Temperature Considerations for Industrial Embedded SSDs

Introduction

Employing NAND flash-based storage has become a mainstay in most of today's industrial embedded applications. Solid state drives (SSDs) satisfy the high performance, low power and strict reliability requirements of a growing list of embedded systems and suppliers have been able to reduce the cost per gigabyte to make these solutions increasingly more viable. This advanced technology, however, is not without its drawbacks.

One key design challenge SSD vendors face is the susceptibility of an SSD to data corruption from an unexpected power failure. This is a critical concern for designers of most embedded systems that require "five nines" uptime - especially those operating in harsh environments. This issue is further compounded by the difficulty in maintaining a steady and uninterrupted power source - especially in light of the proliferation of battery-operated and portable devices in the Internet of Things (IoT), which are particularly sensitive to power loss.

Power failure threats for embedded systems can range from spikes to brown-outs. These less-than-ideal conditions are the cause of data corruption resulting in field failures, which lead to the potential loss of revenue from equipment returns.

Because this is a key issue, it is a critical to understand how a given storage solution will operate in a specific application. This white paper will present the data corruption effects on an SSD when a system loses power during a write operation. It will illustrate the advantages of specifying an SSD that integrates voltage detection and power hold-up circuitry giving users advanced warning of a possible power irregularity. It will also discuss the differences in SSDs designed for use in industrial applications compared to those for the consumer electronic market.

NAND Flash Technology

To understand the issues associated with ungraceful power-downs, it is important to first understand the basics of NAND flash structure. Figure 1 shows a block of NAND flash. NAND is characterized by page size – the minimum write or program unit – and block size – the minimum erase unit.

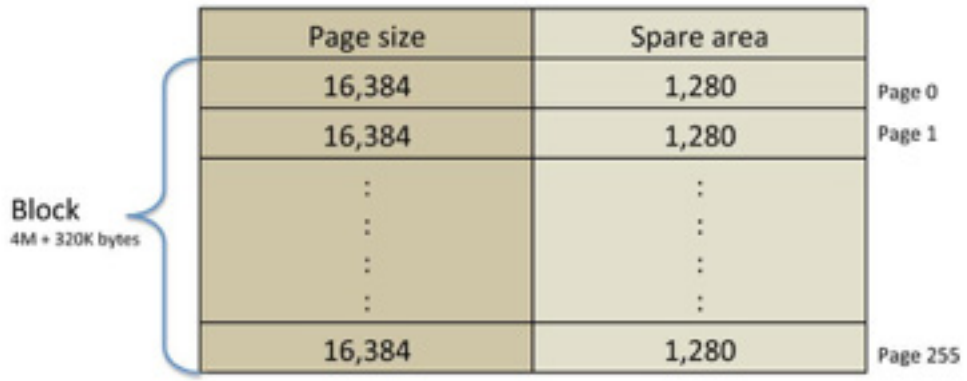

Figure 1: Block and Page Structure of a 16K Page NAND Flash Device

For example, a 16K page NAND has (32) 512-byte sectors per page and 256 pages per block. Each page has an associated spare area where metadata such as Error Correction Code (ECC) is stored. In general, the larger the NAND page size, and the longer the programming time (tprog), the more susceptible it is to power disturbances. 1ynm MLC NAND typically has 16K pages, while SLC NAND ranges from 2KB to 8KB page size. Programming times for SLC are usually much shorter for SLC than for MLC. For these reasons, and for the fact that MLC uses paired pages (two pages open at once) during write operations, SLC SSDs are much more robust than MLC SSDs.

Data Corruption from Unexpected Power Loss

Obviously, the ideal scenario is that the host system always ensures a graceful, deterministic power-down sequence. This can be difficult if not impossible (or at least impractical) for many industrial embedded systems. Since Virtium is focused exclusively on the infrastructure SSD market, we have integrated various techniques for mitigating power-related issues to significantly lessen or eliminate unscheduled downtime. Many suppliers of consumer-based SSDs have not implemented these techniques as they are not seen as economically viable, and often consumer data is not deemed “mission critical.”

Each time the host writes to the SSD, the flash controller within the drive computes an ECC and stores the information in the corresponding page spare area. Sophisticated management technology is used to monitor and compare this ECC information with the corresponding sector to ensure data integrity. In the event of an unexpected power interruption, the ECC signature or data that is being written at the time can become corrupted. If the host tries to read that same sector on the next power-on, the newly computed ECC signature will not match. Similarly, if there is a power disturbance during a large file write, the file may be truncated. For a storage device to be considered robust, the truncation should take place at a sector boundary.

Significance of Errors Due to Power Disturbances

An unexpected power event could occur during the middle of a write operation, resulting in a read sector error. Many applications that have a read sector error will automatically produce a system-level error, which could result in system downtime until the error is corrected. In addition, the SSD may also interpret the read sector error as a defective sector, and would unnecessarily replace it with a spare sector. Once the number of factory-defined spare sectors reaches zero, the drive would need to be replaced unnecessarily.

Corruption as the result of power disturbances can affect host systems differently depending on where and when the disturbance occurs. The following are the various file types that could be corrupted and ways these files can impact the host.

- Master Boot Record (MBR): If the MBR becomes corrupted, it must be repaired because Mit

will be impossible to access any volumes on that drive, rendering it inoperable. To repair the drive, it will need to be repartitioned and re-formatted. Data will be lost.

- File Allocation Table (FAT): Corruption of the FAT causes the user to lose access to the files on the drive and will require that the drive be reformatted. Data will be lost.
- Critical System Files: In an operating system, the loss of one or more critical system files such as config.sys or system.ini result in system errors or may cause the system not to boot. This situation, too, is the source of unscheduled downtime requiring a service call for the drive to be reformatted and the operating system to be re-installed.
- Corruption of User Data Files: User data files can be also be corrupted and can happen without the user being aware until it is too late. After a power disturbance, it is recommended that the entire drive is checked and any sector errors repaired.

In the case of critical file overwrite, the host can re-partition and re-format the drive. In the scenario of running out of spares, the product must be returned to the vendor. After failure analysis, the vendor can re-initialize the drive and send it back. Unfortunately, system downtime is a factor and the data will be lost, but the drive can be used again. It is, therefore, imperative that SSDs include integrated value-add technologies to reduce the overall impact of an abrupt system power-down.

Virtium vtGuard® Technology

Virtium's StorFly® and TuffDrive® SSDs with integrated vtGuard technology provide voltage detection circuitry that acts as an "early warning" system for potential power anomalies. Figure 2 below shows a block diagram of the vtGuard technology. Once the host has requested a data transfer (write), the SSD makes the necessary preparations to receive the data. It must prepare the internal buffer RAM, align data with the particular NAND page size and acknowledge the impending write.

The figure below also presents three different power-down scenarios: power fail before the SSD acknowledges to the host that it "has the data"; power fail after the SSD acknowledges that it "has" the data but before the data has been committed to NAND flash; power fail after the SSD "has" the data in the active block NAND but before it has been committed to the correct logical block address (LBA).

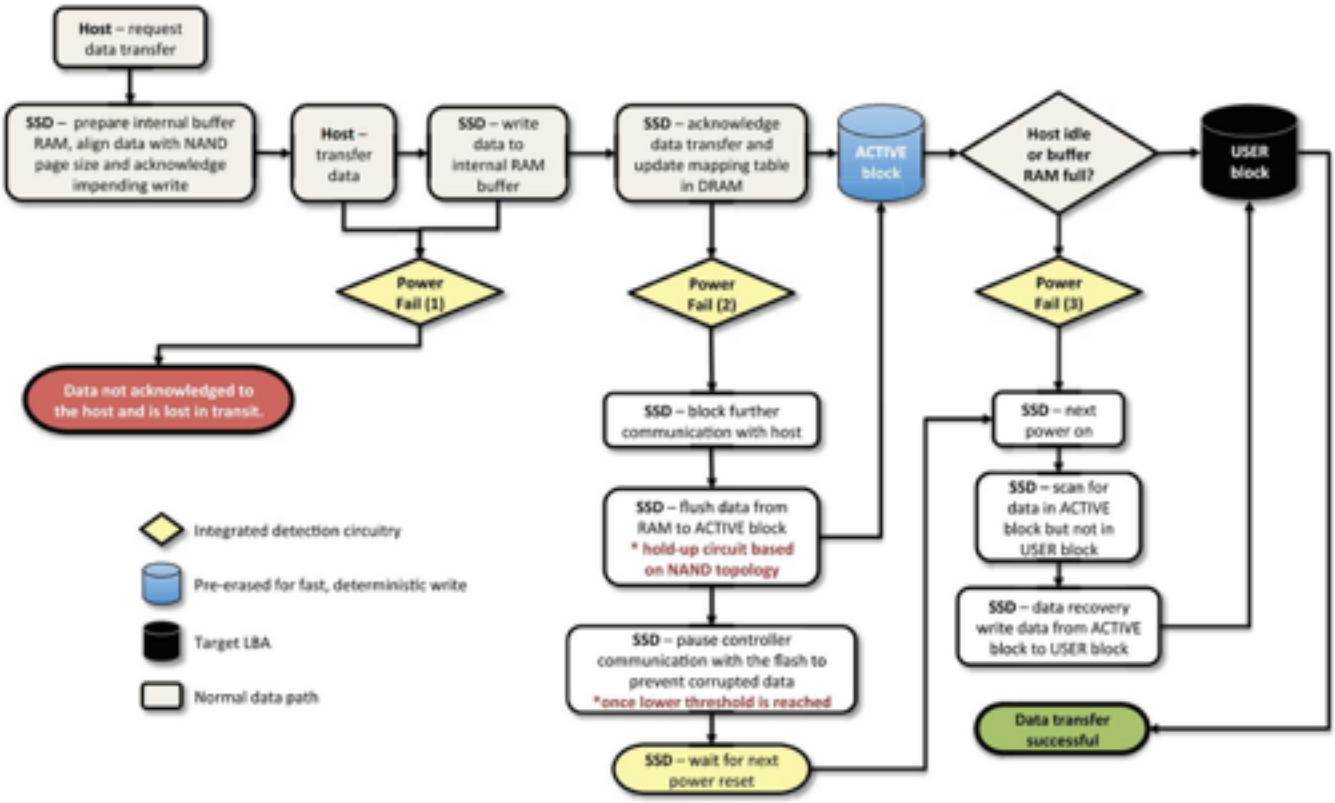

Figure 2: Power Fail Scenarios

It is important to understand two key items prior to reviewing the individual scenarios in the flowchart. First, the SSD must have circuitry to recognize the impending power disturbance. This is where voltage detection circuitry comes in. Based on operating voltage (5V, 3.3V or other) and NAND topology (MLC or SLC, page size), the detection circuitry needs to set the initial voltage threshold to start the protection technology. This detection circuitry can be different for different SSD vendors and is usually considered proprietary intellectual property.

Second, the SSD must have some mechanism for holding up the power long enough to "flush" the data that may be in the internal buffer of the SSD before the lower threshold is reached. Some SSD vendors use a capacitive bank, some a super-cap and still others a battery. Virtium's position is that super-caps and batteries introduce more reliability issues – especially over a wide temperature range – than they solve, so the vtGuard circuit does not use these technologies. Once the lower threshold is reached, the SSD controller and firmware must cut communication with the NAND to mitigate the risk of data corruption.

Normal Operation

During normal operation, the host requests a data transfer, the SSD prepares to receive the data and the host transmits it. Once the data goes "through the interface" it is first written to the internal RAM buffer. The size of this buffer can vary from SSD vendor to SSD vendor. Virtium chooses to keep this buffer as small as possible. While this may sacrifice a bit of performance, it benefits the SSD in that a smaller and more reliable hold-up circuit can be used. The data is then transferred from the buffer to an active block. Since NAND flash must be erased before it can be written, the active block is pre-erased in the over-provisioned area of the SSD and is immediately available to receive data. The data is then transferred from the active block to the target user block at a convenient time as determined by the firmware and the workload.

Power Fail Scenario #1

In scenario number one, the power disturbance occurs during a write, but before the SSD has acknowledged receipt of the data. Therefore, in the case of a small, random write, the host would treat that as if the data was never sent. If this happens in the middle of a large file write, the file would appear to be truncated.

Power Fail Scenario #2

In scenario number two, the power disturbance occurs after the SSD has acknowledged receipt of the data but before the data has been written to the active block. Once the power failure has been detected, the SSD controller and firmware immediately cut off communication with the host, and the hold-up circuit kicks in. The hold-up circuit keeps enough voltage available for the data to be flushed from the internal RAM buffer to the active block. This is where the pre-erase of the active block is really important. Once the lower threshold is reached, the controller pauses communication with the NAND flash to prevent unstable signals that could result in data corruption. On the next power-up cycle the SSD behaves like scenario #3.

Power Fail Scenario #3

In scenario number three, the power disturbance occurs after the data has been written to the active block but before it has been written to the correct LBA (the user block). On subsequent power-up the SSD scans for any data in the active block that has not been written to the proper user block. Any data found will be committed to the proper LBA and the active block will be recycled for new data.

Power-Down Testing and Validation

There are several factors to be considered to ensure the power down testing is most effective and accurately evaluates the storage media for a given application.

- Power-down ramp rate and all I/O pins: the steeper the ramp, the faster the voltage falls below the storage devices' minimum programming voltage. This presents an increased chance for error.
- Power-down timing: a comprehensive power-down test requires power to be removed at various intervals of the write cycle.
- Cut power to I/O and Vcc at the same time: eliminates any Vcc feedback that may occur if power is cut to Vcc and not to the I/O.
- Validation and verification: read sector errors may occur at any point in time yet the drive may continue to operate. Obviously, the worst-case scenario would be if the first read sector error occurred in a critical system file.
- Check for read errors: scan the previously written sector confirm the read sector status
- Validate data in other sectors: this ensures that data was not written to another location by mistake, or in the case of MLC, that the paired page was not corrupted.

Virtium uses ULINK hardware and an in-house developed script to test and validate its vtGuard technology to 5,000 power-down cycles during initial product qualification. This qualification covers the controller and firmware platform (for example Gen2 StorFly SATA 6G) with both SLC and MLC NAND flash. Virtium also performs an additional 1,500-cycle report for each form factor and capacity point both at initial release and during ongoing reliability test.

Power cycle test outputs are available under NDA. Contact your Virtium representative for more information.

Virtium manufactures memory and storage solutions for the world's top industrial embedded OEMs. For two decades we have designed, built and supported our products in the USA - fortified by a network of global locations. Our world-class technology and unsurpassed support provide a superior customer experience that continuously results in better industrial embedded products for our increasingly interconnected world.

Virtium
Solid State Storage and Memory

30052 Tomas | Rancho Santa Margarita, CA 92688
Phone: 949-888-2444 | Fax: 949-888-2445
www.virtium.com