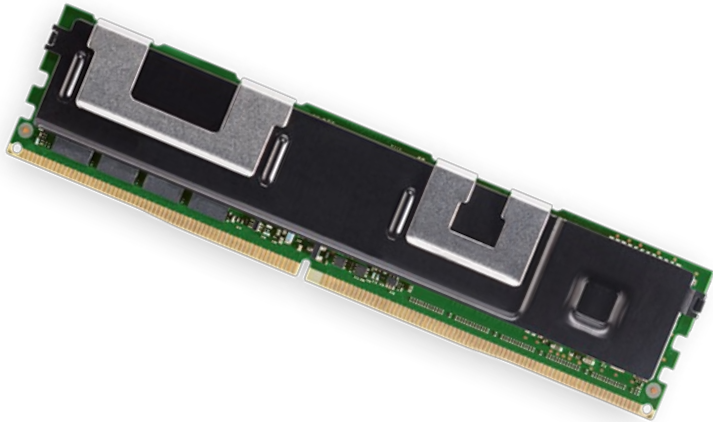# Writes Hurt: Lessons in Cache Design for Optane NVRAM

Alexandra Fedorova, Keith A. Smith, Keith Bostic, Susan LoVerso, Michael Cahill and Alex Gorrod

THE UNIVERSITY OF BRITISH COLUMBIA

mongoDB®

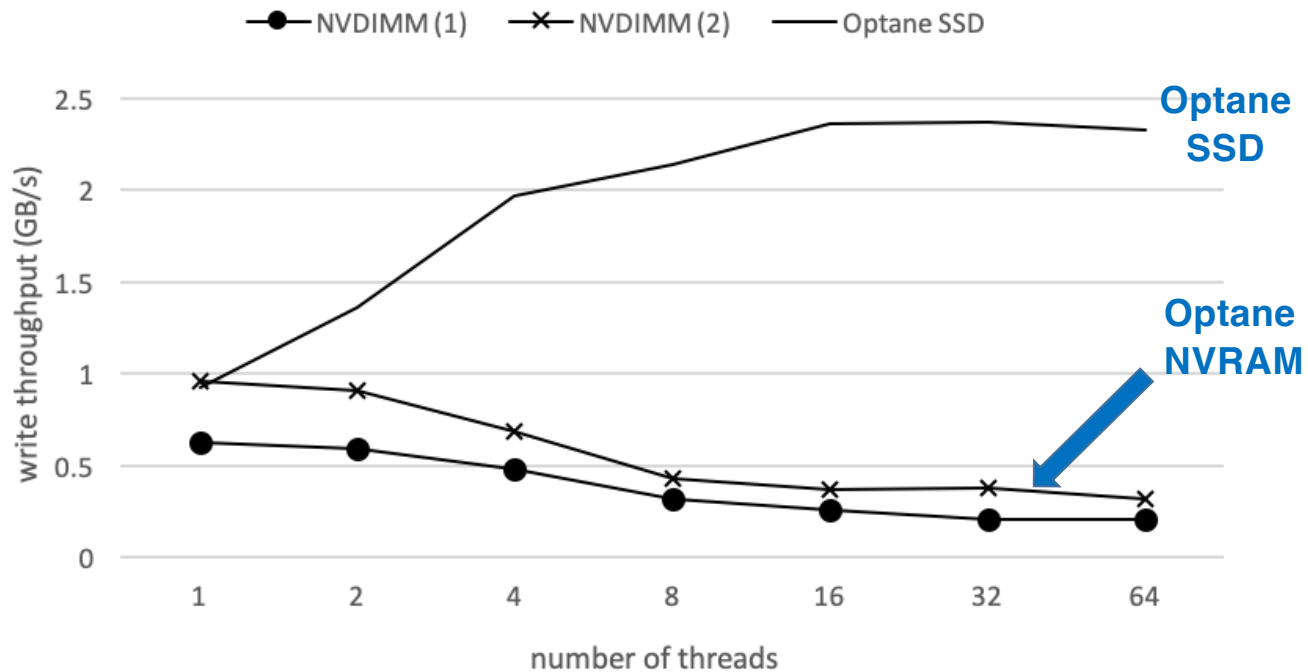# A cache that lives in Optane NVRAM*

Just another cache?

Throwing spanner in the works

NVRAM revolts…

-- Famous Japanese Haiku, unknown author
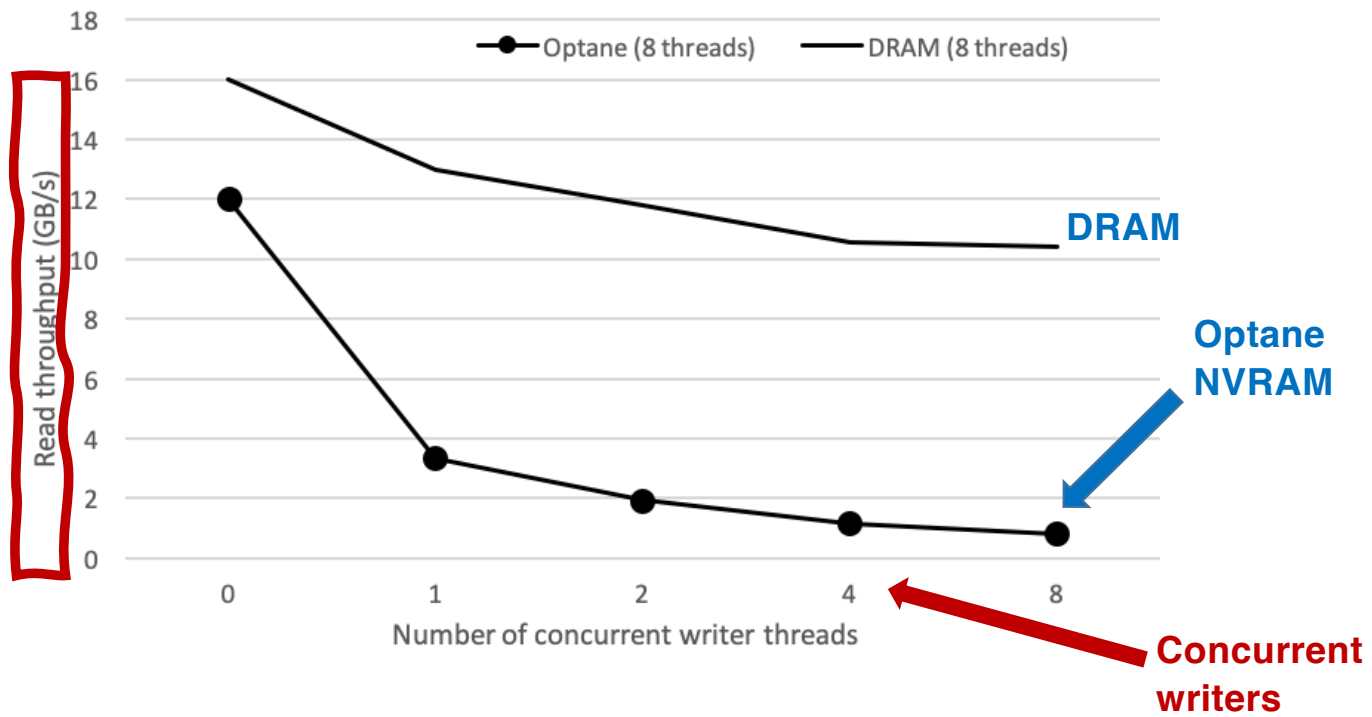
*Non-volatile memory
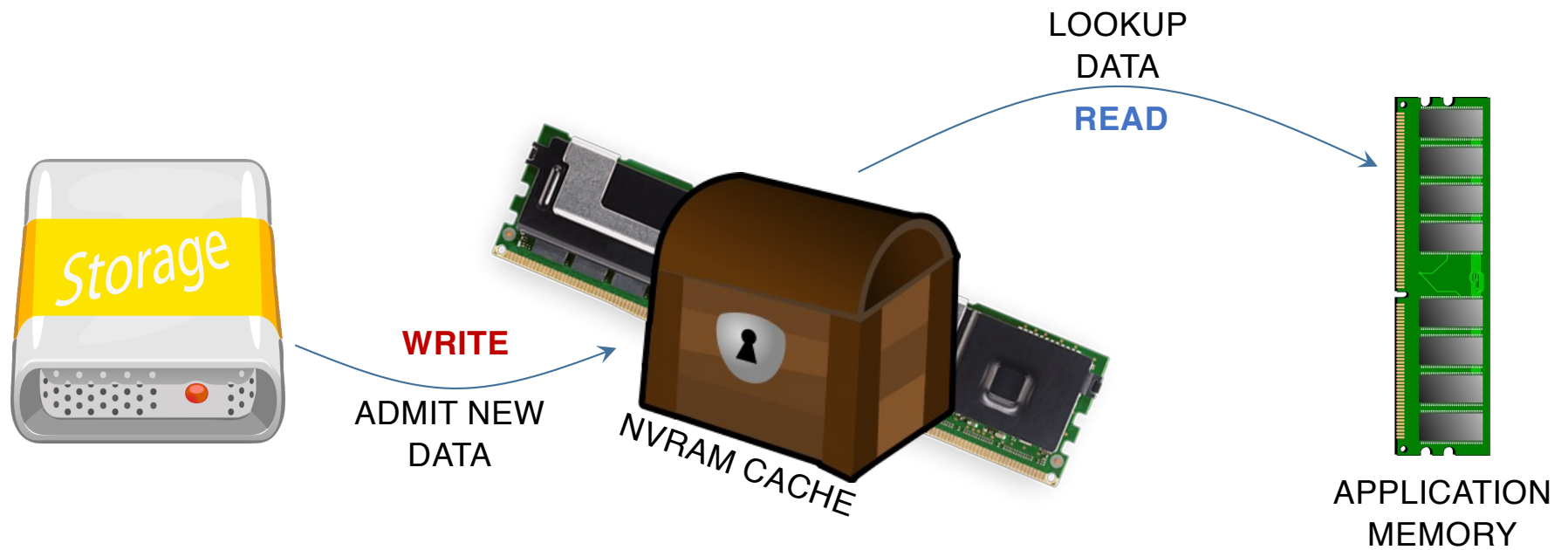
Limited write throughput

Well known phenomenon:

Yang et al. An Empirical Guide to the Behavior and Use of Scalable Persistent Memory, FAST'20
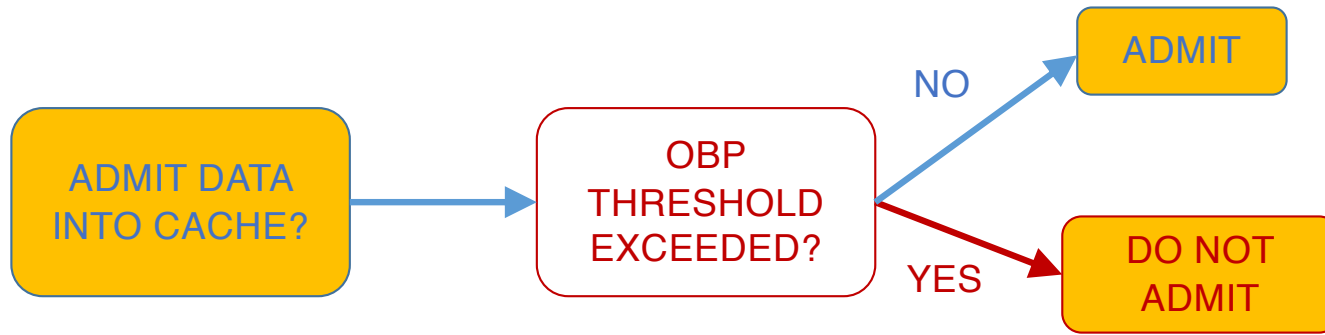
Writes hurt reads

Cache must ration the writes

LOOKUP
DATA
READ

WRITE

ADMIT NEW
DATA

NVRAM CACHE

APPLICATION
MEMORY

Storage

Overhead Bypass heuristic (OBP)

ADMIT DATA INTO CACHE? → OBP THRESHOLD EXCEEDED?

NO → ADMIT

YES → DO NOT ADMIT

$$OBP = \frac{blocks\_inserted + blocks\_removed}{blocks\_looked\_up}$$

COST

BENEFIT

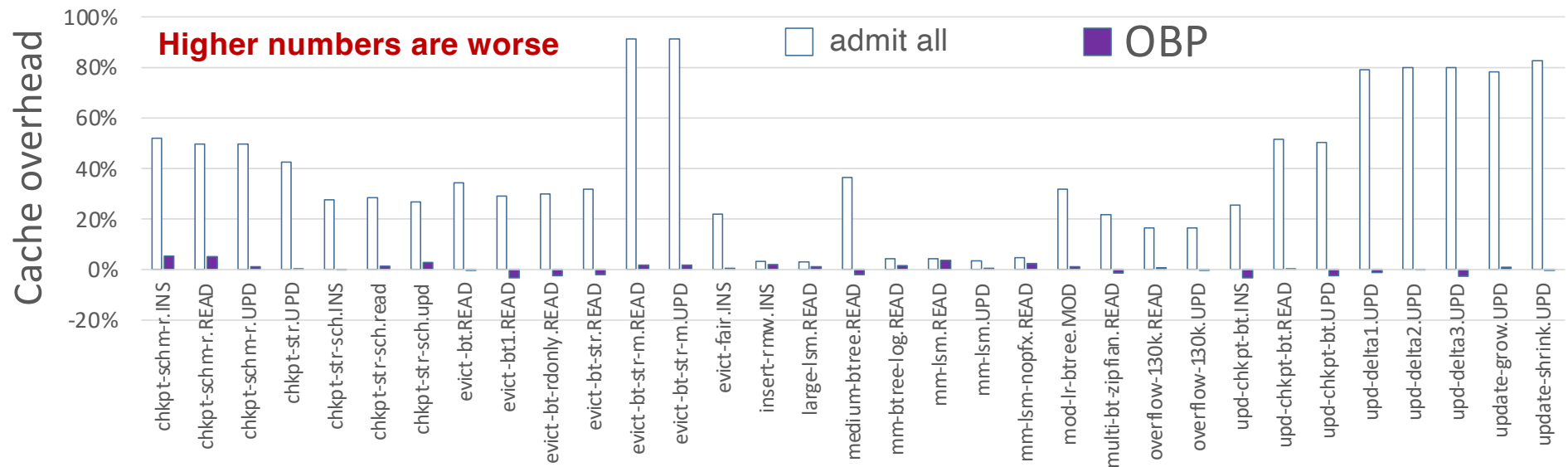**OBP captures dynamic workload properties**

How to set OBP threshold?

WRITES REALLY HURT? ↓ **LOW THRESHOLD**

WRITES DON'T HURT? ↑ **HIGH THRESHOLD**

OBP protects from the pain of writes

- An experiment where caching has no benefit (small datasets and/or lots of writes)
- WiredTiger -- the MongoDB storage engine
- Caching should have zero effect on performance

OBP needed for read-only workloads too

Higher is better

WL	Eager eviction		No eviction	
	ops/sec	hit rate	ops/sec	hit rate
evict-btree-large	61,699	48%	162,690	44%
evict-btree-scan.read	97,491	45%	134,404	36%
medium-btree-large	62,012	48%	164,644	44%

- No OBP in this experiment
- Eager eviction **admits more data**
- Generates **more writes**

More predictable than Memory Mode
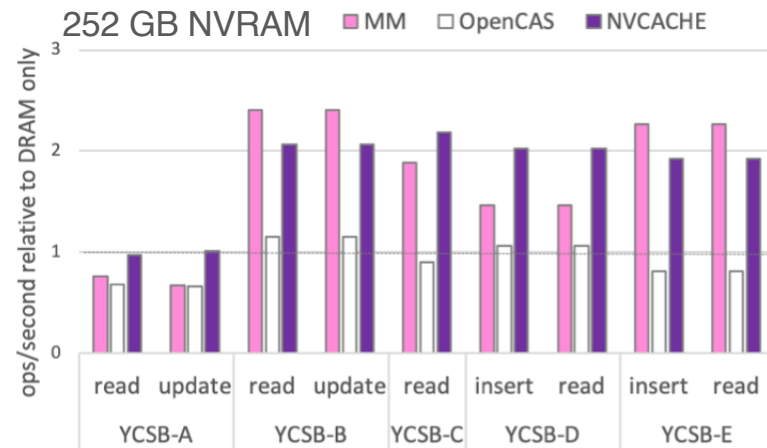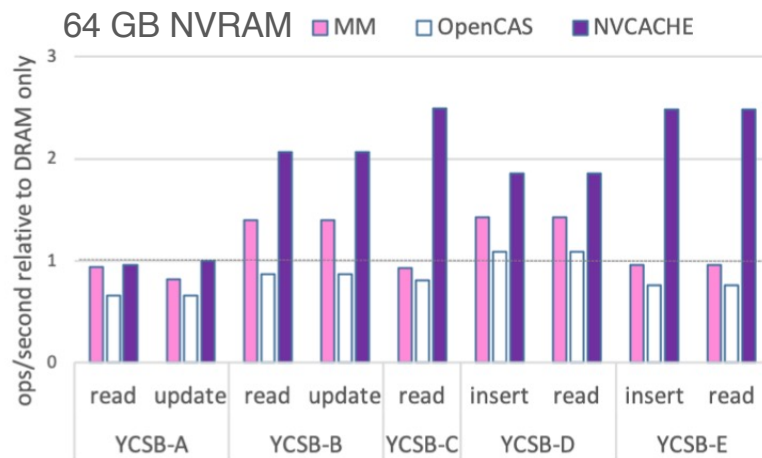
- YCSB over WiredTiger – the MongoDB storage engine
- Higher numbers are better
- MM = Intel Memory Mode (NVRAM is an extension of DRAM)
- OpenCAS = a cache for file systems in Linux kernel

Summary

- Writes hurt reads on Optane NVRAM

- Cache admission policy must ration writes

- Overhead Bypass threshold implements write rationing

- Useful for any memory technology, where writes significantly hurt reads