A Look Under the Hood at Some Unique SSD Features

Jeremy Werner
Director, Marketing
SandForce, Inc.

Overview

- Firmware Design must consider key SSD attributes
  - Performance
  - Reliability
  - Endurance
  - Power Consumption
  - Security and Integrity
  - Flexibility

Native Command Queuing (NCQ)

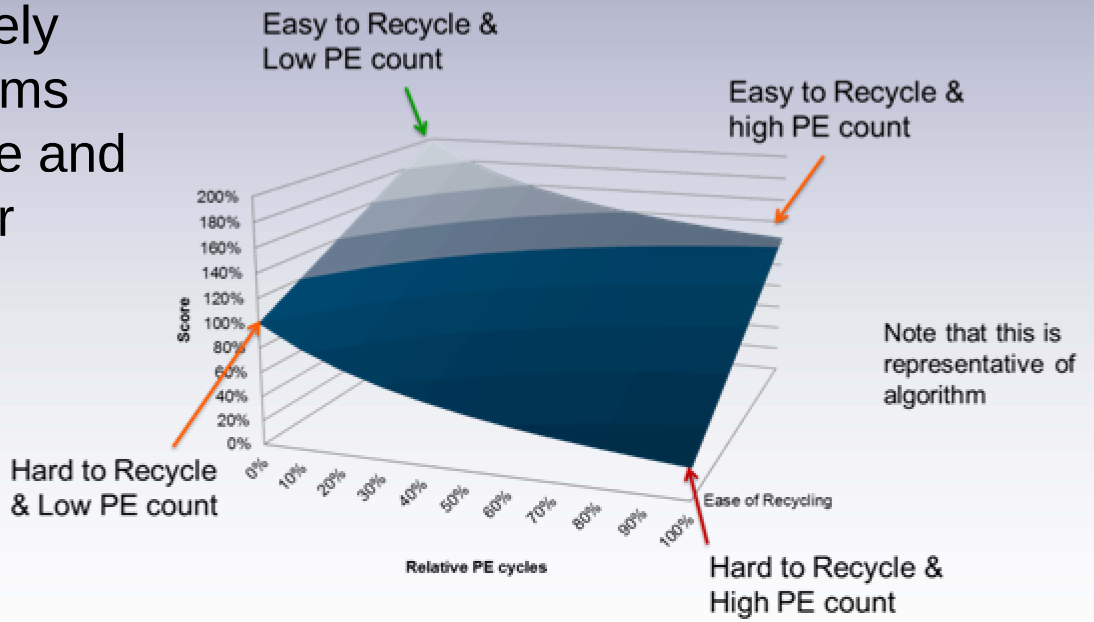
- SATA NCQ supports sending up to 32 commands to SSD prior to completion
- Supporting NCQ can significantly improve random read and write performance
- Firmware must be able mange the queue
- Service commands strategically out of order to achieve greatest performance
- Queued commands complicate error handling

Make Flash your Bottleneck

- Flash parallelism is a major factor in SSD performance
- Multi-plane – allows programming more than one Flash plane at a time
- Multi-LUN – allows programming/reading from more than one die per CE
- Goal – Keep all dies active at all times

Recycling/Garbage Collection

- The process of moving data in flash to create empty blocks to write to
- Steals Flash bandwidth from the host
- Efficient Recycling is a MUST
- Support TRIM to reduce the amount of data which must be moved
- Choose your blocks wisely
  - Block picking algorithms must factor free space and factor cycle counts for Wear leveling

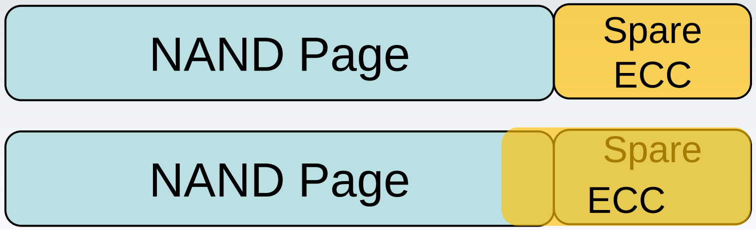


Enhancing the Media

- Low cost (MLC/TLC) is optimized for consumer applications (USB, card, MP3)
  - Just enough Spare area per page to meet these requirements
- SSDs need more ECC for lower UBER (10^{-17})
  - Up to 80bit/1KByte for 20nm class NAND

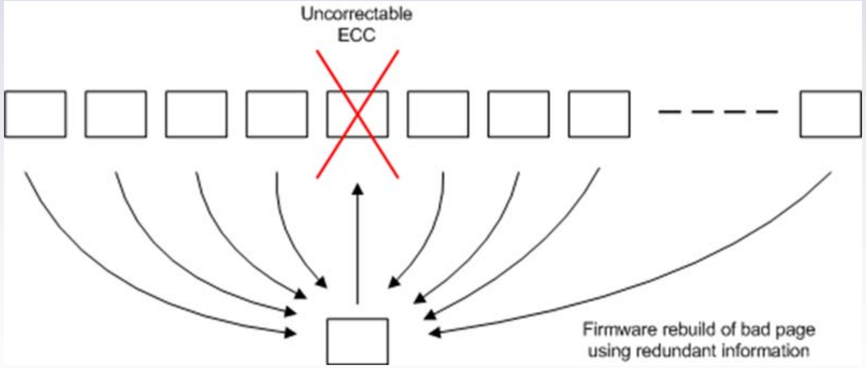
Conventional Error Correction:
Stores ECC in spare field

Adaptive Error Correction:
Stores ECC in spare field and uses some of the NAND page



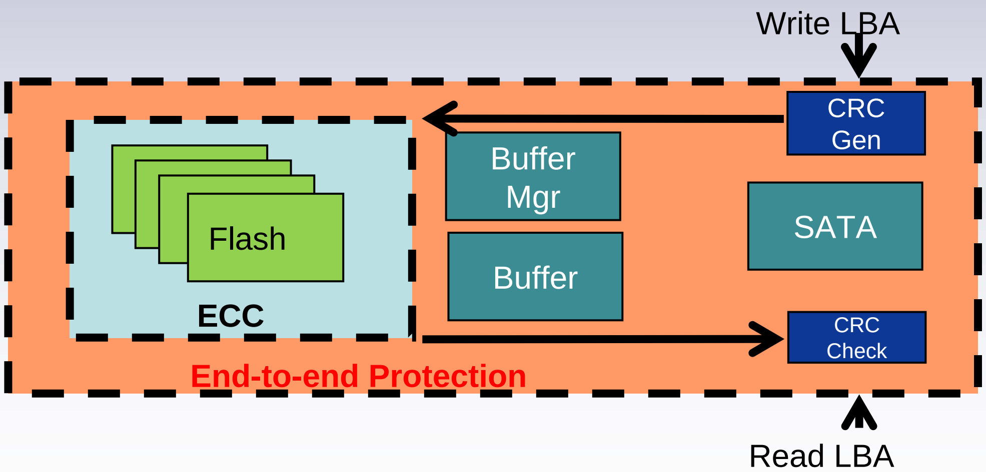
Advanced Error Recovery

- Advanced Error Recovery Techniques are needed
- Ability to recover completely lost sectors, pages, blocks
  - SandForce supports RAISE™
- Handle error recovery and notification
  - Rewrite recovered data
- May need to overprovision to enable extra reliability



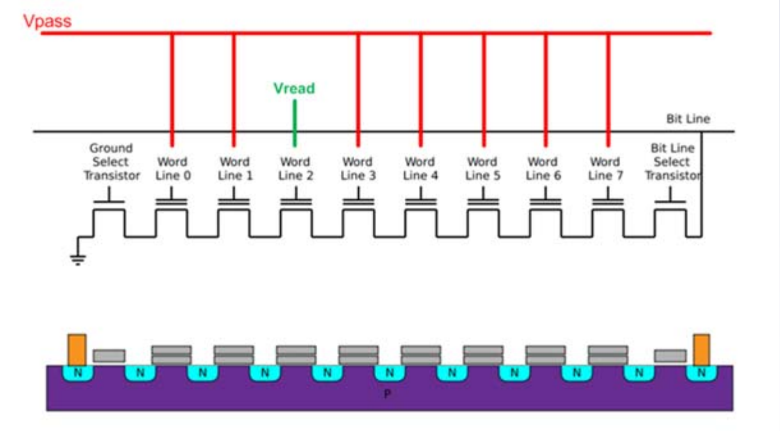
End-to-End Protection

- End-to-End Cyclic Redundancy Check (CRC) must be supported for Enterprise SSDs
  - Manage the remainder
  - Error handling



Read Disturb Management

- Read Disturb is a growing source of flash errors
- Must track reads to blocks <u>efficiently</u>
- Must move data when a block is read too often
- Beginning of life and End of life behaviors differ greatly
- 100K → 10K → ? Reads
- Page and Block sizes increasing
  - More Data to move means increased WA and overhead



Designing FW for Endurance

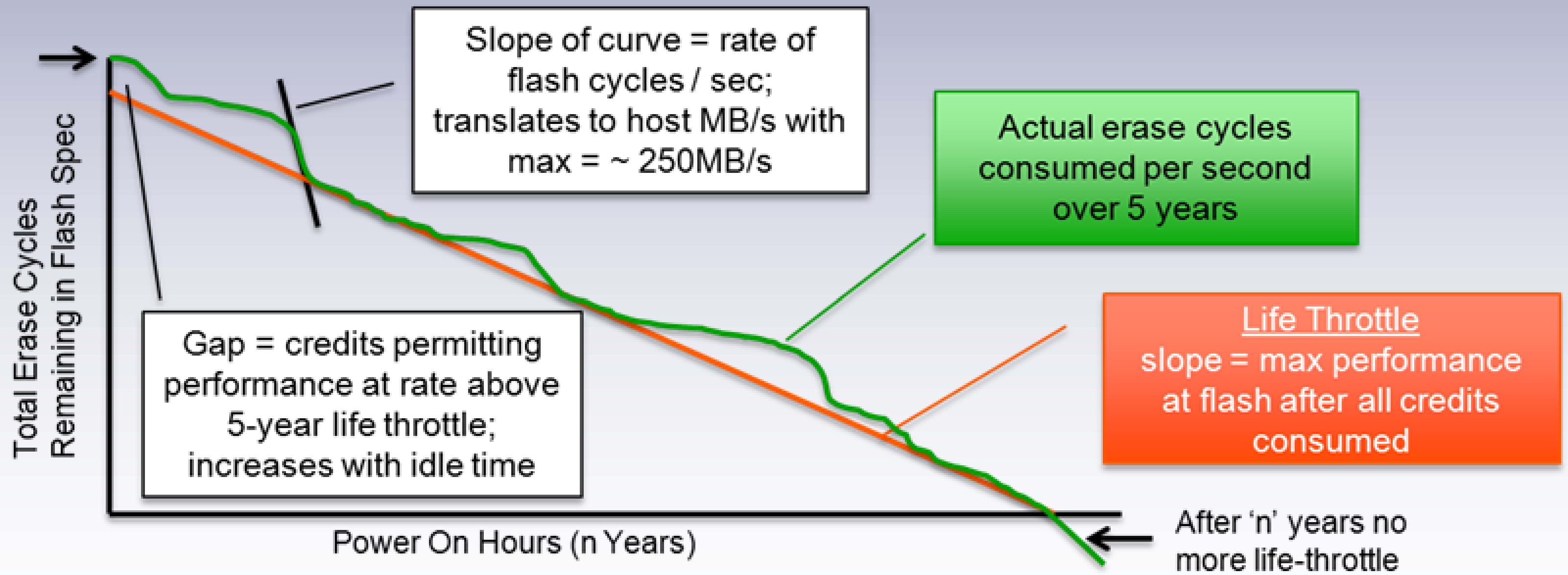
- Write Amplification can extend Flash life or kill it!
  - Page Based vs. Block Based Volume Manager
  - DuraWrite™ SandForce Write Reduction Technology
  - TRIM
  - Background Garbage Collection

$$\frac{\text{Flash Write}}{\text{Host Write}} = \text{Write Amplification}$$

$$\frac{\text{Flash Endurance} * \text{Capacity} * \text{WLE}}{\text{Write Speed} * \text{D/C} * \text{Write \%} * \text{Write Amp}} = \text{SSD Total Life}$$

Life Curve Throttling

- Some users may want to guarantee flash usage will meet a calendar duration
  - FW can limit writes if drive is used too heavily



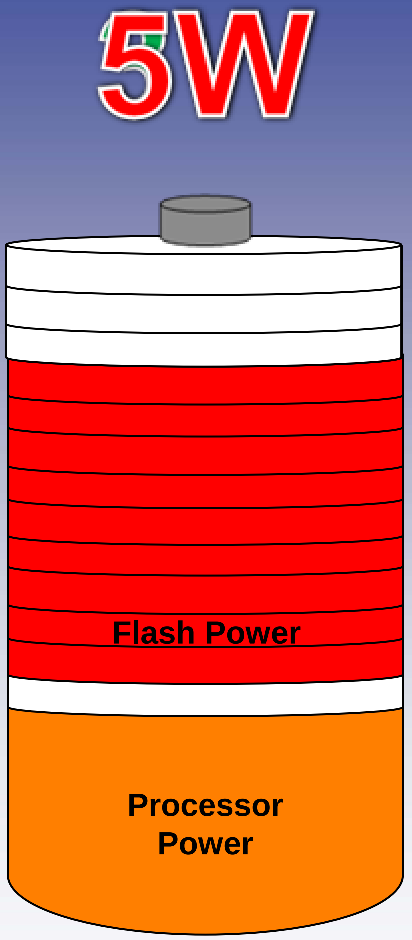
Designing FW for Power Consumption

- SSDs must fit within their power envelope
- Designing an SSD without DRAM helps but complicates FW
  - Must meet all requirements with limited volatile memory
  - Means Mapping algorithms must be outstanding
- 2 possible limiting factors
  - Peak Power (3W-25W)
  - Thermal Dissipation (Varies Greatly)

Basic Power Management

- State based power Management Techniques
  - Serial ATA Power Management
    - Active, Idle, Standby, Sleep
    - PUIS – Power Up in Standby
  - PHY Power Management
- Generally these:
  - reduce power consumption
  - increase wake-up latency

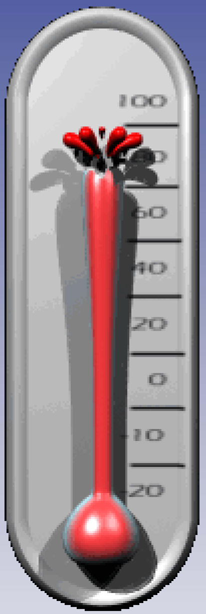
Controlling Peak Power



- Two main contributors to power consumption: Flash and Processor
- To control max power limit the number of active Flash die
- Different applications have different requirements
  - SATA – 2 to 5W typical
  - SAS - 9W typical
  - PCIe - 25W typical
  - Give the power to the customer!

Temperature Management

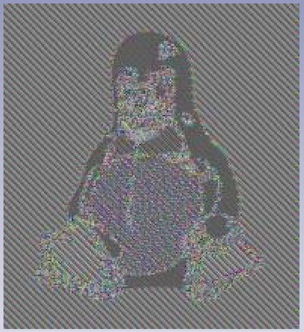
- Monitor surface temperature with onboard sensors
- Allow maximum performance in typical conditions
- Manipulate behavior under extreme temperatures



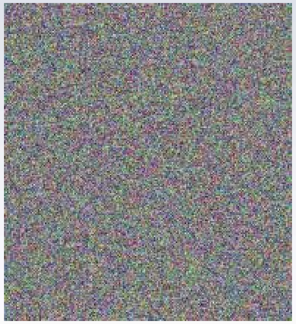
Secure User's Data

- Use encryption to ensure confidentiality
  - Not all encryption methods are created equally
  - SW often uses Electronic Codebook (ECB) mode
  - SW encryption has a high cycle overhead vs. HW
  - SW must manage keys that HW uses securely
- Destroy User Data
  - Secure Erase
  - Sanitize
- Multiple Standards to Implement
  - ATA Security, TCG Opal/Enterprise, IEEE 1667
  - TCG Enterprise Requires multiple user bands
- Industry wide collaboration is critical
  - Ecosystem must work together

Original



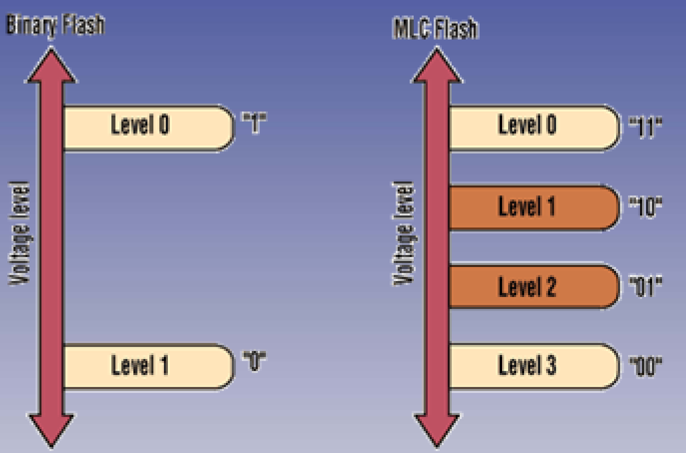
Encrypted using ECB mode



Other modes than ECB results in pseudo-randomness

Source: Wikipedia

Protect User's Data



Source: Electronic Design: MLC Challenges Mobile-Entry Barriers

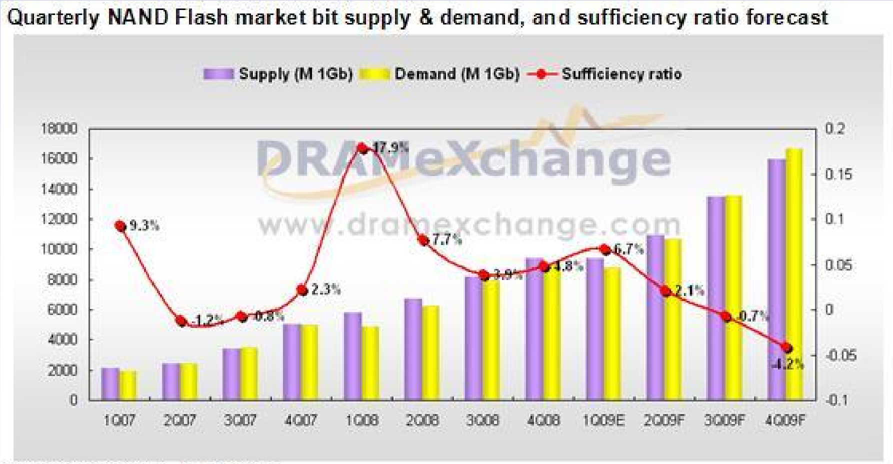
- Guaranteeing data integrity is difficult
  - MLC much harder than SLC
  - Lower Page Corruption is little known issue
- Absolutely Required in Enterprise applications
  - Previously written data
  - Data in flight
- Use supercap to protect against sudden power loss
  - Designing for no DRAM is an important element
- Monitor supercap health to ensure capability

Mapping Strategies for Integrity and Availability

1. Full Map resident in Flash – Best Approach
   - Must end there eventually
   - Easier power failure recovery techniques
   - Very Complex Algorithms
2. Cache full map in DRAM
   - More costly and power intensive
   - Difficult to power fail
   - ECC on DRAM components is not typical
   - Easier lookup and algorithm design
3. Cache partial map in DRAM
   - Similar to full map with better cost and power but more complex algorithms
4. Onload to Host CPU and DRAM
   - Custom Drivers, every OS and even HW is unique driver
   - Resource intensive on host
   - Very long recovery times after power failure

Design for Media Flexibility

- Support for many flash devices is critical
  - Component availability fluctuates greatly
- Every NAND is different
  - Makes software 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.

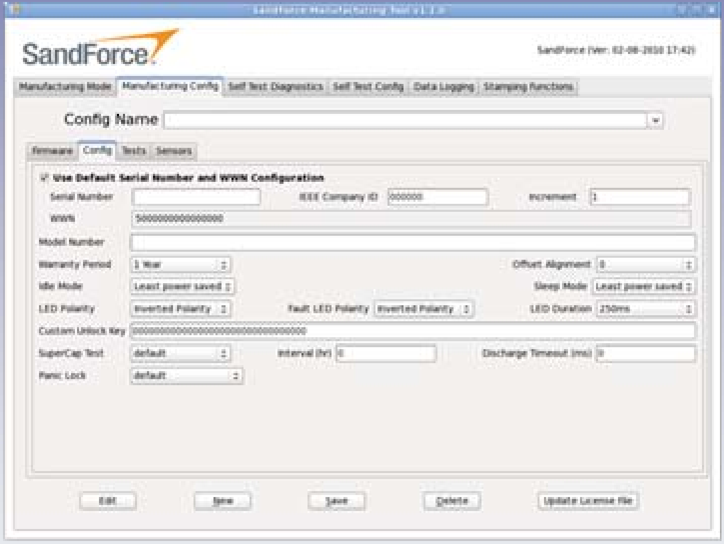


資料來源：集邦科技，2009年 1 月

Source: DRAMeXchange, Jan. 2009.

Design for Manufacturing Flexibility

- SSD vendors differentiate by optimizing for different markets
- Achieved through Manufacturing time customizations
  - Ideal settings vary across applications
- Firmware must be flexible to provide support



Wrap Up

Putting it all together is what makes it great!

SSD Related Standards Bodies

- JEDEC – Flash standards, Form Factors, Rel. + End.
- ONFi (Open NAND Flash Interface) – Flash standards
- Serial ATA – SATA Standards
- PCI SIG – PCIe Standards
- SNIA – SSD Performance Testing

Contact Andy Tomlin – atomlin@SandForce.com

Contact Jeremy Werner – jwerner@SandForce.com

SandForce SSD Processors

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 30nm MLC flash technology
  - MO297 – Ultra compact high-density industrial standard SATA form factor with up to 30K read/write IOPS* per watt
  - PCIe – Industry leading IOPS* for read, write, and mixed workloads
  - SAS – 25K sustained write IOPS* through Emulex's high-performance SAS-SATA embedded storage bridge technology
  - SATA – Updates on SSDs, testing, and infrastructure equipment collaboration
- Stop by to enter our **free drawing** for one of three Corsair Force 120GB SandForce Driven SSDs
- See other SandForce Driven SSDs in our partners' booths as well

***Random 4K transfers**