Beyond Bandwidth Doubling: Embrace Bit-Flips and Unlock Processing-in-NAND

Max Berens, Yun-Chih Chen, Jian-Jia Chen, Jens Teubner TU Dortmund University, Germany ICDE — 22.05.2025

Data-Centric Processing

"How to evaluate ... FROM A > 42?"



Data-Centric Processing

"How to evaluate push-down ... FROM A > 42?"



Data-Centric Processing

"How to evaluate push-down ... FROM A > 42?"



This is not enough!

Processing inside NAND Flash Memory

"How to push-down ... FROM A > 42?"





Processing-in-NAND + DBMS = ...?



Processing-in-NAND + DBMS = ...?

• So many **limitations**...



Processing-in-NAND + DBMS = ...?

- So many **limitations**...
- Much more **complicated**...

Interfaces

Processing-in-NAND + DBMS = ...?

- So many **limitations**...
- Much more **complicated**...
- Suddenly bit flips!?

	$0 \rightarrow 1$
0 7	
	I H

Processing-in-NAND + DBMS = ...?

- So many **limitations**...
- Much more **complicated**...
- Suddenly bit flips!?

Is it worth the trouble?

Performance?
00













The Illusion of Error-Free Memory

Speaking of ECC...

- ECC Guarantee: ≤ 1 per 1.25**PiB**
- Without ECC: \approx 1 per 100-1000 bit
- powerful ECC with read retry (LDPC) mandatory!
- latency **fluctuate** heavily



The Illusion of Error-Free Memory

Speaking of ECC...

- ECC Guarantee: ≤ 1 per 1.25**PiB**
- Without ECC: \approx 1 per 100-1000 bit
- powerful ECC with read retry (LDPC) mandatory!
- latency **fluctuate** heavily



The Illusion of Error-Free Memory

Speaking of ECC...

- ECC Guarantee: ≤ 1 per 1.25**PiB**
- Without ECC: \approx 1 per 100-1000 bit
- powerful ECC with read retry (LDPC) mandatory!
- latency **fluctuate** heavily

Example: Tian et al. TODAES 2024 ECC latency difference: $85\mu s \rightarrow 1ms$ latency (6 retries)



MANY Factors, e.g.

• Age — data is leaking over time...

MANY Factors, e.g.

- Age data is leaking over time...
- Endured writes cycles
 - "Fail slow" Old SSDs become slower...

Impact of Errors and Heat

MANY Factors, e.g.

- Age data is leaking over time...
- Endured writes cycles
 - "Fail slow" Old SSDs become slower...
- Temperature!

Example: Ye et al. ASPLOS 2024

- Up to 6 read-retries after holding data for 2 days at 85°C
- Equivalent: 168 days at 40°C

Not just a matter of cooling:

- Performance/bus speed \sim power consumption
- Power-budget per area is **fixed**
- Chip is always partially "dark"/unused
- Technology scaling makes it worse



Not just a matter of cooling:

- Performance/bus speed \sim power consumption
- Power-budget per area is **fixed**
- Chip is always partially "dark"/unused
- Technology scaling makes it *worse*

Bottom Line

Chip designs are *power-efficiency limited*



Not just a matter of cooling:

- Performance/bus speed \sim power consumption
- Power-budget per area is fixed
- Chip is always partially "dark"/unused
- Technology scaling makes it worse

Bottom Line

Chip designs are *power-efficiency limited*

Bandwidth

Dark Silicon and the End of Mutticore Scaling

Hadi Esmaelizadeh Emily Blemi: Renée St. Amant: Karthikeyan Sankaralingam: Doug Burger: University of Washington University of Wescnish-Madason "The University of Taxas at Autor "Microsoft Research hadareh@ex.washington.edu bem@cxw.isedu stami@ex.utexas.edu deuper@microsoft.com

Not just a matter of cooling:

- Performance/bus speed \sim power consumption
- Power-budget per area is fixed
- Chip is always partially "dark"/unused
- Technology scaling makes it worse

Bottom Line

Chip designs are *power-efficiency limited*

Bandwidth

Dark Silicon and the End of Multicore Scaling

Hadi Esmaelizadeh Emily Blem: Renée St. Amant: Karthikeyan Sankaralingam: Doug Burger University of Washington University of Wisconsin-Adation "The University of Toxas at Austin "Microsoft Research hadareh@cs.washingtonedu.biem@cs.wiscod.stani@cs.wisc.ed..dturget@microsoft.com

2012!

The Bleak Future of NAND Flash Memory

Laura M. Grupp^{*}, John D. Davis[†], Steven Swanson^{*} *Department of Computer Science and Engineering, University of California, San Diego [†]Microsoft Research, Mountain View

NAND Design Goal

(Endurance-limited) mass-storage memory needs to be affordable (GB/\$)

NAND Design Goal

(Endurance-limited) mass-storage memory needs to be affordable (GB/\$)

Instead of Technology Scaling:

• TLC & QLC, 3D-stacking, "SLC"-Mode, Larger pages, DRAM caches, . . .

NAND Design Goal

(Endurance-limited) mass-storage memory needs to be affordable (GB/\$)

Instead of Technology Scaling:

• TLC & QLC, 3D-stacking, "SLC"-Mode, Larger pages, DRAM caches, ...

Bottom Line

Capacity keeps scaling, but **neither** latency nor plane-count keeps up!

NAND Design Goal

(Endurance-limited) mass-storage memory needs to be affordable (GB/\$)

Instead of Technology Scaling:

• TLC & QLC, 3D-stacking, "SLC"-Mode, Larger pages, DRAM caches, ...

Bottom Line

Capacity keeps scaling, but **neither** latency nor plane-count keeps up!

So, how does Processing-in-NAND help?

PiN Improves Goodput

Core Idea Save data movement \Rightarrow safe power!



PiN Improves Goodput

Core Idea Save data movement \Rightarrow safe power!

What kind of logic is possible?

- Proximity implies limitations
- First: Get data out (sensing)
- Peripheral circuitry gives some options



PiN Improves Goodput

Core Idea Save data movement \Rightarrow safe power!

What kind of logic is possible?

- Proximity implies limitations
- First: Get data out (sensing)
- Peripheral circuitry gives some options



New chip design is expensive...



New chip design is expensive... \Rightarrow re-purpose existing circuits!



New chip design is expensive... ⇒ re-purpose existing circuits!

FlashCosmos: Park et al. MICRO 2022

Bit-wise &&, ||, . . ., across pages



New chip design is expensive... \Rightarrow re-purpose existing circuits!

FlashCosmos: Park et al. MICRO 2022

Bit-wise &&, ||, . . ., across pages

Search-in-Memory: Chen at al. TCAD 2024

Word-wise equality check and partial access



New chip design is expensive... \Rightarrow re-purpose existing circuits!

FlashCosmos: Park et al. MICRO 2022

Bit-wise &&, ||, . . ., across pages

Search-in-Memory: Chen at al. TCAD 2024

Word-wise equality check and partial access

But... what about **bit-flips**?



Application: Probabilistic Data Structures



Embrace bit-flips with bulk index evaluation:

• Bloom filter — set membership

Performance degrades gracefully

Application: Probabilistic Data Structures



Embrace bit-flips with *bulk index evaluation*:

- Bloom filter set membership
- Binary Sketch approximate similarity search

Performance degrades gracefully

What can you do?

• Rely less on raw bandwidth..

- Rely less on raw bandwidth..
- Find new applications for PiN

- Rely less on raw bandwidth..
- Find new applications for PiN
- Relax error requirements, become error-aware

- Rely less on raw bandwidth..
- Find new applications for PiN
- Relax error requirements, become error-aware
- Keep (your SSD) cool!

