Cloud Computing Is Driving Infrastructure Innovation

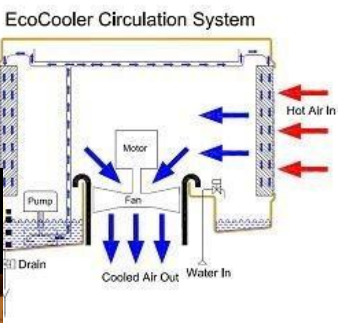
Western Digital

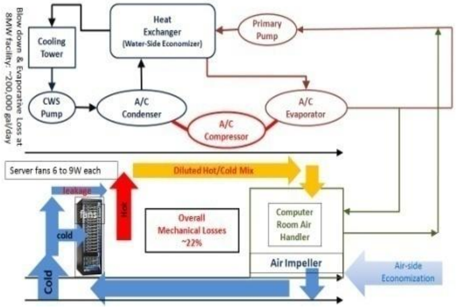
Board of Directors Meeting

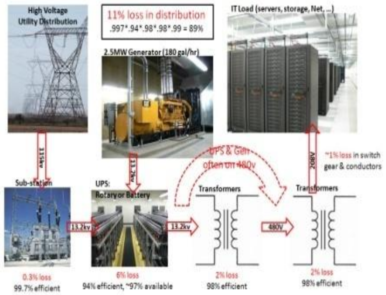
James Hamilton, 2011/5/17
VP & Distinguished Engineer, Amazon Web Services
email: James@amazon.com
web: mvdirona.com/jrh/work
blog: perspectives.mvdirona.com

Agenda

- Quickening Pace Infrastructure Innovation
  - Influence of Cloud computing
- Technology Backdrop
  - Infrastructure innovation
  - Memory Wall & Storage Chasm
- Impact on Storage Market
  - Client Storage Migration to SSD & Cloud
  - Enterprise Migration to Cloud
  - Accelerating computation & storage growth
  - Disk is tape







Pace of Innovation

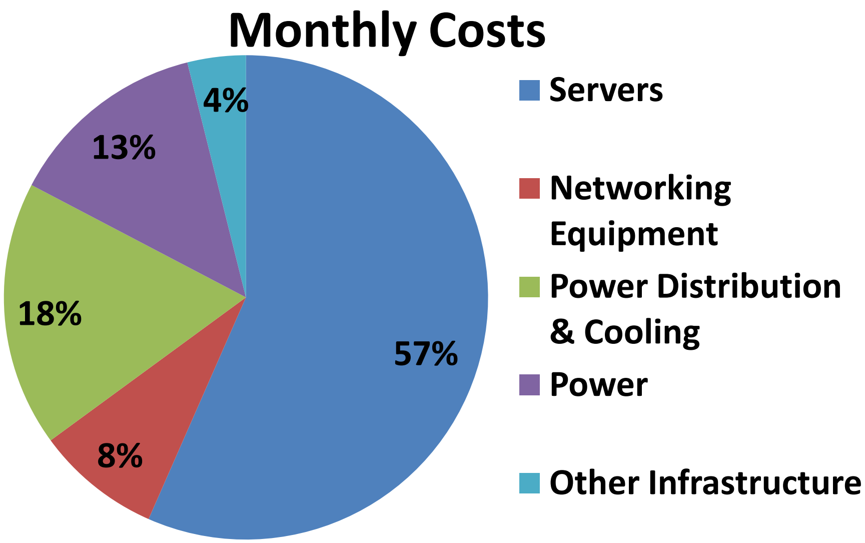
- Datacenter pace of innovation increasing
  - More innovation in last 5 years than previous 15
  - Driven by cloud service providers and very high-scale internet applications like search
  - Cost of infrastructure dominates service cost
  - Not just a cost center
- High focus on infrastructure innovation
  - Driving down cost
  - Increasing aggregate reliability
  - Reducing resource consumption footprint

Perspective on Scaling

Each day Amazon Web Services adds enough new capacity to support all of Amazon.com's global infrastructure through the company's first 5 years, when it was a $2.76B annual revenue enterprise

Where Does the Money Go?

- **Assumptions:**
  - Facility: ~$88M for 8MW critical power
  - Servers: 46,000 @ $1.45k each
  - Commercial Power: ~$0.07/kWhr
  - Power Usage Effectiveness: 1.45



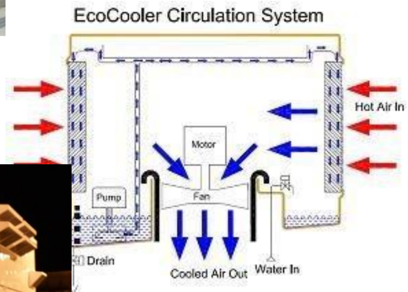
3yr server & 10 yr infrastructure amortization

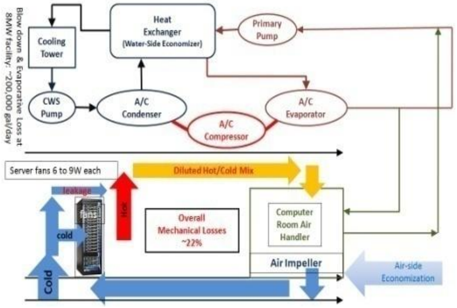
- **Observations:**
  - 31% costs functionally related to power (trending up while server costs down)
  - Networking high at 8% of overall costs & 19% of total server cost (many pay more)

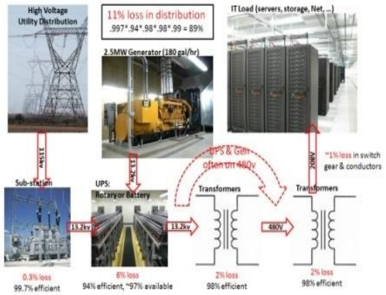
From: http://perspectives.mvdirona.com/2010/09/18/OverallDataCenterCosts.aspx

Agenda

- Quickening Pace Infrastructure Innovation
  - Influence of Cloud computing
- Technology Backdrop
  - Infrastructure innovation
  - Memory Wall & Storage Chasm
- Impact on Storage Market
  - Client Storage Migration to SSD & Cloud
  - Enterprise Migration to Cloud
  - Accelerating computation & storage growth
  - Disk is tape





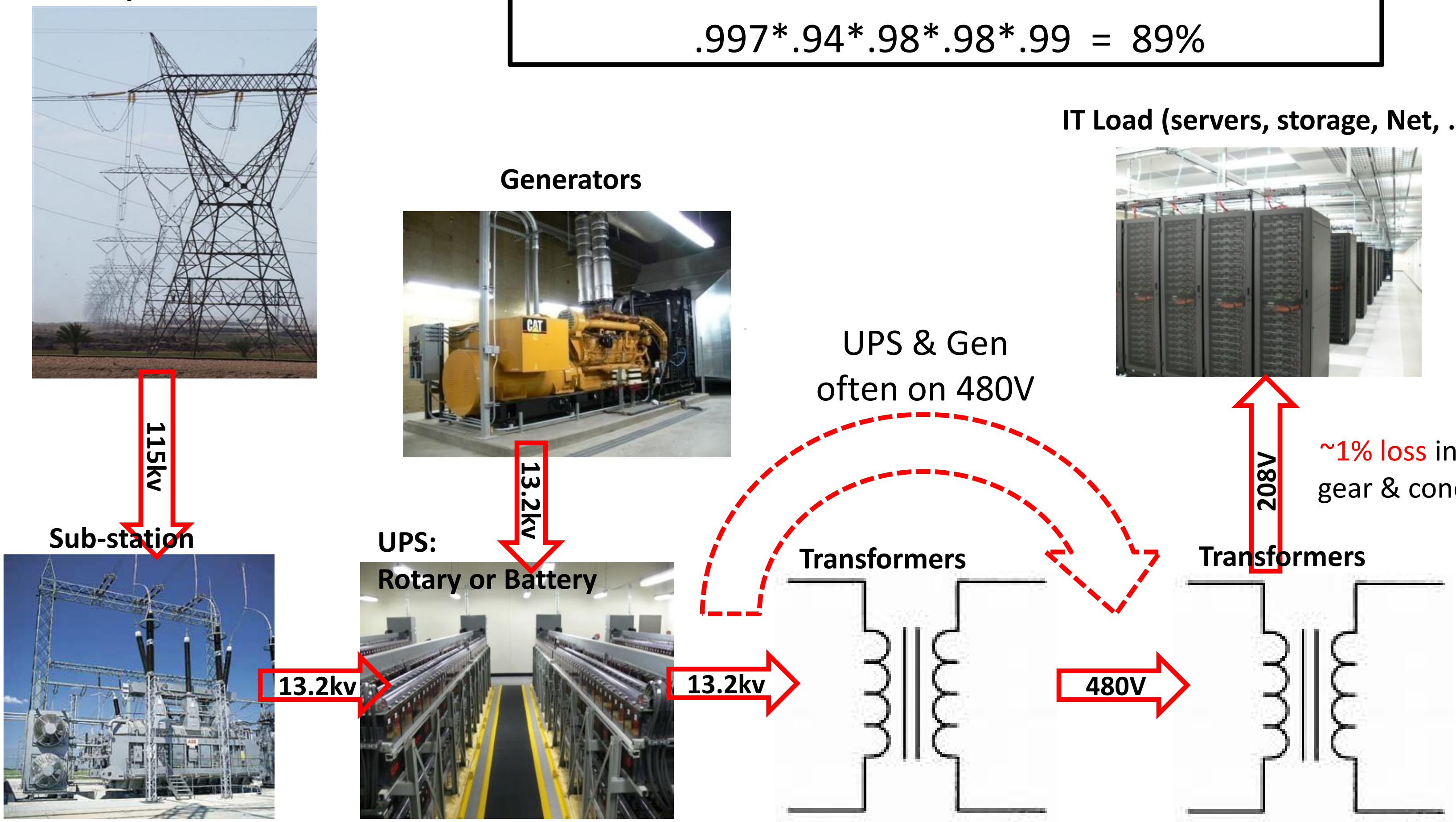


Power Distribution

~11% lost in distribution

.997*.94*.98*.98*.99 = 89%

High Voltage
Utility Distribution



0.3% loss
99.7% efficient

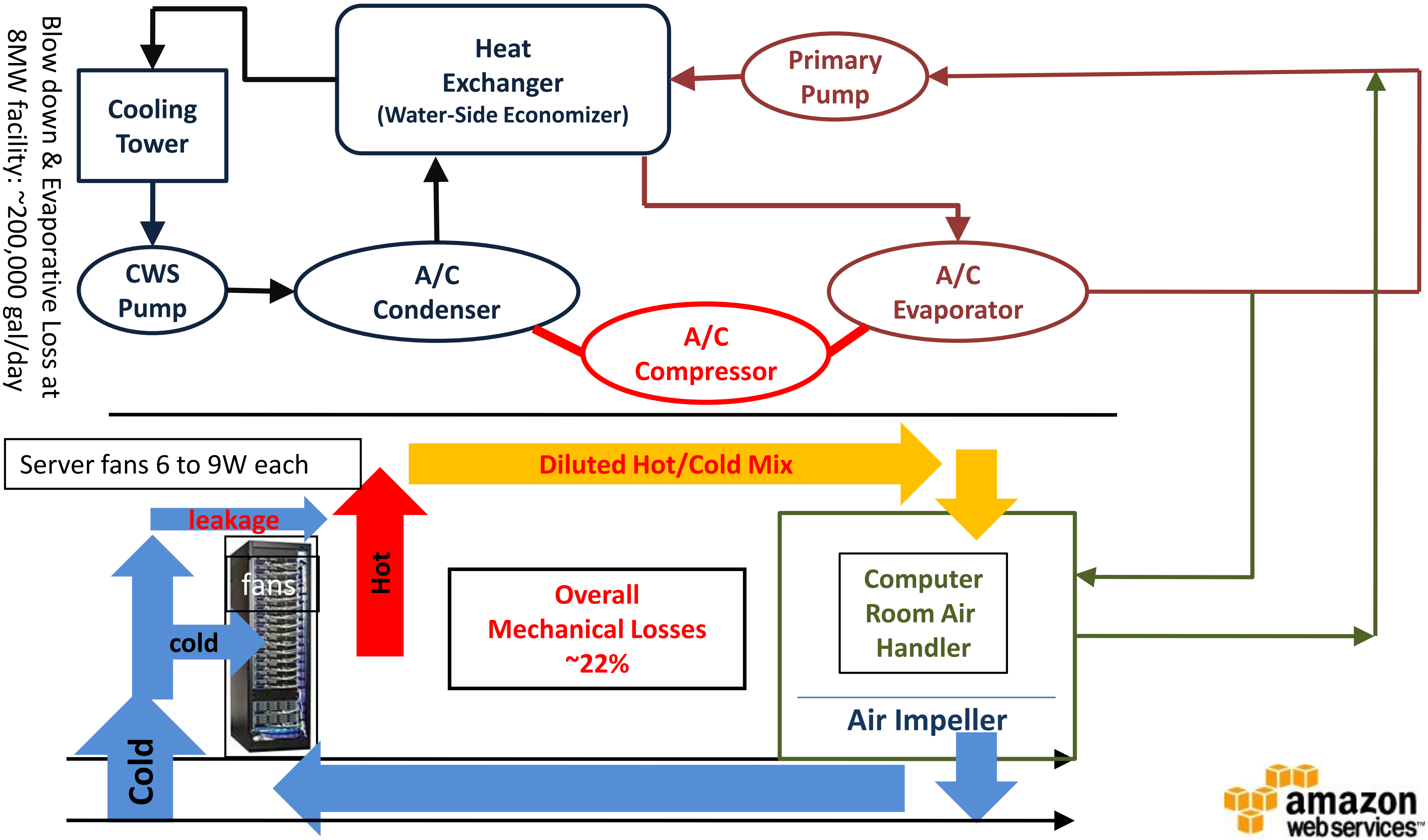
6% loss
94% efficient, ~97% available

2% loss
98% efficient

2% loss
98% efficient

Note: Two more levels of power conversion at server

Mechanical Systems



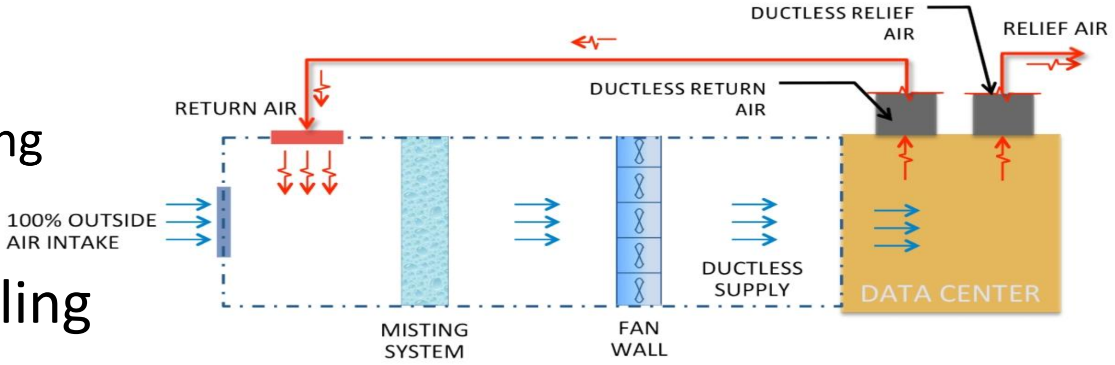
Blow down & Evaporative Loss at 8MW facility: ~200,000 gal/day

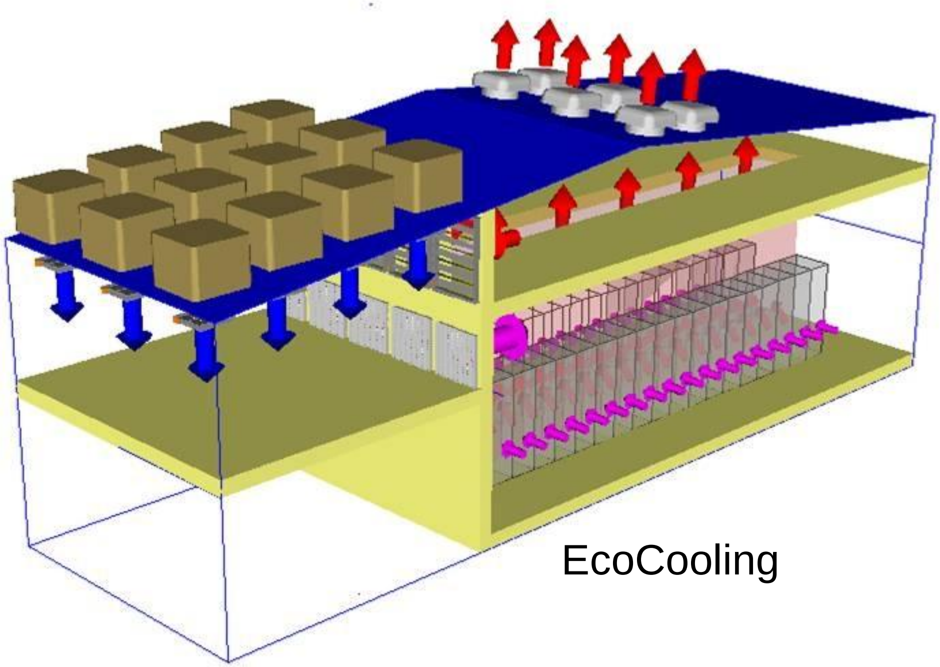
Server fans 6 to 9W each

Overall Mechanical Losses ~22%

Innovative Building Designs

- Evaporative cooling only
  - Right: High pressure misting
  - Below: Wet media cooler
- Ductless full building cooling





EcoCooling

Facebook Prineville above & below

Modular and Pre-fab DC Designs

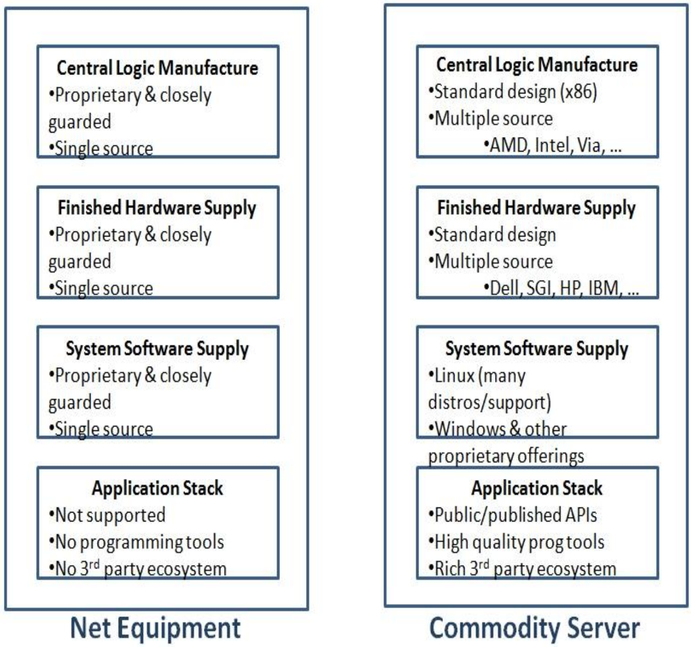
Microsoft ITPAC

Amazon Perdix

- Fast & economic deployments
- Sub-1.15 PUE designs
- Air-side economized
  - No mechanical cooling
- ISO standard shipping containers offered by Dell, HP, SGI, IBM, …

Sea Change in Networking

- Current networks over-subscribed
  - Forces workload placement restrictions
  - Goal: all points in datacenter equidistant
- Mainframe model goes commodity
  - Competition at each layer over vertical integ.
- Get onto networking on Moores Law path
  - ASIC port count growth at near constant cost
  - Competition: Broadcom, Marvell, Fulcrum,…



Central Logic Manufacture
- Proprietary & closely guarded
- Single source

Finished Hardware Supply
- Proprietary & closely guarded
- Single source

System Software Supply
- Proprietary & closely guarded
- Single source

Application Stack
- Not supported
- No programming tools
- No 3rd party ecosystem

Net Equipment

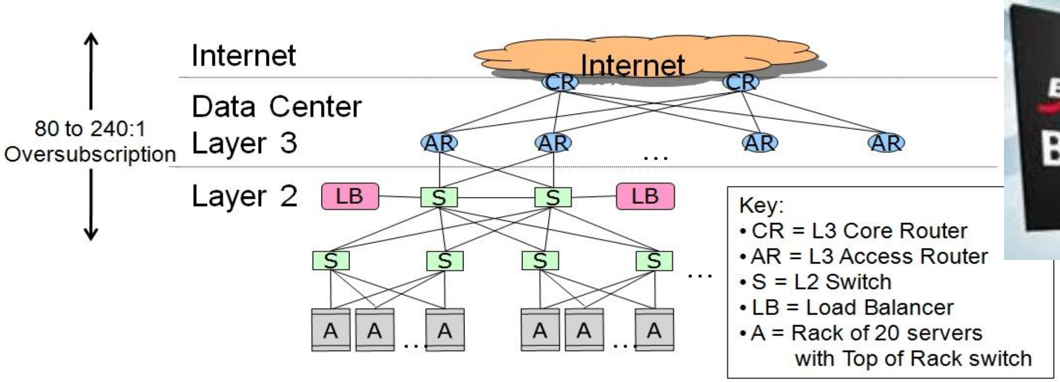
Central Logic Manufacture
- Standard design (x86)
- Multiple source
  - AMD, Intel, Via, …

Finished Hardware Supply
- Standard design
- Multiple source
  - Dell, SGI, HP, IBM, …

System Software Supply
- Linux (many distros/support)
- Windows & other proprietary offerings

Application Stack
- Public/published APIs
- High quality prog tools
- Rich 3rd party ecosystem

Commodity Server



Key:
- CR = L3 Core Router
- AR = L3 Access Router
- S = L2 Switch
- LB = Load Balancer
- A = Rack of 20 servers with Top of Rack switch

Storage & Memory B/W lagging CPU

	CPU	DRAM	LAN	Disk
Annual bandwidth improvement (all milestones)	1.5	1.27	1.39	1.28
Annual latency Improvement (all milestones)	1.17	1.07	1.12	1.11

Memory wall

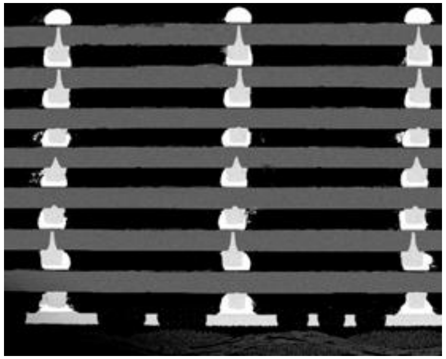
Storage Chasm

- CPU out-pacing source data rates
- Disk & memory getting "further" away from CPU
  - Core limiting factor: power consumption & data availability
  - Powered CPU cores have no value without data
- Large sequential transfers better for both memory & disk
- Lets look first at efficient memory solutions

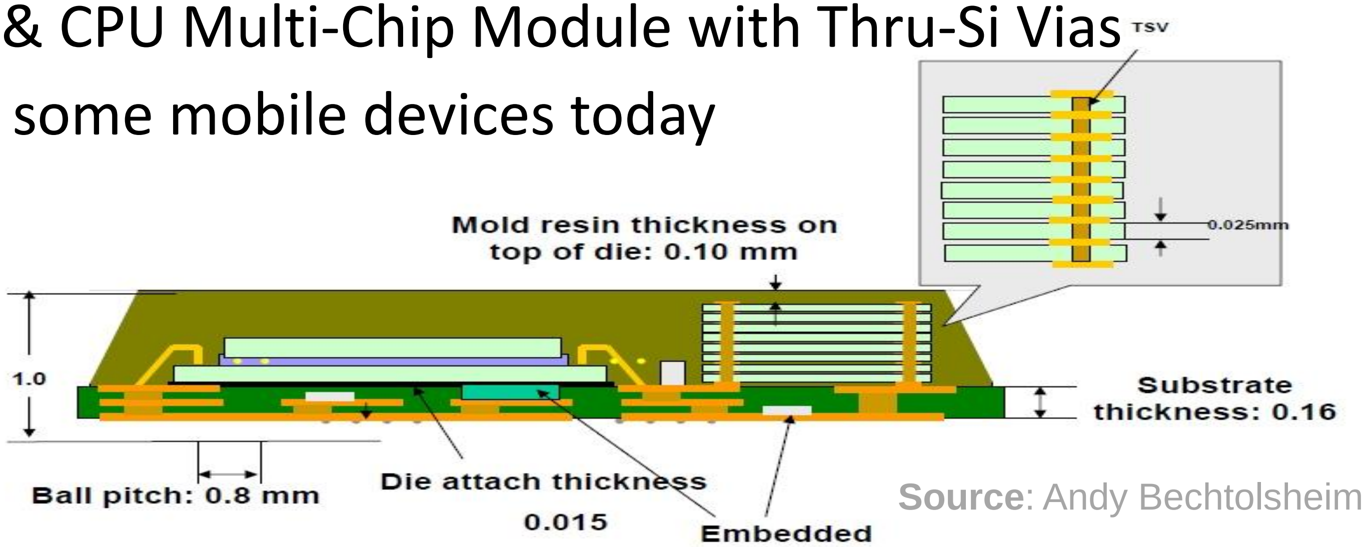
Source: Dave Patterson: Why Latency Lags Bandwidth and What It Means to Computing

Memory Wall

- Adding processor I/O pins has a positive impact but at significant power cost
  - Positive but bounded impact
- Most probable memory wall solution:
  - Mem & CPU Multi-Chip Module with Thru-Si Vias
  - Lab & some mobile devices today



Multi-Chip Module

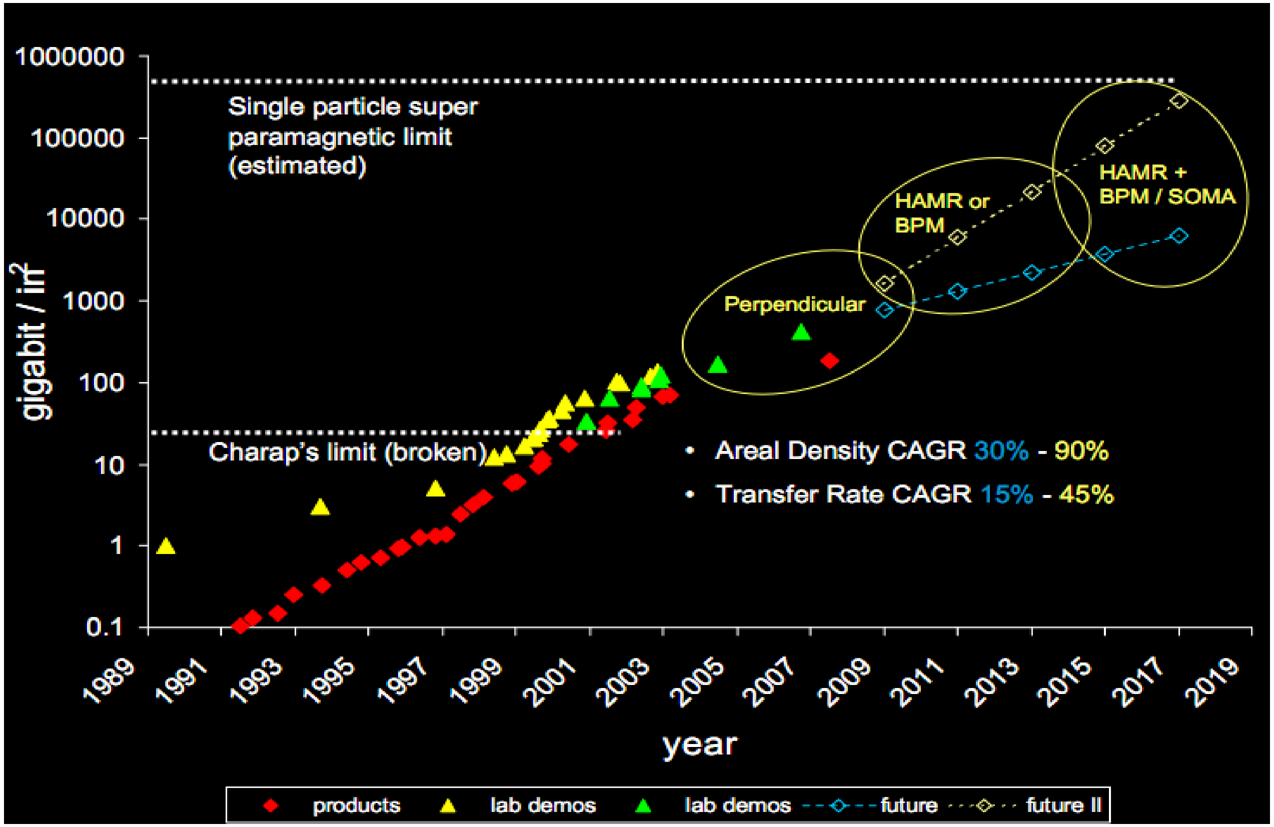

TSV
0.025mm
Mold resin thickness on top of die: 0.10 mm
1.0
Substrate thickness: 0.16
Ball pitch: 0.8 mm
Die attach thickness 0.015
Embedded

Source: Andy Bechtolsheim

- But what about HDD & storage chasm

HDD: Capacity

- Capacity growth continues unabated



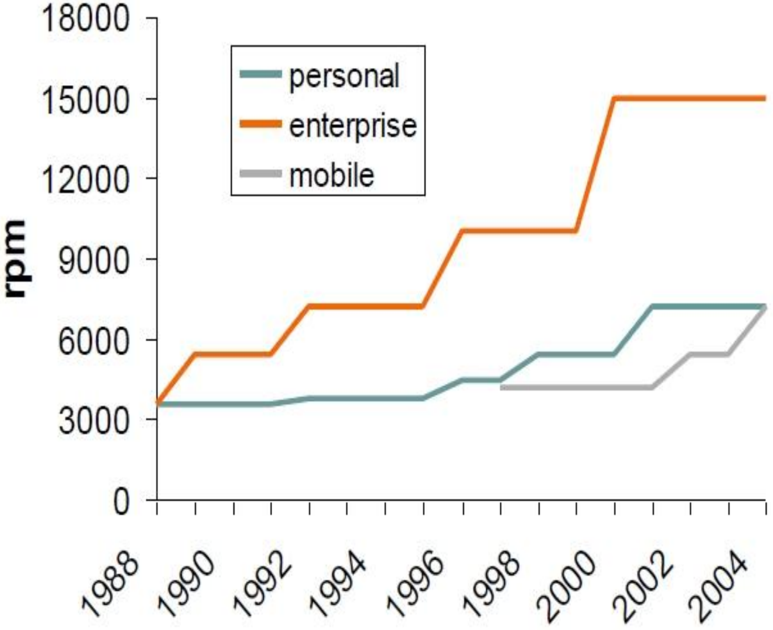
- Areal Density CAGR 30% - 90%
- Transfer Rate CAGR 15% - 45%

Source: Dave Anderson

- Capacity isn't the problem
  – What about bandwidth and IOPS?

HDD: Rotational Speed

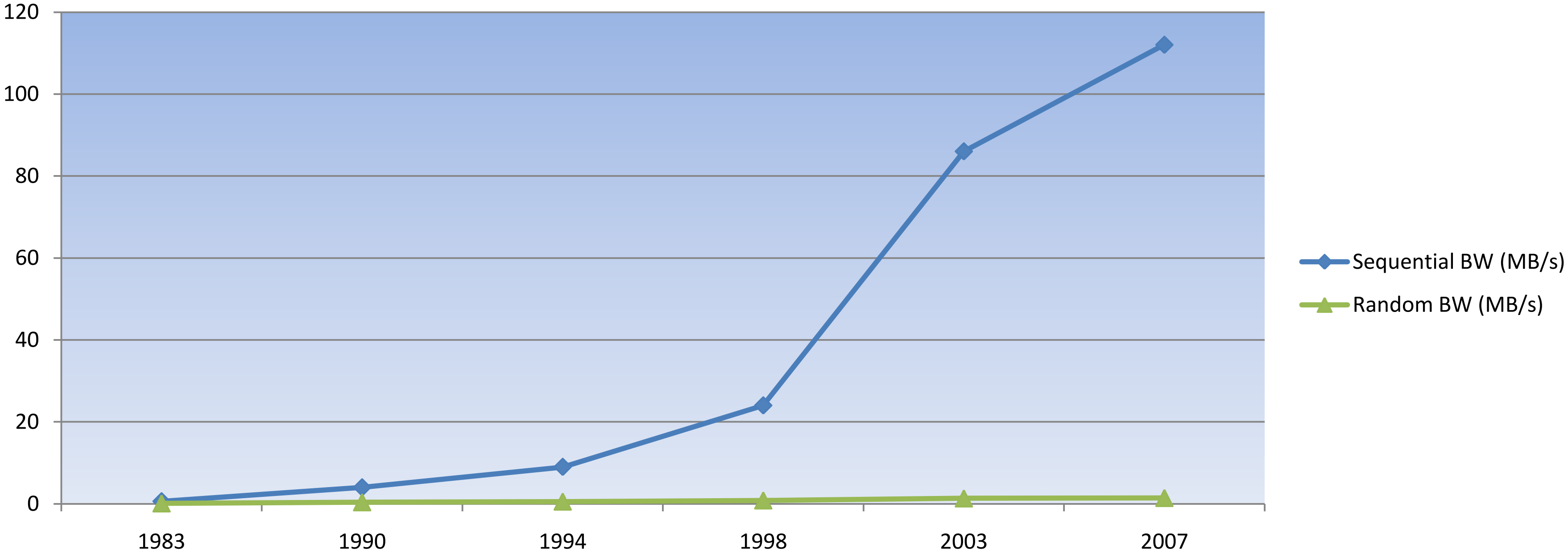
- RPM contributes negatively to:
  - rotational vibration
  - Non-Repeating Run Out (NRRO)
- Power cubically related to RPM
- >15k RPM not economically viable
  - no improvement in sight
- RPM not improving & seek times only improving very slowly
- IOPS improvements looking forward remain slow
- Even sequential BW growth insufficient



product information for Seagate and Control Data disc drives since 1988, mobile includes Toshiba drives since 1997

Source: Dave Anderson

HDD Random BW vs Sequential BW



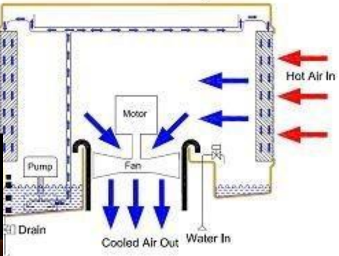
- Disk sequential BW growth slow
- Disk random access BW growth roughly 10% of sequential
- Storage Chasm widening
  - BW a long term problem &IOPS growth very slow

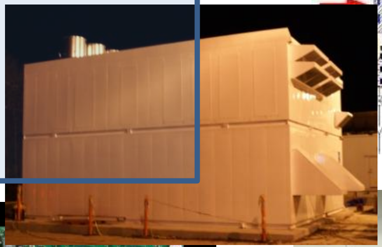
Source: Dave Patterson with James Hamilton updates

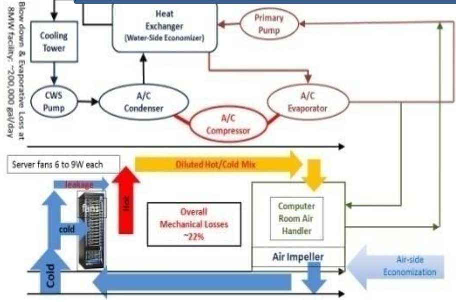
Agenda

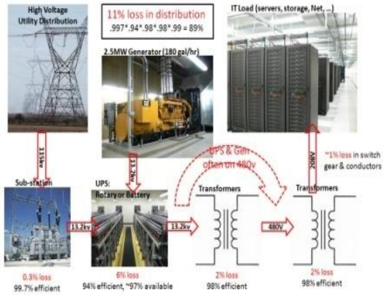
- Quickening Pace Infrastructure Innovation
  - Influence of Cloud computing
- Technology Backdrop
  - Infrastructure innovation
  - Memory Wall & Storage Chasm
- Impact on Storage Market
  - Client Storage Migration to SSD & Cloud
  - Enterprise Migration to Cloud
  - Accelerating computation & storage growth
  - Disk is tape

EcoCooler Circulation System









Disk Becomes Tape

- Hubble’s Expanding Universe:
  - Everything is getting further away from everything else [Pat Helland]
- Non-persistent memory and cache
  - Data is being pulled up the memory hierarchy
  - Thru Si Via for very large on-package memories
- Persistent Storage
  - Data is being pulled up the storage hierarchy
  - Latency of disk random access increasingly impractical
  - Random read 2TB disk:
    - 20.6 days @ 140 IOPS with 8kb page
  - Disk increasingly impractical for random workloads

Tape is Dead
Disk is Tape
Flash is Disk
RAM Locality is King

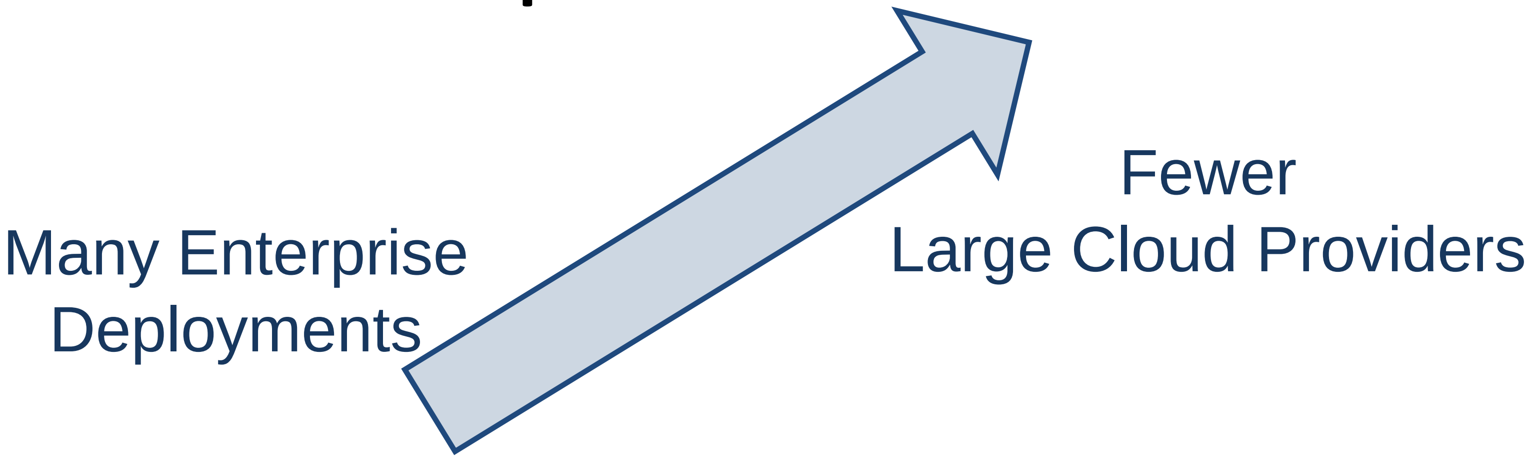
Jim Gray
Microsoft
December 2006

Client Storage Migration

- Client disk rapidly replaced by local semiconductor caches
  - Much higher performance, Lower power dissipation, smaller form factor, greater shock resistance, scale down below HDD cost floor, greater humidity range, wider temp range, lower service costs, …
  - Flash becoming primary client storage media

- Same trend in embedded devices
  - Well connected with cloud-hosted storage
- Clients storage drives cloud storage
  - Value added services, many data copies, shared access, indexed, classified, analyzed, monetized, reported, …
  - Overall HDD-based client storage continuing to expand rapidly but primarily off device in the cloud

Enterprise to Cloud



Many Enterprise Deployments

Fewer Large Cloud Providers

- Direct component supplier relationship with major operators rather than via distribution channel
- Cloud computing 5x to 10x improved price point
  - Low margin, high volume business
  - Yet still profitable, sustainable, & supporting re-investment
  - Incompatible with on-premise enterprise S/W & H/W profit margins
  - Good for customers & good for providers
- Expect many cloud winners rather than single provider

Accelerating Compute & Storage Growth

- Rapidly declining cost of computing
  - Driven by technology improvements & cloud computing economies of scale
- Traditional transactional systems scale with business
  - Purchases, ad impressions, pages served, etc.
  - Computational trading & related machine-to-machine systems limited only by value of transaction & cost of computing
- Warehousing & analytical systems scale inversely with cost
  - Cheaper storage allows more data to be analyzed
  - Lower compute costs allows deeper analysis

Cloud Storage Market is Different

- Will trade warrantee & frills for cost reduction
  - Help address the 50% “no problem found” disk RMA issue
- Will trade reliability for lower cost
  - “Reliable” never good enough so we store redundantly
  - With redundancy, we can manage failure impact
  - Good at disk replacement
- Specialization: will tailor H/W & S/W
  - Most systems internally developed
  - Willing to change any systems aspects to achieve goals
  - e.g. large disk sector sizes

Summary

- Client and Device Storage
  - Disk resident data migrating to Flash and the cloud
  - Client produced data rapidly expanding and most will end up on HDD in cloud
- Server Storage
  - Historically disk resident data migrating up storage hierarchy
  - HDD storage lost to data migration up hierarchy more than made up by:
    - Overall super-Moore growth of persistent storage exceeds losses
    - Tape losing to HDD due to volume market dynamics
  - Smaller number of very large buyers