HPC2014

14R33003 – Jian Guo (Matsuoka Lab)

@ Tokyo Institute of Technology (2014/12/1 & 2014/12/8)

Today’ s Paper

UniFI: Leveraging Non-Volatile Memories for a Unified Fault Tolerance and Idle Power Management Technique

Somayeh Sardashti[1], David A. Wood[1]

1. University of Wisconsin-Madison

In Proceedings of the 26th ACM international conference on Supercomputing (ICS ‘12). ACM, New York, NY, USA, 59-68. DOI=10.1145/2304576.2304587http://doi.acm.org/10.1145/2304576.2304587

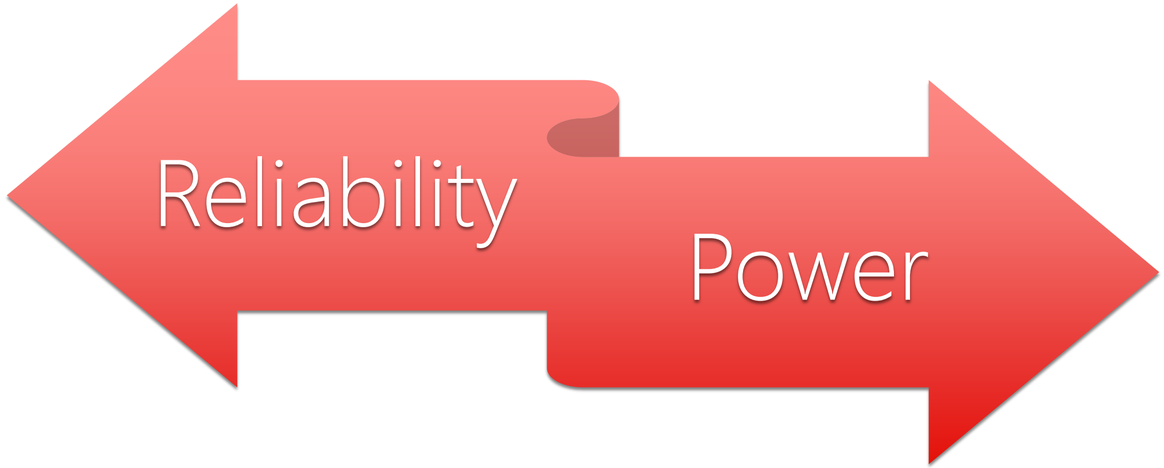
Outline

1. Introduction
2. Synergy between fault tolerance and power management
3. Background in non-volatile memory technologies
4. Unifi: a unified technique for fault tolerance and idle power management
5. Evaluation setup and Evaluation
6. Conclusions
7. Discussion

Contradiction for fault tolerance and power management

Decreasing device sizes ⟹ Increasing the likelihood of both transient and permanent faults

Increasing device count ⟹ Using more power



UniFI

- A technique for fault tolerance and idle power management in shared memory multi-core systems
- **Emerging** non–volatile memory technologies **to provide an energy–efficient lightweight checkpoint** mechanism to recover from a wide range of transient and permanent faults.

Very low performance and energy overheads

Introduction of UniFI

- The synergies between fault-tolerance and idle power management.
- **Using a novel combination of lazy flushing, in–cache logging, and safe replacement—that incurs low performance and energy overheads in the common case of fault–free execution, allowing frequent checkpointing.**
- UniFI has the unique characteristics of resistive memories to efficiently recover from a wide range of permanent and transient faults and to provide efficient idle power management.

Synergy between fault tolerance and power management

Table 1. Taxonomy of fault tolerance and power management techniques.

		System Reliability		
		Tends to hurt	**Neutral**	**Tends to help**
System Power	**Tends to hurt**	N/A	N/A	High-overhead global checkpointing [7],[24],[25], and redundancy [40] mechanisms
	Neutral	N/A	N/A	Power-aware reliability techniques [41], Rebound [45]
	Tends to help	Aggressive power management techniques [38][42]	Razor [39], MS-ECC [47], Word-disable and Bit-fix schemes [48], heterogeneous LLC [46]	UniFI [New]

Background in non-volatile memory technologies

Power has become a primary design constraint for multicore processors and systems.

- Phase-Change Memory (PCM)
- Spin-Torque Transfer Magneto-resistive RAM (STT-MRAM)

All promise scalability, non-volatility, high density, and energy-efficiency.

UniFI System Model

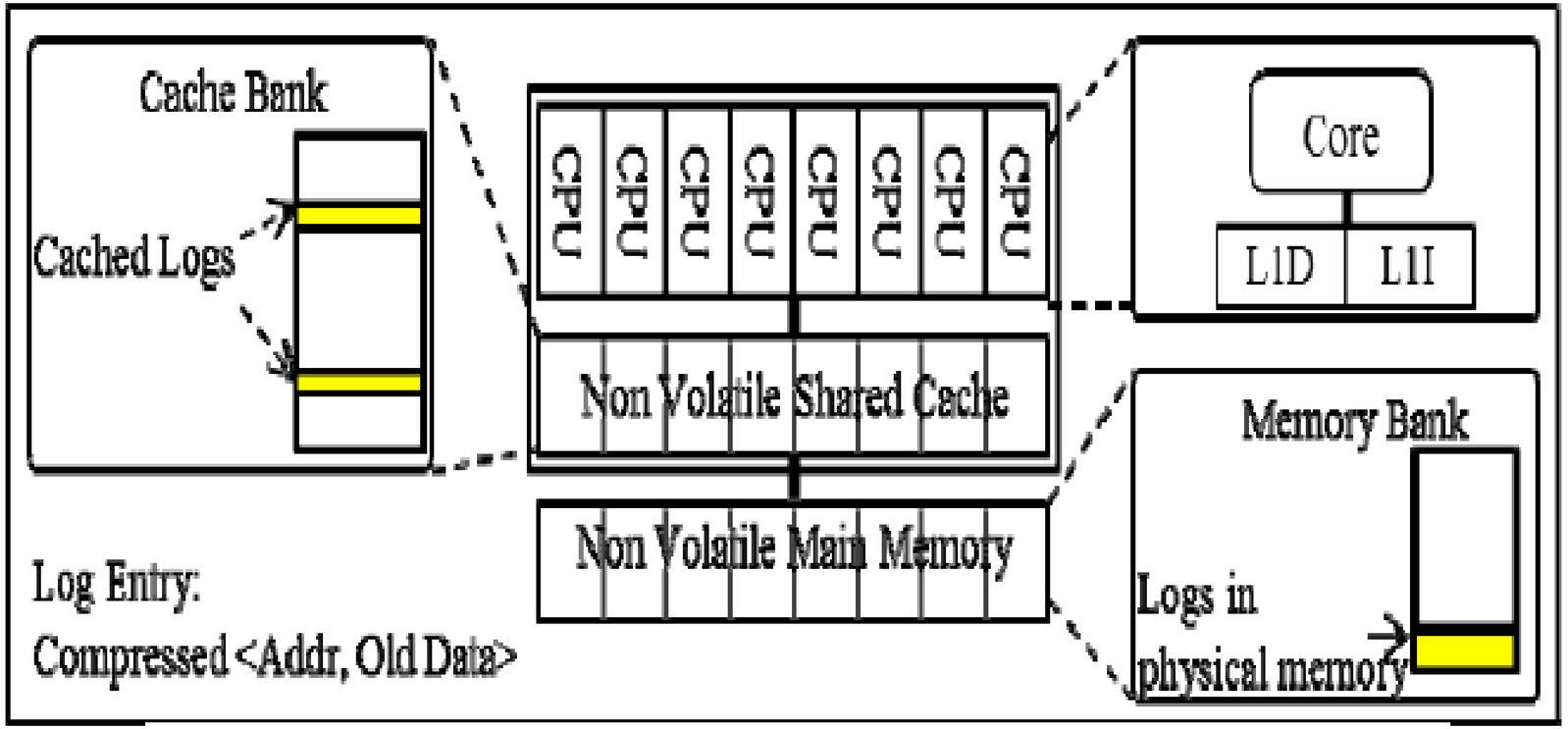


Figure 1. A general view of a multicore system with UniFI's support.

Idle Power Management Mechanisms

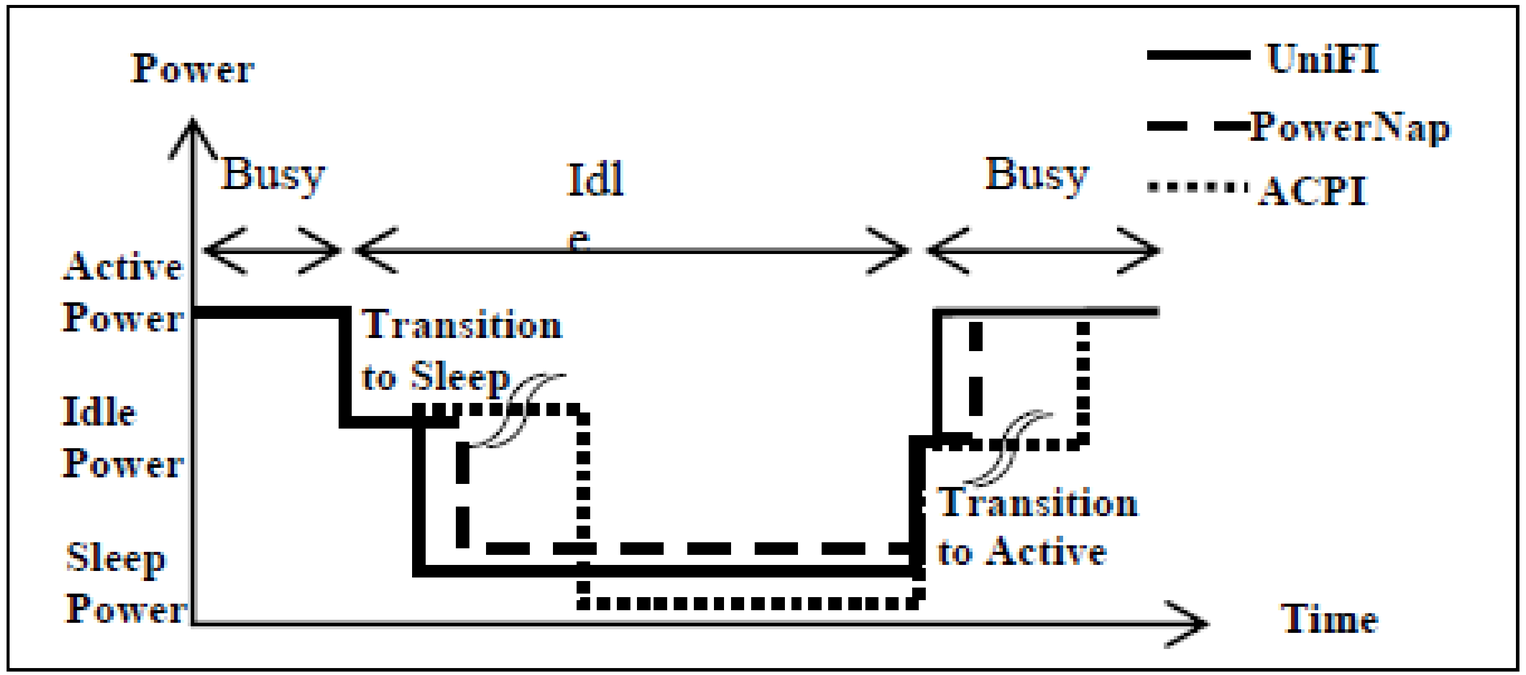


Figure 2. Different idle power management techniques.

Using non-volatile memory technologies and its efficient checkpointing mechanism to provide both fast transitioning and very low power system-level sleep mode.

UniFI Checkpointing Mechanisms

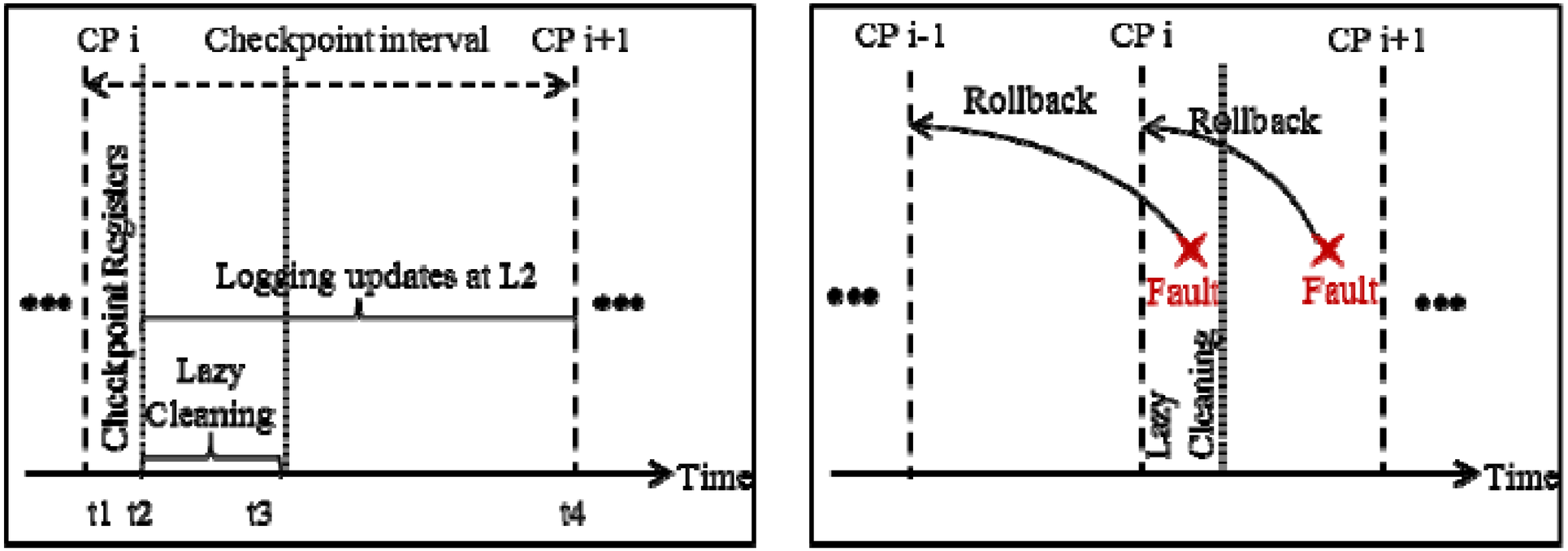


Figure 3. different phases of creating a checkpoint (a) and rollback recovery from different faults.

Checkpoint Mechanisms(Checkpointing at Processors)

UniFI Checkpointing Mechanism at caches

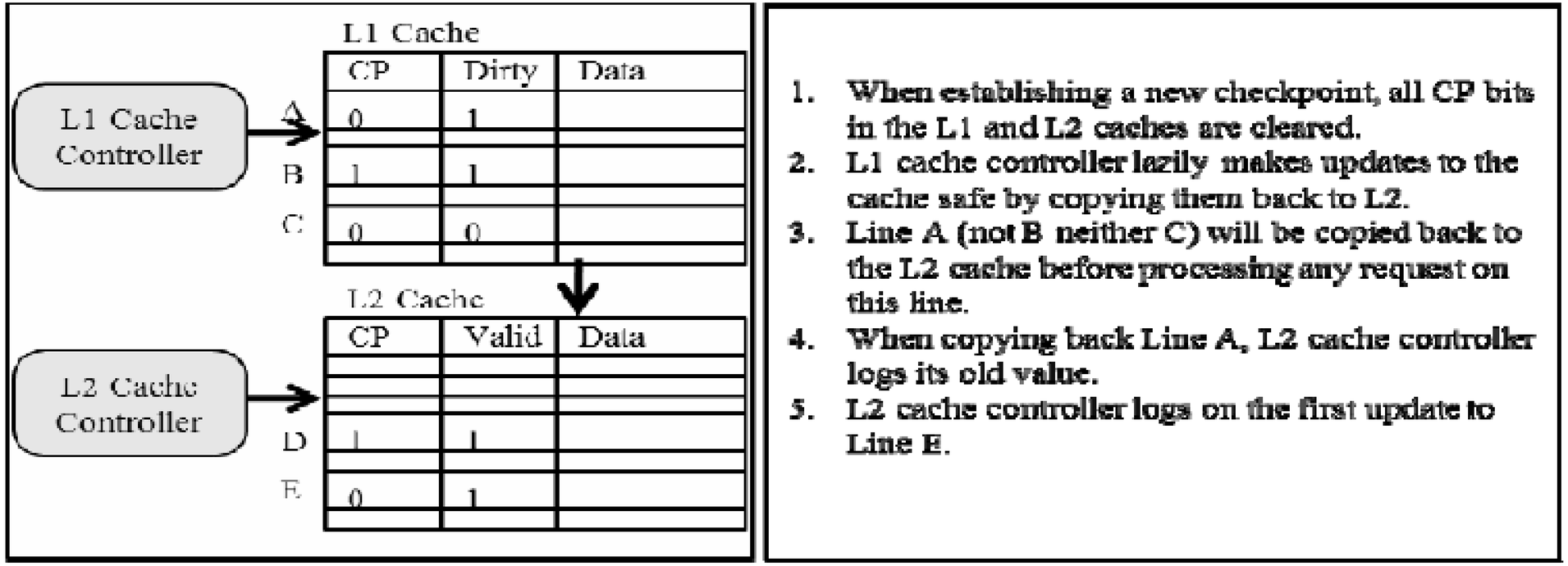


Figure 4. UniFI checkpointing mechanism.

Checkpoint Mechanisms (Logging Updates at the Shared Cache)

UniFI logs data updates to the L2 cache by storing their physical addresses and old data as log entries. It also caches logs in the L2 cache, in the same memory bank.

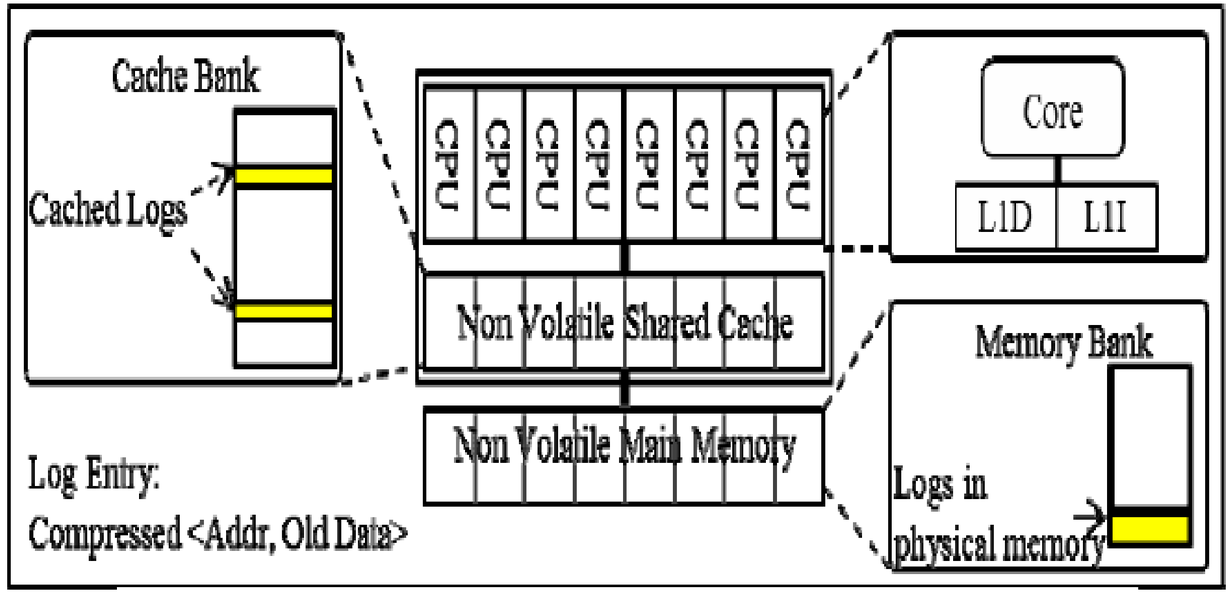
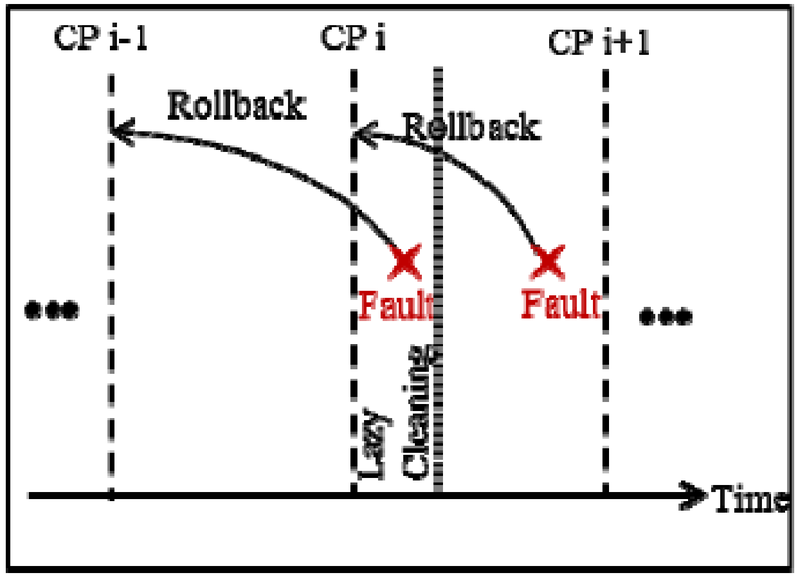


Figure 1. A general view of a multicore system with UniFI's support.

- UniFI avoids extra costs of special log buffers, high energy and performance overheads of storing logs in the main memory.
- Recovering is fast and energy efficient.

Checkpoint Mechanisms (Synchronization and Rollback Recovery)



Restore checkpoints in parallel

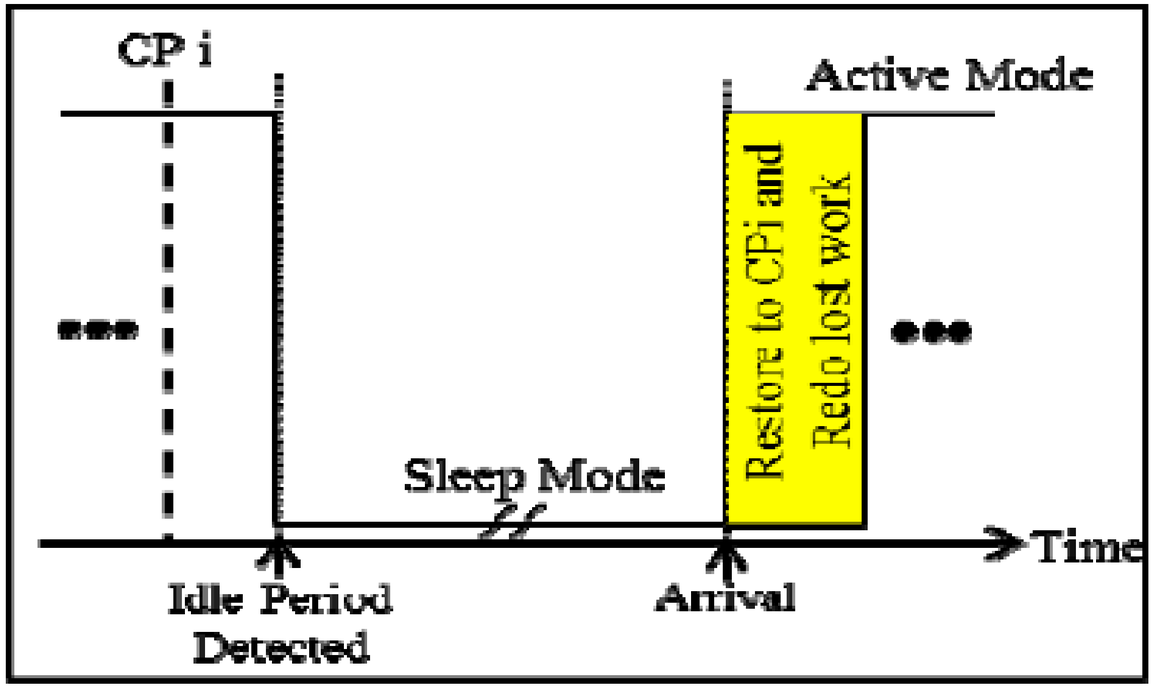
⇩

Reading and unrolling them via load/store units

⇩

Reads and undoes the checkpoint logs

UniFI Idle Power Management Mechanisms(a)

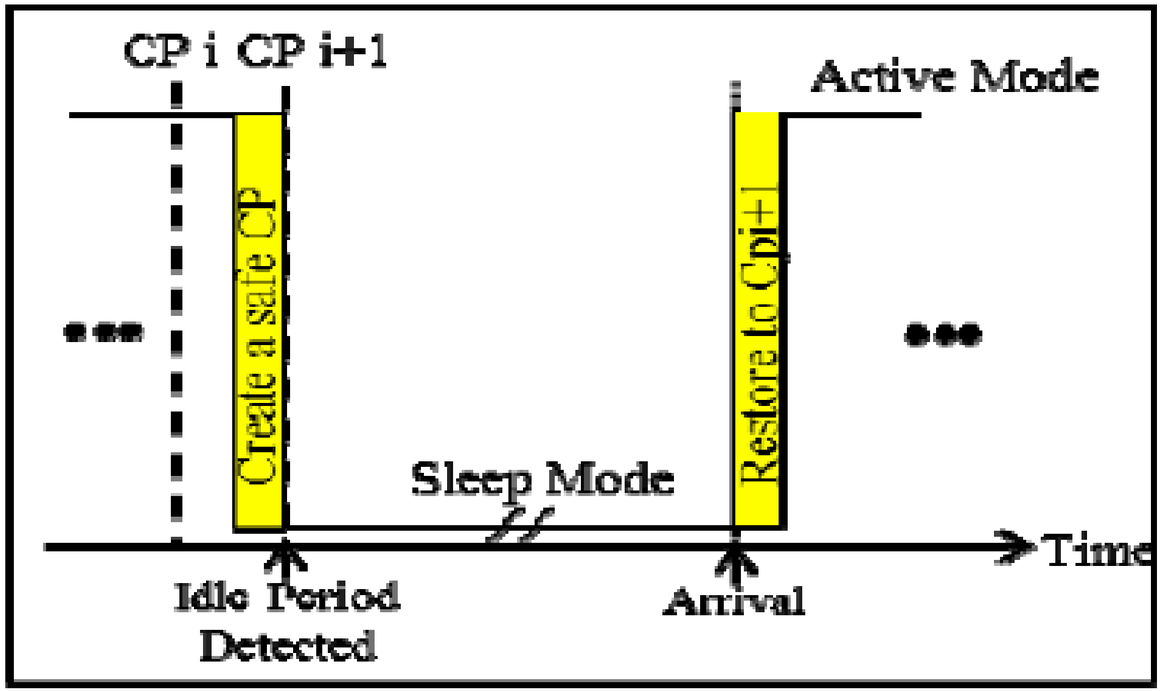


(a)

This mechanism is mostly useful for emergencies, such as thermal emergencies and power outage, since it does not let the previous job commit

power emergency or a detected idle period.

UniFI Idle Power Management Mechanisms(b)



(b)

no emergency

Evaluation setup

Table 2. Simulation Parameters

Cores	8 OOO cores, 4GHz frequency, 4-wide issue, 192 physical registers
L1 Caches	Private, 32-KB iL1/dL1, 4-way associative, 3-cycle access latency
L2 Caches	Shared, 4-MB, 8-way associative, 8 banks, 8-cycle Read latency, 24-cycle Write latency, MESI Coherency
Main Memory	4GB, 16 banks, 400MHz bus frequency
Checkpointing Parameters	2 checkpoints, 400K cycles (0.1ms) checkpointing interval

Table 3. STT-MRAM cache and PCM memory parameters

L2 Cache Parameters	**4MB SRAM**	**4MB STT-MRAM**
Rd/Wrt Latency (core cycle)	10/10	8/24
Energy per Rd/Wrt (pJ/64B)	1268/1268	798/952
Leakage Power (mW)	6578	3343
Memory Parameters	**DRAM**	**PCM**
tCL/tRCD/tWTR/tWR/tRTP/ tRP/tCCD/tWL (mem cycle)	5/5/3/6/3/5/ 4/4	5/22/3/6/3/60/4 /4
Energy per Rd/Wrt (pJ/bit)	1.17/0.39	2.47/16.82

Table 4. Component power consumption of a typical blade with/without power management techniques

Blade Components	Active Power	Idle Power	Sleep Power with PowerNap	Sleep Power with UniFI
CPU Chip	80-150W	12-20W	6.8W	0W
DRAM DIMMs	3.5-5W	1.8-2.5W	1.6W	0W
Other modules (PSU, SSD, etc.)	110-262W	210-230W	2W	2W
Total	**450W**	**270W**	**10.4W**	**2W**

Evaluation setup

PARSEC
(Blackscholes, Bodytrack, Canneal, Fluidanimate, Freqmine, Streamcluster, and Swaptions),

Benchmarks

SPEComp
(Ammp, Applu, Equake, Fma3d, Gafort, Mgrid, Swim, and Wupwise), and commercial workloads (Apache, Oltp, Jbb, and Zeus)

Evaluation (Baseline)

Table 5. Baseline Configurations

SRAM-DRAM	4-MB SRAM L2$, DRAM memory
STT-PCM w/o techs	4-MB STT-MRAM L2$, PCM memory
STT-PCM w techs (our baseline)	4-MB SRAM L2$, PCM memory, extra techniques applied
STT-PCM w large cache	8-MB SRAM L2$, PCM memory, extra techniques applied

⇦ SRAM-DRAM baseline

1. evaluate our baseline system and study impacts of using non-volatile memories.

2. evaluate UniFI’ s checkpointing mechanism and idle power management.

Performance and energy of baseline configurations normalized to SRAM-DRAM baseline system

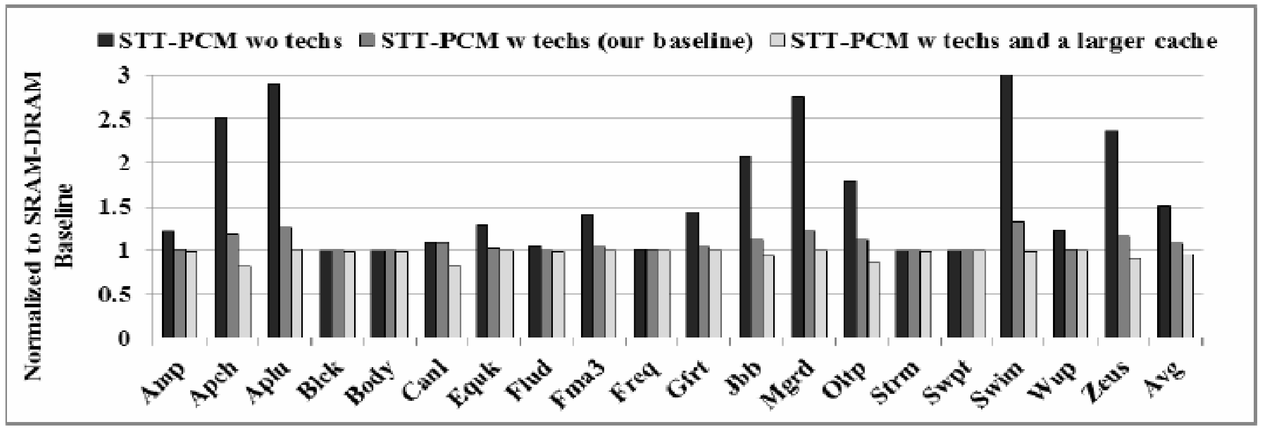


Figure 6. Performance of different baseline configurations.

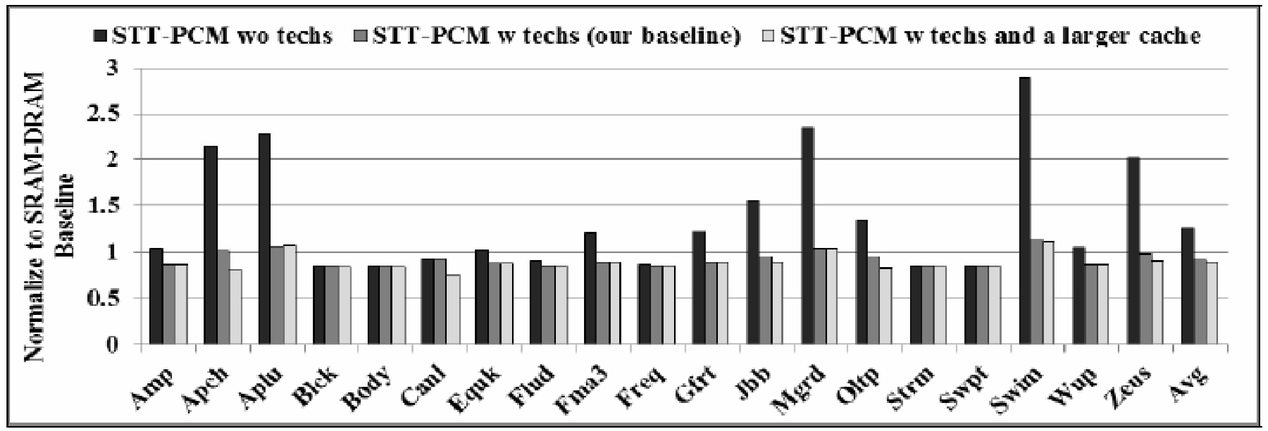


Figure 7. Energy of different baseline configurations.

STT-MRAM and PCM system without any extra technique (the second configuration) is on average 1.5x slower and requires 1.2x more energy than a SRAM-DRAM system.

Evaluation: UniFI

The main sources of UniFI's overheads are:

caching logs in the L2 cache &

cleaning L1 caches.

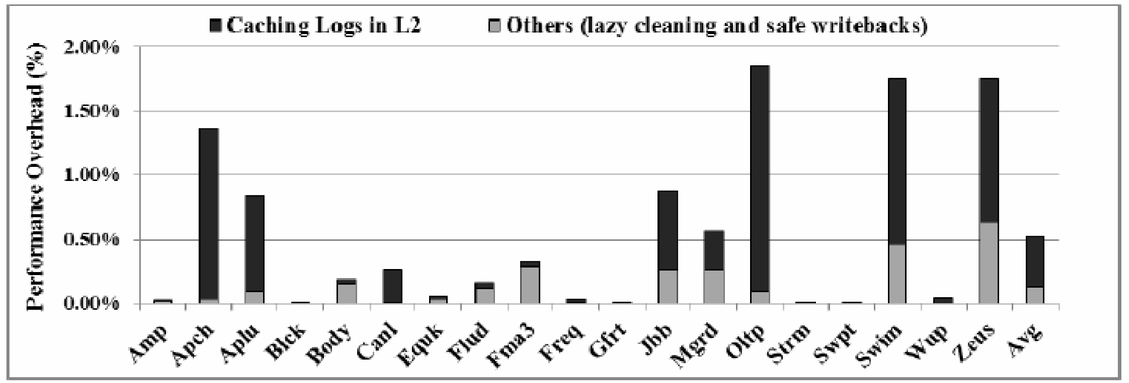


Figure 8. Breakdown of UniFI's performance overhead over the baseline.

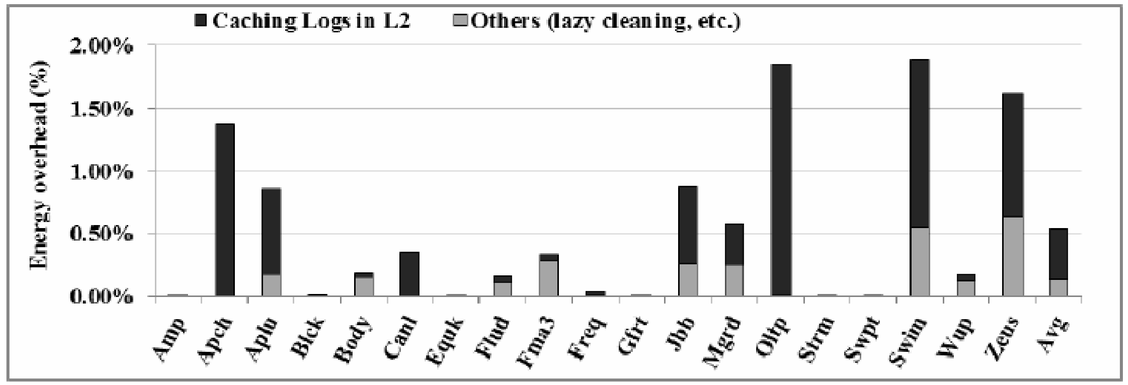


Figure 9. Breakdown of UniFI's energy overhead over the baseline.

Evaluation: UniFI’ s checkpointing mechanism

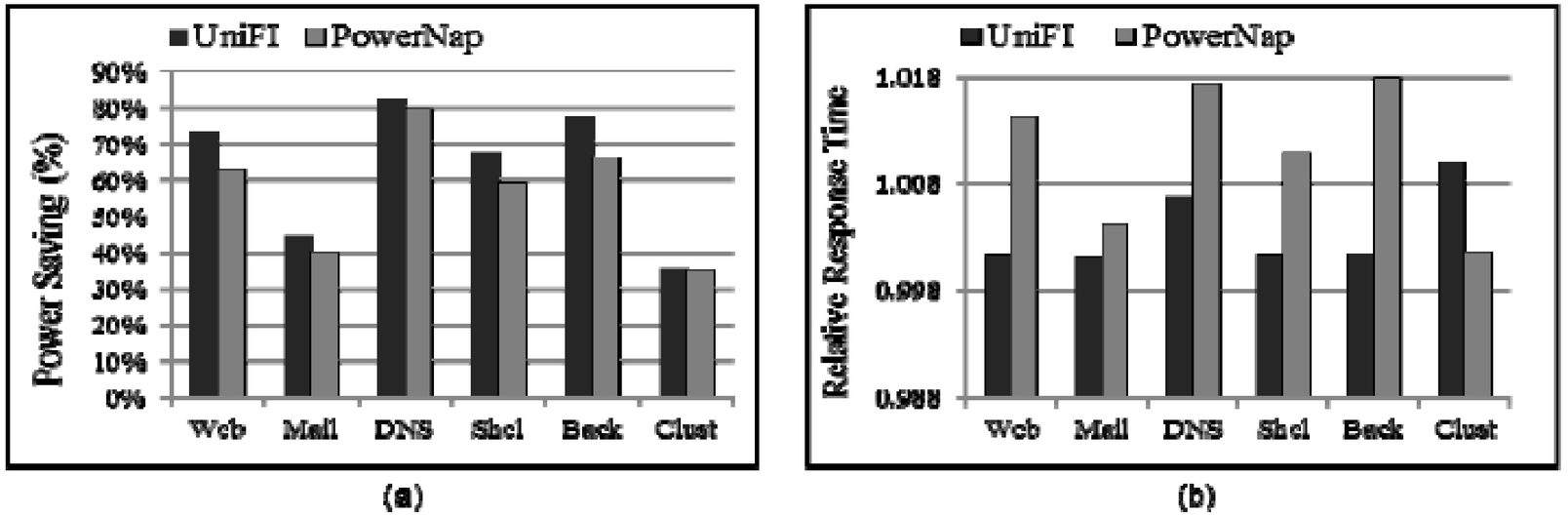


Figure 10. UniFI’s power savings and relative response time for server workloads Breakdown of UniFI’s

How it saves power by eliminating idle power.

Conclusions

- UniFI, a unified technique that addresses two critical challenges of reliability and power management together.
- UniFI less than 2% performance and energy overheads for a wide range of applications.
- UniFI can reduce average power by up to 82% by leveraging its low overhead checkpointing mechanism and non-volatile memories.

Discussion

- This is a coarse-grained system level checkpoints which may tolerate much higher checkpointing overheads.
- Not enough information about logging updates.
- The UniFI needs larger cache for better performance, however, more cache also lead to more power overhead.

Thanks for your listening!