

The Storage Stack of Future Data Centers

Sudhanva Gurumurthi
gurumurthi@cs.virginia.edu

Many Applications are Data Intensive

Traditional Application Domains

amazon.com.

Expedia.com®

eBay®

GMail™
by Google BETA

Emerging Application Domains

You Tube™
Broadcast Yourself™

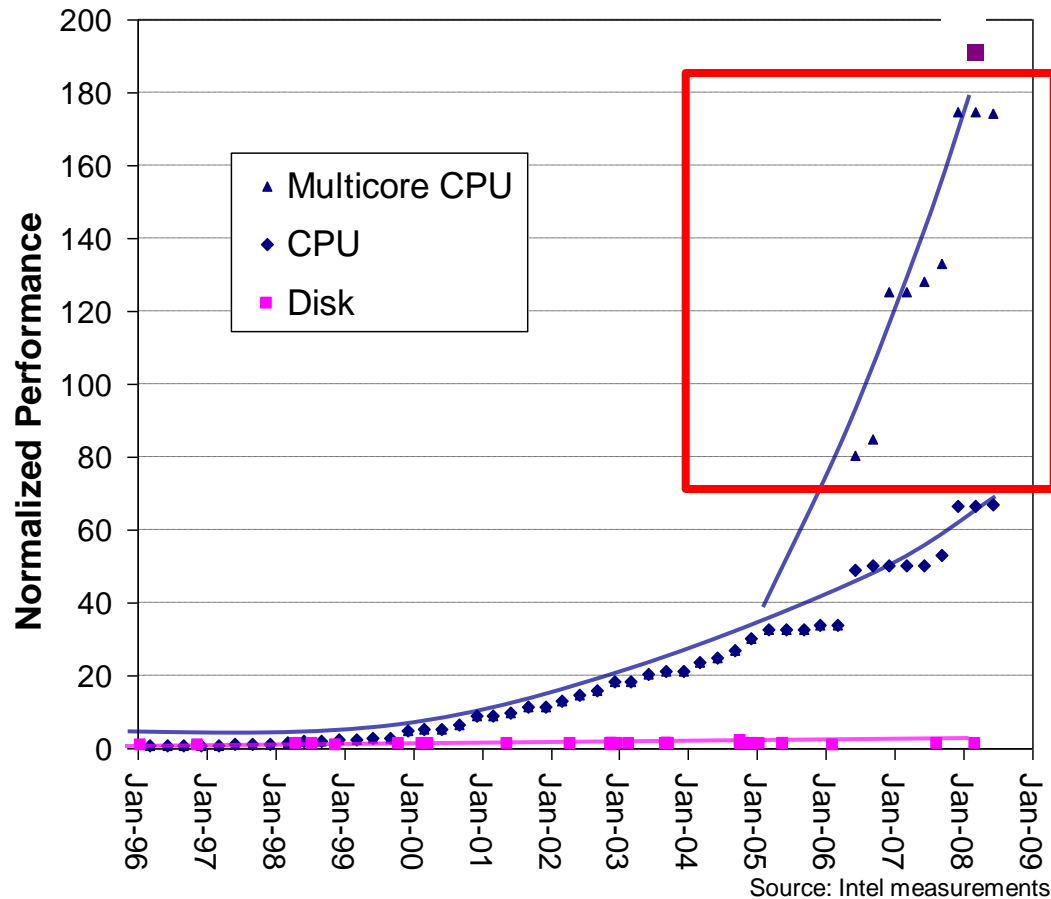
flickr™
LOVES YOU™

facebook

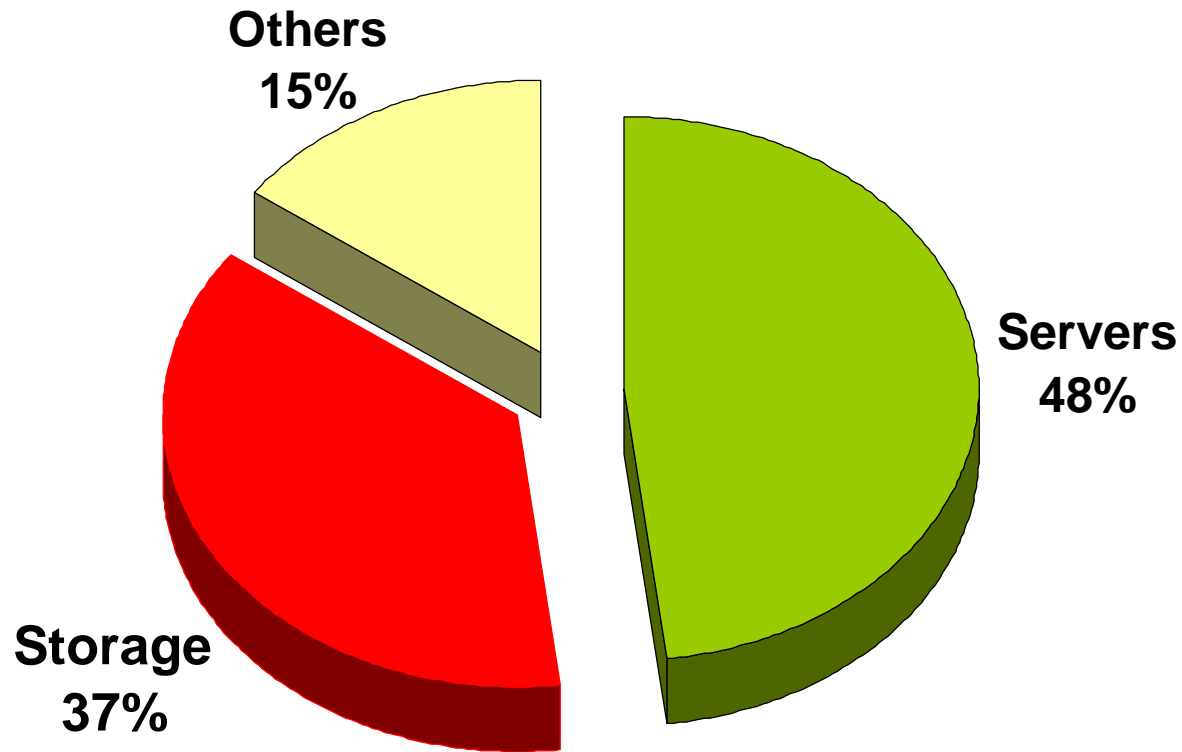
Google™
Maps

Normalized CPU Performance

Normalized Media Access Time for 20K Read



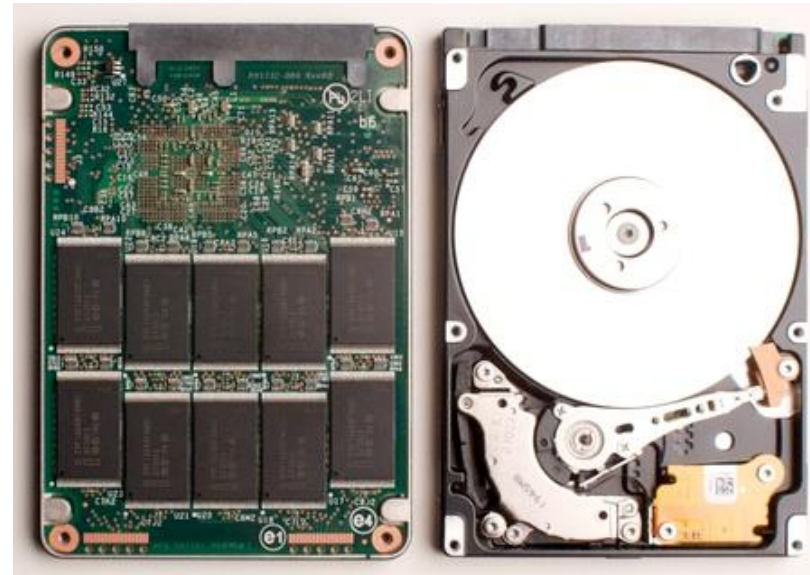
Data Center Computing Equipment Power Consumption



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

Flash-Based Solid-State Disks (SSDs)

- NAND Flash is a commoditized memory technology
- High random I/O performance
- Low power consumption
- Flash reliability can be managed, even for server workloads



SSD

HDD

NAND Flash Scaling is Uncertain

2009Tables_FOCUS_C_ITRS [Read-Only] [Compatibility Mode] - Microsoft Excel

	B	C	D	E	F	G	H	I	J
2	<i>Table PIDS5 Non-Volatile Memory Technology Requirements</i>								
3	<i>Year of Production</i>	2009	2010	2011	2012	2013	2014	2015	2016
4	<i>DRAM ½ Pitch (nm) (contacted)</i>	50	45	40	35	32	28	25	22
5	<i>MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)</i>	54	45	38	32	27	24	21	19
6	<i>(ORTC) NAND Flash poly 1/2 Pitch (nm)</i>	38	32	28	25	23	20	18	16
7	<i>(PIDS) NAND Flash poly 1/2 Pitch (nm)</i>	34	32	28	25	22	20	19	18
8									
9	<i>NAND Flash</i>								
10	NAND Flash technology node – F (nm) [1]	34	32	28	25	22	20	19	18
11	Number of word lines in one NAND string [2]	64	64	64	64	64	64	64	64
12	Cell type (FG, CT, 3D, etc.) [3]	FG	FG	FG	FG/CT	FG/CT	CT–3D	CT–3D	CT–3D
13	3D NAND number of memory layers	1	1	1	1	1	2	4	4
14									
15	<i>A. Floating Gate NAND Flash</i>								
16	Cell size – area factor a in multiples of F ² SLC/MLC [4]	4.0/1.3	4.0/1.3	4.0/1.3	4.0/1.0	4.0/1.0	4.0/1.0	4.0/1.0	4.0/1.0
17	Tunnel oxide thickness (nm) [5]	6–7	6–7	6–7	6–7	6–7	6–7	6–7	4
18	Interpoly dielectric material [6]	ONO	ONO	ONO	High-K	High-K	High-K	High-K	High-K
19	Interpoly dielectric thickness (nm)	10–13	10–13	10–13	9–10	9–10	9–10	9–10	9–10
20	Gate coupling ratio (GCR) [7]	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7
21	Control gate material [8]	n-Poly	n-Poly	n-Poly	Poly/metal	Poly/metal	Poly/metal	Metal	Metal

HDDs Will Still Be Around

- Magnetic technology has scaled even when confronted by potential show-stoppers (e.g., superparamagnetic limit)
- There is still a cost gap between SSDs and HDDs
- ☞ Use HDDs for capacity and SSDs for performance

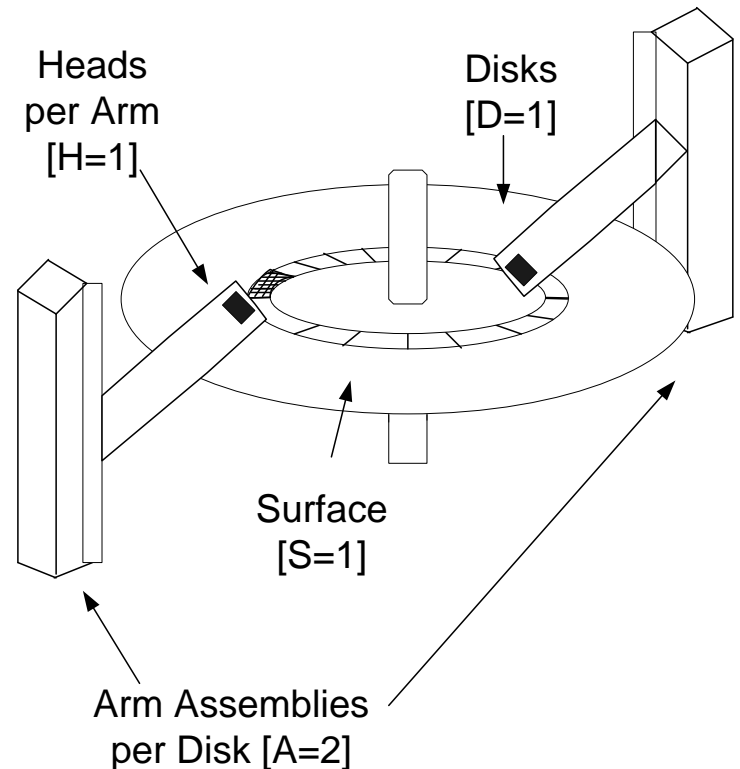
Heterogeneous Storage Systems



HDDs Need To Be Energy Efficient

Energy Efficient HDDs

- Multi-RPM/DRPM Disks
[Gurumurthi et al., ISCA'03]
- Intradisk Parallelism
[Sankar et al., ISCA'08]
- Need to develop techniques for optimally configuring and efficiently managing data locality in such heterogeneous storage systems



Multi-Actuator Parallel HDD