Middleware and Flash Translation Layer Co-Design for the Performance Boost of Solid-State Drives

Chao Sun[1], Asuka Arakawa[1], Ayumi Soga[1], Chihiro Matsui[1] and Ken Takeuchi[1]

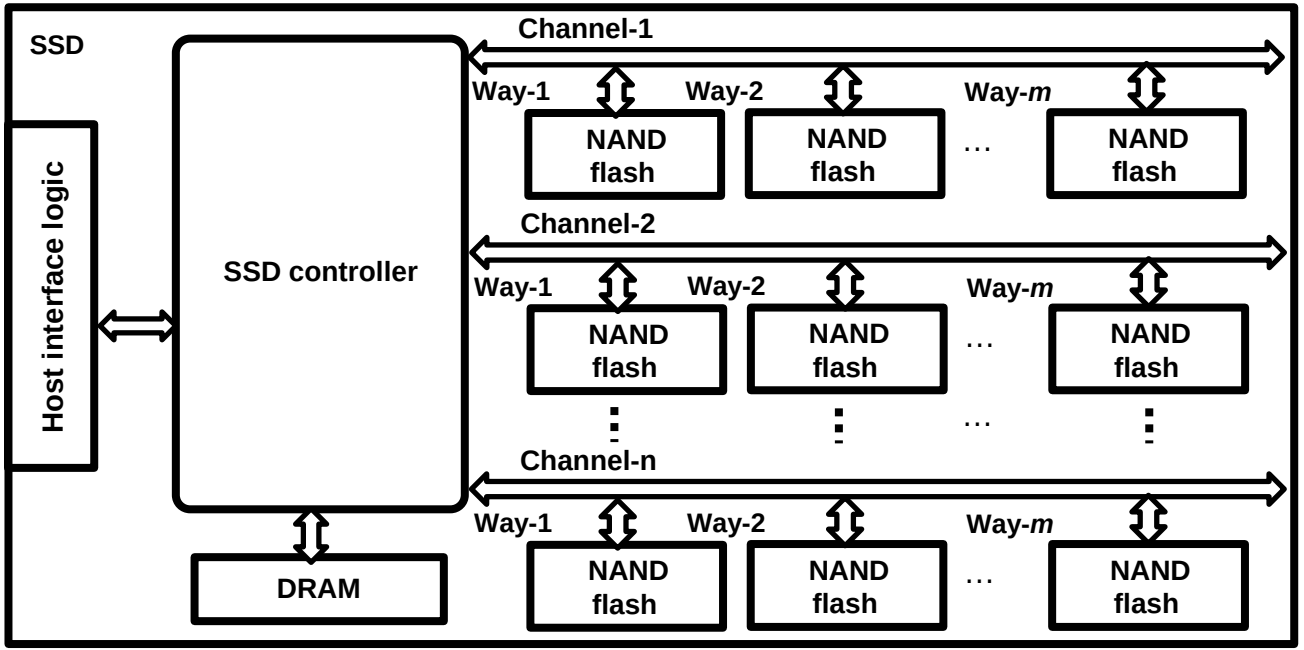
[1]Chuo University

Outline

- Introduction
- Three Techniques to Boost SSD Performance
  - Storage Engine Assisted SSD（SEA-SSD)
  - Logical Block Address (LBA) Scrambler
  - Storage Class Memory (SCM)/NAND Flash Hybrid SSD
- Conclusion

Solid-State Drive

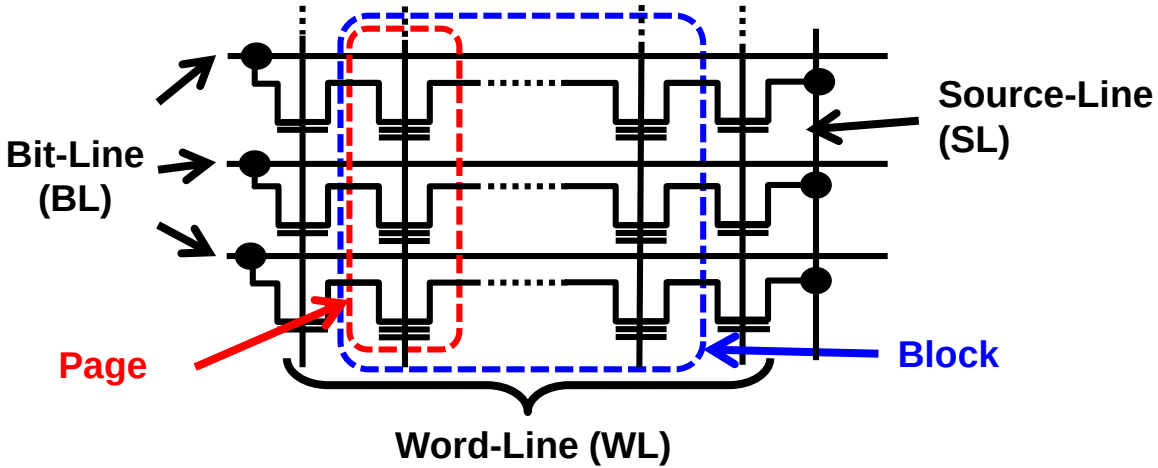
- **SSD is composed of NAND flash memories.**
- **SSD controller manages the data storage.**



[1] C. Sun et al., *TCAS-I*, vol. 62, no. 1, pp. 120-129, Jan. 2015.

NAND Flash Memory

- **The read/write unit of the NAND flash memory is a *page*.**
- **The erase unit of the NAND flash memory is a *block*.**
- **In-place overwrite is prohibited: overwrite involves a page read + a page write.**

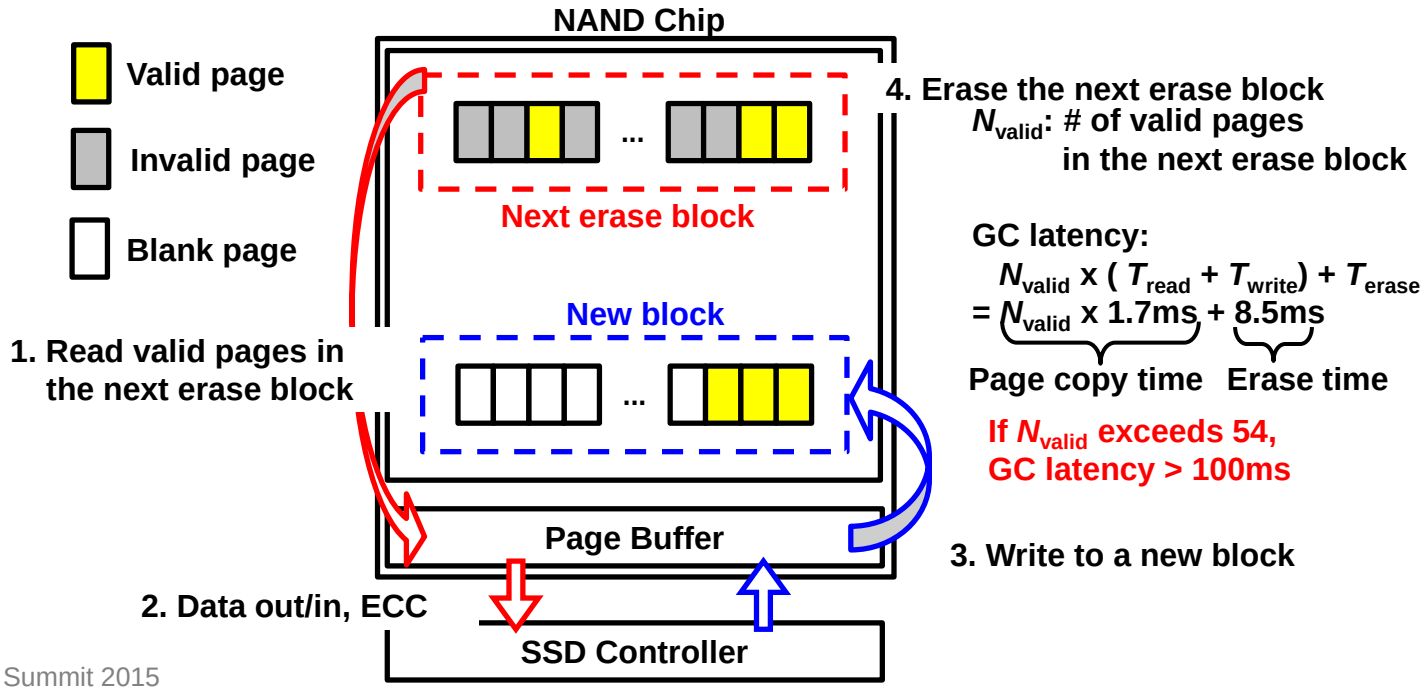


Frequent overwrite create massive number of invalid pages.

[1] K. Takeuchi, *JSSC*, vol. 44, no. 4, pp. 1227-1234, 2009.

Garbage Collection of SSD

- **GC is triggered when SSD runs out of free spaces.**
- **GC can be the bottleneck of SSD write throughput.**



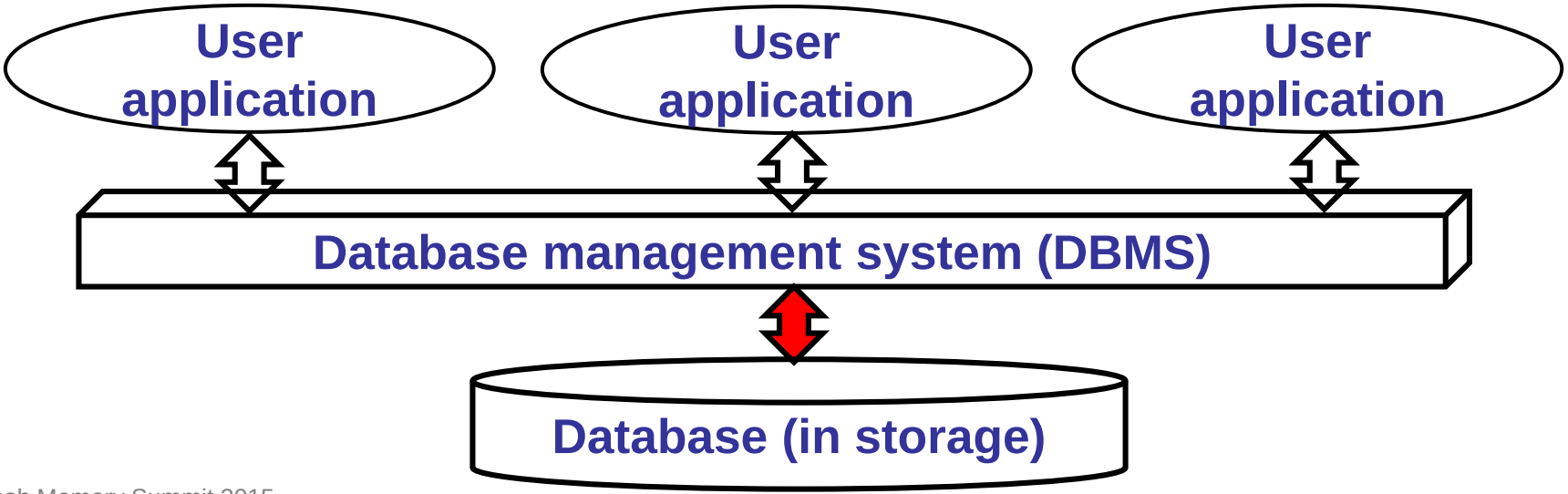
[1] K. Takeuchi, *JSSC*, vol. 42, no. 1, pp. 219-232, 2007.

Outline

- Introduction
- Three Techniques to Boost SSD Performance
  - Storage Engine Assisted SSD（SEA-SSD)
  - Logical Block Address (LBA) Scrambler
  - Storage Class Memory (SCM)/NAND Flash Hybrid SSD
- Conclusion

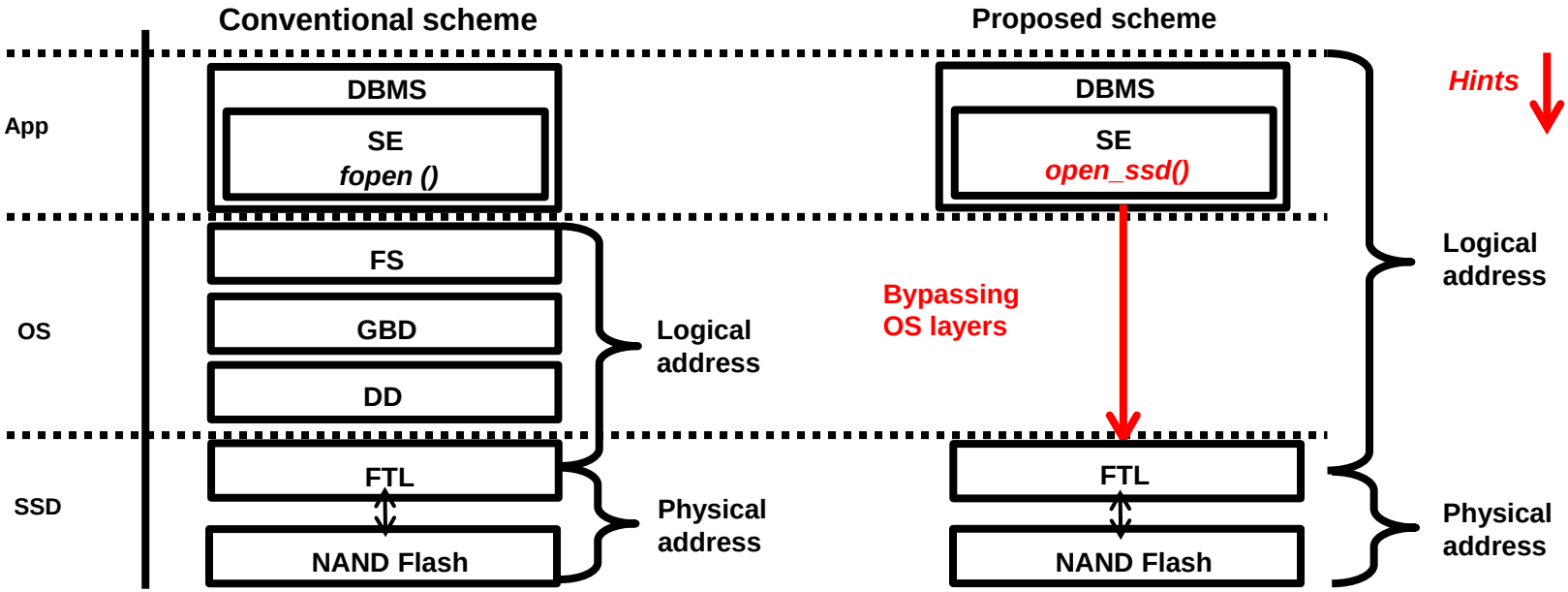
Database Management system (DBMS)

- To solve the IO bottleneck of the DBMS system, high-speed and low-power storage is required.
- Solid-state drive (SSD) is a good fit.



Storage Engine Assisted SSD (SEA-SSD)

- **OS is bypassed to reduce the communication overhead.**

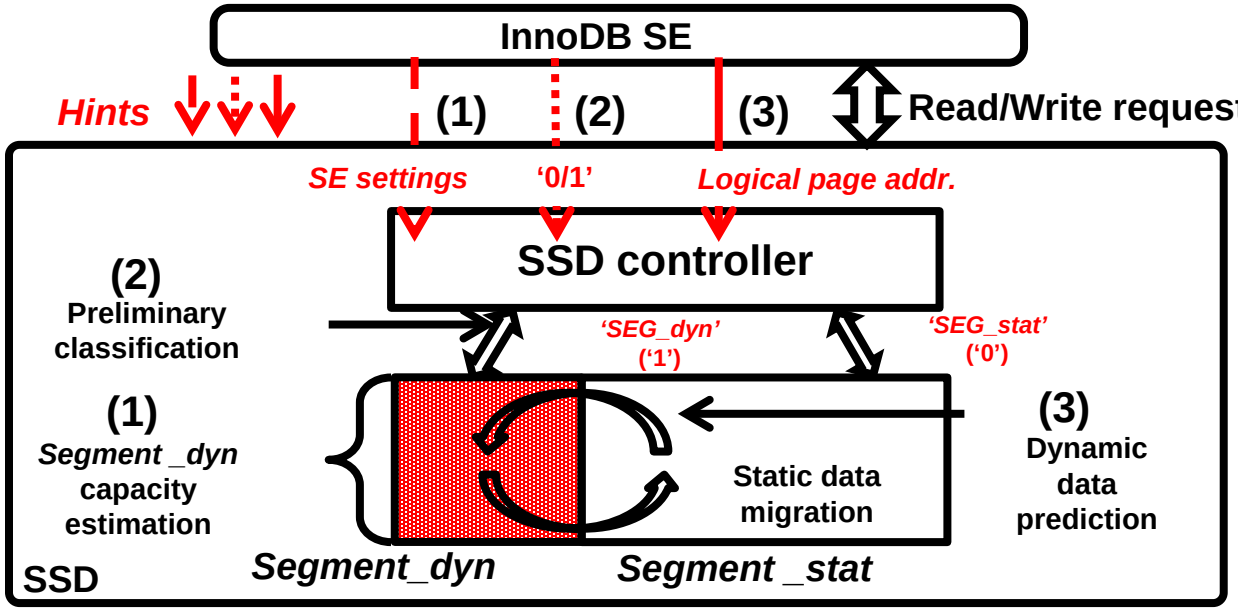


App: Application, DBMS: database management system,
SE: Storage engine, FS: File system, GBD: Generic block device,
DD: Device driver, FTL: Flash translation layer (SSD controller)

[1] C. Sun et al., *TCAS-I*, vol. 62, no. 1, pp. 120-129, Jan. 2015.

SEA-SSD Architecture

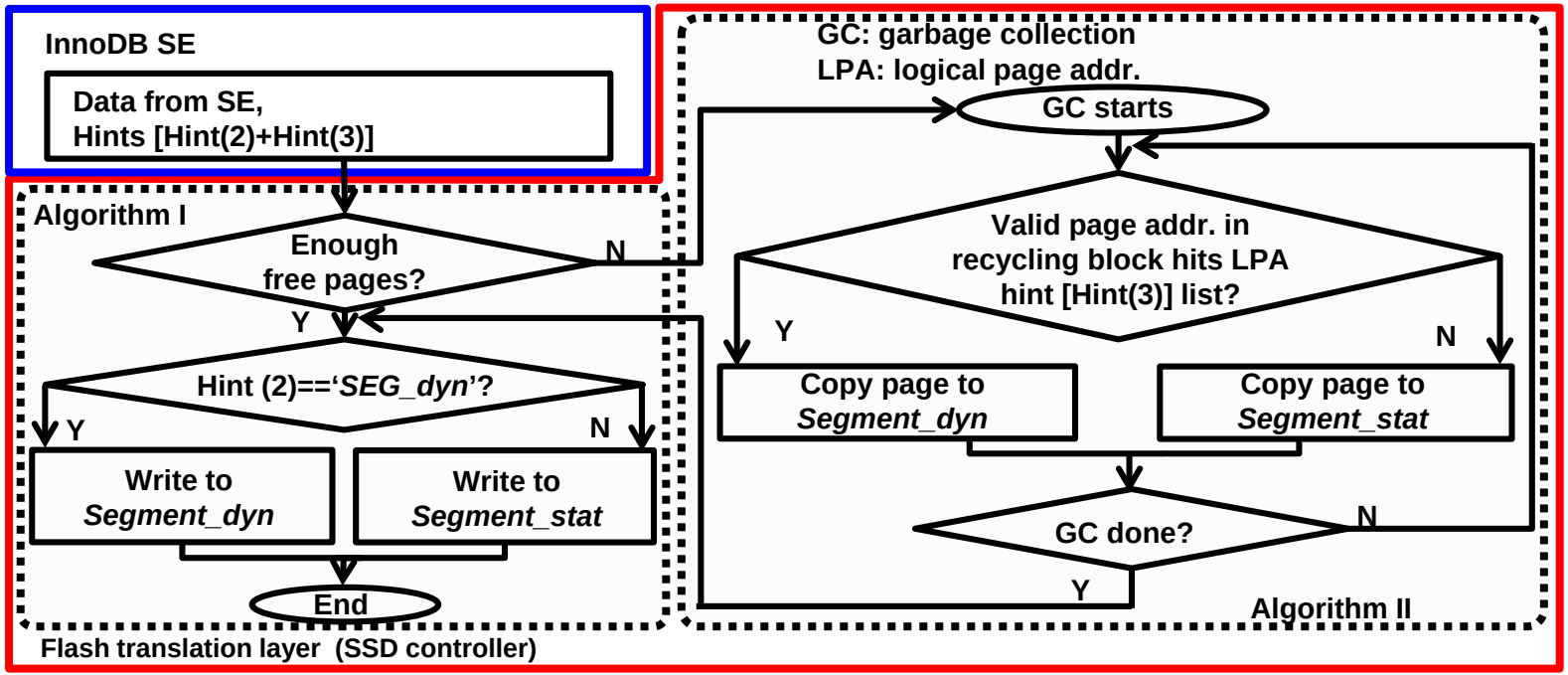
- Three hints are passed from the storage engine (SE) to SSD controller.
- Data are classified by activity and stored to the dynamic and static segments of SSD, respectively.



[1] C. Sun et al., *TCAS-I*, vol. 62, no. 1, pp. 120-129, Jan. 2015.

SEA-SSD Management Algorithms

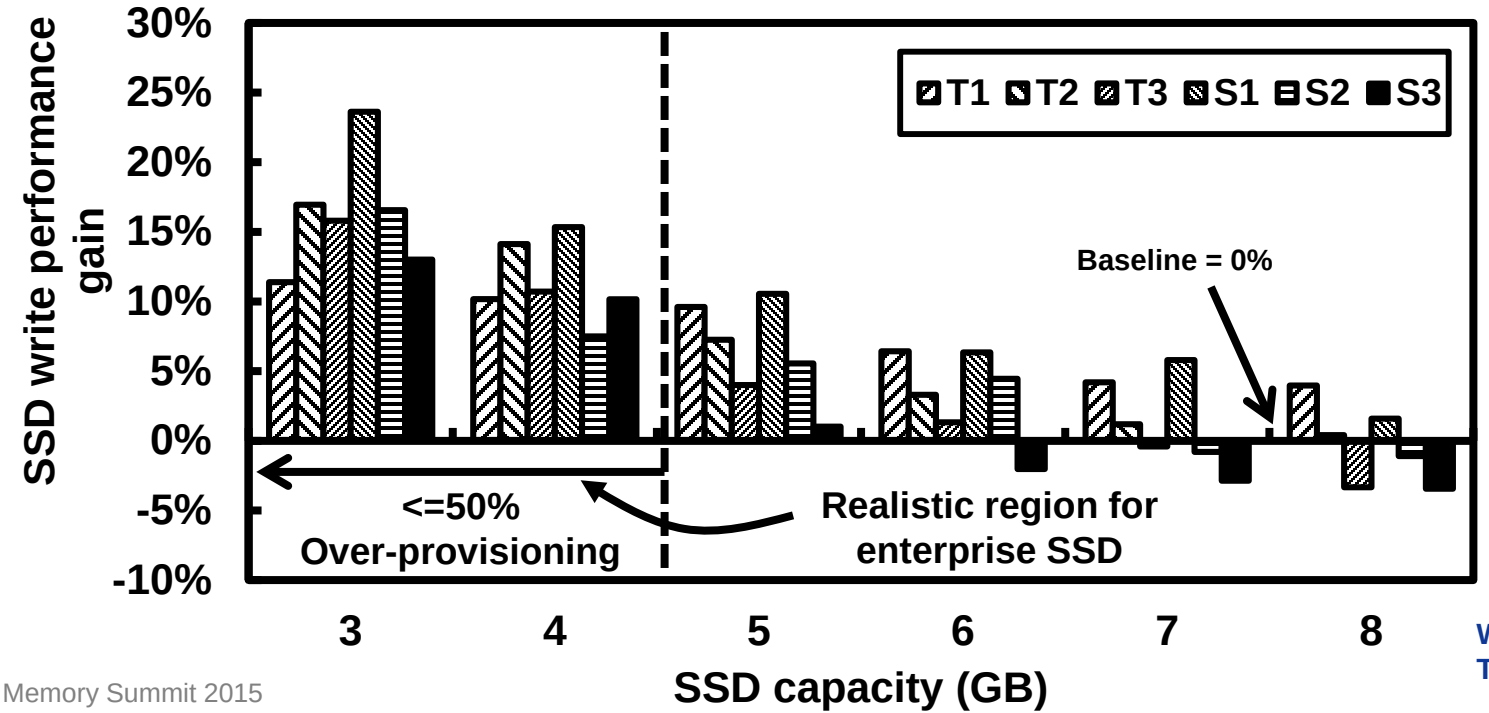
- **Data activity is judged by Hint (2) and Hint (3) from SE.**
- **Dynamic data -> *Segment_dyn*; Static data -> *Segment_stat.***



[1] C. Sun et al., *TCAS-I*, vol. 62, no. 1, pp. 120-129, Jan. 2015.

SSD Performance Evaluation

- **Over 10% performance improvement is achieved for all workloads at 3 GB SSD capacity (33% NAND flash over-provisioning).**



Workload:
TPC-C & Sysbench

[1] C. Sun et al., *TCAS-I*, vol. 62, no. 1, pp. 120-129, Jan. 2015.

Short Summary

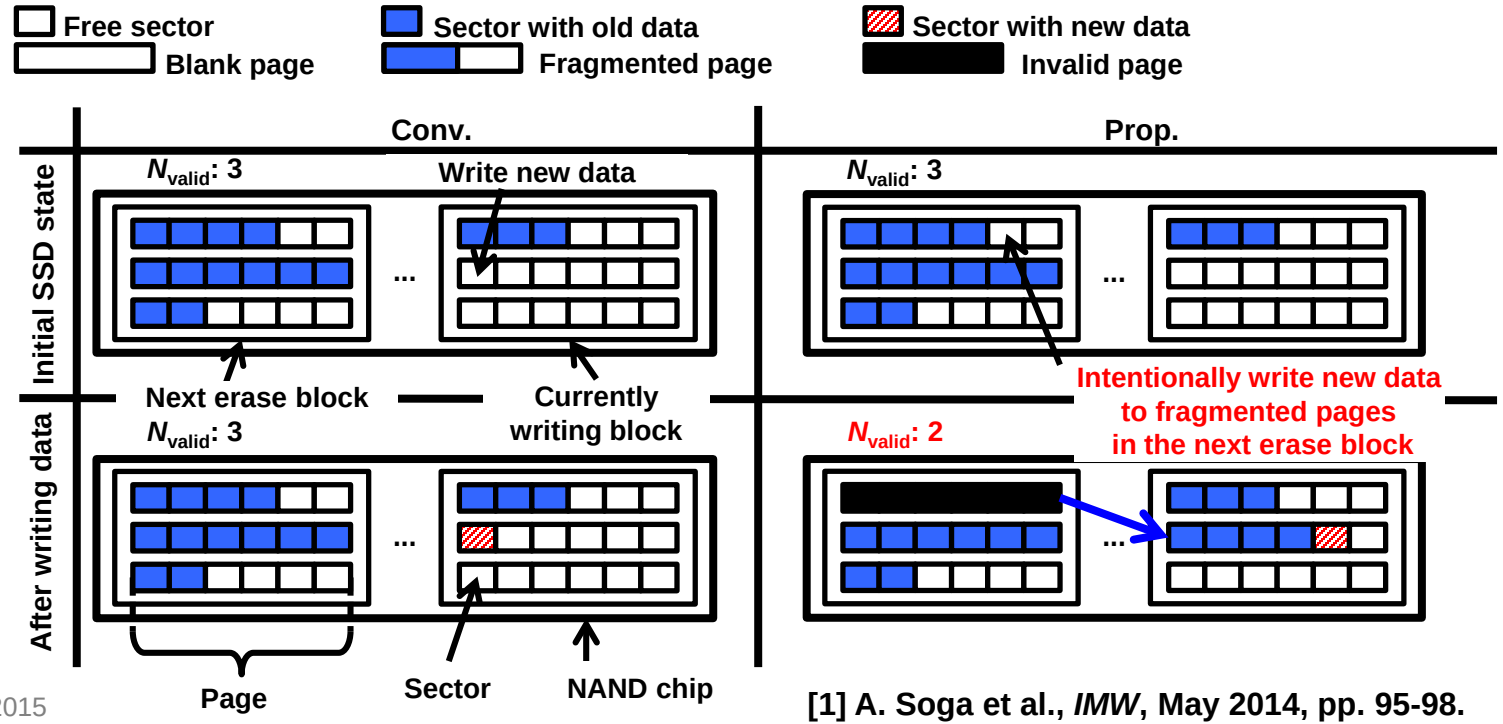
- SEA-SSD has been proposed to improve the write performance of the SSD for the database application.
- Hint information is passed from the SE to SSD controller for better classifying and predicting the data activity.
- A database application coupled simulation platform has been developed for such a cross-layer work, which accelerates the simulation speed by over 20-times, compared with the all virtualized simulation platform.
- Max. 24% performance improvement, 16% energy consumption reduction and 19% lifetime extension are achieved without requiring a cache layer for the SSD.

Outline

- Introduction
- Three Techniques to Boost SSD Performance
  - Storage Engine Assisted SSD（SEA-SSD)
  - Logical Block Address (LBA) Scrambler
  - Storage Class Memory (SCM)/NAND Flash Hybrid SSD
- Conclusion

LBA scrambler-Concept

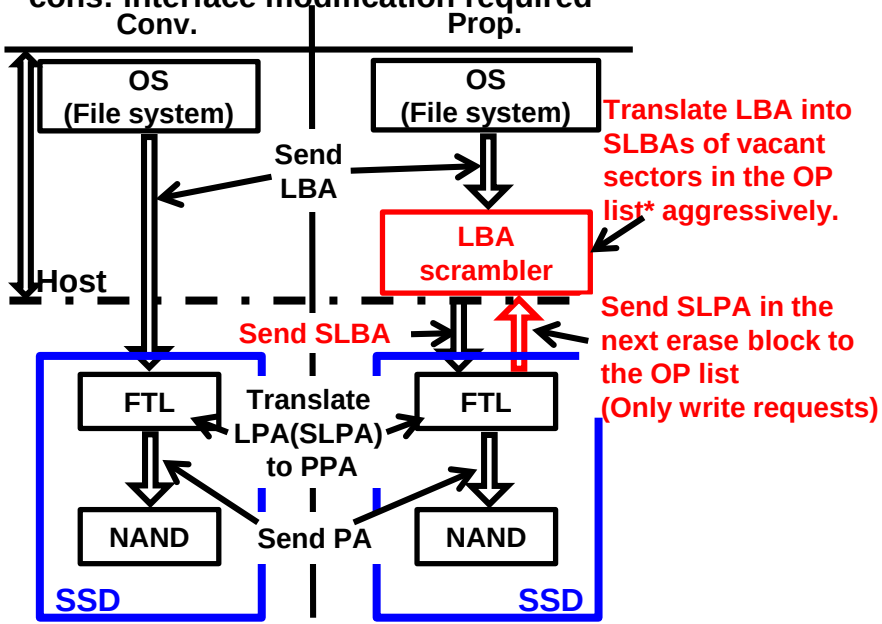
- **Actively write small data to fragmented pages (space utilization < 100%) so as to minimize the GC page-copy overhead.**



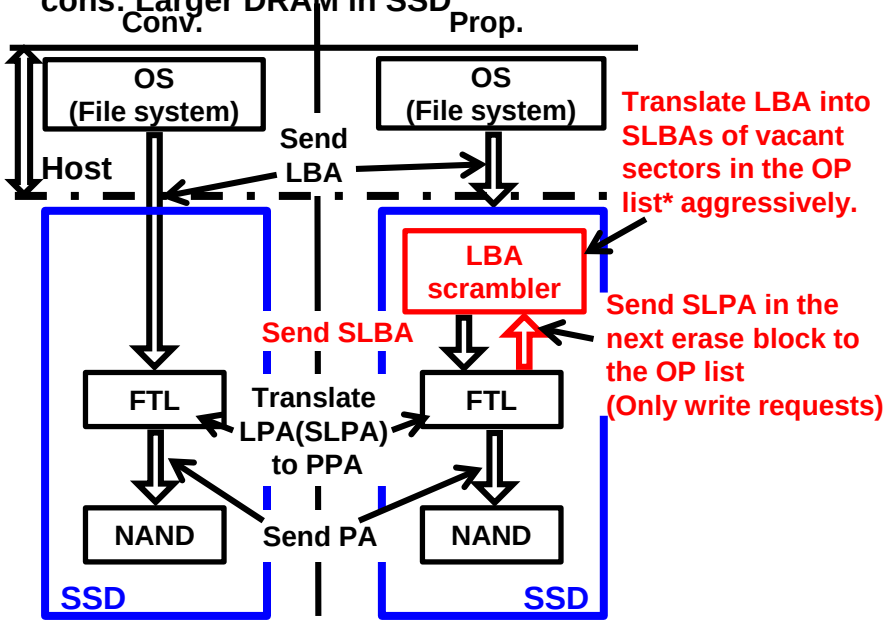
[1] A. Soga et al., *IMW*, May 2014, pp. 95-98.
[2] C. Sun et al., *TVLSI*, in-press.

LBA Scrambler-System View

(a) LBA scrambler in the host
pros: smaller DRAM in SSD
cons: interface modification required



(b) LBA scrambler in SSD
pros: no need to modify the interface
cons: Larger DRAM in SSD



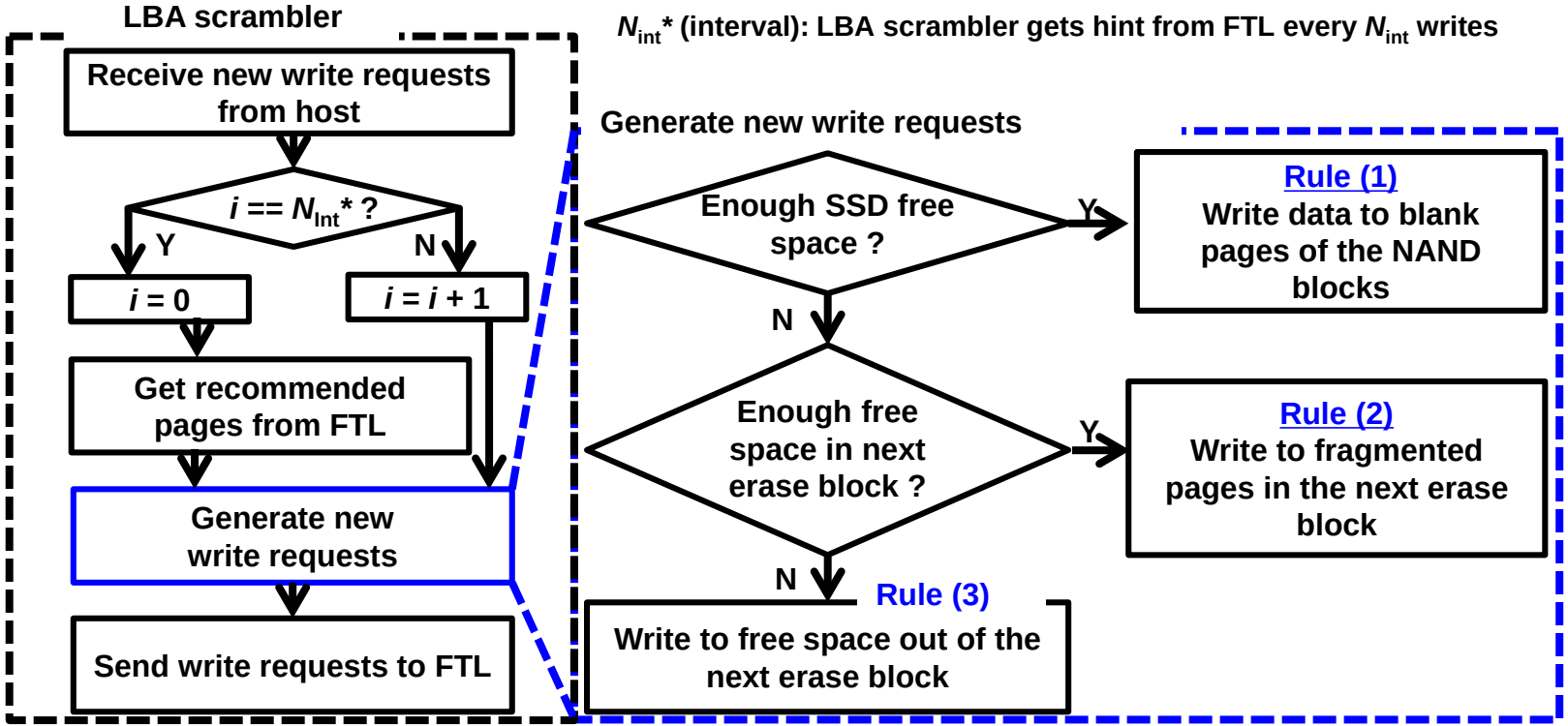
*Recommended writing pages are stored in the OP list
LBA: Logical block address SLBA: Scrambled LBA
SLPA: Scrambled LPA PA: Physical address
OP list : overwrite_preferred list

*Recommended writing pages are stored in the OP list
LPA: Logical page address
PPA: Physical page address
FTL: Flash translation layer(SSD controller)

[1] C. Sun et al., *TVLSI*, in-press.

LBA Scrambler-Algorithm Flow

- **LBA Scrambler is based on the address-remapping with three rules.**



[1] C. Sun et al., *TVLSI,* in-press.

Short Summary

- **LBA scrambler is able to improve the SSD write performance for all applications.**
- **With the LBA scrambler, small data are actively written to the remaining space of the fragmented page (the page utilization is less than 100%) in the next erase block of the SSD.**
- **Thus, the page-copy overhead is reduced when the GC is triggered.**
- **As a result, 35%-394% write speed boost, 27%-56% energy consumption reduction and 25%-55% endurance enhancement are achieved.**

Outline

- Introduction
- Three Techniques to Boost SSD Performance
  - Storage Engine Assisted SSD（SEA-SSD)
  - Logical Block Address (LBA) Scrambler
  - Storage Class Memory (SCM)/NAND Flash Hybrid SSD
- Conclusion

Storage Class Memory (SCM)

- **Memory overview**

	Mature		Prototypical memory		Emerging
	DRAM	NAND flash	STT-MRAM	PCM	ReRAM
Feature size (nm)	36	16	65	45	<40
Cell area	$6F^2$	$4F^2$	$20F^2$	$4F^2$	$4F^2$
Read latency	<10ns	0.1ms	35ns	12ns	N/A
W/E latency	<10ns	0.1/1ms	35ns	100ns	<10ns
Retention	64 ms	10 year	>10year	>10year	>=10year
Endurance	>1E16	1E5	>1E12	1E9	>1E6
W/E voltage (v)	2.5	15-20	1.8	3	<3
Read voltage (v)	1.8	4.5	1.8	1.2	<0.5
Single cell write energy (J/bit)	4E-15	4E-16	2.5E-12	6E-12	<=1E-12

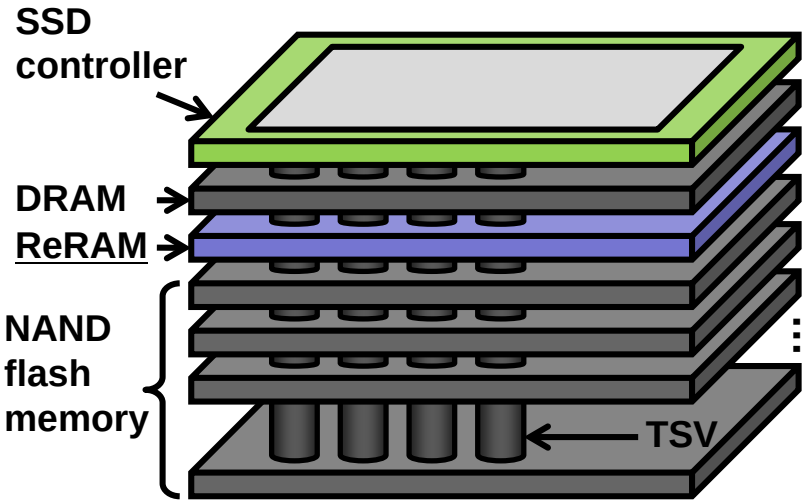
SCM

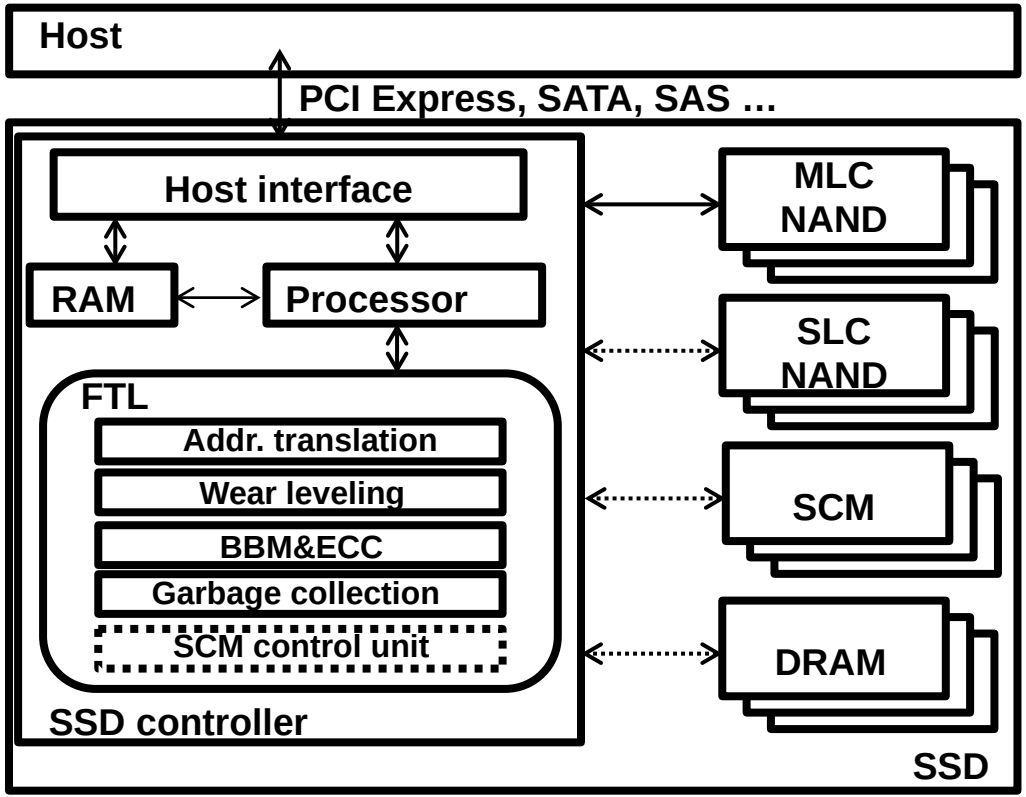
Limited endurance

[1] International Technology Roadmap for Semiconductors. http://www.itrs.net/Links/2013ITRS/Summary2013.htm

SCM/NAND Flash Hybrid SSD

- A 3D TSV-integrated SCM/NAND flash hybrid SSD.
- DRAM is used to store management tables.

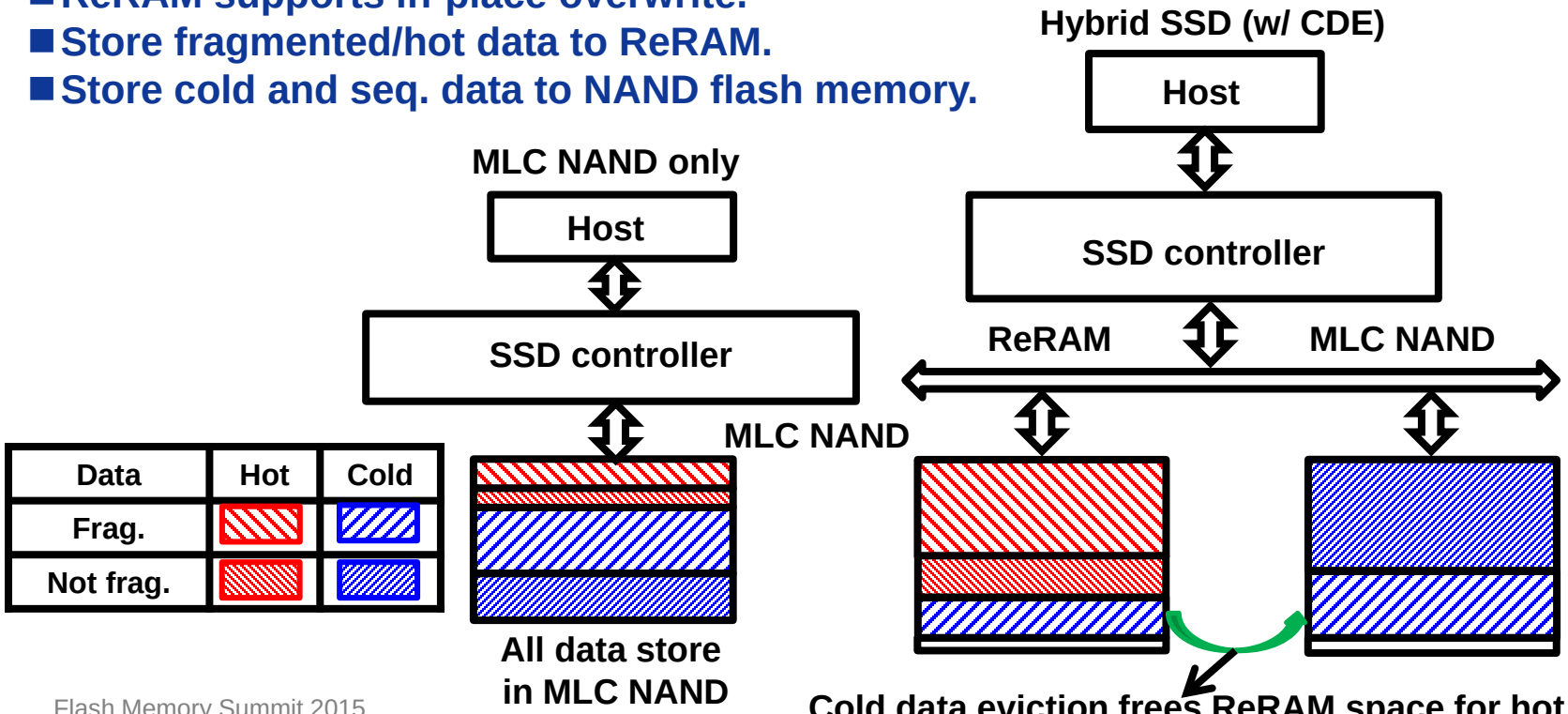




[1] H. Fujii et al., *VLSI Circuits*, 2012, pp. 134-135.
[2] C. Sun et al., *TCAS-I*, vol. 62, no. 2, pp. 382-392, 2014.

Data Storage in Hybrid SSD

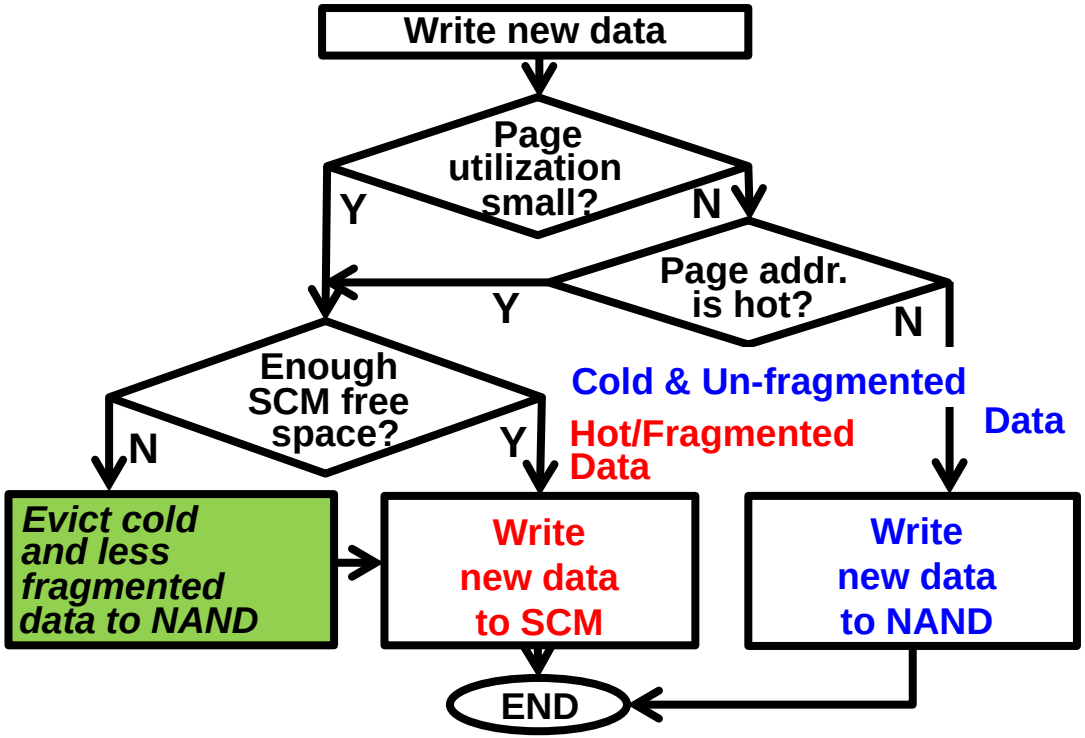
- ReRAM supports in-place overwrite.
- Store fragmented/hot data to ReRAM.
- Store cold and seq. data to NAND flash memory.



[1] C. Sun et al., *TCAS-I*, vol. 62, no. 2, pp. 382-392, 2014.

SSD Data Management Algorithm

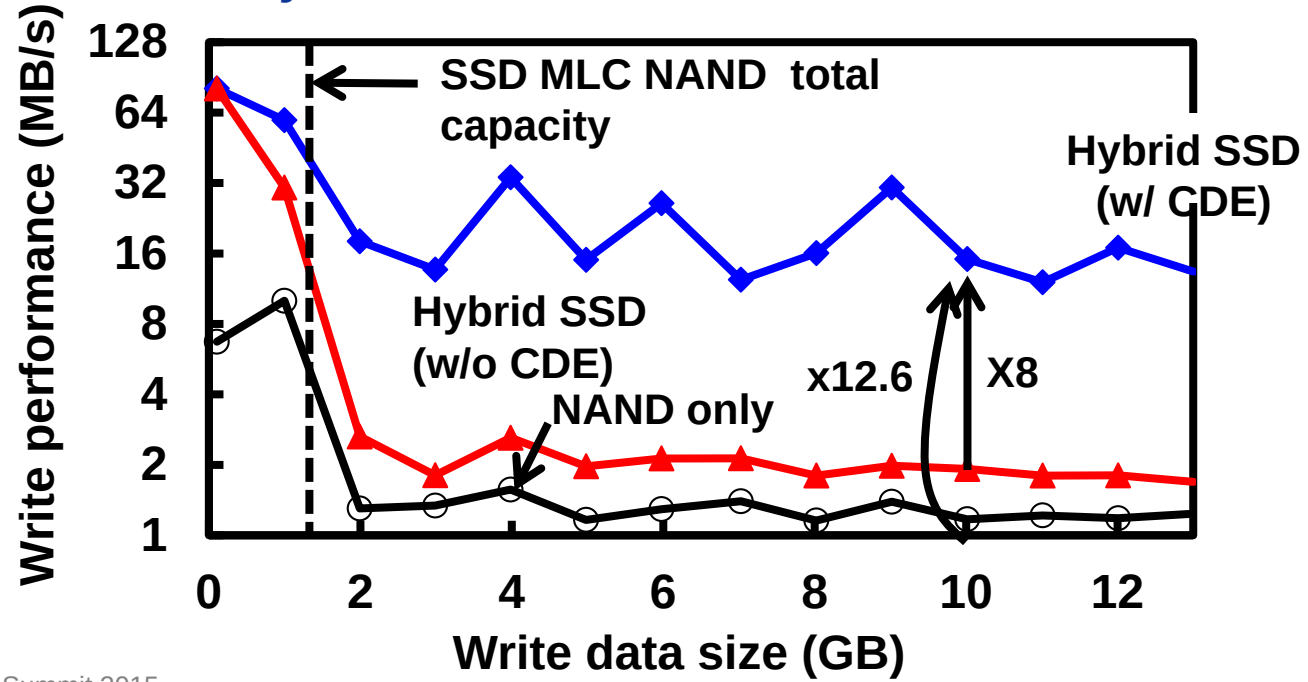
- **A cold data eviction (CDE) algorithm is proposed as the replacement algorithm.**



[1] C. Sun et al., *TCAS-I*, vol. 62, no. 2, pp. 382-392, 2014.

SSD Write Performance Boost

- **Over 10-times performance gain is achieved, compared with the NAND flash only SSD.**



[1] C. Sun et al., *TCAS-I*, vol. 62, no. 2, pp. 382-392, 2014.

Short Summary

- SCM is used as a storage device in SSD rather than a simple cache.
- Data placement strategy is to store fragmented data or hot data in SCM while cold and sequential data is stored in NAND flash memory. It is achieved by the intelligent data management algorithms.
- For financial server application , over 10-times write performance boost can be achieved, compared with NAND flash only SSD.
- TSV technology has been proved to be efficient of reducing the energy consumption of such hybrid SSD by over 90%.

Outline

- Introduction
- Three Techniques to Boost SSD Performance
  - Storage Engine Assisted SSD（SEA-SSD)
  - Logical Block Address (LBA) Scrambler
  - Storage Class Memory (SCM)/NAND Flash Hybrid SSD
- Conclusion

Conclusion

- **Three techniques have been proposed to improve the SSD write performance.**
- **SEA-SSD is proposed to improve the SSD write performance for the database application, by co-designing the database storage engine and SSD controller.**
- **LBA scrambler boosts the SSD write performance for general applications. It is based on the address-remapping technique. The middleware obtains the data storage information from the FTL.**
- **SCM/NAND flash hybrid SSD could own over 10-times faster write performance than the NAND flash only SSD as a future hybrid solution.**

More Research

- http://takeuchi-lab.org/research_e.html#5
  - SSD system: SSD intelligent data management (FTL)
  - 3D-LSI circuit: boost converter-based adaptive voltage generator for 3D-SSD.
  - Device: ReRAM, PCM, 0.5V low-power SRAM, Fe-NAND flash etc.

Thank you very much for your attention!

This work is partially supported by NEDO.