



Education

# **The Benefits of Solid State in Enterprise Storage Systems**

David Dale, NetApp

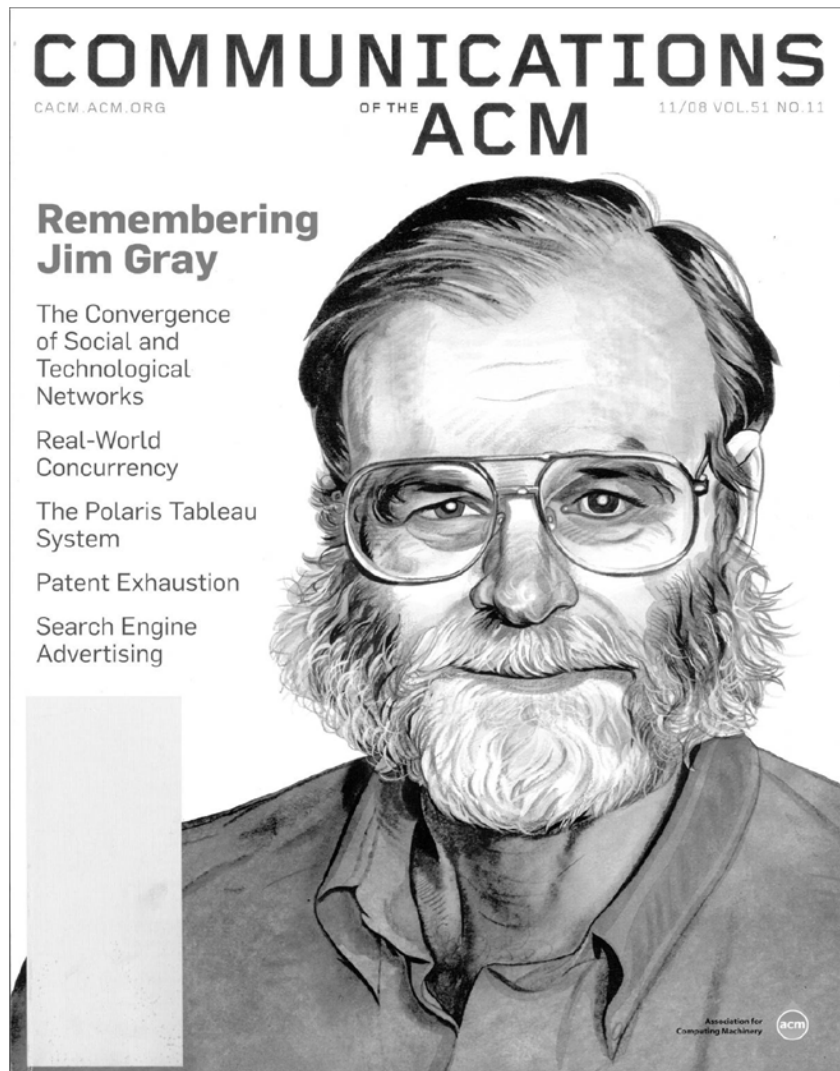
- The material contained in this tutorial is copyrighted by the SNIA.
  - Member companies and individual members may use this material in presentations and literature under the following conditions:
    - ◆ Any slide or slides used must be reproduced in their entirety without modification
    - ◆ The 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.**

## ➤ 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.

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

# Remembering Jim Gray



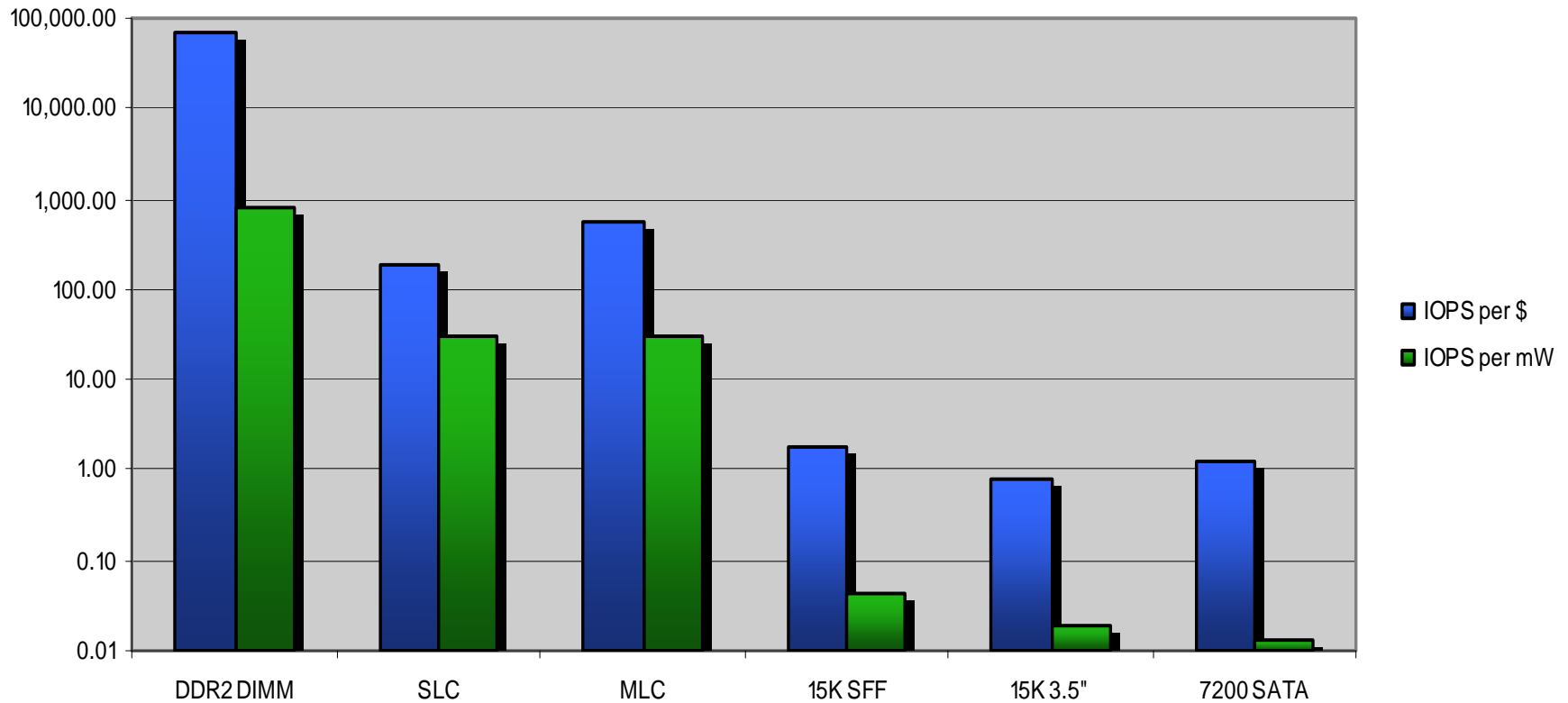
Database and systems design pioneer, and co-creator of the Five Minute Rule (1987)

“Flash is a better disk ..., and disk is a better tape”  
~2006

Lost at sea January 2007

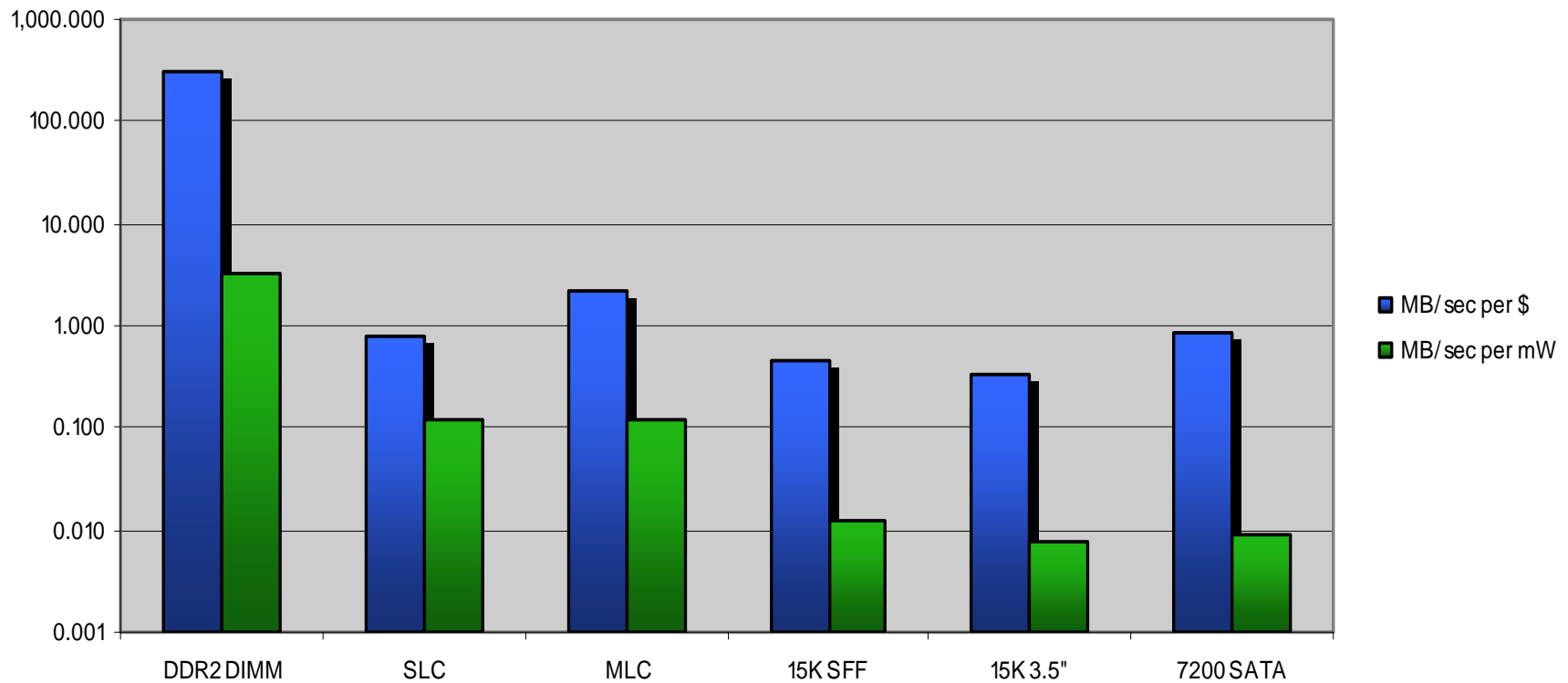
# Why Flash? IOPS efficiency vs. HDD

### Random read efficiency



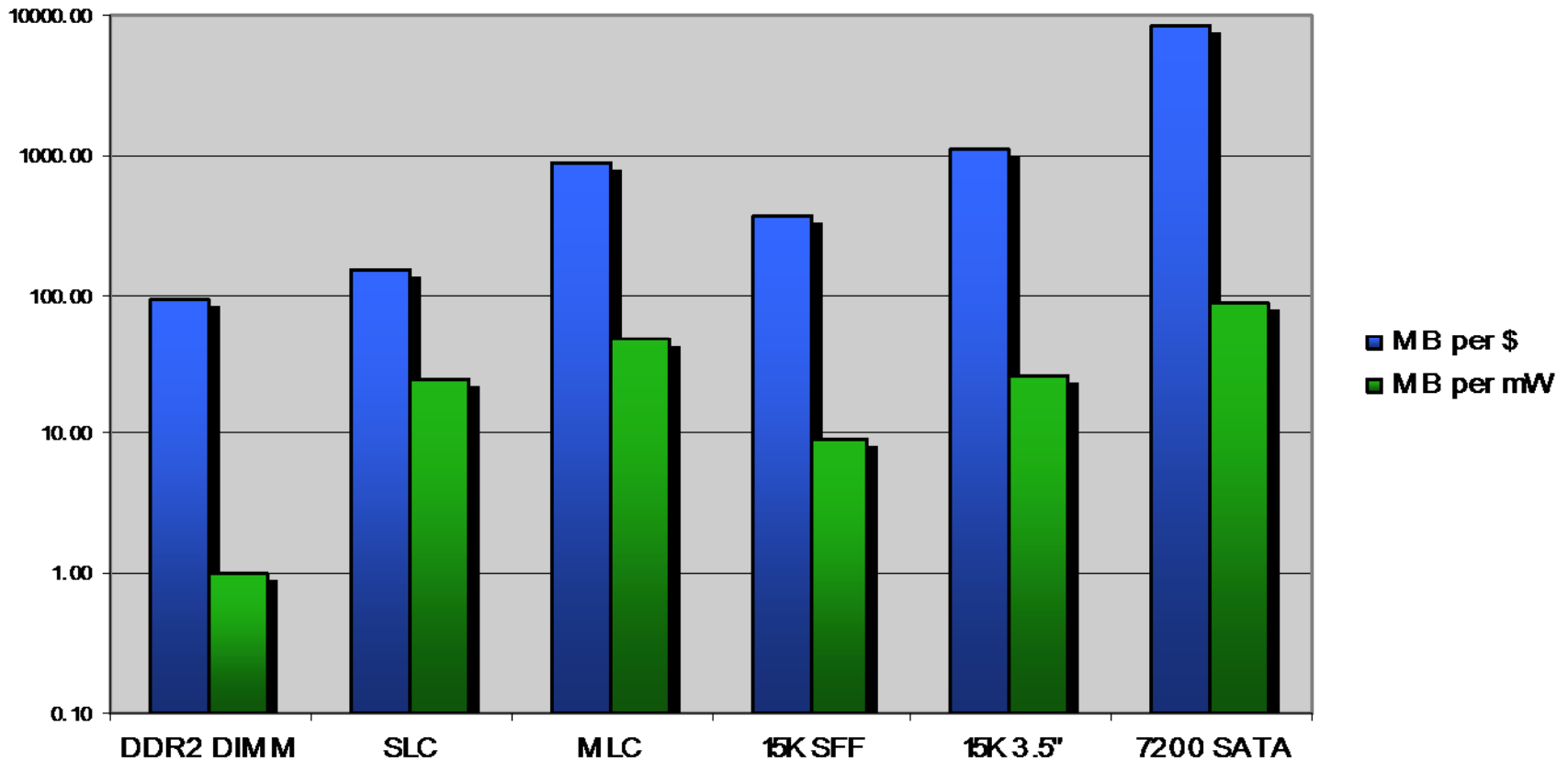
# Why Flash? Bandwidth/Watt vs. HDD

## Sequential throughput efficiency

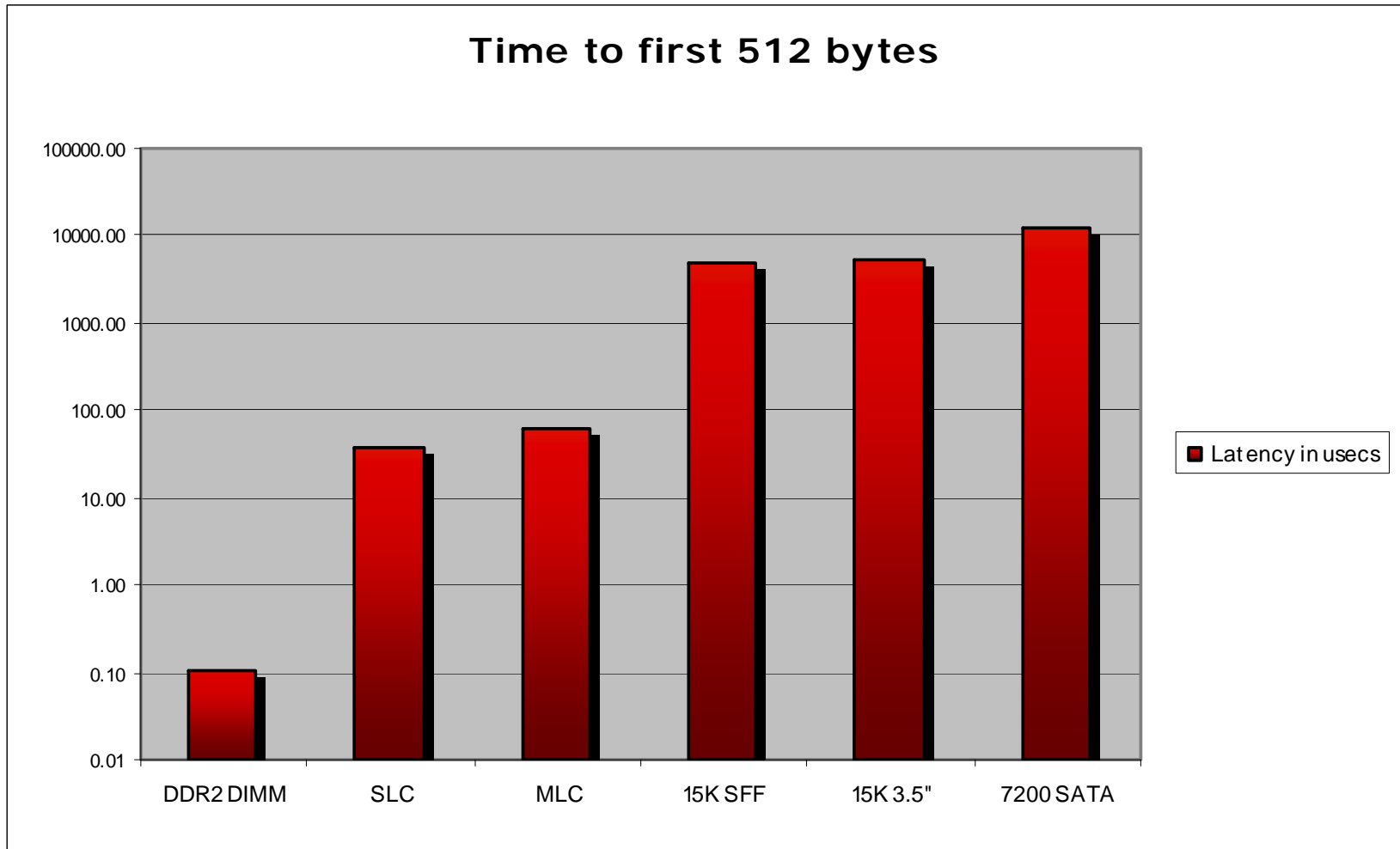


# Why Flash? Capacity/Watt vs. DRAM

### Capacity efficiency



# Why Flash? Latency vs. HDD



## ➤ Why flash?

- ◆ Capacity efficiency versus DRAM
  - > ~10x better \$ per GB
  - > ~30x better power per GB
- ◆ IOPS efficiency versus HDDs
  - > ~100x better \$ per IOPS
  - > ~1000x 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

- 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 keeping a page of data in cache is given by:

Break-Even-Interval =

$$\frac{\text{Backing-Store-Cost-Per-IOPS}}{\text{Cache-Cost-Per-Page}}$$

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

- Disk \$1 / IOPS (2,000x reduction)
- DRAM \$50 / GB (100,000x reduction)  
\$0.05 / MB, \$0.0025 / 50 KB
- ➔ 50 KB breakeven  $\approx$  5 minutes
- ➔ 4 KB breakeven  $\approx$  1 hour
- ➔ 1 KB breakeven  $\approx$  5 hours *as Gray predicted*
- $100,000x / 2,000x = 50$ -fold increase in size of “page” to cache for breakeven at 5 minutes

- HDD \$1 / IOPS (2,000x reduction)
- SLC flash ~\$10 / GB (packaged)
- MLC flash ~\$4 / GB (packaged)
  
- ➔ SLC 250 KB breakeven  $\approx$  5 minutes  
    SLC 4 KB breakeven  $\approx$  5 *hours*
- ➔ MLC 625 KB breakeven  $\approx$  5 minutes  
    MLC 4 KB breakeven  $\approx$  13 *hours*
  
- $100,000x / 2,000x = 50$ -fold increase in size of “page” to cache for breakeven at 5 minutes

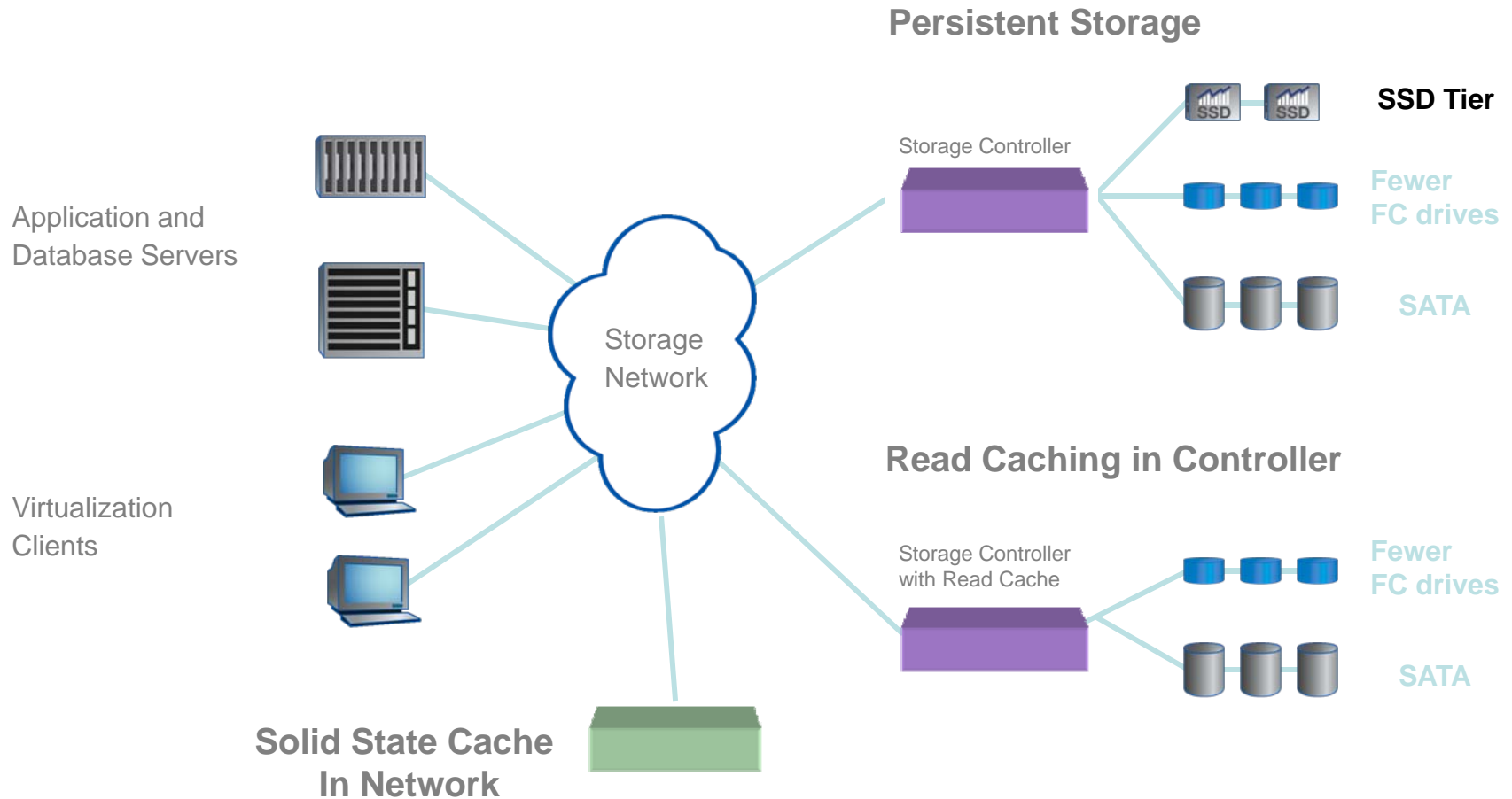
- SLC flash ~\$0.05 / IOPS (4 KB, enterprise SSD)
- MLC flash ~\$0.02 / IOPS (4 KB, enterprise SSD)
  
- DRAM \$20 / GB (enterprise DIMMs)
  
- ➔ SLC 6 KB breakeven  $\approx$  5 minutes
- ➔ MLC 2 KB breakeven  $\approx$  5 minutes

Need to consider however that cost/capacity of flash (at least SLC) is a large fraction of that of DRAM

- 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)

- 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: Pro and Con

Technology	Pros	Cons
Solid State Drives	<ul style="list-style-type: none"><li>• Response times consistently fast for reads</li><li>• Low cost per IOP</li><li>• Administrator has direct control over data stored in SSD tier</li></ul>	<ul style="list-style-type: none"><li>• High cost per gigabyte</li><li>• Requires software tools and administration to move hot data into and out of SSD tier</li><li>• Limited apps today</li></ul>
Controller Read Cache	<ul style="list-style-type: none"><li>• Hot data automatically flows into read cache—no administration required</li><li>• Deployment is relatively non-disruptive</li><li>• Viable for common enterprise applications</li></ul>	<ul style="list-style-type: none"><li>• Cache must be populated before it becomes effective</li></ul>
Network Cache	<ul style="list-style-type: none"><li>• Hot data automatically flows into the caching tier</li><li>• Deployment is relatively non-disruptive</li><li>• Scalable solution for high performance applications</li></ul>	<ul style="list-style-type: none"><li>• Cache must be populated before it becomes effective</li><li>• May be limited by protocol choice available from vendor</li></ul>

# SSD Is Much Faster than HDD



**SSD tier**



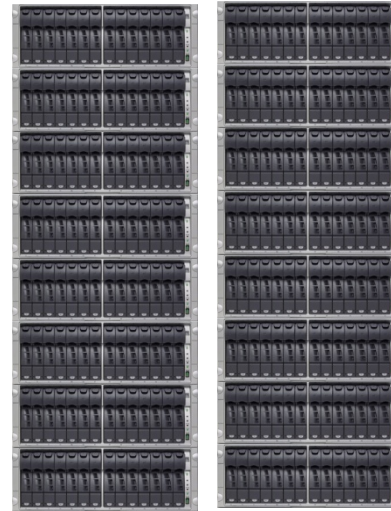
**HDD tier**

	<b>SDD</b>	<b>HDD</b>	<b>Comment</b>
<b>Capacity</b>	<b>Equal (2TB)</b>	<b>Equal (2TB)</b>	<b>Hold capacity constant</b>
<b>IOPs</b>	<b>~50,000</b>	<b>~3,600</b>	<b>Order of magnitude faster for flash</b>
<b>Latency</b>	<b>Better</b>	<b>Worse</b>	<b>Order of magnitude lower for flash</b>
<b>Carbon Footprint</b>	<b>Better</b>	<b>Worse</b>	<b>Same per TB, better per IOPs</b>

# For a 50,000 IOP System, SSD Has Much Better Latency and Footprint



**SSD tier**



**HDD Tier**

	<b>SSD Tier</b>	<b>HDD Tier</b>	<b>Comment</b>
<b>IOPs</b>	<b>50,000</b>	<b>50,000</b>	<b>Hold IOPs constant</b>
<b>Capacity</b>	<b>Worse (2TB)</b>	<b>Better (27 TB)</b>	<b>Significantly more for disk</b>
<b>Latency</b>	<b>Better (1ms)</b>	<b>Worse (10ms)</b>	<b>Order of magnitude lower for flash</b>
<b>Rack Space</b>	<b>12U</b>	<b>54U</b>	<b>Significantly more for disk</b>

## ➤ Advantages:

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

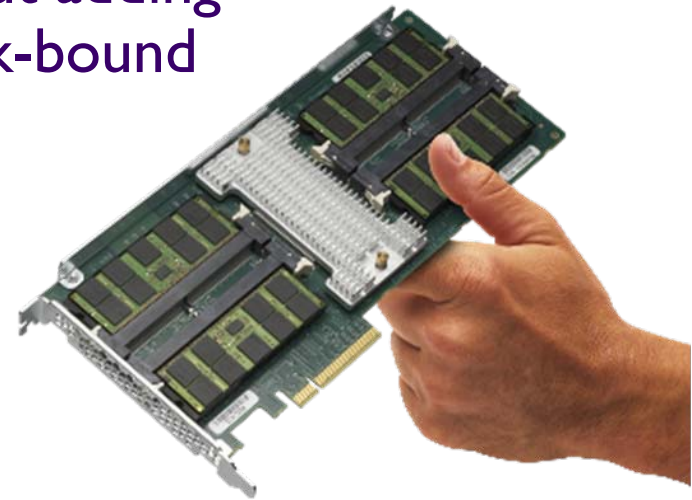
## ➤ Disadvantages:

- ◆ 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

# SSD-based Solutions

- **Database acceleration solution**
  - ◆ Entire DB on SSD tier
  - ◆ Or hot files on SSD and rest of DB on standard disk
    - › Redo logs, indexes, temp space
- **Large scale virtual machine environments**
  - ◆ Solves “boot storm” problem for large numbers of virtual machines
- **Entire data set on SSD tier**
  - ◆ Multiple apps: Virtual Servers, VDI, and so on
  - ◆ Any app where entire data set can fit into memory of SSD array
  - ◆ Works well in NAS environments
- **Network cache solutions**
  - ◆ All files on HDD in shared storage array
  - ◆ Accelerated by SSD-based network cache
  - ◆ Self-tuning write-through cache
  - ◆ Applications include
    - › Rendering, seismic, financial modeling, ASIC design

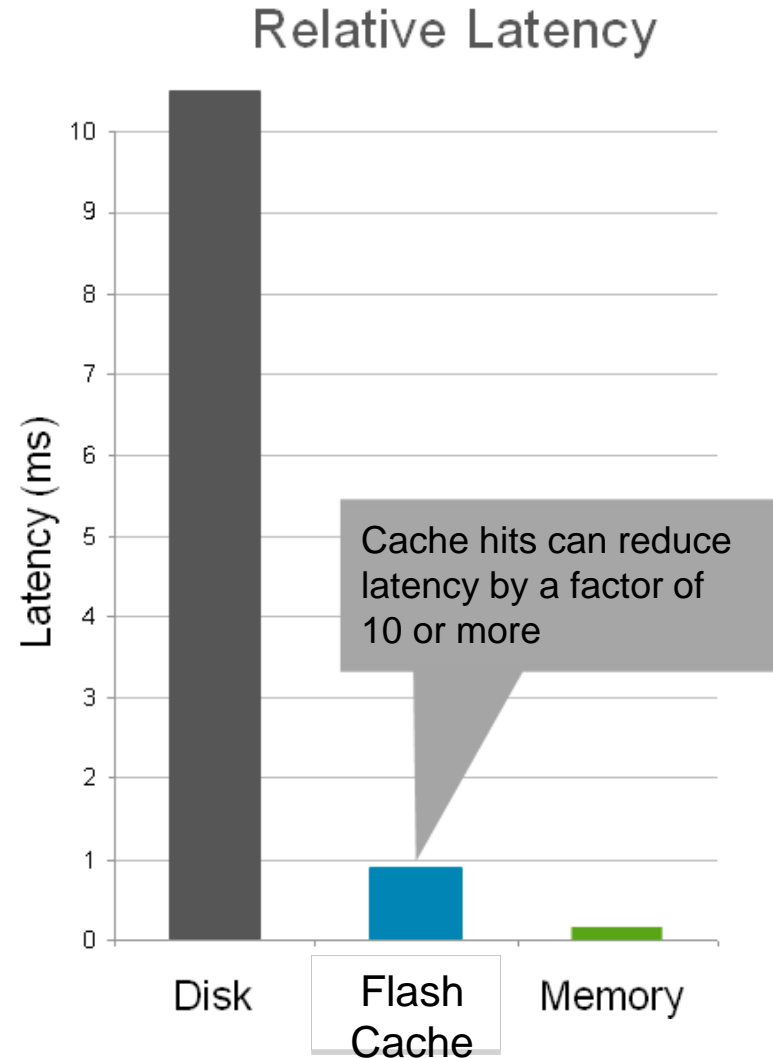
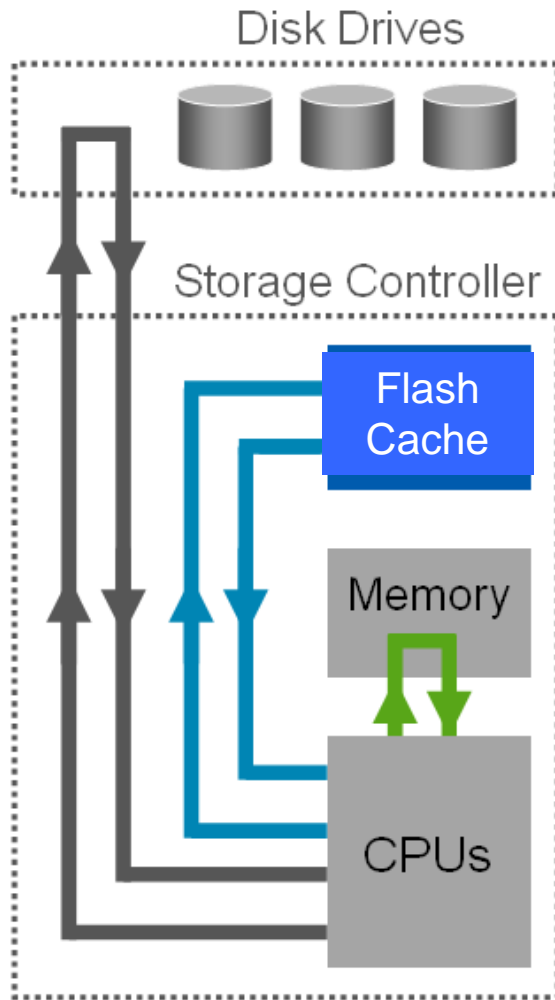
- Functions as an intelligent read cache for data and metadata
- Automatically places active data where access can be fast
- Provides more I/O throughput without adding high-performance disk drives to a disk-bound storage system
- Effective for file services, OLTP databases, messaging, and virtual infrastructure

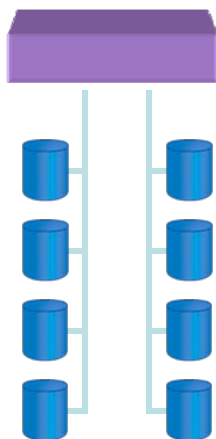


# Deciding Between SSD and Cache

<b>SSD Persistent Storage</b>	<b>Controller-Based Read Caching</b>
<p><b>Good Fit When ...</b></p> <ul style="list-style-type: none"><li>■ Random I/O intensive workload</li><li>■ Every read must be fast</li><li>■ Active data is known and fits into the SSD tier</li><li>■ Active data is known, is dynamic, and ongoing administration is OK</li><li>■ Upside of write acceleration desired</li></ul>	<p><b>Good Fit When ...</b></p> <ul style="list-style-type: none"><li>■ Random read intensive workload</li><li>■ Improving <i>average</i> response time is adequate</li><li>■ Active data is unpredictable or unknown</li><li>■ Administration-free approach is desired</li><li>■ Start small and scale up</li></ul>

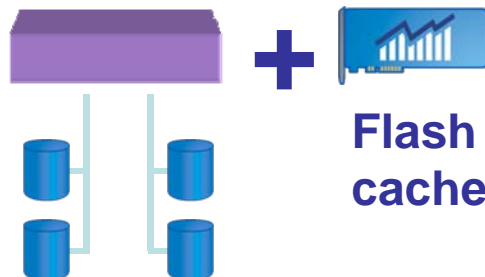
# Reduce Latency with Flash Cache





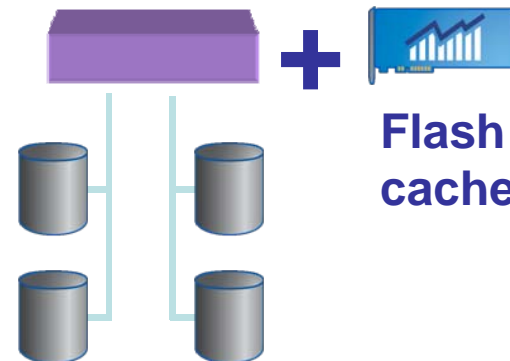
## Configure with **FC** Disks Only

- Additional disk drives provide IOPs
- Inefficient use of storage capacity, power, and space



## Configure with **FC** Disks And **Flash Cache**

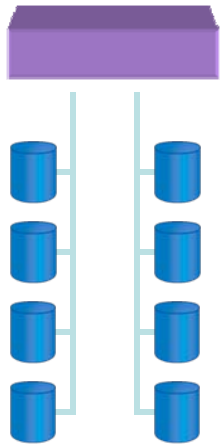
- Disks provide capacity/IOPs
- Cache delivers more IOPs and speeds response times
- Achieve cost savings for storage, power, and space



## Configure with **SATA** Disks and **Flash Cache**

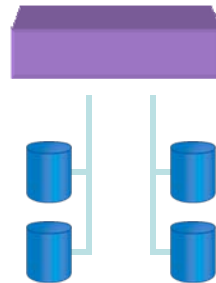
- More storage capacity
- Cache provides IOPs boost for SATA disk drives
- Achieve cost savings for storage, power, and space

# Use case: Scale Performance of Disk-bound Systems



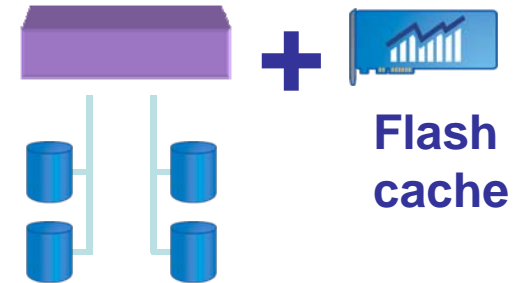
## Add Spindles

- 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

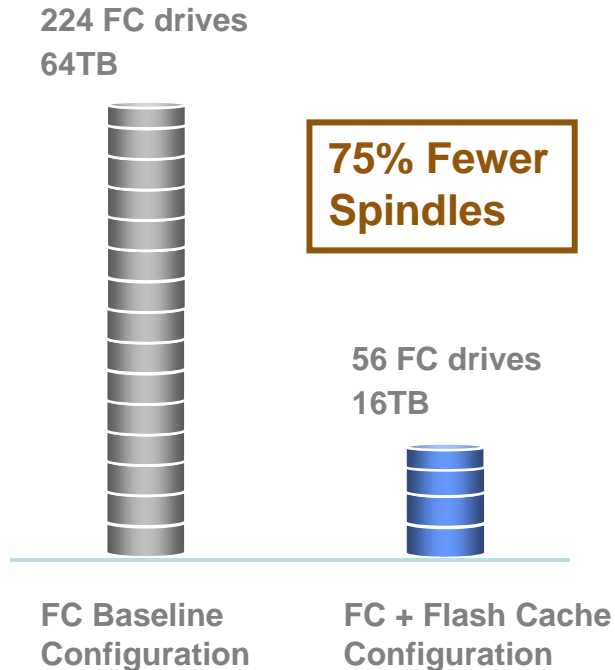


## Add Flash Cache

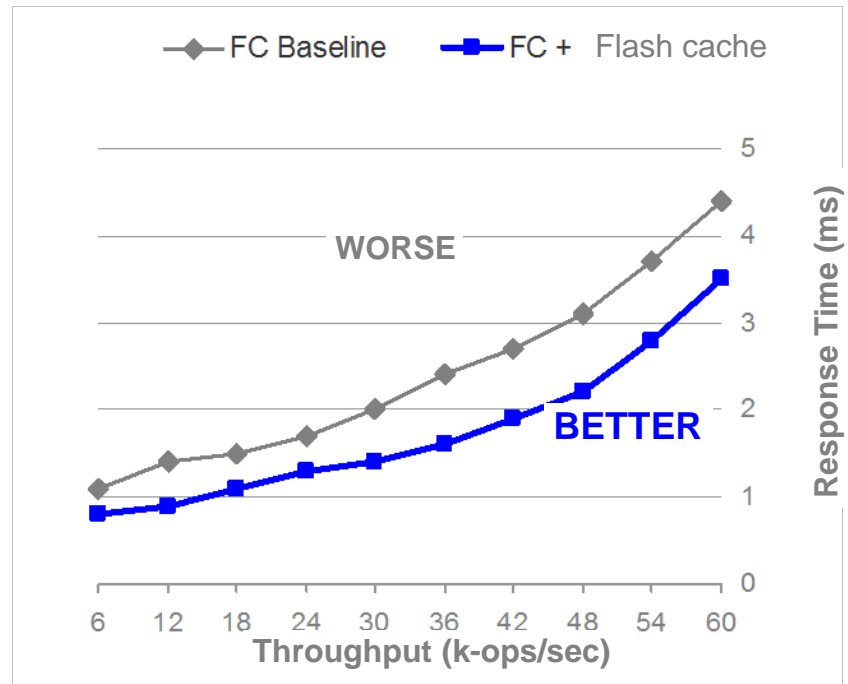
- Use cache to provide more IOPs
- Improves response times
- Uses storage efficiently
- Achieves cost savings for storage, power, and space

# FC HDD plus Flash Cache Example

## Benchmarked Configurations



## SPECsfs2008 Performance



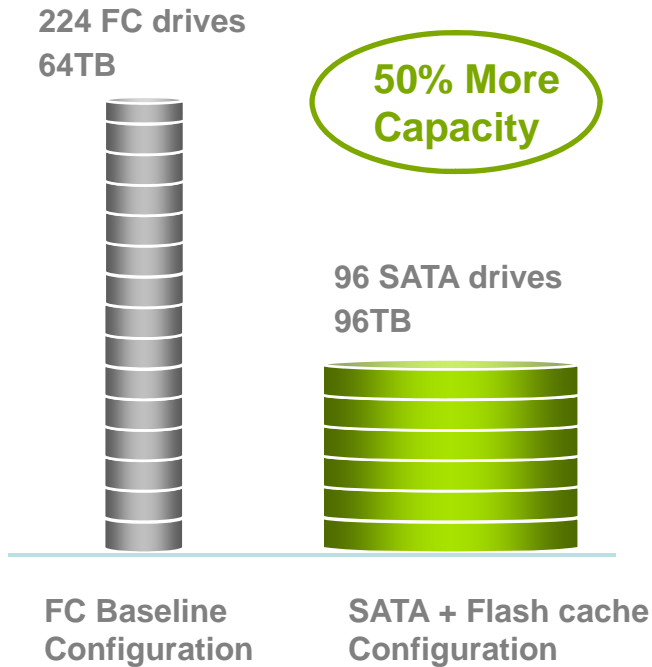
- Purchase price is **50% lower** for FC + Flash cache compared to Fibre Channel baseline
- FC + Flash cache yields **67% power savings** and **67% space savings**

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

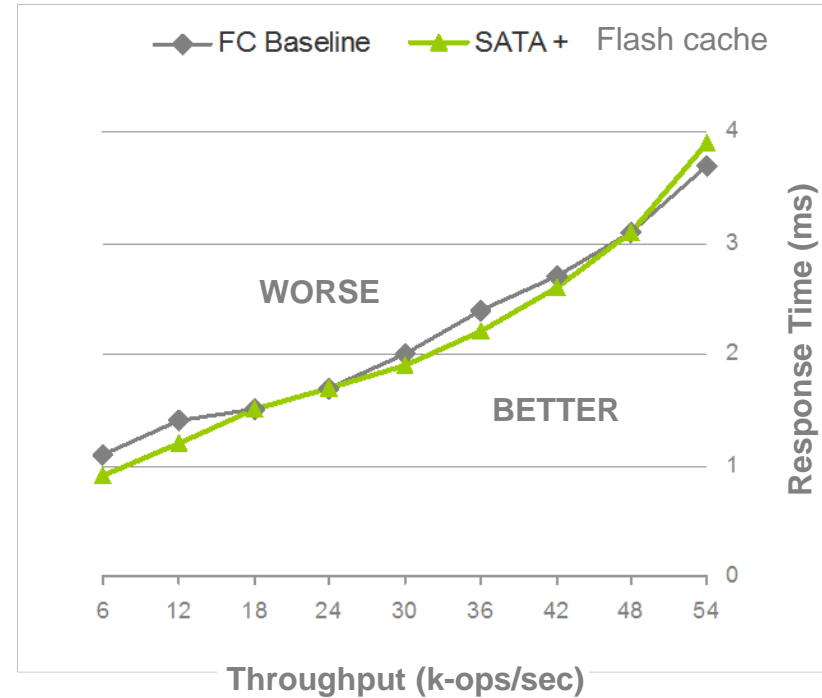
SPEC® and SPECsfs2008® are trademarks of the Standard Performance Evaluation Corp.

# SATA HDD plus Flash Cache Example

## Benchmarked Configurations



## SPECsfs2008 Performance

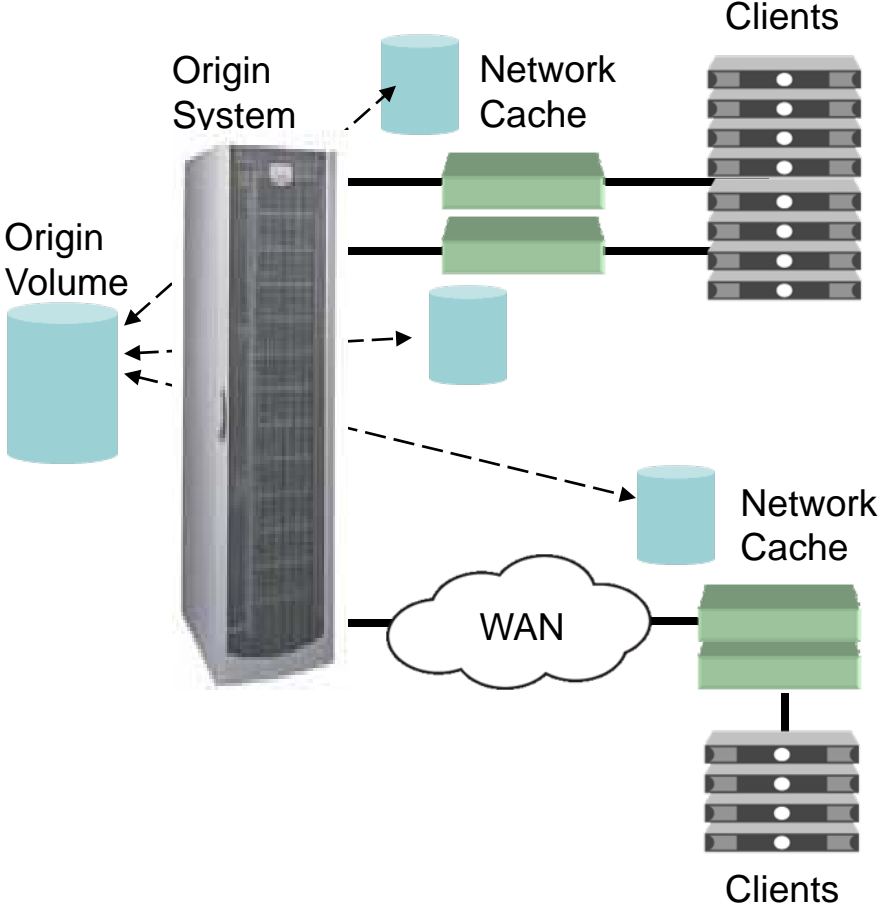


- Purchase price is **39% lower** for SATA + Flash cache compared to Fibre Channel baseline
- SATA + Flash cache yields **66% power savings** and **59% space savings**

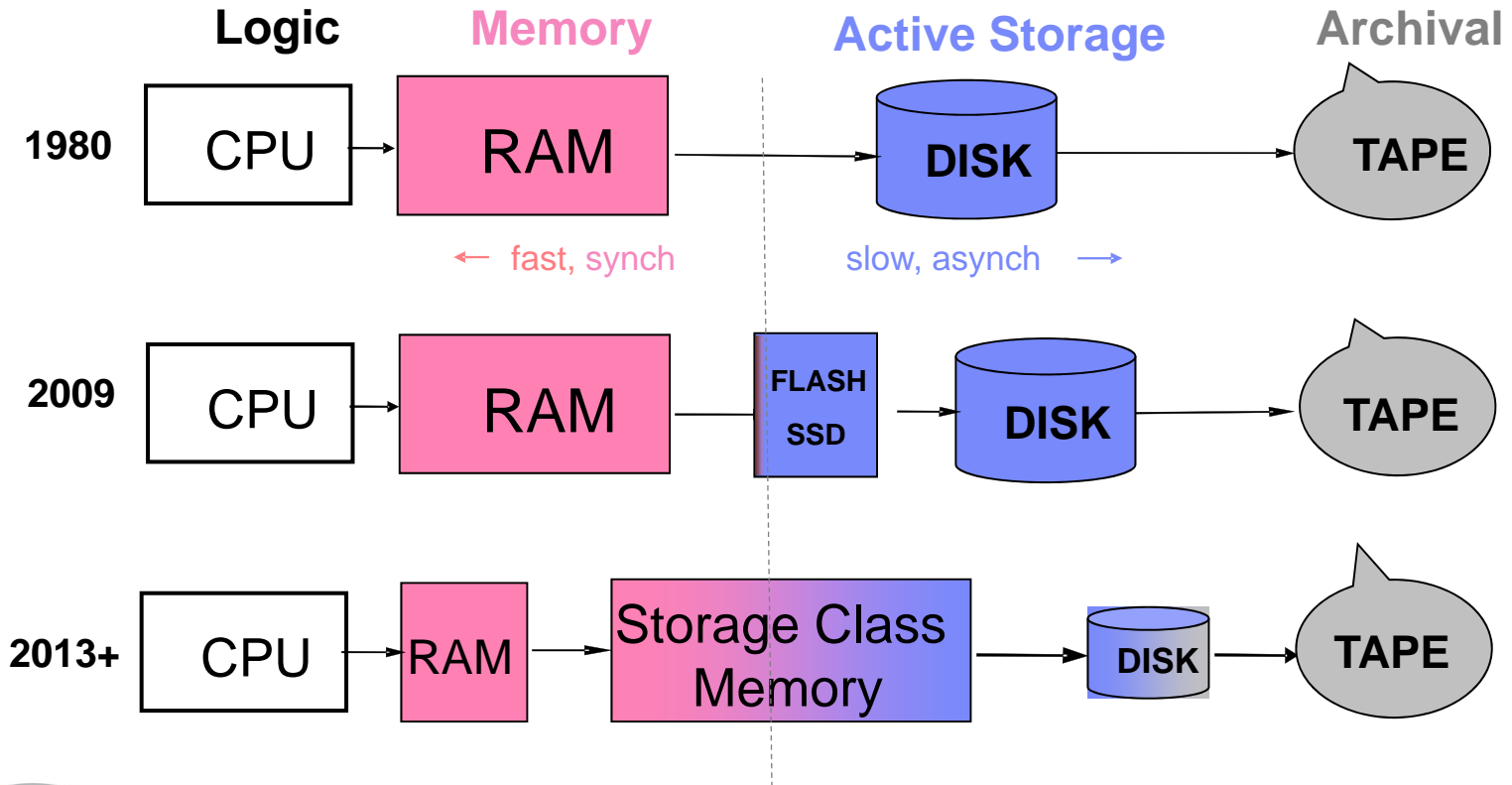
For more information, visit <http://spec.org/sfs2008/results/sfs2008nfs.html>.

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

# Network Cache Topology



# System Evolution

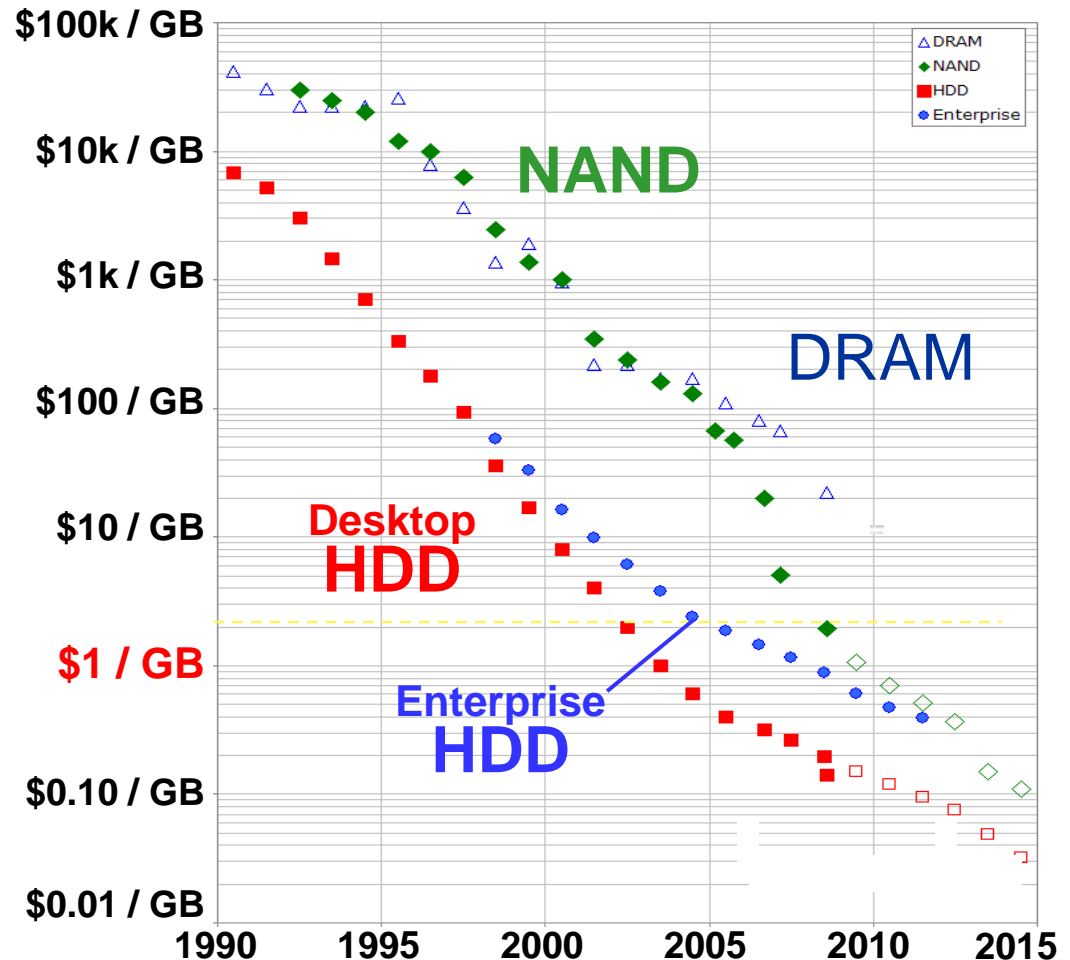


**Check out SNIA Tutorial:**

**“The Future of Solid State Storage”**

## Cost determined by

- cost per wafer
- # of dies/wafer
- memory area per die [sq.  $\mu\text{m}$ ]
- memory density [bits per  $4F^2$ ]
- patterning density [sq.  $\mu\text{m}$  per  $4F^2$ ]



Check out SNIA Tutorial:  
“The Future of Solid State Storage”

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

- 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...

- Please send any questions or comments on this presentation to SNIA: [tracksolidstate@snia.org](mailto:tracksolidstate@snia.org)

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

**- SNIA Education Committee**

**David Dale  
Jeff Kimmel  
Mark Woods**

**Phil Mills  
Chris Lionetti  
Amit Shah**