Intra-Disk Parallelism: A Green Storage Solution for Data Centers

Sudhanva Gurumurthi

University of Virginia

E-mail: gurumurthi@cs.virginia.edu

University of Virginia
Department of Computer Science

Research Team

- Faculty: Mircea R. Stan
- Students
  - Sriram Sankar
  - Yan Zhang

Data-Centric Computing

- Store and process massive datasets
- Concurrent accesses from many users

Traditional Application Domains

amazon.com. Expedia.com®

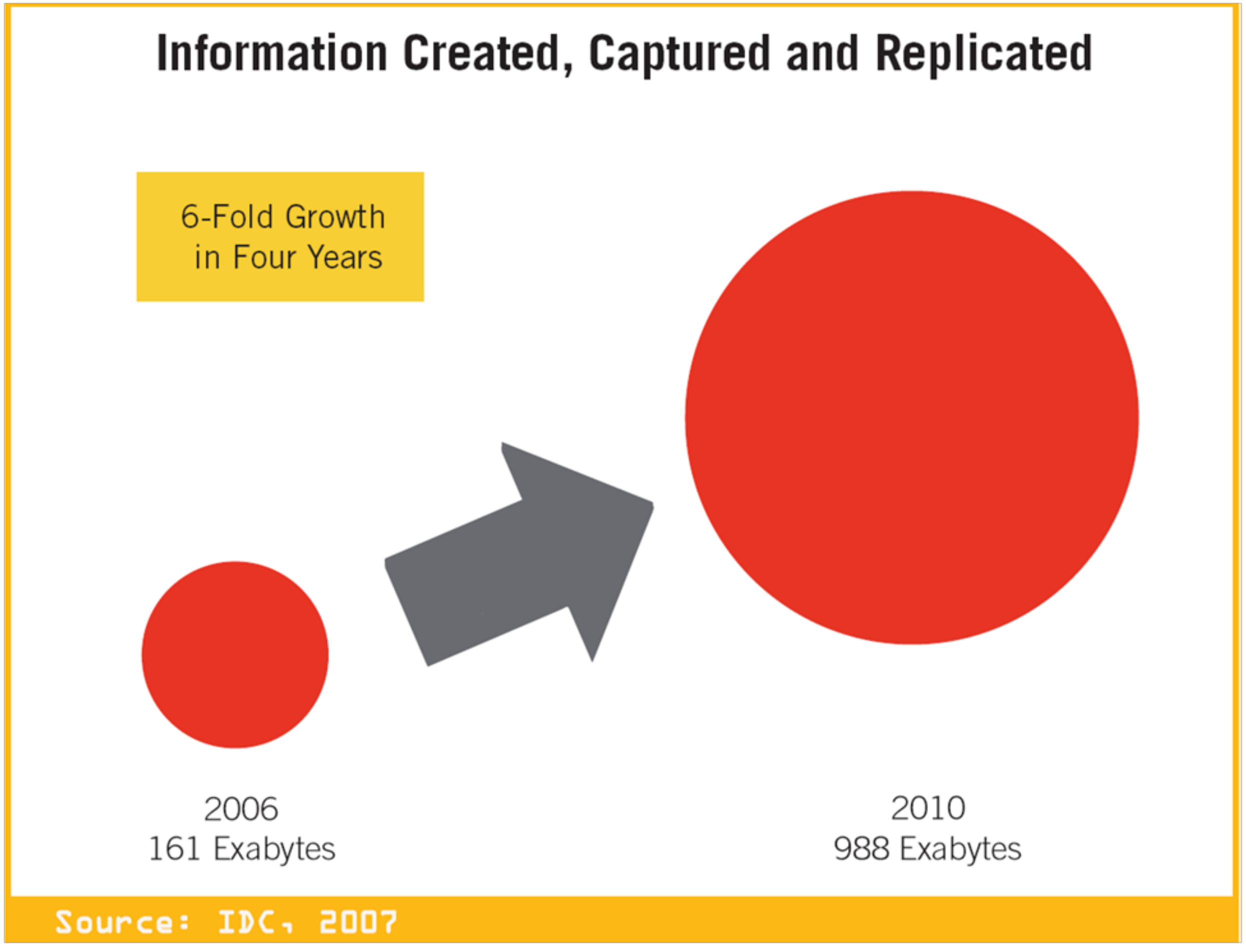
ebaY® Gmail™ by Google BETA

Emerging Application Domains

YouTube Broadcast Yourself™ flickr™ LOVES YOU

facebook Google™ Maps

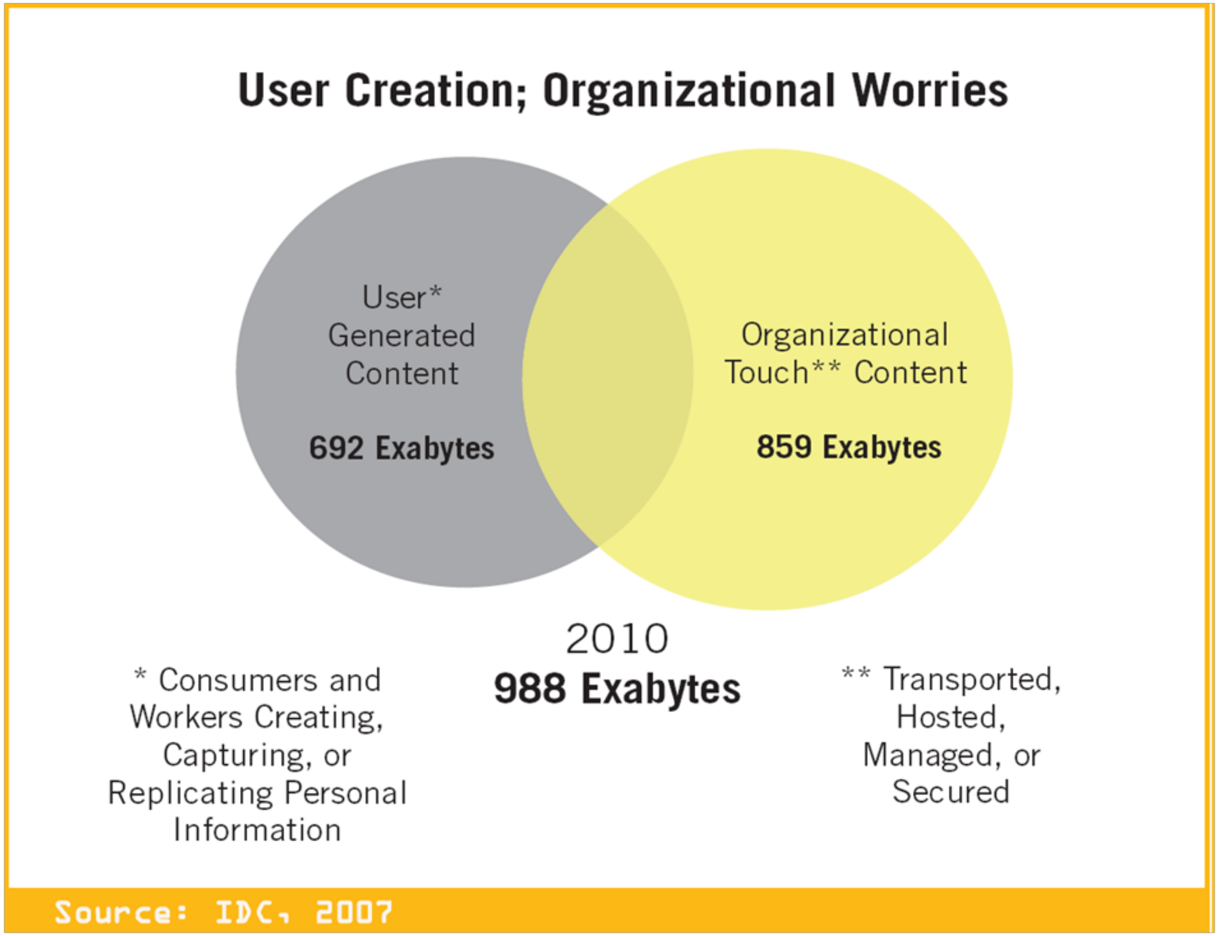
Data Keeps Growing



1 exabyte = 10^{18} bytes

Source: IDC Whitepaper, “The Expanding Digital Universe”, March 2007.

Need High-Performance and High-Capacity Storage Systems



User Creation; Organizational Worries

User* Generated Content

692 Exabytes

Organizational Touch** Content

859 Exabytes

2010

988 Exabytes

* Consumers and Workers Creating, Capturing, or Replicating Personal Information

** Transported, Hosted, Managed, or Secured

Source: IDC, 2007

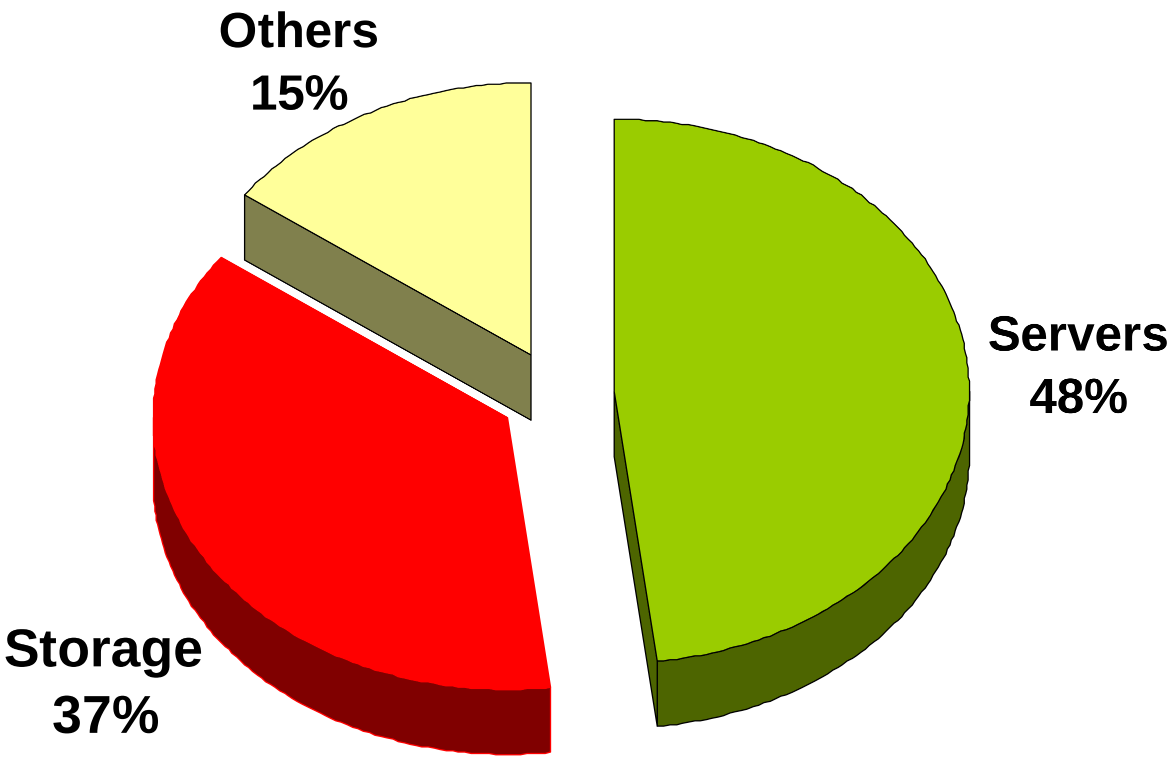
Source: IDC Whitepaper, “The Expanding Digital Universe”, March 2007.

Disks Are Slow

- Disk access time ~ **milliseconds**
  - Random I/O exposes these delays
- Approaches to Boost Storage Performance
  - Use faster disk drives
    - Not scalable due to thermal design reasons
  - Build storage arrays
    - Increases storage system power consumption
  - Short-stroking to trade disk capacity for storage performance

☞ **Higher Power Consumption**

Data Center Computing Equipment Power Consumption

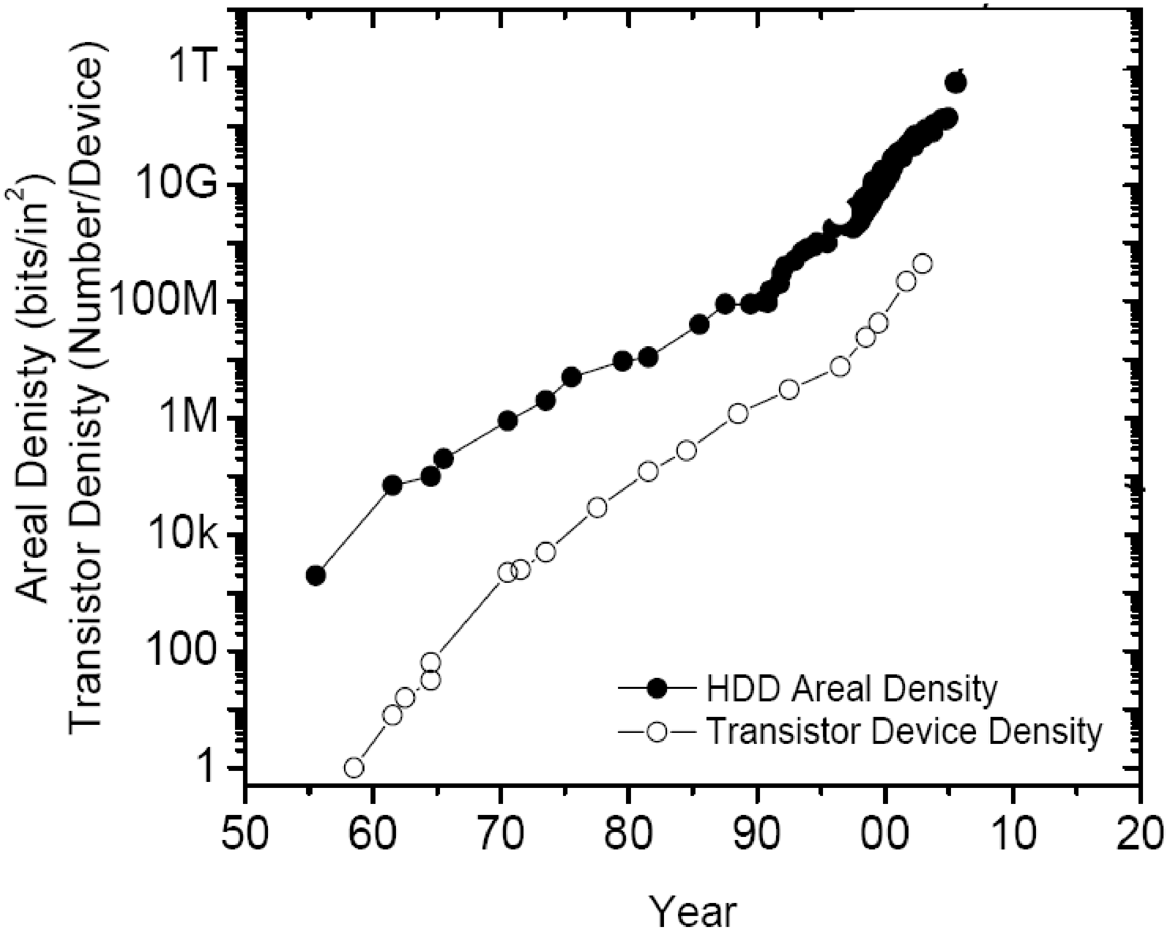


80% of the storage power is consumed by the disk arrays

Source: The green data center: Energy-efficient computing in the 21st century, Chapter 4, 2007.

Disk Drive Capacity

Moore's Law vs. Kryder's Law



Source: Mark Kryder, "Future Storage Technologies: A Look Beyond the Horizon", SNIA Storage Networking World, 2007.

Green Storage

- We would like to the storage system to:
  - Deliver high performance
  - Utilize the disk capacity to the fullest
  - Consume lower power
- Our Approach
  - Extend the architecture of conventional disk drives
  - **Intra-Disk Parallelism**

Why Not Just Use SSDs?

- SSD Benefits: Power, Performance
- Cost/GB of SSDs
  - Flash: **$3.58/GB**
  - HDD: **38¢/GB**
- HDDs will be an integral part of enterprise storage systems for a while
  - Look at HDD solutions that can **complement** SSDs

Source: Computerworld Storage, “Seagate plans SSD, 2TB hard drive for next year”, May 30, 2008.

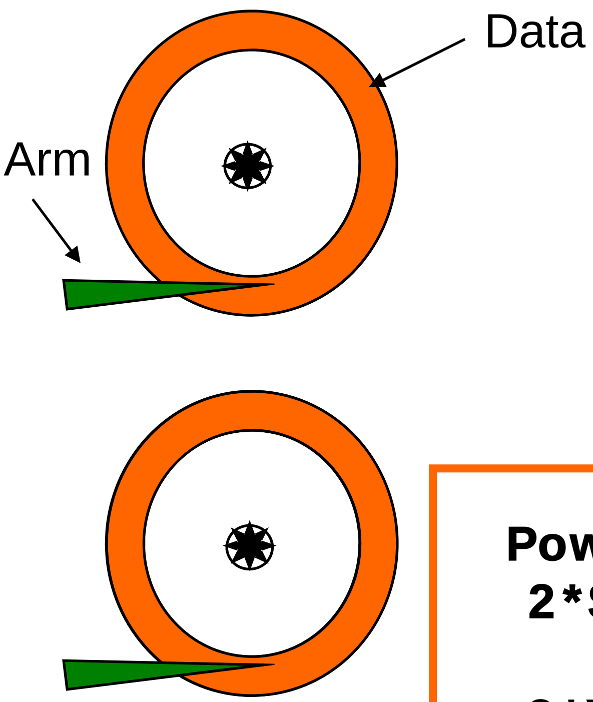
Outline

- ❐ Intuition Behind Intra-Disk Parallelism
- ❐ Historical Retrospective
- ❐ Experimental Results
- ❐ Cost and Engineering Issues
- ❐ Other Green Storage Research
- ❐ Conclusions

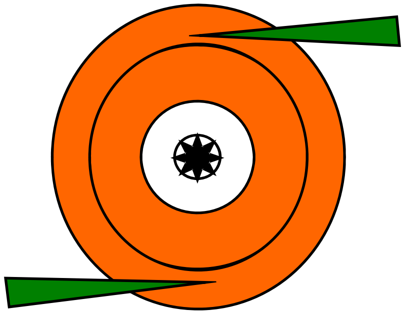
Parallelism in Storage Systems

- **Disk Request:** Seek, Rotational Latency, Data Transfer
  - All disk resources (arms, heads, channel) are dedicated for **each** request
- "Intra-Disk Parallelism" in current drives
  - Tagged command queuing
  - Read-ahead buffering
- Good parallel I/O performance in servers requires **multiple disks**

Inter-Disk vs. Intra-Disk Parallelism



Power =
2*SPM
+
2*VCM



Power =
1*SPM
+
2*VCM

History of Intra-Disk Parallelism

Disk Drive Characteristics	IBM 3380 AK4 (**1980**)	Seagate Barracuda ES (**2006**)	4-Actuator Parallel HDD (**Future?**)
Areal Density (Mb/in^2)			
Disk Diameter (in)			
Capacity (GB)			
No. of Actuators			
HDD Power (Watts)			

IBM 3380 AK4

Disk Drive Characteristics	IBM 3380 AK4 (1980)
Areal Density (Mb/in^2)	**12**
Disk Diameter (in)	**14**
Capacity (GB)	**7.5**
No. of Actuators	**4**
HDD Power (Watts)	**6,600**

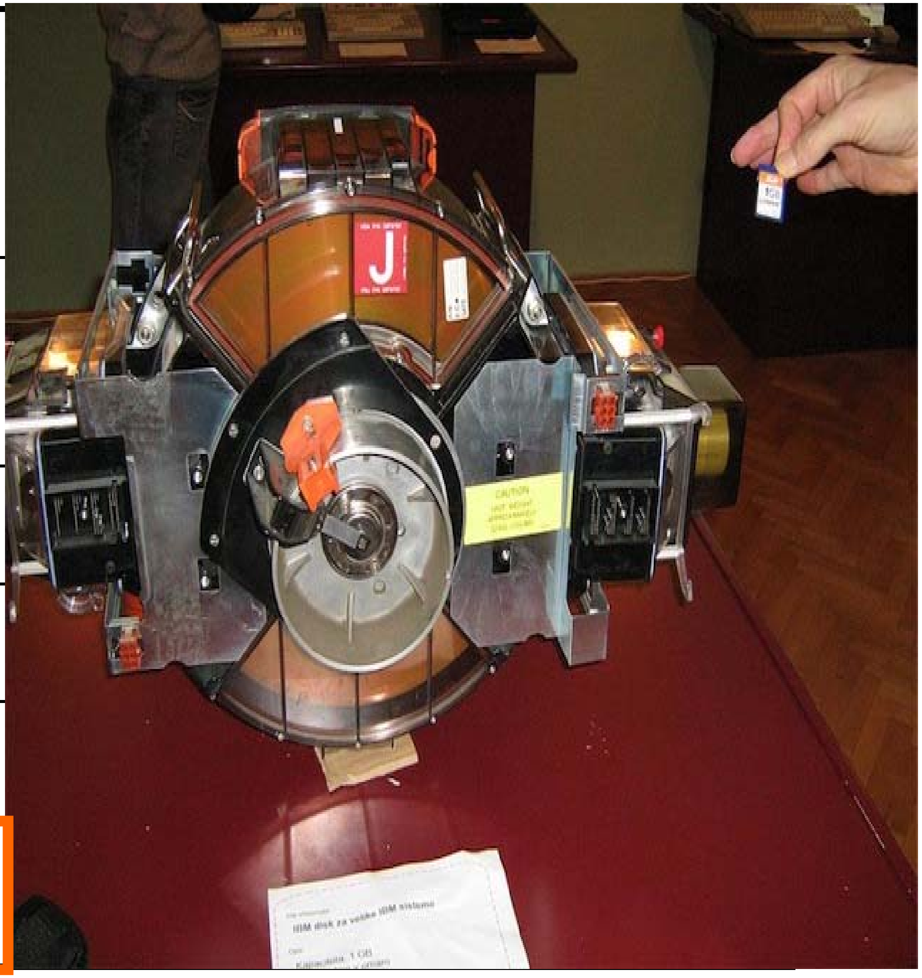


Image Source: newlaunches.com

Seagate Barracuda ES

Disk Drive Characteristics	IBM 3380 AK4 (1980)	Seagate Barracuda ES (2006)
Areal Density (Mb/in^2)	12	**128,000**
Disk Diameter (in)	14	**3.7**
Capacity (GB)	7.5	**750**
No. of Actuators	4	**1**
HDD Power (Watts)	6,600	**13**

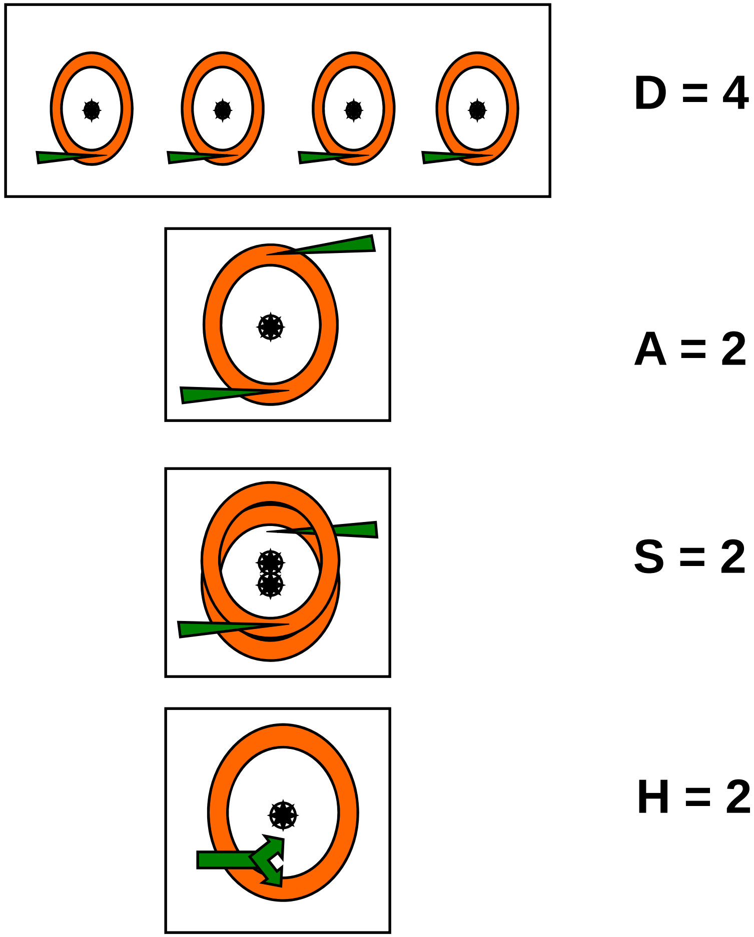
Image Source: Seagate

Hypothetical Modern Parallel HDD

Disk Drive Characteristics	IBM 3380 AK4 (1980)	Seagate Barracuda ES (2006)	4-Actuator Parallel HDD
Areal Density (Mb/in^2)	12	128,000	
Disk Diameter (in)	14	3.7	
Capacity (GB)	7.5	750	
No. of Actuators	4	1	
HDD Power (Watts)	6,600	13	

Intra-Disk Parallelism Taxonomy

- ❒ Disk/Spindle **[D]**
- ❒ Arm Assembly **[A]**
- ❒ Surface **[S]**
- ❒ Head **[H]**



Experimental Setup

- Simulator
  - Disksim with power models
- Commercial workload traces
  - Financial
  - Websearch
  - TPC-C
  - TPC-H

Impact of Data Consolidation

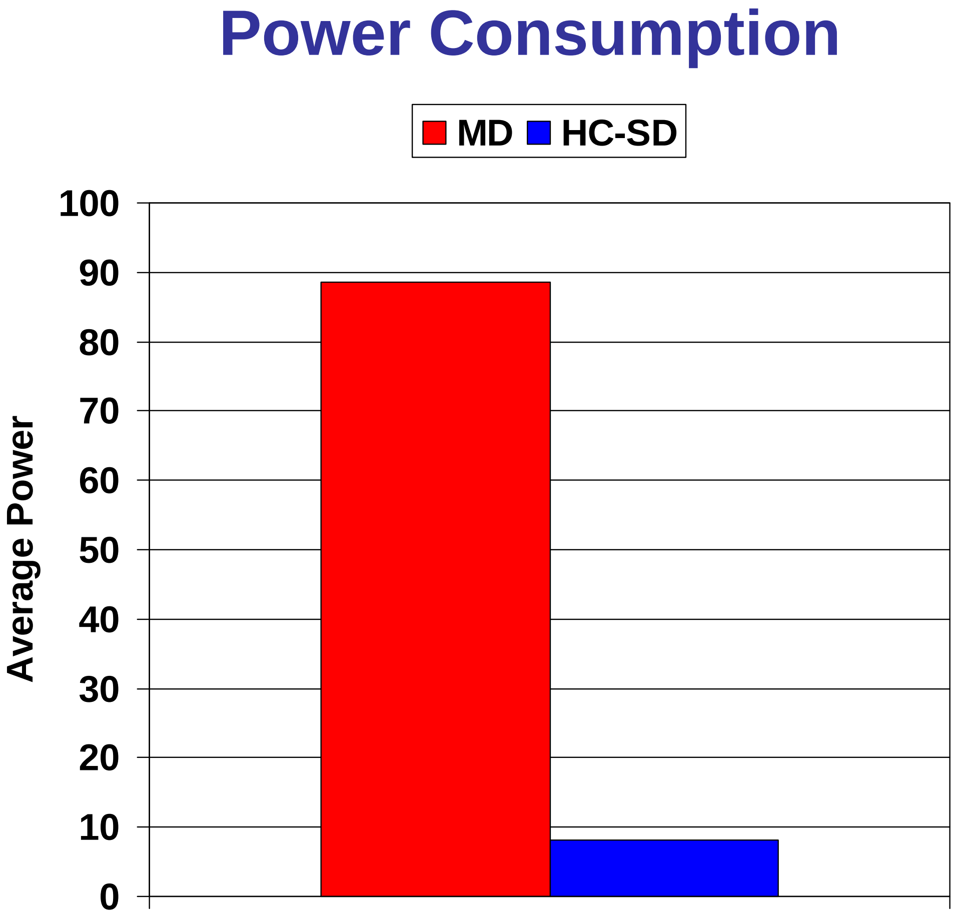
- **What if we migrate data from multiple disks on to one high-capacity drive?**
- Baseline **M**ultiple-**D**isk Configuration (***MD***)

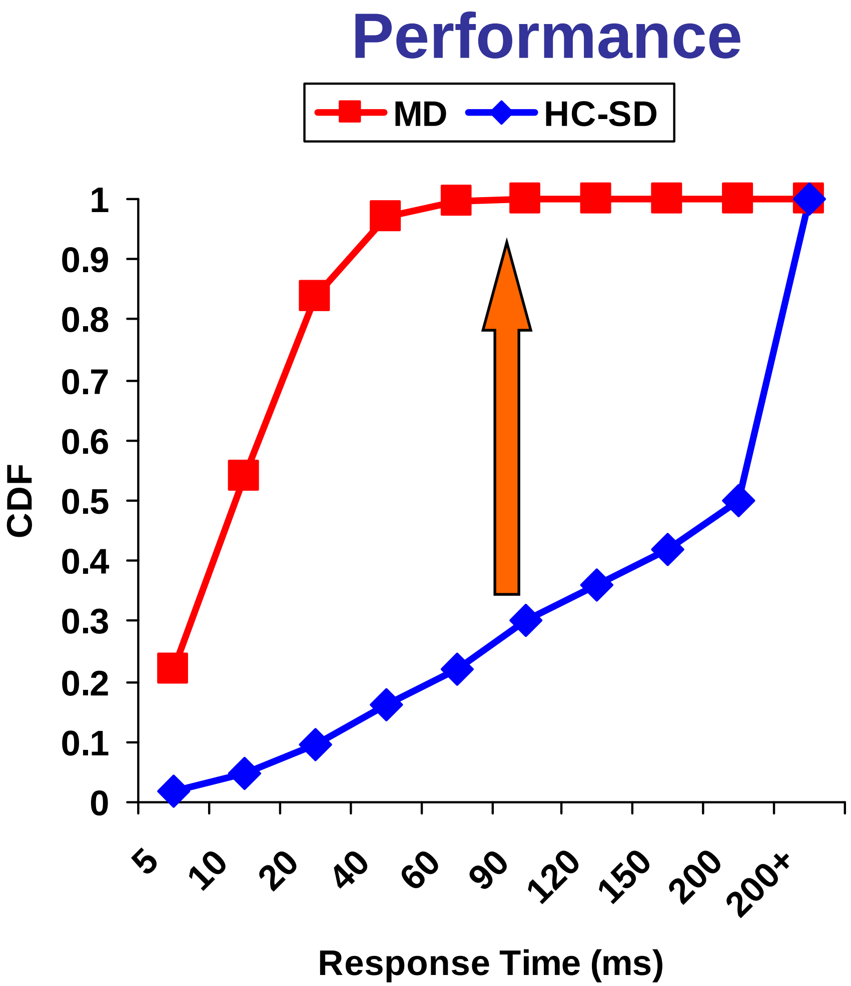
Workload	Disks	Capacity (GB)	RPM	Platters
Financial	24	19.07	10,000	4
Websearch	6	19.07	10,000	4
TPC-C	4	37.17	10,000	4
TPC-H	15	35.96	7,200	6

- Migrate data from **MD** to a **H**igh-**C**apacity **S**ingle **D**isk (**HC-SD**) drive
  - Modeled based on Seagate Barracuda ES (**750 GB**)

Impact of Data Consolidation

Websearch



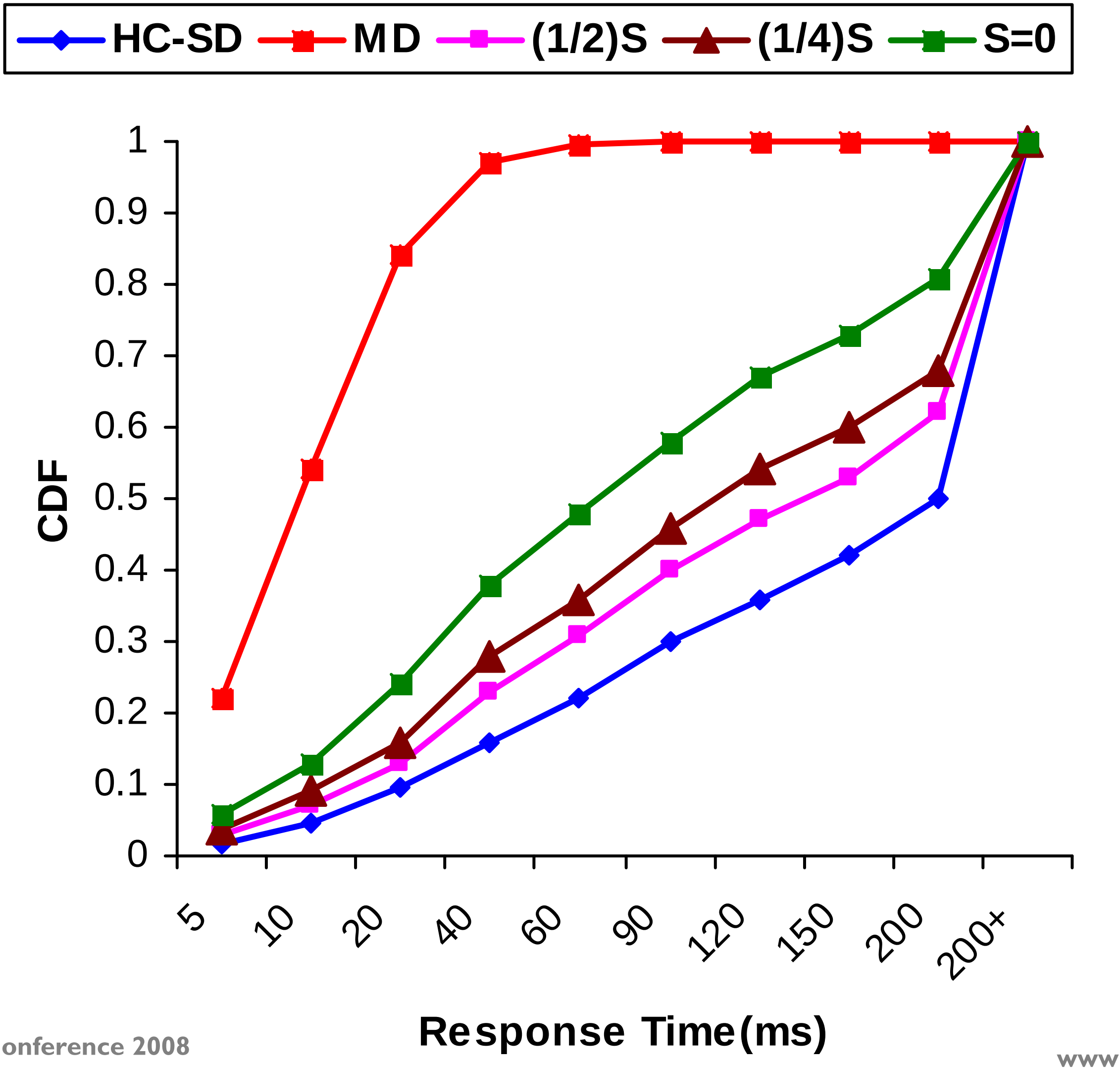


***MD* = 6 Disks**

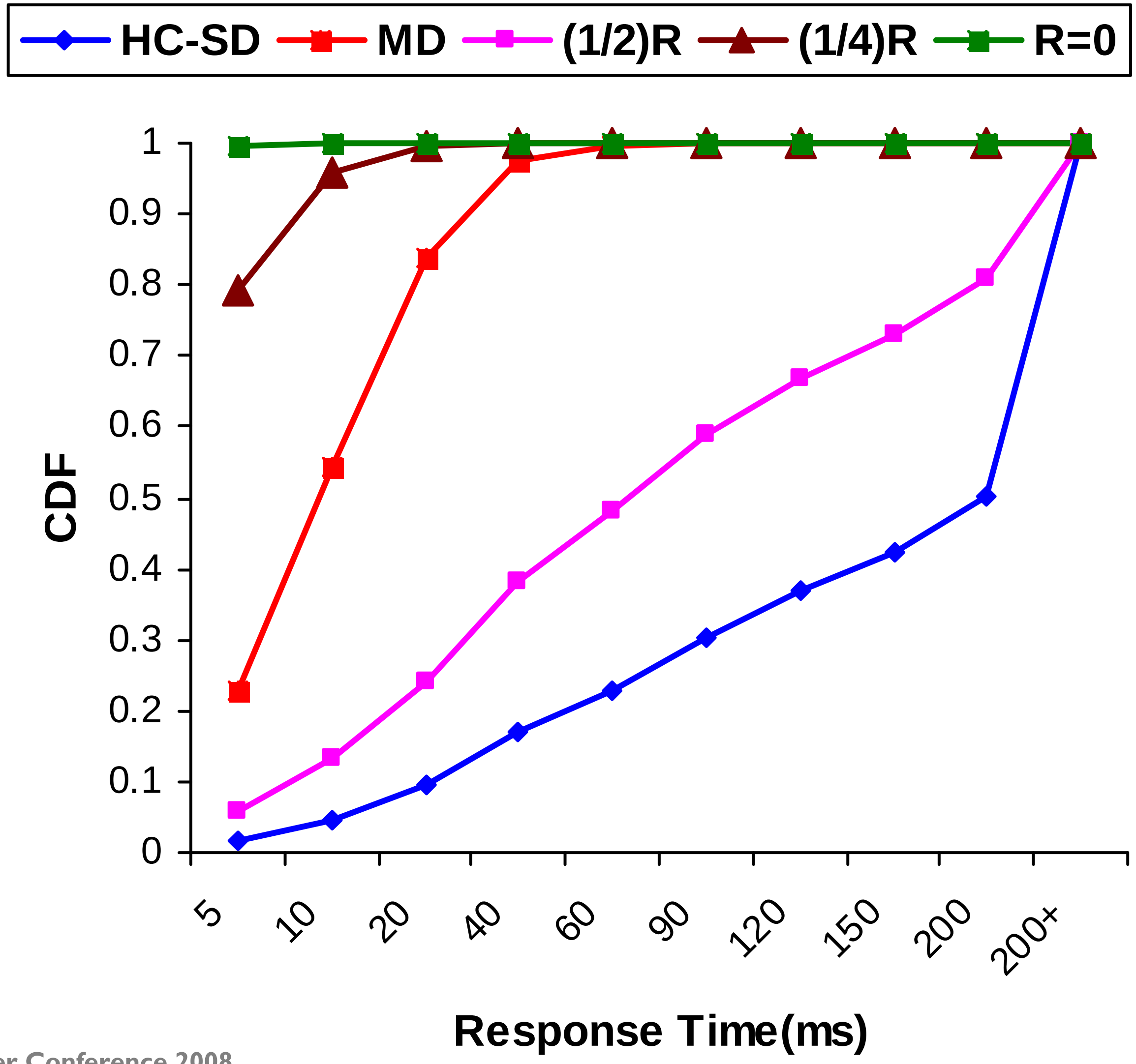
Bottleneck Analysis

- **Transfer time**
  - Much smaller compared to seek time and rotational latency
- **Cache Size**
  - Increased from 8 MB to 64 MB – Negligible impact
- **Seek Time and Rotational Latency**
  - Progressively reduced latencies for each to be:
    - ½ of original value
    - ¼ of original value
    - Latency = **0** (Eliminate performance impact)

Impact of Seek Time



Impact of Rotational Latency



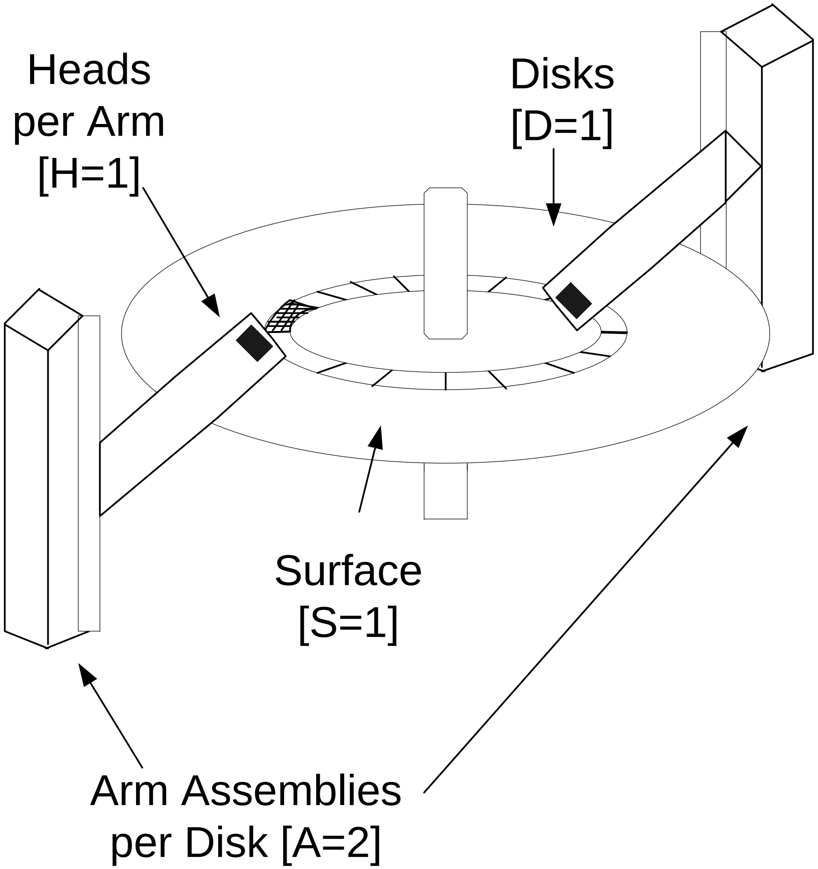
Multi-Actuator Drives

- ❐ **S**ingle-**A**rm Movement:

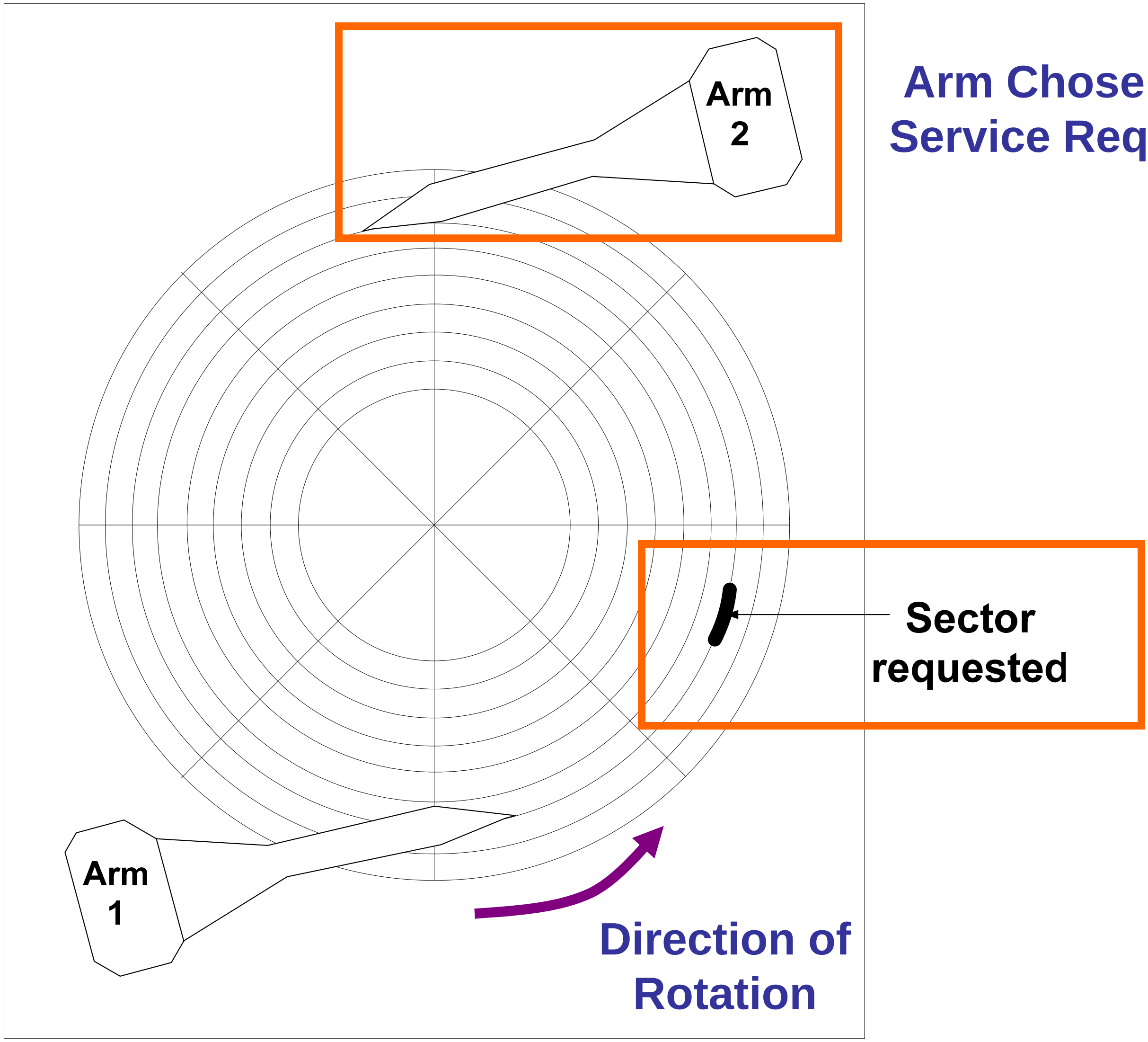
HC-SD-SA(n)

- ❐ **Peak power ~** conventional HDD
- ❐ Number of Actuators:

n = 1, 2, 3, 4



SPTF-Based Disk Arm Scheduling



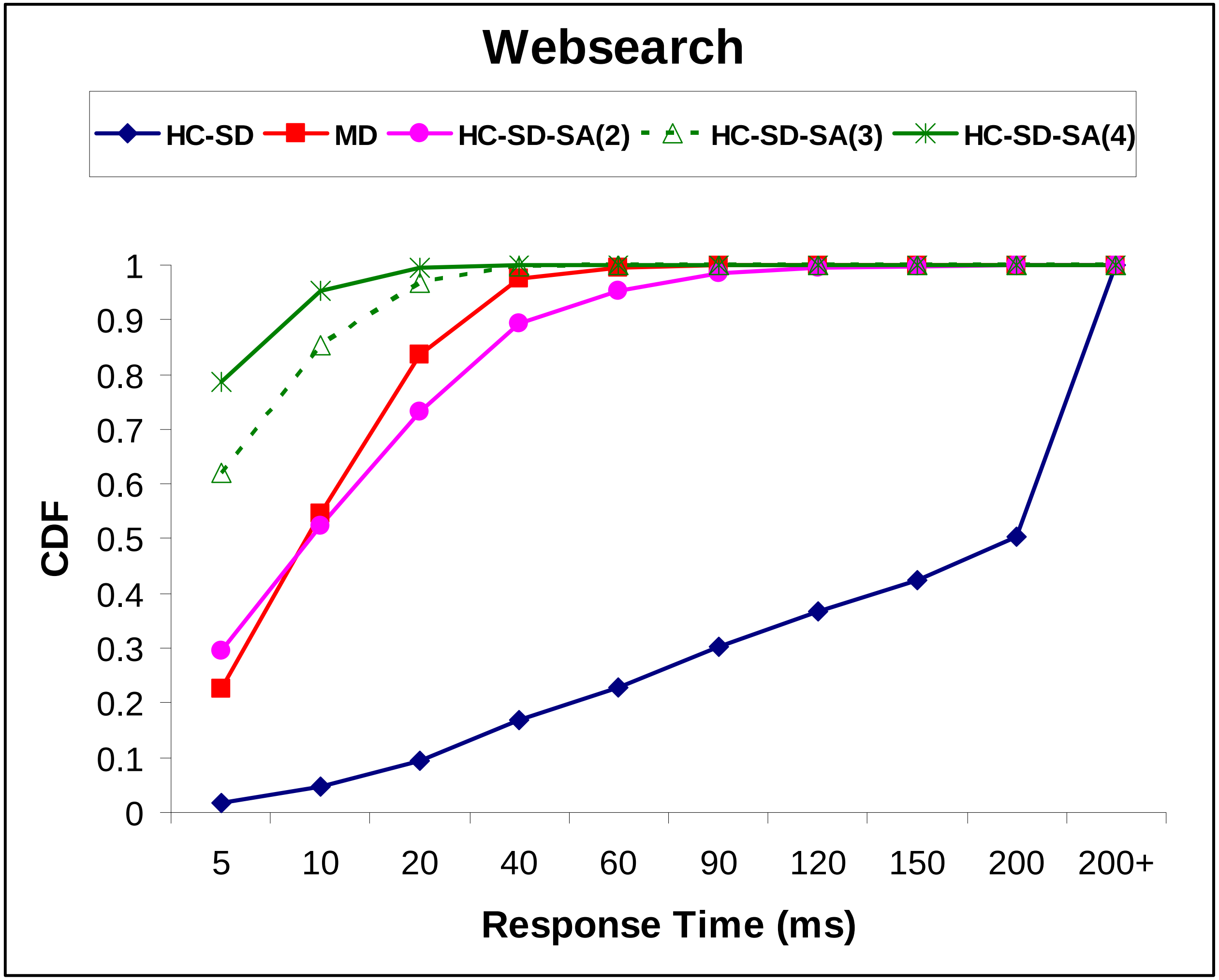
Arm Chosen to Service Request

Sector requested

Direction of Rotation

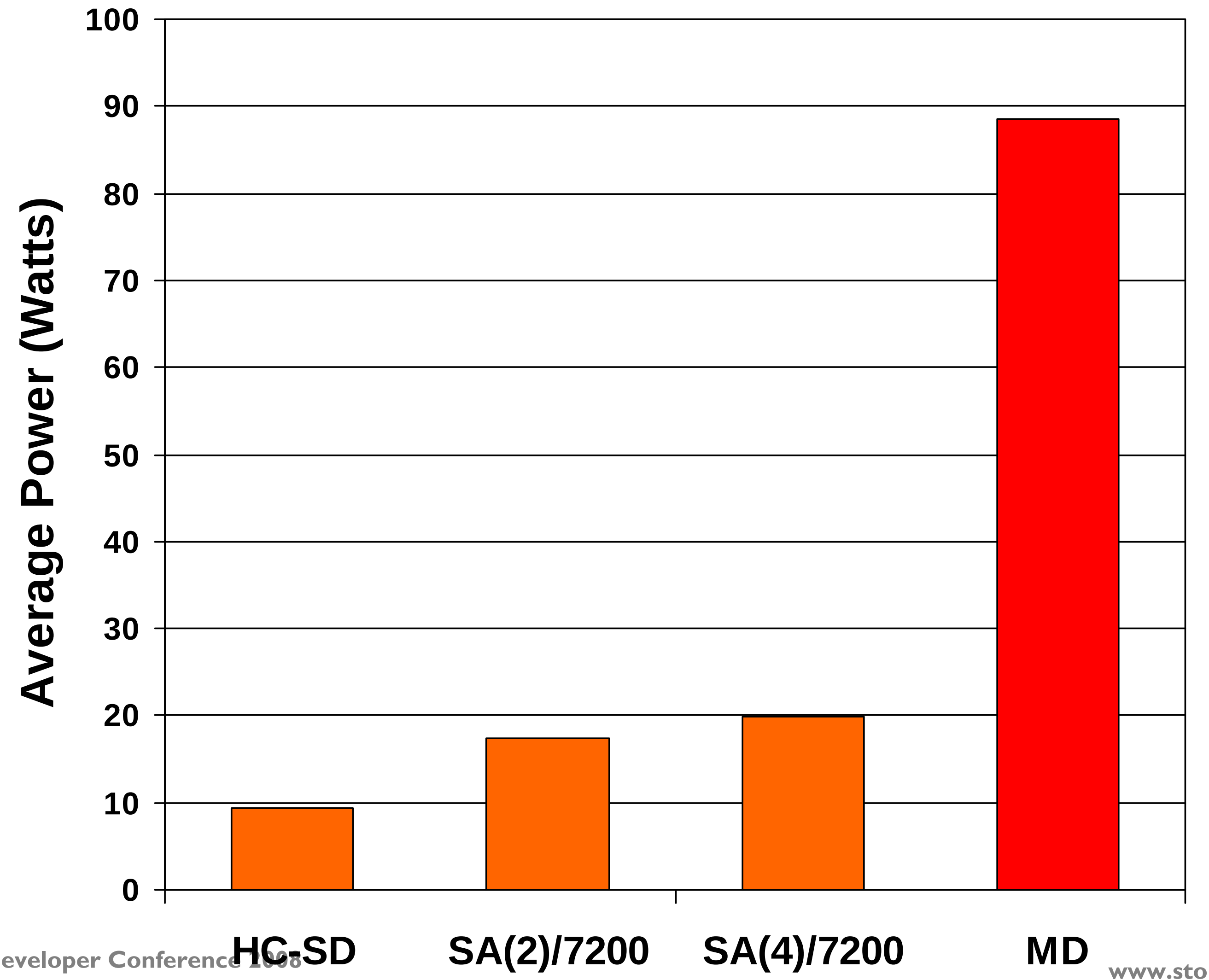
HC-SD-SA(2) Drive

HC-SD-SA(*n*) Performance



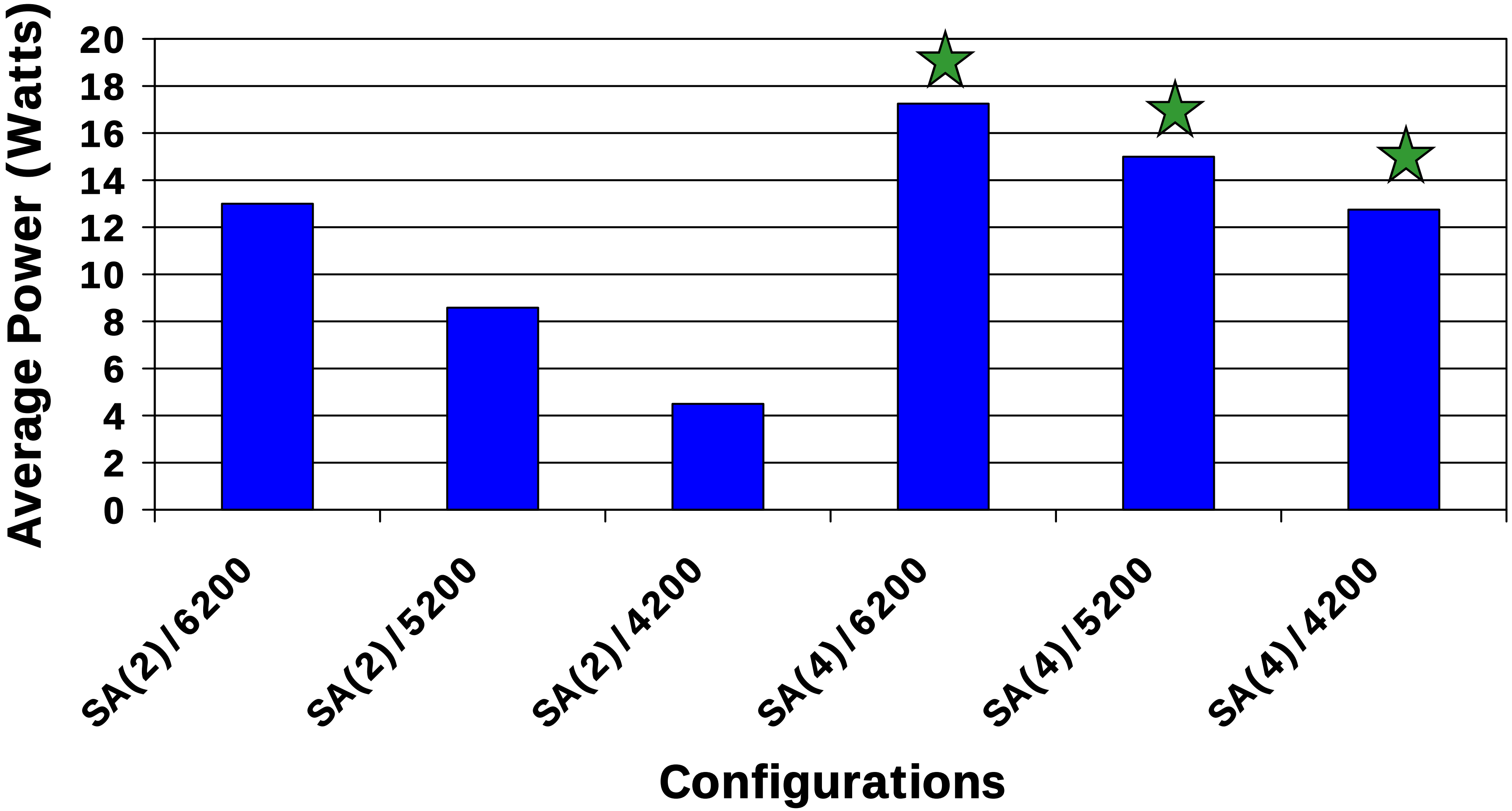
HC-SD-SA(n) Power Consumption

Websearch



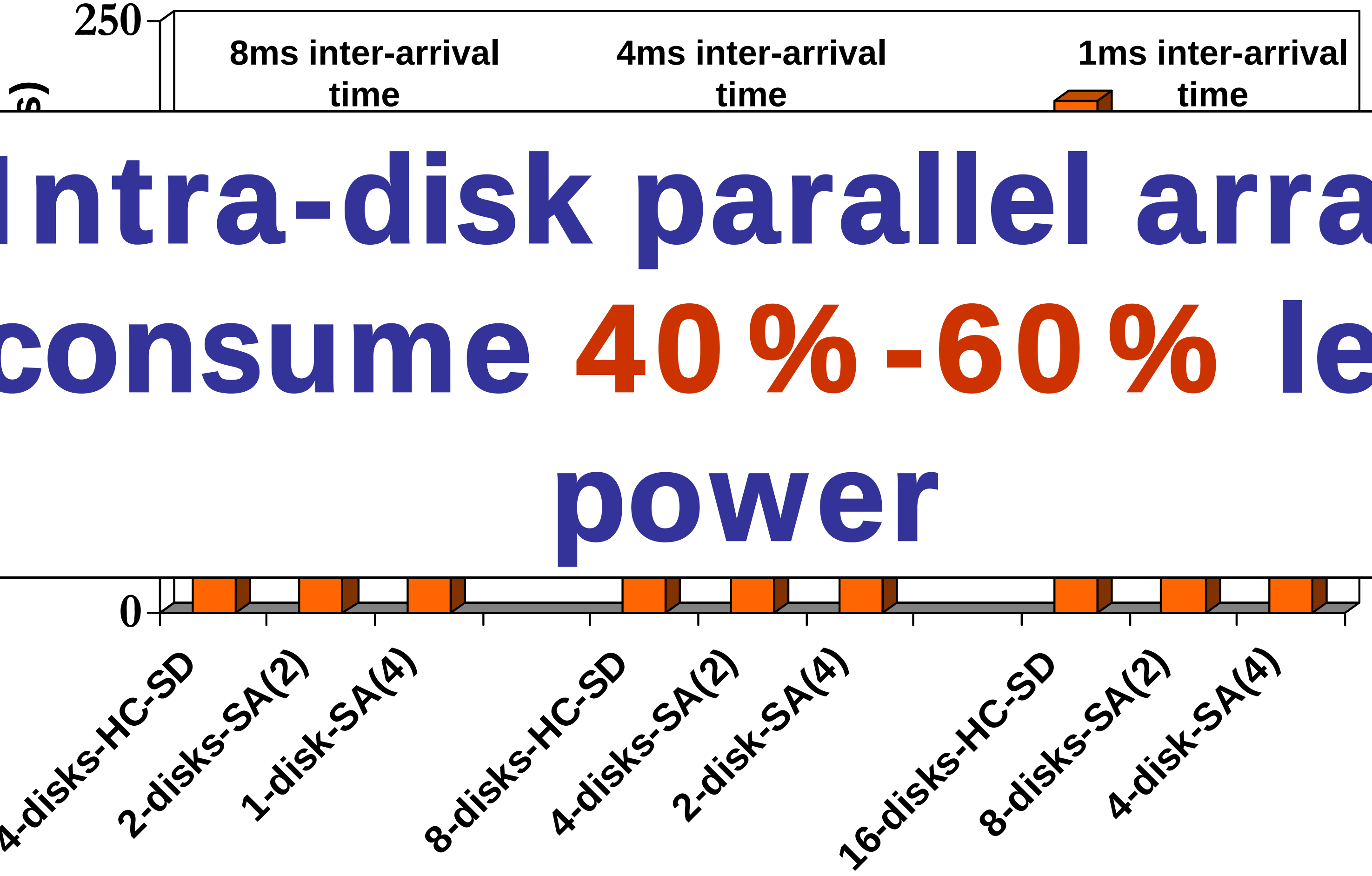
Lower the RPM to Reduce Power

Websearch



Storage Array Power Comparison

Iso-Performance Configurations



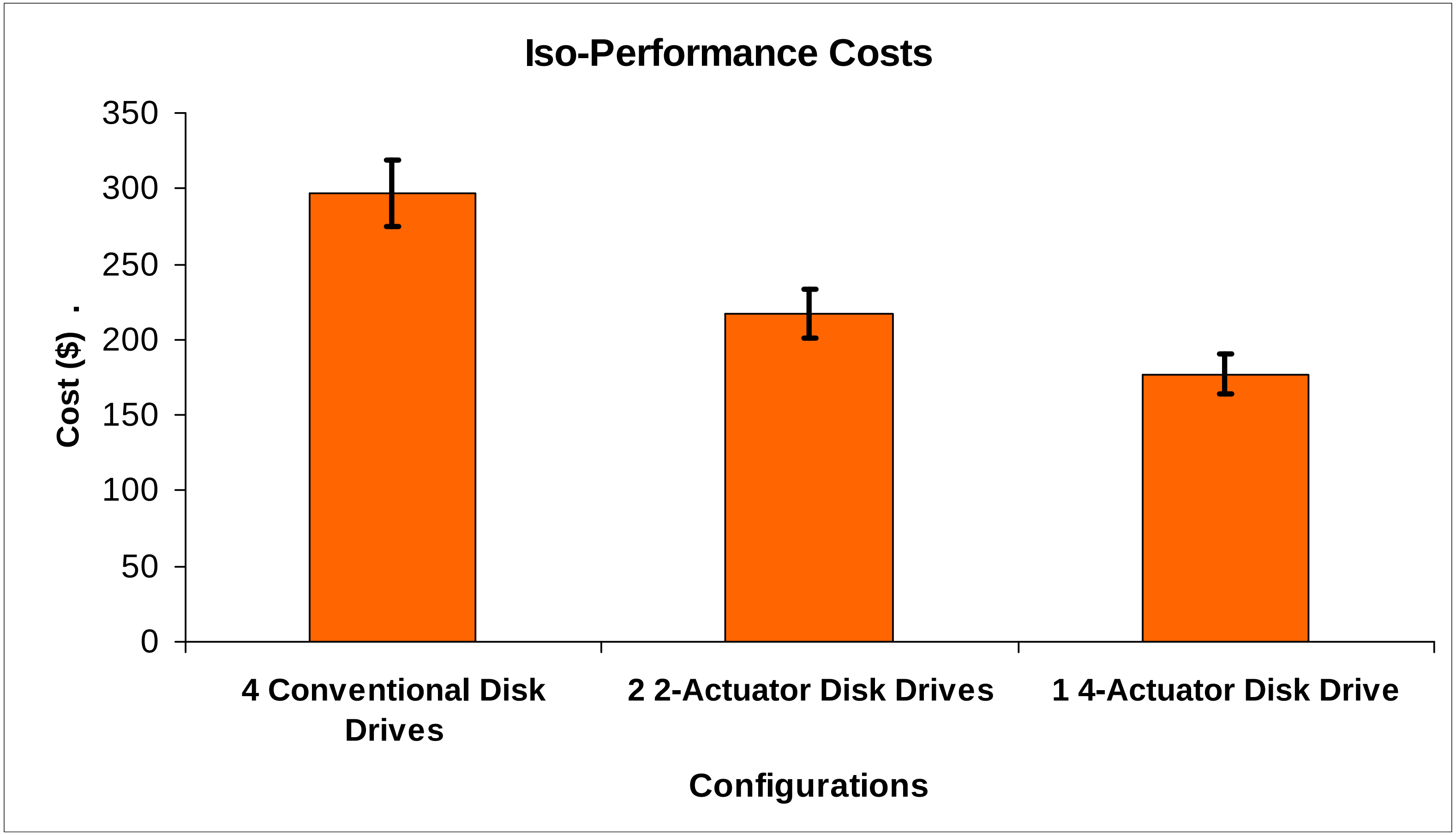
Intra-disk parallel arrays consume 40%-60% less power

Iso-performance datapoints determined via simulation using synthetic workloads

Preliminary Cost Analysis

- Material costs dominate manufacturing costs
- Identified the key HDD components
- Contacted several component manufacturers for price quotes
  - Data provided as price ranges

Cost Comparison



Engineering Issues

- Air Turbulence and Vibration
  - Use vibration-sensors and servo-based compensation techniques
- Disk Drive Reliability
  - Modify drive firmware to allow for graceful degradation

More Details and Future Work

- Paper at the 2008 International Symposium on Computer Architecture (ISCA)
- **Future Work**
  - Explore other points in the intra-disk parallelism design space (hardware, scheduling policies)
  - Build a prototype

Other Green Storage Research

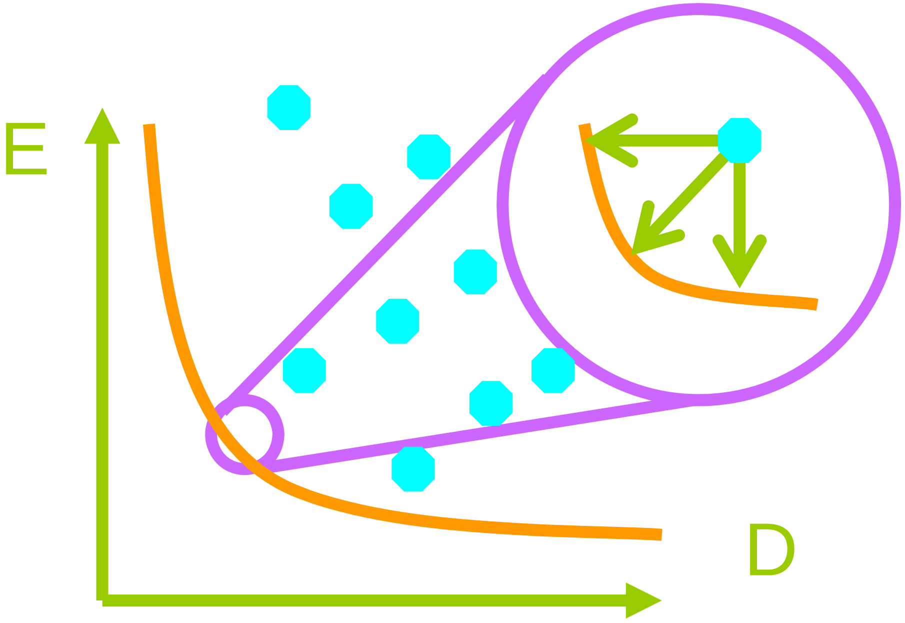
Power Management is Challenging

- **Figures of Merit:** Performance, Energy, Capacity
- **Optimization Knobs**
  - **Static knobs:** Platter size, number of platters
  - **Dynamic knobs:** Voltages of the spindle and arm motors
  - Optimal knob settings are workload-dependent

☞ **Need tools to help in the design and optimization of storage systems**

Sensitivity-based Optimization of Disk Architectures (SODA)

- ❒ **Figures of Merit:** Energy (E), Performance (D)
- ❒ **Knobs:** x, y



$$\boxed{\frac{\frac{\partial E}{\partial x}}{\frac{\partial D}{\partial x}} = \frac{\frac{\partial E}{\partial y}}{\frac{\partial D}{\partial y}}}$$

Optimality requires balancing the ratio of sensitivities with respect to each knob

Using SODA

- **Design Time:** Exploring workload-dependent tradeoffs between performance, energy, and capacity using static knobs [DAC’07]
- **Run Time:** To craft disk power management policies and analyze the effectiveness of existing policies [MASCOTS’08]

SSDs in Enterprise Storage

- SSDs provide significant performance and power benefits but Cost/GB is not yet competitive with HDDs at high capacities
  - Flash: **$3.58/GB**
  - HDD: **38¢/GB**
- **Hybrid Enterprise Storage**
  - Storage systems with mix of SSD and HDD-based devices

Hybrid Enterprise Storage Systems Design and Management

- ❐ **Device Design:** What SSD design would maximize performance for a given cost constraint?
- ❐ **System Deployment:** What mix of SSDs and HDDs would give me the best energy savings for given cost, capacity, and performance constraints?
- ❐ **System Management:** Is my power management policy providing the best energy savings for a given performance target?

Conclusions

- Storage power is a growing problem in data centers
- Intra-disk parallelism [ISCA'08]
  - 40%-60% reduction in power consumption
  - Preliminary analysis suggests that such drives are viable
- Sensitivity-based optimization [DAC'07, MASCOTS'08]
  - Allows us to systematically design and optimize storage systems
- PDF of papers available at: **http://www.cs.virginia.edu/~gurumurthi**

Thank You

E-mail: gurumurthi@cs.virginia.edu

University of Virginia
Department of Computer Science