The State of Solid State Storage Industry Report 2010

IMEX RESEARCH.COM

IMEX RESEARCH	INDUSTRY REPORTS	EXECUTIVE SUMMARY	TABLE OF CONTENTS	ORDER FORM

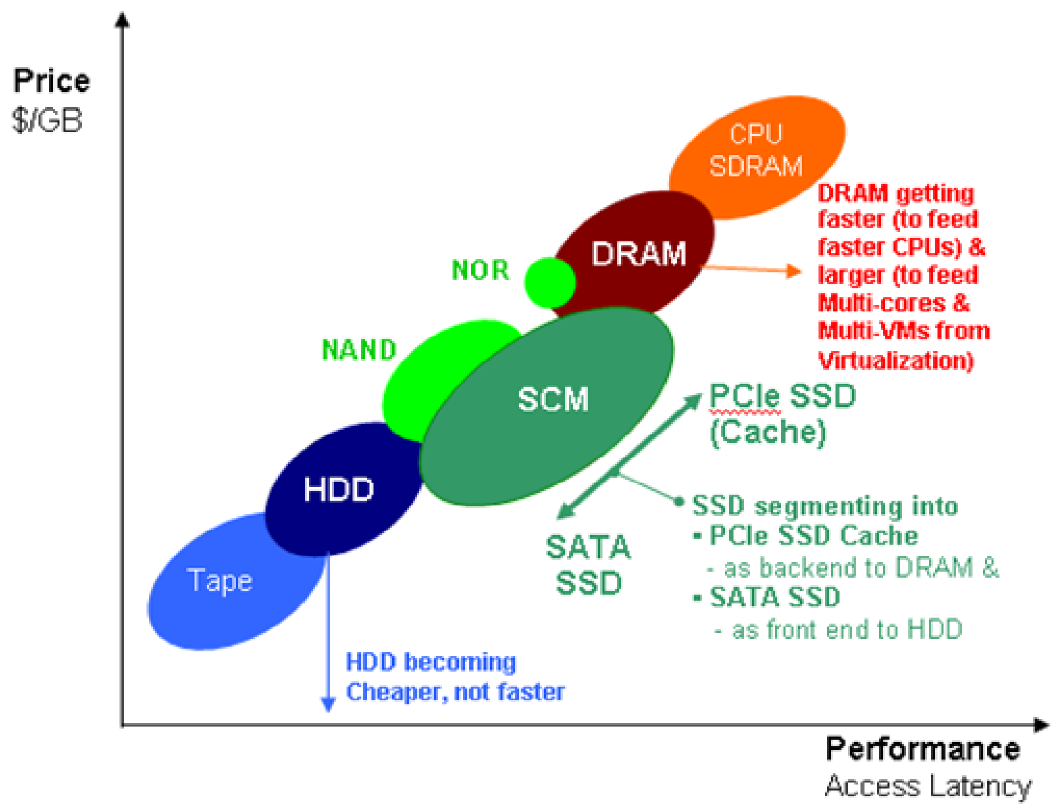
- State of Memory Storage – Leveraging Price/Performance Gaps
- SCM- A new Storage Class Memory
- Economics of SSDs
- "Enterprise-Ready SSDs"©IMEX – SSD for NextGen Data Centers & Cloud Computing
- Product Requirements for NextGen SSDs by Market Segments
- New Controllers – Key to MLC SSD Adoption
- Targeting Killer Applications for SSDs- Workloads best suited for SSDs
- Future SSD Technologies and prognosis of their success

State of Memory Storage – Industry Dynamics

Fundamental to computing are three elements – CPUs, Memory and I/O (Storage I/O & Network I/O). In the last two decades these computing elements have progressed at breakneck speed. Today we have CPUs that are ~1,000x faster, DRAMs with ~ 1,000,000x better access and Storage capacity 3,000,000x larger, than two decades ago. The remaining problem is I/O.

In a perfect world, Storage I/O would not be necessary since what applications/ workloads really want is infinite cheap storage capacity ($/GB) and immediate access (i.e. low response time or low latency) from this first level storage, in effect, get very high IOPS at a minimal cost of storage (IOPS/$/GB). That has long been the Holy Grail for computer architects.

Price/Performance Gaps in Storage Technologies



But architects (and applications/ workloads) had to yield to accommodate the real life constraints and tradeoffs of cost, access, reliability and other factors, resulting in the attached Price/Performance positioning of different storage technologies.

Industry Report

Table of Contents

IMEX RESEARCH.COM
1474 Camino Robles Ct.
San Jose, CA - 95120
Tel: 408.268.0800
Fax: 408.268.2300

Click for email address
www.imexresearch.com

Sophisticated Controllers to overcome NAND SSD's shortcomings

SSD Compétitive Technologies - Pros & Cons		
CAPSIMS[©IMEX] Metrics: Cost, Availability, Performance, Scalability, Interoperability, Manageability, Security		
Tech	**Basic Characteristics**	**Usage Potential in Storage Systems**
HDD	Cost - Cheap	Established Interfaces, large Installed Base
	Performance - Poor I/Os	Fastest HDDs can only sustain about 350 IOPS
DRAM	Cost - Expensive	Cost - Very High $/GB, Poor TCO
	Performance - Very fast, Dense Availability - Volatile	Performance - Can be used as Fast Cache Availability - Needs Complex Battery Backups
	Interoperability- No internal file system,	Interoperability - Use as Cache or Disk? DRAM Disk as Cache Replace
	Scalability - Poor Expandability/ flexibility	
NAND Flash	Cost - Reasonably Cheap, Dense	Cost - Cost/GB between HDD & RAM,, TCO Reasonable
	Availability - Non-volatile	Availability - Write cycles, media lifetime ,
	Performance -	Performance - Slow Writes,
	Interoperability -	Interoperability - Controller emulates it as HDD
		Scalability -

With the use of new sophisticated controllers, SSDs are getting closer to having best of both worlds – HDD costs and DRAM like performance for certain IOP intensive storage workloads such as Databases and OLTP with SSD models now able to sustain over 40,000 IOPS.

SCM – a new Storage Class Memory

SCM (Storage Class Memory) is a solid-state memory that is filling the gap between DRAM and HDDs by being low-cost, fast, and non-volatile. The marketplace is quickly segmenting SCMS into SATA and PCIe based SSDs

Requirements for SCMs - Key Metrics

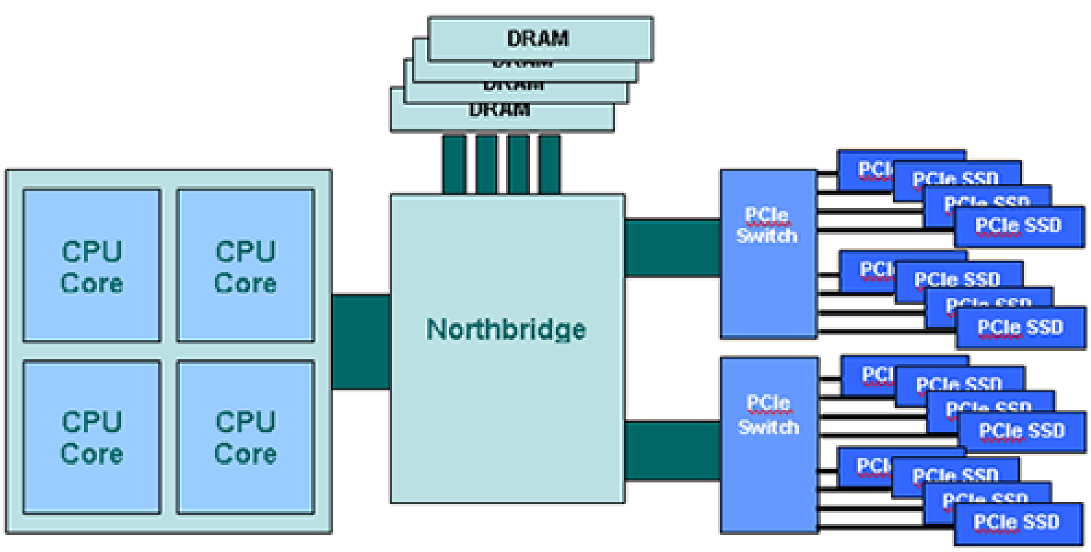
- Device - Capacity (GB), Cost ($/GB),
- Performance - Latency (Random/Block RW Access Time - ms); Bandwidth (R/W - GB/sec)
- Data Integrity – BER (Better than 1 in 10^17)
- Reliability - Write Endurance (No. of writes before death); Data Retention(Years); MTBF (millions of Hrs),
- Environmental – Power Consumption (Watts); Volumetric Density (TB/cu.in.); Power On/Off Time (sec),
- Resistance - Shock/Vibration (g-force); Temperature/Voltage Extremes 4-Corner (°C,V); Radiation (Rad)

PCIe Value Proposition

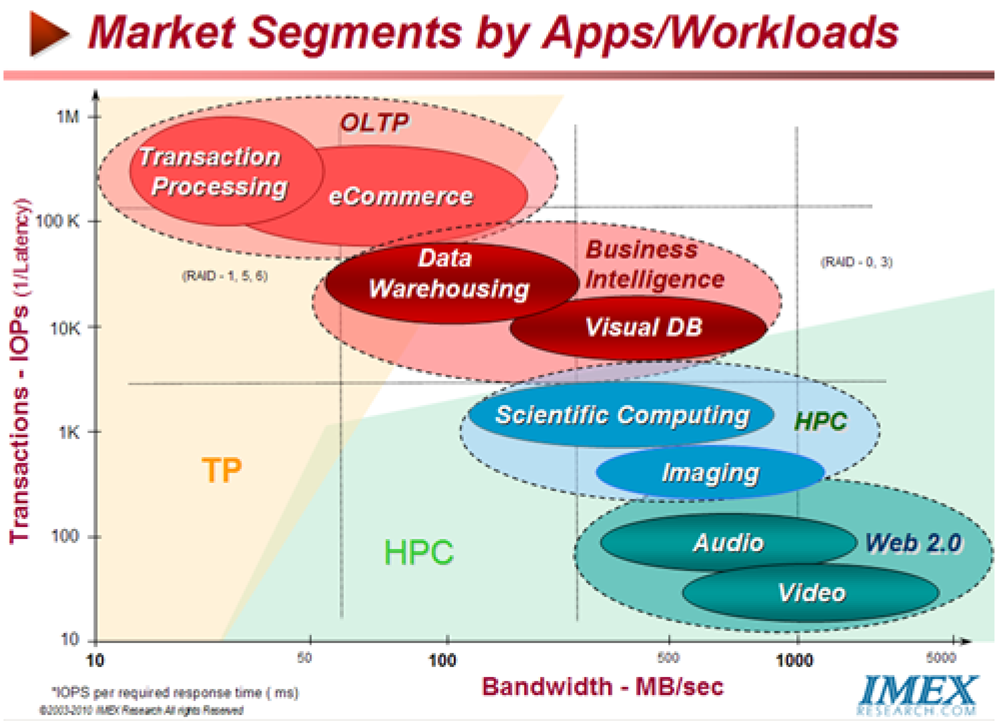
- SSD as backend storage to DRAM as the front end
- 36 PCIe Lanes Availability,
- 3/6 GB/s Performance (PCIe Gen2/3 x8),
- Low Latency in micro sec,
- Low Cost (via eliminating HBA cost)

SATA Value Proposition

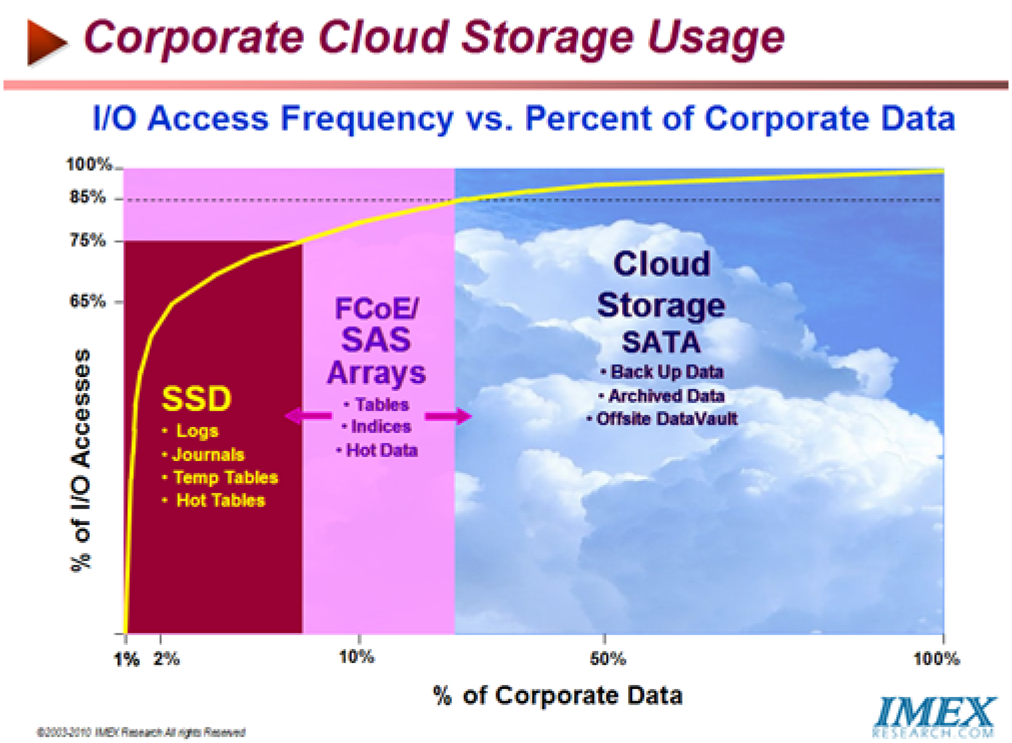
See IMEX Research Industry Report "**SSDs in the Enterprise**" with exhaustive use cases and market forecast SATA SSDs vs. PCIe SSDs.



SSDs in NextGen Data Centers and Cloud Computing



Different workloads impose different requirements on **Latency** (1/IOPs or 1/Transactions per minute or tpm) and **Bandwidth** (MB/sec) – two fundamental metrics in designing IT infrastructure be it Servers, Storage or Networking. Access Method selection, in turn, are generally dependent on best latency designable in an infrastructure architecture within constraints of cost, availability, interoperability (to achieve backward compatibility with installed base for investment protection purposes and easy migration to avoid forklift redo) and manageability (integrating into one pane of glass to view and operate from) and of course, security.

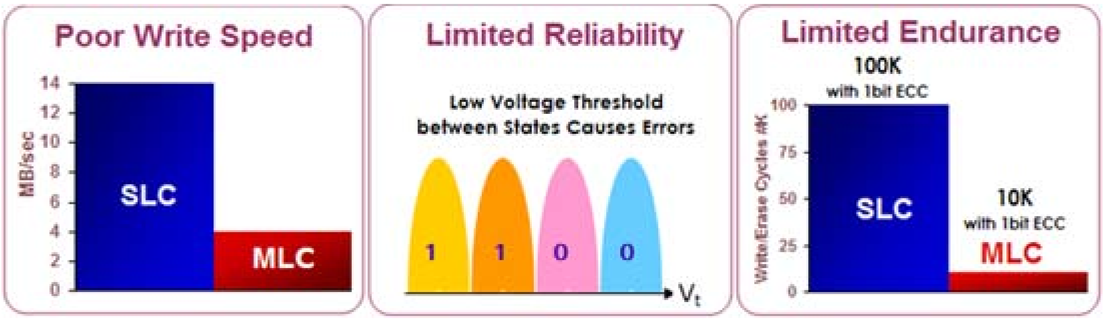


Drivers & Challenges in Developing Next Gen SSDs

SLC vs. MLC vs. TLC SSD Technologies

By using 2 bits/cell in MLC (multi-level cell) against 1 bit/cell used in SLC (single level cell), MLC NAND stores 2x the capacity. As a result MLC offers a higher density and lower cost/bit than SLC. With the cost almost the key decision metric for adoption of Flash Storage in the PC and Consumer Computing gear, lower cost/GB MLC based SSDs became the drivers necessary to accelerate SSD adoption. But issues related to reliability (endurance, data retention…), performance, and adaptability to existing storage interfaces, ease of management etc. became the challenges to overcome.

Challenges with enabling MLC SSDs



Raw Media Reliability

- No moving parts
- Predictable wear out
- Post infant mortality catastrophic device failures rare
- Higher density of MLC increases bit error rate
- High bit error rate increases with wear
- Program and Read Disturb Prevention
- Partial Page Programming
- Data retention is poor at high temperature and wear

Media Performance

- Performance is excellent (compared to HDDs)
- High performance per power (IOPS/Watt)
- Low pin count: shared command / data bus, good balance
- NAND not really a random access device
  - Block oriented
  - Slow effective write, erase/transfer/program) latency,
  - Imbalanced R/W access speed
- NAND Performance changes with wear
- Some controllers do read/erase/modify/write
- Others use inefficient garbage collection

Controller

- Transparently converts NAND Flash memory into storage device
- Manages high bit error rate
- Improves endurance to sustain a 5-year life cycle
- Interconnect
- Number of NAND Flash Chips (Die)
- Number of Buses (Real / Pipelined)
- Data Protection (Int./Ext. RAID; DIF; ECC…)
- Write Mitigation techniques
- Effective Block (LBA; Sector) Size
- Write Amplification
- Garbage Collection (GC) Efficiency
- Buffer Capacity & Management
- Meta-data processing

The Real Endurance numbers…

One serious drawback of MLC has been its lower endurance to withstand data write/erase cycles (typically at 10,000 vs. 100, 000 for SLC), besides slower write speeds and higher bit error rates compared with SLC NAND. Thus

- Moving from HDD and mechanical issues to SSD with "hard" limits on writing can be complex
- Different vendors show different wear levels on raw NAND
- As geometry shrinks so do Endurance and Reliability

Retaining Customer Data…

- Raw NAND retention is inversely proportional to cycles
- NAND media types also have different wear out factors
- How long is good enough for Enterprise SSDs
- more»

New Controllers - Key to MLC SSDs Adoption

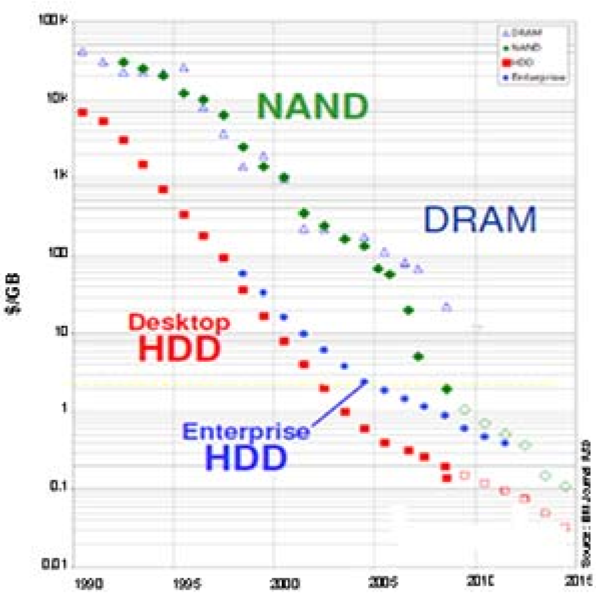
Now with the industry on a solid roadmap for the future through a continuous cost reduction by increasing the bit density by adopting 2, 3, and 4 bits per cell (bpc) propels it towards mass adoption of MLC technology based SSDs.
To leverage Flash NAND with its genesis as Non-Volatile Memory capable of semiconductor based mass production techniques and use them as self contained storage devices required an interface to connect to the host, an advanced device controller besides the NAND Flash semiconductor components and packaged them in a single device ready to plug into computers.
To meet the rigorous requirements of their use in the enterprise where reliability and performance requirements supersede cost, new sophisticated controllers and firmware had to be devised before they could be adopted as mission critical applications in the enterprise. Now sophisticated controllers with advanced architectures are being made available from a number of manufacturers (for an exhaustive industry updates see IMEX Research's Industry Report "SSD in the Enterprise") to mitigate the key challenges posed by MLC SSDs.

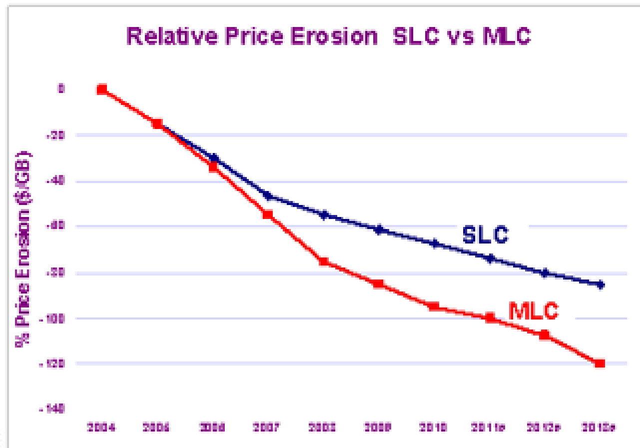
Earlier Shortfalls

- High cost due to use of low density single bit SLC NANDs
- Using Higher density MLC increased bit error rate
- Relatively high bit error rate increases with wear
- Program and Read Disturbs
- Partial Page Programming
- Data retention poor at high temperature and wear

Shortfall mitigation by Modern Controllers

Today MLC NAND is able to overcome above shortfalls experienced in previous years and now meet the cost/performance/ reliability requirements of SSDs for use in the enterprise through techniques such as:





COST

- Using 2 and 3 bit per cell MLC NANDs for cost reduction

PERFORMANCE

Factors Impacting Performance

- Hardware (CPU, Interface, Chipset …)
- SW (OS, App, Drivers, Various caches, SSD specific TRIM, Purge, …)
- Device (Flash Generation, Parallelism, Caching Strategy, Wear-Leveling, Garbage Collection, Warranty Strategy…)
- Write History (TGW, spares…)
- Workload (Random, Sequential, R/W Mix, Queues, Threads…)
- Pre-Conditioning (Random, Sequential, Amount …)
- Short "Burst" performance when First On Board (FOB)
  FOB state not important unless drive can return to FOB like performance somehow. Performance can change dramatically with time
- Can have many transition phases
- Performance comparison valid only under same condition

Using interleaved memory banks, caching and other techniques being designed in modern controllers, the performance of MLC SSDs today matches and even outshines performance offered by SLC SSDs.

MANAGING ENDURANCE

To overcome NAND's earlier endurance shortfalls due to its limitation in number of write/erase cycles per block, new controllers manage NAND using

- **Error Correction Techniques** – To correct and guard against bit failures, same that has been commonly used in hard disk drives for years.
- **Built-in Wear Leveling Algorithms** - Writing data in a way that evenly distributes over all of available cells so it avoids a block of cells being overused and cause failures.
- **Over-provisioning Capacity** - Extra spare raw blocks are designed-in as headroom and included to replace those blocks that get overused or go bad and additionally provide enough room for wear-leveling algorithms to operate, thus enhancing the reliability of the device over its life-cycle. A typical SSD device's specified GB device will actually contain 20-25% extra raw capacity to meet these criterions.

RELIABILITY MANAGEMENT

Multiple techniques are being used to improve the reliability, such as:

- **In-Flight** -Corruption upstream disk controllers, Corruption in SSD controller itself, Flush at power loss, using large cap elements

- **At-Rest**
- **ECC,** - Scanning & scrubbing, - Redundancy
- **Meta-Data**
- **Error correcting memory,** - Data integrity field

"Enterprise-Ready SSD" ©2010 IMEX Research

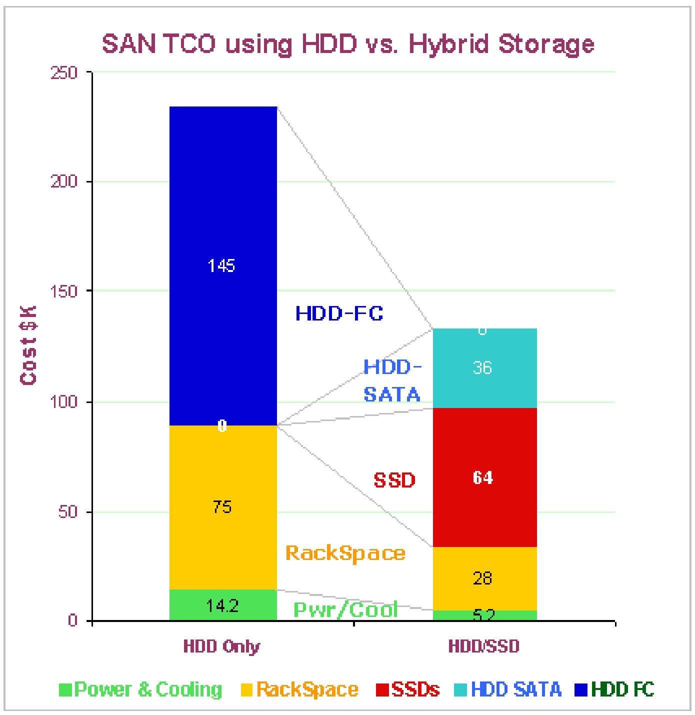
These advanced controllers manage the above features to help make NAND Flash suitable as "Enterprise-Ready SSD" (©2010 IMEX Research) to meet the expected:

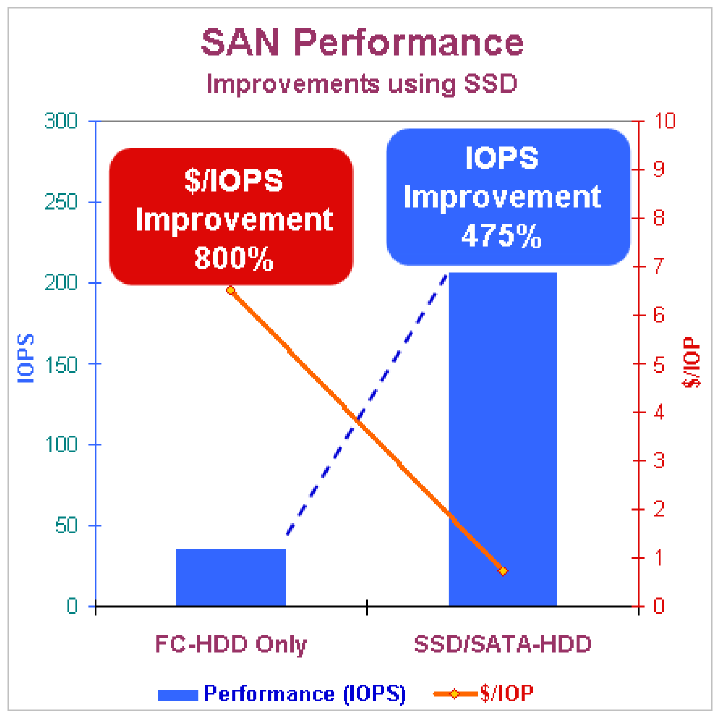
- **Fast I/O Performance** required by business-critical applications and
- **5-Yr. Life Cycle Endurance** required by mission-critical applications in the enterprise.
-

Hybrid Storage

To combine the best of features of SSDs - outstanding Read Performance (Latency, IOPs) and Throughput (MB/s) and the extremely low cost of HDDs has given rise to a new class of storage - Hybrid Storage Devices (brought to market by Seagate, EMC, Nvelo, Violin Memory etc)

For an exhaustive in-depth study of markets, adoption rates, newer technologies, newer standards, vendor offerings and their competitive strategies and positioning plus future directions see IMEX Research's detailed report on Solid State Storage in the Enterprise 2010.



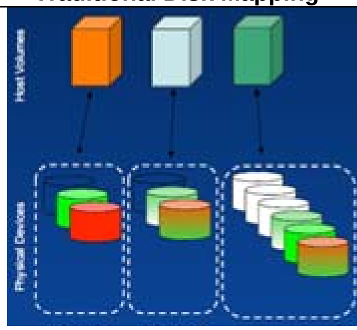
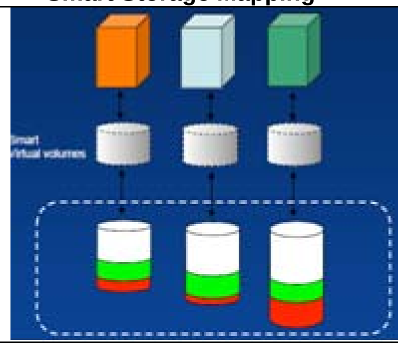


Automated Storage Tiering – The Killer App for SSDs

Automated Tiered Solid State Storage is the next killer application for SSDs

EMC – FAST (Fully Automated Storage Tiering)
- Continuously monitor and analyze data access on the tiers
- Automatically elevate hot data to "Hot Tiers" and demote cool data/volumes to "Lower Tiers"
- Allocate and relocate volumes on each tier based on use
- Automated Migration reduces OPEX to otherwise manage SANs manually

IBM – Smart Tiering Technology

Traditional Disk Mapping	Smart Storage Mapping
	

Volumes have different characteristics - Applications need to place them on correct tiers of storage based on usage.
All volumes appear to be "logically" homogenous to apps. - But data is placed at the right tier of storage based on its usage through smart data placement and migration

Workload I/O Monitoring & Smart Migration to SSD

Every workload has its unique IO access signatures and behavior over time. IBM has a Smart Monitoring and Analysis Tool that allows customers to develop deeper insight into the application's behavior over time to allow optimization of storage infrastructure supporting it. A typical historical performance data for a LUN over time is shown that reveals performance skews and hot data regions in three LBA ranges.
Smart Tiering Technology identifies these hot LBA regions and non-disruptively migrates "hot data" from HDD to SSD. Typically about 4-8% of data becomes candidate for migration from HDD to SSD depending on the workload. Result: Response time reduction of 60-70+ % at peak loads.

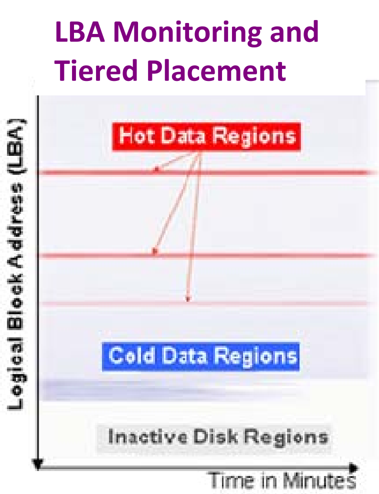
Response Time Improvement - Productivity Enhancements for OLTP Transactions using SSDs

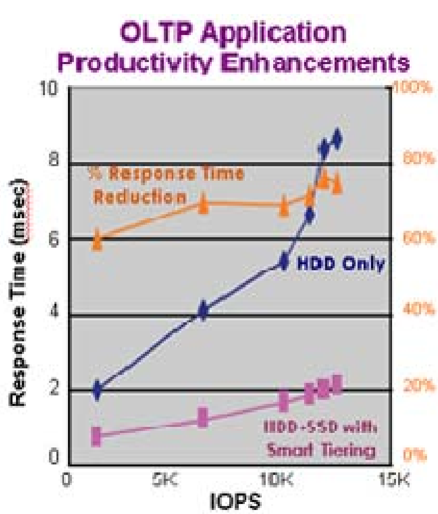
Using Smart Tiering Technology Monitoring and using automated reallocation of hot spot data (typically 5-10% of total data) to SSDs organizations can typically achieve performance improvement benefits in:
Response time reduction of around 70+% or
o Through put (IOPS) increase of 200% for any I/O intensive loads as experience by Time-Perishable Online Transactions such as Airlines Reservations, Wall Street Investment Banking Stock Transactions, Financial Institutions Hedge Funds etc. as well as Low Latency seeking High Performance Clustered Systems etc.

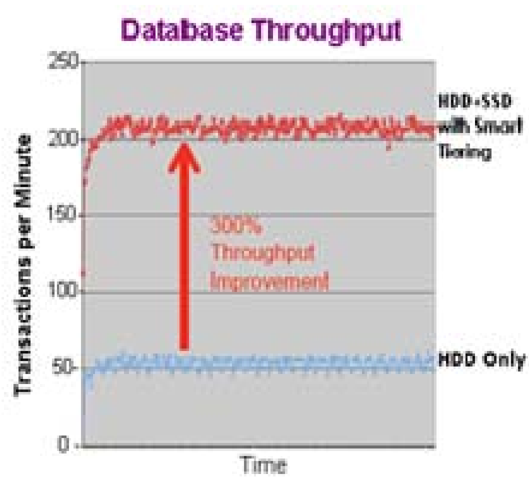
Brokerage Workload Optimization Using Smart Tiering

- Identify hot "database objects" and smartly placed in the right tier.
- Scalable Throughput Improvement - 300%
- Substantial IO Bound Transaction Response time Improvement - 45%-75%

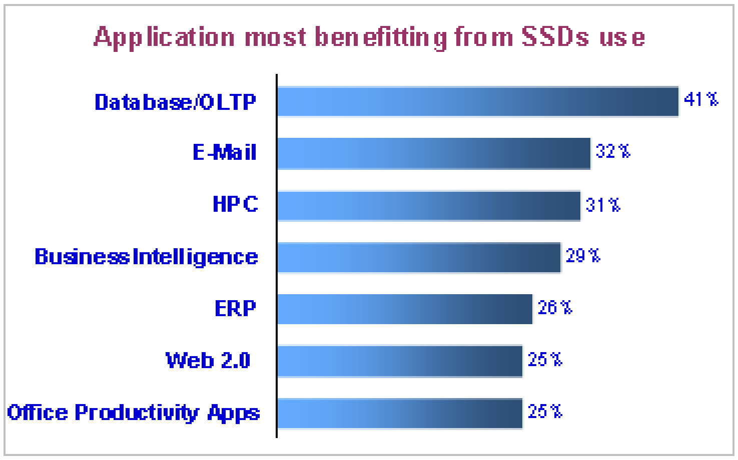
Enhancement using IBM Smart Tiering







Workloads best suited for SSD



- Database
  - Databases have key elements of Commit files – logs, redo, undo, tempDB
- **Structured data**
  - Structured data access is an excellent fit for SSD
  - Exception – large, growing table spaces
- **Unstructured data**
  - Unstructured data access is a poor fit for SSD
  - Exception – small, non-growing, tagged files
  - OS images – boot-from-flash, page-to-DRAM

Economics of SSDs

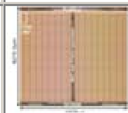
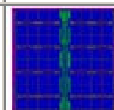
Multiple companies have achieved outstanding results through using SSDs in combination with HDDs to achieve the best of both worlds – excellent read performance of SSDs with cost effective low cost $/GB of HDDs. In the process they have been able to achieve

In a typical SAN environment attached graph typically depicts cost reductions - $230K using large number of Fibre Channel HDDs most commonly used in enterprises to achieve better performance vs. cost of $130K using SSDs with lower cost SATA achieving a TCO reduction of 76%, as shown. In the process IOPS performance improvements of 475 % and $/IOP reductions of a whopping 800% have been achieved. For more details refer to **IMEX Research Industry Report**.

Future SSD Device Technologies - Status & Success Prognosis

(Courtesy: J.Freitas, IBM)

New technologies currently under development in research labs around the world that promise to replace today's NAND Flash technology. These new technologies - collectively called Storage Class Memory (SCM) – are being targeted to provide higher performance, lower cost, and more energy efficient solutions than today's SLC/MLC NAND Flash products.

..	**Improved Flash**	**FeRAM**	**MRAM**	**Racetrack**	**RRAM**	**Memristor**	**Solid Electrolyte**	**PCRAM**
	 64Mb FeRAM (Prototype) 0.13um 3.3V				 4Mb PCRAM (Product) 0.25um 3.3V		 512Mb PCRAM (Prototype) 0.1um 1.8V	 4Mb MRAM (Product) 0.18um 3.3V
Knowledge level	advanced development	product	product	basic research	Early development	Early development	development	advanced development
Smallest Cell demonstrated	**4F2** (**1F2** per bit)	15F2 (@130nm)	**25F2** @180nm	—	—	—	**8F2** @90nm (4F2 per bit)	**5.8F2** (diode) **12F2** (BJT) @90nm
Prospects for ……Scalability	**maybe** (enough stored charge?)	**Poor** (integration, signal loss)	**Poor** (high currents)	**Unknown** (too early to know, good potential)	**unknown**	**unknown**	**promising** (filament-based, but new materials)	**promising** (rapid progress to date)
…fast readout	yes	yes	yes	yes	yes	yes	yes	yes
…fast writing	**NO**	yes	yes	yes	**sometimes**	**sometimes**	yes	yes
…low Switching Power	yes	yes	**NO**	**uncertain**	**sometimes**	**sometimes**	yes	**poor**
…high endurance	**Poor** (1e7 cycles)	yes	yes	should	**poor**	**poor**	**unknown**	yes
…non-volatility	yes	yes	yes	**unknown**	**sometimes**	**sometimes**	**sometimes**	yes
…MLC operation	yes	**Difficult**	**NO**	yes **(3-D)**	yes	yes	yes	yes
Cos. pursuing	Spansion Infineon Macronix Samsung Toshiba NEC Nano-x'tal Freescale Matsushita	Fujitsu STMicro TI Toshiba Infineon Samsung NEC Hitachi Rohm HP Cypress Matsushita Oki Hynix Celis Fujitsu Seiko Epson	IBM Infineon Freescale Philips STMicro HP NVE Honeywell Toshiba NEC Sony Fujitsu Renesas Samsung Hynix TSMC		IBM Sharp Unity Spansion Samsung		Axon Infineon	Ovonyx BAE Intel STMicro Samsung Elpida IBM Macronix Infineon Hitachi Philips