Unioning of the Buffer Cache and Journaling Layers with Non-volatile Memory

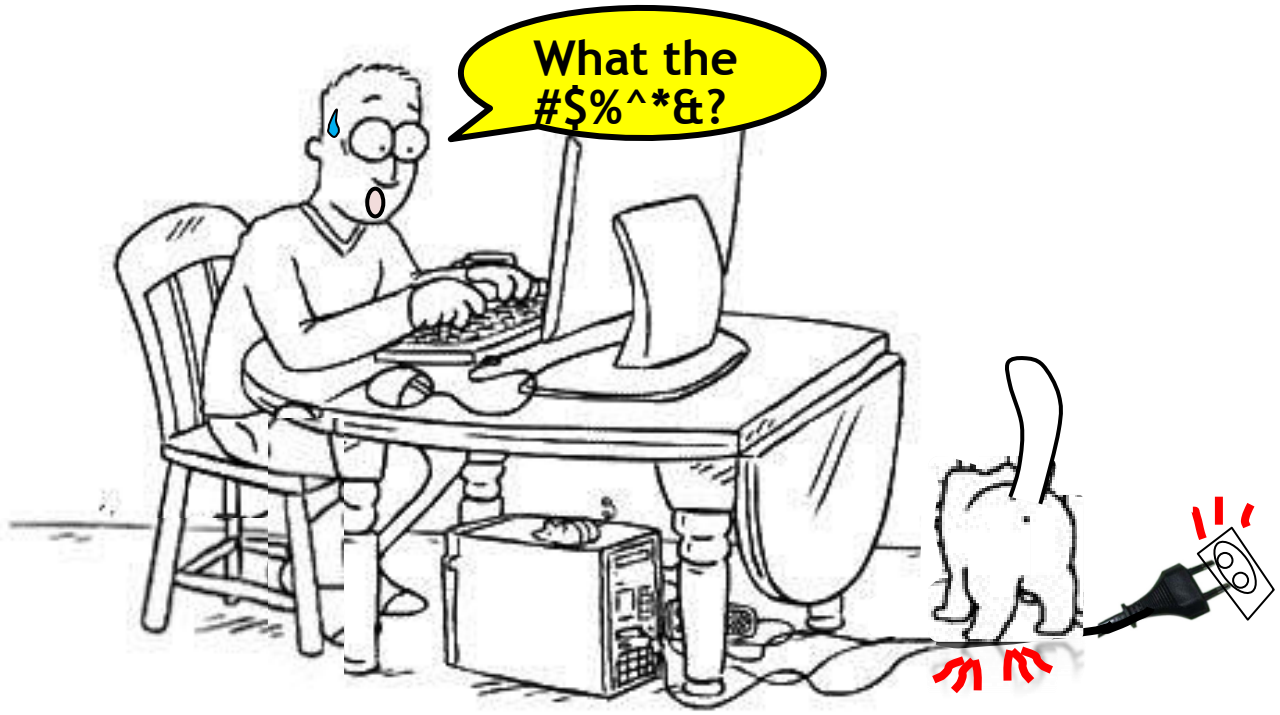
USENIX FAST ‘13

Eunji Lee (Ewha University, Seoul, Korea)
Hyokyung Bahn (Ewha University)
Sam H. Noh (Hongik University)

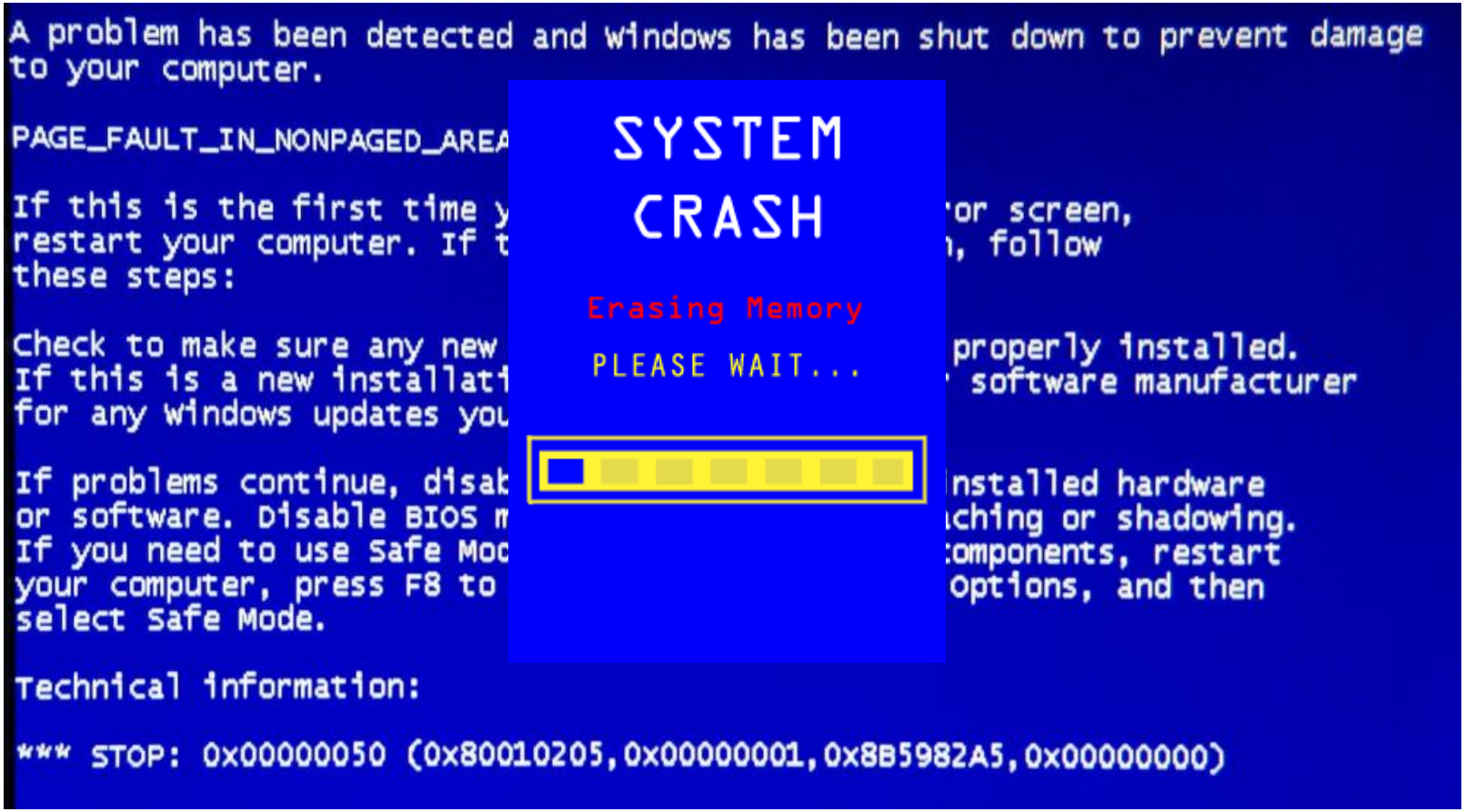
Outline

- Reliability issues in storage systems
- Non-volatile memory as a solution
- UBJ: Unioning of Buffer cache and Journaling
- Performance evaluation

A man working hard …

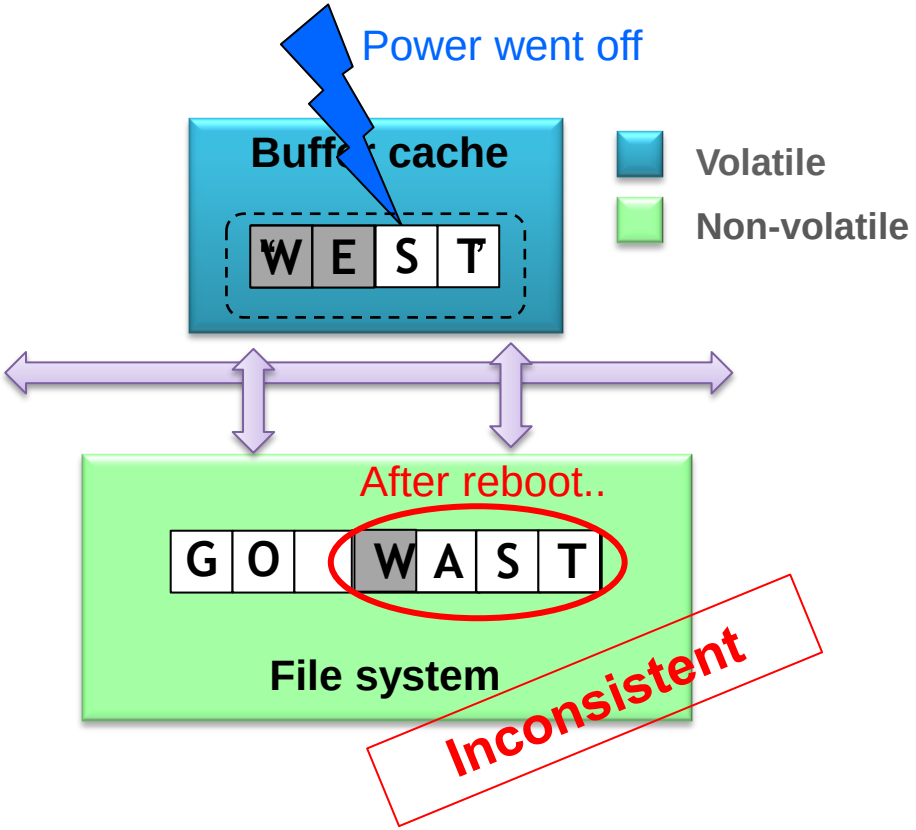


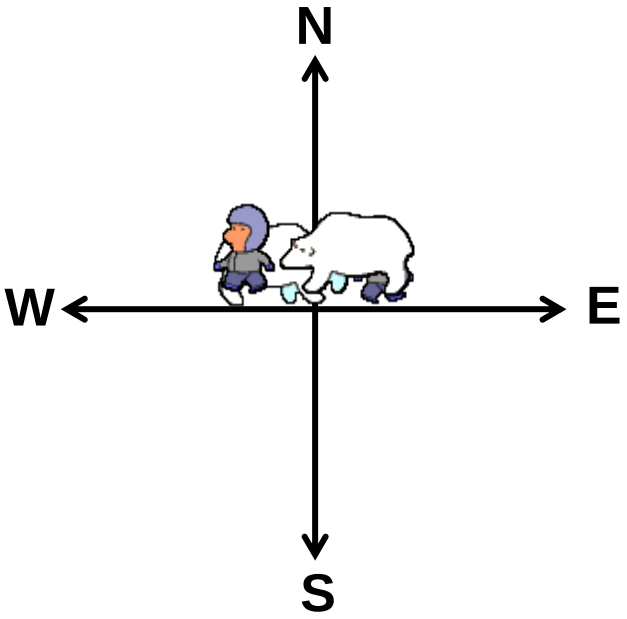
A man working hard …



So what happened?

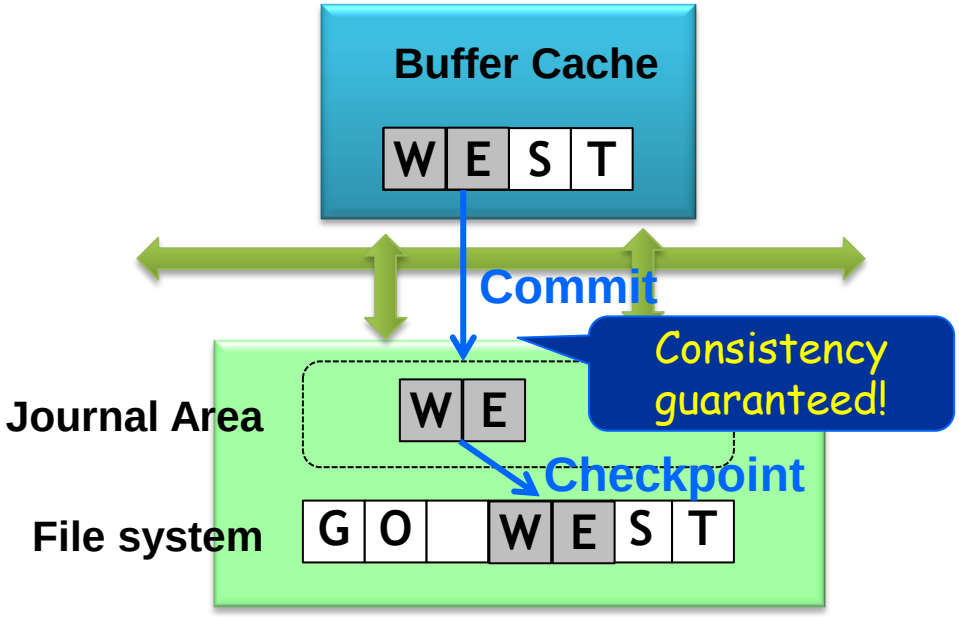
- Sudden power failure incurs file system inconsistency

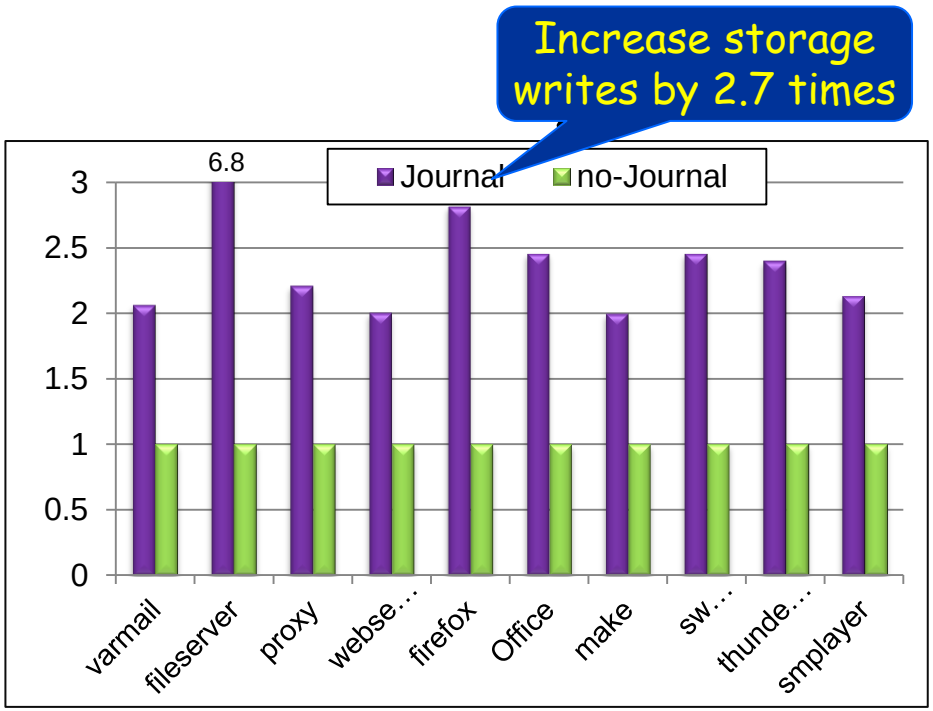




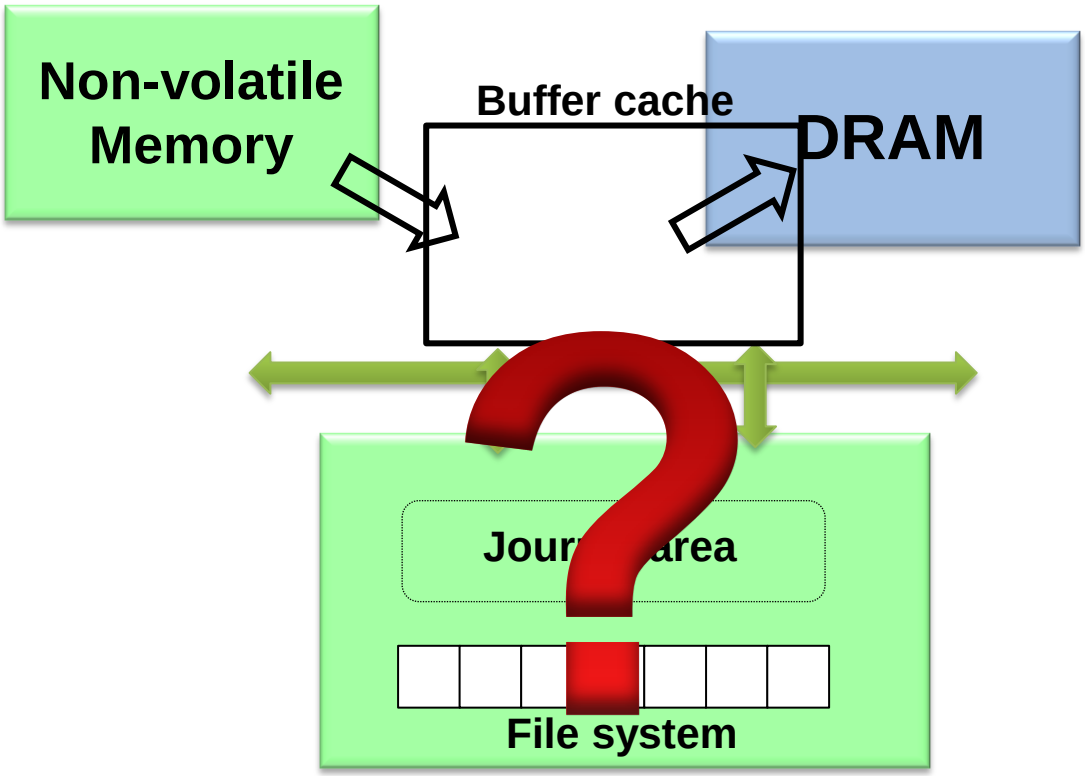
Journaling as a solution

- Prevent data inconsistency through write-twice
  - ext4, ReiserFS, XFS, btrFS …



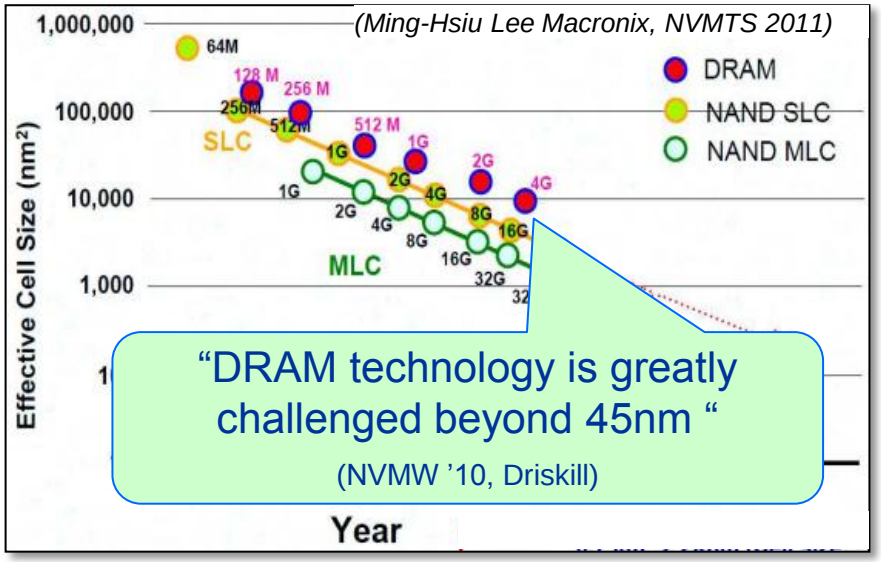


Non-volatile memory as a solution



Non-volatile memory as main memory

1. Scaling Limit of DRAM



2. Power consumption

As much as 40% of the total system energy is consumed by the main memory subsystem in a mid-range IBM eServer machine. (Querish, ISCA 2009)

Replacing DRAM with STT-RAM in data centers can reduce power by up to 75%
(NVMW '10, Driskill)

3. Demand for fast memory access

As critical applications are becoming more data-centric, memory performance is fast becoming the key bottleneck

Non-volatile Memory Technology

Source: T. Perez, C. A. F. D Rose, Technical Report, PUCRS, 2010

	SRAM	DRAM	Disk	NAND Flash		PCRAM	RRAM (Memristor)	MRAM (STT-RAM)
Maturity	Product	Product	Product	Product		Advanced development	Early development	Advanced development
Cell Size	>100 F^2	6-8 F^2	(2/3) F^2	4-5 F^2		8-16 F^2	>5 F^2	37 F^2
Read Latency	<10 ns	10-60 ns	8.5 ms	25 µs		48 ns	<10 ns	<10 ns
Write Latency	<10 ns	10-60 ns	9.5 ms	200 µs		40-150 ns	~10 ns	12.5 ns
Energy per bit access	>1 pJ	2 pJ	100-1000 mJ	10 nJ		100 pJ	2 pJ	0.02 pJ
Static Power	Yes	Yes	Yes	No		No	No	No
Endurance	$>10^{15}$	$>10^{15}$	$>10^{15}$	10^4		10^8	10^5	$>10^{15}$
Nonvolatility	No	No	Yes	Yes		Yes	Yes	Yes
	Current Memory Technologies					Emerging NVM Technologies		

Scalability ☐ Low-power ☐ High-performance ☐

Non-volatile Memory Technology

Source: T. Perez, C. A. F. D Rose, Technical Report, PUCRS, 2010

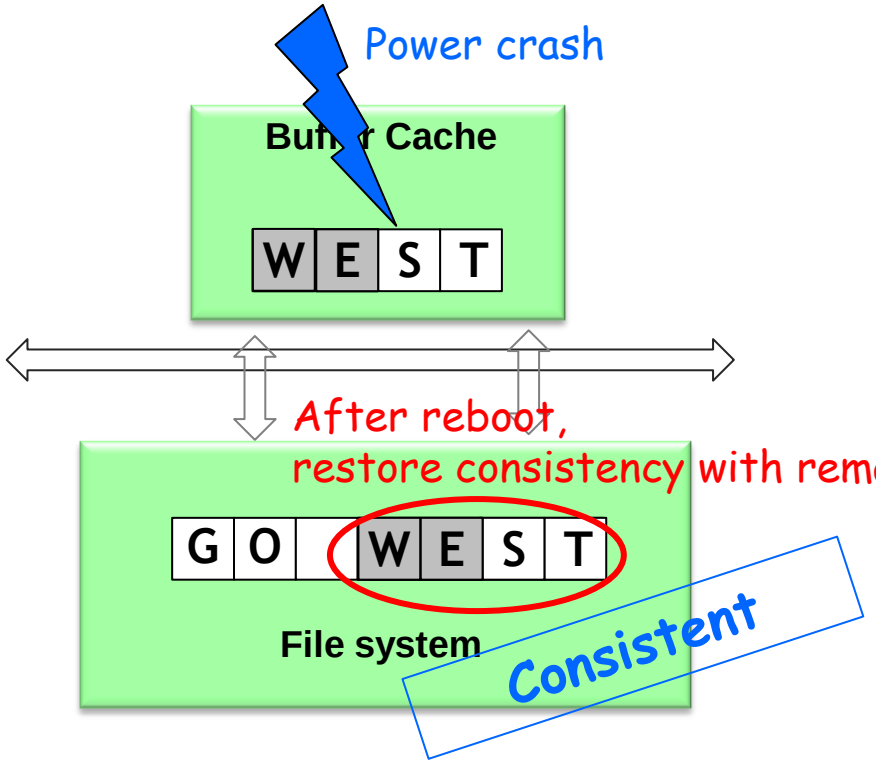
	SRAM	DRAM	Disk	NAND Flash	PCRAM	RRAM (Memristor)	MRAM (STT-RAM)
Maturity	Product	Product	Product	Product	Advanced development	Early development	Advanced development
Cell Size	>100 F^2	6-8 F^2	(2/3) F^2	4-5 F^2	8-16 F^2	>5 F^2	37 F^2
Read Latency	<10 ns	10-60 ns	8.5 ms	25 µs	48 ns	<10 ns	<10 ns
Write Latency	<10 ns	10-60 ns	9.5 ms	200 µs	40-150 ns	~10 ns	12.5 ns
Energy per bit access	>1 pJ	2 pJ	100-1000 mJ	10 nJ	100 pJ	2 pJ	0.02 pJ
Static Power	Yes	Yes	Yes	No	No	No	No
Endurance	$>10^{15}$	$>10^{15}$	$>10^{15}$	10^4	10^8	10^5	$>10^{15}$
Nonvolatility	No	No	Yes	Yes	Yes	Yes	Yes
	Current Memory Technologies				Emerging NVM Technologies		

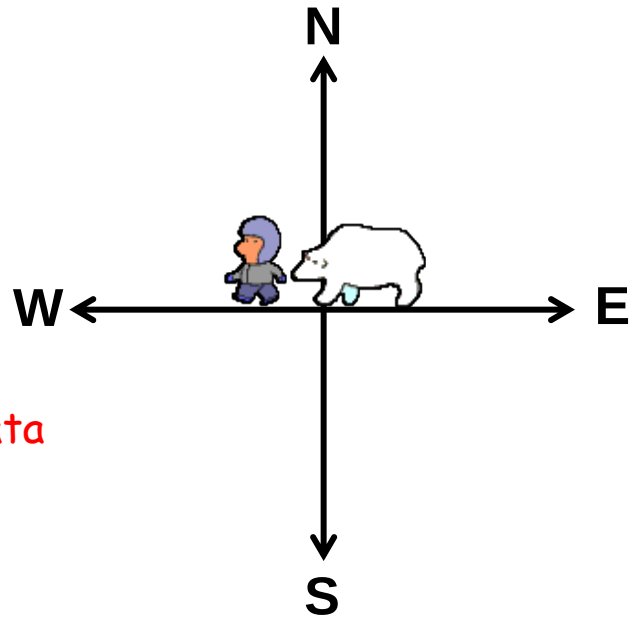
Scalability ☑ Low-power ☑ High-performance ☑

(Optimistic expectations)

Non-volatile memory as a solution

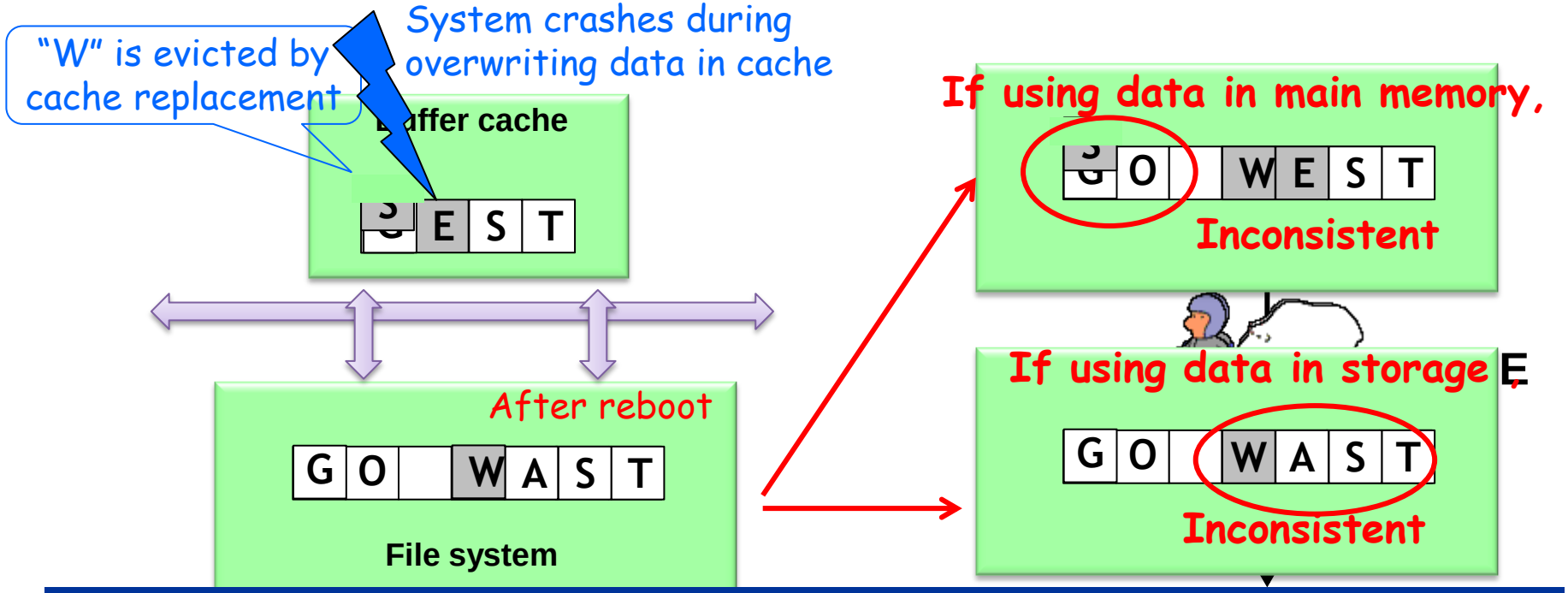
- Seems to provide data consistency





Non-volatile memory as a solution?

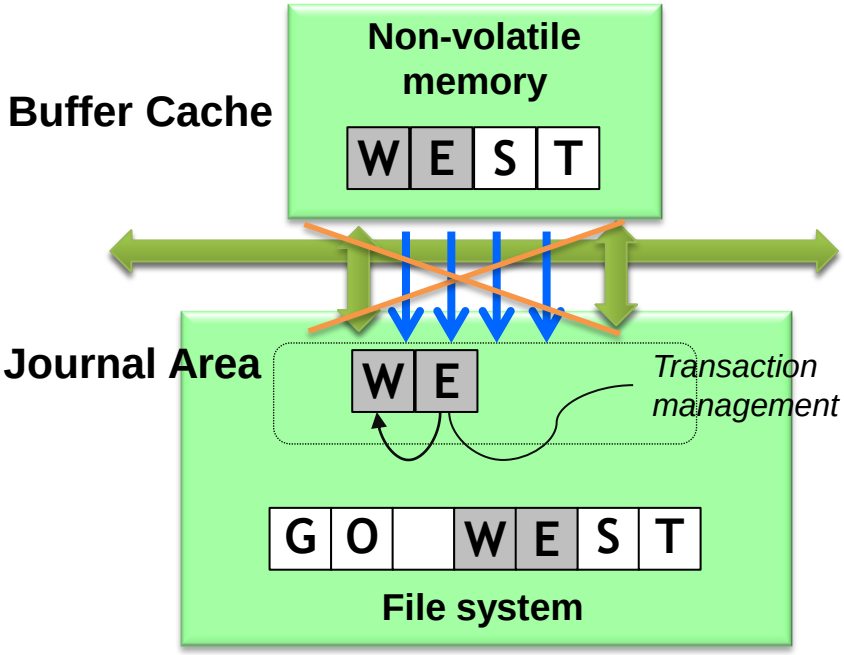
- ✓ Inconsistency problem still exists with NVM



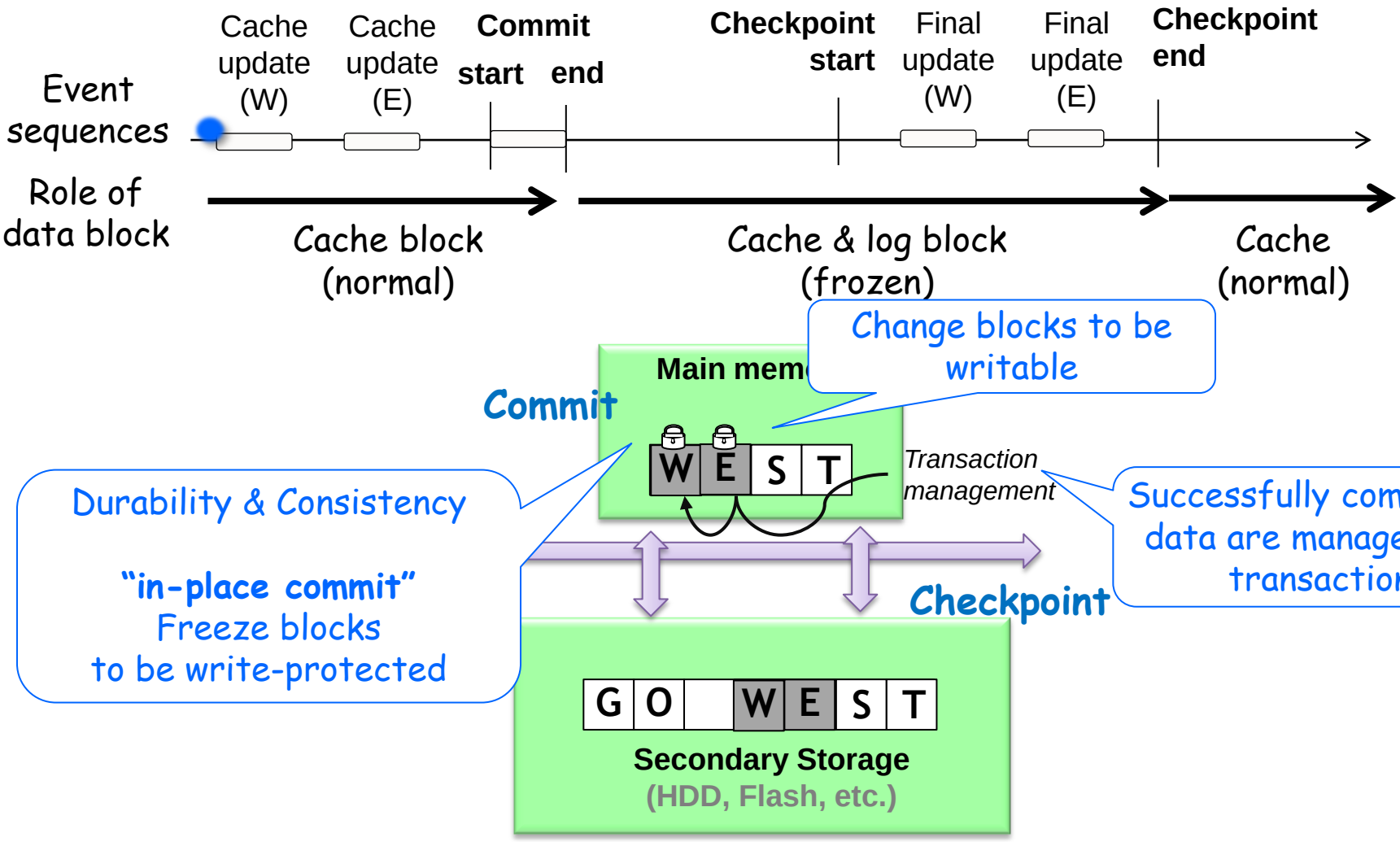
Drop-in replacement of non-volatile memory does not suffice

Unioning of Buffer cache and Journaling Layers (UBJ)

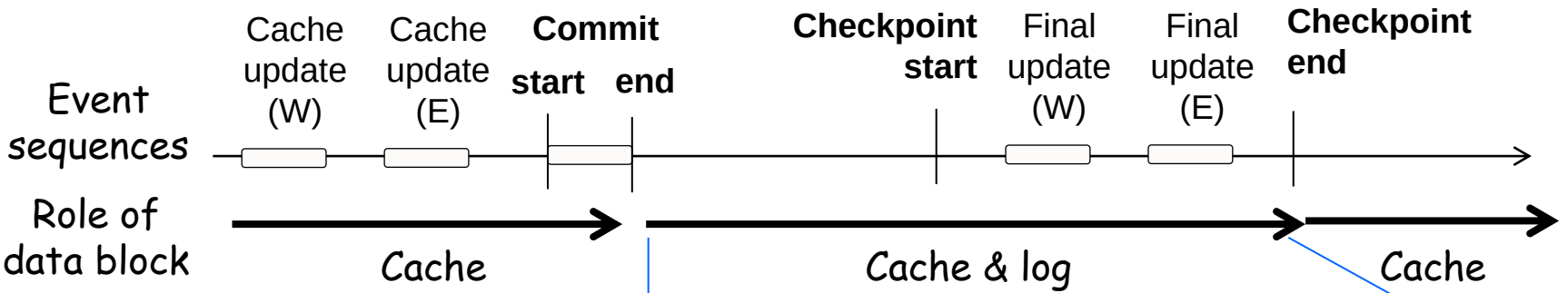
- Provide data consistency without sacrificing performance
- Design a novel buffer cache architecture “UBJ”
- Subsume functions of caching and journaling
  - Use data block for dual purposes
  - Provide journaling effect through transition of cache block state



Workings of UBJ



Workings of UBJ



Please refer to our paper for details!

Unioning of the Buffer Cache and Journaling Layers with Non-volatile Memory

Eunji Lee[1], Sam H. Noh[2], and Hyokyung Bahn[1]

[1]Ewha University, bahn@ewha.ac.kr

[2]Hongik University, http://next.hongik.ac.kr

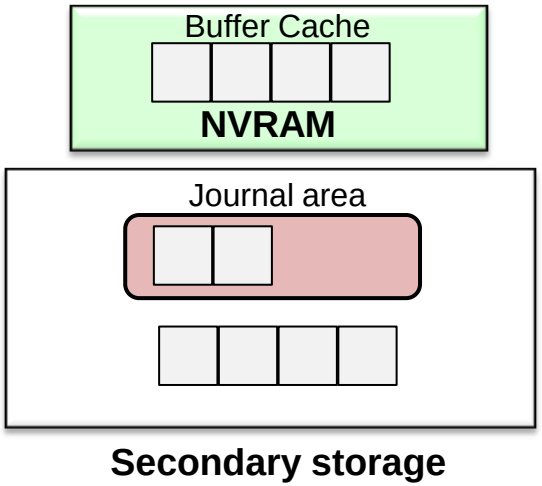
Abstract

1. Introduction

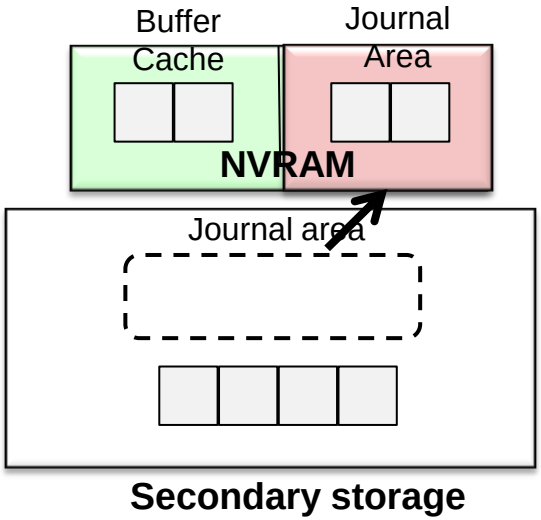
Log blocks

- ✓ Transaction Management
- ✓ Protected from replacement
- ✓ Copy-on-write for write request
- ✓ Serve read requests as cache blocks

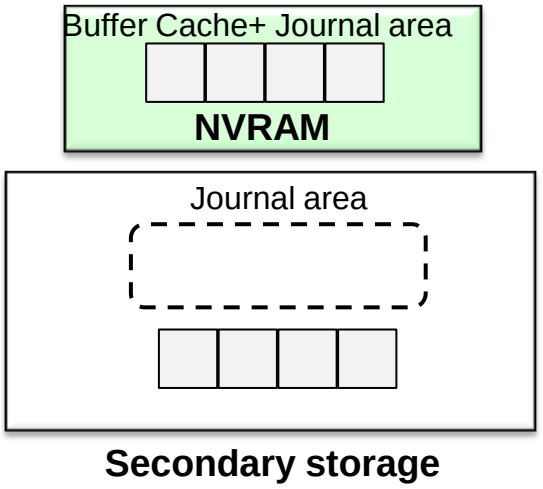
Cache performance of UBJ



1. Original buffer cache

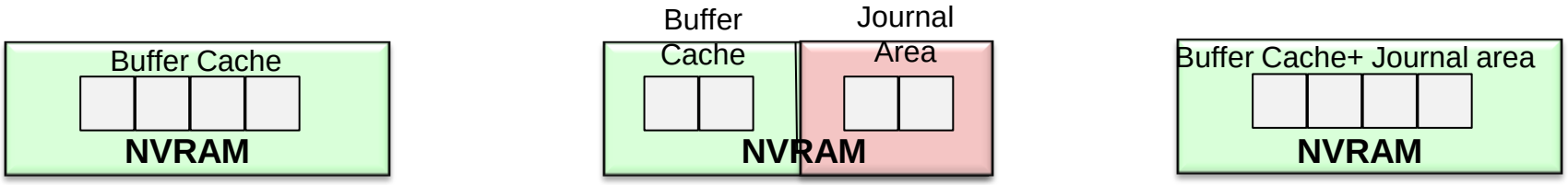


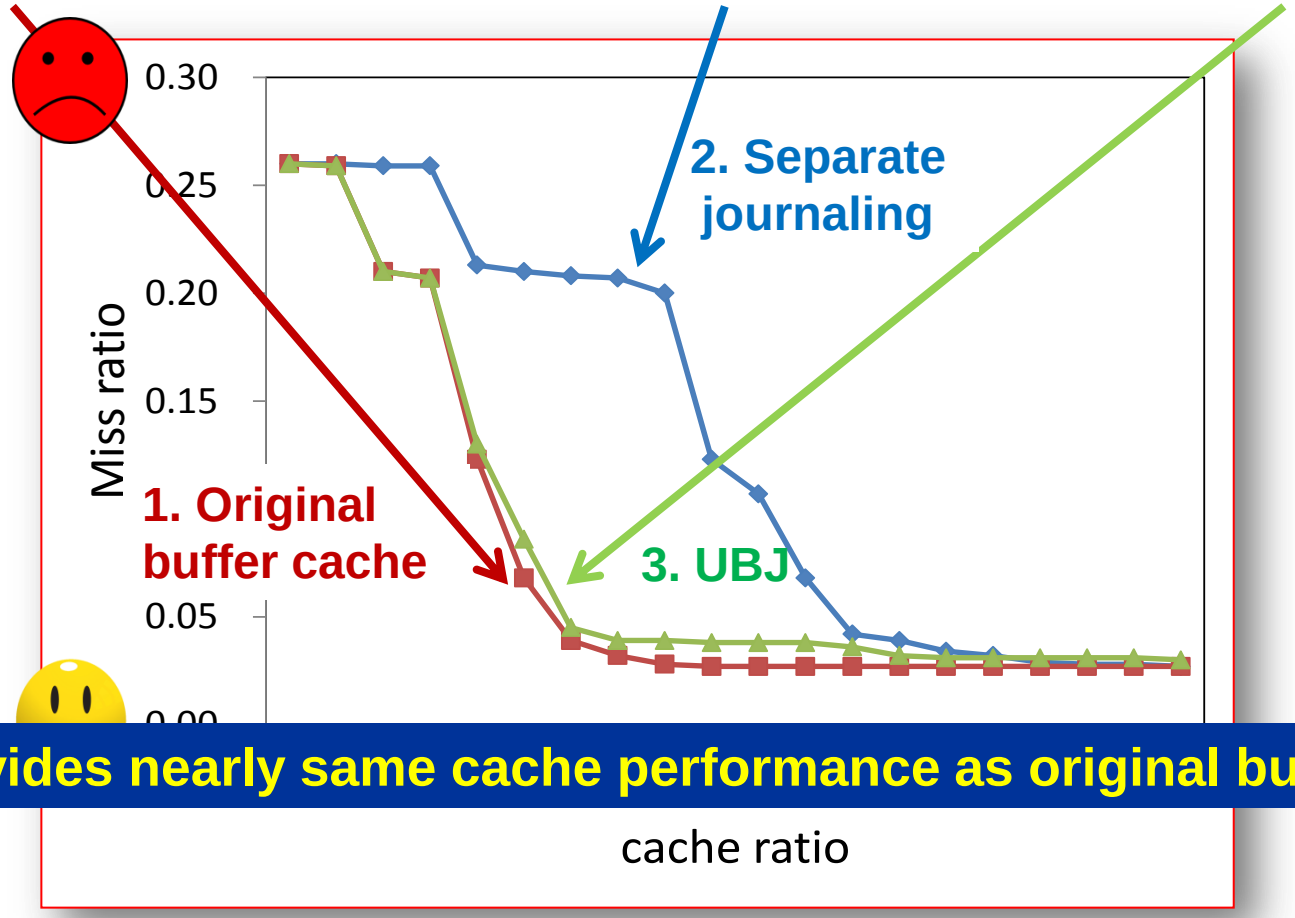
2. Separate journaling



3. UBJ

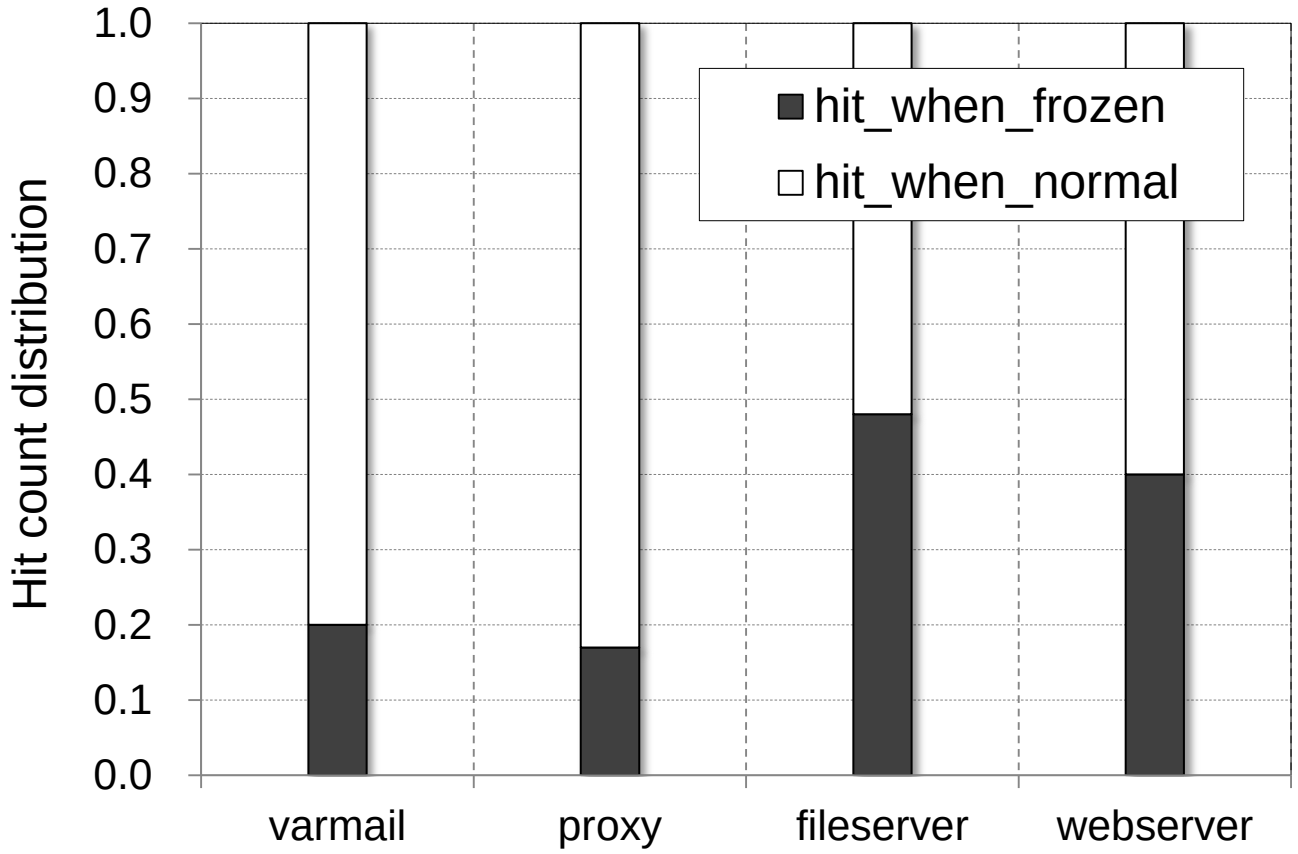
Cache performance of UBJ



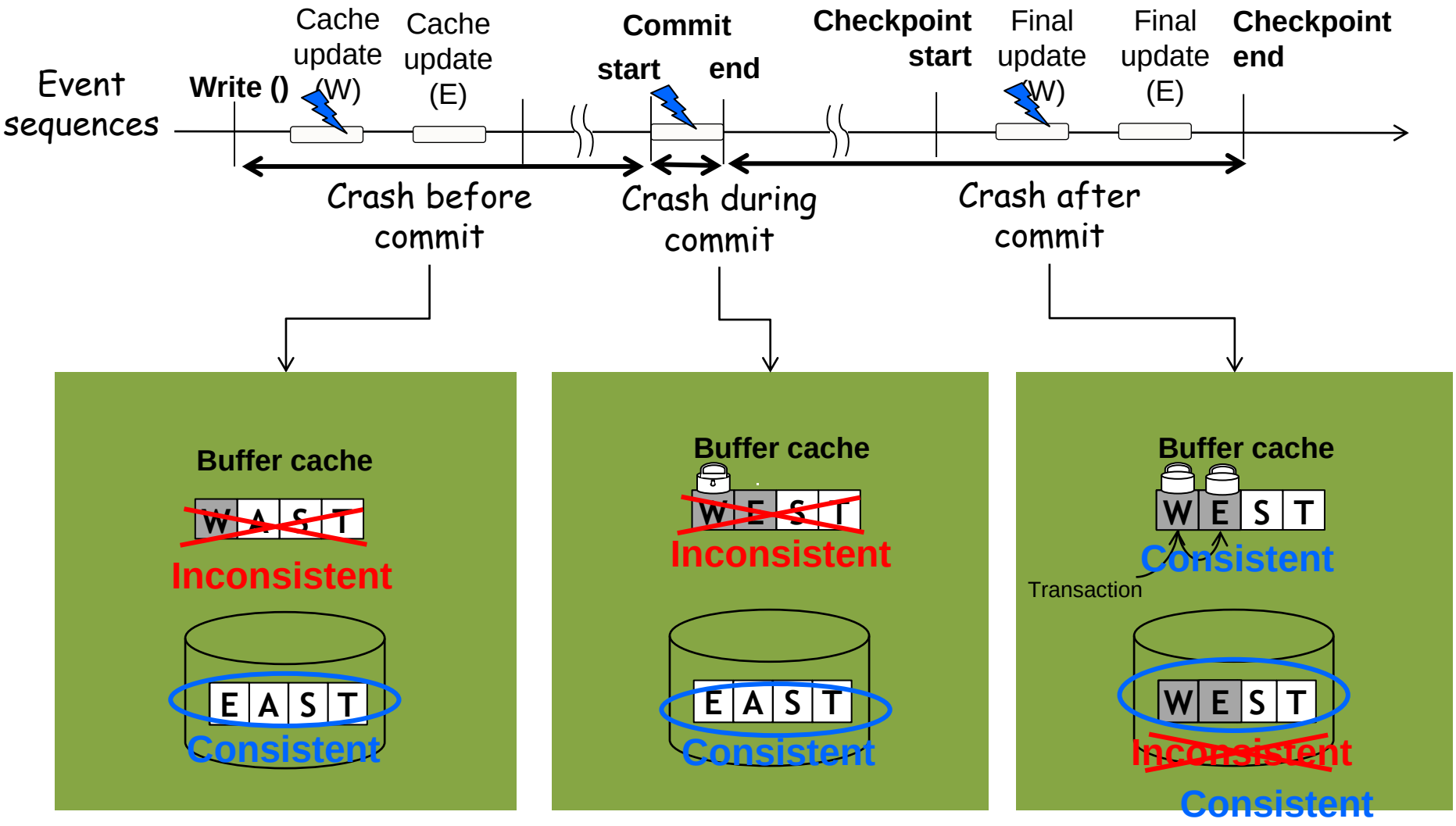


UBJ provides nearly same cache performance as original buffer cache

Cache hits on frozen data blocks



System recovery

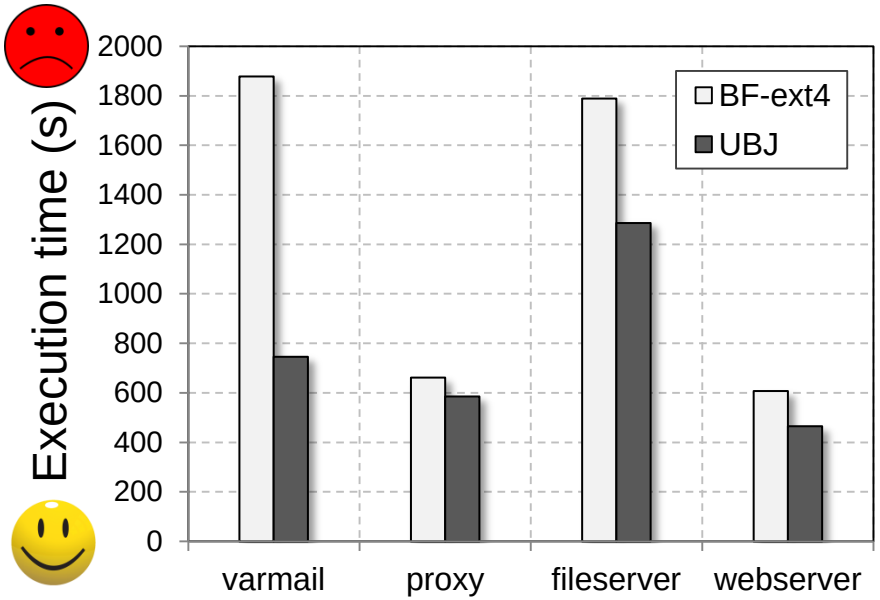


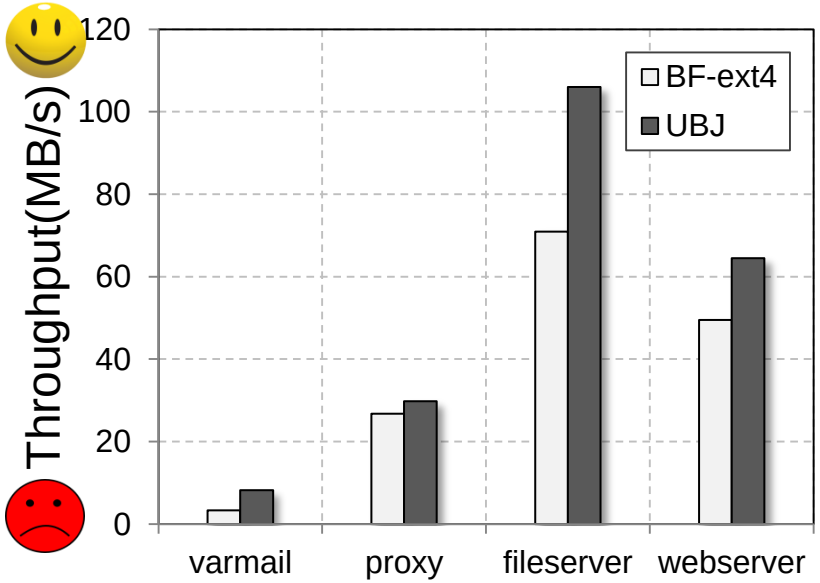
Performance Evaluation

- Prototype of UBJ on Linux 2.6.38
- Intel Core i3-2100 CPU
  - 3.1GHz and 4GB of DDR2-800 memory
- Emulate non-volatile memory with DRAM
- Compare with ext4 in journal-mode
  - logs both data and metadata
- Three benchmarks
  - Filebench, IOzone, Postmark

Performance Evaluation

- Filebench

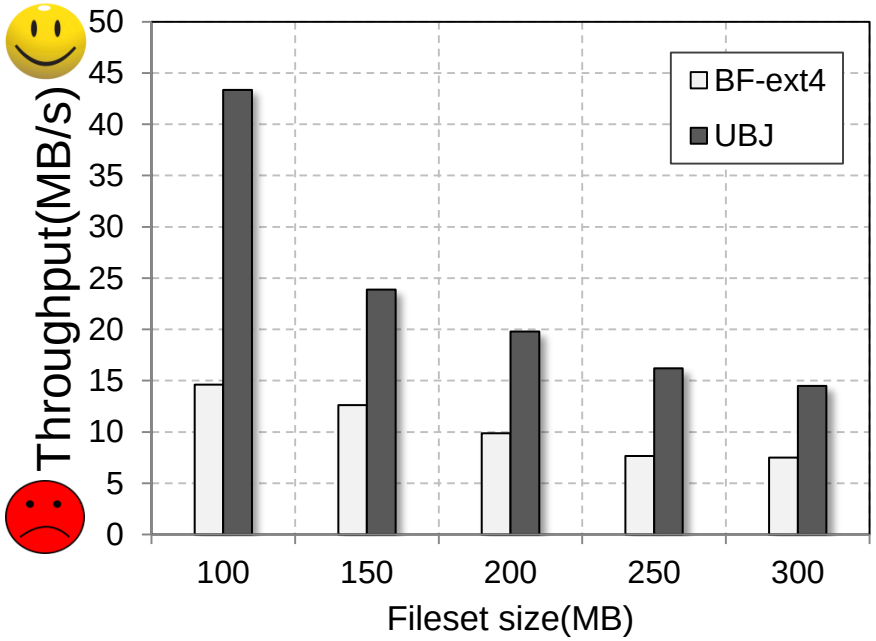




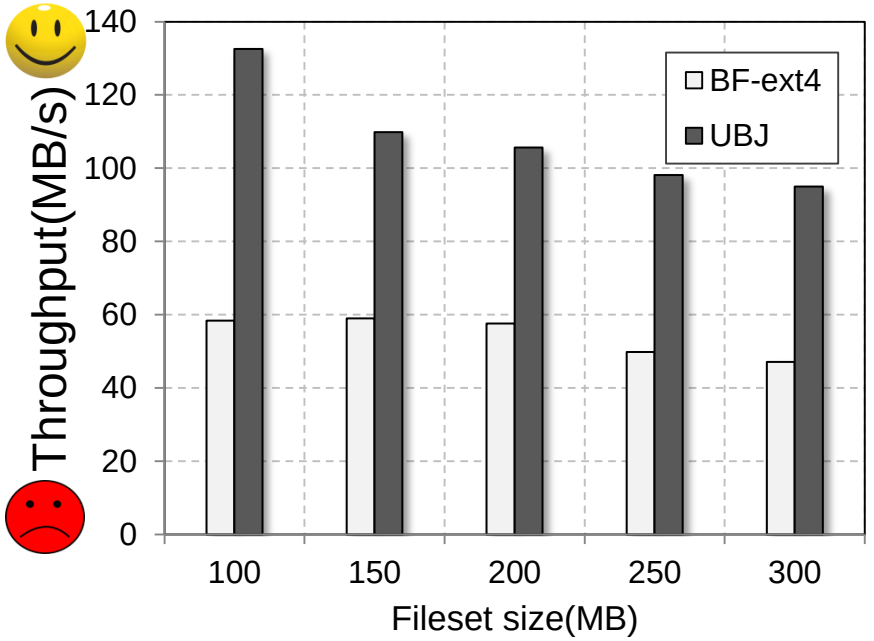
Improve execution time and throughput by 30.7% and 59.8% on average

Performance Evaluation

- IOzone



(a) Random write

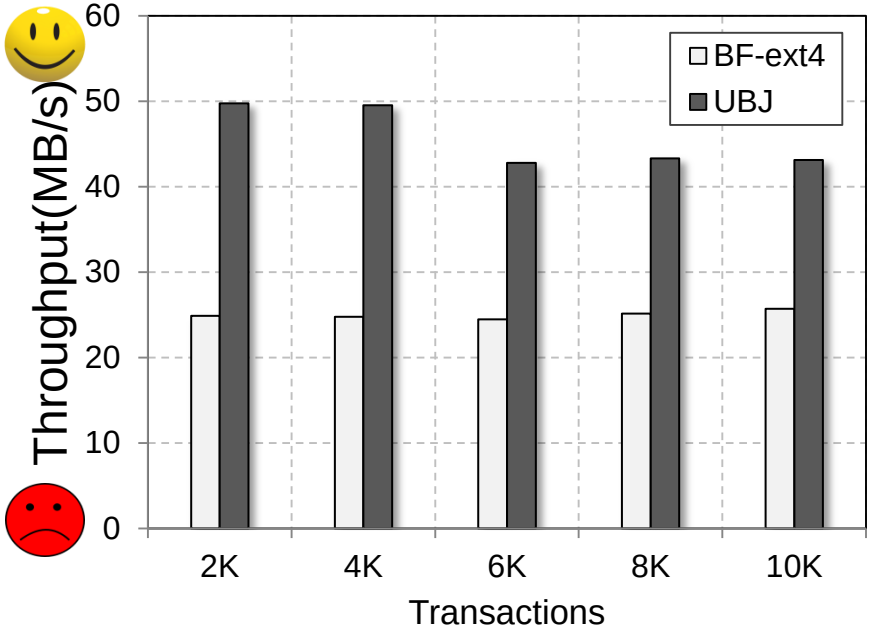


(b) Sequential write

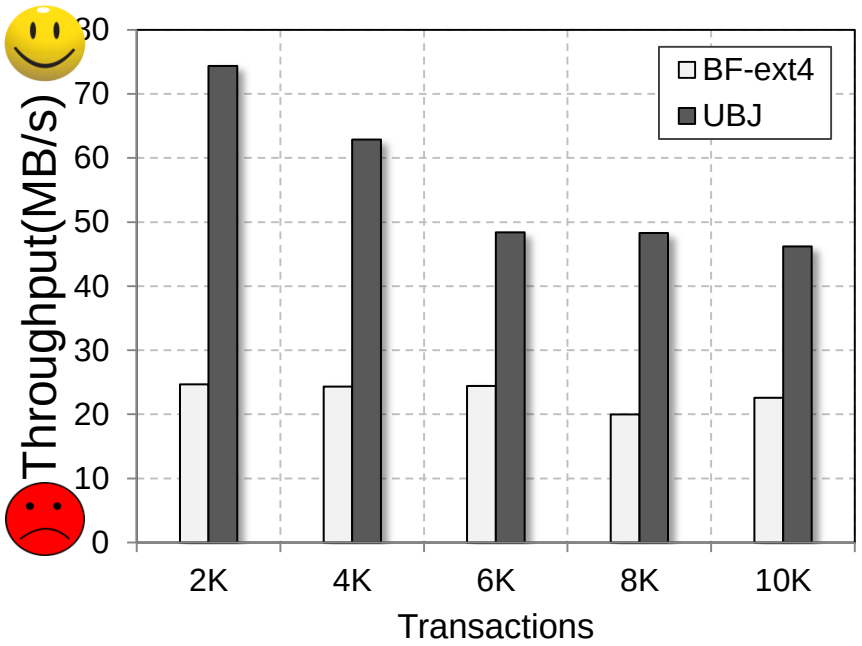
Improve performance by 110% on average, up to by 240%

Performance Evaluation

- Postmark



(a) Read

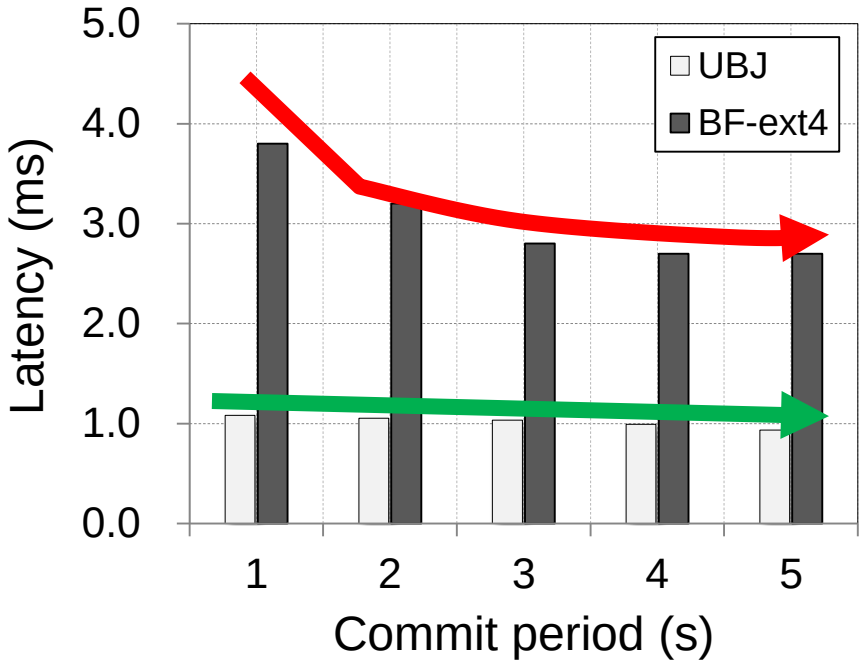


(b) Write

Improve performance by 109% on average

Performance Evaluation

- Effectiveness of UBJ on performance as the commit period changes



Latency of ext4 becomes smaller as the commit period is longer

Latency of UBJ is not sensitive to the commit period changes

Conclusion

- Novel non-volatile memory buffer cache architecture
- Subsumes the functions of caching and journaling
  - Buffer cache blocks ← → Journal logs
  - In-place Commit
    - Notion of a frozen state
- Performance results
  - Implemented on Linux 2.6.38
  - Compared to ext4 in journal mode
  - Improve I/O performance by 76% and up to 240%

Thank you

Eunji Lee
https://sites.google.com/site/alicia0729

Hyokyung Bahn
https://home.ewha.ac.kr/~bahn

Sam H. Noh
https://next.hongik.ac.kr