Designing Enterprise SSDs with Low Cost Media

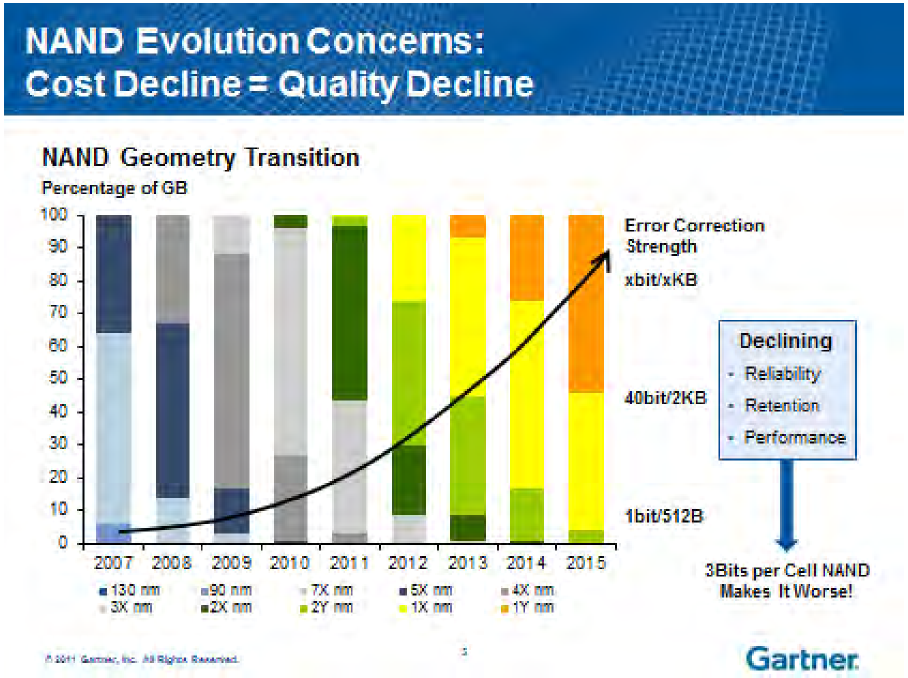
Jeremy Werner

Director of Marketing

SandForce

Everyone Knows…

- Flash is migrating:
  - To smaller nodes
  - 2-bit and 3-bit MLC



NAND Evolution Concerns:
Cost Decline = Quality Decline

Source: Gartner June 2011

- $/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!

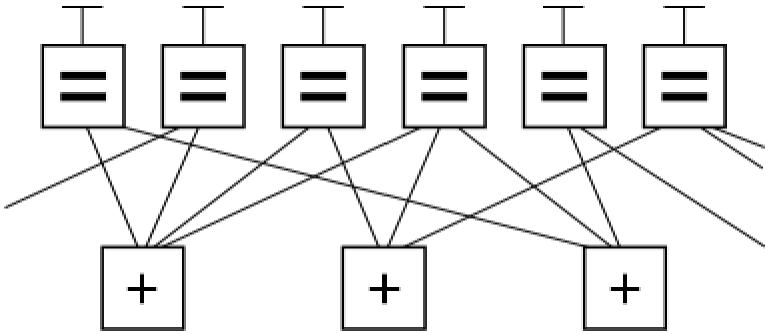
Stepping up to the challenge

- 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

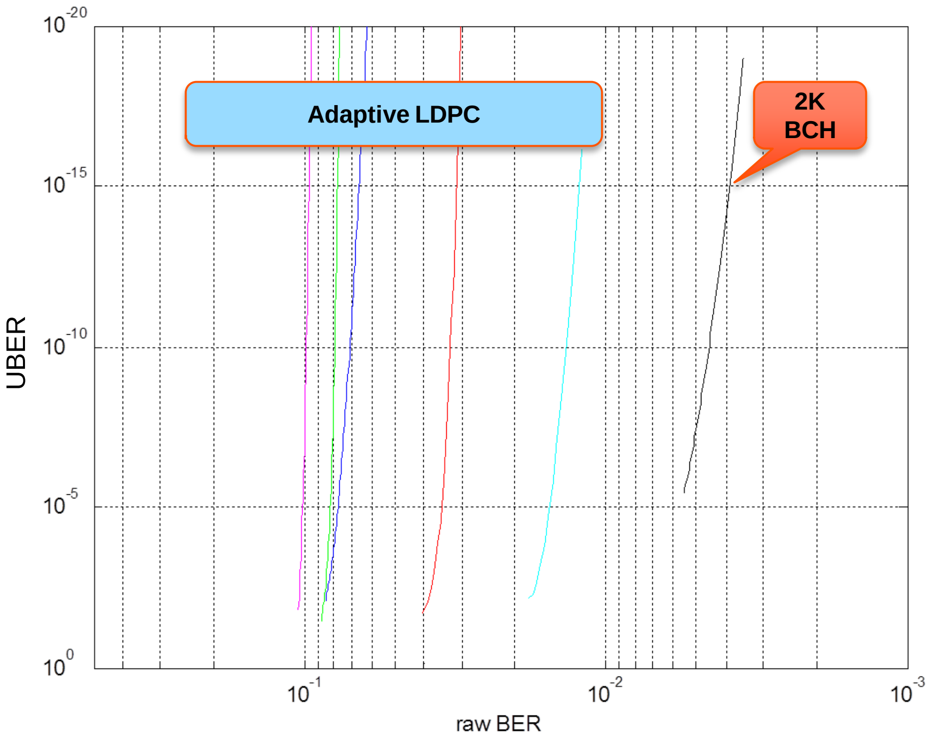


Source: Wikipedia http://en.wikipedia.org/wiki/Ldpc

- 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

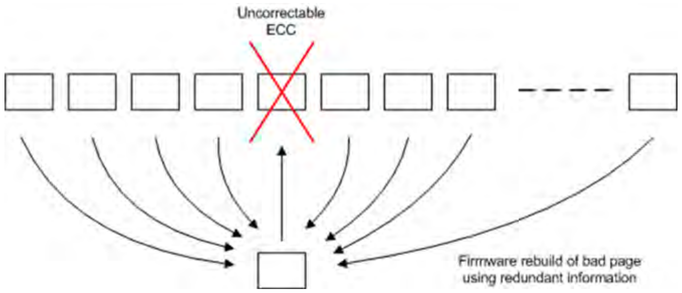
Soft and Hard LDPC and DSP Challenges

- Enables remarkable bit error correction capabilities ~10% RBER but design challenges include:
  - Adequate Error Floor (10^-16 UBER)
  - Efficient Iteration Requirements (1-1.5x)
  - Very high throughput (>2GB/s)
  - Low Latency (<2us)
  - 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



The Soft Error Problem

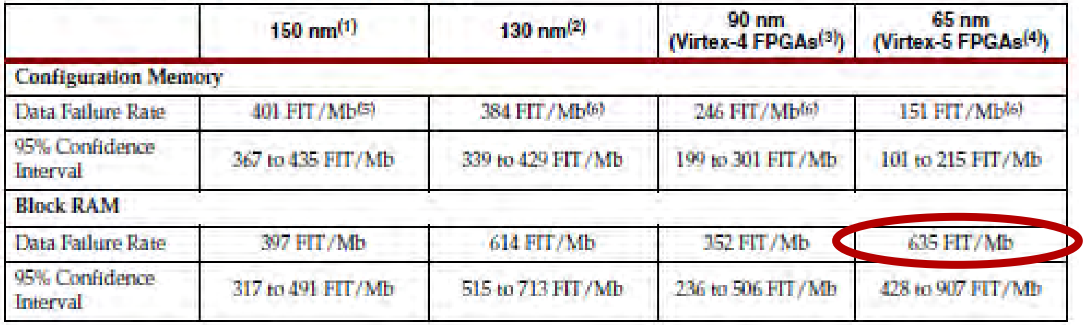
- Soft errors are caused by a charged particle striking a semiconductor memory or a memory-type element
- High-energy cosmic rays and solar particles react with the upper atmosphere generating high-energy protons and neutrons
  - Neutrons they can penetrate most man-made construction (a neutron can easily pass through five feet of concrete).
- This effect varies with both latitude and altitude
  - 14x worse at 10,000 Feet vs. Sea Level
  - 200x worse at 30,000 Feet vs. Sea Level
- Alpha particles are emitted by radioactive isotopes present in IC packaging materials

Source: Actel http://www.actel.com/documents/SER_FAQ.pdf

Soft Error Protection, ASIC protection

- High quality Enterprise SSD Processors have SER protection
  - ECC on memory, CRC on data path for flop protection
  - External DRAM can introduce failure if no ECC…
- FPGAs suffer from configuration memory, Block RAM, and Flip-flop errors

Table 3: **Atmospheric Test Results by Technology**



	150 nm[1]	130 nm[2]	90 nm (Virtex-4 FPGAs[3])	65 nm (Virtex-5 FPGAs[4])
Configuration Memory				
Data Failure Rate	401 FIT/Mb[5]	384 FIT/Mb[6]	246 FIT/Mb[6]	151 FIT/Mb[6]
95% Confidence Interval	367 to 435 FIT/Mb	339 to 429 FIT/Mb	199 to 301 FIT/Mb	101 to 215 FIT/Mb
Block RAM				
Data Failure Rate	397 FIT/Mb	614 FIT/Mb	352 FIT/Mb	635 FIT/Mb
95% Confidence Interval	317 to 491 FIT/Mb	515 to 713 FIT/Mb	236 to 506 FIT/Mb	428 to 907 FIT/Mb

1000 FIT = 0.88% (AFR) Annual Failure Rate

- Up to 16.4 Mbits of Block RAM on a Virtex-5 FPGA

Source: Xilinx http://www.xilinx.com/support/documentation/white_papers/wp286.pdf
http://www.xilinx.com/support/documentation/data_sheets/ds100.pdf

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

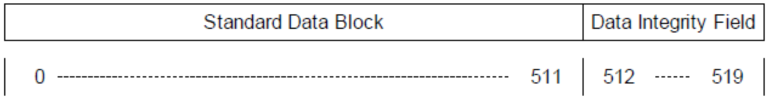
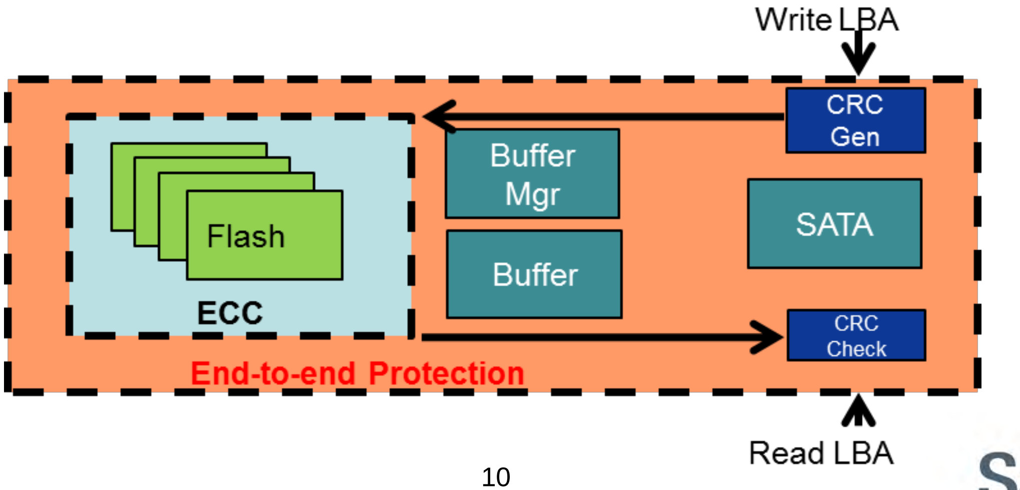


Figure 1 – Data Integrity Field Appended to 512-Byte Standard Block

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

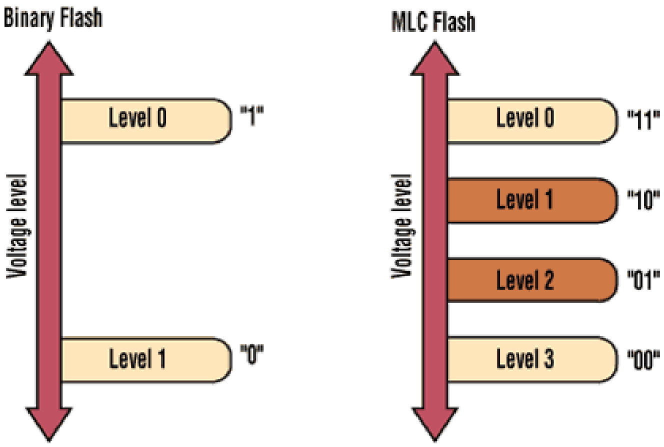
End-to-End CRC

- 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 protects Flops in the data path from SER and other errors

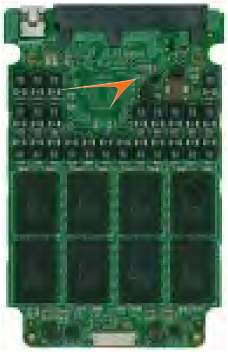


Power Fail Protection

- 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



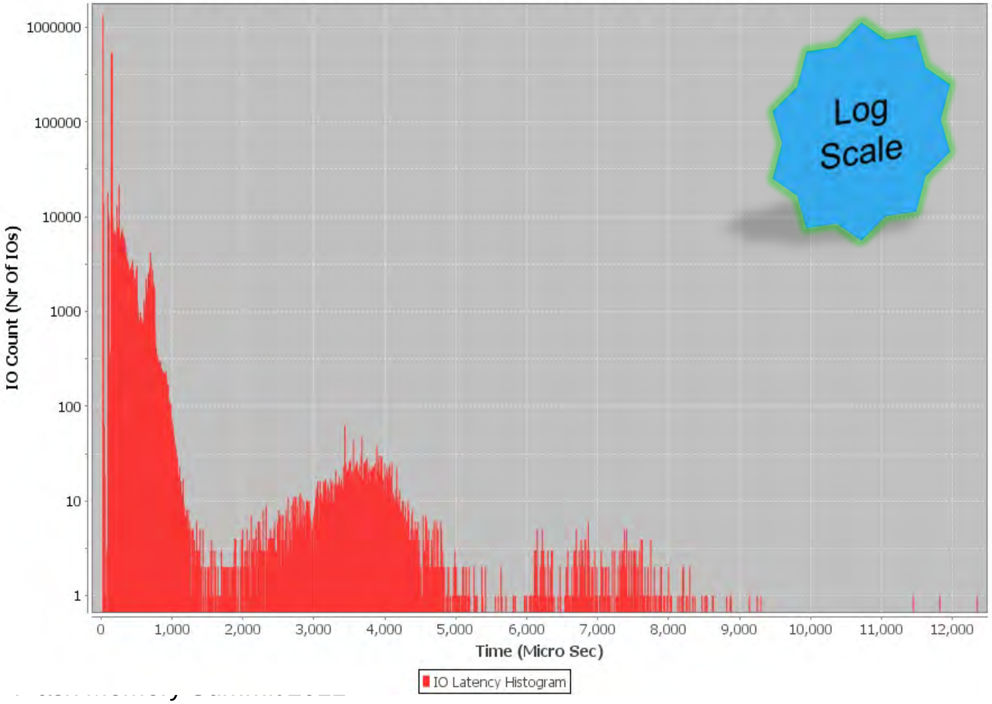
Polymer Capacitor Based Backup Power 2.5" SATA Drive

SuperCap Based Backup Power 2.5" SATA Drive

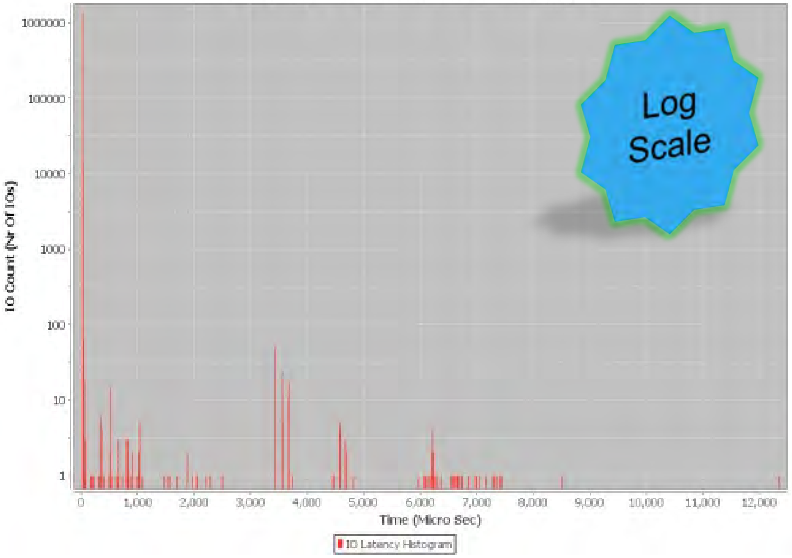
Mixed-I/O Latency Distribution

- Average Latency = OIO/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

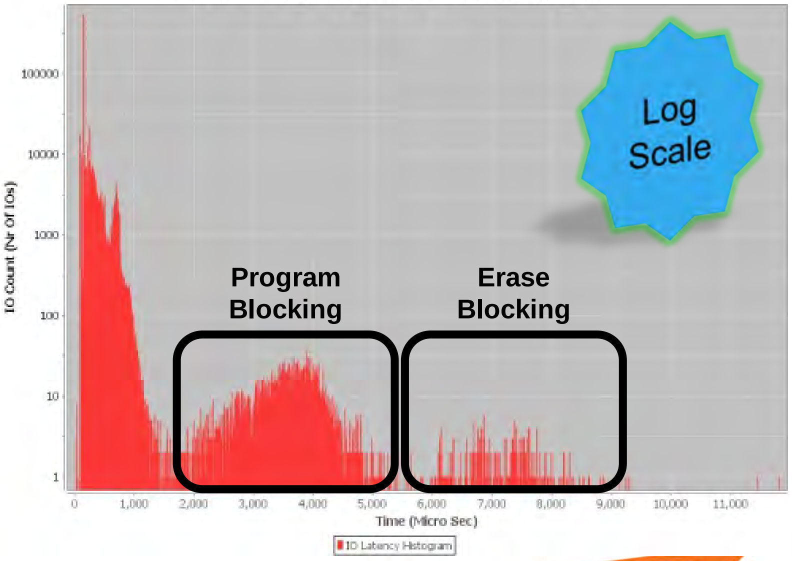
Read/Write Latency: 8K Random - 50% Write, QD 1



Write Latency: 8K Random - 50% Write, QD 1



Read Latency: 8K Random - 50% Write, QD 1



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):

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

  - E_a is the intrinsic activation energy (eV)
  - k is Boltzmanns' constant
    - 8.617 x 10 –5 eV/K
    - K = –273.16° 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

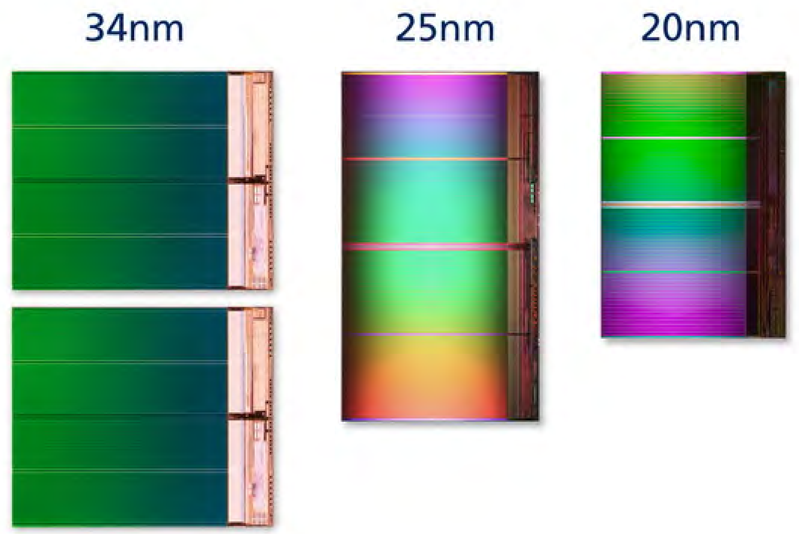
- The Acceleration Factor highlights the potential differences between nominal and hot operation
- At 70°Celsius – Retention may be <1/35th of Retention at 40°Celsius
  - 1 year becomes 10 days!
- Dynamic Read Scrub acts like a Flash refresh to ensure data retention when power is on
- Temperature aware 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, Kingspec, 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