SanDisk®

White Paper

Unexpected Power Loss Protection

SanDisk Corporation | Corporate Headquarters | 951 SanDisk Drive Milpitas, CA 95035, U.S.A.
Phone +1.408.801.1000 | Fax +1.408.801.8657 | www.SanDisk.com

Products, samples, and prototypes are subject to update and change for technological and manufacturing purposes. SanDisk Corporation general policy does not recommend the use of its products in life support applications wherein a failure or malfunction of the product may directly threaten life or injury. Without limitation to the foregoing, SanDisk shall not be liable for any loss, injury, or damage caused by use of its products in any of the following applications:

- Special applications such as military related equipment, nuclear reactor control, and aerospace.
- Control devices for automotive vehicles, train, ship and traffic equipment.
- Safety system for disaster prevention and crime prevention.
- Medical-related equipment including medical measurement device.

Accordingly, in any use of SanDisk products in life support systems or other applications where failure could cause damage, injury, or loss of life, the products should only be incorporated in systems designed with appropriate redundancy, fault tolerant, or back-up features. Per SanDisk Terms and Conditions of Sale, the user of SanDisk products in life support or other such applications assumes all risk of such use and agrees to indemnify, defend, and hold harmless SanDisk Corporation and its affiliates against all damages. Security safeguards, by their nature, are capable of circumvention. SanDisk cannot, and does not, guarantee that data will not be accessed by unauthorized persons, and SanDisk disclaims any warranties to that effect to the fullest extent permitted by law. This document and related material is for information use only and is subject to change without prior notice. SanDisk Corporation assumes no responsibility for any errors that may appear in this document or related material, nor for any damages or claims resulting from the furnishing, performance or use of this document or related material. SanDisk Corporation explicitly disclaims any express and implied warranties and indemnities of any kind that may or could be associated with this document and related material, and any user of this document or related material agrees to such disclaimer as a precondition to receipt and usage hereof. EACH USER OF THIS DOCUMENT EXPRESSLY WAIVES ALL GUARANTIES AND WARRANTIES OF ANY KIND ASSOCIATED WITH THIS DOCUMENT AND/OR RELATED MATERIALS, WHETHER EXPRESSED OR IMPLIED, INCLUDING WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR INFRINGEMENT, TOGETHER WITH ANY LIABILITY OF SANDISK CORPORATION AND ITS AFFILIATES UNDER ANY CONTRACT, NEGLIGENCE, STRICT LIABILITY OR OTHER LEGAL OR EQUITABLE THEORY FOR LOSS OF USE, REVENUE, OR PROFIT OR OTHER INCIDENTAL, PUNITIVE, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES, INCLUDING WITHOUT LIMITATION PHYSICAL INJURY OR DEATH, PROPERTY DAMAGE, LOST DATA, OR COSTS OF PROCUREMENT OF SUBSTITUTE GOODS, TECHNOLOGY OR SERVICES.

Executive Summary

Unexpected SSD power loss can cause critical data loss. IO commands (user data and host system data) that are in transit from the host to the SSD non-volatile flash memory (flash media) or are temporarily in the SSD volatile cache components and not fully committed to the flash media are dangerously exposed during unexpected power loss.

This document describes the challenges SSDs face with respect to protecting the exposed IO commands (data) during an unexpected power loss and how SanDisk SSD products address these challenges.

Overview

Storage devices require a graceful removal of power to ensure data integrity is preserved. Graceful removal of power includes commands to signal to the storage device that power might be imminently removed. The device in turn flushes all its volatile cache contents or in-transit data to the flash media. This command sequence is handled transparently by the operating systems' storage driver during a power down sequence.

In some cases, power cannot always be gracefully removed from the storage device. These cases include: users unplugging power from the system without prior notification, unexpected power outages, sudden battery loss, or users unplugging devices from a system. For these cases, a storage device like a SSD needs to have a mechanism in place to ensure data integrity. Host operating systems also include a mechanism to handle unexpected power loss, such as journaling or disk recovery.

The risk during an unexpected power loss is mostly relevant to IO write commands and device metadata table contents. The IO write command includes new or updated data which, for a period of time when in transit, is not saved in a safe location and therefore its loss during unexpected power loss can lead to data integrity problems. Due to the frequent access and quick response time requirements, device metadata tables are usually stored in a volatile cache; as such, the device metadata tables are subject to integrity problems when an unexpected power loss occurs. An IO read command is usually a request to read data that is already stored in a safe location; therefore, unexpected power loss should not have an impact on IO read commands.

SSD architecture includes solutions to improve performance and endurance, however, these solutions present more challenges for protecting IO write commands during unexpected power loss.

SanDisk SSD devices have been designed with solutions to minimize or completely eliminate the potential of data loss/corruption or device metadata table corruption in the event of an unexpected power loss. This document discusses the different mechanisms SanDisk uses to overcome the challenges posed by unexpected power loss.

Overview

IO write commands come from the host OS storage driver to the SSD controller for processing, moved to the volatile cache to increase overall SSD performance, as well as IO write optimization, and eventually committed into the flash media.

The following diagrams provide an overview of the SSD components and the IO write commands flow within two types of SSD products, and the risk points when an unexpected power loss occurs.

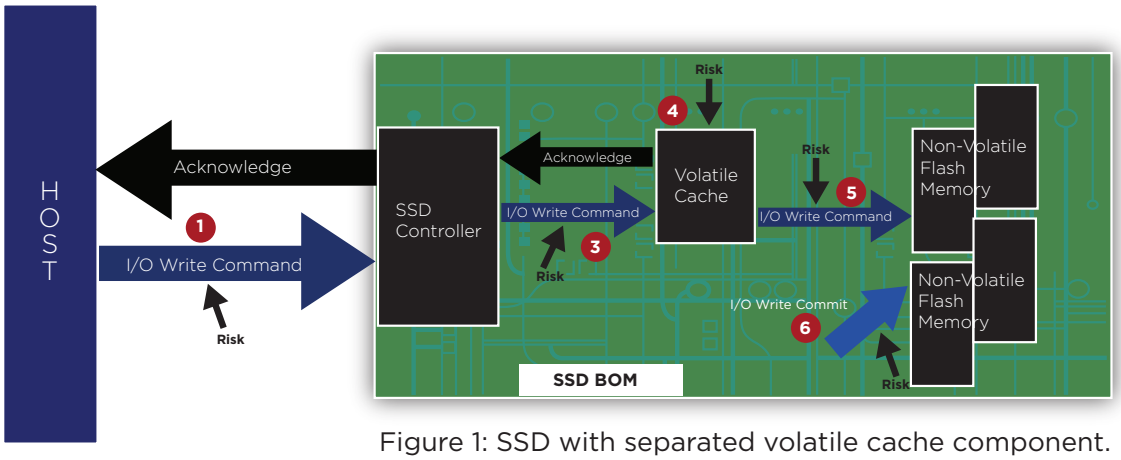


Figure 1: SSD with separated volatile cache component.

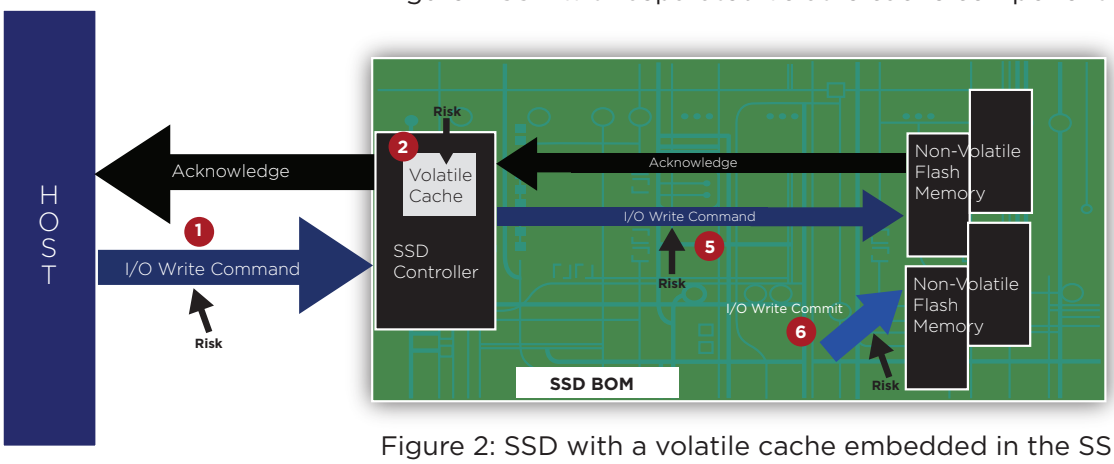


Figure 2: SSD with a volatile cache embedded in the SSD controller.

As shown in the above diagrams, during unexpected power loss there are risk points at each of the stages/locations, potentially exposing user data to risks such as data corruption and/or data integrity problems.

Terminology

In the ensuing discussion of risk points, the following terminology is used.

SSD Components

- SSD controller – The controller is an embedded processor that executes firmware level code.
- Volatile cache – SRAM/DRAM volatile memory is part of a SanDisk proprietary, sophisticated, tiered device cache-management system used mostly to improve endurance and increase overall SSD performance. The volatile cache also contains the SSD internal device metadata tables. The volatile cache is either a separated component or part of the SSD controller. The volatile cache does not retain the data when power is off.
- Flash media – The non-volatile flash memory (NAND) physical component that retains data even when power is off.
- SSD BOM – SSD build of material, physical assembly of all SSD components.

IO Write Command

IO write commands are chunks that include new or modified data.

SSD Internal Device Metadata Tables

These tables hold the virtual to physical mapping information.

Risk Points Summary

- IO write commands in transit from the host to the SSD controller.
- IO Write and device metadata tables stored temporary in the volatile cache which is part of the SSD controller in some of SanDisk SSD products.
- IO write in transit from the SSD controller to the volatile cache which is a separated component in some of SanDisk SSD products.
- IO write and device metadata tables stored temporary in the volatile cache which is a separated component in some of SanDisk SSD products.
- IO write in transit from the volatile cache to the flash media.
- IO write commit process into the flash media.

IO Write Commands In Transit

IO write commands in transit are commands that have left the host but have not yet arrived at their final destination are at risk during an unexpected power loss. At the code level, in transit actually comprises buffers with content at risk during an unexpected power loss. The route between the host and the final destination is assembled from segments, or in other words, from many buffers.

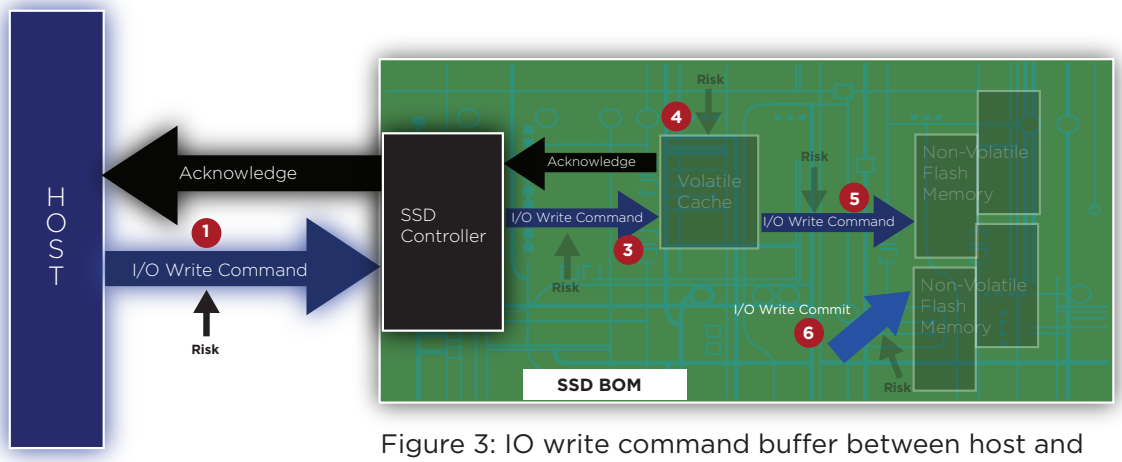


Figure 3: IO write command buffer between host and SSD controller.

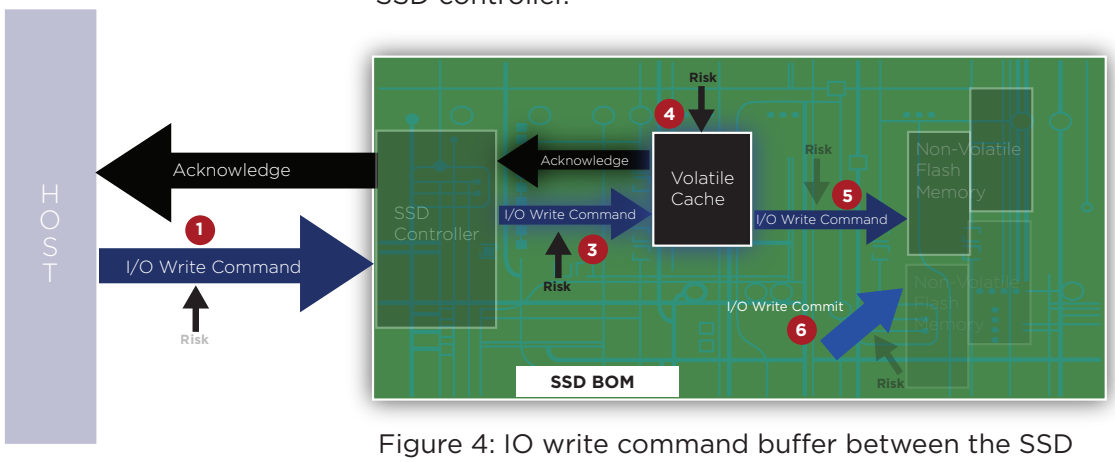


Figure 4: IO write command buffer between the SSD controller and the volatile cache (a separated component for some SanDisk SSD products).

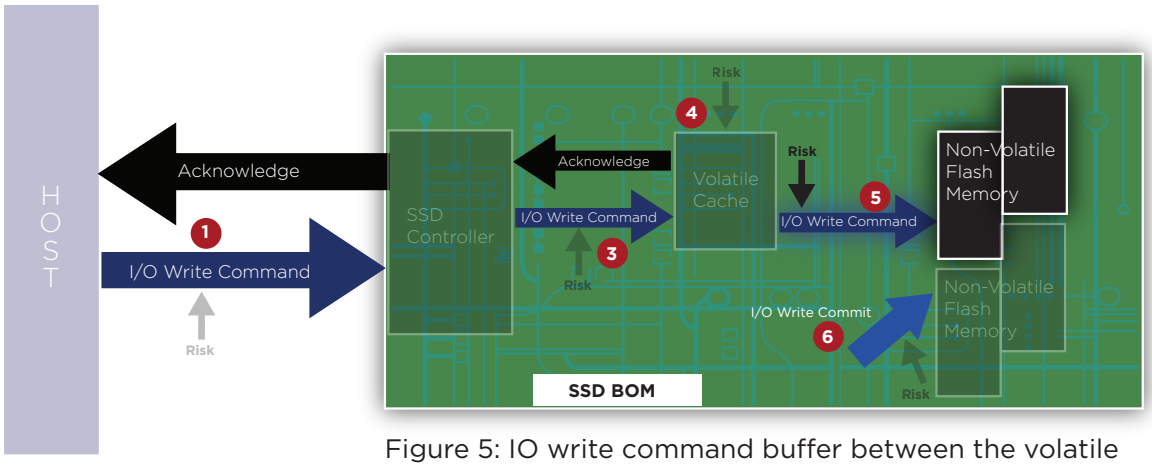


Figure 5: IO write command buffer between the volatile cache (a separated component for some SanDisk SSD products) and the flash media.

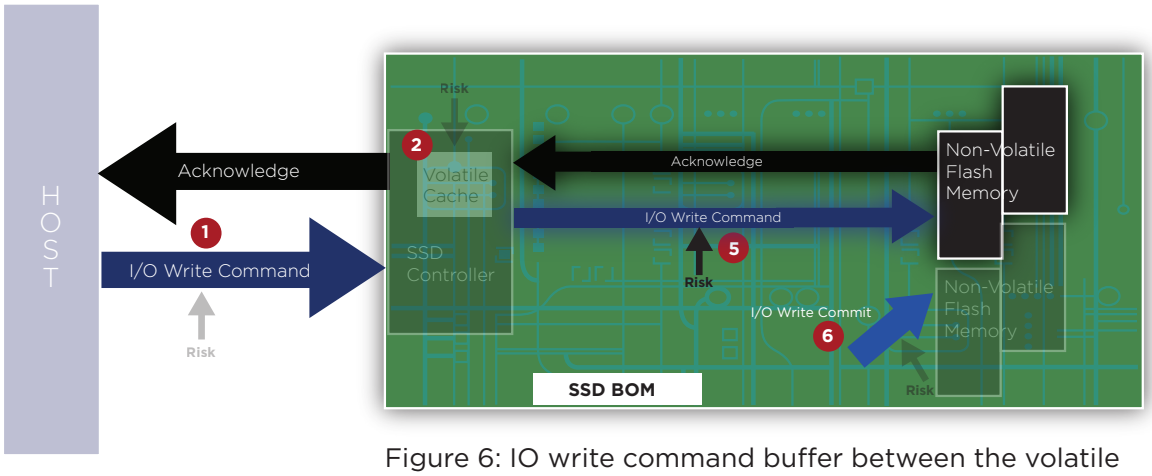


Figure 6: IO write command buffer between the volatile cache (part of the SSD controller for some SanDisk SSD products) and the flash media.

Volatile Cache

Volatile cache is a high speed Random Access Memory (RAM) component that can retain data as long as power is supplied. SanDisk SSD products use two types of RAM components:

SRAM – Static Random Access Memory, which is very fast, holds data as long as power is supplied, but is very expensive, with small densities.

DRAM - Dynamic Random Access Memory, which is less expensive, with higher densities. Due to its internal structure, DRAM cannot hold data for more than a few milliseconds and requires constant refreshing.

The SSD volatile cache is used to:

- Hold device metadata tables
- Improve endurance as part of SanDisk's proprietary, sophisticated, tiered device cache-management system
- Improve SSD write performance

All SanDisk SSD products include the volatile cache component, in varying sizes, using either SRAM (as part of the SSD controller) or DRAM (as a separated component).

Volatile Cache – FTL Device Metadata Tables

The constant dynamic allocation of where the data is physically stored in the SSD flash media requires presenting virtual fixed data location information to the OS and constantly translating this virtual location information to the constantly changing physical location on the flash media. This mechanism is called the flash translation layer and is implemented using tables that hold the virtual to physical mapping information. These tables are called device metadata tables as they include non-user data. These device metadata tables are normally stored on the non-volatile flash media. Due to its very frequent access and modification, some or all of the device metadata tables are loaded from the flash media into the SSD volatile cache to speed up the device metadata tables' access time; this in turn improves the overall SSD performance.

This mechanism presents a risk during unexpected power loss as some or all device metadata tables stored in the flash media are not updated or in sync with the device metadata tables on the volatile cache, which may lead to data integrity problems.

Volatile Cache – Improve Endurance

SanDisk SSD products use a proprietary, sophisticated, tiered device cache-management system. Modern operating systems mostly send small chunks to the storage device, for the most part 4KB chunks. The small chunks conflict with the larger chunks. To bridge these differences, SanDisk has developed a tiered device cache management system to minimize these conflicts using volatile cache memory and non-volatile memory (also known as nCache™) to aggregate small IO in different sizes to improve the overall SSD endurance. Nonetheless, processed IO writes in the volatile cache are exposed to the risk of loss due to an unexpected power interruption.

Volatile Cache – Improve SSD Write Performance

When IO writes arrive at the SSD volatile cache, an acknowledgement is sent to the host without waiting for a full commit into the flash media, which improves the overall SSD write performance.

Operating systems do not consider the IO write to be safe until it is fully commited into the flash media (even if an aknowledegment has received from the SSD).

Volatile Cache – Flush User Data and Synchronize Device Metadata Tables

The device metadata tables in the volatile cache must be constantly synchronized with the device metadata tables in the flash media. In addition, the IO writes (user data) in the volatile cache must be fully committed into the flash media as quickly as possible, ideally after being aggregated to an optimized IO size. Committing the IO writes must be combined with syncing the device metadata tables that contain the relevant information for the new/modified IO.

To this end, two combined processes take place:

1. The SSD controller flushes user data from the volatile cache to the flash media in the event of one of the following triggers/requests:
   - Flush command issued by the OS storage driver following a flush cache request from an application or from the file system.
   - Event driven flush cache command issued by the host, for example, during graceful shutdown.
   - SSD firmware initiate flush cache.
2. At the same time, the SSD controller syncs the device metadata tables in the volatile cache.

IO Write Commit into the Flash Media

The final destination of the IO writes is the SSD flash media, which is also known as write commit.

In Multi-Level Cell (MLC) NAND, each word line (collection of cells) of the memory array consists of two pages of data; a lower page and an upper page. The lower and upper pages are logical concepts that reside on the same physical word line.

During a write to the upper page, the upper and lower pages' state is continuously altered until the write is complete. If unexpected power loss occurs, the current IO write commit is at risk together with the stored data in the lower page.

As previously discussed, at any point in time the IO write commands are either:

- In transit between the host, SSD controller, SSD volatile cache, and the flash media
- In the SSD volatile cache waiting for IO writes optimization process.
- In process of being committed into the physical flash media.

In each of the above phases, the IO write commands are exposed to the risk of data integrity/corruption during unexpected power loss.

Modified device metadata tables that have not yet been synchronized are also exposed to data integrity risks during unexpected power loss.

The following sections describe general and SanDisk specific solutions for user data and device metadata tables' protection during an unexpected power loss.

Volatile Cache

The content of the volatile cache (device metadata tables and IO writes) is protected by using one or more of the following protection mechanisms:

Disable the Use of SSD Volatile Cache

Most operating systems provide an interface from which the user can disable the use of the SSD volatile cache either for all or some host IO writes. The host OS uses, in combination or separately, the following methods:

- Increasing Flush commands frequency - This does not really eliminate the use of the SSD volatile cache. IO writes still go through the SSD Volatile Cache but the OS issues more frequent flush cache commands, minimizing the time when the IO write is not stored safely in the flash media and device metadata tables are not in sync with the tables on the flash media.
- Using ATA 'set features' – In most SSD devices the volotile cache can be disabled by using ATA set features command. IO writes bypass the volatile cache and goes directly to the flash media.

Note: Metadata tables stored on the volatile cache are not affected.

Pros: Minimizes or eliminates the risk of losing/corrupting IO writes in the volatile cache during unexpected power loss.

Cons: Cache disabled configuration significantly reduces the overall SSD performance (device metadata tables are still exposed).

Device Metadata Table Recovery – SanDisk Specific

During normal operation, the SanDisk SSD firmware periodically syncs the device metadata tables in the SSD volatile cache with the device metadata tables safely stored in the flash media. Modified device metadata tables in the volatile cache that have not yet been synchronized are exposed to the risk of corruption in the event of an unexpected power loss. To meet this challenge, SanDisk SSDs use a proprietary journaling mechanism that helps recover the last version of device metadata tables present in the SSD volatile cache when an unexpected power loss occurred.

Pros: Ensures very minimal or no loss of data.

Cons: Longer SSD power-on to ready time resuming full operation (can take several seconds).

IO Write Commands in Transit

As previously noted, a Host OS cannot assume the IO write is "safe" until it is fully committed into the Flash Media. Therefore, the Host OS uses the following methods:

Host IO Write Retries

The host treats the IO write command as an incomplete operation until it receives an acknowledgment from the SSD controller. Most applications and operating systems include a mechanism that waits for the acknowledgment for configured amount of time; if the acknowledgement is not received, it resends the same IO write command for n number of times. Eventually, if acknowledgement is not received after n number of times (for example, due to unexpected power loss), the operating system invalidates the SSD drive.

Pros: The retry mechanism is a good solution if the unexpected power loss is very short within the defined OS timeout.

Cons: The retry mechanism is not effective if power downtime duration is longer than the application/OS timeout.

It does not help in the case of an IO write command that was acknowledged by the SSD volatile cache but was en route between the SSD volatile cache and the flash media during an unexpected power loss.

IO Write Commit

As described in the previous section (IO Write Commit into the Flash Media), SanDisk SSD products utilize MLC technology, therefore, when the write commit is to the upper page, the content of the corresponding lower page is at risk as well.

SanDisk SDD products use either or both of the following mechanisms:

Backup and Recovery

- During Normal Operation - As part of the normal write commit process into an upper page, the data in the lower page is copied to the safe location in the Flash Media. The actual write into the upper page is not performed until this copy operation completes.
- During SSD Power Up Process after unexpected power loss – The following process occurs:
  1. The open block is scanned starting from the last sync point onward, looking for the last write.
  2. The last write that was committed successfully is verified; if not, it is declared invalid.
  3. If the last write in the open block is invalid, the safe location blocks are searched. If a safe copy of the invalid write is stored there, it is recovered from the safe location block.

Pros: Minimizes the chance of data corruption.

Cons: There is no guarantee that safe zone blocks hold a copy of the corrupted data. Longer SSD power on to ready time.

Capacitors – SanDisk Specific

Some SanDisk SSD products include electrical capacitors with sufficient capacitance to provide enough power during unexpected power loss to safely complete the current write commit into the flash media.

The use of the electrical power from the capacitors is triggered by a power monitoring unit residing on the electrical power input line. Once the power monitoring unit detects the power loss, it triggers the SSD firmware, which in turn completes only the current write commit process and rejects any others.

Some SanDisk embedded SSD products provide detail specifications for the capacitors and power monitoring unit, to be built by the system integration vendor on the system. These SSD products' firmware includes an interface to trigger the appropriate commands.

Pros: Minimizes the chance of data corruption.

Cons: Costly solution; it increases the overall SSD BOM cost.

Data Protection Per Product Series

The following sections include descriptions of unexpected power loss data protection mechanisms, implemented with each of SanDisk SSD product series.

Additional details on each data protection mechanism can be found in previous sections: Volatile Cache, IO Write Commands In Transit, and IO Write Commit.

SanDisk SSD X and A Series

SanDisk X and A series include the following unexpected power loss protection mechanisms:

Disable the SSD Volatile Cache using the following ATA command:

Command Name	Code
Set features	0xEF
Sea features sub-commands	
Disable write cache	0x82

IO Write Commit

SanDisk X and A series use the safe location mechanism to protect existing data in the lower page while writing new data to the upper page. It also includes the mechanism to detect unsuccessful writes during SSD power up after unexpected power loss to attempt recovery from the safe location.

SanDisk SSD U Series

SanDisk U series includes the following unexpected power loss protection mechanisms:

Disable SSD Volatile Cache using the following ATA set features:

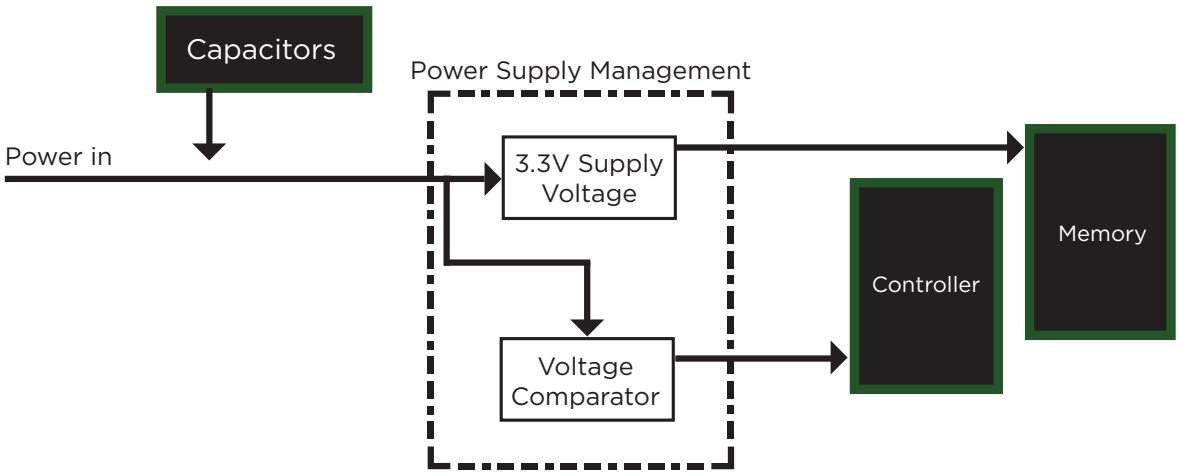
Command Name	Code
Set features	0xEF
Sea features sub-commands	
Disable write cache	0x82

IO Write Commit

SanDisk U series uses the safe location mechanism to protect existing data in the lower page while writing new data to the upper page. It also includes the mechanism to detect unsuccessful writes during SSD power up after unexpected power loss and the attempt to recover from the safe location.

Capacitors

SanDisk U series includes the following capacitor system as part of the product:



During an unexpected power loss:

1. The Voltage Comparator detects the power loss and sends a signal to the SSD controller.
2. The SSD controller and the Flash Media use the electricity from the capacitors to complete the current write commit and reject any other write requests.

To ensure enough electrical power above 2.5V for at least 2.5ms, depending on the SSD capacity, the capacitors hold the following capacitance:

SSD Capacity	32GB	64GB	128GB	256GB
Capacitance	500µF	570µF	700µF	700µF

SanDisk SSD i Series

SanDisk i series includes the following protection mechanisms against unexpected power loss:

Disable the SSD Volatile Cache using the following ATA command:

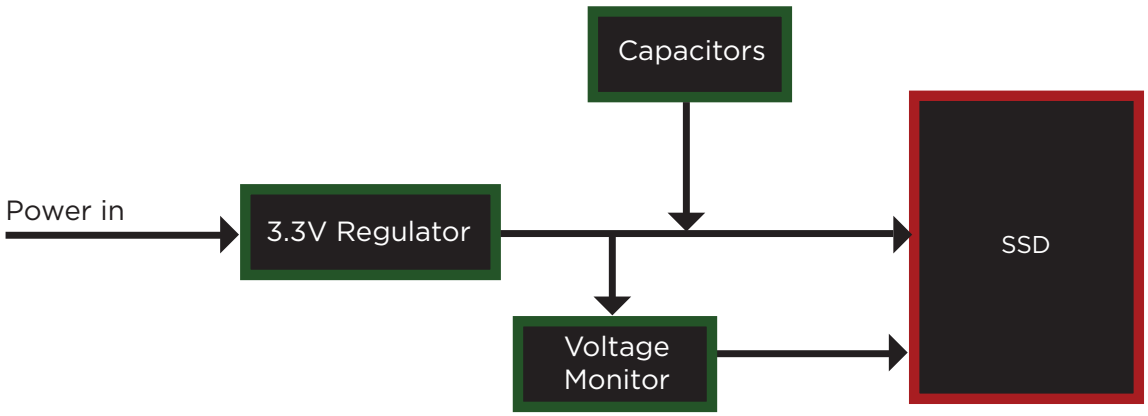
Command Name	Code
Set features	0xEF
Sea features sub-commands	
Disable write cache	0x82

IO Write/Programming

The i series uses the safe location mechanism to protect existing data in the lower page while writing new data to the upper page. It also includes the mechanism to detect unsuccessful writes during SSD power up after unexpected power loss and the attempt to recover from the safe location.

Capacitors

The i series firmware includes the capability to use capacitors, however, as an embedded product, the i series does not include capacitors instead it provides specific requirements to the system integrator vendor for capacitors and a voltage monitor:



It is up to the system integrator vendor to provide the regulator, capacitors, and voltage monitor according to SanDisk specifications.

During an unexpected power loss:

1. The voltage monitor detects the power loss and sends a signal to the SSD controller.
2. The SSD controller and the flash media use the electricity from the bulk decoupling capacitors to complete the current write commit and reject all other write/programing requests.

To ensure enough electrical power above 2.5V for at least 2.5ms, depending on the SSD capacity, the capacitors provided by the system integrator vendor should hold the following capacitance:

SSD Capacity	32GB	64GB	128GB	256GB
Capacitance	400µF	440µF	470µF	470µF

SanDisk SSD products use different methods to protect the IO writes during unexpected power loss. Some of these methods are generic and some are SanDisk specific. By using these methods, SanDisk SSD products minimize or eliminate the risk of data loss/data corruption during unexpected power loss.

