Long-term Research Issues in SSD

NVRAMOS ’2011

한양대학교

강 수 용

Research Issues: At-a-Glance

- Inside SSD
- Inside Computer Systems
- Inside Independent Storage Systems
- Inside Large/Networked Systems

Inside SSD (1)

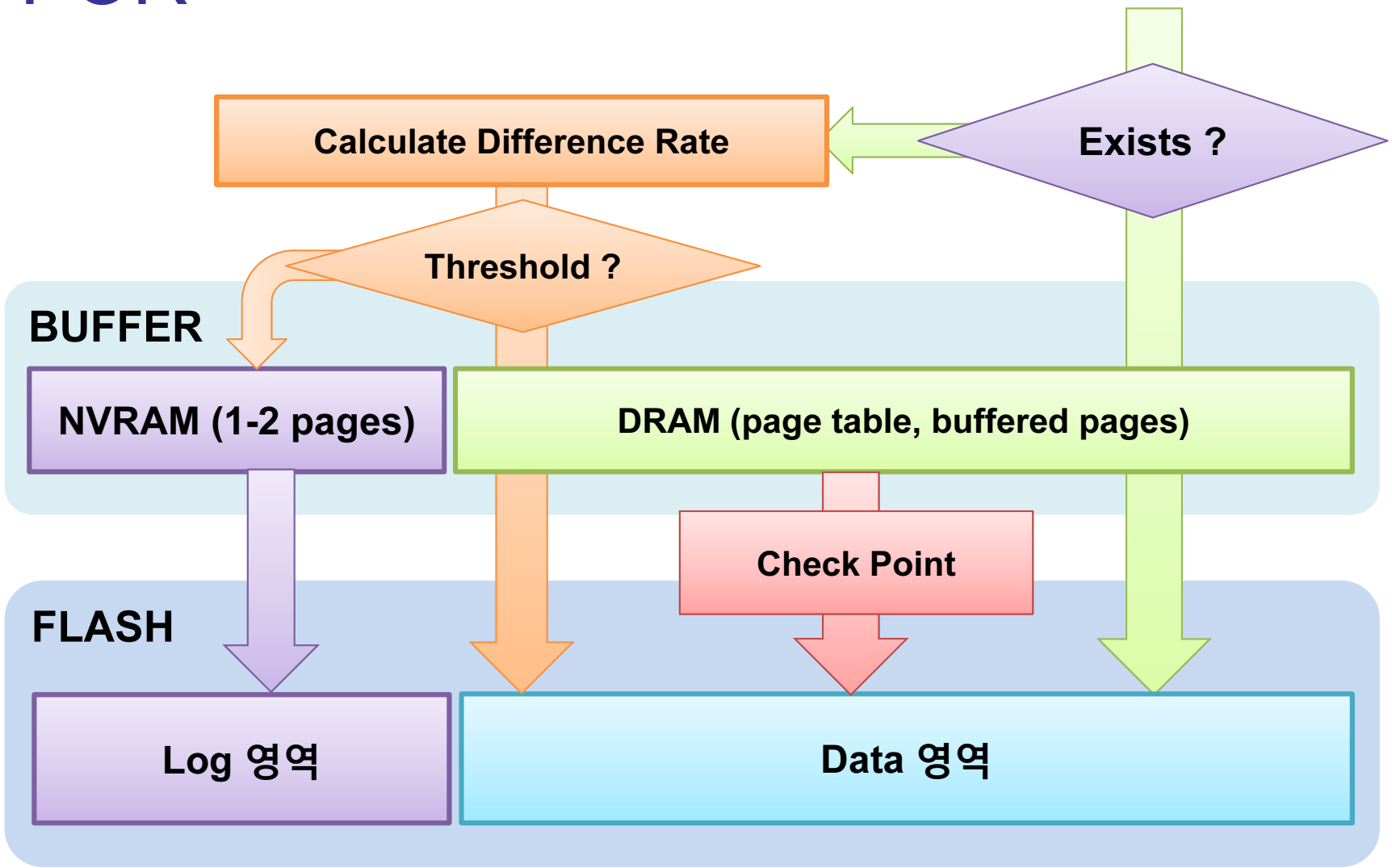
- Mapping for TB-scale SSDs
  - Page mapping with caching is enough?
    - For TB-scale workloads (MS exchange server, TPC-E), 64MB DRAM could accommodate the entire working set
  - When subpage (sector) mapping is used?
    - Multiple granularity mapping is worth investigating
  - Mapping for compressed/deduplicated data

Inside SSD (2)

- Reliability
  - In-Flash data reliability
    - ECC/CRC-based short-term reliability
      - Adaptive ECC : SandForce
      - E-MLC : SMART
      - “Using flash memories as SIMO channels for extending the lifetime of solid-state drives”, ICECS, 2010

        ➔ Read an erroneous page multiple times and correct errors
    - Redundancy-based long-term reliability
      - RAID-5 based data reliability: SandForce
  - DRAM data reliability (POR)
    - Cached metadata and buffered data
      - Logging & Checkpointing-based approach
      - High speed NVRAM-based approach

POR


Calculate Difference Rate
Exists ?
Threshold ?
BUFFER
NVRAM (1-2 pages)
DRAM (page table, buffered pages)
Check Point
FLASH
Log 영역
Data 영역

Inside Computer Systems

- Traditional short-term issues
  - Intelligent device driver: Fusion-IO
    - "Beyond block I/O: rethinking traditional storage primitives", HPCA'11
      - ➔ 'Atomic Write' primitive implemented in the device driver
  - Enriching interface commands set
- Traditional long-term issues
  - SSD Filesystem
    - "DFS: A file system for virtualized flash storage", FAST'10
      - ➔ removed duplicated functions (block allocation, free block management, file mapping, etc) from filesystem
  - All-New Memory-Storage stack in OS considering both SSD and Next-Generation NVRAM

Inside Computer Systems

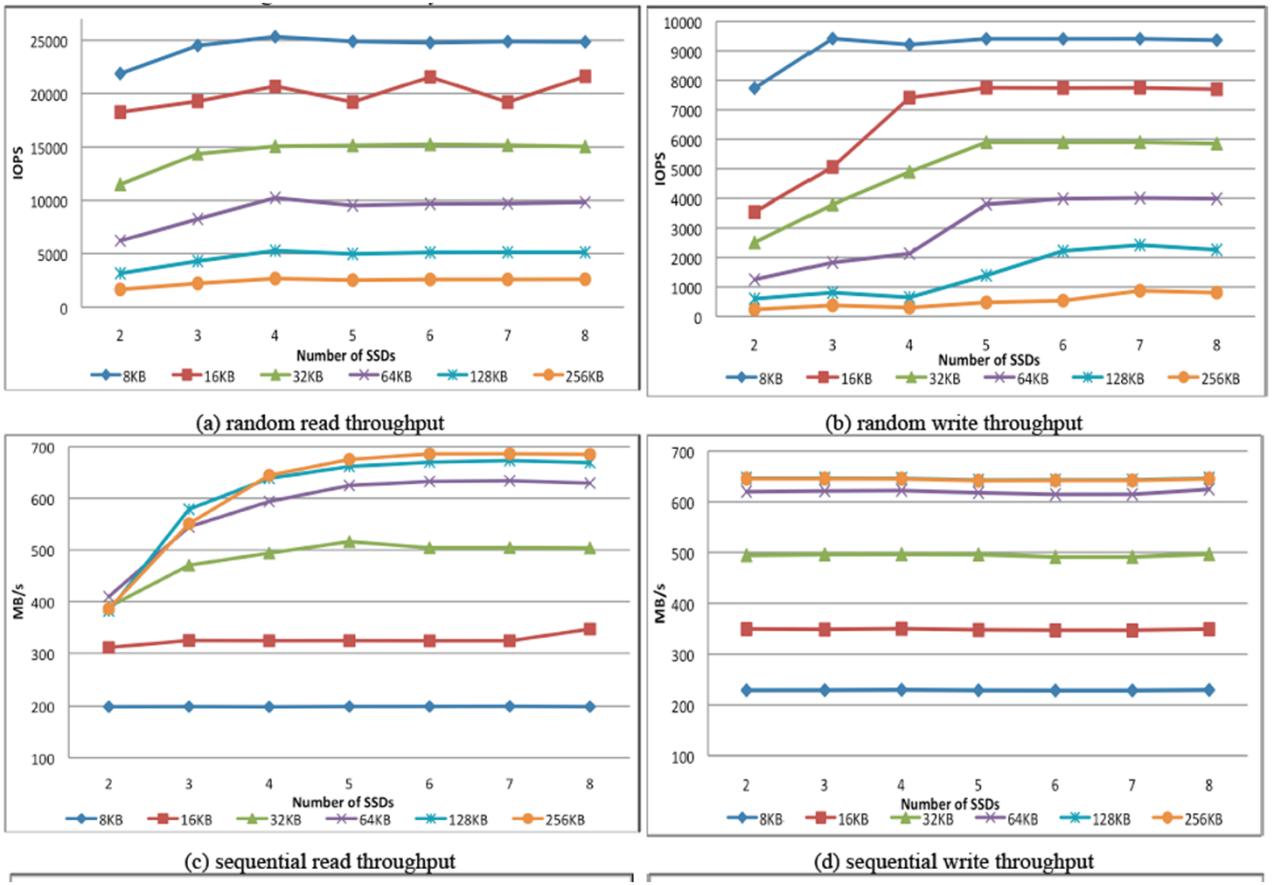
- New issue
  - Object-based Storage Device and Filesystem
    - “Block management in solid-state devices”, USENIX ATC, 2009
    - “Object-based SSD (OSSD) : Our Practice and Experiences”, Linuxcon 2010

Inside Storage Systems

- SSD Array
  - “Differential RAID: Rethinking RAID for SSD Reliability”, TOS ,2010
    - ➔ Unbalanced allocation of parity blocks across SSDs in RAID
  - “Building Large Storage based on Flash Disks”, ADMS, 2010
    - ➔ The bottleneck of the SSD RAID is controller
  - “Flash-Aware RAID Techniques for Dependable and High-Performance Flash Memory SSD”, TOC,2011

SSD RAID – Scalability Problem

- RAID 0, Intel X25-E 64GB



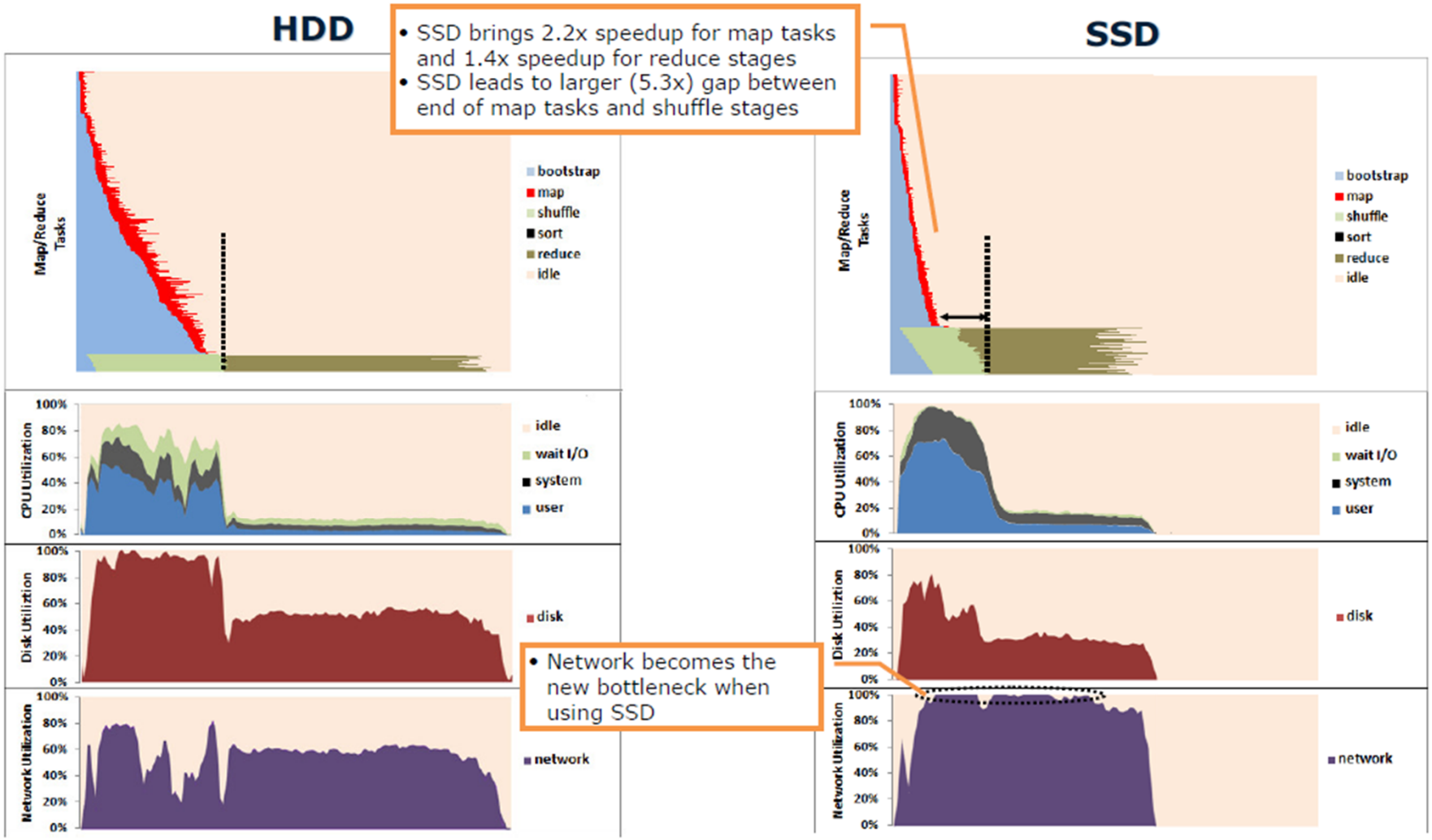
Inside Storage Systems

- Hybrid Array
  - SSD + HDD
    - "Reliability and Performance Enhancement Technique for SSD array storage system using RAID mechanism", ISCIT 2009
      ➔ Parity blocks for Hot blocks make unbalanced write counts across SSDs in RAID. Completely contradictory motivation with Differential RAID.
    - "Hybrid RAID With Dual Control Architecture for SSD Reliability", AIP 2009
      ➔ Use HDD (instead of SSD) for parity disk of RAID-4 SSD array
  - NVRAM + SSD (or HDD)
    - "Using a Shared Storage Class Memory Device to Improve the Reliability of RAID Arrays", PDSW 2010
      ➔ Use SCM as a shared additional parity store among multiple RAID-5 arrays
  - NVRAM + SSD + HDD
  - Combined LBA space or Separated LBA space
    - SSD/NVRAM as a cache? or a final store?
    - Same issue in Hybrid disks

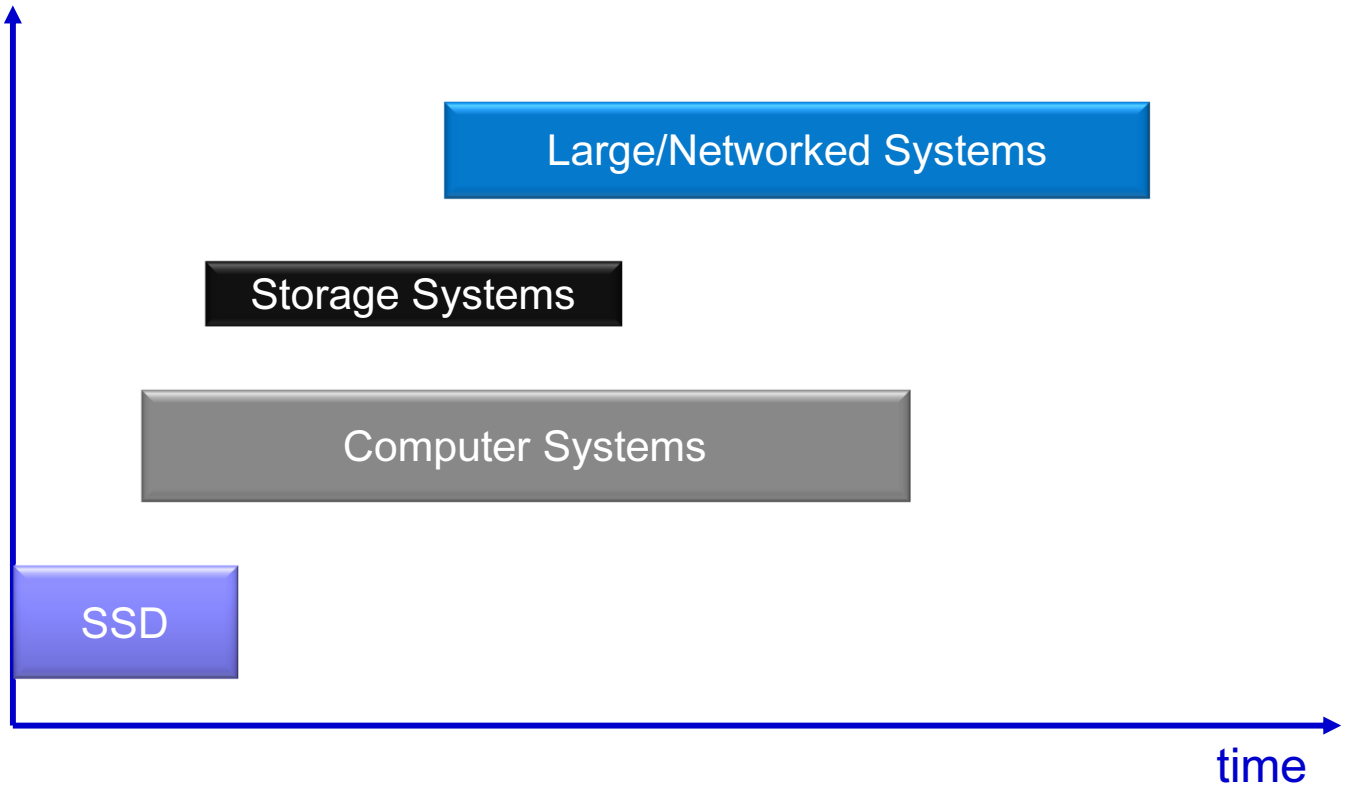
Inside Large/Networked Systems

- SSD as a Storage for High-Performance Computing Systems
  - Data-Intensive computing
  - Storage for Map-Reduce Framework
- SSD as a Networked Cache/Buffer
  - SSD as a metadata store in the Cloud

HDD vs. SSD for Hadoop Sort



Predicted Future Research Trends



§Thickness of each bar represents the popularity of the issue