Technical white paper

Risks of unexpected power loss on solid state drives

Table of contents

Introduction 2
Memory hierarchy 2
Write operations 2
Shutting down cleanly 3
Unexpected power loss (dirty shutdown) 3
Metadata loss 3
Wear leveling 3
Background processes (garbage collection) 3
Mitigation techniques 4
    Turning drive cache off 4
    Capacitor hold-up circuitry 5
    Uninterruptible Power Supply (UPS) 5
Testing SSDs for the effects of unexpected power loss 5
The HP Workstations differentiation 6
Summary 6
References 6

Introduction

The purpose of this white paper is to inform customers about the risks associated with unexpected power loss on the data integrity of solid state drives (SSDs). These risks can range from severe, in the case of the loss of an entire drive, to mild, in the case of a few kilobytes of lost data, to none, in the case of a drive with power loss protection circuitry. An overview of the common SSD memory hierarchy, along with the typical data path for a write operation, are discussed with areas of risk highlighted. Modern industry solutions which are currently being used to mitigate, or eliminate, the effects of unexpected power loss are discussed and the pros and cons of each method are highlighted.

Memory hierarchy

Modern SSDs implement a memory hierarchy system comprised of both volatile and non-volatile storage mediums. Volatile storage mediums, such as DRAM, offer very fast access times, but require a constant source of power to retain data. Non-volatile storage mediums, such as NAND, do not require a constant source of power to store data but have much slower access times. SSDs store the bulk of their data in non-volatile storage to maintain data integrity, and store a small amount of data which is used most frequently or most recently in volatile storage. This method offers a compromise between access times and data retention.

Write operations

Figure 1 shows a typical write operation from host to SSD. Data starts on the host and is sent to the SSD's volatile cache. The SSD storage controller monitors this operation and is responsible for directing the flow of data as well as various other tasks. Once data is transmitted, the host waits for the SSD to provide an acknowledgement stating that the data was received before proceeding. Under most configurations the SSD will acknowledge receipt of the data as soon as the data reaches the SSD's volatile cache. This improves performance and helps to lower the number of writes that need to go to the non-volatile storage, but it also puts the data at risk if power is unexpectedly removed. After the data is in the volatile cache and the host receives the acknowledge signal, the host continues processing and the drive firmware later decides when the data should be copied to the drive's non-volatile storage.

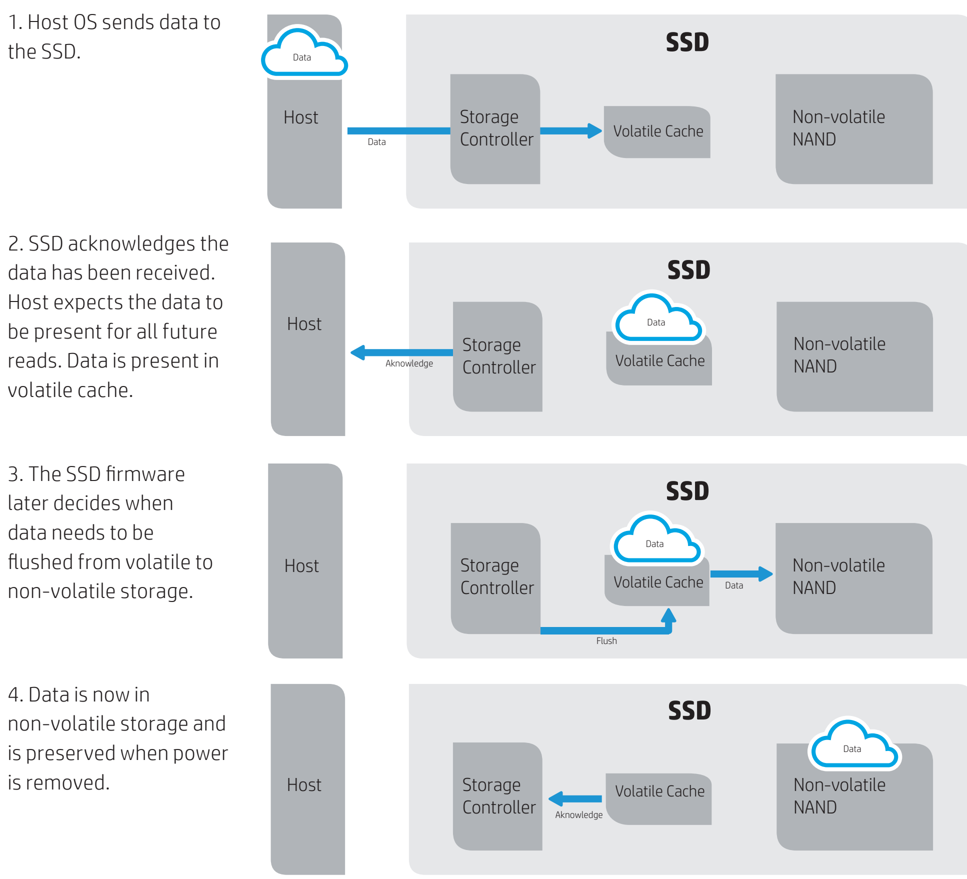

Figure 1. Simplified Write Operation from Host to SSD.

Shutting down cleanly

During a normal "clean" system shutdown the host signals to the storage device that a shutdown is occurring and power will soon be removed. Upon receiving this shutdown signal, the storage device is expected to flush any data from volatile storage to non-volatile storage and then send an acknowledgement signal to the host stating that the drive is ready for power removal. The host is expected to wait for this acknowledgement before proceeding with the shutdown process. This process ensures that power is removed only after data is stored safely on the device's non-volatile storage.

Unexpected power loss (dirty shutdown)

A "dirty shutdown" occurs anytime power is removed from a drive without the clean shutdown process described above. Some common examples of this occurrence are: unexpected power outages, accidental power removal from a computer, users unplugging storage device from a host system while power is on, and battery power loss. When power is removed from a non-power loss protected drive, without the clean shutdown process, the data which is stored in the drive's volatile cache will be lost. Since the data residing in the volatile cache is data which was used most recently, or data which is used most often, the data which is stored in the volatile cache could be very important user data, data used by the host operating system, or even data which is used by the SSD for normal operation.

Metadata loss

In addition to the loss of the data written to the drive from the host, it is also possible for a drive to lose important metadata pertaining to how data is mapped on the drive. Each time data is accessed or modified on the SSD, metadata is modified pertaining to the state of the data. This information allows for the storage controller to choose what data should be in the drive's volatile cache at any given moment as well as the ability to implement techniques which increase performance and endurance. A loss, or corruption, of metadata can result in seemingly lost data, in which the drive no longer has a pointer to the location of the actual data. More dramatically, a loss of metadata can result in an SSD which can no longer function for read or write operations [1].

Wear leveling

One of the most common storage improvement techniques is wear leveling. This technique allows for the drive to age evenly by mapping new writes to physical NAND cells which have had the fewest number of total writes. This is necessary due to NAND cells having a finite life determined by the number of writes occurring on them. This technique is transparent to the host which uses a Logical Block Address (LBA) when executing read/write operations to the drive. The drive then translates the LBA to the Physical Block Address (PBA) by way a flash translation table (FTL). The FTL is stored in cache and it is the responsibility of the controller to flush this table to non-volatile memory at times it deems appropriate or during a clean shutdown. The metadata associated with each page of data written also has information that can be used to rebuild the FTL if it is lost, but this rebuild takes time at the next power-on. If both the FTL and metadata are corrupted due to a sudden power loss event, the data stored on the SSD can become lost or even worse, the SSD, as a whole, can become inaccessible. SSD vendors go to great lengths to try and prevent inaccessibility even in the event of a sudden power loss.

Background processes (garbage collection)

Another area of concern is unexpected power removal during execution of background processes such as garbage collection. Due to technical limitations, to overwrite existing data in NAND memory requires an erase of the old data. Additionally, NAND memory can only be erased in large collections of bits referred to as blocks and written in smaller collections of bits called pages, see figure 2. When a command is sent to overwrite existing data in NAND, instead of overwriting the data on the drive which would require a copy, erase, and write cycle, the drive can instead write the new data to unused pages and remap the LBA to the new PBA. The old data is then flagged as invalid. Later, while the drive is not busy servicing the host, the valid pages located in the block which holds invalid pages can be copied over to free pages such that the block can be erased and reused. To do this, the mapping data from LBA to PBA is being updated. For this reason, if power is unexpectedly removed while garbage collection is taking place, faults can occur in the translation capabilities of the drive.

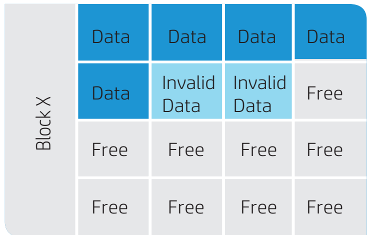

Figure 2. NAND block and page division with data, invalid data, and free space [2].

Mitigation techniques

The effects of unexpected power loss can be lessened, or even eliminated, with the aid of hardware and/or software solutions. A few of the most commonly used options are described below.

Turning drive cache off

The host is capable of issuing a command to the SSD instructing it to not use the volatile cache. This results in the host not receiving an acknowledgement from the SSD until the data has been committed to the non-volatile storage. Since the host can only logically expect data retention after an acknowledgment is received, this method ensures data integrity will always be maintained. This feature applies only to the actual data being transmitted from the host to the drive; the recent FTL data is still stored in the drive's volatile write cache [3]. SSDs need to keep this data in volatile cache for performance reasons but most SSDs implement methods which allow for the automated rebuilding of the FTL upon powering up after an unexpected power loss event. Figure 3 shows how this write operation differs from that of the standard write operation shown in Figure 1. It is important to note that under most applications the host process cannot proceed until receiving the acknowledgement from the drive.

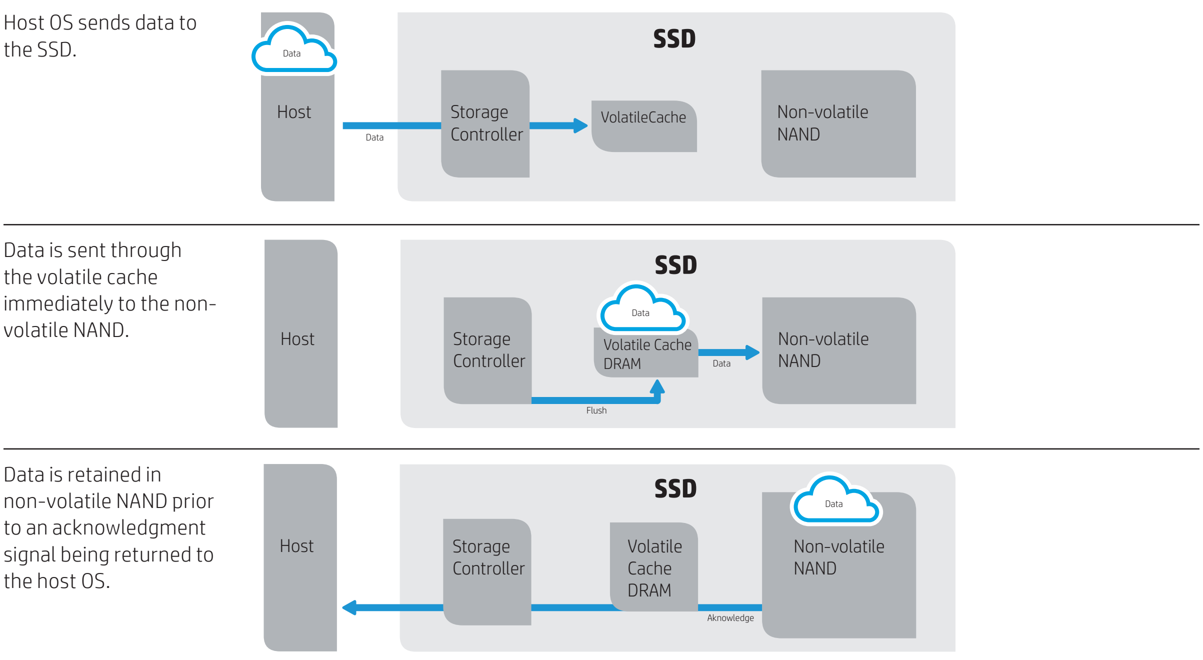

Figure 3. Write Operation from Host to SSD with Write-Cache Turned Off.

Pros: Data integrity is preserved in the event of a sudden power loss.

Cons: Dramatic decrease in read and write performance.
Increased writes to NAND resulting in lower write endurance and decreased drive life.
FTL data is still stored in the volatile cache, which leaves it at risk.

Capacitor hold-up circuitry

Capacitors located on the SSD can be used to supply power during unexpected power loss events. The SSD can monitor the incoming voltage and trigger a flush of the volatile cache whenever the voltage decreases past a designated threshold value. The capacitors must be sized appropriately such that enough power can be sourced to complete the flushing sequence and cleanly shut down the drive. An example of an SSD with capacitors for power loss protection is shown in Figure 4. The capacitors for this particular drive are located on the bottom of the PCB.

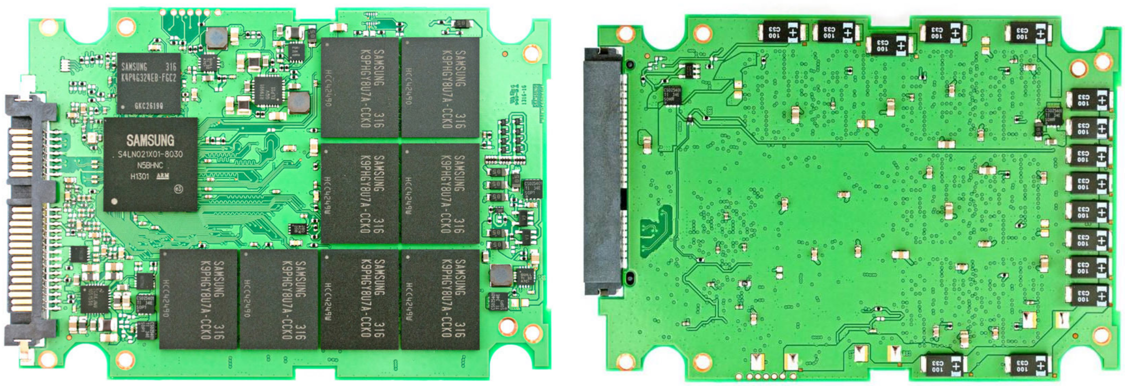

Figure 4. SSD with capacitors: Samsung SM843T, top view left, bottom view right [4].

Pros: Data integrity is preserved in the event of a sudden power loss.
Cons: Increased cost due to inclusion of additional circuitry including capacitors.
Additional board space is required.

Uninterruptible Power Supply (UPS)

An external battery backup power supply can be used to power the host system which in turn powers the SSD. This method protects the drive from unexpected power loss events occurring from power outages, but not from the host system's internal power loss events such as unexpected removal or system failures. UPS systems tend to be very costly but can offer protection from the most common forms of unexpected power loss.

Figure 5. Uninterruptible Power Supply [5].

Pros: Allows entire system protection from unexpected power loss (as long as system is shutdown cleanly before battery supply is depleted).
Cons: Expensive.
Does not protect drive against host system power failures or unexpected removal.

Testing SSDs for the effects of unexpected power loss

HP has developed a test which allows drives to be subjected to many cycles of unexpected power loss events. This test allows HP to assess the effect of unexpected power loss on the drive's data integrity. The results of these tests have shown that drives possessing no power loss protection are highly susceptible to loss of the data which is in flight between the host system and the drive's non-volatile storage. This behavior is to be expected, but more interestingly HP has also found drives which are prone to complete corruption when faced with unexpected power loss. These are drives which are no longer accessible when hooked up to a host system. Through a rigorous testing process, HP is able to filter out drives which are prone to this sort of behavior and ensure only the most reliable of drives are allowed into the HP Desktop and Mobile Workstation line of products.

The HP Workstations differentiation

HP Workstations offer several solutions to protect customers from the ill effects of sudden power loss. First, through a rigorous testing process, HP Workstations ensure that only the most reliable of commercial-grade drives are allowed into the Workstation class product line. Second, the HP Workstation product line offers enterprise-class SSDs which can withstand the most demanding of workloads and incorporate capacitor hold-up circuitry to protect against sudden power loss. Third, HP Workstations has developed a new PCIe storage device, designated the HP Z Turbo Quad Pro, which not only offers next generation PCIe access speeds, but also incorporates HP specific firmware and capacitor hold-up circuitry to ensure the highest level of protection against sudden power loss.

Summary

Under normal operating conditions all SSDs used in HP products provide a high-quality, high performance storage solution. For customers with very demanding data integrity concerns a solution with power-loss protection is recommended such as HP Workstations Enterprise Drives or the new HP Z Turbo Quad Pro.

References

[1] R. Micheloni et al., 2013. Inside Solid State Drives (SSDs). Springer Series in Advanced Microelectronics 37. Section 8.5: Reliable Data Management in Power Failure Scenarios. [ONLINE] Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB4QFjAA&url=http%3A%2F%2Fwww.springer.com%2Fcda%2Fcontent%2Fdocument%2Fcda_downloaddocument%2F9789400751453-c1.pdf%3FSGWID%3D0-0-45-1354508-p174543000&ei=kbtHVeaSLoilgwTa5IGIAw&usg=AFQjCNEefXGYFjMmxxKQy7H-fUDrJiLPlQ&sig2=DvojHbu0FrD98A4-iZXwXQ&bvm=bv.92291466,d.eXY. [Accessed 03 September, 2015]

[2] Les Tokar. 2012. Garbage Collection and TRIM in SSDs Explained - An SSD Primer. [ONLINE] Available at: http://www.thessdreview.com/daily-news/latest-buzz/garbage-collection-and-trim-in-ssds-explained-an-ssd-primer/. [Accessed 03 September 15].

[3] SanDisk Corporation. Whitepaper: "Unexpected Power Loss Protection." 2013. Available at: http://www.sandisk.com/Assets/docs/Unexpected_Power_Loss_Protection_Final.pdf. [Accessed 03 September 2015]

[4] Lyle Smith. 2014. Samsung SM843T SSD Review. [ONLINE] Available at: www.storagereview.com/samsung_sm843t_ssd_review. [Accessed 03 September 15].

[5] www.Newegg.com. 2015. Newegg Product Page. [ONLINE] Available at: http://www.newegg.com/Product/Product.aspx?Item=N82E16842102163. [Accessed 03 September 15].

4AA6-1470ENW, March 2016