Overprovisioning and the SanDisk® X400 SSD

Improving Performance and Endurance with Overprovisioning

Rev 1.2

October 2016

CSS Technical Marketing

Western Digital Technologies, Inc.

951 SanDisk Dr. • Milpitas, CA 95035

Phone (408) 801-1000 • Fax (408) 801-8657

www.sandisk.com

Overview

Enterprise workloads are diverse and can create different levels of fragmentation of low level NAND storage. For example, pure sustain random writes, create internal media fragmentation and results in many additional writes to the NAND media, while sequential write workload generates very little fragmentation and keep the ratio between the host writes to NAND writes closer to 1.

Workloads such as pure random write can significantly affect the drive endurance as well as sustain write performance and consistency. While enterprise-class SSDs contain higher level of spare NAND to mitigate the internal write fragmentation, client and read-intensive SSDs tend to maximize the storage capacity provided to the user rather than keeping more NAND as spare for internal management.

In order to lessen the gap between full enterprise-grade SSDs and read-intensive drives like the SanDisk® X400, SSDs comes with internal features to increase the spare blocks while decreasing the user capacity; this process is called **overprovisioning**. Overprovisioning is the practice of reserving an additional portion of the overall NAND capacity to boost performance and endurance of an SSD--two important attributes for users--beyond its typical specification.

This application note will compare performance and endurance for the SanDisk X400 SSD at 0%, 7%, and 28% overprovisioning. This will help customers determine whether overprovisioning is needed for their respective workloads.

How SSDs organize and manage data

SSDs are based on NAND flash technology. Unlike magnetic memory, NAND cannot be updated in place (direct over-write) – it must first be erased before it can be written to. NAND memory is organized in erasable units called **blocks**. Blocks are many MBs in size while user writes are typically only a few KBs at a time. Therefore, today's SSDs organize blocks into small **pages**. Typical page sizes are 8KB and 16KB. For an SSD to function, it has a minimum number of extra blocks of NAND memory beyond the stated capacity of the drive.

To perform a data write from the host, an SSD writes to clean or previously erased but unwritten pages in a block. At the same time, the locations of the old data are marked as containing invalid data. To allow the SSD to continue to receive and store new data from write commands, the drive needs to have one or more free blocks with no valid data that can be erased for re-use. This is accomplished by taking a block which contains both valid data and invalid data and moving all valid data to other block(s) until the original block is completely empty and can be erased. This process is commonly referred to as **garbage collection**.

Write Amplification Factor

There are several different garbage collection memory management schemes or strategies. Regardless of which strategy is implemented, the net effect is that host writes will lead to greater amount of writes to the NAND. **Write Amplification Factor (WAF)** is simply the ratio of total GB of NAND writes to total GB of host initiated writes.

WAF is dependent on the workload that an SSD is being used for. For example, high sequential write usages would commonly be seen in *multimedia* being written to a catalog or to a cache-kind of storage and has a WAF very close to 1. Commercial applications where **Trim** is not utilized, result in very high WAF. For example, in *Multi-Function Printers* where Trim is not used, storage capacity may be reduced where it is not a critical resource and thus gain benefits of OP to level WAF.

WAF and Performance

Since garbage collection is extra work that an SSD needs to perform, the added work can degrade performance. The higher the WAF, the more extra work is needed, and the more performance is degraded. Conversely, the lower the WAF, the less extra work is needed, and the drive can provide better performance as a result.

WAF and Endurance

Like all NAND based devices, an SSD has a physical limit as to how much data can be written to it over its lifetime. This is known as the **endurance** of the SSD. The underlying NAND memory of an SSD has a limit on how many erase and program cycles it can go through before it no longer can perform those functions reliably. In the Computing and Commercial world, endurance is measured as **Terabytes Written (TBW)**, meaning the number of terabytes of data that the host can reliably write to the storage device over its lifetime and expect to be able to retrieve the data without errors.

In the Enterprise world endurance is measured in **Data Write Per Day (DWPD)**. It is simply how much data a user is able to reliably write to the drive each day over the guaranteed life of the drive. It is commonly expressed as a fraction of the drive's user capacity. The relationship between TBW and DWPD is

DWPD = TBW x 1024 / drive capacity in GB / life of drive in days.

Garbage collection, as represented by WAF, is extra writes that an SSD needs to perform. The added non-host initiated writes consume the endurance of a drive. The higher the WAF, the more extra write is done, and the more TBW is shortened. Conversely, the lower the WAF, the less extra write is done, and TBW is extended as a result.

Overprovisioning

An SSD must have a minimum amount of extra NAND memory blocks for it to function. Overprovisioning is a design strategy to reduce WAF and hence improve performance and endurance by reserving additional NAND memory beyond the minimum requirement. For a given SSD with a fixed amount of total physical NAND memory, this can be accomplished by reducing the capacity made available to the user.

For example, even a “binary capacity” 256GB SSD may be built with 262GB of actual NAND memory, of which 6GB is reserved as the minimum amount of extra memory needed for the drive to function. This drive is said to have 0% of overprovisioning, even though in reality it has 2% of extra memory. By configuring the same 256GB drive to 240GB, an additional 16GB of free space becomes available for the drive to use internally. This reduced capacity drive is said to have 7% (16/240) overprovisioning. If the same drive is configured as a 200GB drive so that an additional 56GB of extra memory becomes available for the drive to use internally, then the 200GB drive is said to have 28% (56/200) overprovisioning.

Capacity Point	Configured to	Ratio	% of Overprovisioning[1]
256GB	256GB	0/256	0%
256GB	240GB	16/240	7%
256GB	200GB	56/200	28%

Overprovisioning reduces WAF by providing more space where invalid data can be kept. This increases the chance of finding blocks with a high percentage of invalid data to be selected for reclaim. Such blocks are reclaimed with minimal amount of valid data which needs to be relocated elsewhere. Thus, the amount of extra work necessitated by garbage collection is reduced. Hence, WAF is lowered with increased overprovisioning. As explained in the previous section, lowering the WAF will extend the TBW. Increasing an SSD’s overprovisioning will also improve its performance.

1 The table as shown is only focused on OP in relation to improvement in performance and endurance and does not include other capacity that may not be available to the user that is reserved for other functions.

Configuring Overprovisioning

There are multiple ways to configure an SSD to any level of overprovisioning. They all require the drive to start in an empty and secure erased state. Fresh out of the box (FOB) will also work.

One simple method is to create a logical partition of the desired size using the disk management tool that comes with Windows OS. The unallocated space on the drive will not be used by the user and becomes free space for the SSD firmware to take advantage.

Another method is to use third party tools such as HDAT20 or HDPARM. (See Reference section)

Overprovisioning SanDisk X400 SSD

The SanDisk X400 family of SSDs is available in several capacities up to 1TB. They can be formatted to different percentages of overprovisioning by reducing the drive capacity, as in the following table.

% of Overprovisioning	Resulted Capacity[2]			
	128GB	256GB	512GB	1024GB
7%	120GB	240GB	480GB	960GB
28%	100GB	200GB	400GB	800GB

Endurance Impact

These WAF numbers are used to determine the endurance of various capacity points:

JEDEC workload		Overprovisioning		
	TBW	0%	7%	28%
Capacity at 0% over-provisioning	128GB	72	74	83
	256GB	80	95	107
	512GB	160	189	198
	1024GB	320	378	396

[2] The table as shown is only focused on OP in relation to improvement in performance and endurance and does not include other capacity that may not be available to the user that is reserved for other functions.

The equivalent DWPD numbers for a 3 year product life are:

3 year life		Overprovisioning		
	DWPD	0%	7%	28%
Capacity at 0% over-provisioning	128GB	0.53	0.58	0.78
	256GB	0.29	0.37	0.50
	512GB	0.29	0.37	0.46
	1024GB	0.29	0.37	0.46

The benefit of overprovisioning on TBW is clearly illustrated.

Performance Impact

To fully illustrate the power and benefit of overprovisioning on performance, we illustrate here using a 4KB pure random write workload instead of the JEDEC workload. It is the most demanding workload that can be applied to a storage device. Instead of estimating the impact of overprovisioning using estimation and modeling, we did actual run on the devices and took actual measurement results. The full test procedure is described in the appendix.

The following table shows how IOPS performance is improved for each capacity of X400 when the overprovisioning is increased:

4KB random writes		Overprovisioning		
	IOPS	0%	7%	28%
Capacity at 0% over-provisioning	128GB	651	2,255	5,737
	256GB	2,131	4,547	9,948
	512GB	2,696	5,693	9,957
	1024GB	2,176	4,066	8,264

Summary

Whether your data needs are saving a large multimedia application, or using a Multi-Function Printer operation, increasing the overprovisioning of the X400 will reduce its WAF and hence extend the endurance and lifetime of the SSD. Increasing the overprovisioning of the X400 will also improve the overall performance – higher throughput (IOPS) and lower latencies.

References

DLAT2 download: http://www.hdat2.com/download.html

HDPARM download: https://sourceforge.net/projects/hdparm/

Appendix

To maximally demonstrate the overprovisioning effect on WAF, a pure random write workload of 4KB command block size is applied. All the capacities from the X400 family are tested, with 0%, 7% and 28% of overprovisioning tested for each capacity. A queue depth of 32 is used to drive the SSD as hard as possible. IOMeter is used as the tool for generating and injecting this workload to the drive under test. The test procedure for each capacity and each overprovisioning is as follows:

- Secure erase the drive
- Set the overprovisioning to be tested (see previous section)
- Write twice the user capacity of the drive with 128KB sequential writes
- Write twice the user capacity of the drive with 4KB random writes
- Dump and save the SMART log of the drive (to record host write and NAND write parameters)
- Test the drive with 4KB random writes @ QD32 for 24 hours
- Save the IOMeter results at the end of the test (to obtain average IOPS)
- Dump and save the SMART log of the drive (to record host write and NAND write parameters)
- From the two SMART log dumps, compute the total number of GB of host writes and the total number of NAND write. Divide the two to obtain WAF.