



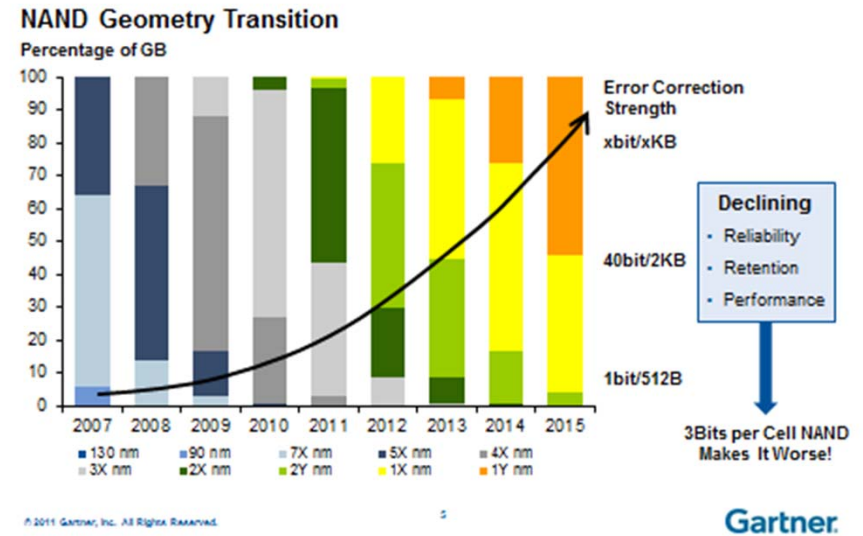
Designing Enterprise SSDs with Low Cost Media

Jeremy Werner
Director of Marketing
SandForce

Everyone Knows...

NAND Evolution Concerns: Cost Decline = Quality Decline

- Flash is migrating:
 - To smaller nodes
 - 2-bit and 3-bit MLC
- \$/GB decreases due to increasing transistor density (lower geometries)
 - Addressing demands of the consumer market
- Major trade-off in terms of reliability, endurance, and performance
- Yet more than ever organizations want lower cost flash in the Enterprise Computing market!



Source: Gartner June 2011



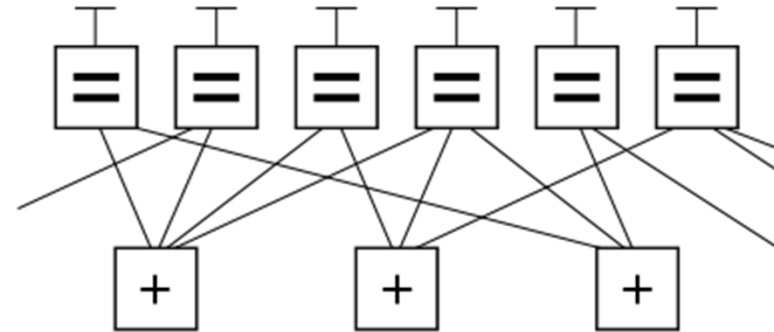
Stepping up to the challenge

- Enterprise SSD Technology enable analyst's forecasted growth
- Full system ASIC and FW co-design is critical
- Advanced critical, required capabilities including:
 - ▶ Advanced Flash ECC
 - ▶ RAID-like protection
 - ▶ Soft Error protection
 - ▶ End-to-End CRC
 - ▶ Native non-512 Byte sector support
 - ▶ Write Reduction Technologies
 - ▶ Power-Fail Protection
 - ▶ Consistent Low Latency Performance
 - ▶ Temperature Intelligent Technology
 - ▶ Predictive Failure Capabilities

Advanced Flash ECC

- Today's State of the art
 - ▶ High Powered BCH
 - ▶ Data Randomizer
 - ▶ Advanced Read-Retry
 - ▶ 512Byte and 1KByte code words

Example LDPC code using Forney's factor graph notation

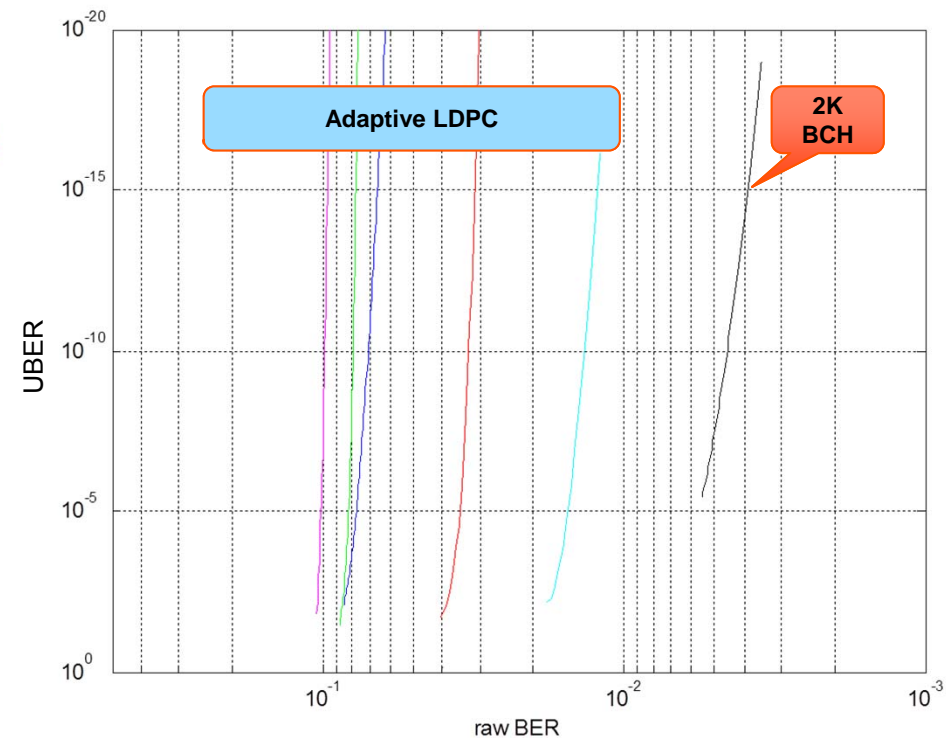


Source: Wikipedia <http://en.wikipedia.org/wiki/Ldpc>

- 2013 Requirements
 - ▶ Soft and Hard LDPC (Low Density Parity-Check)
 - 10-100x more correction than traditional BCH
 - ▶ DSP-Aided Intra-cell and Inter-cell equalization
 - ▶ Adaptive code-rates
 - ▶ 2KByte code words

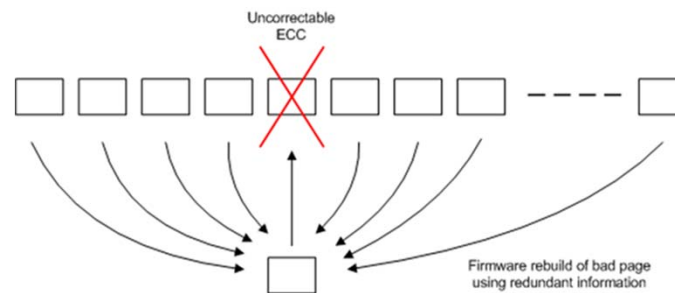
Soft and Hard LDPC and DSP Challenges

- Enables remarkable bit error correction capabilities $\sim 10\%$ RBER but design challenges include:
 - ▶ Adequate Error Floor (10^{-16} UBER)
 - ▶ Efficient Iteration Requirements (1-1.5x)
 - ▶ Very high throughput ($>2\text{GB/s}$)
 - ▶ Low Latency ($<2\mu\text{s}$)
 - ▶ Low Power (Zero-power idle, 50-100mW active)
 - ▶ Small silicon footprint
 - ▶ Flexible code rates (BOL v. EOL adaptive)



RAID-like Protection

- SandForce introduced RAISE in 2009
 - ▶ Redundant Array of Independent Silicon Elements
 - ▶ Page and Block level protection against uncorrectable errors
 - ▶ RAID-5 like protection (single uncorrectable per stripe)
- In 2013 more advanced RAID-like protection will be needed
 - ▶ Multi-die failure
 - ▶ RAID-6 like protection (two uncorrectable per stripe)
 - ▶ Advanced internal rebuild capabilities





End-to-End Data Protection

- Data Protected at System Level by End to End Data Protection
 - ▶ Although older concept not universally adopted yet
 - ▶ T10 DIF (520-Byte Sectors) is the most common for SCSI devices
 - Also proprietary solutions e.g. 524, 528 Byte
 - 4K + DIF sectors coming
 - ▶ NVM Express has Data Integrity support for PCIe SSDs
- Critical to support the larger sectors without performance or ECC loss
- Also must handle fancy pattern generation to account for multiple heterogeneous host and initiator infrastructure



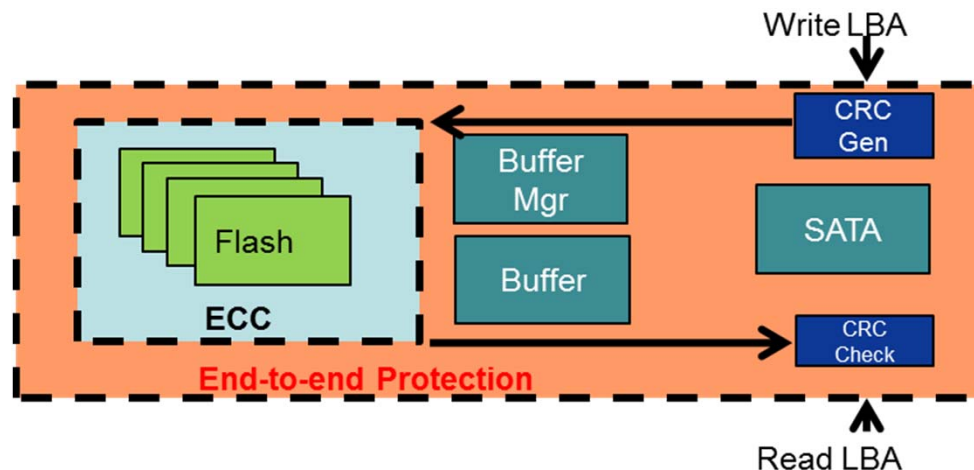
Figure 1 – Data Integrity Field Appended to 512-Byte Standard Block

Source: T10 <http://www.t10.org/ftp/t10/document.03/03-224r0.pdf>

Flash Memory Summit 2011
Santa Clara, CA

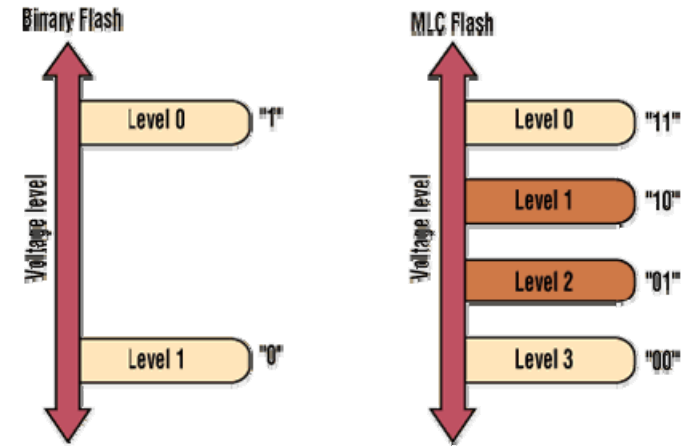
End-to-End CRC

- End-to-End Cyclic Redundancy Check (CRC) must be supported for Enterprise SSDs
 - ▶ Pre-requisite to prevent silent data corruption
 - ▶ Apply and Remove as early as possible
 - ▶ Manage the remainder and handle errors
- A Good CRC solution is LBA seeded
- Data Protection inside the drive
 - ▶ Can protect Flops in the data path from SER and other errors

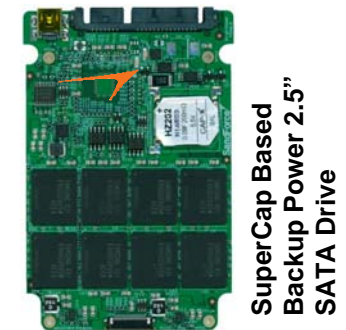
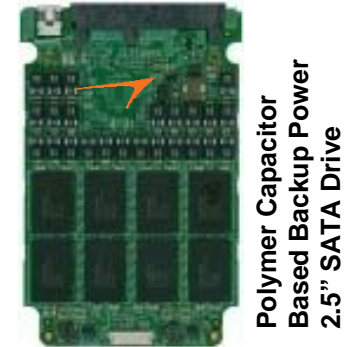


Power Fail Protection

- Guaranteeing data integrity is difficult
 - ▶ MLC much harder than SLC
 - ▶ Lower Page Corruption is a gotcha
 - Getting more complex – more than 1 page can be corrupted
- Absolutely Required in Enterprise applications
 - ▶ Previously written data
 - ▶ Data in flight
- Use backup power to protect against sudden power loss
- Designing for no DRAM simplifies the solution
- Monitor supercap health to ensure capability



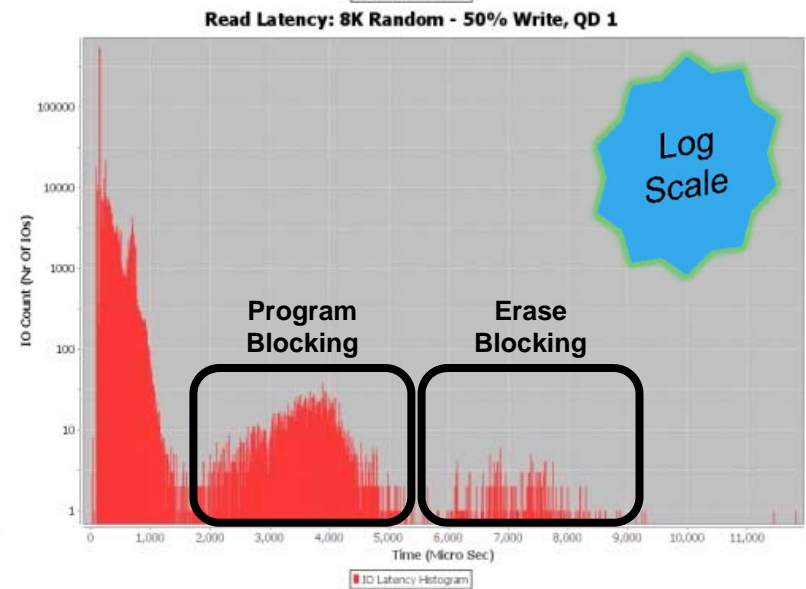
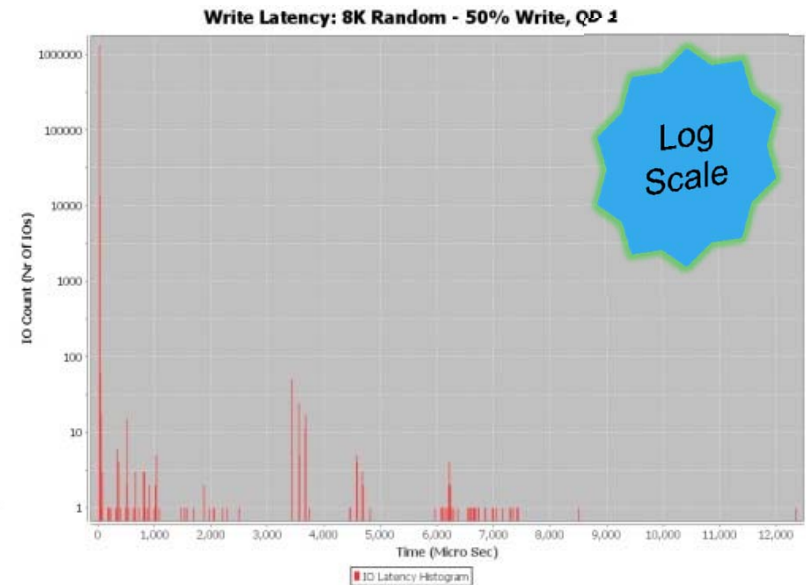
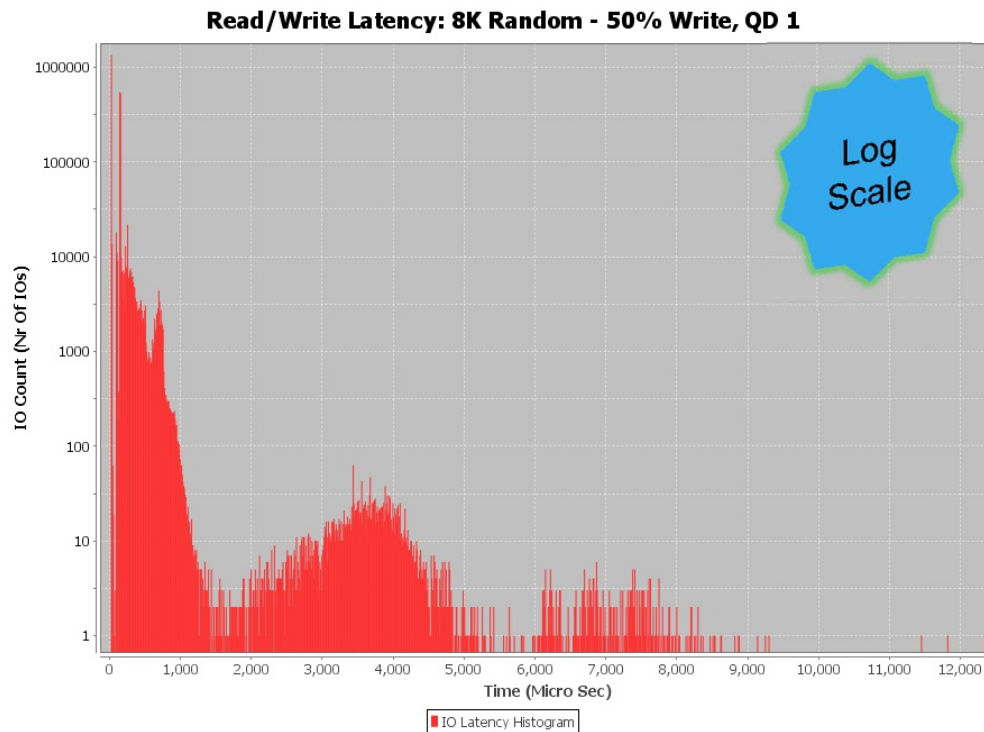
Source: Electronic Design: MLC Challenges Mobile-Entry Barriers





Mixed-I/O Latency Distribution

- Average Latency = OI/O/IOPS (simple)
- The latency distribution is critical for Enterprise QoS
- MLC/3-bit harder to guarantee read latency
 - ▶ Longer program + erase times
 - ▶ More ECC recovery events



Data Retention

- The Arrhenius model is an industry standard for estimating data retention life of floating gate technologies
- Used to derive the acceleration factor between a stress temperature and a use condition
 - ▶ Can be used to de-rate data retention
- Acceleration Factor Equation (AF):

$$AF = e^{\left[\left(\frac{E_a}{k}\right) \times \left(\frac{1}{T_{Use}} - \frac{1}{T_{Stress}}\right)\right]}$$

- ▶ E_a is the intrinsic activation energy (eV)
- ▶ k is Boltzmann's constant
 - 8.617×10^{-5} eV/K
 - $K = -273.16^\circ$ C
- ▶ T_{Use} = use temperature (K)
- ▶ T_{Stress} = stress temperature (K)

Source: Freescale http://www.freescale.com/files/microcontrollers/doc/eng_bulletin/EB618.pdf



Data Retention Continued

- The Acceleration Factor highlights the potential differences between nominal and hot operation
- At 70° Celsius – Retention may be $<1/35^{\text{th}}$ of Retention at 40° Celsius
 - ▶ 1 year becomes 10 days!
- Dynamic Read Scrub acts like a Flash refresh to ensure data retention when power is on
- Temperature aware technology can mitigate temperature and aid in optimizing management algorithms

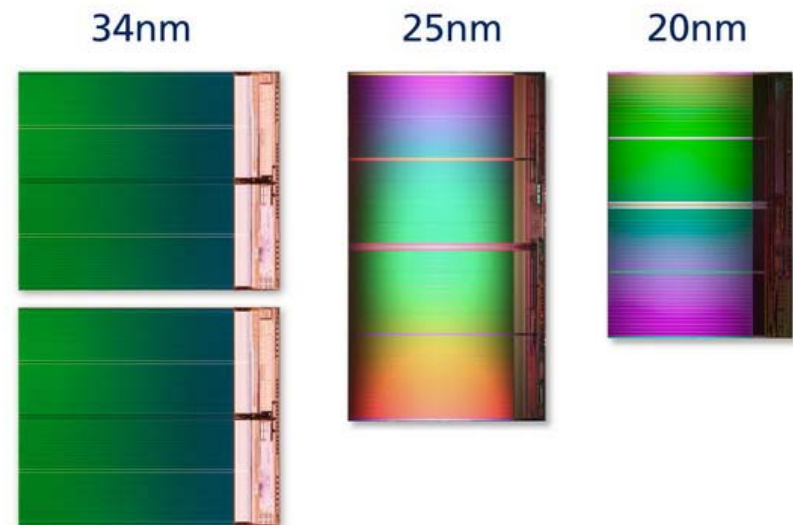


Predictive Failure Analysis

- More intelligent large scale data centers, public and private cloud implementations are changing the classic paradigm
 - ▶ Failures not catastrophic because of architectural data redundancies (country, data center, rack, server, drive)
- Willing to run way past warranty or specification
 - ▶ Must be able to accurately predict drive failure
- Requires diagnostic, statistics and reporting features never capable on HDDs
 - ▶ Up to the second reporting provides users a means to predict a failure
- Trade warranty liabilities for lower TCO and more intelligent usage model

Design for Media Flexibility

- Support for many flash devices is critical
 - ▶ Component availability fluctuates greatly
 - ▶ Early node support means lower cost and longer life!
- Every NAND is different
 - ▶ Makes solutions complex to design and qualify!
 - Page/Block size
 - Page/Block count
 - Spare Area
 - Planes
 - Commands
 - Interfaces
 - Reliability characteristics
 - Multi-LUN support
 - Performance/Response Times
 - Etc. etc. etc.



<http://www.pcper.com/news/Storage/Intel-Micron-jointly-release-20nm-flash-memory>



Visit the SandForce Exhibition Booth

World-class reliability, performance, & power efficiency for enterprise, client, and industrial SSD applications



- Visit us at **booth #407** to see our DuraClass™ technology in action with the latest 24nm MLC flash technology and new non-HDD form factors like the JEDEC MO-300
- Stop by to enter our **free drawings** to win one of many different SandForce Driven SSDs from Corsair, Kingston, OCZ, OWC, Patriot Memory & Viking
 - Free drives given away approximately every 30 minutes!
- See other SandForce Driven™ SSDs in our partners' booths
- Visit the many SandForce Trusted™ SSD ecosystem members

