M.2 SSD

Technical White Paper

Issue 01

Date 2017-07-09

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions

Notice

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://e.huawei.com

Email: support@huawei.com

Contents

1 Overview

M.2, formerly known as the Next Generation Form Factor (NGFF), is an interface for solid state disks (SSDs) and other expansion cards. M.2 SSDs evolved from mSATA SSDs, with more flexible physical specifications. These SSDs support both SATA and PCIe protocols.

M.2 SSDs use flash memory as storage media. NAND flash memory is the current industry favorite, and stores data by storing electrons on the floating gate. However, electrons frequently passing through the gate weaken and eventually inhibit the gate's ability to store them. Because of this inherent problem, the volume of written data must be monitored to prevent component failure from premature write-through.

2 Application Constraints

An M.2 SSD can only be used to boot a Linux OS.

M.2 SSDs have weak endurance and are unsuitable as data storage devices, which tend to be enterprise-level HDDs or SSDs with high drive writes per day (DWPD). Because of the write-through risks M.2 SSDs carry over short periods, they cannot be used in frequent data erase operations.

Write-intensive service software installed on M.2 SSDs reduces the lifespan faster, resulting in lasting SSD damage. M.2 SSDs are not recommended for such services. They are also contraindicated in cache scenarios.

3 Working Principles

3.1 Basic Features

3.1.1 Port Compatibility

SATA-based M.2 SSDs support SATA 3.0 port transfer of up to 6.0 Gbit/s.

3.1.2 Form Factor

The initial line-up of the commercially available M.2 SSDs is 22 mm (0.87 inch) wide, with varying lengths of 30, 42, 60, 80 and 110 mm (1.19, 1.65, 2.36, 3.15, and 4.33 inch). M.2 SSDs also come in various capacity specifications, as listed in the product documentation.

3.1.3 Secure Erase

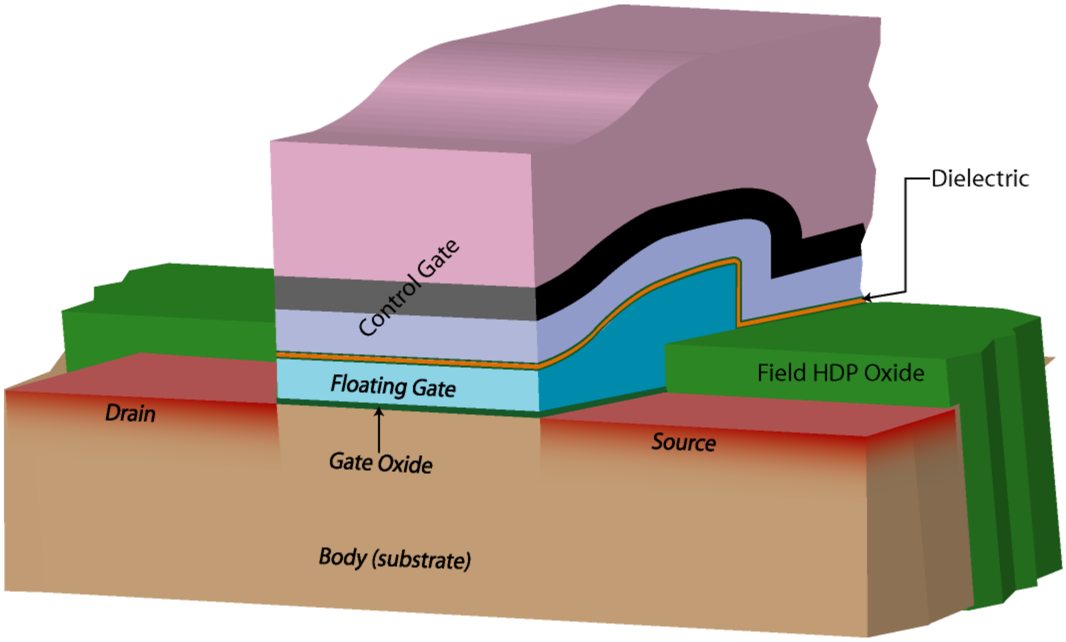
Secure erase is a common feature for low-level data formatting.

3.2 Flash Chips

3.2.1 Flash Memory Cells

An individual flash memory cell is a component with three parts. The parts have same names as those on a field effect transistor: source, drain, and gate. An insulation layer of silicon dioxide between the gate and the silicon substrate prevents electric charge from leaking out of the floating gate. This structure enables the storage unit to hold electric charge.

Like a field effect transistor, flash memory is voltage-controlled. NAND flash erase and write are tunnel effects: The current passes through the insulation layer between the floating gate and the silicon substrate layer to charge (writing data) or discharge (erasing data) the floating gate.



3.2.2 Cell Classification

A single-level cell (SLC) stores one bit of data. Flash memory using such cells is called SLC flash memory.

A multi-level cell (MLC) stores at least two bits of data. Multiple bits are stored because an MLC has multiple energy states, or voltage values, during charging.

A triple-level cell (TLC) is similar to an MLC in architecture and stores three bits of data. It writes more slowly than an SLC and MLC, with a smaller P/E cycle (500 to 1000 times).

3.2.3 Write Lifespan Analysis

Write amplification (WA) is an undesirable phenomenon of flash memory and SSDs. Flash memory must be erased before it can be rewritten, so data is moved more than once. These repeated operations shorten the lifespan and consume the bandwidth of the flash memory, affecting random write performance for the same volume of written data.

The following is a simple formula to calculate the WA of an M.2 SSD:

WA = Data written to flash memory/Data written by the host

If 4 KB data is to be written to the currently full flash memory, invalid data must first be erased from it. The host will write all the data to the cache, erase the block, update the entire block in the cache, and then write back the new data. In this operation of writing 4 KB data, the entire block (1024 KB, for example) is rewritten and the WA value is 256. This turns a simple operation of writing 4 KB data into four operations: flash memory read (1024 KB), cache memory update (4 KB), flash memory erase (1024 KB), and flash memory write (1024 KB), resulting in increased delay and a slower speed. This example illustrates how WA affects the random write performance and lifespan of SSDs.

In terms of WA, 100% write from low LBA to high LBA has a possible value of 1. 100% random writing of 4 KB of data to SSDs can exceed 100 in worst case scenarios. In practice, the WA ranges between 1 and 256.

Factors affecting WA are as follows:

- Garbage collection: WA increases during garbage collection and decreases after garbage collection.
- Over-provisioning: WA decreases with increased over-provisioning.
- TRIM: A TRIM command reduces WA.
- Secure erase: reduces WA.
- Sequential writes: reduces WA.
- Random writes: increases WA.
- Separating static and dynamic data: reduces WA.
- Wear leveling: increases WA.

4 Write Lifespan and Risk Assessment

4.1 Write Lifespan

An M.2 SSD is a read-intensive storage component with a low DWPD. It can only be used as a system boot disk and not for write-intensive services. Take the typical M.2 SSD model with a 240 GB capacity and DWPD of 0.3 as an example. It supports a maximum of 131 TBW (total bytes written) within its five-year warranty.

The formula for calculating its lifespan is as follows:

TBW = (DWPD x Capacity x 365 x Warranty years)/1000

4.2 Risk

According to the formula specified in section 4.1 "Write Lifespan", M.2 SSDs are not suitable for write-intensive services, since their maximum TBW is easily exceeded within the warranty and they carry high write-through risk.

4.2.1 Lifespan Determination

To find out the remaining lifespan of an M.2 SSD, query its SMART information. Take an Intel M.2 SSD as an example. The initial percentage of its remaining lifespan falls linearly from 100 to 1 with increasing erase count, indicating that the SSD has reached the end of its expected lifespan.

4.2.2 When to Replace M.2 SSD

When its remaining lifespan falls below 5%, the M.2 SSD is about to experience write-through and must be replaced immediately.