

Processing in Storage Class Memory

Joel Nider

Andrada Zoltan

Craig Mustard

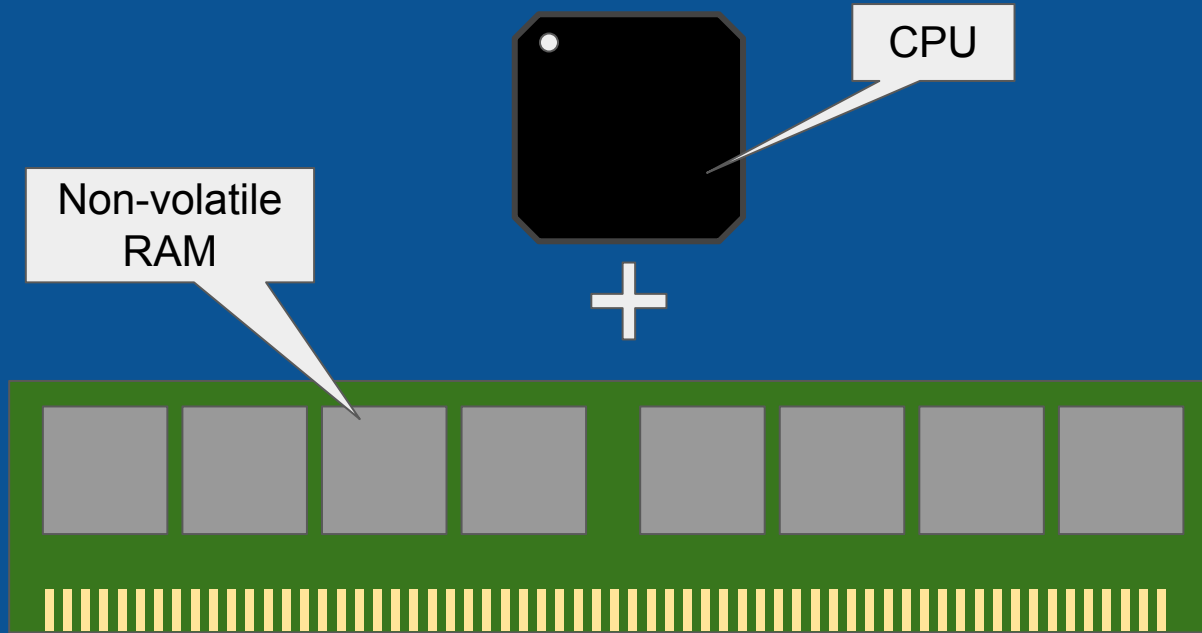
Alexandra Fedorova



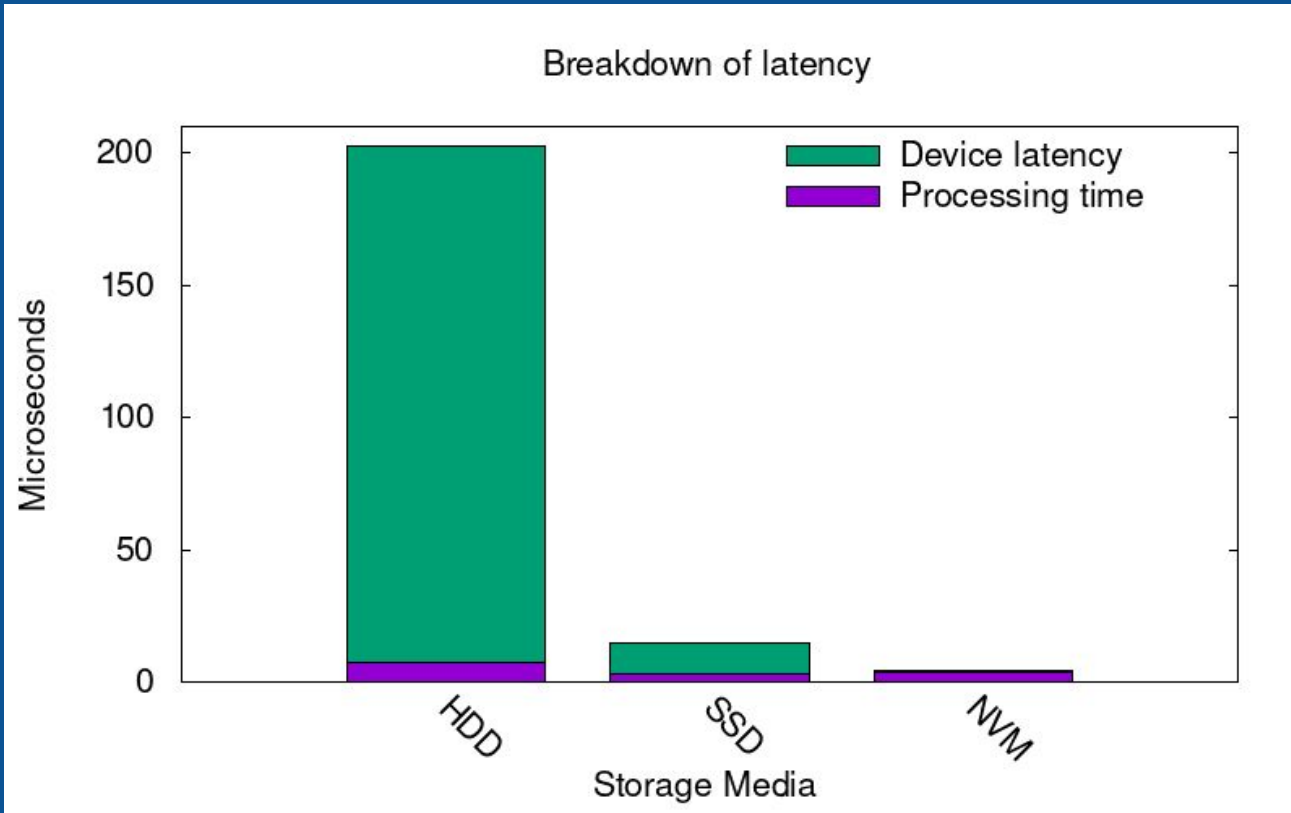
THE UNIVERSITY
OF BRITISH COLUMBIA

Electrical and Computer
Engineering

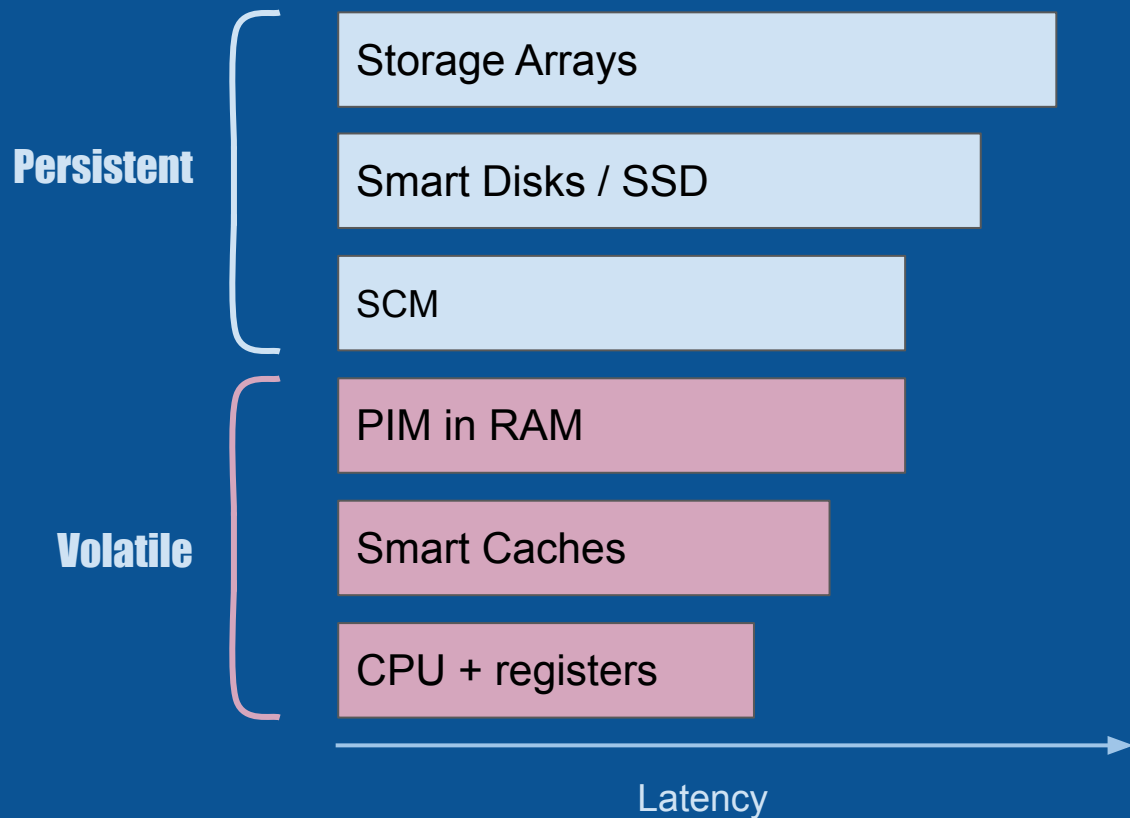
Embedding Processors in SCM



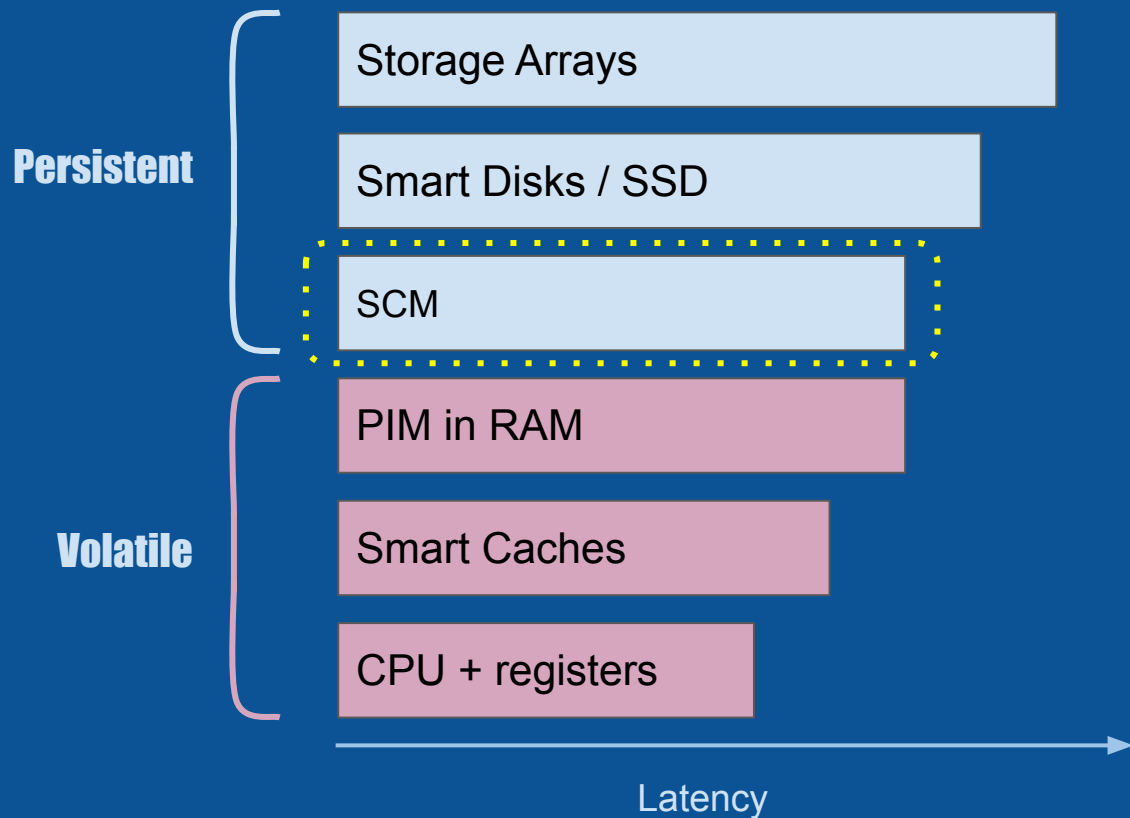
Storage Latency Is Decreasing



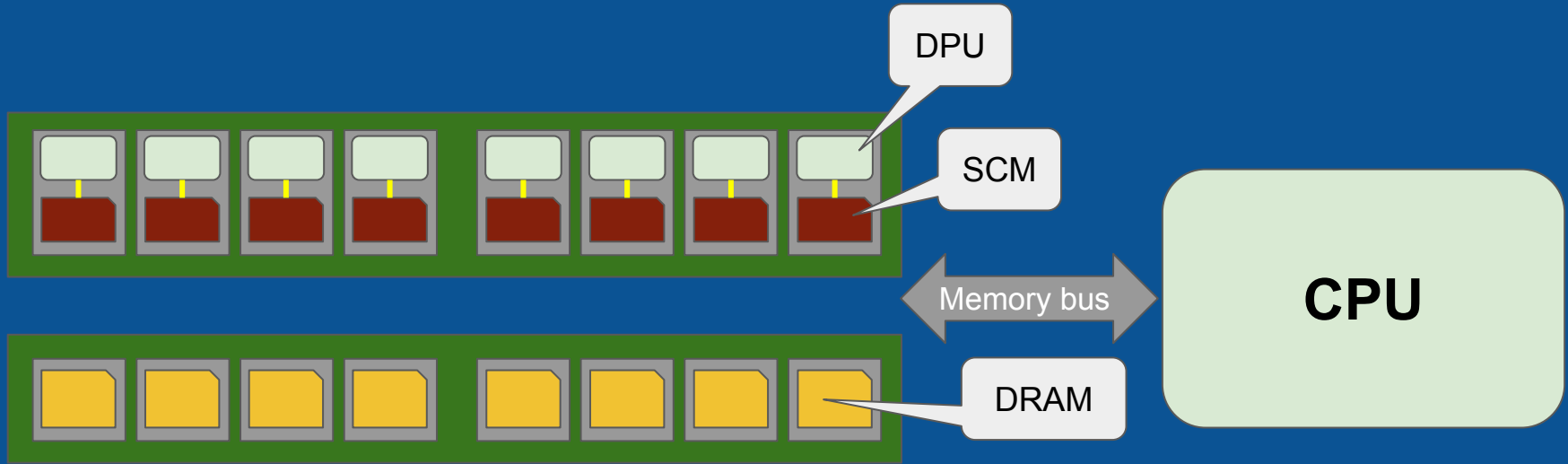
Scaling Compute with Storage



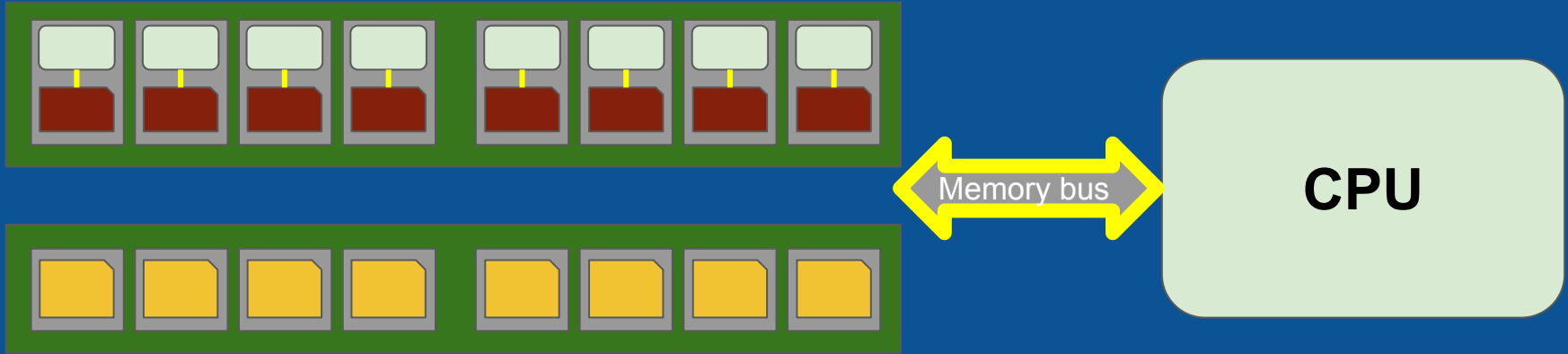
Scaling Compute with Storage



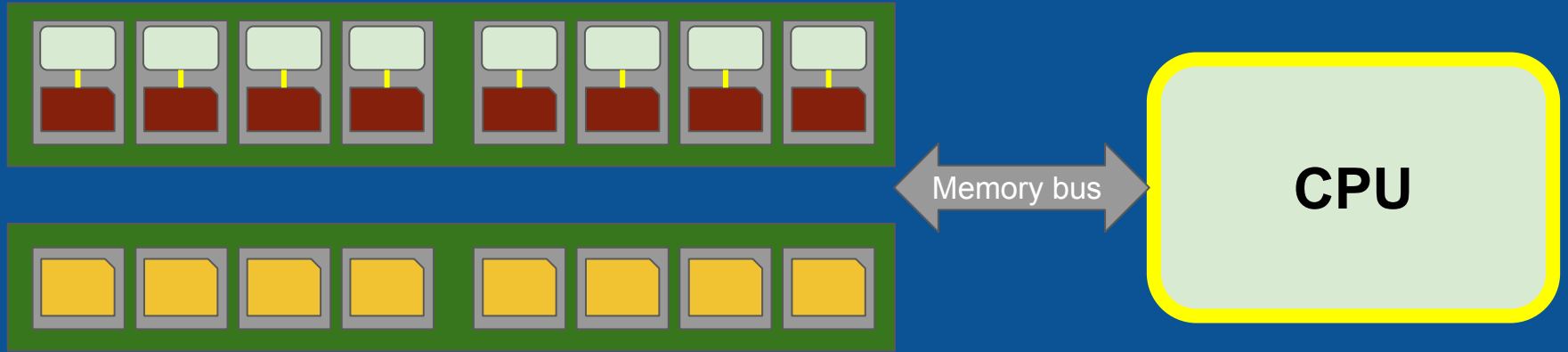
Benefits of PIM on SCM



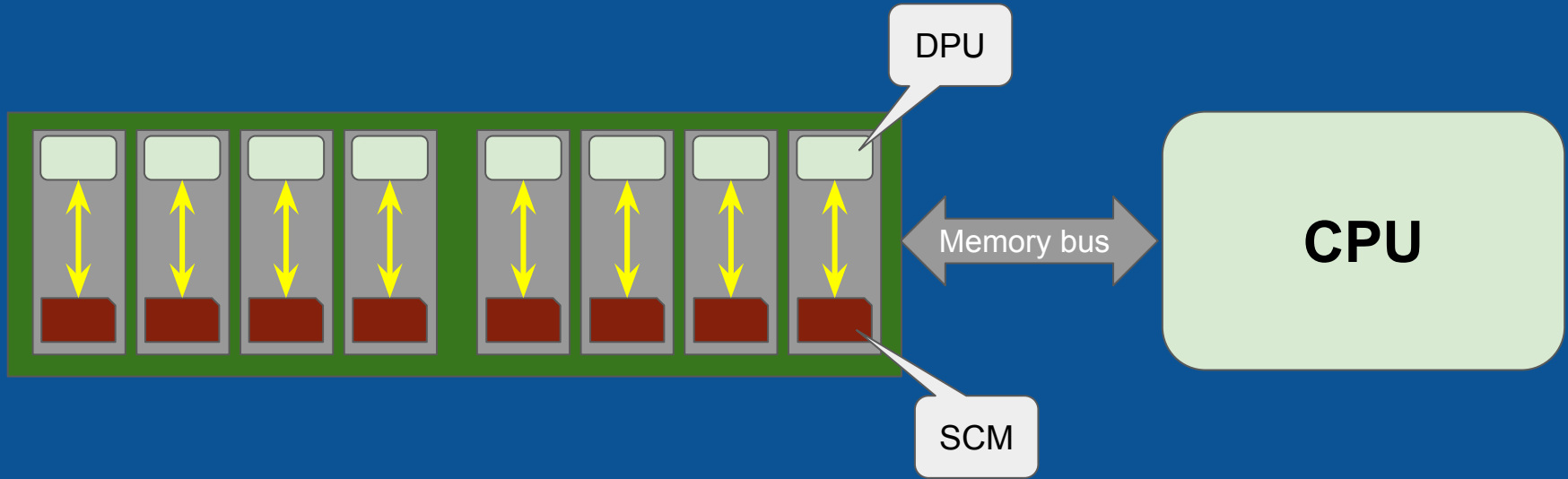
Benefits of PIM on SCM



Benefits of PIM on SCM



Benefits of PIM on SCM



Benefits of PIM on SCM

DPU Count:

64

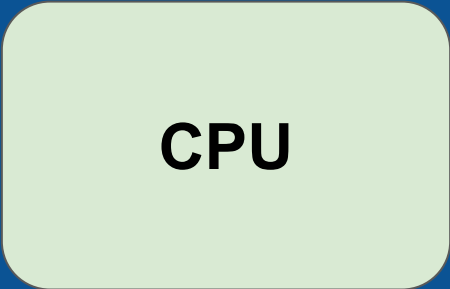
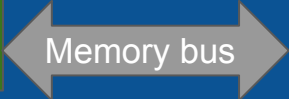
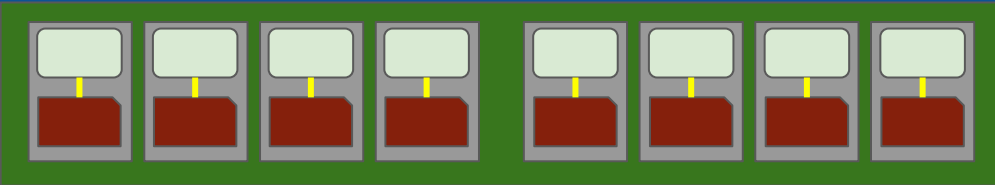
SCM Capacity:

4 GB

Ratio:

1:64 MB

Core Density

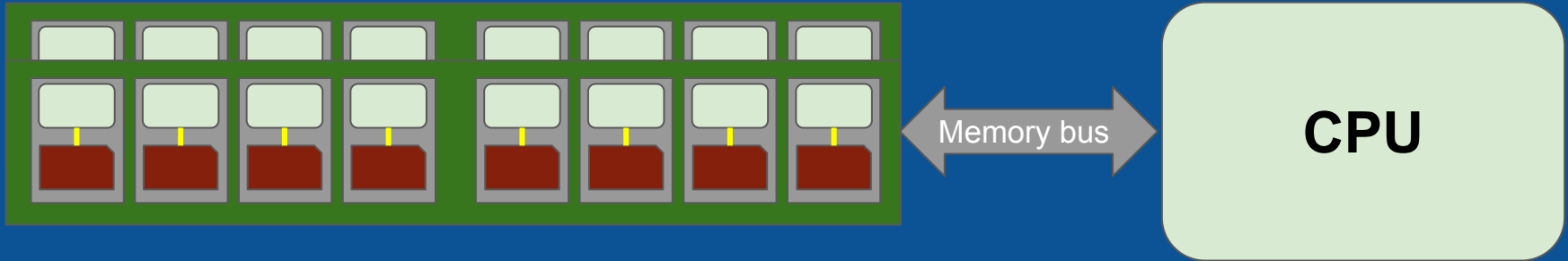


Benefits of PIM on SCM

DPU Count: **128**

SCM Capacity: **8 GB**

Ratio: **1:64 MB**

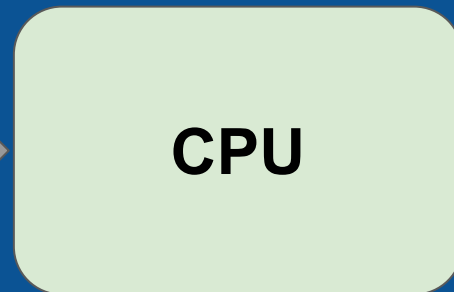
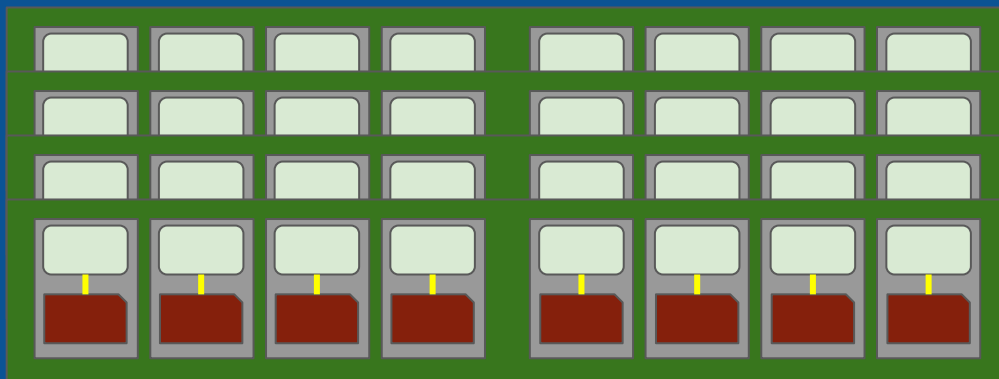


Benefits of PIM on SCM

DPU Count: **256**

SCM Capacity: **16 GB**

Ratio: **1:64 MB**

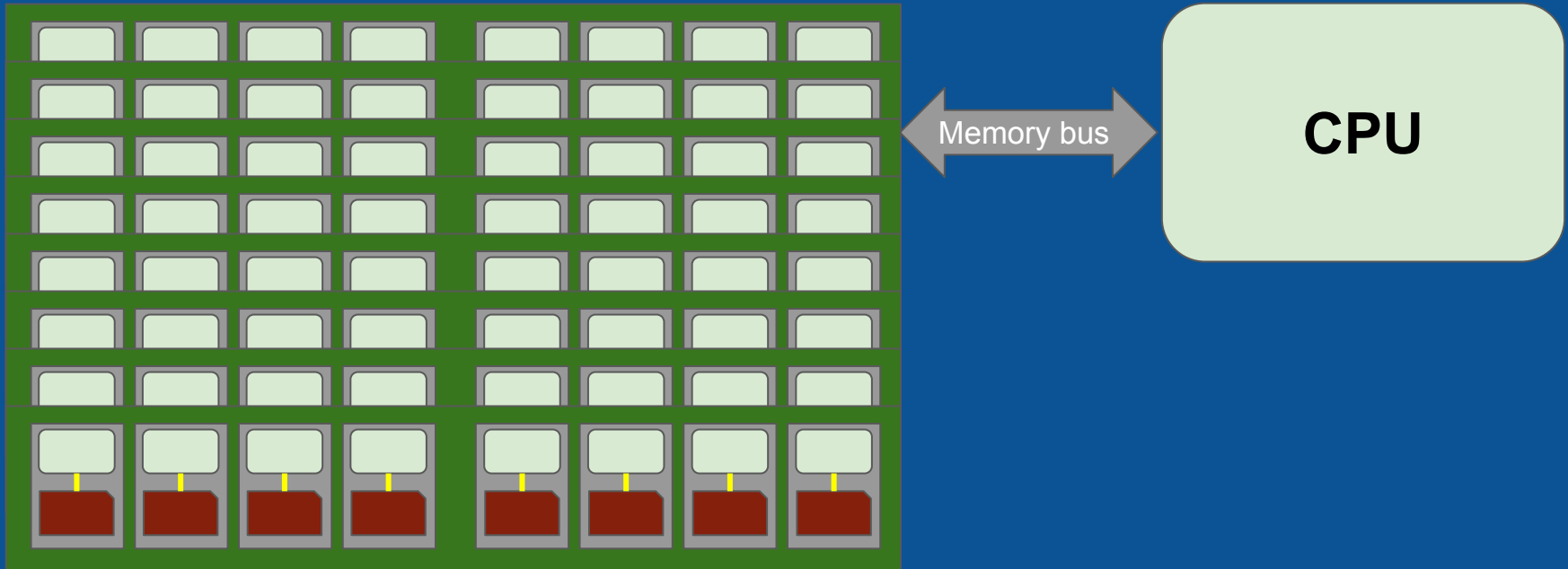


Benefits of PIM on SCM

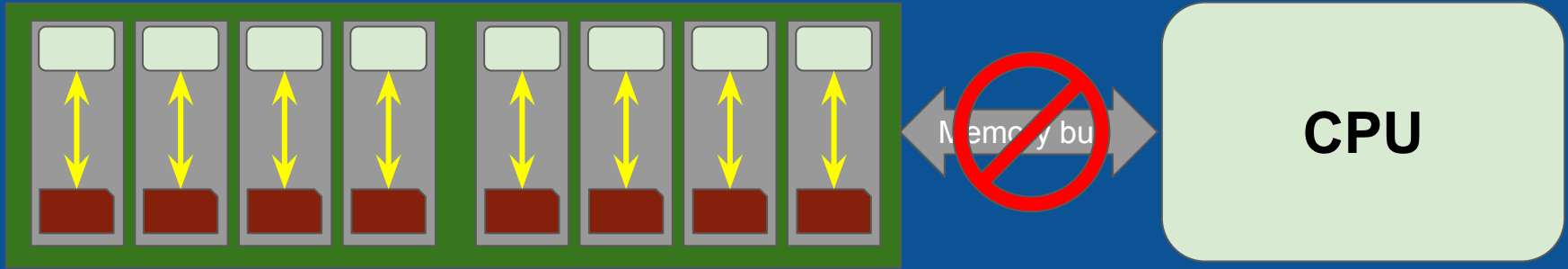
DPU Count: **512**

SCM Capacity: **32 GB**

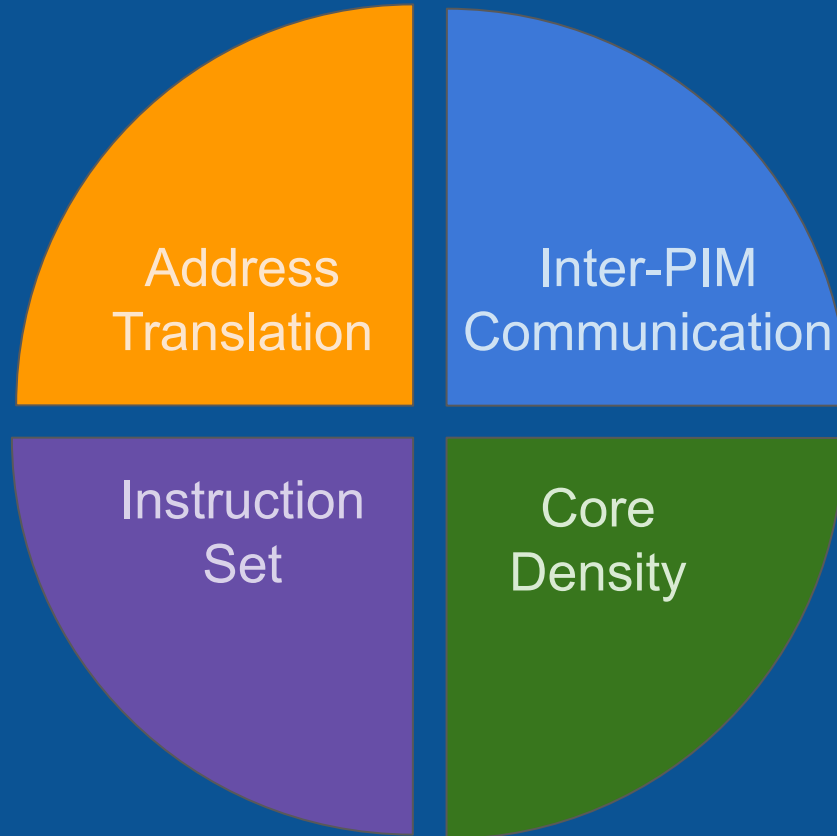
Ratio: **1:64 MB**



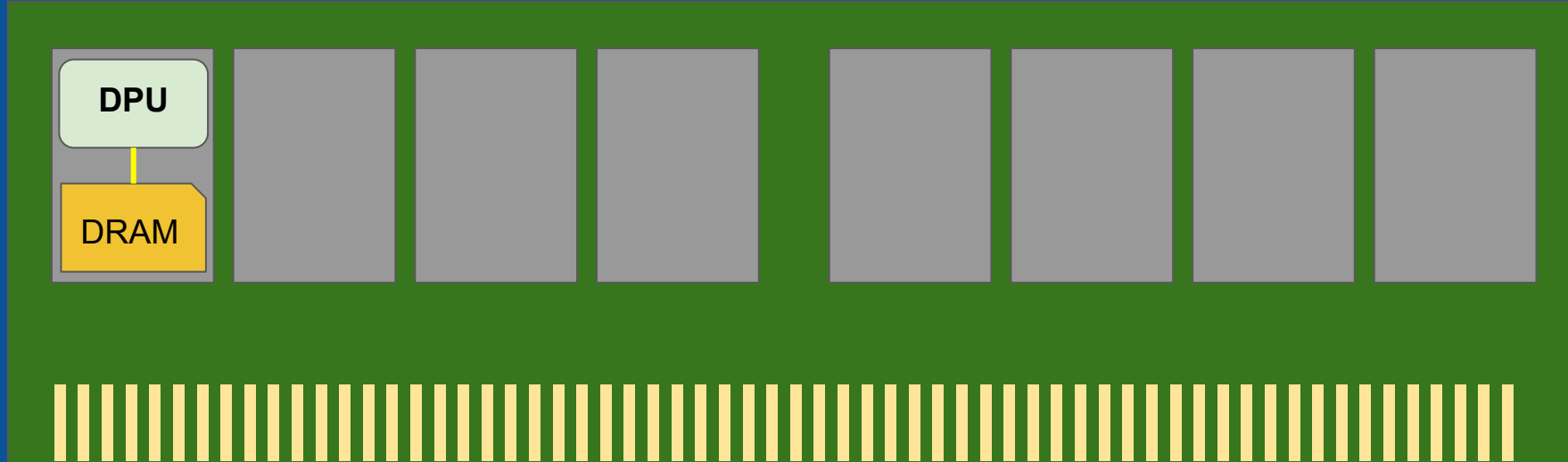
Benefits of PIM on SCM



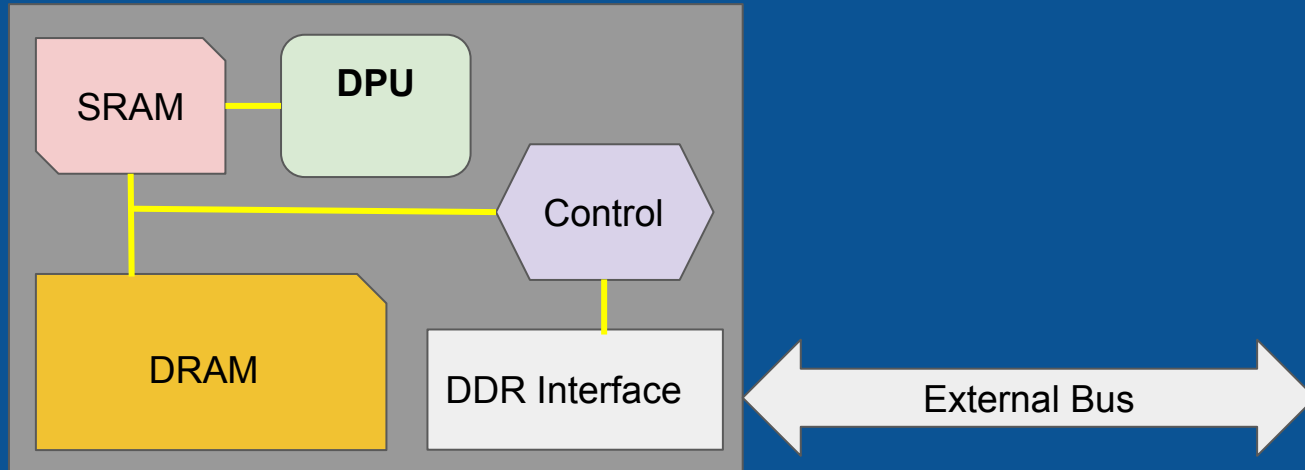
PIM Design Points



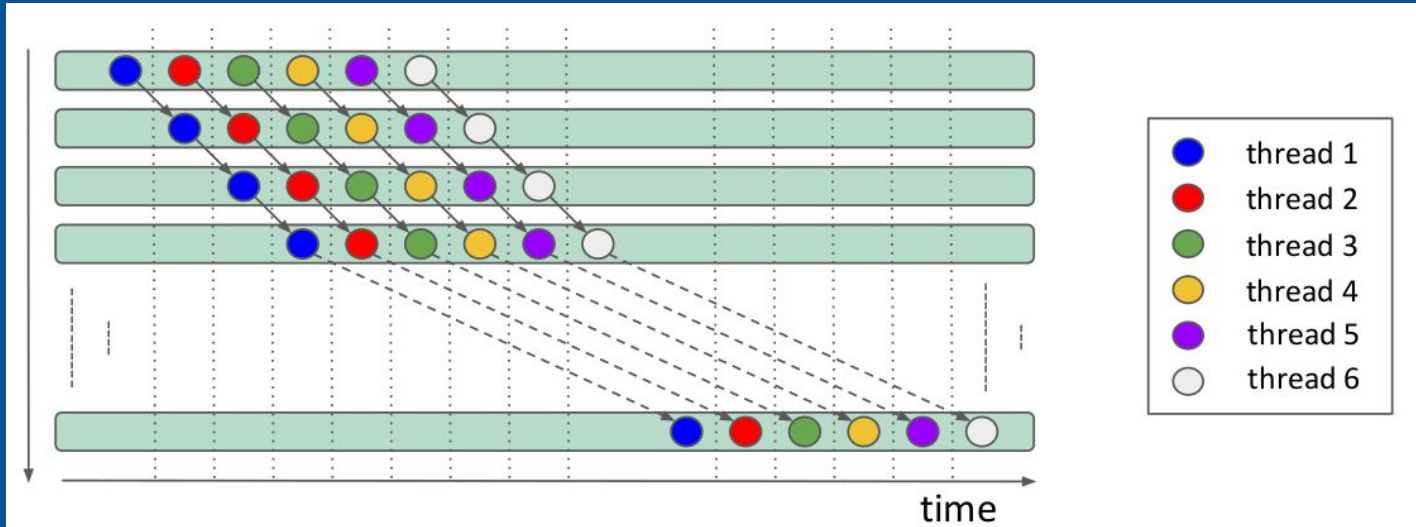
UPMEM Architecture and Limitations



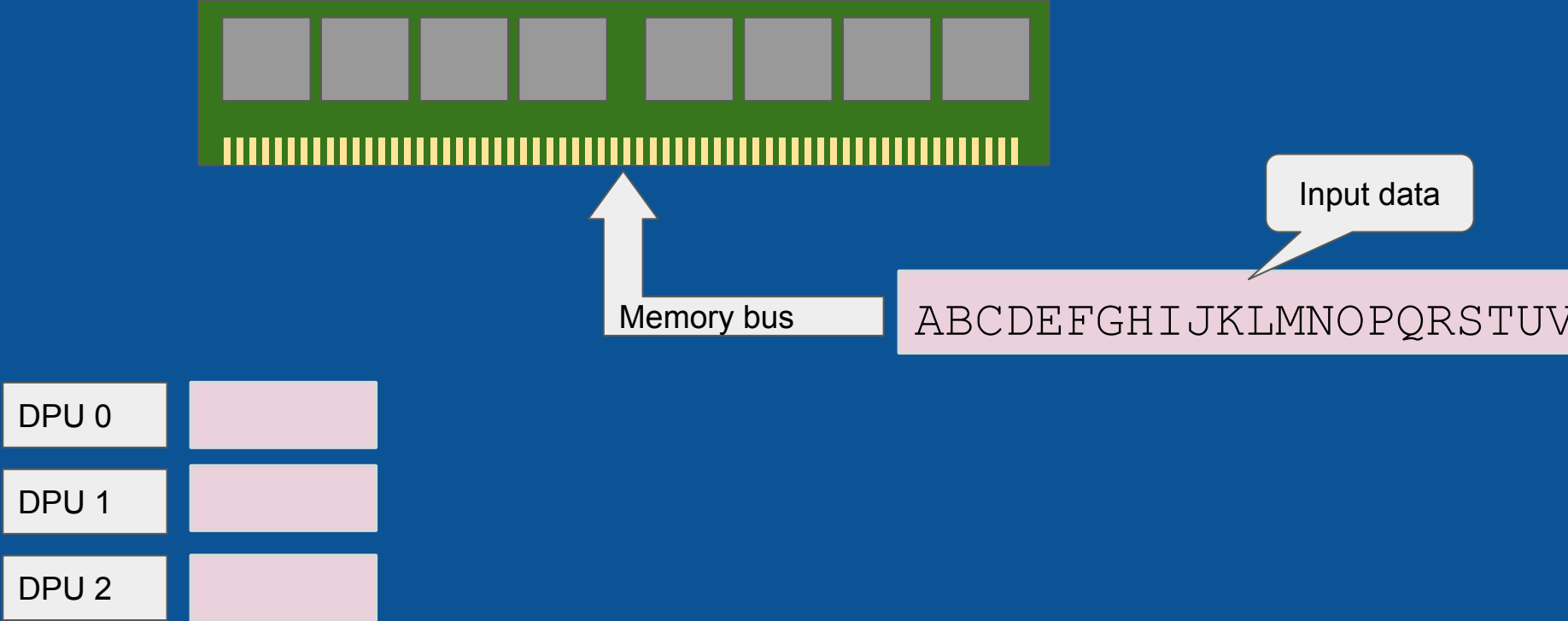
UPMEM Architecture and Limitations



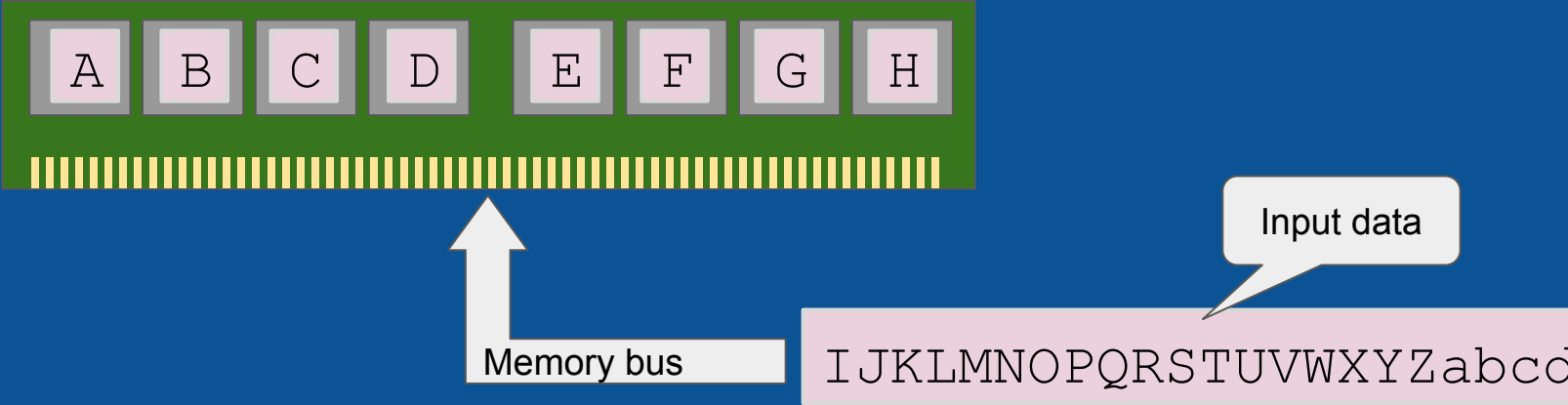
Interleaved Multithreading



UPMEM Architecture and Limitations

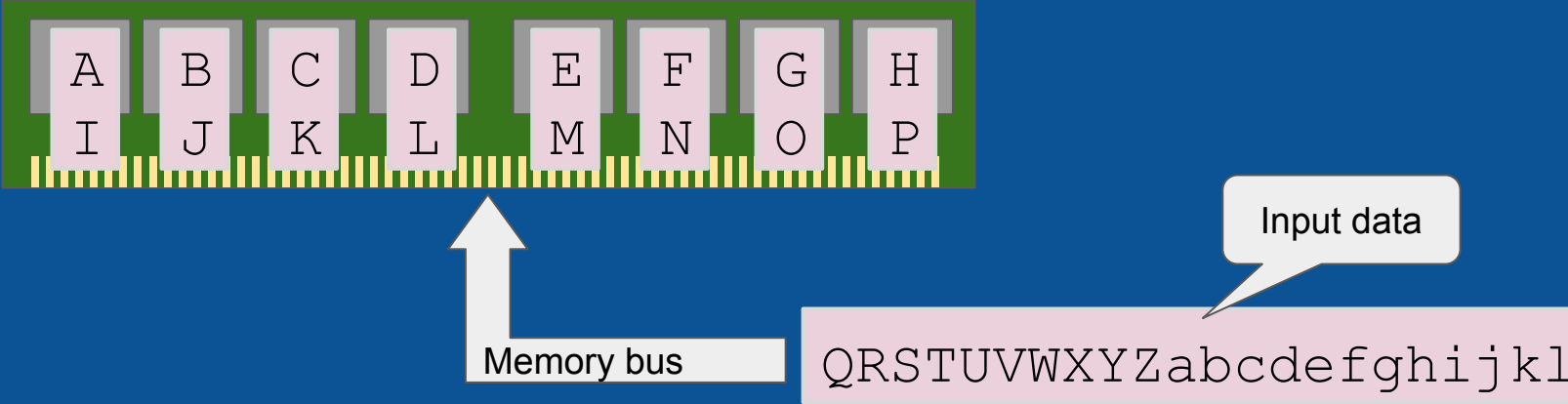


UPMEM Architecture and Limitations



DPU 0	A
DPU 1	B
DPU 2	C

UPMEM Architecture and Limitations



DPU 0	AI
DPU 1	BJ
DPU 2	CK

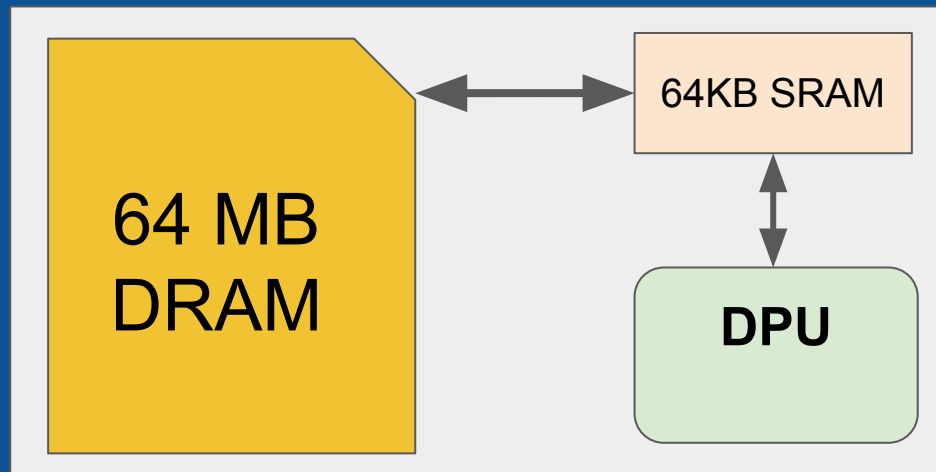
Raw Performance: Throughput

9 ranks x 64 DPUS = **576 DPUs**

576 DPUs x 64MB = **36GB DRAM**

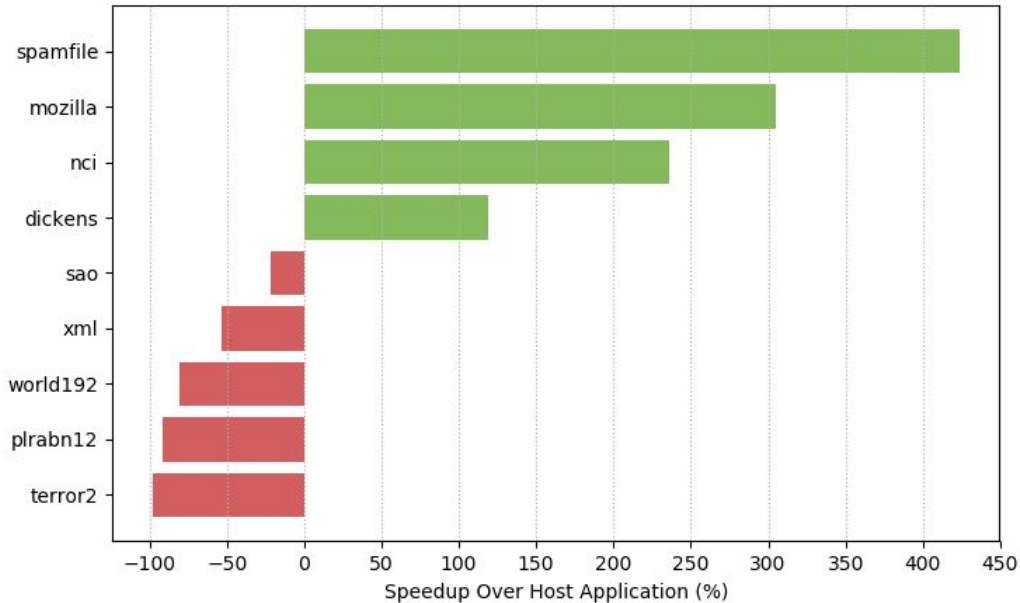
36 GB in 0.16 s = **252 GB/s**

**Top speed of DDR4-2400
channel: 19GB/s**



16 threads @ 2KB per transfer

Use Case: Compression



File	Size	DPUs
spamfile	84 MB	172
mozilla	50 MB	105
nci	30 MB	64
dickens	10 MB	35
sao	7 MB	21
xml	5 MB	15
world192	1 MB	4
plrabn12	0.5 MB	2
terror2	0.1 MB	1

Wishlist

**Data Triggered
Functions**

**Concurrent
Memory Access**

**Mix Of
Memory Types**

**Tuning For
Performance**

Future Directions

Hyperdimensional
Computing

Regular
Expression
Search

?

Thank you for watching

Joel Nider joel@ece.ubc.ca

Craig Mustard craigm@ece.ubc.ca

Andrada Zoltan zoltandrada@gmail.com

Alexandra Fedorova sasha@ece.ubc.ca