# Flash vs Flash:

## What's in it for me?

DATAKINETICS
DATA PERFORMANCE & OPTIMIZATION

www.DKL.com

JBaker@DKL.com

# Agenda

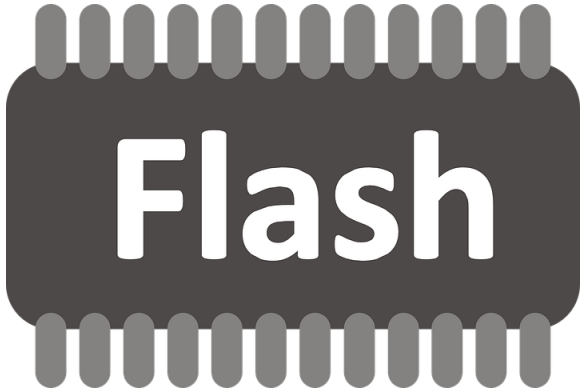
- Flash: What's in a name?

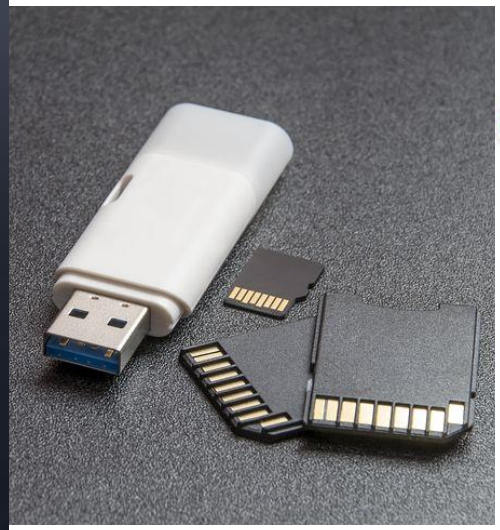
- Flash memory:
  - History
  - How does it work?  Benefits and drawbacks
  - Uses today

- Flash as a storage medium
  - Comparison to spinning disks
  - Measurements from real data

# Flash: What's in a name?





# Memory “Lingo”

- RAM: Random Access Memory
  - Fastest and most common
  - Requires constant power source (= “volatile”)
- ROM: Read-Only Memory
  - Slow and read only (data may not be changed)
  - non-volatile
- EEPROM: Electronically Erasable Programable Read-Only Memory
  - non-volatile*
  - Faster than ROM but slower than RAM

- FLASH is essentially a type of EEPROM

# Flash memory: history

- Invented by Fujio Masuoka of Toshiba in 1984
  - 1989: first Solid State Drive (SSD) fully emulating HDD
  - 1991: Flash revenue at $170 million
  - 1992: PC’s begin using flash for BIOS
  - 1995: first flash-based digital camera
  - 1997: flash-based cell phones, PDA, MP3 player, USB sticks
  - 2000: Flash revenue exceeds $10 billion
  - 2008: EMC offers “Enterprise Flash Drives” in storage arrays
  - 2011: All-flash arrays
  - 2017: Flash revenue exceeds $50 billion

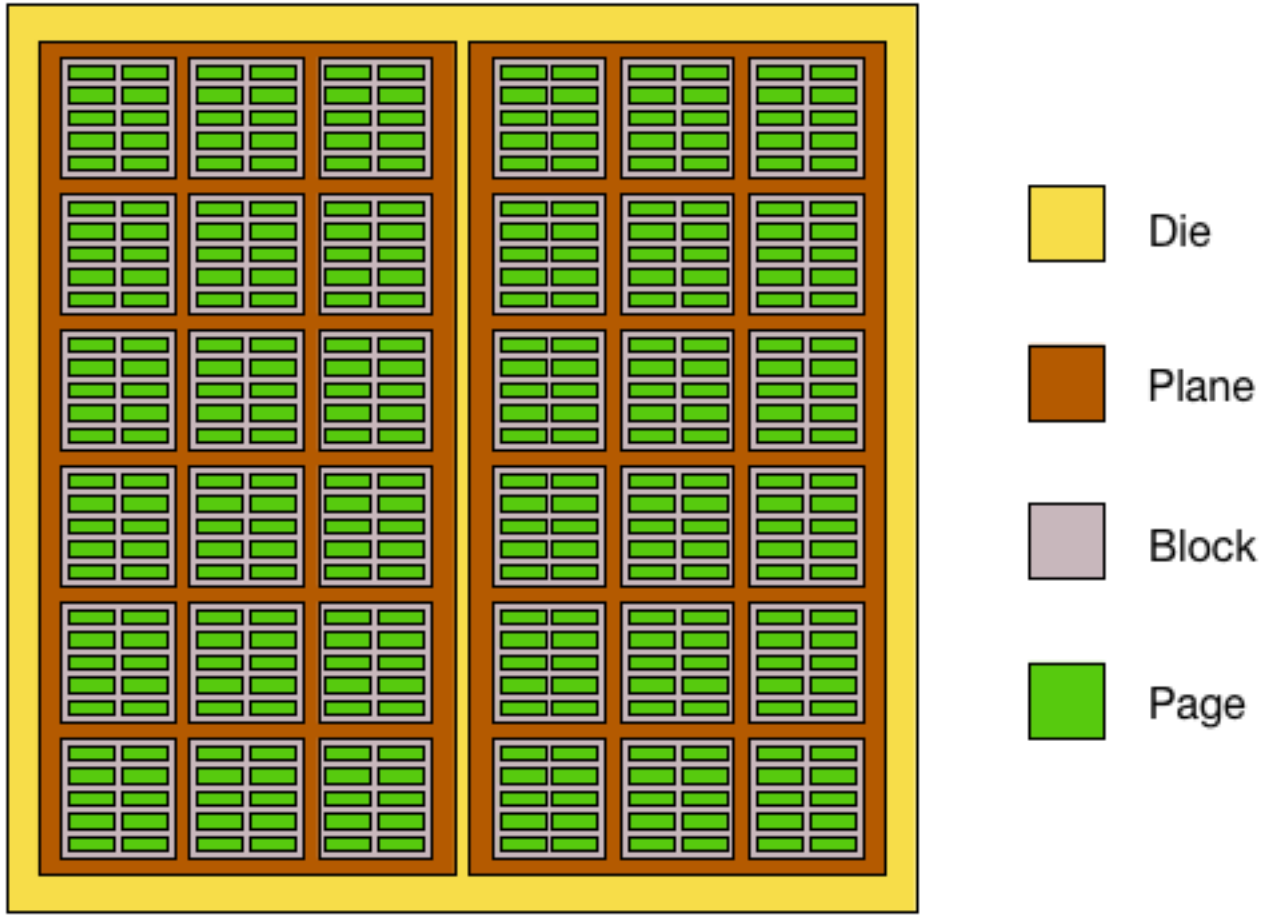
# NAND vs NOR

- Named after their respective gates
- NOR-based flash
  - byte-addressable (suitable for code execution)
  - Relatively slow erase and write speeds (suitable for static data, e.g. router memory)
  - Low density (smaller memory footprints, e.g. BIOS)

- NAND-based flash
  - block addressable (data only; not code)
  - Significantly faster, higher density and cheaper
  - The bulk of common flash today (USB sticks, SD cards, SSD’s, etc.)

# Flash memory: principles

- Flash is a series of electronic memory cells
- Individual pages may be read or written to – BUT page must be empty before writing
- If page contains data (even old, unused data, entire block must be erased




https://www.mikroe.com/blog/this-nand-nor-that-nand

# NAND flash flavours

- Flash is a series of electronic memory cells
  - Single-Level Cells (SLC) store a single bit per cell
  - Multi-level Cells (MLC) store 2 bits per cell
  - Triple-level (or three-level, TLC) Cells store 3 bits per cell


https://www.mikroe.com/blog/this-nand-nor-that-nand

# Comparison

| | SLC | MLC | TLC |
|---|---|---|---|
| Bits per cell | 1 | 2 | 3 |
| P/E Cycles | 100,000 | 3,000 | 1,000 |
| Read Time | 25 µs | 50 µs | ~75 µs |
| Program Time | 200-300 µs | 600-900 µs | ~900-1350 µs |
| Erase Time | 1.5-2 ms | 3 ms | 4.5 ms |


Source: http://www.anandtech.com/show/6337/samsung-ssd-840-250gb-review/2

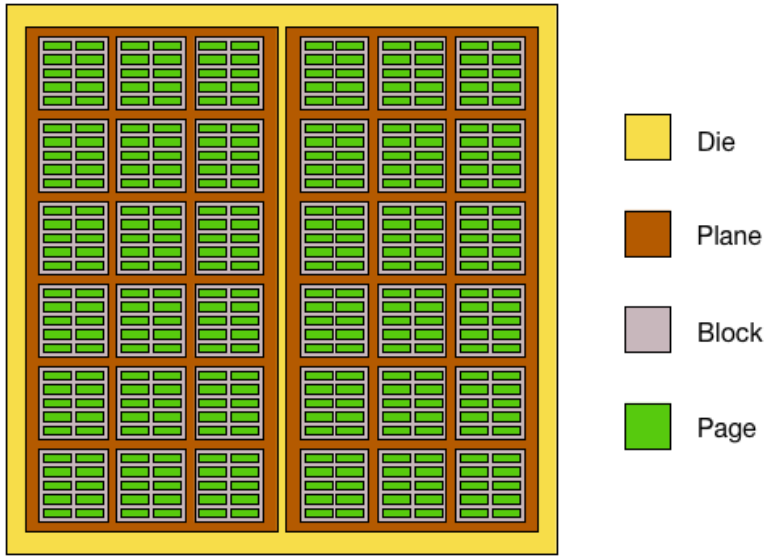
https://www.mikroe.com/blog/this-nand-nor-that-nand

# Program/Erase (P/E) Cycles

- Writing ("programing") and erasing cells requires a fairly high voltage which stresses the transistors
- Eventually, they wear out
- P/E cycles range from 100 or so for low-duty NOR to 100,000 or more for enterprise NAND

# Addressing the P/E Problem

- ***Wear leveling***: logical writes (even successive writes to the same logical page/track) will be distributed across the die rather than to the same physical page
- Block erases will be deferred as long as possible (until most/all pages are marked for erasure
- Controller maintains track tables/pointers




- MLC for enterprise, TLC for consumer/commodity
  - SLC? Originally yes (STEC*). MLC half price (overprovision)
- Capacity: A 500GB SSD will contain more than 500GB of flash
- RAID and RAIM

# Don’t throw away tape

- Flash memory is listed as non-volatile (i.e. it does not require a constant power source to retain its data)
- This is not ***literally*** true
- Left without power (i.e. sitting on a shelf) for several years, the cells may “drain”

**FLASH is not for archives**

# Mainframe Paging: Flash vs Flash

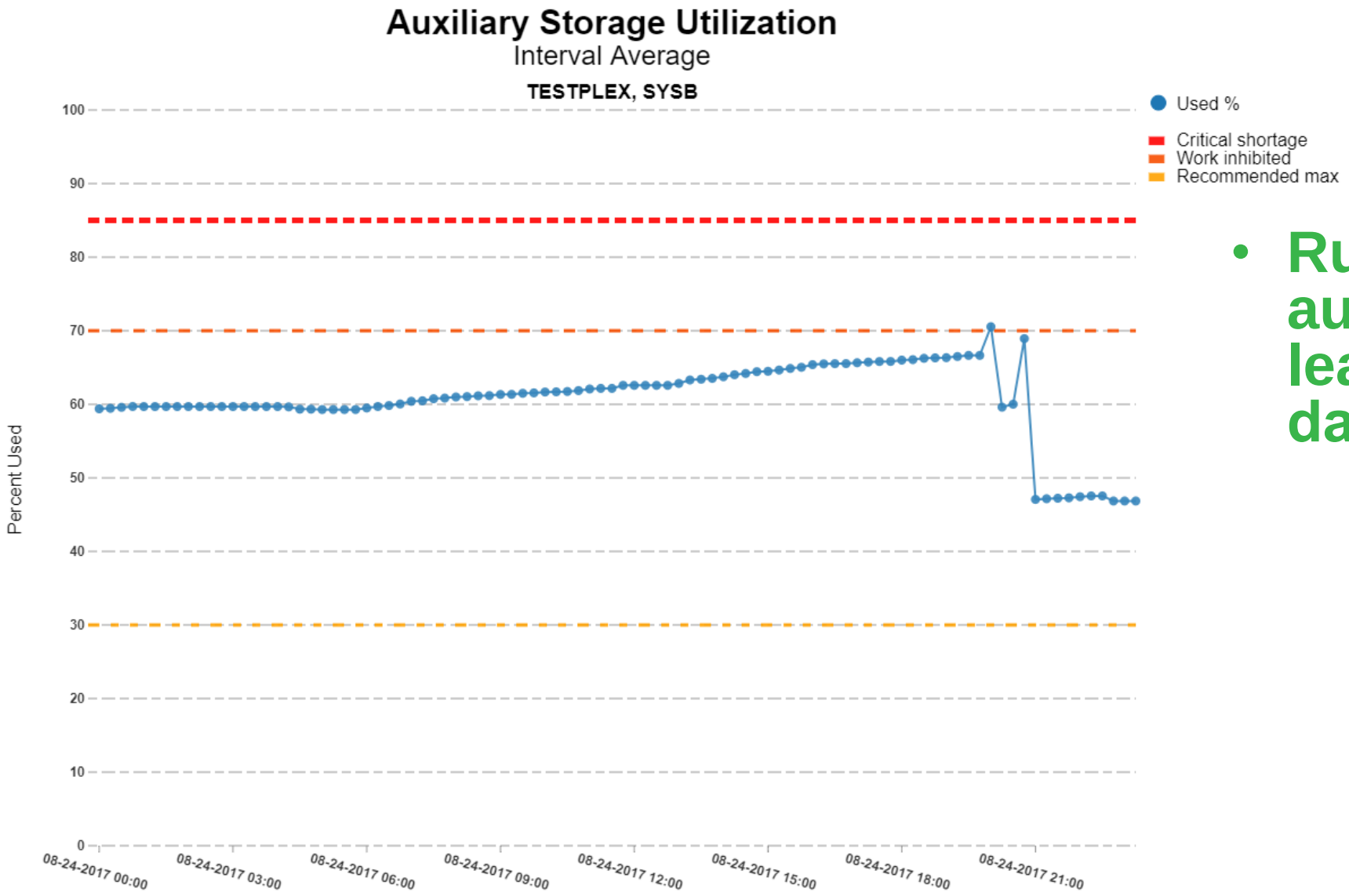
## EC12 and z13
### "Flash Express"

- Flash Express are SSD drives connected to the PCIe bus
- Like traditional paging volumes, the Auxilliary Storage Manager (ASM) still manages access via SSCH instruction but no physical channels are used
- Primary benefits are reduced latency as external I/O is avoided for paging as well as speed benefits of SSD vs spinning disk

## z14
### "Virtual Flash Memory (VFM)"

- Not actually flash but rather the same DRAM chips as main memory (faster but volatile)
  - literally the same hardware but less expensive than main memory
  - Still, only an auxiliary data pool; data must be transferred to main memory for execution (not directly addressable)
- Faster access than Flash Express as no SSCH required
- Like Expanded Memory from the past

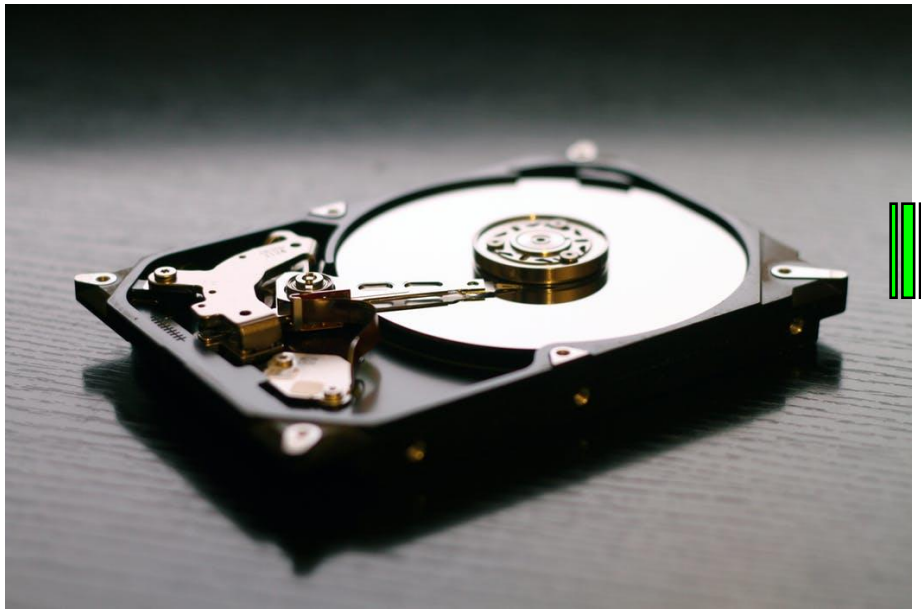
# Aux Storage Use




- **Running out of auxiliary storage can lead to a very bad day…**

# Flash as primary storage

- The main plan for Flash was always to replace spinning disks
- NAND Flash memory is packaged, with a controller, inside a traditional form-factor HDD enclosure
- To any computer, the device identifies and responds just like a spinning disk drive – just much faster

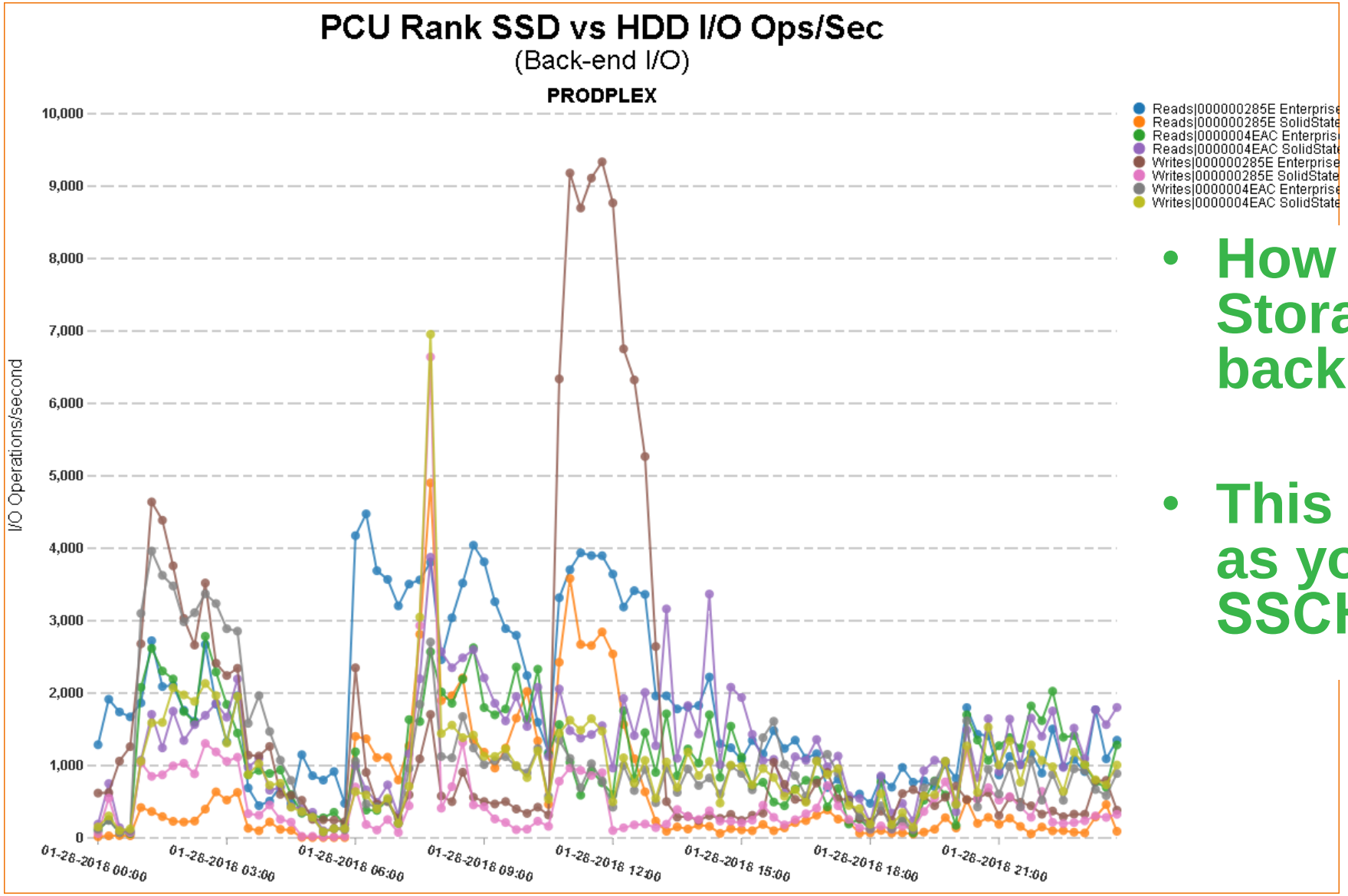


# Flash Candidates

- Random reads
- High disconnect
- When tuning, cache and buffers haven’t helped

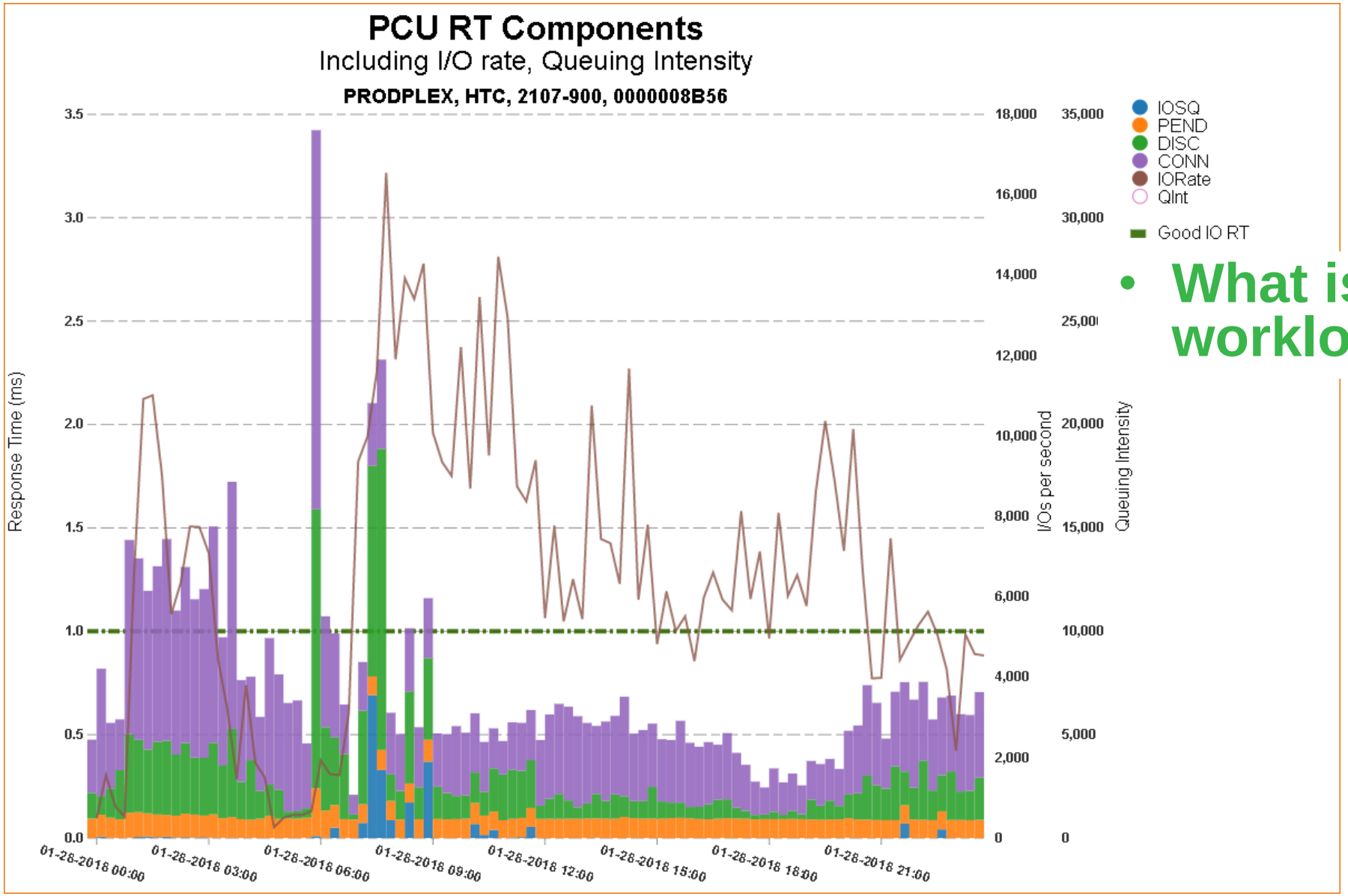
= cache unfriendly

# Disk Activity by Type



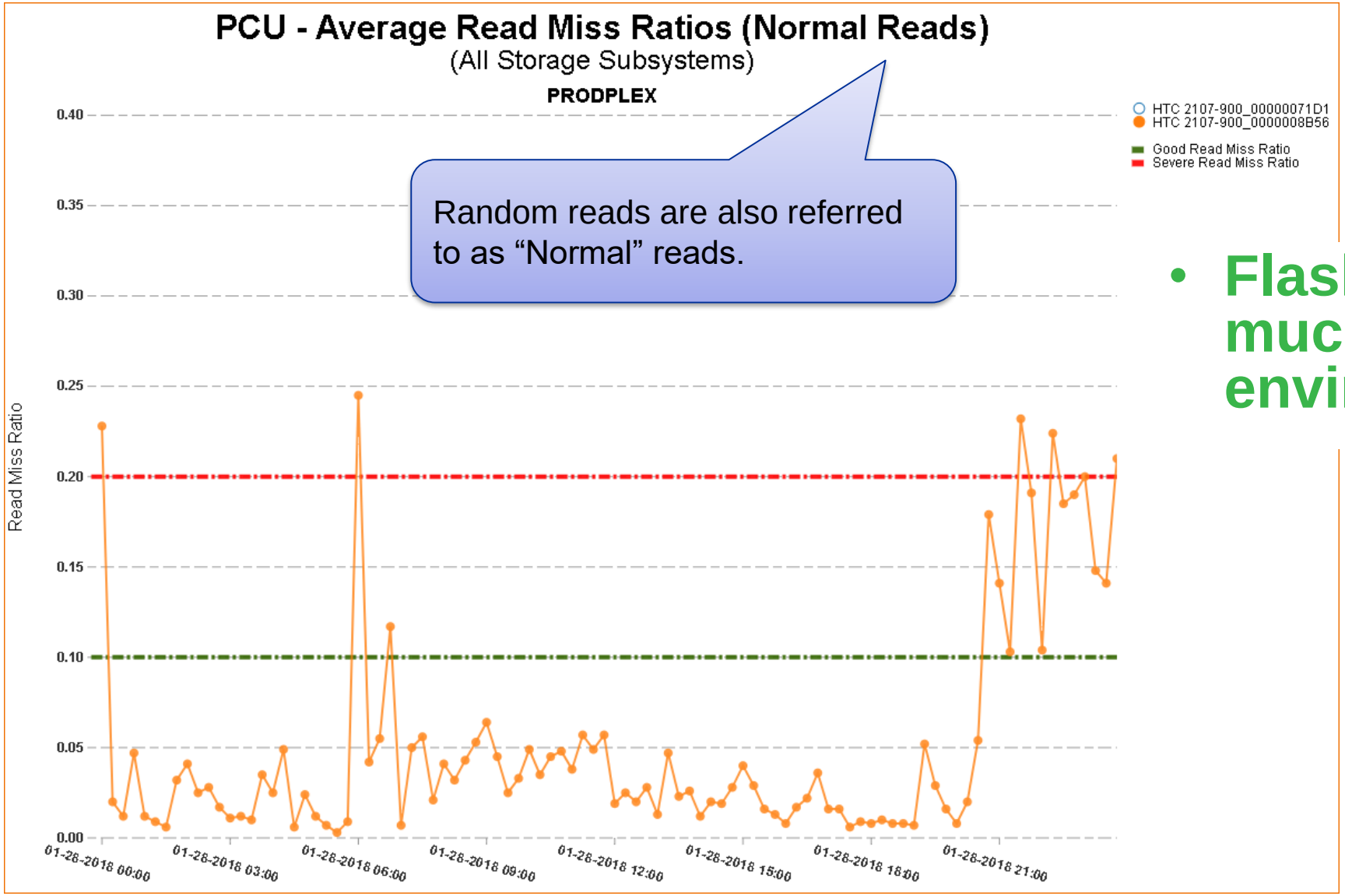

- **How busy is your Storage Subsystem back-end?**
- **This is not the same as your traditional SSCH or "IOPS" rate**

# Response Time Components




- **What is your workload profile?**

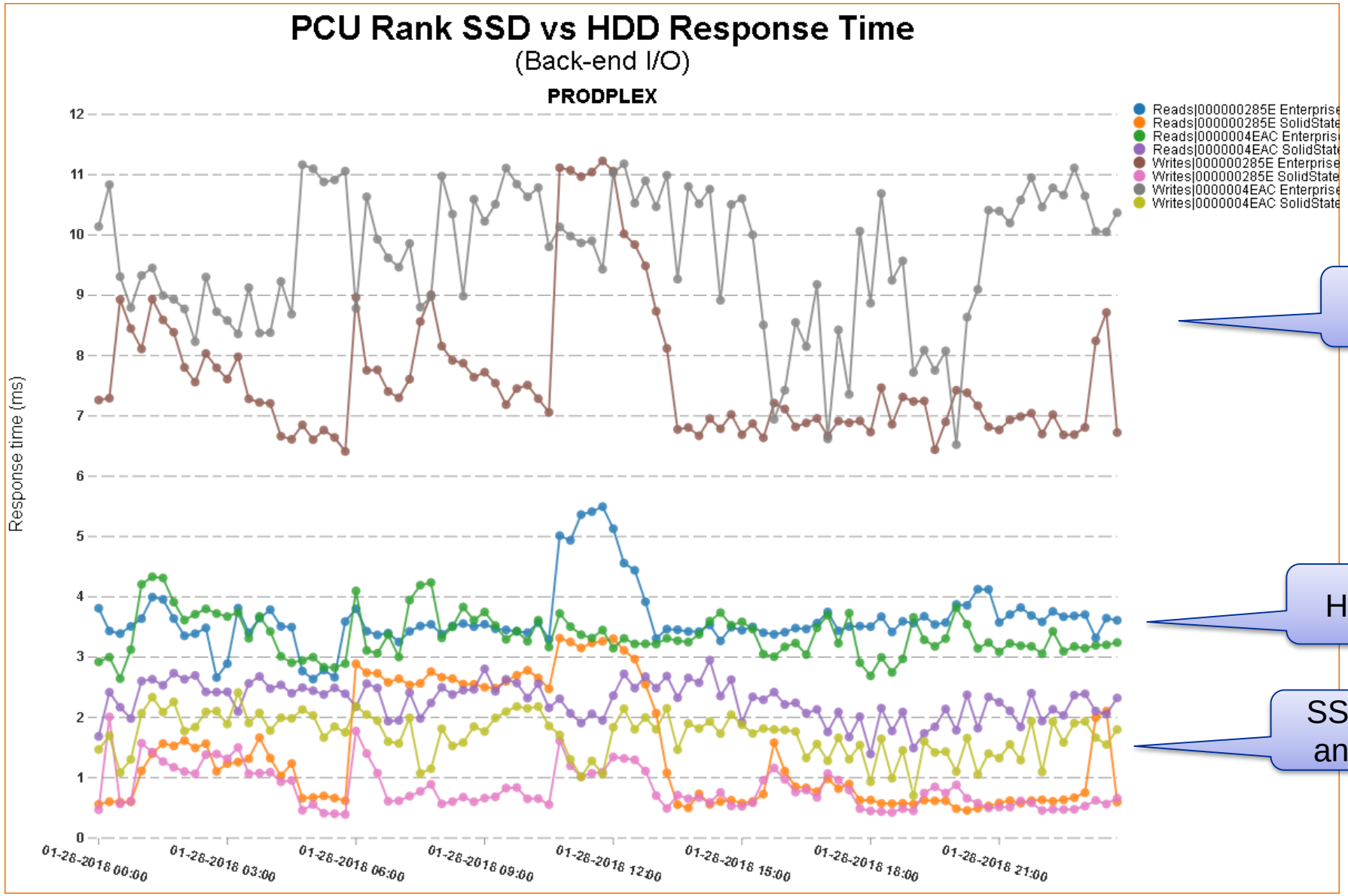
# Cache Hit Ratio




Random reads are also referred to as “Normal” reads.

- **Flash will not help much in high hit-ratio environments**

# Response by Disk Type




HDD Writes

HDD Reads

SSD Reads and Writes

# SSD vs HDD: The Economics

- $/GB has been the traditional metric
  - HDD still cheaper: SSD around 10x price of HDD today (average)

- "I personally will never foresee the day, […] when there is a crossover [between the per-gigabyte cost of NAND and rotating magnetic media].
  - Dave Morton. Seagate SVP Finance, 2015
  - https://www.kitguru.net/components/hard-drives/anton-shilov/seagate-ssds-will-never-match-per-gigabyte-costs-of-hdds/

  - https://www.computerworld.com/article/3182207/data-storage/cw50-data-storage-goes-from-1m-to-2-cents-per-gigabyte.html

# SSD vs HDD: $/GB

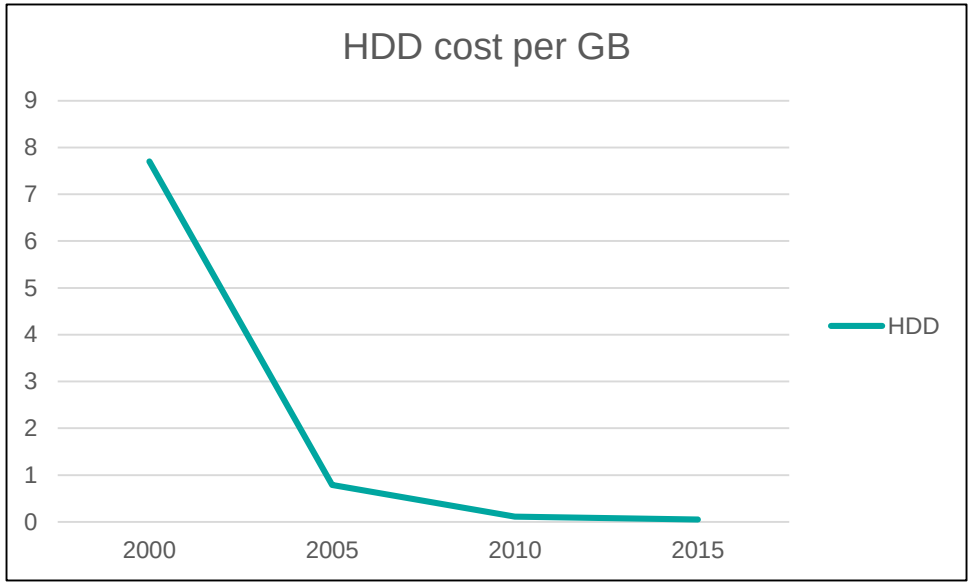
## Storage and memory costs through the years

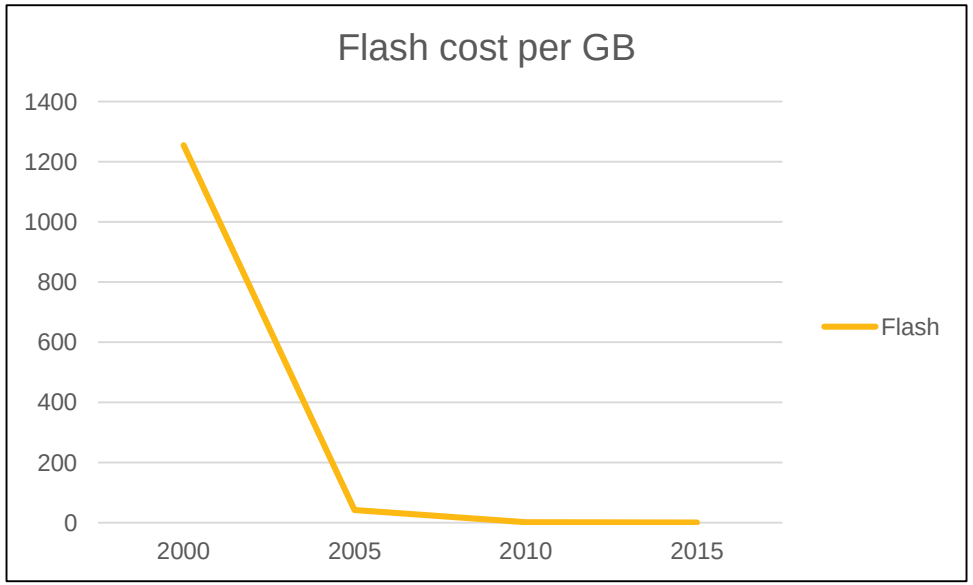
BY THE NUMBERS | Source: John C. McCallum, IDC

[ Price per gigabyte ]

| Year | Disk Drives | DRAM | NAND flash |
|---|---|---|---|
| 1965 | n/a | $2.64B | – |
| 1966 | $1.05M | n/a | – |
| 1970 | $185,000 | $734M | – |
| 1975 | $180,000 | $180M | – |
| 1980 | $202,000 | $6.48M | – |
| 1985 | $40,000 | $477,500 | – |
| 1990 | $4,400 | $78,400 | – |
| 1995 | $277 | $31,633 | – |
| 2000 | $7.70 | $1,031 | $1,255 |
| 2005 | $0.79 | $158 | $42 |
| 2010 | $0.11 | $18.87 | $1.76 |
| 2015 | $0.05 | $5.22 | $0.38 |
| 2017 | $0.02 | $4.90 | $0.32 |

© COMPUTERWORLD



# SSD vs HDD: Other considerations

- Both HDD and SSD prices continue to drop
- Greater SSD adoption fuels demand which fuels innovation
  - TLC, QLC, 3D NAND
  - Economies of scale – it worked with RAID!

- Don’t discount $ per I/O
- Consider electricity and cooling costs as well

# Another perspective: Performance vs space requirement

- Access Density = I/O per second per GB of storage

- Consider a requirement for 10,000 I/O's per second
  - A 15K RPM disk can handle about 200 I/O's per second with decent response
  - You will need 50 disks to meet this requirement – ***whether you need the space or not***
  - The smallest available 15K HDD's are 300 GB. 50 * 300 = 15 TB

- Perhaps 5 SSD's at 3 TB each could also accomplish this

- HDD's cannot compete as capacities increase

# All-Flash Arrays

- Traditional controllers were designed for spinning disks
- Internal structure (primarily device adapters) limit SSD throughput

- Vendor implementations vary but all are addressing the issue
  - ***Flash Enclosures, Bricks, Flash Packs***
  - Purpose-built adapters designed to overcome previous limits
  - Higher cost offset by performance, electricity and cooling benefits

- Look for the new bottleneck!

# Thank you!

www.DKL.com

JBaker@DKL.com