Contents

1. Document Purpose............................................................................................................................ 6
2. PLI (Power Loss Imminent)................................................................................................................ 7
3. NCQ (Native Command Queuing) ..................................................................................................... 8
4. AES Encryption .................................................................................................................................. 9
5. End-to-End Data Protection ............................................................................................................. 10
6. XOR (Exclusive “OR”) ....................................................................................................................... 11
7. HET® (High Endurance Technology) ................................................................................................. 12
8. OPAL* .............................................................................................................................................. 13
9. Trim ................................................................................................................................................. 14
10. Power Safe Write Cache .................................................................................................................. 15
11. DIPM and HIPM (Device and Host Power Management) ............................................................... 16
12. Write Amplification ......................................................................................................................... 17
13. RAS (Reliability, Availability, & Serviceability) ................................................................................. 18

Figures

Figure 1: PLI Functional Description Block Diagram............................................................................... 7
Figure 2: AES Security Diagram ........................................................................................................... 9
Figure 3: “End-to-End” Data Flow Diagram .......................................................................................... 10
Figure 4: ETE Appended Protection Information ................................................................................... 10
Figure 5: Write with “XOR” Operation Illustration .............................................................................. 11
Figure 6: HET (High Endurance Technology) Diagram ........................................................................ 12
Figure 7: SSD Trim Operation Flow Diagram ....................................................................................... 14
Figure 8: Energy Supply Capacitors for Power Safe Write Cache on Intel® Solid-State Drive 710 Series 15
Figure 9: SATA Power Management States ......................................................................................... 16
Figure 10: Write Amplification Formula ............................................................................................... 17
## Revision History

<table>
<thead>
<tr>
<th>Document Number</th>
<th>Revision Number</th>
<th>Description</th>
<th>Revision Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>327756</td>
<td>000-US</td>
<td>Initial Release</td>
<td>August 2012</td>
</tr>
</tbody>
</table>

§
## Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOS</td>
<td>Basic Input / Output System</td>
</tr>
<tr>
<td>Chipset</td>
<td>A term used to define a collection of integrated components required to make a PC function.</td>
</tr>
<tr>
<td>DOS</td>
<td>Disk Operating System</td>
</tr>
<tr>
<td>HBA</td>
<td>Host Bus Adapter</td>
</tr>
<tr>
<td>HDD</td>
<td>Hard Disk Drive</td>
</tr>
<tr>
<td>I/F</td>
<td>Interface</td>
</tr>
<tr>
<td>LBA</td>
<td>Logical Block Address</td>
</tr>
<tr>
<td>NAND</td>
<td>Negated “AND” – A NVM Flash Memory Architecture</td>
</tr>
<tr>
<td>NVM</td>
<td>Non-Volatile Memory</td>
</tr>
<tr>
<td>OEM</td>
<td>Original Equipment Manufacturer</td>
</tr>
<tr>
<td>ONFI</td>
<td>Open NAND Flash Interface</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System</td>
</tr>
<tr>
<td>RAID</td>
<td>Redundant Array of Independent Disks</td>
</tr>
<tr>
<td>ROM</td>
<td>Read Only Memory</td>
</tr>
<tr>
<td>SATA</td>
<td>Serial Advanced Technology Attachment</td>
</tr>
<tr>
<td>SSD</td>
<td>Solid-StateDrive</td>
</tr>
<tr>
<td>UI</td>
<td>User Interface</td>
</tr>
</tbody>
</table>
1. Document Purpose

The purpose of this document is to provide interested readers explanations of many Intel and industry Solid-StateDrive technology functions, features, and acronyms.

The document is structured to limit the explanations to one per each SSD technology item. For each item, overview and context is provided along with the “short” and “long” explanations suited to the user’s need and interest. Feel free to pull individual topic sheets out as needed.
2. **PLI (Power Loss Imminent)**

**Overview and Context**

Worried about data loss during unplanned power shutdowns or inadvertent drive removal in data centers? Data safety features in SSDs can prepare for unexpected power-loss and protect system and user data.

**Definition and Explanation**

**Short:**
PLI is a hardware and firmware feature on SSDs that provides enough stored energy for the SSD to safely store user and system data in temporary buffers to the non-volatile NAND flash storage during an unexpected loss of power. Not all SSDs have the PLI feature, but the Intel® SSD 320 Series and Intel® SSD 710 Series do. PLI is sometimes referred to as “power safe write cache” (also included in this document; see section 10).

**Long:**
During a “clean” shutdown, most host systems initiate a command (the STANDBY IMMEDIATE command) to an SSD to give the SSD enough time to prepare for the shutdown. This allows the SSD to save data currently in transition (in temporary buffers) to the non-volatile NAND media.

However, during an unsafe power shutdown or a loss of power, the SSD abruptly loses power before the host system can initiate the ATA STANDBY IMMEDIATE command. This prevents data in the temporary buffers from being saved in the non-volatile NAND.

The Intel SSD 320 Series and Intel SSD 710 Series contain hardware and firmware-based power-loss data protection features. These SSDs includes a power-fail detection circuit, which sends a signal to the ASIC controller in the SSD indicating there is an imminent drop in power level. The SSD then relies on its on-board power-loss protection capacitors to provide enough energy for the SSD firmware to safely move user and system data from the transfer buffer and other temporary buffers to the NAND.

![Figure 1: PLI Functional Description Block Diagram](image-url)
3. **NCQ (Native Command Queuing)**

**Overview and Context**

Native Command Queuing (NCQ) was originally a technology designed to increase performance of SATA hard disks. It does this by allowing a hard disk to internally optimize the order in which it executes the read and write requests it received from the host. For hard drives, this can reduce the amount of unnecessary drive head movement and resulting in better performance, in particular for workloads where multiple read/write requests are outstanding at the same time. This situation most often occurs in server-type applications. Since NCQ helps hard disk drives performance by optimizing the internal order of execution, NCQ will further the performance of solid-state drive even more so.

**Definition and Explanation**

**Short:**

Native Command Queuing (NCQ) is a technology designed to increase performance of SATA hard disk and solid-state drives by allowing them to internally optimize the order in which it executes read and write requests received from the host. All Intel solid-state drives support the Native Command Queuing (NCQ) command set, which includes: “READ FPDMA QUEUED” and “WRITE FPDMA QUEUED”. Maximum Queue Depth is 32.

**Long:**

Native Command Queuing (NCQ) was originally a technology designed to increase performance of SATA hard disks. It allows a hard disk to internally optimize the order in which it executes read and write requests received from the host. Since NCQ helps hard disk drives performance by optimizing the internal order of execution, NCQ will further the performance of solid-state drive even more so.

For NCQ to be enabled, it must be supported and enabled in the SATA host bus adapter and in the hard drive or SSD. Additionally, the appropriate device driver must be loaded into the operating system to enable NCQ. Many newer chipsets support the Advanced Host Controller Interface (AHCI), which should allow a generic driver supplied by the operating system to control them and enable NCQ. Newer mainstream Linux kernels support AHCI natively.
4. AES Encryption

Overview and Context

Encryption of data in storage devices such as PC hard drives, USB keys, or SD removable cards is increasingly important to protect sensitive personal or proprietary company information. Encryption protects the data from being read without a decryption key if the device is lost or stolen. Advanced Encryption Standard (AES) is one of many specifications for the encryption of electronic data and it has been adopted by the U.S. government and widely used in industry.

Definition and Explanation

Short:
Advanced Encryption Standard (AES) is a specification for the encryption of electronic data. It has been adopted by the U.S. government and is now used worldwide. The algorithm described by AES is a symmetric-key algorithm, meaning the same key is used for both encrypting and decrypting the data. AES is the first publicly accessible and open cipher approved by the National Security Agency (NSA) for top secret information when used in an NSA approved cryptographic module.

Long:
The Advanced Encryption Standard (AES) is the United States Government’s Federal Information Processing Standard for symmetric encryption, defined by FIPS Publication #197. AES is a block cipher that encrypts a 128-bit block (plaintext) to a 128-bit block (ciphertext), or decrypts a 128-bit block (ciphertext) to a 128-bit block (plaintext). AES uses a key (cipher key) whose length can be 128, 192, or 256 bits. AES-128, AES-192, and AES-256 process the data block in, respectively, 10, 12, or 14 iterations of pre-defined sequences of transformations, which are also called AES rounds. The rounds are identical except for the last one, which differs slightly from the others by skipping one of the transformations.

Steps to Enable AES Encryption on an SSD:

1. Set ATA password (also known as the drive password) in the BIOS to enable the drive AES-128 /AES-256 encryption (The ATA password is stored on the drive as a non-reversible hash, and is used to encrypt the encryption keys on the drive)
2. Perform a secure erase on the SSD if a new password is desired or needed
3. Set your new password

Figure 2: AES Security Diagram

The following Intel SSDs support 128-bit AES encryption:

Intel® SSD 320 Series, Intel® SSD 330 Series, Intel® SSD 520 Series and Intel® SSD 710 Series

References:
5. **End-to-End Data Protection**

**Overview and Context**

Data integrity is extremely important, especially in data center storage environments where there are many layers of hardware devices and software that the data must traverse.

![Figure 3: “End-to-End” Data Flow Diagram](image)

Chances for user data to get corrupted as it is passed from device to device are possible and problematic. Therefore, there are mechanisms and algorithmic schemes that ensure that saved and retrieved data is indeed correct without error from the beginning end to the furthest end, where the data is stored, thus the name “end to end” data protection.

**Definition and Explanation**

**Short:**

End-to-end Data Protection is a feature for Solid-State Drives and hard drives that extend error detection to cover the entire path from the computer system to the hard drive media and back. Data protection information is appended to the data in the computer system. It stays with the data from the computer, through connections, through RAID controllers, HBAs, and through drive electronics to the storage device. When read, the same data protection information returns with the data to the computer system. The protection information is used to verify the correctness.

**Long:**

End-to-End Data Protection is a feature for storage hard drives that extends error detection to cover the entire path from the computer system to the hard drive media and back. Data protection information is appended to the data in the computer system, and stays with the data from the computer, through connections, through RAID controllers, and through drive electronics to the storage device. The appended “end to end” data integrity field (DIF) is 8 Bytes that is composed of CRC, App Tag, and Ref Tag (See Figure 4).

![Figure 4: ETE Appended Protection Information](image)

**References:**

6. XOR (Exclusive “OR”)

Overview and Context

Striving to improve the reliability of SSD is always of paramount value at Intel. Improving the reliability of the NAND flash components is critical to ensure that user data is preserved and performance is sustained for the life of product. Many schemes exist to improve the reliability of the NAND components; Intel has chosen XOR reliability enhancement. The “XOR (Exclusive “OR”)” reliability enhancement operation is named after the Boolean logical function (See logical gate and truth table), and allows ability to replace up to an entire defective NAND die in a component through XOR operations.

Definition and Explanation

Short:

XOR significantly improves the NAND component reliability by providing protection against die failure and extrinsic UBER (Uncorrectable Bit Error Rate) events by replacing entire or defective portions of a NAND die array utilizing spare NAND die array that can be built with reconstructed data via the XOR parity data bits. This scheme would fail in the event that there are 2 or more die failures simultaneously, or if the SSD runs out of spare XOR NAND die space. The detailed “XOR” function will be explained in the long explanation below.

Long:

XOR’s primary goal in an SSD is to REDUCE any catastrophic data loss failure and AFR (average failure rate). This is accomplished by using the Boolean XOR logical function that A XOR B XOR B = A which becomes the underlying principle on rebuilding data. The XOR-ing of incoming data packets from the host generates XOR parity bits that later can be used to rebuild data if needed. This ability to rebuild or replace bad NAND bytes with spare good NAND bytes significantly enhances the SSD’s reliability while providing consistent performance. The illustration below shows the XOR operation with some example packets of data.

![Figure 5: Write with “XOR” Operation Illustration](image)

Since every SSD has some spare NAND area, the XOR rebuilds—if needed—uses this spare area.
7. **HET® (High Endurance Technology)**

**Overview and Context**

Improving the reliability of the core storage elements of an SSD—its NAND flash components—is critical to ensure that user data is retained along with optimum performance for the life of product. Intel has developed technologies that combine the enhancements of NAND wafer/die sorting, component testing, and software at the SSD system level that provide a very effective solution to the overall reliability of the SSD.

**Definition and Explanation**

**Short:**

HET® (High Endurance Technology) is Intel proprietary technology that combining NAND silicon enhancements and SSD NAND management techniques to extend the write endurance of MLC-NAND-based SSDs.

**Long:**

Intel developed “HET,” High Endurance Technology, which comprises enhancements at both the NAND die/wafer and component testing levels with the SSD firmware enhancements to improve the endurance and reliability of its NAND components and the SSD product.

At the NAND component level, the NAND is optimized with fine-tuned read, program and erase voltages and extensively tested at temperature and voltage skews to ensure the highest-binned products.

In the SSD, firmware enhancements are implemented to minimize bit errors and to reduce write amplification. Innovative and efficient bundling of writes to minimize excessive background data manipulation and management was created to reduce write amplification. In the real world, SSDs with HET technology from Intel will provide reliable performance far beyond the expected lifetimes of standard MLC-NAND-flash-based devices.

![Figure 6: HET (High Endurance Technology) Diagram](image)

**Figure 6: HET (High Endurance Technology) Diagram**
8. OPAL*

Overview and Context

Secure protection of data on storage devices such as solid state drives and hard disk drives has become very important given the broadly available user connection points such as WiFi, Broadband, LTE, etc. that dramatically increases vulnerability to malware and virus entering our devices.

The Opal* (named after the precious gem, rather than an acronym) Storage Specification originated from the Trusted Computing Group* (TCG: http://www.trustedcomputinggroup.org/). TCG is an international organization involved in setting security standards for computing environments.

Definition and Explanation

Short:
TCG’s Opal SSC (Security Subsystem Class) specification has been adopted as an international standard with the goal of creating more secure IT environments. When used in combination with supporting application software, SSDs and HDDs that support Opal SSC will enable sophisticated security solutions for a wide range of computing platforms, including notebook PCs.

Long:
When Opal is used in combination with application software supporting Opal SSC, it will enable advanced security features such as pre-boot authentication and secure partition.

Pre-boot authentication performs user authentication when starting up the computer. The Opal SSC standard will allow for the use of advanced authentication techniques that include biometric authentication or smart-card authentication, even before starting up Windows or other operating systems that rely on traditional keyboard password entry access.

Secure partition is a technology enabling SSD or HDD storage to be partitioned into a number of secure storage regions—each with its own encryption key—giving access to only the valid owner of partition and users given access by the owner. This allows, for example, for content and its licensing data to be stored and protected in separate partitions, so that different partitions can be used depending on the nature of the data stored, enabling more secure data management.

There are many independent software vendors to implement these functions, including McAfee*, SECUDE*, Wave Systems*, and WinMagic*, that will work in conjunction with the Opal SSC. Many vendors also provide their own consoles to manage these features.

Other features that Opal enables with independent software suppliers are:

- Security Provider Support
- Interface Communication Protocol
- Cryptographic Features
- Authentication
- Table Management
- Access Control & Personalization
- Issuance
- SSC Discovery
9. **Trim**

**Overview and Context**

Erasing data is different between solid-state drives and hard drives. In hard drives, existing data can simply be overwritten. On the other hand, SSDs cannot overwrite old data in NAND components until that old data has been erased with a separate operation. Therefore, SSDs must efficiently aggregate the erase operations without interrupting any active SSD read and write operations. To facilitate these NAND erase operations, also known as “garbage collection,” the Windows 7* OS issues a TRIM command to the SSD when files or data are no longer needed by the user.

**Definition and Explanation**

**Short:**

TRIM is a command issued by the operating system to inform the solid-state drive which blocks of data are no longer in use and can be wiped or erased internally. TRIM enables the SSD to handle its garbage collection to free up space for future writing of new data at a high sustained rate.

**Long:**

The TRIM command is designed to enable the operating system to notify the SSD which pages of data are now invalid due to erases by the user or operating system itself. During a delete operation, the OS will not only mark the sectors as free for new data, but it will also send a TRIM command to the SSD with the associated LBAs (Logical Block Address) to be marked as no longer valid. After that point, the SSD knows not to relocate the data from those LBAs during garbage collection. This will result in fewer writes to the flash, reducing write amplification and increasing drive life. Different SSDs will act on the TRIM command somewhat differently so the final performance can vary based on the SSD models.

TRIM is not supported in older Windows generation operating systems and in RAID configurations.

![Figure 7: SSD Trim Operation Flow Diagram](image-url)
10. Power Safe Write Cache

Overview and Context

Data integrity is the utmost concern with all computing and storage devices and environments. Unexpected events such as a power loss can cause serious data integrity issues, especially for data that is “in flight”—in temporary volatile DRAM write caches—and not securely saved in the storage device. SSDs providing a means to capture and save data “in flight” add to the overall system computing and storage platform integrity; this is the benefit that a power safe write cache provides.

Definition and Explanation

Short:
In the event of a power failure, a power safe write cache will have energy-storing capacitors to ensure that there is no data loss by providing enough energy to complete all writes to the NAND flash memory.

Long:
In the event of a power failure, a power safe write cache will have energy storing capacitors to ensure that there is no data loss by providing enough energy to complete all writes to the NAND flash memory.

Please see Section 2 on “PLI” (Power Loss Imminent) for more details of the capacitor operation in supplying enough energy to complete write operation.
11. DIPM and HIPM (Device and Host Power Management)

Overview and Context
SATA allows “PHY” Power Management to be Host Initiated (HIPM) or Device Initiated (DIPM), thus providing the flexibility to optimize the SATA components like Solid-State Drives and hard drives for a wide range of usages and applications. The host will have the ability to put SATA peripherals directly into Idle, Standby, and Sleep modes, and report the current power management mode of SATA peripherals.

SATA Link Power Management requires cooperation between the host and the device. Either can request the link to enter a low-power state, but the corresponding host or device must accept or reject the link state change request. Each of these provides power savings by themselves; maximum power savings, however, are achieved when both are implemented together.

Definition and Explanation

Short:
In HIPM (Host Initiated Power Management), the Host either in hardware or software manages the power state of SATA PHY to enter a low power state. In DIPM (Device Initiated Power Management) the Device manages the power state.

Long:
AHCI Link Power Management is a technique where the SATA AHCI controller puts the SATA link to the internal HDD and/or SSD into a very low power mode when there is no IO (input/output) activity for an extended period. The controller automatically puts the link back into active power state when there is real work to be done. This is done to save power consumption by the HDD and/or SSD.

Host-initiated power management can be implemented either in the host hardware or the host software. In the first case, the host controller requests a link power management transition immediately after all outstanding commands to the drive have been completed. This allows the link to enter a low-power state immediately upon completion of the commands to the disk. Since the host has the best knowledge of what commands have been posted, or will be posted to the device, the host is able to make an immediate link power state change without invoking a time-out period.

Figure 9: SATA Power Management States

Device-initiated power management is implemented by the drive. The drive knows best how long a specific command might take to complete, and is best equipped to request a link power management state change while processing the command. The host controller can automatically put the link into either Slumber or Partial after the command completes, typically, this will be Partial. However, after some extended period of idleness, the link will transition from Partial to Slumber. This can be done either by the host software or the device. Since the host is best equipped to manage the PHY between commands and the best device within a command, the best power management is obtained when the host and device cooperate.

References:
1. Intel Technology Journal, Volume 9, Issue 1, 2005
2. SATA Article Brief: SATA Power Management: “It’s Good to Be Green”; April 8, 2009
12. **Write Amplification**

**Overview and Context**

SSDs that use NAND Flash Memory as storage will have some data write amplification due to the nature of NAND Flash memory—that it must be erased before new data can be written—which requires extra NAND operations to move existing data possibly more than once. These extra NAND operations produce a multiplying effect that increases the number of writes required, producing an “amplification” effect; thus the term, “write amplification.” The write amplification factor constantly changes over the life of the SSD. Write amplification has many implications to the read/write performance and the reliability of the SSD. Depending on the SSD’s intelligence in managing its data, write amplification can cause extra wear and extra read/write/erase cycles on the NAND components reducing the life of the NAND Flash component. Additionally, the extra erase and write operations could cause an IOP latency outlier if these operations were done at inopportune times.

Because write amplification is very impactful to the life of a SSD, SSD controller companies, such as LSI-SandForce*, have developed compression algorithms where the amount of data written to the NAND Flash is less for every host write. Therefore, with write compression techniques, an SSD can achieve a write amplification that is less than 1.0, which enhances the reliability of the NAND components because less data is written. While there are merits to compression techniques, it is not a panacea to all the challenges of developing great Solid-State Drives.

**Definition and Explanation**

**Short:**

Due to the nature of NAND Flash Memories in that it must be erased before new data can be written may require extra NAND operations to move existing data more than once. This extra movement of data may involve erases and writes to accommodate the single host write request. These extra NAND write operations create a multiplying effect producing an “amplification” effect; thus the term, “write amplification”. Write amplification is typically measured by the ratio of writes coming from the host system and the number of actual writes required of the flash memory.

**Long:**

The following illustration shows the extra operations that NAND flash must do to accommodate one host write request:

1. Host Requests to Write 3 Pages of NAND Data:
2. Resultant NAND Component Operations:
   1) 3 copy or write operations to move pages “A”, “B” and “C” from Block 2 to Block 1, 3, and 4 respectively
   2) 1 erase operation of Block 2
   3) Page Program Operations of Page 1, 2, and 3 into Block 2
   4) Grand Total: 7 Program and 1 Erase Operation

![Figure 10: Write Amplification Formula](image)
13. RAS (Reliability, Availability, & Serviceability)

Overview and Context

RAS is a term created by IBM* to describe the robustness of their mainframe computers. In the past when mainframe computers were the hub of all computing servicing many remote terminal users, “uptime” was key feature or metric to the value of the mainframe computer. All users of such computers dreaded to hear that the computer (mainframe) was “down”. IBM wanted to differentiate their mainframes from others in that they had RAS features to ensure their computers would be “up” more than “down”. So, computers designed with higher levels of RAS have a host of features that help them stay available for long periods of time without failure—with some computer vendors offering uptimes on the order of years! While RAS was a term created for hardware, it is also being applied to software.

Definition and Explanation

Short:

RAS (Reliability, Availability, Serviceability) was a term created by IBM to differentiate their mainframe computers that offered higher levels of reliability features to detect and avoid crashing faults, availability even with a fault occurrence, and serviceability of repairs is fast and easy to ensure higher levels of “uptime”. Now this term once applied to mainframes is also being used for servers and data centers.

Long:

Computers designed with higher levels of RAS have a host of features that help them be Reliable, Available, and Serviceable.

Reliability means features that help avoid and detect faults. A reliable system does not silently continue and deliver results that include uncorrected corrupted data. Instead, it detects and corrects the corruption when possible.

Availability is the amount of time a device is actually operating as the percentage of total time it should be operating. Availability features allow the system to stay operational even when faults do occur.

Serviceability is the simplicity and speed with which a system can be repaired or maintained, and includes various methods of easily diagnosing the system when problems arise.

RAS features are available for most computer/server components, including:

Processor: Processor instruction error detection and instruction retry, including alternative processor recovery
Memory: Parity or ECC protection of memory components as well as memory bus
I/O: Cyclic redundancy check checksums for data transmission/retry and data storage
Storage: RAID configurations for Solid-state or magnetic disk storage; Journaling file systems for file repair after crashes; Checksums on both data and metadata, and background scrubbing
Power/cooling: Duplication of components to avoid failures (for example power-supplies). Systems are over-designed for the specified operating ranges of clock frequency, temperature, voltage, vibration. Temperature sensors are included to throttle operating frequency if temperatures are exceeded. Surge protector, uninterruptible power supply, and auxiliary power are also provided.
System: Hot swapping of components capability is provided. Predictive failure analysis is done to predict which intermittent correctable errors will lead eventually to hard non-correctable errors.

References: