Desktop, Nearline & Enterprise Disk Drives

- *Deltas by Design*

Willis Whittington, Seagate Technology
SNIA Legal Notice

• The material contained in this tutorial is copyrighted by the SNIA.
• Member companies and individuals may use this material in presentations and literature under the following conditions:
  – Any slide or slides used must be reproduced without modification
  – The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
• This presentation is a project of the SNIA Education Committee.
Agenda

• Fundamentals & Firmware
• SAS & SATA Compatibility
• Rotational Vibration
• Data Error Rate
• SATA Error Recovery – NCQ Limitations
• Error Correction Capability
• Data Integrity
• Performance
• Annualized Failure Rate

~ Q & A Along the way ~
Desktop/Enterprise Differences

- **Desktop Storage**
  - 7,200 rpm
  - SATA

- **Enterprise Storage**
  - 15,000 rpm
  - SAS/FC

- **Cost**
- **Power**
- **Capacity**
- **Duty Cycle**
- **Reliability**
- **Scalability**
- **Environment**
- **Performance**
- **Data Integrity**
- **System Architecture**
Nearline Enhancements to Desktop

**Nearline SATA**

- **Desktop**
- **Enterprise SAS/FC**

**I/O Active**
- Low / Med
- High

**Reliability**
- Low: 0.6
- Medium: 1.2
- High: 1.6

**Environment**
- Tolerance to RV (Radians/sec²)
- Low: 6
- Medium: 12.5
- High: >21

**Data Integrity**
- Drive Electronics
- Nearline: n/a
- Enterprise: IOEDC & IOECC

*Command in Execution

The right drive for the right application
Desktop compared to Nearline

**DESKTOP**

- **Reliability**: Standard Desktop
- **Firmware**: Low
- **Duty Cycle**: 8 Hours x 5 Days
- **Data Integrity**: Standard
- **Power On Hours**: 6 Radians/sec²
- **Rotational Vibration**: Standard
- **Micro Code Download**: Standard
- **Error Recovery Control**: None
- **Workload Management**: None
- **Power Management**: None

**NEARLINE**

- **Reliability**: 1,200,000 Hrs
- **Firmware**: Enterprise Features
- **Duty Cycle**: Low / Medium
- **Data Integrity**: IOEDC
- **Power On Hours**: 24 Hours x 7 Days
- **Rotational Vibration**: 12½ Radians/sec²
- **Micro Code Download**: Segmented
- **Error Recovery Control**: Time Selectable
- **Workload Management**: Standard
- **Power Management**: Standard
Enterprise H/W Differences to NL & Desktop

Electronics
- Dual processors
- Multi host
- Dual port
- Twice the firmware
- High rpm control
- Command scheduling
- Superior error protection
- Superior error correction
- Smart servo algorithms
- Perform. optimization
- Data integrity checks
- Sequential h/w assist

Motor
- Higher rpm
- Less runout
- More expensive

Discs
- More platters
- Smaller diameter
- Full media certification
- Fully characterized

Head Stack
- More heads
- Low mass, high rigidity
- Higher cost design
EDUCATION

SAS & SATA Compatibility
• Connectors and cables are required to:
  – Support 3.0 Gbps speed with headroom for 6 Gbps
  – Be suitable for both 3½” and 2½” storage devices
  – Be blind-mateable
  – Be hot-pluggable (staggered pins)
SAS/SATA Connector Compatibility

Connector Flip Side

Port B
Keyway

SATA
Pluggable
SAS
SAS/SATA Compatibility

- SATA drives are plug compatible with SAS backplanes
- SAS Topology allows SATA device addressing
- This is done with Serial Tunneling Protocol (STP)
SAS/SATA Compatibility

- SATA drives are **plug compatible** with SAS backplanes
- SAS Topology allows SATA device addressing
- This is done with Serial Tunneling Protocol (STP)
SATA to SAS Interposer Card

- SATA drives are designed for a desktop/cabled interconnect
- Onus is on the Integrator to assure use in Backplane applications
- Interposer card
  - Provides additional Transceiver drive capability in SAS cabinets
    - Necessary to preserve signal integrity in medium to large backplanes
  - Provides additional electronics for “Port Selector” feature
    - Buffer and Mux for active-active access
Rotational Vibration
RV Emitted by a Seeking Drive

Neighboring Drive’s Servo needs to compensate for externally induced RV

HDA subjected to rotational force

Reactive Force

RV is Proportional to Seek Current

'Scope Picture, Seagate Prod. Dev.
Rotational Vibration

Impact on Performance*

*Source: STX Competitive Analysis.
Rotational Vibration

- More stringent RV spec. needed for SATA cabinets
- RV aggravated by system fans, random operation and “bursty” workloads
Data Error Rate
UER* on Very High Capacity RAID Sets

• The UER for SATA desktop products is 1 in $10^{14}$ bits read
  – $10^{14}$ bits = 12½ terabytes
• A 500 Gbyte drive has $1/25 \times 10^{14}$ bits
• Rebuilding a SATA drive in a RAID 5 set of 5 drives means transferring $5/25 \times 10^{14}$ bits
  – This means there is a 20% probability of an Unrecoverable Error during the rebuild.
• Better odds would be available with RAID 1 or 6
  – Higher redundancy, faster Rebuild, more expensive
• Risks can be reduced with good error management
  – Intelligent rebuild (e.g. ignore unused capacity)
  – Background media scan

*Unrecoverable Error Rate
Desktop / Nearline / Enterprise UER*

Probability of Unrecoverable Errors during RAID Rebuild

- **Desktop Drives UER = 10^{-14}**
- **Nearline UER = 10^{-15} (2007)**
- **Enterprise UER = 10^{-16} (2006)**

*Unrecoverable Error Rate
Limitations of SATA
Error Recovery
with
Native Command Queuing
Creating the Command Queue
Optimize Queue and Execute

Commands will not be executed

Execute

Status = Error
Consequences of an Error

- Command Queue processing halts on an ERROR
- All commands remaining in the Queue are cleared
Consequences of an Error

- Command Queue processing halts on an ERROR
- All commands remaining in the Queue are cleared
- Host must reissue outstanding commands

Complicates Dual Port Interposer Applications

Reissue Cmd 2, 3, 4, & 5 during Error Recovery

- Command 1
- Command 6
Error Correction Capability
Standard vs Reverse ECC

(Write Command)

**Standard ECC**
- User Data
- ECC Generator
- User Data
- Randomizer
- Encoded

**Reverse ECC**
- User Data
- Randomizer
- Randomized
- RLL Encoder
- Encoded
- ECC Generator
- Encoded

Desktop, Nearline & Enterprise Disk Drives – Deltas by Design
© 2007 Storage Networking Industry Association. All Rights Reserved.
Standard vs Reverse ECC

(Read Command)

**Standard ECC**
- User Data
- ECC Correction
- De-randomize
- EC
- 120 Bit Error (Propagates in Decoder)
- RLL Decoder
- User Data

**Reverse ECC**
- User Data
- De-randomize
- RLL Decoder
- ECC Correction
- Error Corrected on the fly
- User Data
- EC
- 20 Bit Error
Sync Mark Errors on Desktop/NL Drives

- The Sync Field is used to get the read channel in frequency sync with the data recorded on the media.
- The Sync Mark is used to define the beginning of the User Data Field.
- Failure to recognize the Sync Mark (due to a thermal asperity or a grown media defect) means the User Data Field is not delineated and the data is lost.
Sync Mark Errors on Enterprise drives

Sync Mark → Media Flaw → Sync Mark 2 embedded in the data field → Sector Format

Read Operation

Before ECC

DATA BUFFER

00000000000000 0011011000011110101011000110110001010101000111

The missing n bytes are recovered using the ECC

After ECC

11010111010011 0011011000011110101011000110110001010101000111

Read Channel realizes the first Sync Mark is missing:
- Loads Buffer with n zeroes
- Starts searching for SM2

Read Channel finds SM2 and reads the data following it into the Buffer, starting at location n+1

The missing n bytes are recovered using the ECC

0011011000011110101011000110110001010101000111
End to End Data Checking

Data Frame

- SOF
- Header
- User Data
- End to End Data Check
- CRC
- EOF

Host and drive can check the data

- Logical Block Guard: 2 Bytes
- Logical Block Application Tag: 2 Bytes
- Logical Block Reference Tag: 4 Bytes

Not available on Desktop and SATA Nearline drives with fixed 512 byte sectors
Data Integrity Protection in Enterprise Drives

I/O Connector

Data Frame from Host

- Data
- CRC
- Data
- P
- Data
- IOEDC
- Data
- IOEDC
- IOECC

I/O Controller

DDR SDRAM

Buffer

- Data checked against CRC
- CRC Discarded, temporary Parity added
- LBA Seeded Error Detection Code added to Data
- Error Correction Code added to Data & IOEDC
- Protected Data now processed by drive

Disc Drive Block Diagram

- Data checked against IOECC
- IOECC Discarded
- Data & IOEDC are RLL encoded
- Reverse ECC added
- Final EDC added before writing

Formatter

- Data
- IOEDC
- IOECC
- Data
- IOEDC
- DataRLL
- IOEDC_RLL
- ECC_RLL
- DataRLL
- IOEDC_RLL
- ECC_RLL
- ECC

Eliminates the risk of unreported data miscompare errors
Performance
Performance Comparison

Enterprise Sequential Access

Desktop / NL Sequential Access

Vendor Range

Sequential Transfer Rate \( \propto \) RPM x Disc Dia x Bit Density

Enterprise Transfer Rate

\[
\frac{\text{Enterprise Transfer Rate}}{\text{Desktop Transfer Rate}} = \frac{\text{RPM x Disc Dia x Bit Density}}{\text{RPM x Disc Dia x Bit Density}} = \frac{15000 \times 65}{7200 \times 95} = 142\%
\]

(independent of seek time & Latency)
Performance Comparison

Enterprise Sequential

Desktop / NL Sequential

Enterprise Random

Desktop / NL Random

Operations/sec

Mbytes/sec

1 4 16 32
SPC-1C Performance Comparison

• SPC-1C comprises I/O operations designed to demonstrate small storage subsystem performance (1-16 drives) while performing the typical functions of a business critical application.
AFR

(Annualized Failure Rate)
SATA drives in Enterprise applications run hotter, at higher duty cycle, and for more Power-On-Hours than in desktop applications.
In Conclusion.....

Although technological advances, driven by Enterprise research, will be leveraged into SATA products, there will continue to be functional limitations imposed on these devices by the overriding metric of Low $/GB Storage.
Q&A / Feedback

• Please send any questions or comments on this presentation to SNIA: trackstorage@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

SNIA Education Committee

Willis Whittington