A Close Look at PCI Express SSDs

Shirish Jamthe

Director of System Engineering

Virident Systems, Inc.

August 2011

Macro Datacenter Trends

Key driver: Information Processing

- <u>Data Footprint</u> (PB) CAGR: 100%
- <u>Data Access</u> (IOPS) CAGR: 75%
- <u>Data Access Pattern</u>: More Random
- New Application trends

Challenges

- Increasing infrastructure complexity
  - Data Center Footprint, Power, Cooling
- Increasing demands on traditional storage infrastructure
- Accelerating adoption of Enterprise SSDs

Diverging Performance: CPU vs. HDD
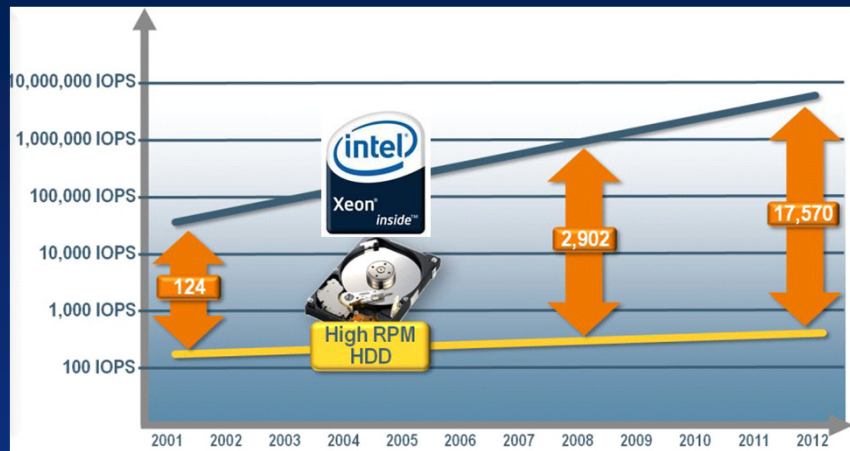
- Moore's Law – transistor density doubling every 18 months:
  - Exponential growth in CPU clock speeds, RAM, Hard Disk Drive storage density
- Mechanical components (HDD) lagging behind:
  - Seek latency: limited by disk circumference and seek arm
 - Throughput: limited by rotation speeds of the disk

Why Applications don't perform?

They are waiting on Data!

- Apps on Multi-core, Multi-Socket Servers limited by slow storage
  - App performance remains sluggish regardless of CPU speed
- Traditional remedy of adding expensive DRAM inadequate
  - Datasets double every 9–12 months
  - Non-linear Price/Density relationship
- Solve the problem by adding lots of spindles at 10% utilization
  - Wasted capacity to deliver the requisite performance
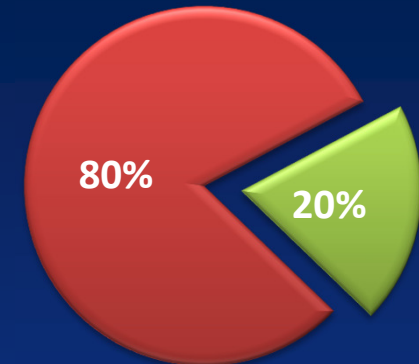 - Expensive infrastructure costs (power/cooling)

Enterprises desperate for Performance Storage tier
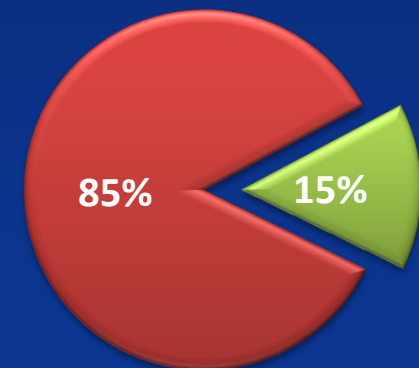
The Consequence…

- SAN build out to support data access not capacity
- Storage expenditure growing as a % of the IT spend
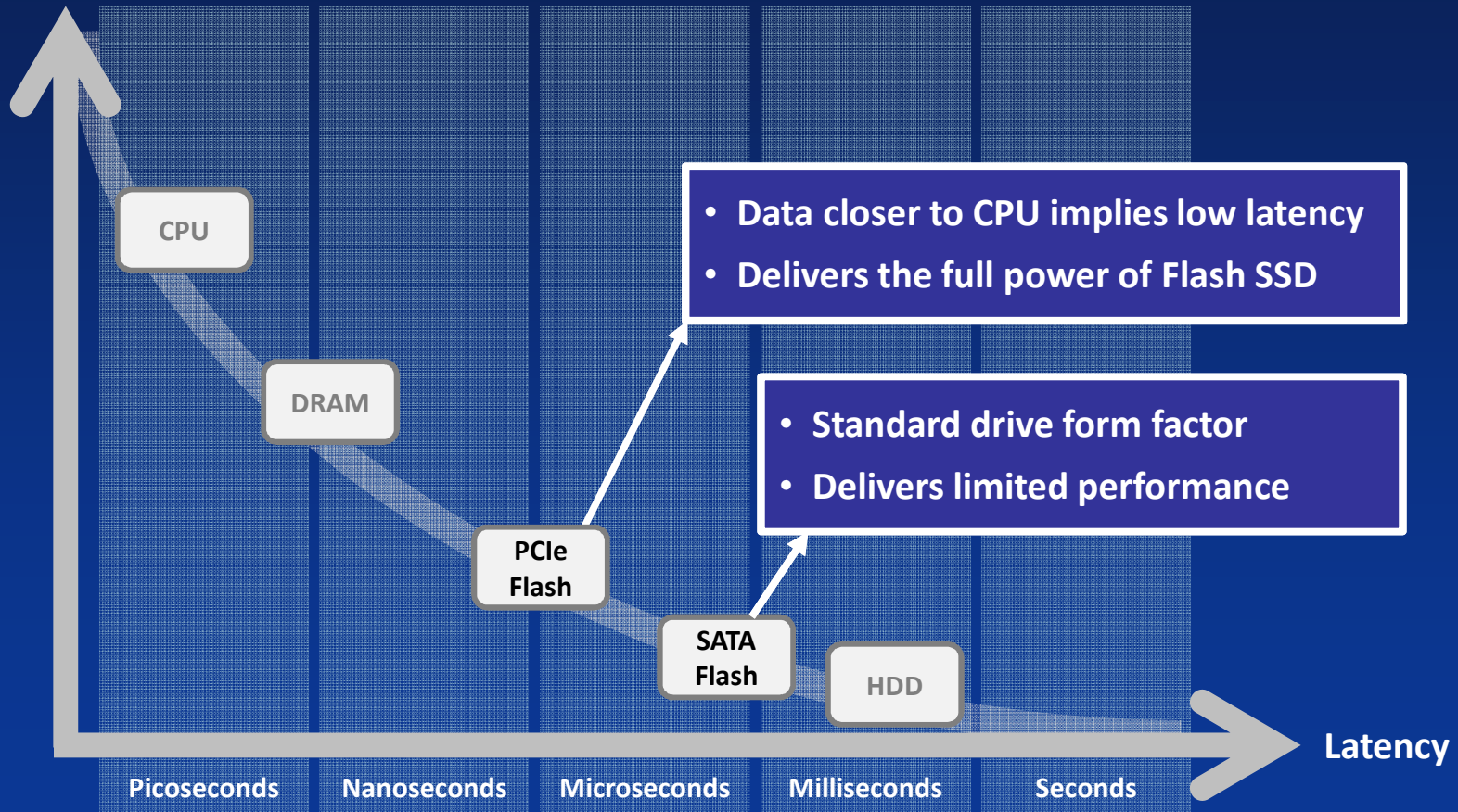- Low utilization of resources
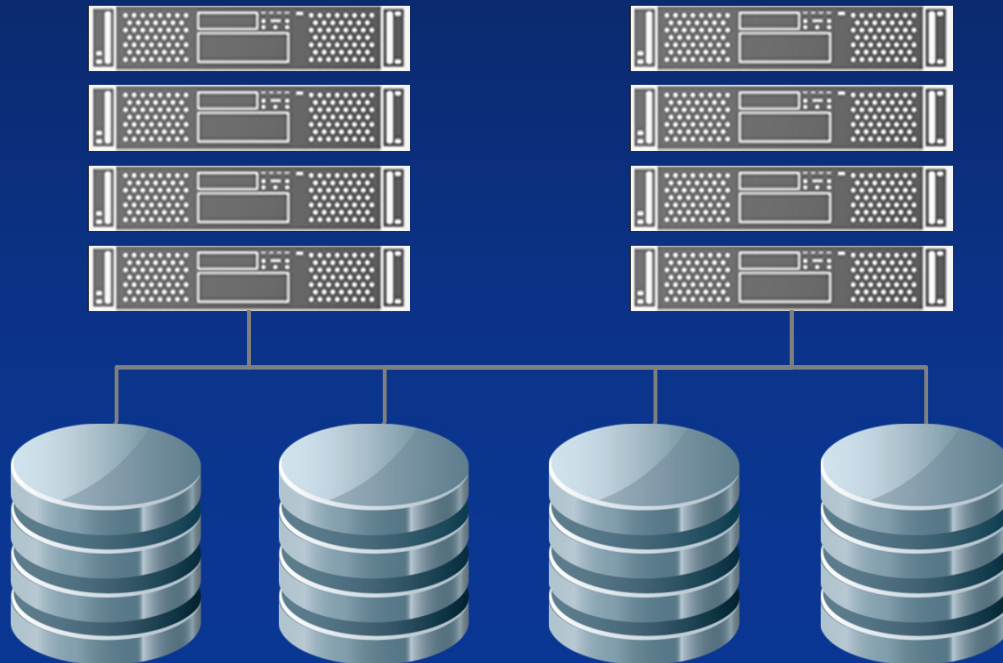
Server Utilization

80% 20%

Storage Utilization

85% 15%

■ **Productive**

Solving the Data Delivery Problem
SSDs Fill the Performance Gap

Ops / Second

CPU

DRAM

PCIe Flash

SATA Flash

HDD

Latency

Picoseconds   Nanoseconds   Microseconds   Milliseconds   Seconds

- **Data closer to CPU implies low latency**
- **Delivers the full power of Flash SSD**

- **Standard drive form factor**
- **Delivers limited performance**

PCIe vs. Non-PCIe SSDs

| PCIe	Non-PCIe (SATA/SAS)
Lower Latency (10's us)	Higher latency (high 100's us)
Direct data delivery by high bandwidth PCIe bus to CPU	Multiple hardware and protocol layers between data and CPU
Primarily DAS or Cache implementation	DAS or Shared storage implementations
Proprietary host based driver	Leverages RAID storage infrastructure

Native PCIe vs. Non-native PCIe

Native PCIe	Non-native PCIe
Lower Latency (10's us)	Higher latency (low 100's us)
High scalability with direct mapping similar to virtual memory	Storage interfaces limits the scaling due to off-load
High application scaling	Application scaling limited by multiple components
Less error-prone due to streamlined data path	Multiple context switches on the card is error prone
DAS or Cache implementation	DAS or Cache implementation
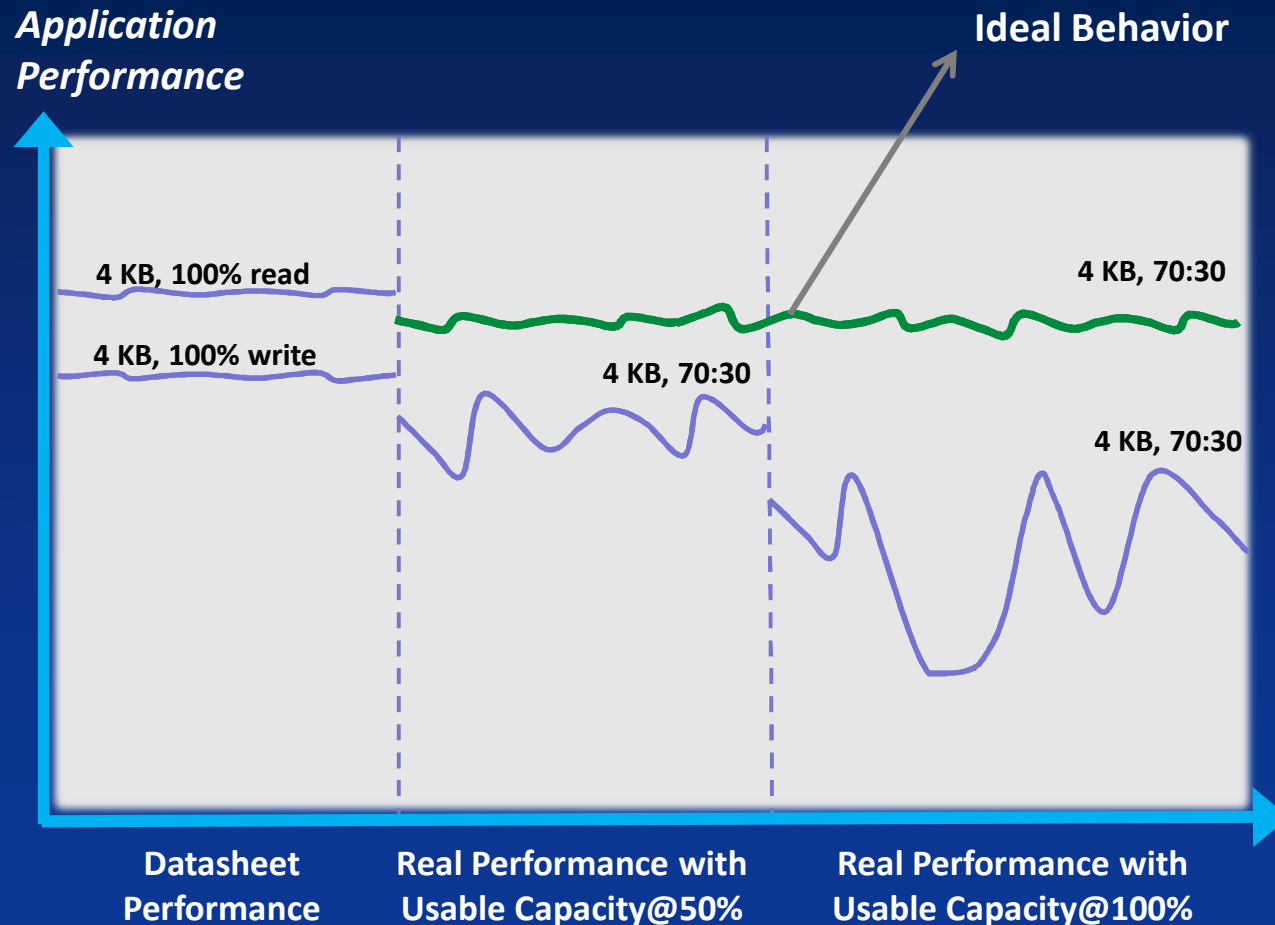
Baselining the Performance

- Operation mix
  - Reads, Writes, R+W mix
  - Sequential, Random
  - Aligned, Unaligned

- Block sizes: 512B–1MB

- Threads/Queue-depth: 1-256

- Configurations
  - Balanced
  - With and without RAID or equivalent

- With and without Garbage collection

- Key Metrics to be measured
  - Bandwidth, IOPS, Latency, Latency Standard Deviation
  - System resource overheads (CPU, Memory footprint)

Enterprises Care About…

- RAS: Reliability/Availability/Serviceability
  - Enterprise class data availability – card as well as system level
  - Even at lower geometries…

- Capacity Utilization with Performance
  - High sustained performance under full capacity utilization, real world block sizes (4K/8K)

- Capacity Density
  - High capacity in a low profile form factor
  - Wide server applicability
  - Within the PCIe power specs

About Virident Systems

- Vision: Deliver best-in-class Solid-State Storage Solutions based on Non-Volatile Memory innovations
  - "Storage Class Memory" architecture for NVM integration
  - Performance of Memory, Capacity & Persistence of Storage Drives
  - Single Server + Cluster/multi-Server
  - Optimized for data- and throughput-centric workloads

- Background
  - Founded in 2006
  - Expertise in Servers, Software & Silicon - from Intel, Sun, Google, Apple, SGI, HP
  - Significant expertise in Flash devices
  - Development in Milpitas, CA and Bangalore, India

- Currently shipping 2nd generation product: tachIOnTM drive
  - PCIe SSD using NAND Flash
  - Best-of-breed sustained performance and reliability

- Second round of funding in November 2010 by Globespan Capital Partners, Sequoia Capital, and Artiman Ventures