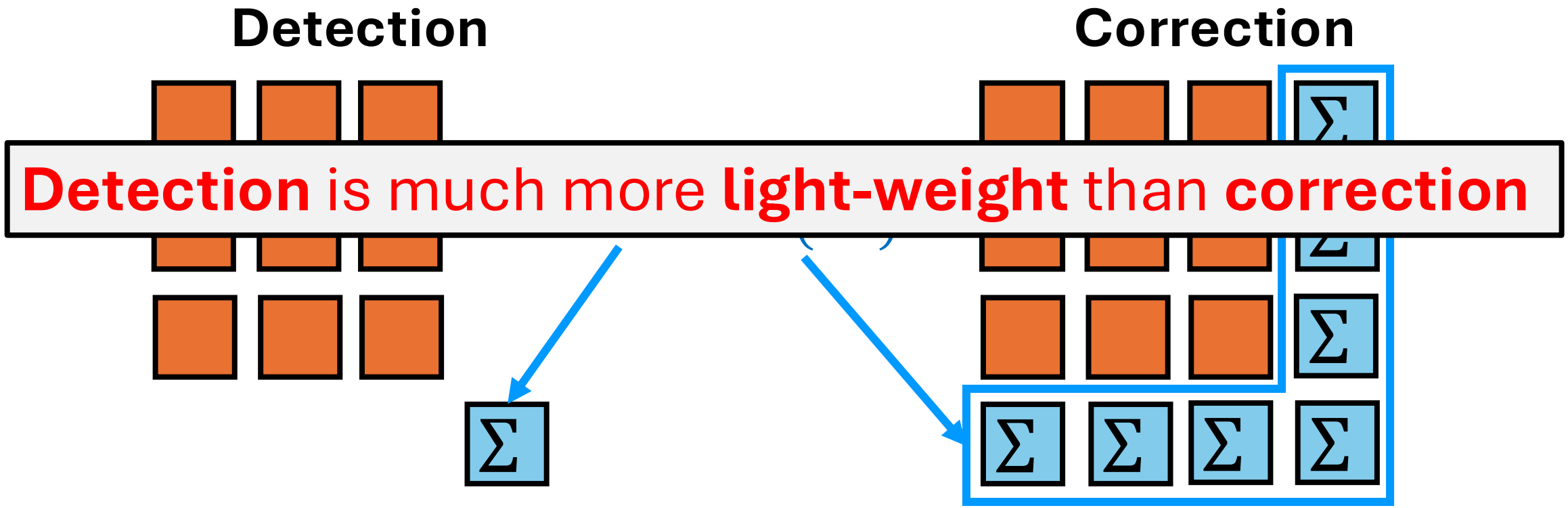




Challenge II: Reliability



Challenge II: Reliability

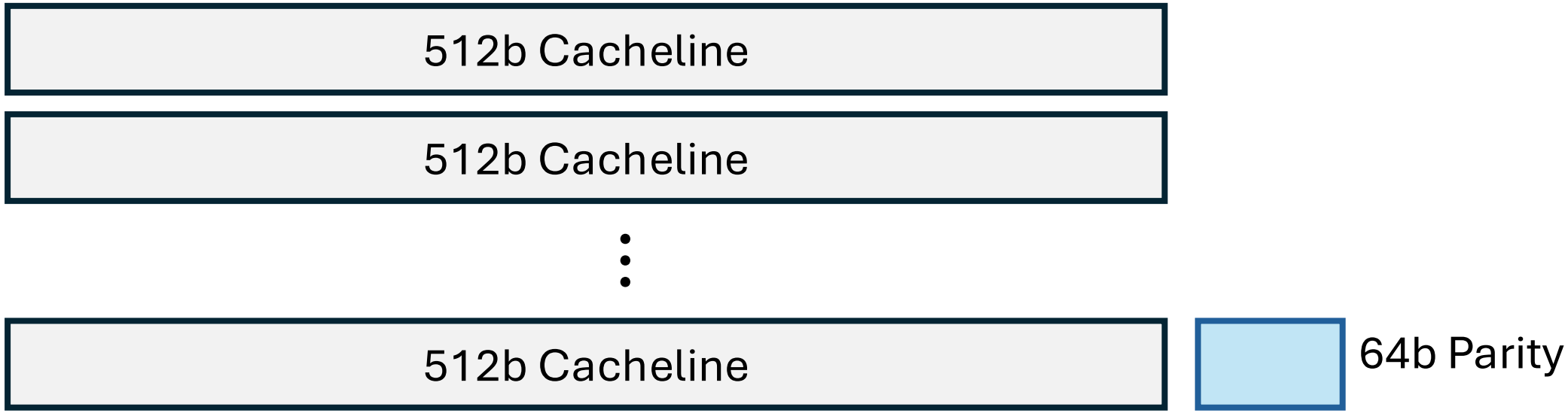


Detection Capability

Collision rate: 2^{-32}, **sufficient** for error detection

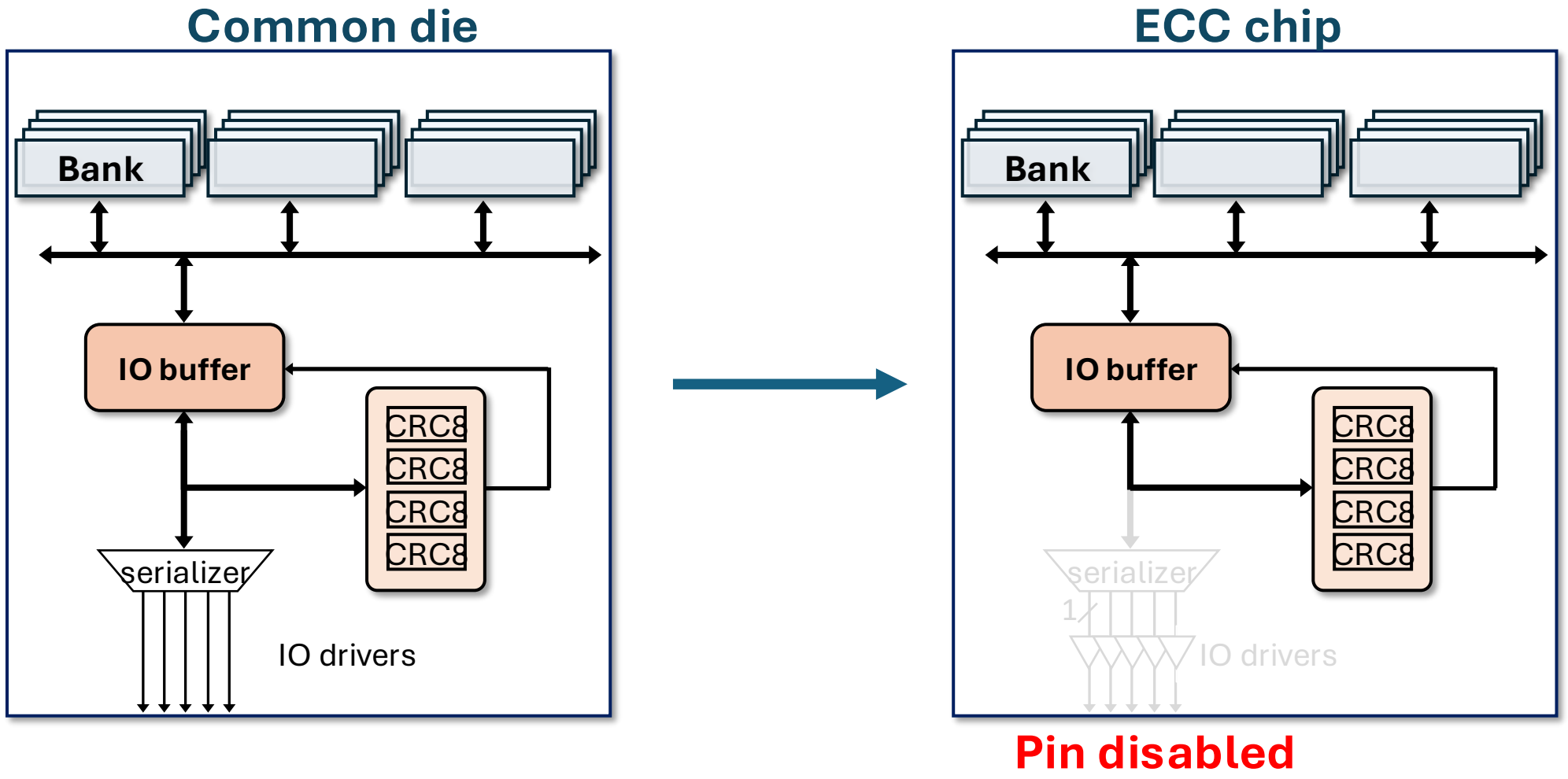
Detection Capability

Collision rate: 2^{-64}, **exponential sufficient** for error detection



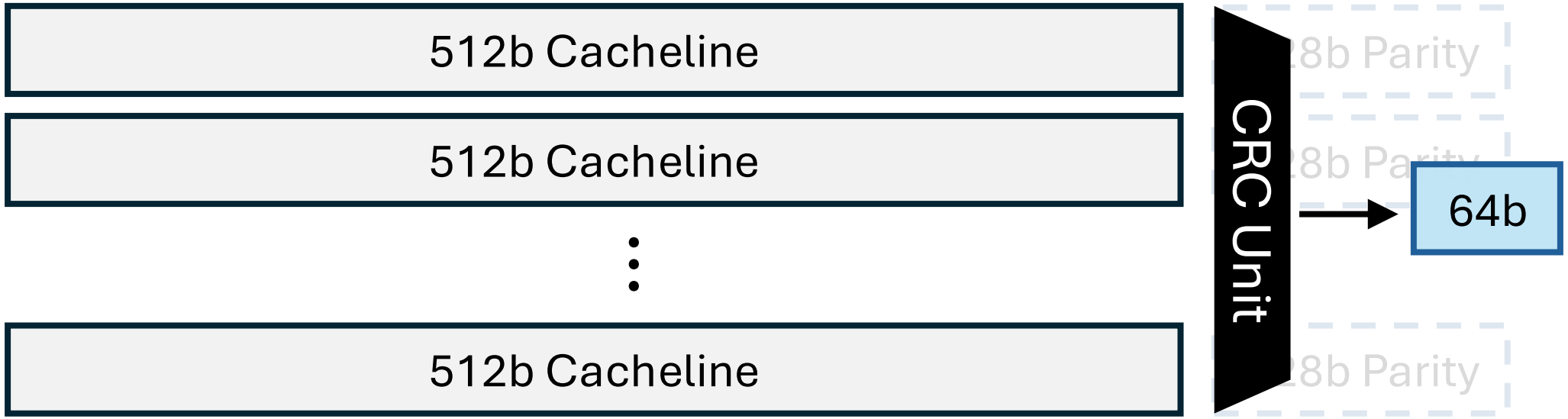
Collision rate: 2^{-64} , **exponential sufficient** for error detection

Detection Capability



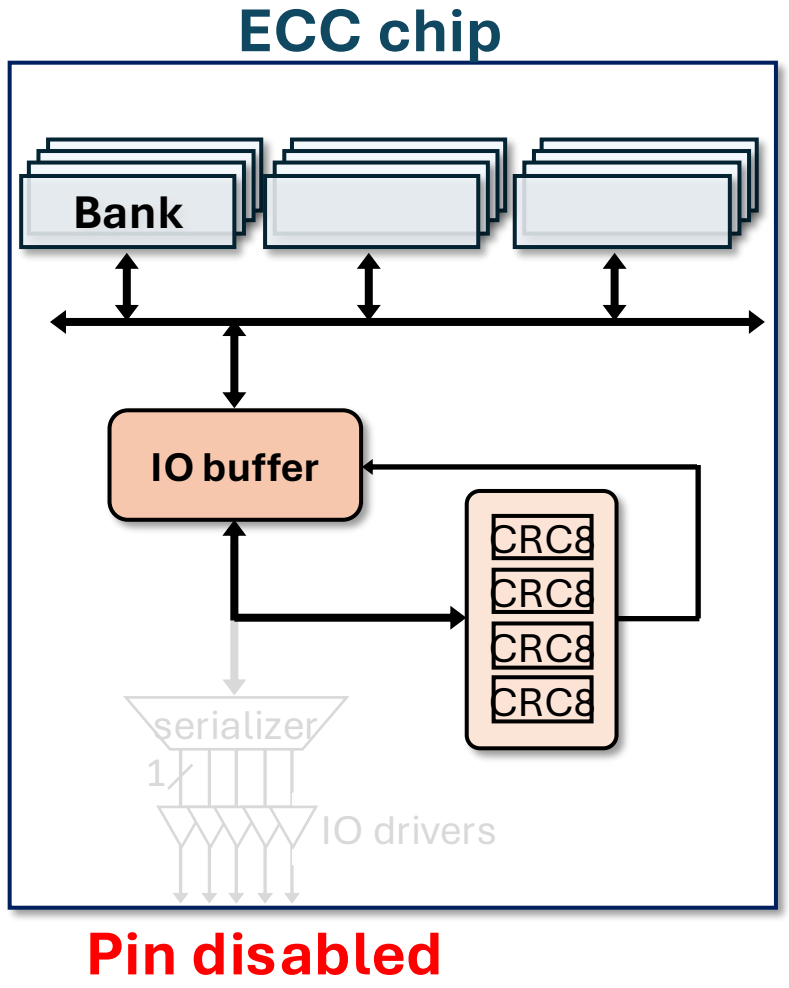
Detection Capability

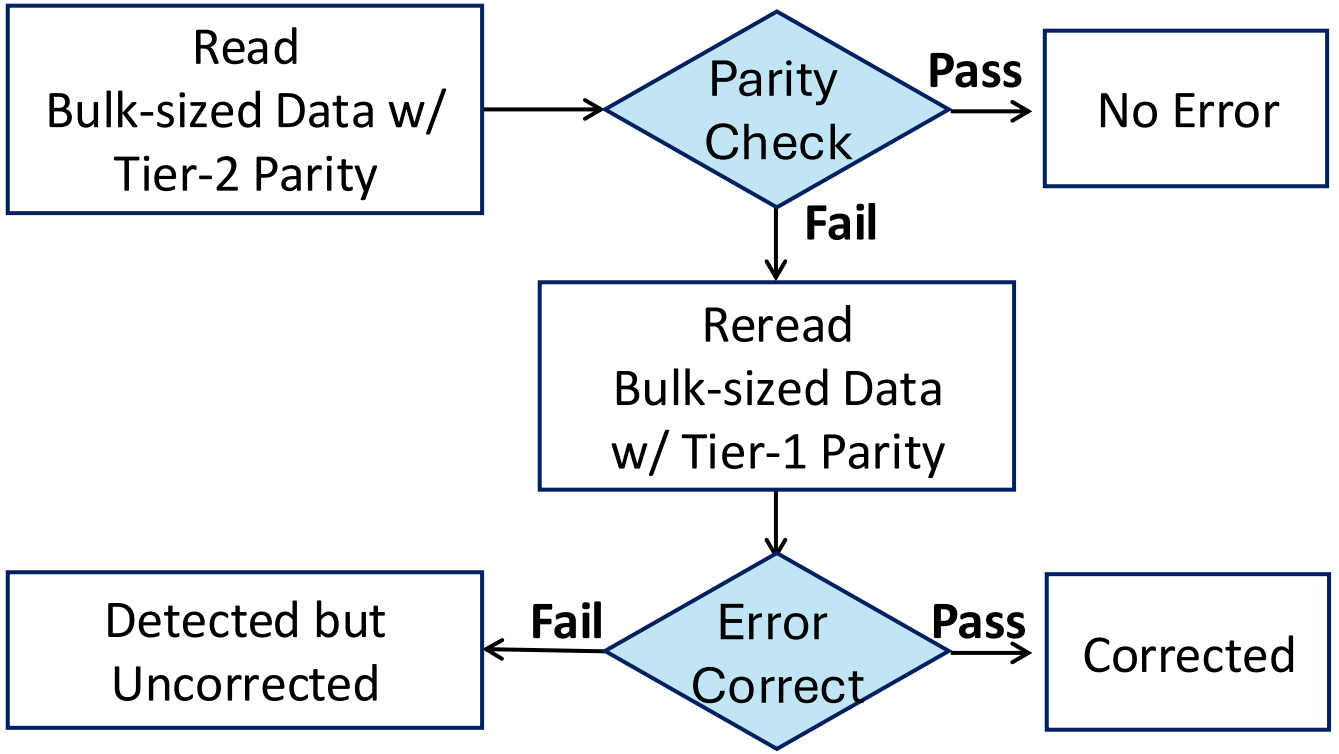
Collusion probability= 2^{-128}>> 512 single bit flip cases



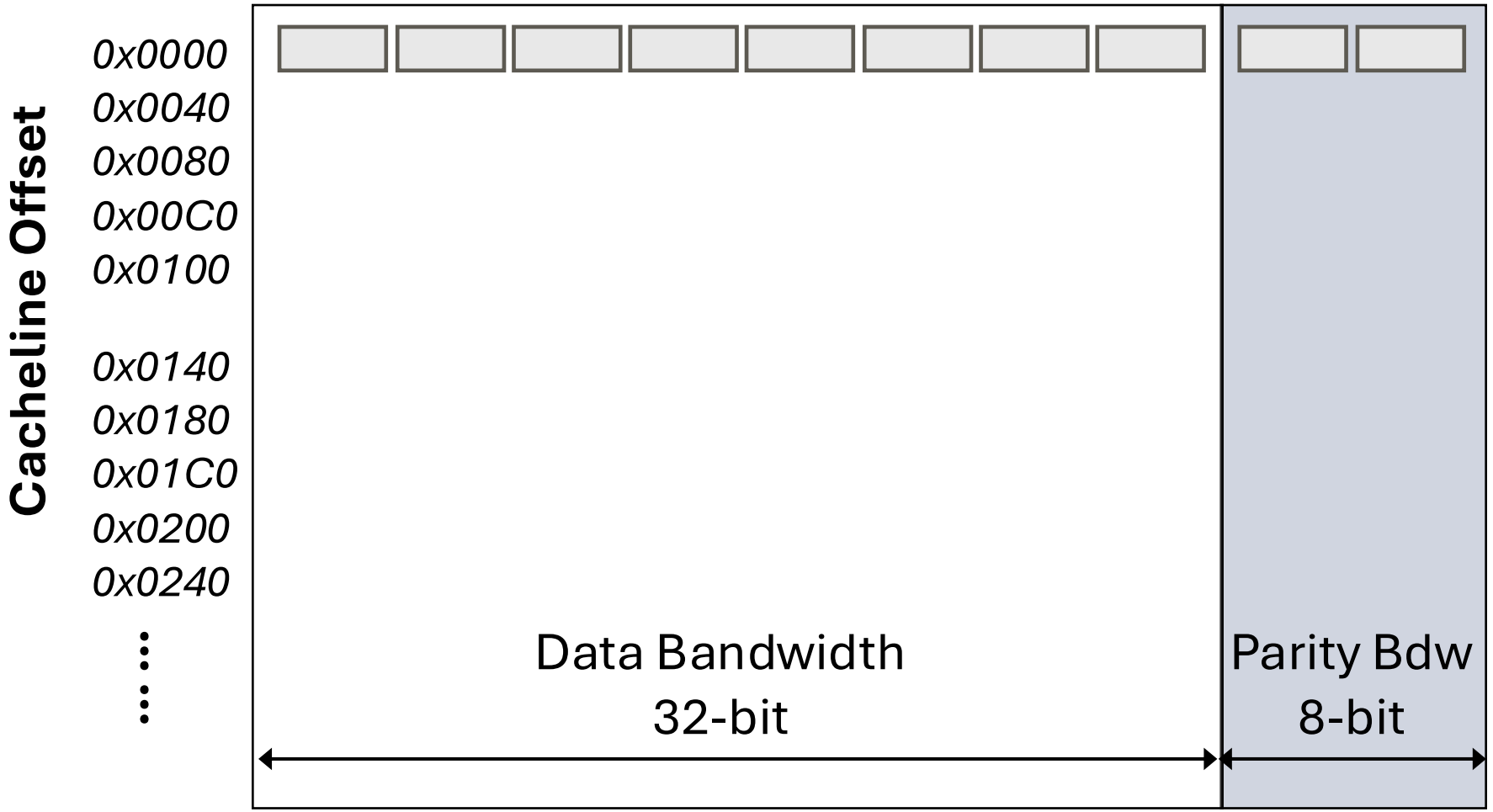
Collusion probability= 2^{-64}>> 512*16 single bit flip cases

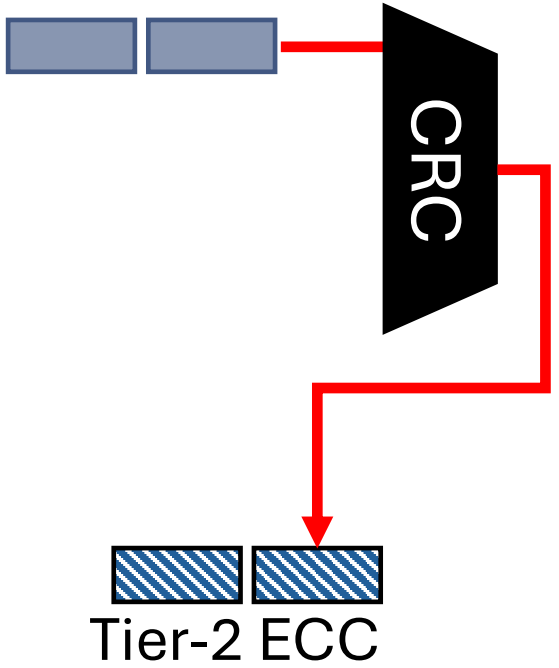
Error Detection & Correction Logic



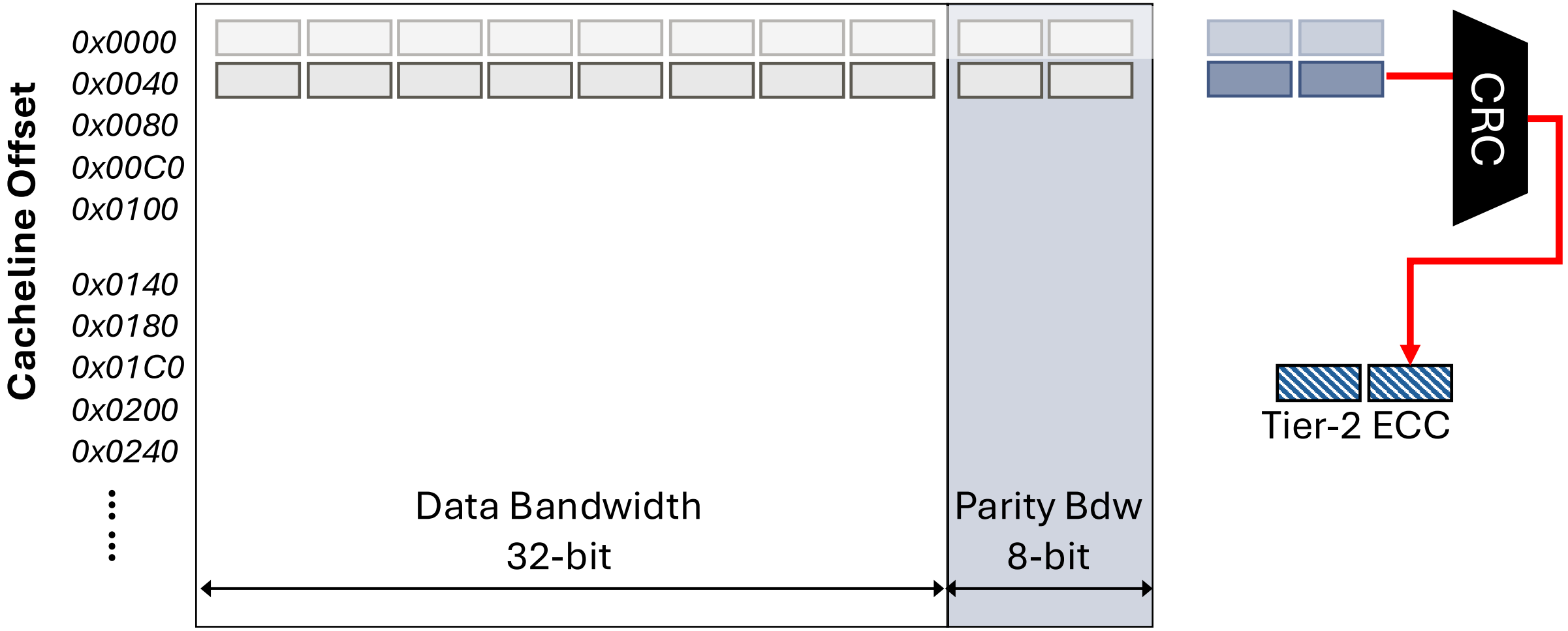


Workflow

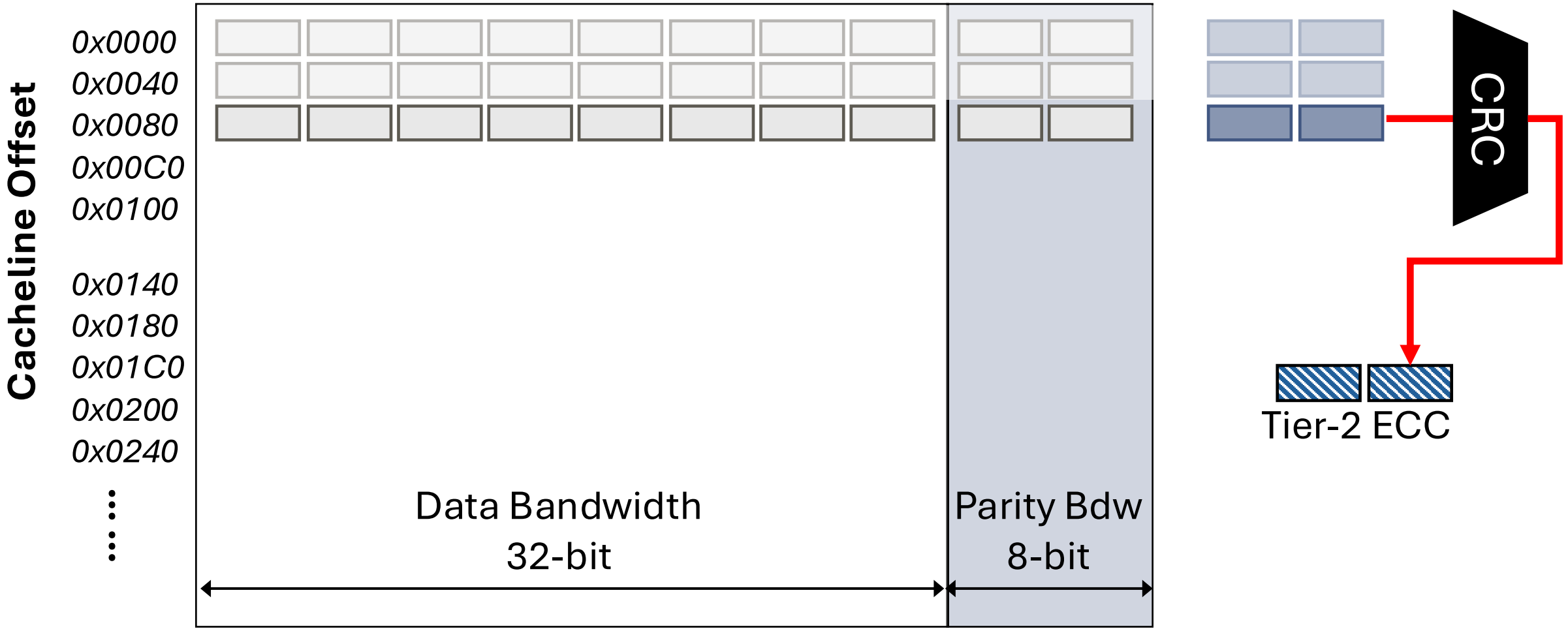

Cacheline Offset
0x0000
0x0040
0x0080
0x00C0
0x0100
0x0140
0x0180
0x01C0
0x0200
0x0240
Data Bandwidth
32-bit
Parity Bdw
8-bit


CRC
Tier-2 ECC

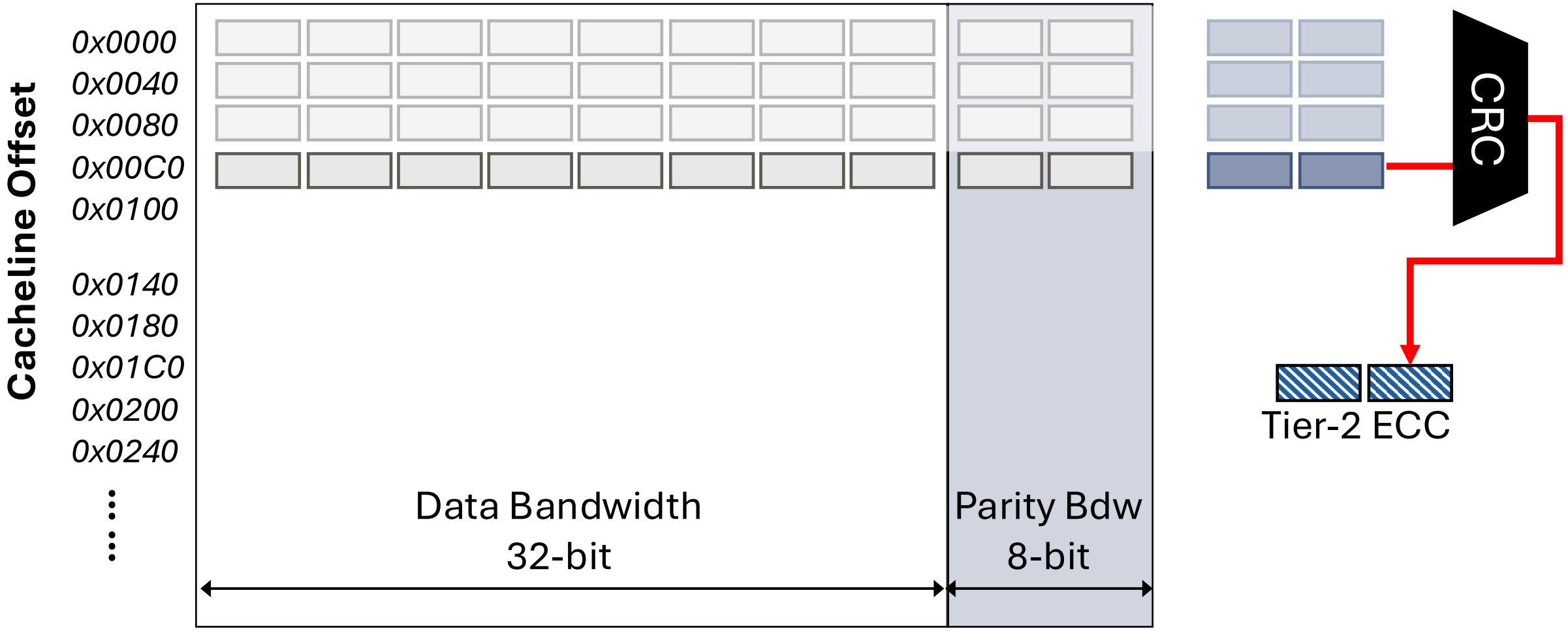
Workflow


Cacheline Offset
0x0000
0x0040
0x0080
0x00C0
0x0100
0x0140
0x0180
0x01C0
0x0200
0x0240
Data Bandwidth
32-bit
Parity Bdw
8-bit
CRC
Tier-2 ECC

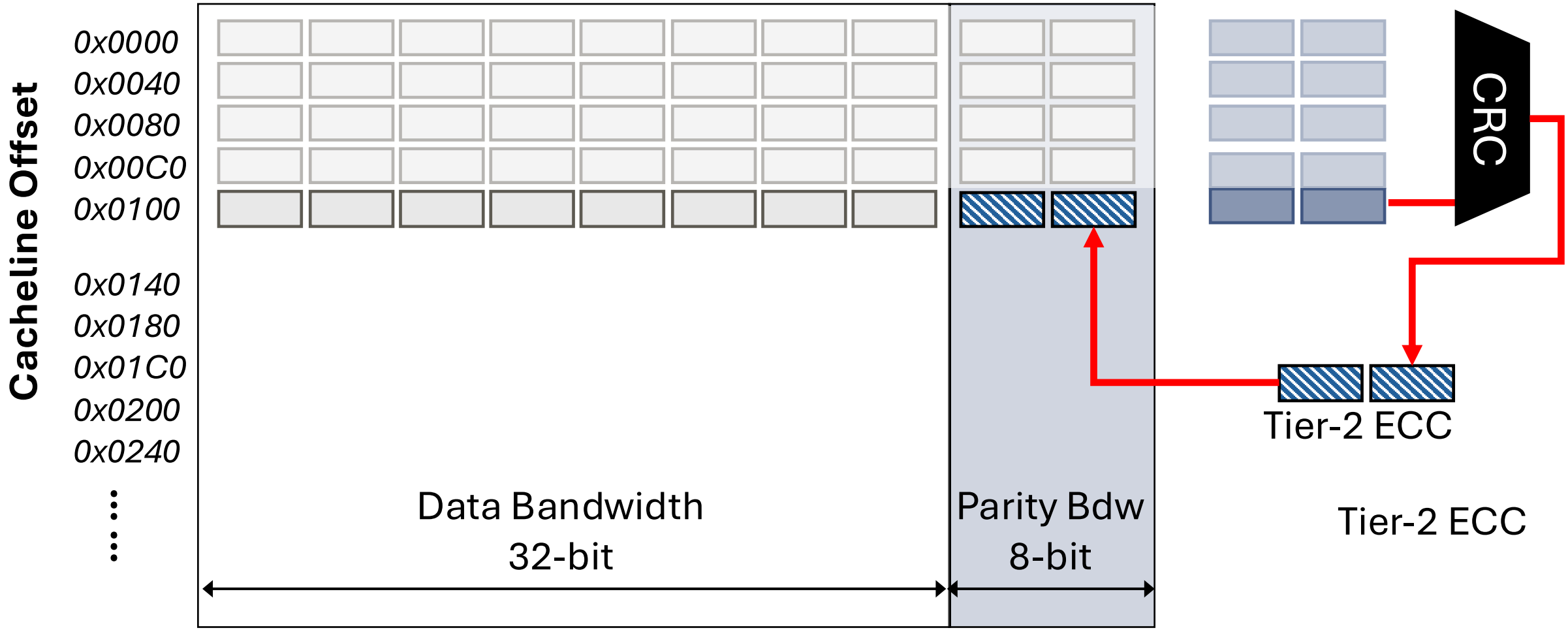
Workflow


Cacheline Offset
0x0000
0x0040
0x0080
0x00C0
0x0100
0x0140
0x0180
0x01C0
0x0200
0x0240
Data Bandwidth
32-bit
Parity Bdw
8-bit
CRC
Tier-2 ECC

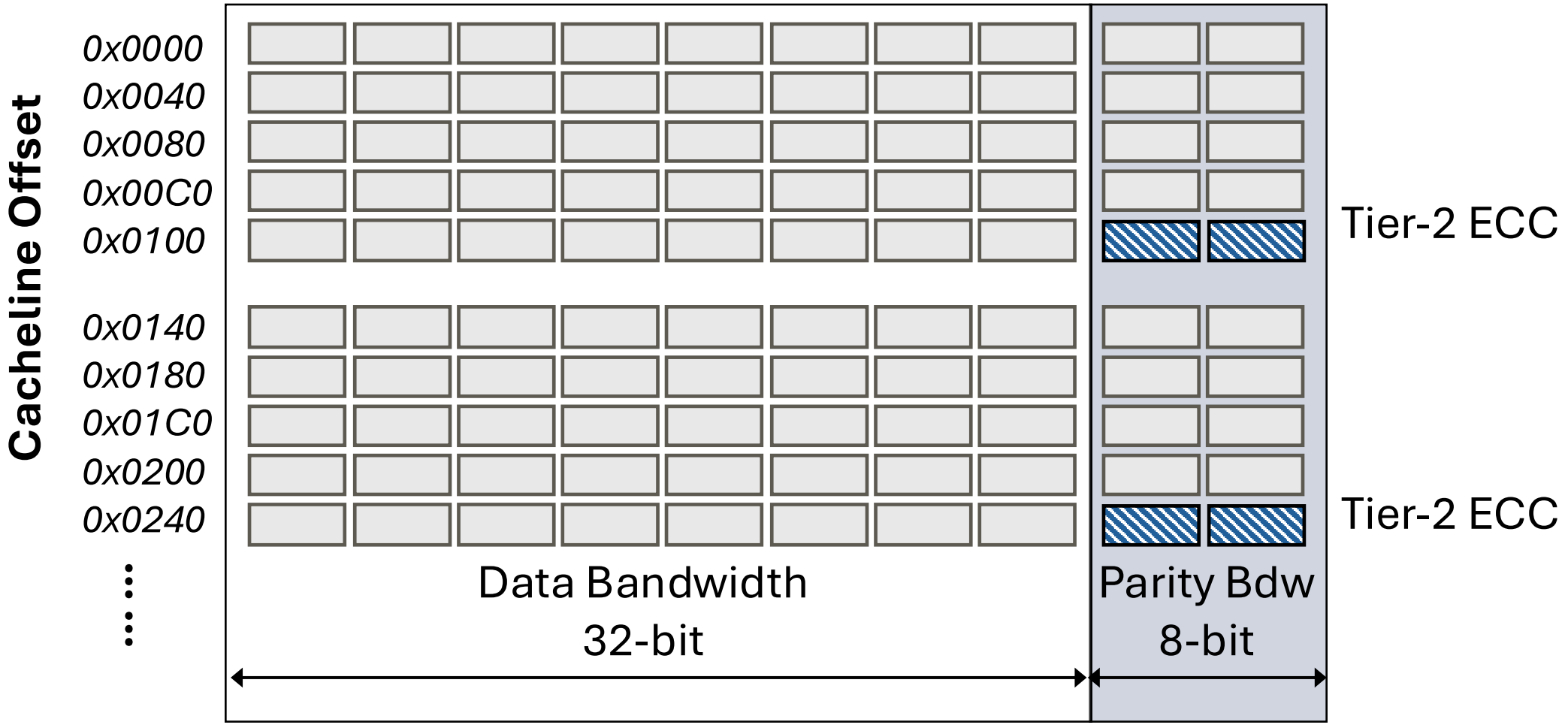
Workflow


Cacheline Offset
0x0000
0x0040
0x0080
0x00C0
0x0100
0x0140
0x0180
0x01C0
0x0200
0x0240
Data Bandwidth
32-bit
Parity Bdw
8-bit
CRC
Tier-2 ECC

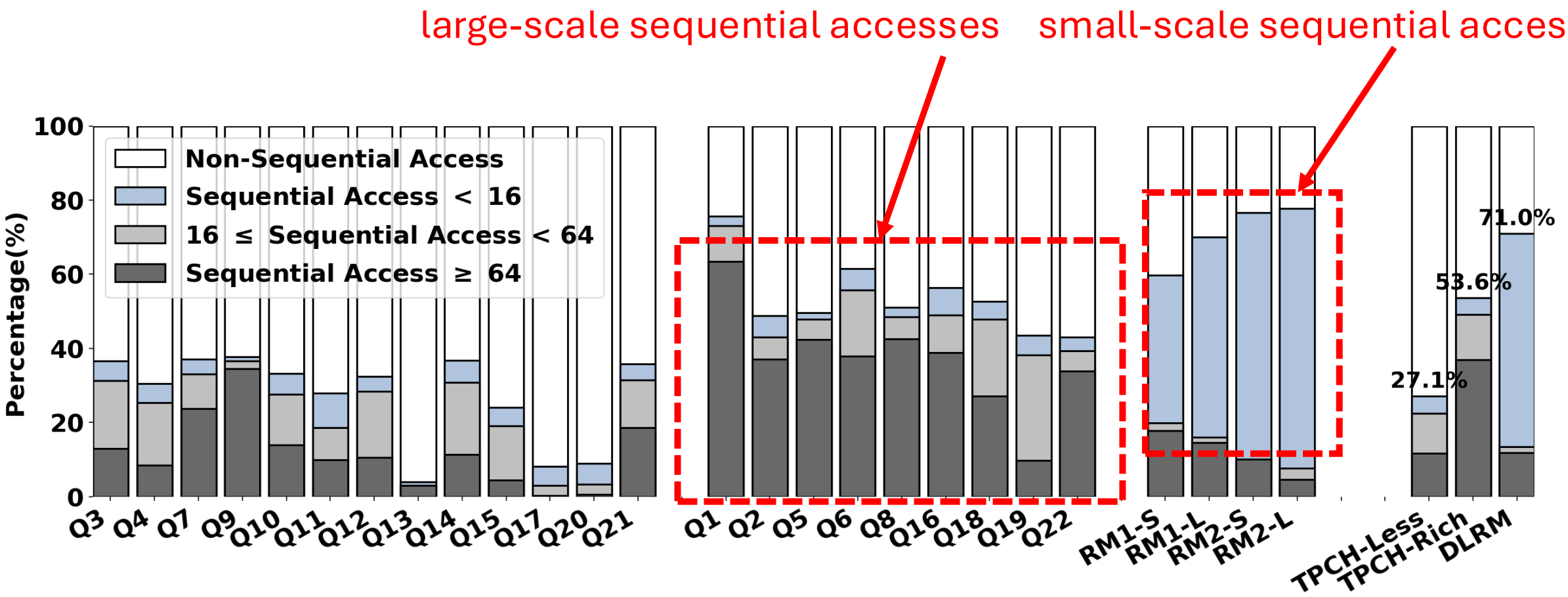
Workflow


Cacheline Offset
0x0000
0x0040
0x0080
0x00C0
0x0100
0x0140
0x0180
0x01C0
0x0200
0x0240
CRC
Tier-2 ECC
Data Bandwidth
32-bit
Parity Bdw
8-bit
Tier-2 ECC

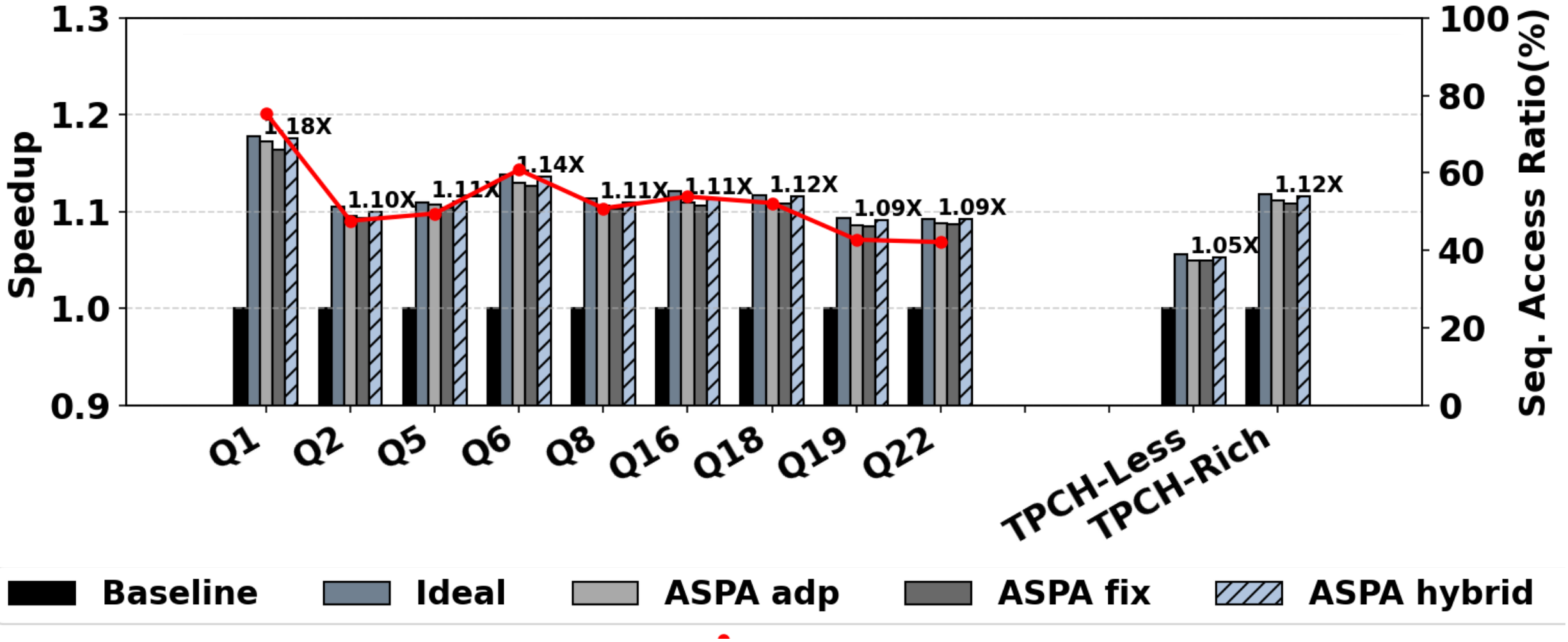
Workflow


Cacheline Offset
0x0000
0x0040
0x0080
0x00C0
0x0100
0x0140
0x0180
0x01C0
0x0200
0x0240
Tier-2 ECC
Tier-2 ECC
Data Bandwidth
32-bit
Parity Bdw
8-bit

Access Pattern of Workloads



Performance



Thank you for listening

Q&A: fan.li@ucf.edu