
Education

# The Benefits of Solid State in Enterprise Storage Systems

David Dale, NetApp

# SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

# Abstract

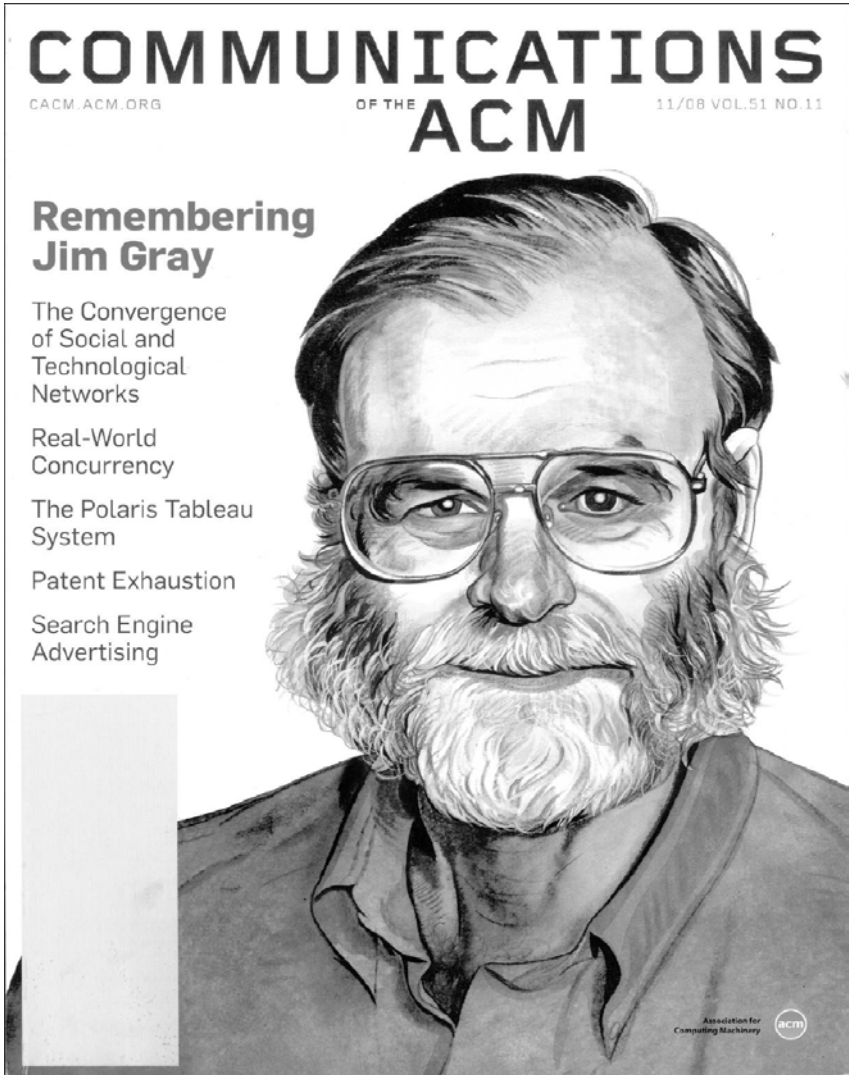
## Solid State in Enterprise Storage Systems

- Targeted primarily at an IT audience, this session presents a brief overview of the solid state technologies which are being integrated into Enterprise Storage Systems today, including technologies, benefits, and price/performance.
- It then goes on to describe where they fit into typical Enterprise Storage architectures today, with descriptions of specific use cases.
- Finally the presentation speculates briefly on what the future will bring.

# Agenda

- Why flash in the datacenter? Why now?
- Memory, cache and storage
- Application opportunities
- Flash in enterprise storage today
  - SSD storage tier
  - Storage controller-based cache
  - Network cache
- What's next
- Conclusion

# Remembering Jim Gray




Database and systems design pioneer, and co-creator of the Five Minute Rule (1987)

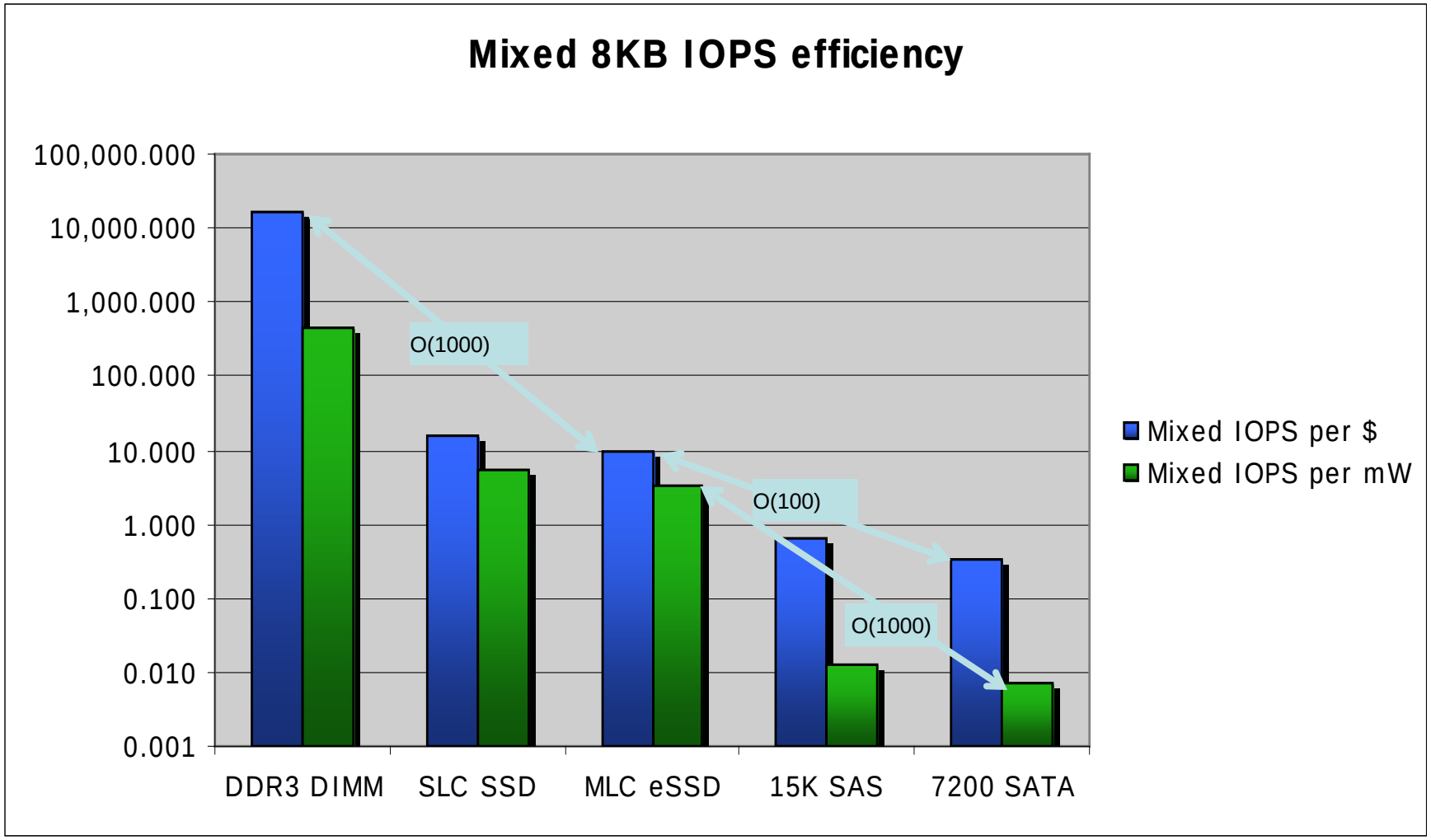
"Flash is a better disk …, and disk is a better tape" ~2006

Lost at sea January 2007

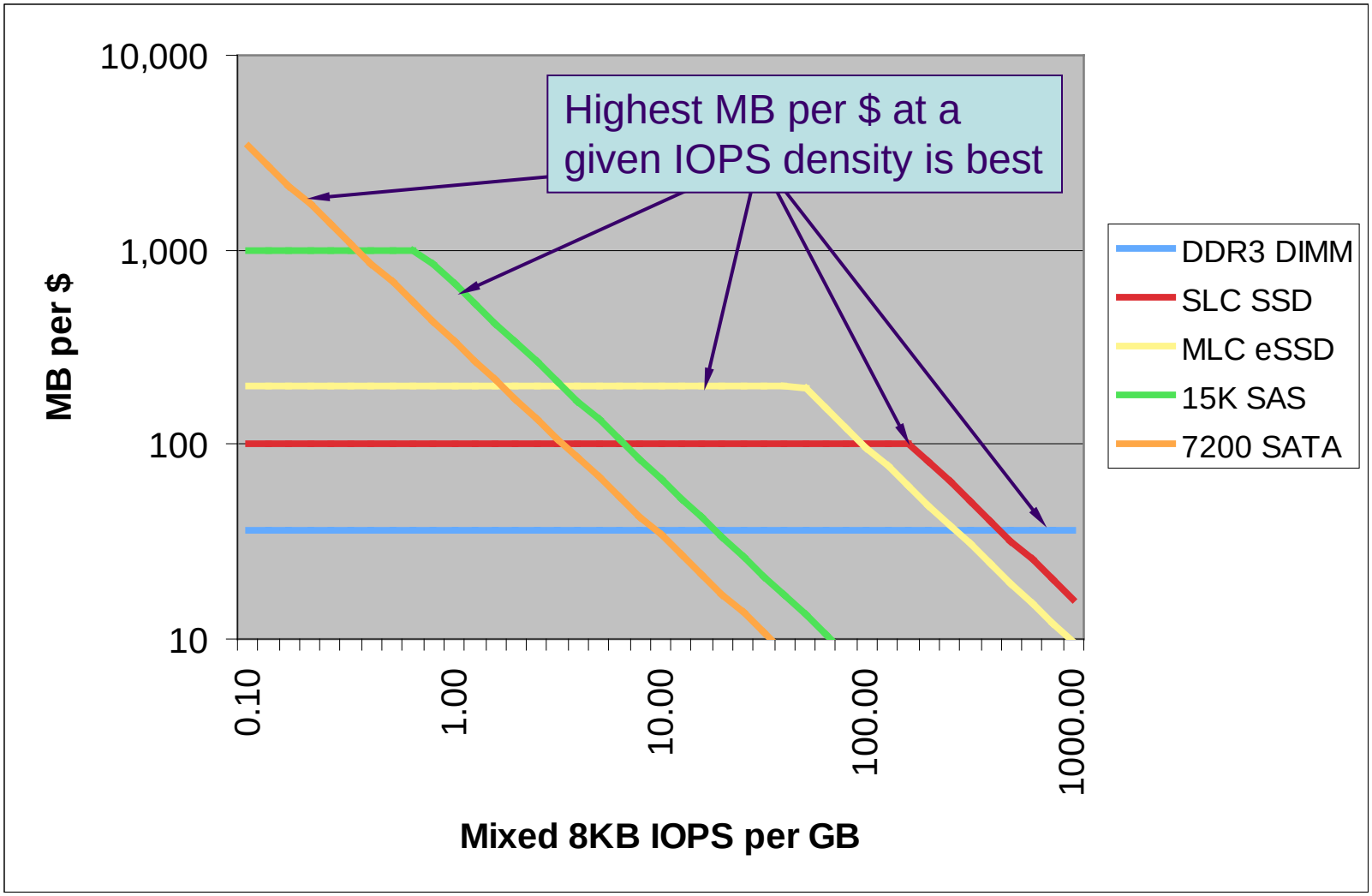
# Why Flash in the Data Center Now?

- Why flash?
  - Capacity efficiency versus DRAM
    - ~5x better $ per GB
    - ~40x better power per GB
  - IOPS efficiency versus HDDs
    - ~40x better $ per IOPS
    - ~600x better power per IOPS
- Why now?
  - Period of rapid density advancements led to HDD-like bit density at lower $/GB than DRAM
  - Innovations in SSD and tiering technology

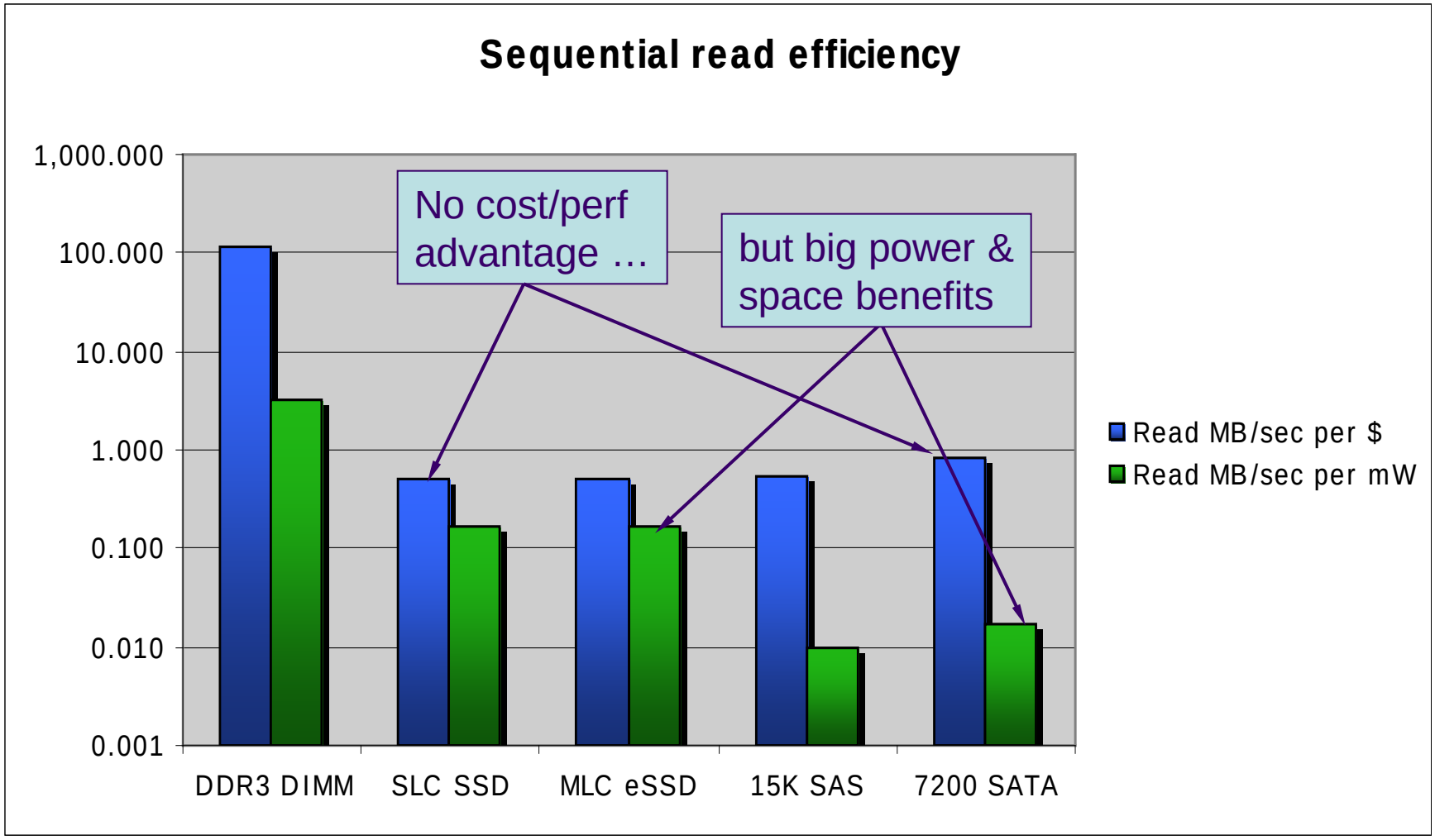
# Why Flash? IOPS Efficiency



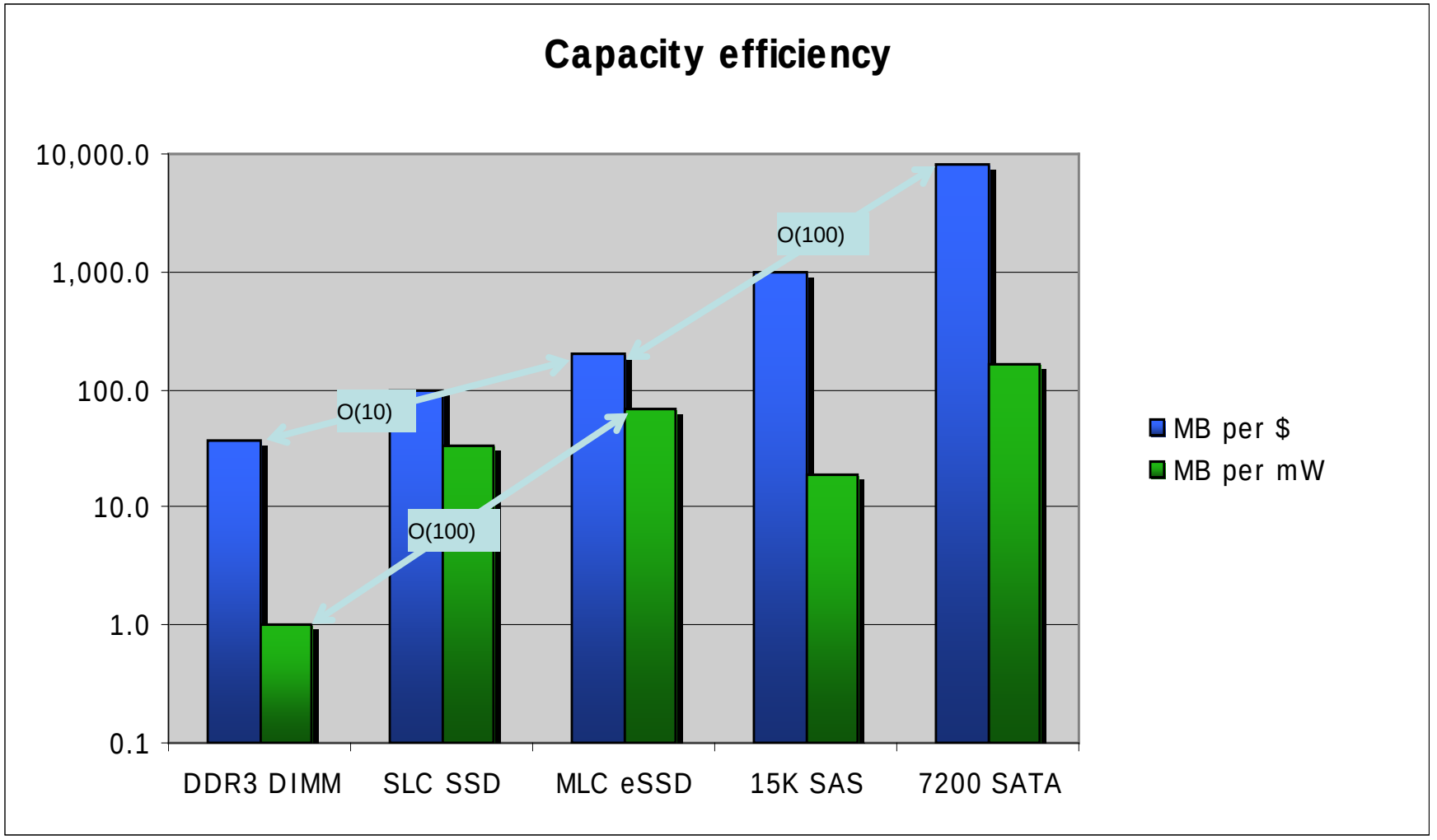
# Why Flash? An IOPS Density View



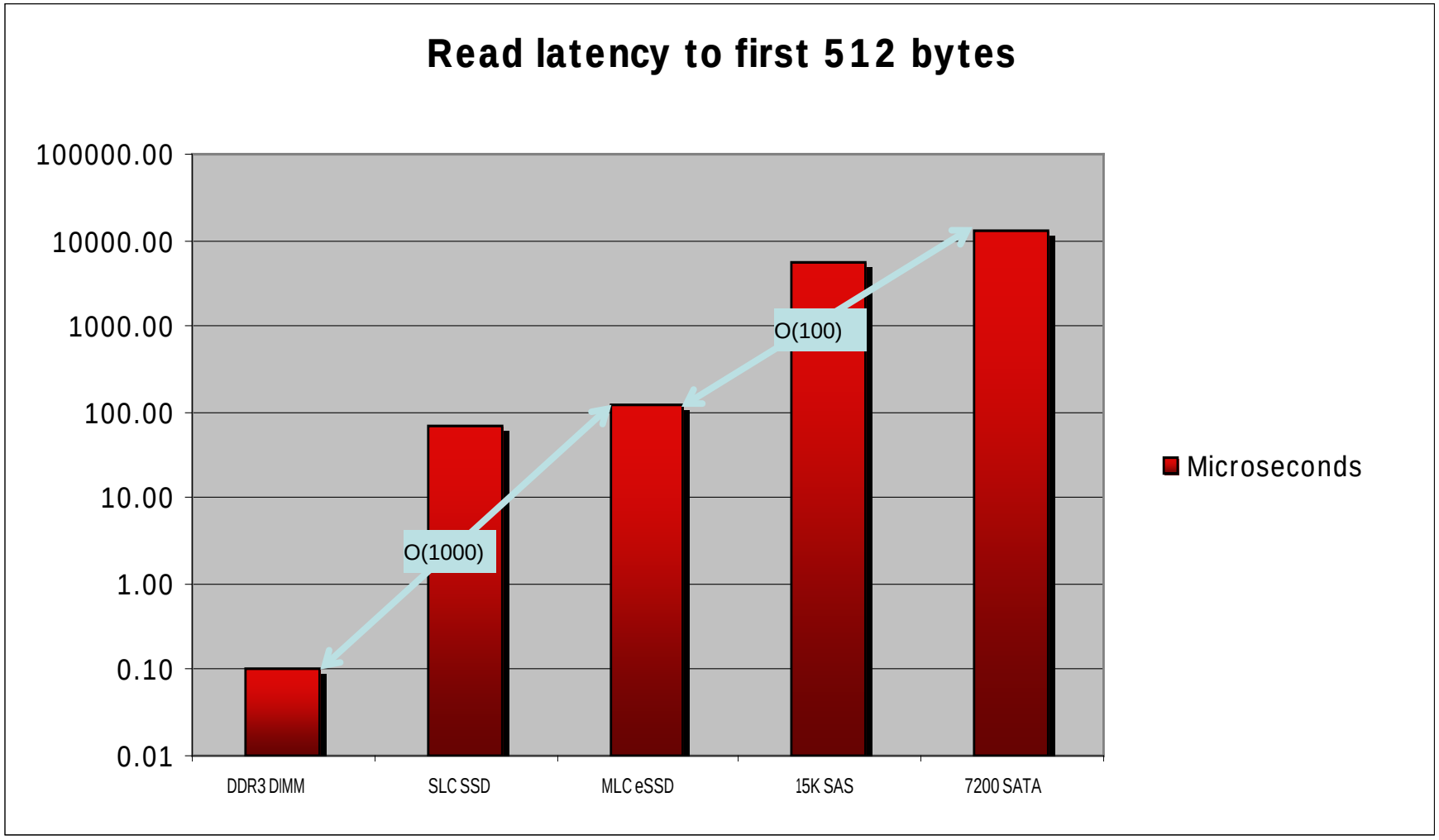
# Why Flash? Read Throughput per Watt



# Why Flash? Capacity Efficiency



# Why Flash? Read Latency



# Five Minute Rule, 1987

- Assuming that the cost of a cache is dominated by its capacity, and the cost of a backing store is dominated by its access cost (cost per IOPS), then the breakeven interval for accessing a page of data in cache is given by:

  Break-Even-Interval =
  Backing-Store-Cost-Per-IOPS /
  Cache-Cost-Per-Page

- 1987: Disk $2,000 / IOPS; RAM $5 / KB ➔
  1 KB breakeven = 400 seconds ~= 5 minutes

# Five Minute Rule, 2010: DRAM & HDD

- Disk $1 / IOPS (2,000x reduction)
- DRAM $25 / GB (200,000x reduction)

➔ 100 KB breakeven ~= 5 minutes
➔ 8 KB breakeven ~= 1 *hour*
➔ 1 KB breakeven ~= 10 hours *as Gray predicted*

- 200,000x / 2,000x = 100-fold decrease in breakeven access rate for a DRAM cache page backed by disk
  ➔ much bigger DRAM caches

# Five Minute Rule, 2010: Flash & HDD

- Disk $1 / IOPS
- MLC eSSD ~$5 / GB

➔SSD 100 KB breakeven ~= 30 minutes
➔SSD 8 KB breakeven ~= 7 hours (5x DRAM)

Flash economically caches working sets with 5x longer access intervals than DRAM.

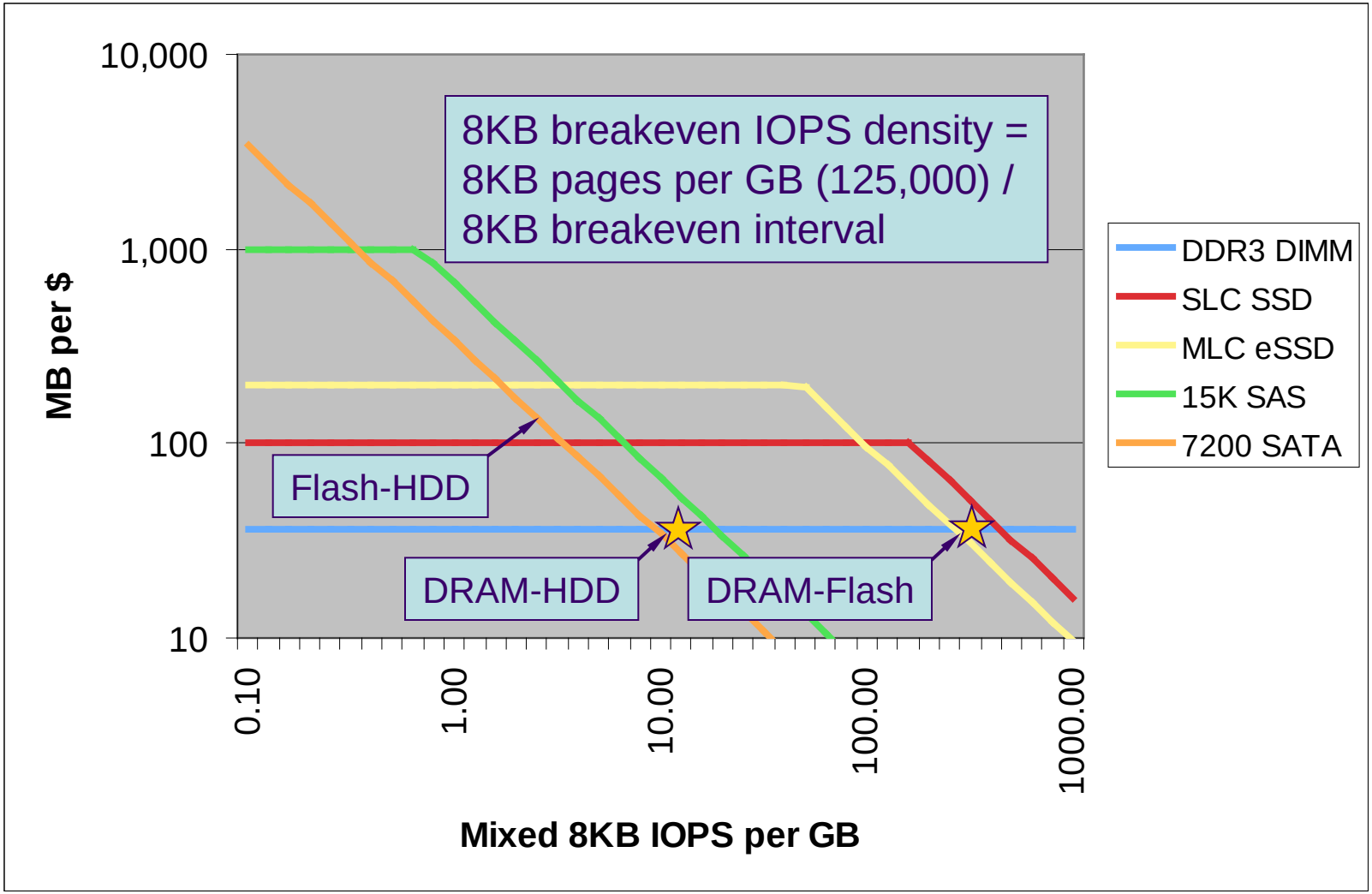
# Five Minute Rule, 2010: DRAM & Flash

- MLC eSSD ~$0.10 / mixed 8 KB IOPS
- DRAM $25 / GB

➔8 KB breakeven ~= 8 minutes (1/10th DRAM)

Adding flash between DRAM and HDD reduces the breakeven access interval for DRAM by 10x, indicating that DRAM capacity could be reduced to hold working sets for data accessed 1/10th as often

# IOPS Density and the Five Minute Rule



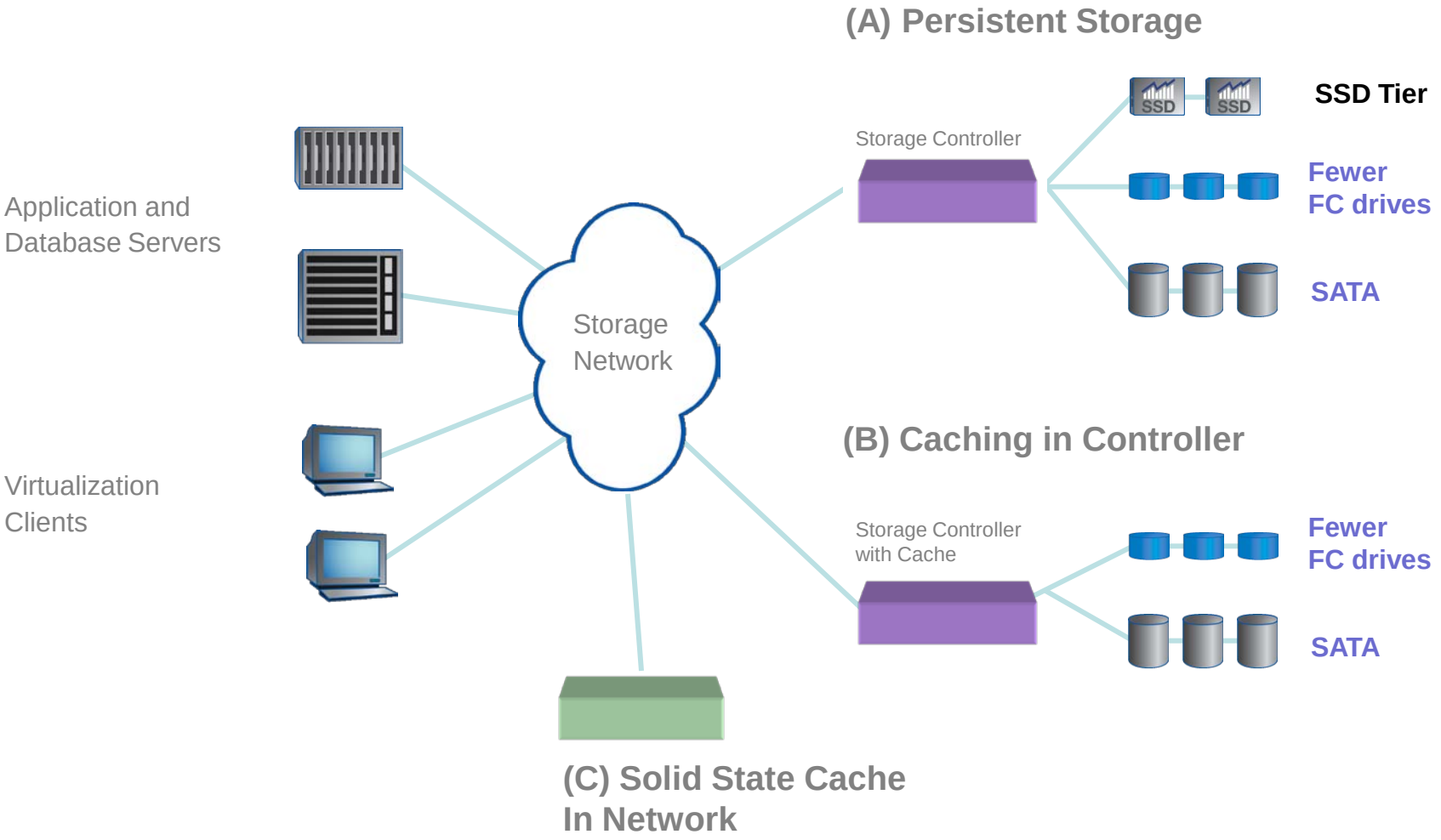
# Implications for Systems Design

- Flash makes it cost-effective to keep more small random data in silicon-based cache versus DRAM:
  ~5+ hour working set versus ~1 hour
- Flash allows small random data working set in DRAM to be reduced, allowing cost, power, space efficiency:
  ~5 minute working set versus ~1 hour
- Assuming appropriate locality of reference, transfer sizes between HDD and flash tiers should increase to preserve expensive HDD IOPS
- Flash tier likely to alter checkpoint processing intervals (shorter), metadata organization (e.g. optimal page size)

# Application Opportunities

- Intense random reads, e.g. OLTP, metadata
- Sequential read after random write
  - Log-oriented writes convert this to random read after sequential write (e.g. FTL)
- Low read latency (~100x better than HDD)
  - Facilitates DRAM extension by allowing high read throughput with limited read concurrency
  - Paging datacenter apps can be practical again
  - Memory capacity to consolidate more servers with underutilized CPU
- Enabling memory-resident datasets, e.g.
  - OLTP
  - Data warehouses (*viz* TPC-H results)
  - Large metadata

# Storage Networking with Flash



# Available Solutions Compared

| | Pros | Cons |
|---|---|---|
| Solid State Drives | ❒ Assured performance levels<br>❒ Low cost per IOPS<br>❒ Administrator has direct control over data stored in SSD tier | ❒ High cost per gigabyte<br>❒ Requires (manual) partitioning of hot data<br>❒ Limited practical applications |
| Controller Cache | ❒ Hot data automatically flows into cache – no administration required ➔ automated efficiency benefit<br>❒ Deployment can be non-disruptive<br>❒ Viable for common enterprise applications – cache "just helps" | ❒ Cache must be populated before it becomes effective<br>❒ Less predictable performance than static placement |
| Network Cache | ❒ Hot data automatically flows into the caching tier<br>❒ Deployment is relatively non-disruptive<br>❒ Scalable solution for high performance applications | ❒ Cache must be populated before it becomes effective<br>❒ Less predictable performance than static placement<br>❒ Placement in front of storage may constrain protocols or use cases |

# (A) Solid State Disk Tier

Education SNIA

## Advantages:

- Fast random I/O for small blocks
- Low read and write latency time
- Low power consumption
- Low noise
- Better mechanical reliability

## Disadvantages:

- Very high price, typically 10-30 X comparable FC drives
- Limited capacities
- Slow random write speeds, e.g. erase of blocks
- Slow sequential write throughput

# Practical Use Cases

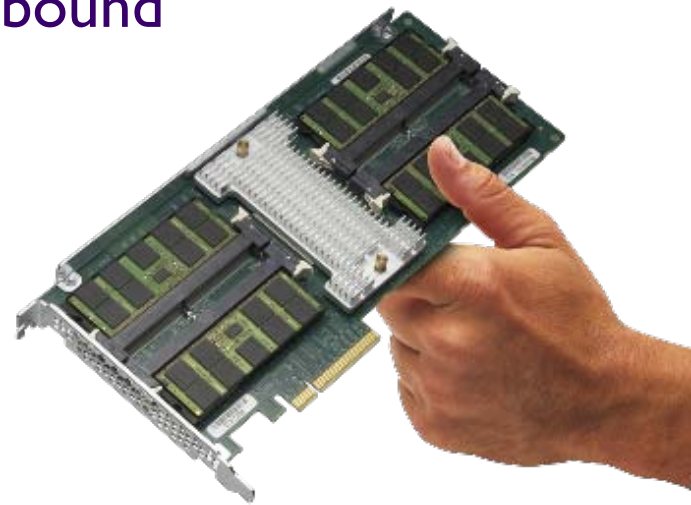
- **Database acceleration solution**
  - Entire database on SSD tier, or
  - Hot random access files on SSD and rest of database on standard disk
    - Indexes and temp space
- **Large scale virtual machine environments**
  - Solves "boot storm" problem for large numbers of virtual machines
  - Deduplication of VM data, e.g. virtual desktops
    - Reduces capacity requirements, increasing IOPS density, potentially making SSD economical

# Automated Tiering or Tier-less

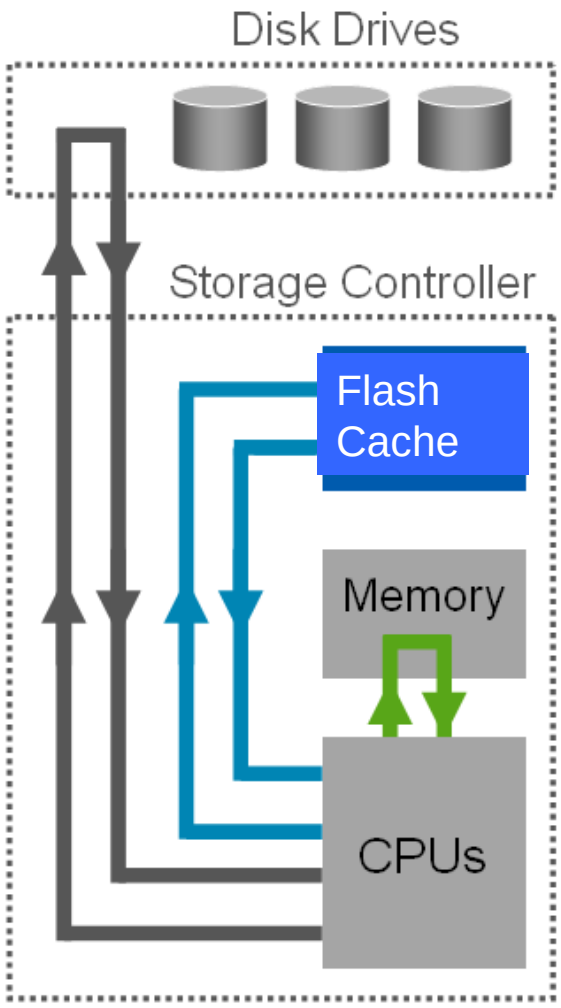
- Mixing SSD and HDD for a particular workload will probably be the most cost-efficient use of SSDs in over the next few years
- Area of intense innovation among enterprise storage vendors
- Issue is to automate data placement and movement
  - Automated tiering
  - Policy-based
  - No administrator overhead imposed
  - Some vendors refer to this as tier-less storage
- As SSD prices fall this will become increasingly important

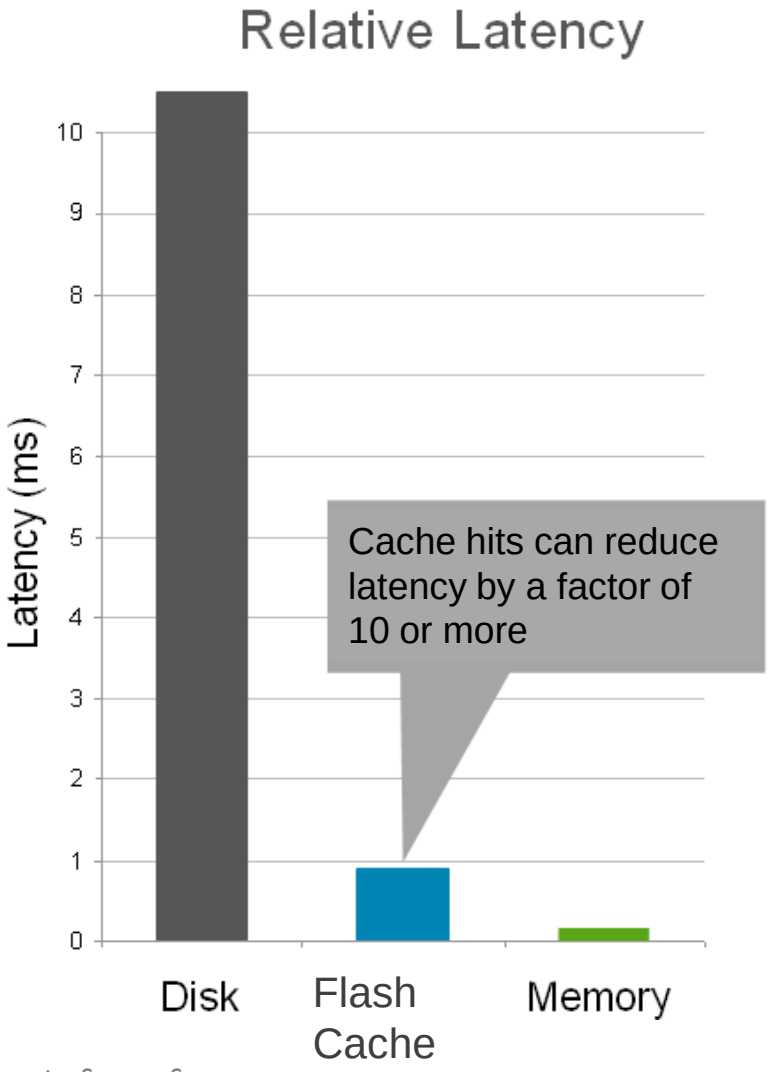
# (B) Controller-based Flash Cache

- Functions as an intelligent read cache for data and metadata
- Automatically places active data where access can be fast
- Provides more I/O throughput without adding high-performance disk drives to a disk-bound storage system
- Effective for file services, OLTP databases, messaging, and virtual infrastructure

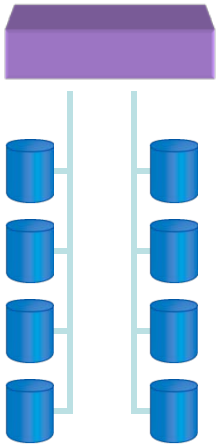


# Reduce Latency with Flash Cache



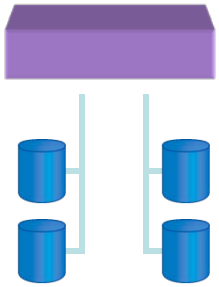



# Use case: Scale Performance of Disk-bound Systems



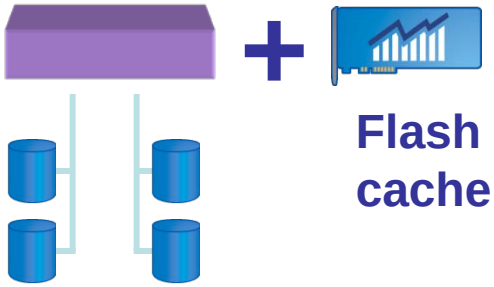
## Add Spindles

- Use more disks to provide more IOPs
- May waste storage capacity
- Consumes more power and space



## Starting Point:
## Need More IOPs

- Performance is disk-bound
- Have enough storage capacity
- Random read intensive workload

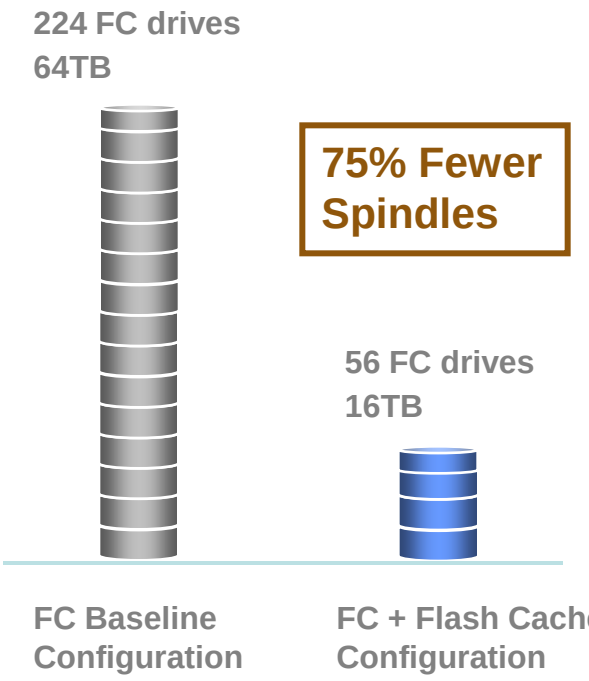



## Add Flash Cache

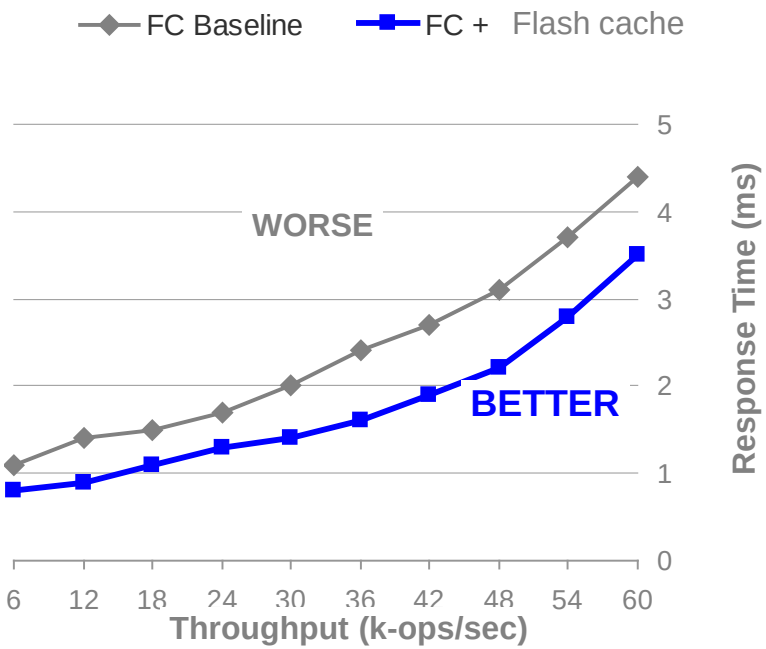
- Use cache to provide more IOPs
- Improves response times
- Uses storage efficiently
- Achieves cost savings for storage, power, and space

# FC HDD plus Flash Cache Example

## Benchmarked Configurations




## SPECsfs2008 Performance

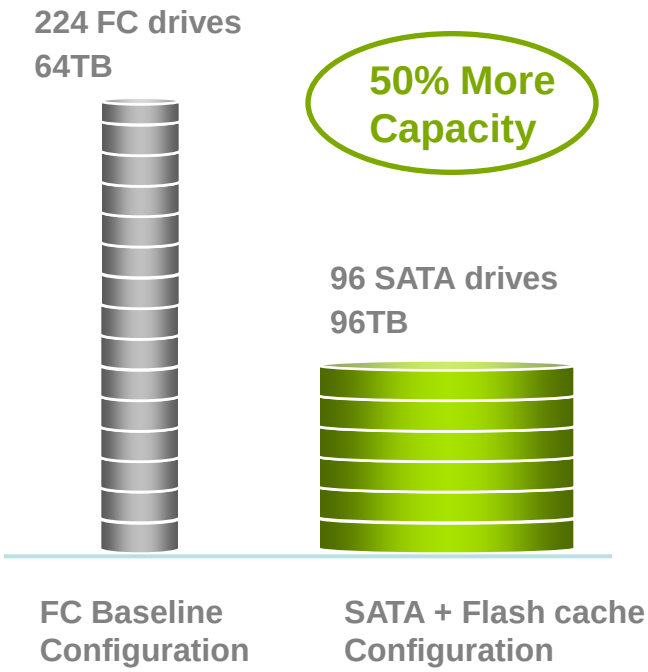



- Purchase price is **50% lower** for FC + Flash cache compared to Fibre Channel baseline
- FC + Flash cache yields **67% power savings** and **67% space savings**

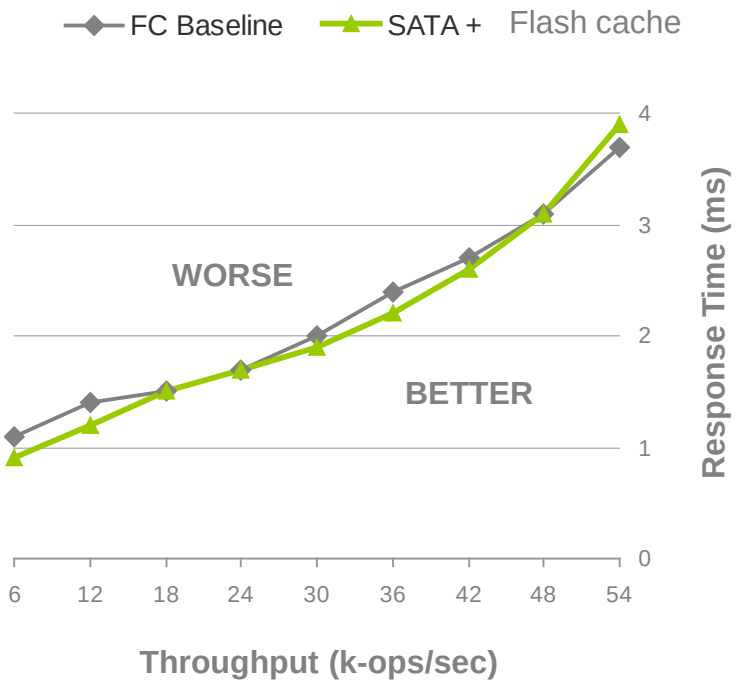
For more information, visit http://spec.org/sfs2008/results/sfs2008nfs.html.
SPEC® and SPECsfs2008® are trademarks of the Standard Performance Evaluation Corp.

# SATA HDD plus Flash Cache Example

## Benchmarked Configurations




## SPECsfs2008 Performance

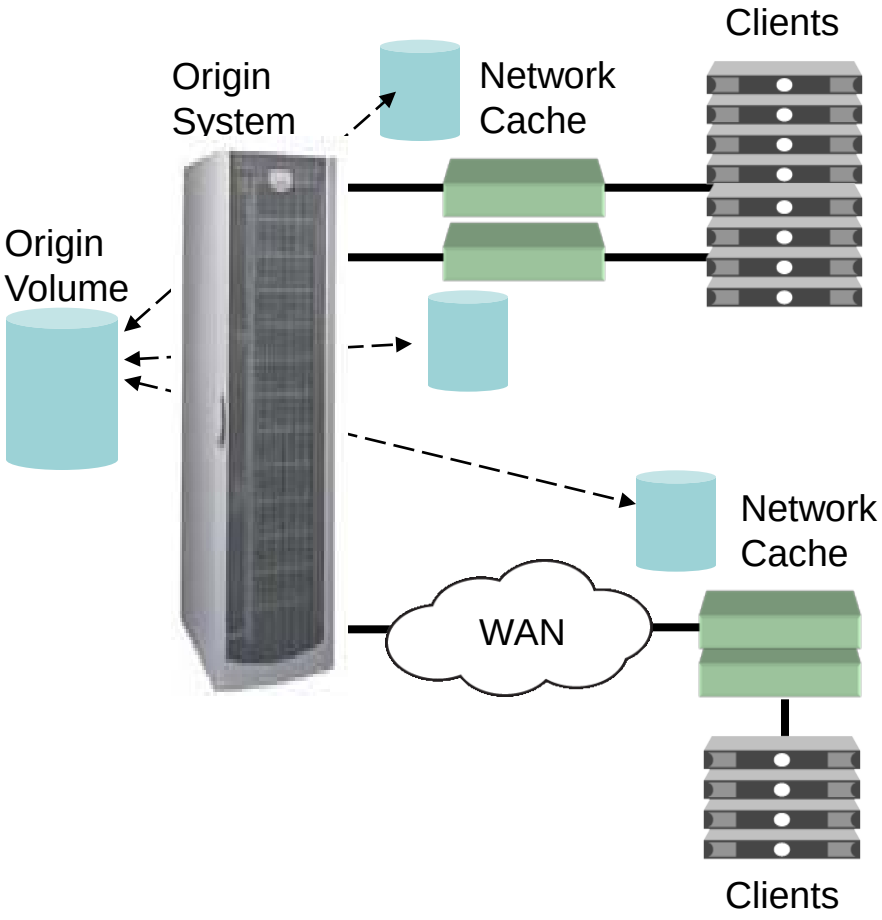



- Purchase price is **39% lower** for SATA + Flash cache compared to FC baseline
- SATA + Flash cache yields **66% power savings** and **59% space savings**

For more information, visit http://spec.org/sfs2008/results/sfs2008nfs.html.

SPEC® and SPECsfs2008® are trademarks of the Standard Performance Evaluation Corp.

# (C) Network Cache Topology



# Practical Use Cases

- **Network cache solutions**
  - All files on HDD in shared storage array
  - Accelerated by SSD-based network cache
  - Self–tuning write-through cache
- **Same pros and cons as SSD tier**
- **Typical applications**
  - Rendering
  - Seismic
  - Financial modeling
  - ASIC design




30

# Cost Structure of Memory/Storage Technologies

**Cost** determined by

- cost per wafer
  - # of dies/wafer
    - memory area per die [sq. μm]
      - memory density [bits per $4F^2$]
        - patterning density [sq. μm per $4F^2$]

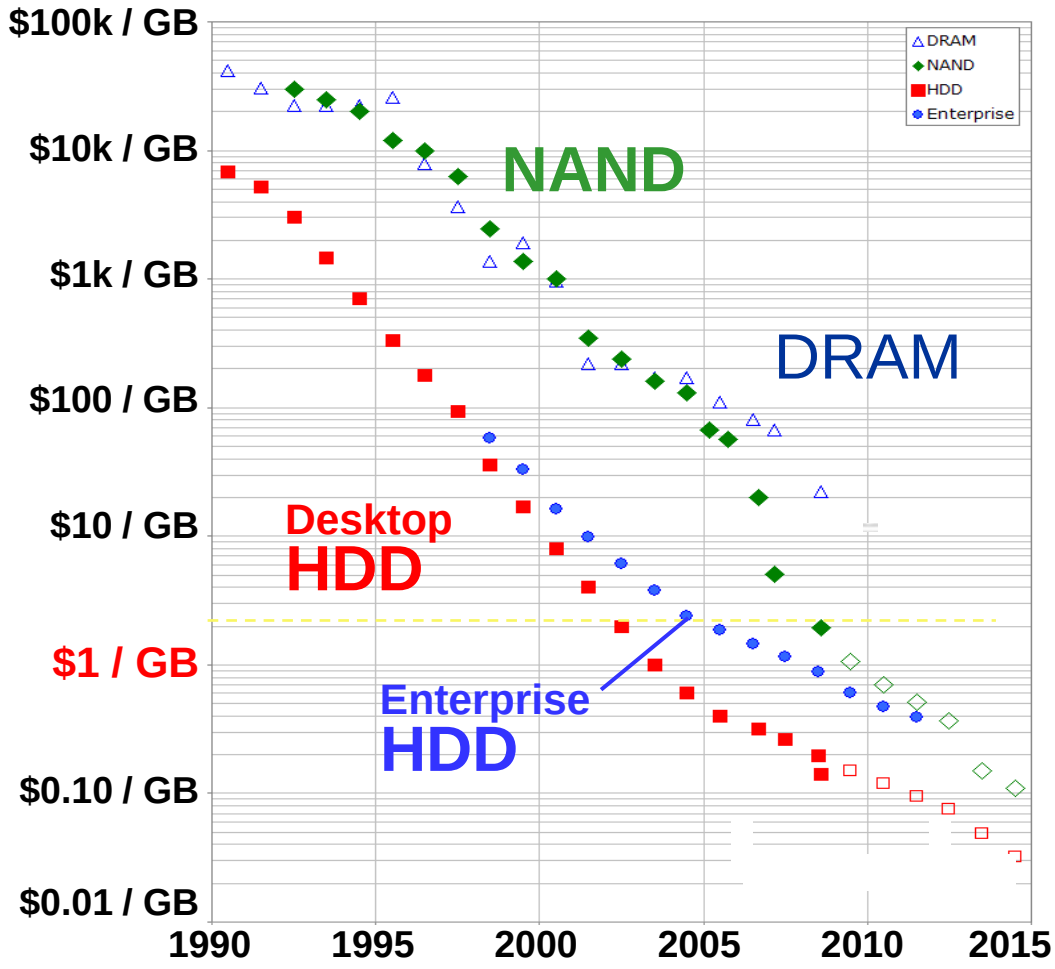
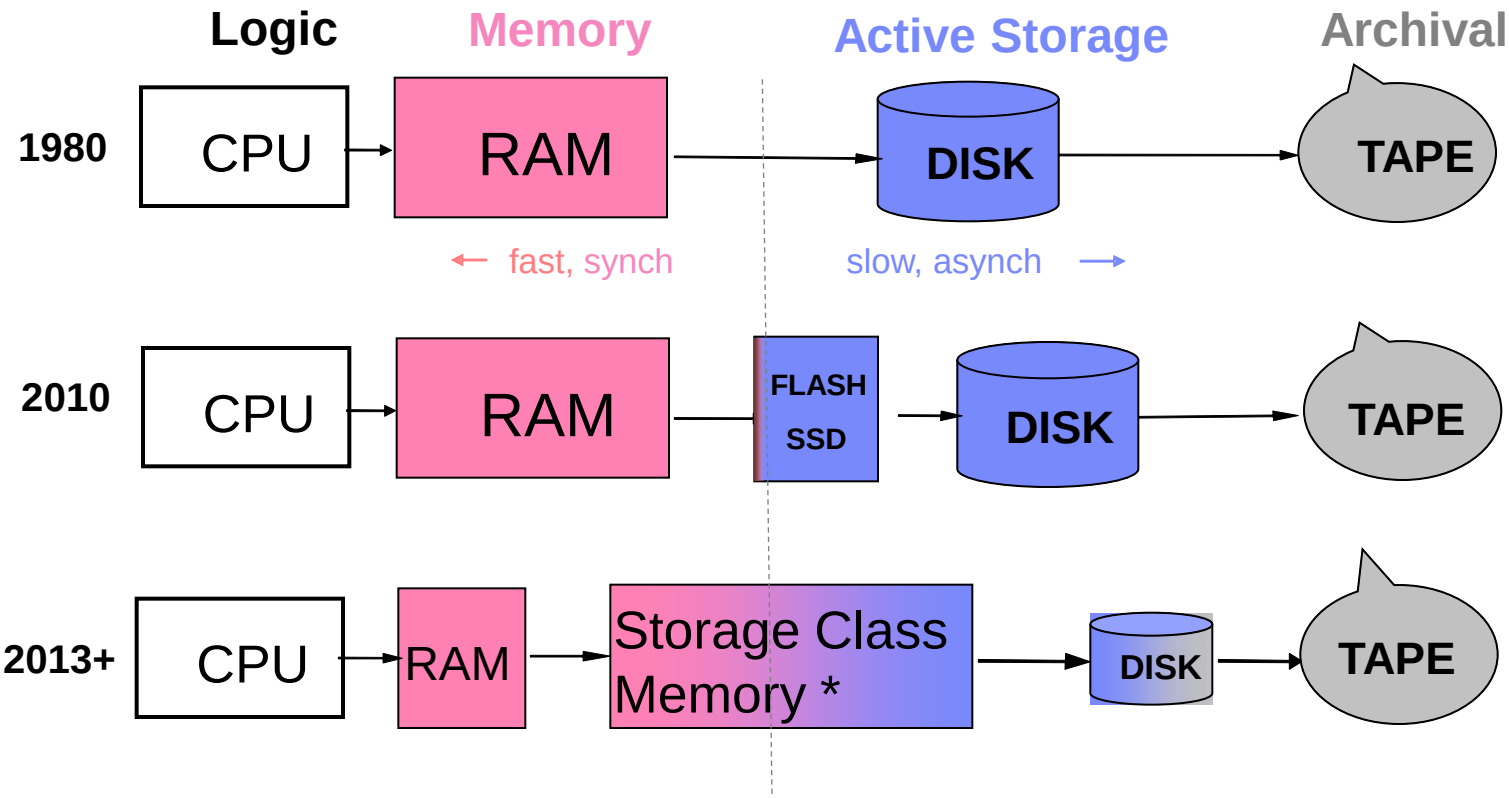



Chart courtesy of Dr. Chung Lam, IBM Research updated version of plot from 2008 *IBM Journal R&D* article

# System Evolution




* e.g. Phase change memory
Memristor
Solid Electrolyte
Racetrack memory

# Summary

- Over the next 5 years solid state technologies will have a profound impact on enterprise storage
- It's not just about replacing mechanical media with solid state media
- The architectural balance of memory, cache and persistent storage will change
- Today's solid state implementations in enterprise storage demonstrate these changes
- It's only the beginning…

# Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: trackexecutive@snia.org

**Many thanks to the following individuals for their contributions to this tutorial.**

**- SNIA Education Committee**

**David Dale**
**Jeff Kimmel**
**Chris Lionetti**
**Phil Mills**
**Amit Shah**
**Mark Woods**
**Alan Yoder**