EDN
NETWORK

Understanding SSD over-provisioning

Kent Smith, LSI Corporation - January 8, 2013

http://www.edn.com/design/systems-design/4404566/Understanding-SSD-over-provisioning

The over-provisioning of NAND flash memory in solid state drives (SSDs) and flash memory-based accelerator cards (cache) is a required practice in the storage industry owing to the need for a controller to manage the NAND flash memory. This is true for all segments of the computer industry—from ultrabooks and tablets to enterprise and cloud servers.

Essentially, over-provisioning allocates a portion of the total flash memory available to the flash storage processor, which it needs to perform various memory management functions. This leaves less usable capacity, of course, but results in superior performance and endurance. More sophisticated applications require more over-provisioning, but the benefits inevitably outweigh the reduction in usable capacity.

The Need for Over-provisioning NAND Flash Memory

NAND flash memory is unlike both random access memory and magnetic media, including hard disk drives, in one fundamental way: there is no ability to overwrite existing content. Instead, entire blocks of flash memory must first be erased before any new pages can be written.

With a hard disk drive (HDD), for example, that act of “deleting” files affects only the metadata in the directory. No data is actually deleted on the drive; the sectors used previously are merely made available as “free space” for storing new data. This is the reason “deleted” files can be recovered (or “undeleted”) from HDDs, and why it is necessary to actually erase sensitive data to fully secure a drive.

With NAND flash memory, by contrast, free space can only be created by actually deleting or erasing the data that previously occupied any block of memory. The process of reclaiming blocks of flash memory that no longer contains valid data is called “garbage collection.” Only when the blocks, and the pages they contain, have been cleared in this fashion are they then able to store new data during a write operation.

The flash storage processor (FSP) is responsible for managing the pages and blocks of memory, and also provides the interface with the operating system’s file subsystem. This need to manage individual cells, pages and blocks of flash memory requires some overhead, and that in turn, means that the full amount of memory is not available to the user. To provide a specified amount of user capacity it is therefore necessary to over-provision the amount of flash memory, and as will be shown later, the more over-provisioning the better.

The portion of total NAND flash memory capacity held in reserve (unavailable to the user) for use by the FSP is used for garbage collection (the major use); FSP firmware (a small percentage); spare blocks (another small percentage); and optionally, enhanced data protection beyond the basic error correction (space requirement varies).

Even though there is a loss in user capacity with over-provisioning, the user does receive two important benefits: better performance and greater endurance. The former is one of the reasons for using flash memory, including in solid state drives (SSDs), while the latter addresses an inherent limitation in flash memory.

Percentage Over-provisioning

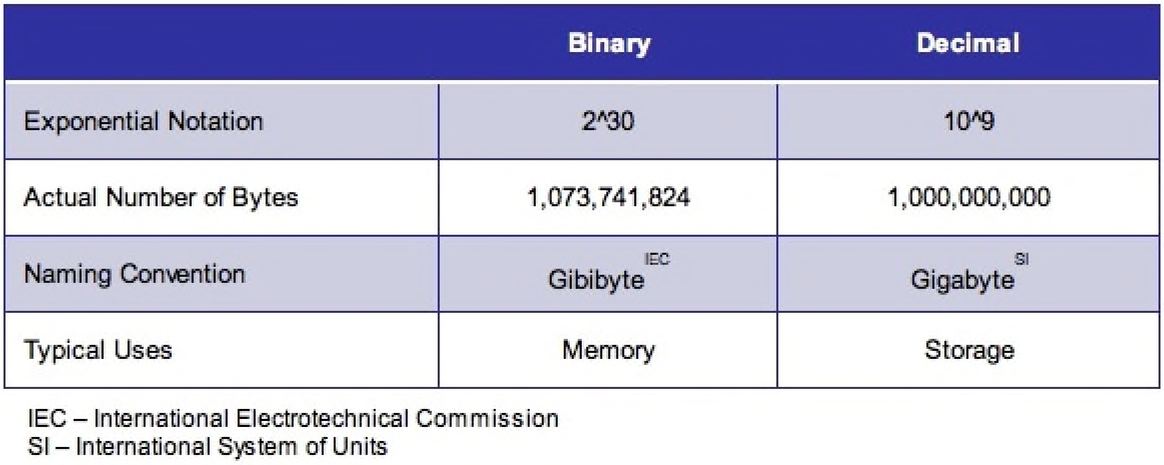
The equation for calculating the percentage of over-provisioning is rather straightforward:

$$\text{Percentage Over-provisioning} = \frac{\text{Physical Capacity - User Capacity}}{\text{User Capacity}}$$

For example, in a configuration consisting of 128 Gigabytes (GB) of flash memory total, 120 GB of which is available to the user, the system is over-provisioned by 6.7 percent, which is typically rounded up to 7 percent:

$$\frac{128\text{ GB} - 120\text{ GB}}{120\text{ GB}} = 0.067$$

It is also important to note another factor that often causes confusion: a binary Gibibyte is not the same as a decimal Gigabyte. As shown in Figure 1, a binary GB is 7.37 percent larger than a decimal GB. Because most operating systems display the binary representation for both memory and storage, this causes over-provisioning to appear smaller because the actual number of bytes is 7.37 percent higher than the number of bytes displayed. This is why an SSD listed as providing 128 GB of user space can still function with 128 GB of physical memory. Using the calculation above, the over-provisioning amount would appear to be zero percent, which is impossible for NAND flash. In reality it is really over-provisioned closer to 0 + 7.37 percent.



	Binary	Decimal
Exponential Notation	2^30	10^9
Actual Number of Bytes	1,073,741,824	1,000,000,000
Naming Convention	Gibibyte[IEC]	Gigabyte[SI]
Typical Uses	Memory	Storage

IEC – International Electrotechnical Commission
SI – International System of Units

Figure 1. The difference between a binary Gigabyte and a decimal Gigabyte

Test Environment

To isolate the over-provisioning variable, the tests were conducted on a single SSD with Toshiba MLC (multi-level cell) 24nm NAND flash memory controlled by an LSI SF-2281 flash storage processor. It is important to note that the FSP used employs the LSI DuraWrite™ technology that optimizes writes to flash memory, and utilizes intelligent block management and wear-leveling to improve reliability and endurance. These capabilities combine to afford over five years of useful life for MLC-based flash memory with typical use cases.

Previous testing performed by LSI revealed that entropy has an effect on performance only for SSDs without data reduction technology. For this reason, the red lines in the graphs showing the results for 100% entropy are labeled “Typical SSDs.” This series of tests, which used SSDs equipped with LSI DuraWrite data reduction technology, were designed to evaluate performance at different levels of both over-provisioning and entropy, and to specifically test the hypothesis that data reduction could improve performance at lower levels of entropy.

Test result data points are based on post-garbage collection, steady state operation. All preconditioning used the same transfer size and type as the test result (e.g. random 4KB results are preconditioned with random 4KB transfers until reaching steady state operation).

VDBench V5.02 was used as the main test software with IOMeter V1.1.0 providing cross-check verification. The test PC was configured with an Intel Core i5-2500K 3.30 GHz processor, the Intel H67 Express chipset, Intel Rapid Storage Technology 10.1.0.1008 (with AHCI Enabled); 4 GB of 1333 MHz RAM; and Windows 7 Professional (32-bit).

Performance Test Results

Sequential writes were uniform across all tested over-provisioning ranging from zero to 75 percent. This flat performance derives from the nature of sequential writes to flash. As data is written to flash memory, it completely fills all of the pages in a block. When the drive becomes filled, blocks of data that are no longer valid need to be erased first via the garbage collection process, which it does by simply erasing entire blocks without needing to move (read then write) any individual pages that might otherwise still be valid. Because there are no incremental writes during garbage collection during this operation, there is no benefit from additional free space. With SSDs that use a data reduction technology like DuraWrite from LSI, the level of flat performance will increase as a function of the entropy (data randomness); the lower the entropy the higher the performance. In this situation, however, the increase in performance is due to the reduced writes being completed sooner and not from the additional free space.

Throughput performance for sustained 4KB random writes improved as the amount of over-provisioning increased. Additionally, for SSDs with DuraWrite data reduction technology, the throughput improvement also increased at all levels of entropy.

Figure 2 shows the results of this test. The reason why the increased over-provisioning improves performance for random writes is due to how garbage collection operates. As data is written randomly, the logical block addresses (LBAs) being updated are distributed across all the blocks of the flash. This causes a number of small “holes” of invalid data pages among valid data pages. During garbage collection those blocks with invalid data pages require the valid data to be read and moved to new empty blocks. This background read and write operation requires time to execute and prevents the SSD from responding to read and write requests from the host, giving the perception of slower overall performance. When the over-provisioning is a higher percentage of the total flash memory, the time required for garbage collection is reduced, enabling the SSD to operate faster.

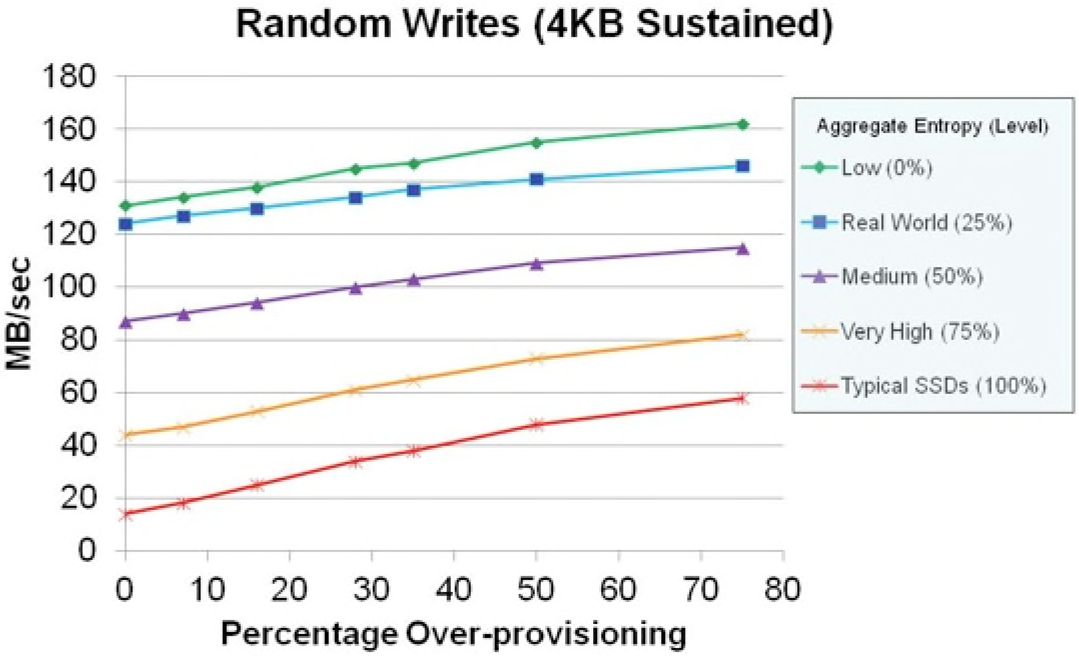


Figure 2. The effect of over-provisioning on write performance throughput

The need for garbage collection and wear-leveling with NAND flash memory causes the amount of data being physically written to be a multiple of the logical data intended to be written. This phenomenon is expressed as a simple ratio called “write amplification,” which ideally would approach 1.0 for standard SSDs with sequential writes, but typically is much higher due to the addition of random writes in most environments. With SSDs that have DuraWrite technology, the typical user experiences a much lower write amplification that is often on average only 0.5. Getting write amplification low is important to extending the flash memory’s useful life.

Random write operations have the greatest impact on write amplification, so to best view the effect of over-provisioning on write amplification, tests were conducted under those conditions. As shown in Figure 3, write amplification for sustained 4KB random writes benefited significantly from a higher percentage of over-provisioning for SSDs that do not include DuraWrite technology. For SSDs that do include DuraWrite or a similar data reduction technology, the throughput improvement increased at a higher rate at higher levels of entropy.

Note also how the use of a data reduction technology like DuraWrite minimizes the benefits of over-provisioning for lower levels of entropy. When the entropy of the user data is low, DuraWrite is able to reduce the amount of space consumed in the flash memory. Because the operating system is unaware of this reduction, the extra space is automatically used by the flash storage processor as additional over-provisioning space. As the entropy of the data increases, the additional free space decreases. At 100 percent entropy the additional over-provisioning is zero, which is the same result as a “Typical SSD” (red line) that does not employ a data reduction technology. Referring again to Figure 3, a standard SSD with 28 percent over-provisioning would have the same write amplification as an SSD with DuraWrite technology at zero percent over-provisioning for data with an entropy as high as 75 percent.

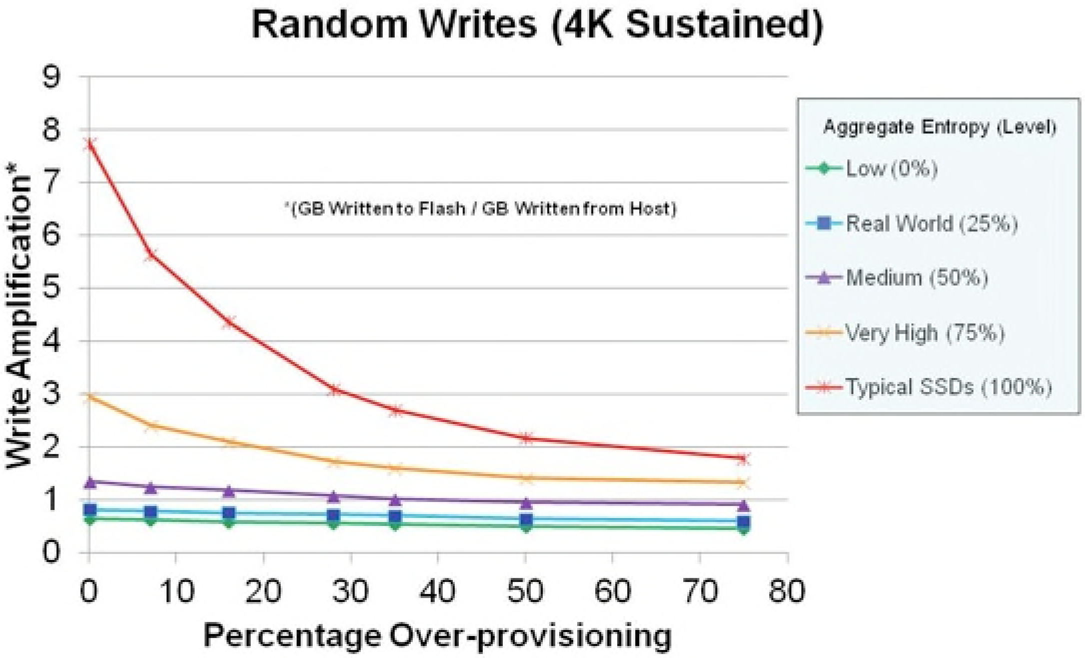


Figure 3. The effect of over-provisioning on write amplification

With the advent of SSDs, and the need to manage them differently from traditional HDDs, a TRIM command was added to storage protocols to enable operating systems to designate blocks of data that are no longer valid. Until the SSD is informed the data is invalid with a write to a currently occupied LBA, it will continue to save that data during the garbage collection process, resulting in less free space and higher write amplification. TRIM enables the SSD to perform its garbage collection and free up the storage space occupied by invalid data in advance of future write operations.

Figure 4 shows the effect of the TRIM command on over-provisioning. For a “marketed” percentage of over-provisioning (28 percent in this example), the amount effectively increases after performing a TRIM operation. Note how the capacity originally designated as Free Space remains consumed as Presumed Valid Data by the SSD after being deleted by the operating system or the user until a TRIM command is received. In effect, the TRIM operation provides dynamic over-provisioning because it increases the resulting over-provisioning after completion.

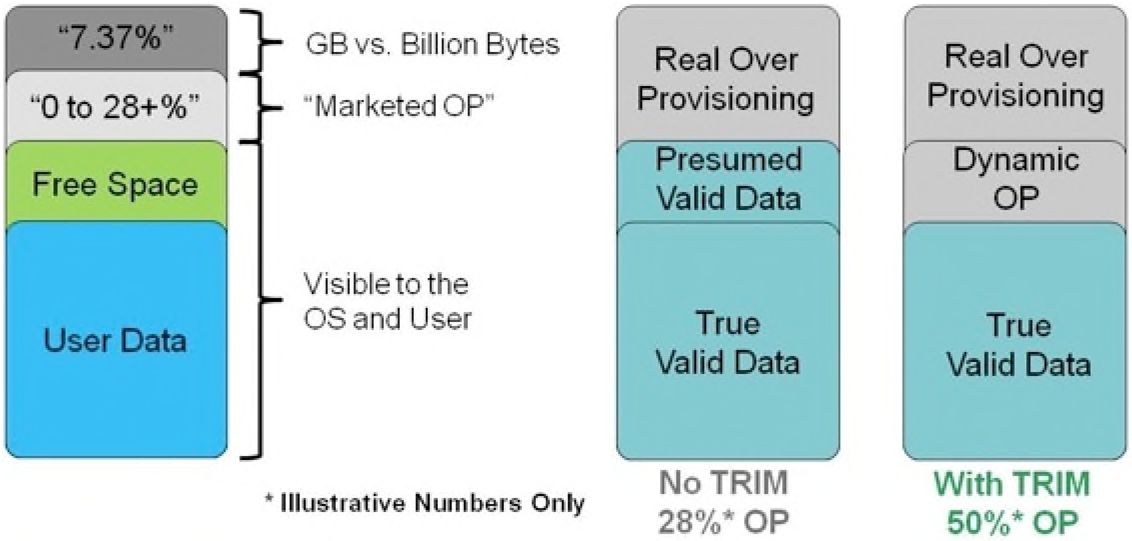


Figure 4. The effect of the TRIM command on over-provisioning percentage

Conclusion

The over-provisioned capacity of NAND flash memory creates the space the flash storage processor needs to manage the flash memory more intelligently and effectively. As shown by these test results, higher percentages of over-provisioning improve both write performance and write amplification. Higher percentages of over-provisioning can also improve the endurance of flash memory and enable more robust forms of data protection beyond basic error correction.

Only SSDs that utilize a data reduction technology, such as DuraWrite in the LSI SandForce flash storage processors, can take advantage of lower levels of entropy to improve performance based on the increase in "dynamic" over-provisioning.

Owing to the many benefits of over-provisioning, a growing number of SSDs now enable users to control the percentage of over-provisioning by allocating a smaller portion of the total available flash memory to user capacity during formatting. With increased capacities based on the ever-shrinking geometries of NAND flash memory technology, combined with steady advances in flash storage processors, it is reasonable to expect that over-provisioning will become less of an issue with users over time.

About the Author

Kent Smith is senior director of Marketing for the Flash Components Division of LSI Corporation, where he is responsible for all outbound marketing and performance analysis. Prior to LSI, Smith was the senior director of Corporate Marketing at SandForce, which was acquired by LSI in 2012, his second company to be sold to LSI. He has over 25 years of marketing and management experience in the storage and high-tech industry, holding senior management positions at companies including SiliconStor, Polycom, Adaptec, Acer and Quantum. Smith holds an MBA from the University of Phoenix.