TECHNOLOGY PAPER

SSD Performance Advancements

Many of us are familiar with the sliding tile game *15-puzzle,* a sliding puzzle that consists of a frame of numbered square tiles in random order with one tile missing. With just one empty square, shifting pieces to new locations requires a fair amount of effort because there is so little free space. Having the puzzle fifteen-sixteenths full makes for challenging and entertaining gameplay, but it's certainly not the kind of performance limitation you want in your solid state drive (SSD). Imagine if the same puzzle were only half-full with eight pieces. It could be solved almost instantly. Obviously, more free space enables faster piece movement and task (game) completion.

Figure 1. 15-puzzle used for representation of how overprovisoning works in SSDs

SSDs work on a very similar principle. Visualize the NAND flash memory inside of an SSD as a puzzle, except that the amount of free space in an SSD is not fixed. Manufacturers utilize various tactics to improve performance, and one of these is to allocate more free space, a process known as *overprovisioning (OP).*

While the minimum amount of overprovisioning for an SSD is set at the factory, users can also allocate more space. Either way, a moderate understanding of overprovisioning is necessary in order to make better SSD purchasing decisions and to configure them in the most advantageous way for each environment.

Background: The Nature of HDD vs. SSD Writes

The fundamental unit of NAND flash memory is typically a 4KB page (the minimum unit to program), and there are usually 128 pages in a block (the entire grid layout of pages). Writes can happen one page at a time, but only on blank (or erased) pages. Pages cannot be directly overwritten. Rather, they must first be erased. However, erasing a page is complicated by the fact that entire blocks of pages must be erased at one time. When the host wants to rewrite to an address, the SSD actually writes to a different blank page and then updates the logical block address (LBA) table. Within the LBA table, the original page is marked as ***invalid*** and the new page is marked as the current location for the new data.

Of course, SSDs must erase these invalid pages of data at some point or the usable space on the SSD would eventually fill up. SSDs, therefore, periodically go through a process called *garbage collection* to clear out these invalid pages of data. During this process, the SSD controller reads all the

good pages of a block (skipping the invalid pages) and writes them to a new erased block. Then the original block is erased, thus preparing it for new data.

Amount of Overprovisioning

All SSDs reserve some amount of space for these extra write operations, as well as for the controller firmware, failed block replacements, and other unique features that vary by the SSD. Typically, 7.37% of memory space is reserved at the factory as a provision for background activities, such as garbage collection.

Even if an SSD appears to be full, it will still have 7.37% of available space with which to keep functioning and performing writes. Often, though, write performance will suffer at this level. (Think in terms of the 15-puzzle with just one free square.)

In practice, an SSD's performance begins to decline after it reaches approximately 50% full. This is why some manufacturers reduce the amount of capacity available to the user and set it aside as additional overprovisioning. For example, 28% may be reserved for overprovisioning. In actuality, this 28% is in addition to the built-in 7.37%, so it's good to be aware of how these terms are used loosely **(Figure 2)**. Users should also be cognizant that an SSD in service is rarely completely full. SSDs take advantage of this unused capacity, dynamically using it as additional overprovisioning.

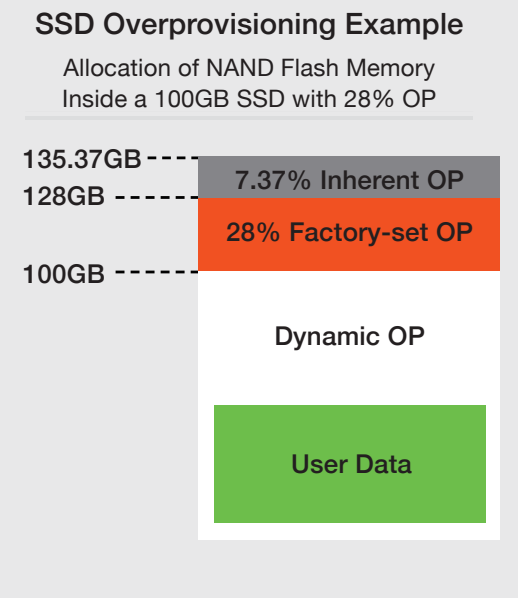

Figure 2. Allocation of NAND flash memory inside a 100GB SSD with 28% OP

Some SSDs include software tools to allow for overprovisioning by the user. Or users can set aside a portion of the SSD when first setting it up in the system by creating a partition that does not use the SSD's full capacity. This unclaimed space will automatically be used by the controller as dynamic overprovisioning.

One obvious drawback to overprovisioning must be addressed. The more unused capacity one reserves to increase writing speeds, the less space is available for actual data.

When an SSD arrives new from the factory, writes will gradually fill the SSD in a progressive, linear pattern until the addressable storage space has been entirely written. Essentially, this reflects an ideal sequential writing condition. No garbage collection has been prompted at this point, and the little pockets of invalid data caused by deletions has yet to impact performance, because there has been no need to write to those pockets with new data.

However, once garbage collection begins, the method by which the data is written—sequentially vs. randomly—will begin to show itself in the performance. Sequentially written data will constantly fill whole blocks, and when the data is replaced, it is generally replacing the entire block of pages. Then, during garbage collection, all of the pages in that block are invalid, and nothing is required to be moved to another block. This is the fastest possible garbage collection—that is, no garbage to collect.

Figure 3. Nytro 1551 SATA SSD with Seagate DuraWrite Technology

What does affect performance is the entropy of the data, provided the SSD is using a flash controller that supports a data reduction technology, such as the Seagate® Nytro® 1000 SATA SSD Series. The entropy of data is the measure of the randomness of that data, not to be confused with it being written randomly vs. sequentially. For example, a completely encrypted data file, MPEG movie, or a compressed ZIP file will have the highest entropy, while database, executable, and other file types will have lower entropy. As the entropy of the data goes down, the write reduction-capable SSD will take advantage of this and provide higher performance. However, the performance remains constant with a given overprovisioning level when written sequentially.

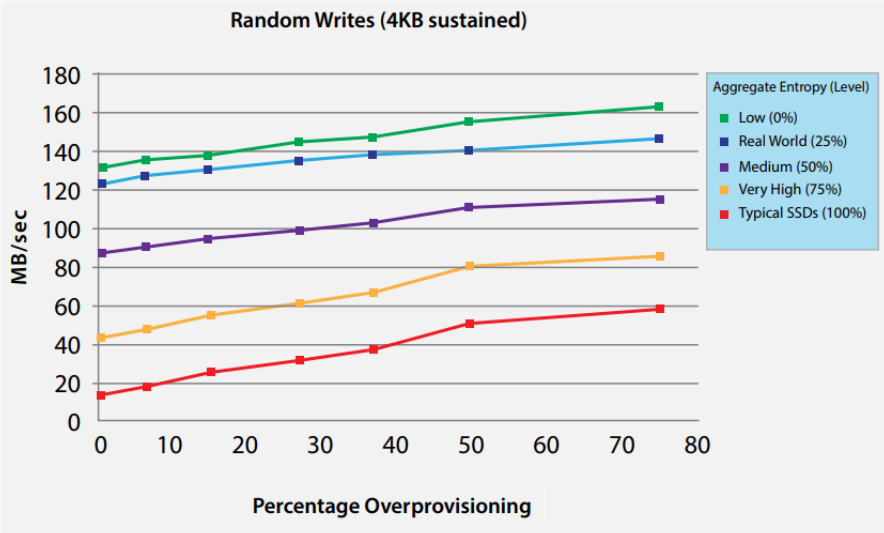

Figure 4. Random writes and overprovisioning

In contrast, when data is written randomly to the SSD, the data that is marked invalid is scattered throughout the entire SSD, so lots of small holes are created in every block. Then when garbage collection acts on a block containing randomly written data, more data must be moved to new blocks before the block can be erased. The red line in the graph above **(Figure 4)** shows how most

SSDs would operate. Note that in this case, as the amount of overprovisioning increases, the gain in performance is quite significant. Just moving from 0% OP to 7% OP will improve performance by nearly 30%. For SSDs with lossless data reduction technology, the performance gains are not as significant, but the performance is already significantly higher for any given level of overprovisioning.

Write Amplification

As mentioned earlier, SSD writes generally involve writing data more than once: initially when saving the data the first time and later when moving valid data during multiple garbage collection cycles.

As a result, it's common for more data to be written to an SSD's flash memory than was originally issued by the host. This disparity is known as ***write amplification***, and it is generally expressed as a multiple. For instance, if 2MB of data gets written to flash while only 1MB was issued from the host, this would indicate a write amplification of 2.0. Observably, write amplification is undesirable as it means that the more data that is being written to the media, then increasing wear and negative performance impacts results by consuming precious bandwidth of the flash memory. Several factors can contribute to write amplification, chief among these being the percentage of data written in a random vs. sequential manner.

Lossless Data Reduction Technology

Surprisingly, it is also possible to write less data to flash than was issued by the host. This would be expressed as a write amplification of, say, 0.5 or 0.7. Seagate DuraWrite™ data reduction technology is probably the most well-known method of accomplishing this through real-time data manipulation. Only SSDs with lossless data reduction technology can create a write amplification of less than one. As the entropy of the data from the host goes down, DuraWrite technology results in less and less data being written to the flash memory, leaving more space for overprovisioning. Without a similar data reduction technology, an SSD would be stuck with higher write amplification.

Note that additional overprovisioning and a data reduction technique such as DuraWrite technology can achieve similar write amplification results with different trade-offs.

Figure 5. Lossless data reduction technology

Competitive SSDs, without a similar technology, are limited to the write amplification from a given overprovisioning level. As an example, an SSD with 28% overprovisioning will exhibit the same write amplification (3.0) as an SSD with DuraWrite technology writing a 75% entropy stream with 0% overprovisioning (all other factors being equal). In other words, this scenario shows how Seagate SSDs equipped with DuraWrite technology could display the same level of write amplification as a standard SSD while reclaiming 28% of the storage capacity.

The Next Efficiency Level

An SSD does not natively know which blocks of data are invalid and available for replacing with new data. It is only when the OS tries to store new data in a previously used location that the SSD will know

a particular location contains invalid data. All free space not consumed by the user becomes available to hold whatever the SSD believes is valid data. This is the reason for the creation of the *TRIM command*. TRIM enables the OS to alert the SSD about pages now containing unneeded data so that they can be tagged as invalid. When completed, there is no need for those pages to be copied during garbage collection and wear leveling. This reduces write amplification and improves performance. **Figure 6** shows just how much difference TRIM can make in allowing more capacity to be available for overprovisioning.

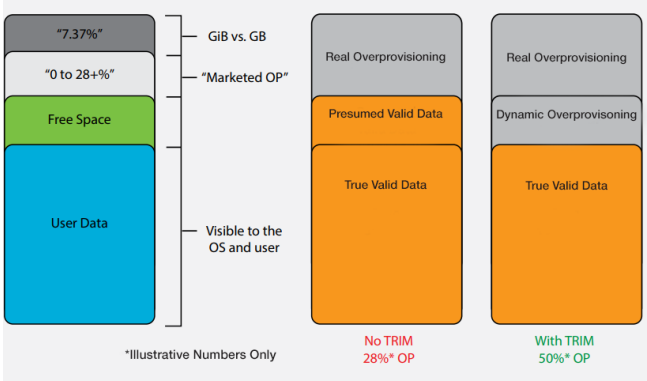

Figure 6. The TRIM command and boosting overprovisioning

TRIM is yet another method that vendors can employ to boost overprovisioning, thereby increasing performance and SSD longevity. It demonstrates a more preferable way to reclaim capacity for acceleration compared to forcing SSDs to permanently surrender large amounts of their capacity. Using TRIM with DuraWrite technology can yield even more impressive results.

Conclusion

Buyers should take a close look at their workloads, assess the typical entropy levels of their data sets, and consider which SSD technologies will provide the greatest benefits for their invested dollars. By reducing write amplification and employing technologies that make SSD operation ever more efficient, such as Seagate DuraWrite, buyers will not only get more storage for their dollar, but the SSD will perform faster and last longer than other options could possibly provide.

AMERICAS Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, United States, 408-658-1000
ASIA/PACIFIC Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapore 569877, 65-6485-3888
EUROPE, MIDDLE EAST AND AFRICA Seagate Technology SAS 16–18, rue du Dôme, 92100 Boulogne-Billancourt, France, 33 1-4186 10 00