NAND Flash Media Management Through RAIN

Scott Shadley, Senior Product Marketing Manager

Micron Technology, Inc.

Technical Marketing Brief

What Is RAIN, and Why Does My SSD Need It?

This brief describes the general concept of RAIN (redundant array of independent NAND) protection in enterprise solid state drives (SSDs) and details how this technology benefits Micron's RealSSD® products.

Existing Protection Techniques in Enterprise Storage

In the early days of using hard disk drives (HDDs), it became apparent that a single device had enough potential for failure to warrant the creation of a protection scheme to prevent unwanted data loss.

A protection scheme was then developed called RAID, or redundant array of independent (inexpensive) disks. This mechanism for protecting data came in many different forms, but the most common mode of RAID used today is RAID 5, which provides not only data protection, but also performance enhancements. Figure 1 shows an example of how RAID 5 works.

In this representation, data is moved around within four unique disks. The parity (in green) represents the data required for the system to recover any of the actual blocks of stored information. However, this parity occupies space on the disks, and therefore affects the overall capacity of the system.

This is an important point to remember as we begin exploring how this technique and protection scheme can be applied to an SSD. Note also that the parity moves around within the architecture, a key benefit in how it can be useful in SSD architectures.

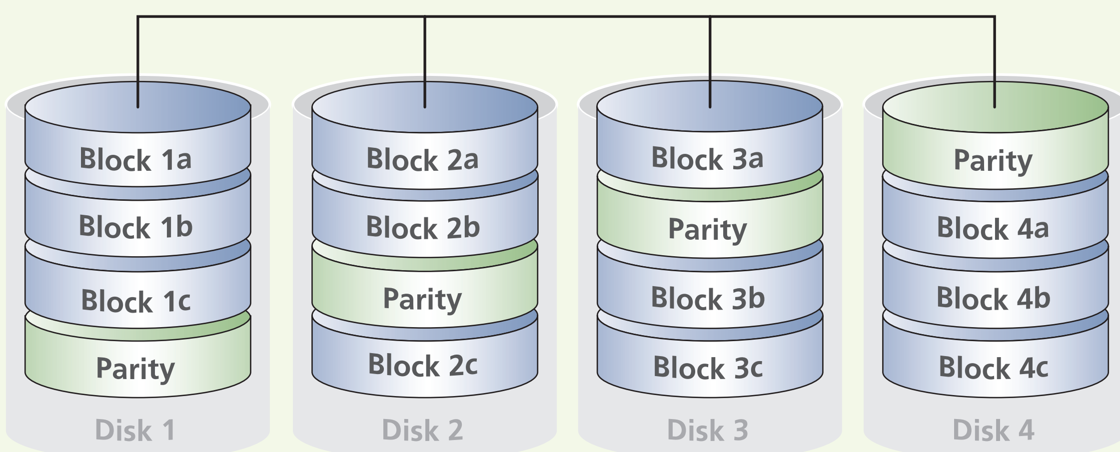

Figure 1: RAID 5 in HDD

ECC Is No Longer Enough to Protect NAND in SSDs

Like HDDs, SSD devices also have a potential for failure and data loss. The original method for protecting data in SSD designs was simply to add the required levels of ECC to the given pages, and then recover data using ECC. When the first SSDs were shipped, this initial level of protection was sufficient to reduce the chance of failure, much like that of the early HDD products.

But as NAND geometries continue to shrink, and controller complexity increases, there is a need to move beyond simple ECC, just like when RAID was invented for HDD.

With HDDs, RAID technology grew beyond the simple initial version RAID 0, and RAID 2, 3, 4 followed. However, these did not prove to be the best implementations, and continued development lead to the RAID 5 protection technology of today.

Similar solutions to protect the NAND are emerging as the market matures, but, unlike the HDD market, in which the industry collectively settled on a specific version of RAID implementation, SSDs utilize approaches at the device level. Thus, no two versions of RAIN-like technology are likely to be the same. With many competitive architectures being created by many SSD manufacturers, it is important to document the details of the approach being used.

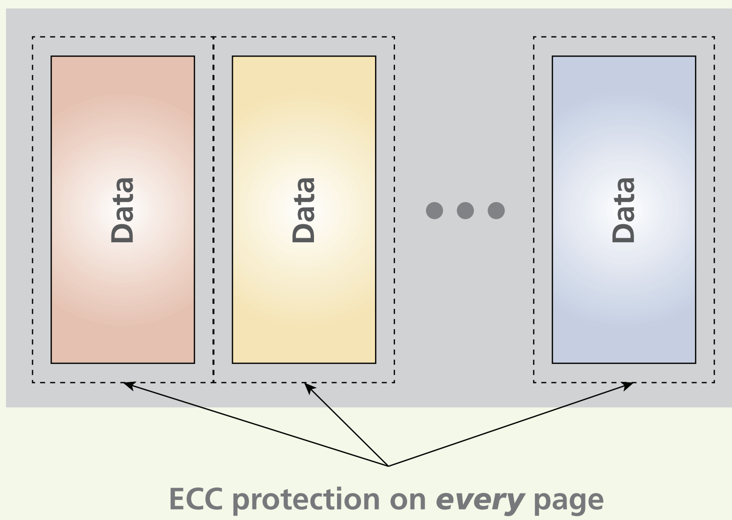

Figure 2: Original method of protection added the required levels of ECC

SSD Architecture Review

To show how RAIN can be beneficial in SSD design and also the relevance of RAID 5, we need to look first at how data can be organized in the SSD. Unlike HDD designs that have rotational storage organizations (e.g. platter, track, sector), SSDs utilize Flash devices as the storage media, enabling some new design flexibility.

The SSD layout in Figure 3 shows some direct similarities between how SSD NAND devices are arranged compared to the multiple-drive arrangement of HDDs in RAID 5. This high level of similarity suggests possible integration of RAID into an SSD single device rather than using many individual drives. This represents the first level at which RAIN comes into play within the Micron RealSSD product.

The SSD controller uses parallelism to move data within the SSD in order to increase performance and locate the stored data across the many smaller devices. For example, to create a 350GB SSD, smaller 8GB NAND devices are used to create this capacity.

This is best represented in the HDD environment as a RAID 0 architecture, also known as striping data.

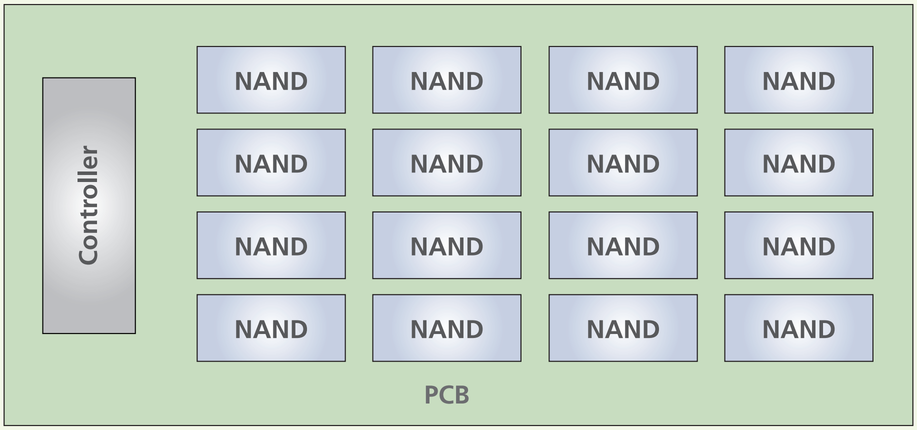

Figure 3: NAND device arrangement

Now that we've covered the board-level layout of an SSD, the next step is to look at the data architecture inside a NAND component. There are many ways to organize the data, and each controller's architecture and firmware can be unique in its implementation. Moving data the most efficient way possible is the key to a successful design. The better the methodology, the faster and more compelling a user will find the end product.

The diagram below shows an example of how the data structure inside a NAND device, as well as the overall SSD, can be designed to take advantage of the RAID 0 architecture.

Figure 4 shows how the controller breaks down the SSD into progressively smaller parts from the device level—represented here as a logical unit, or LUN. These LUNs are then collected in series to create a super LUN.

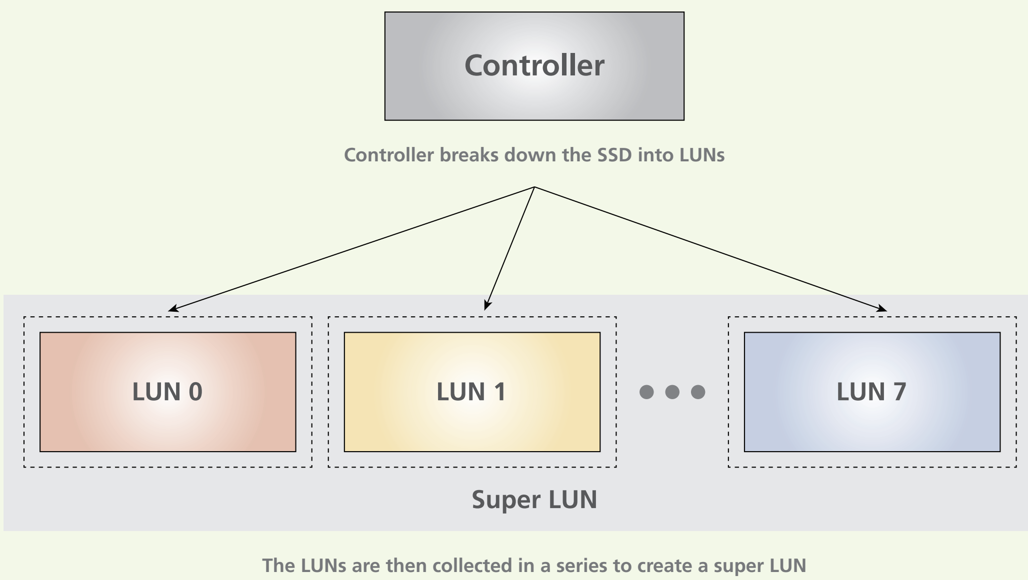

Figure 4: Data architecture inside a NAND component

Within a super LUN, the controller takes individual blocks inside the NAND devices and creates a striped super block. In turn, this super block can also be broken down into super pages.

The super page, or group of individual NAND pages, is what we will be using to discuss how the addition of Micron's RAIN architecture is used.

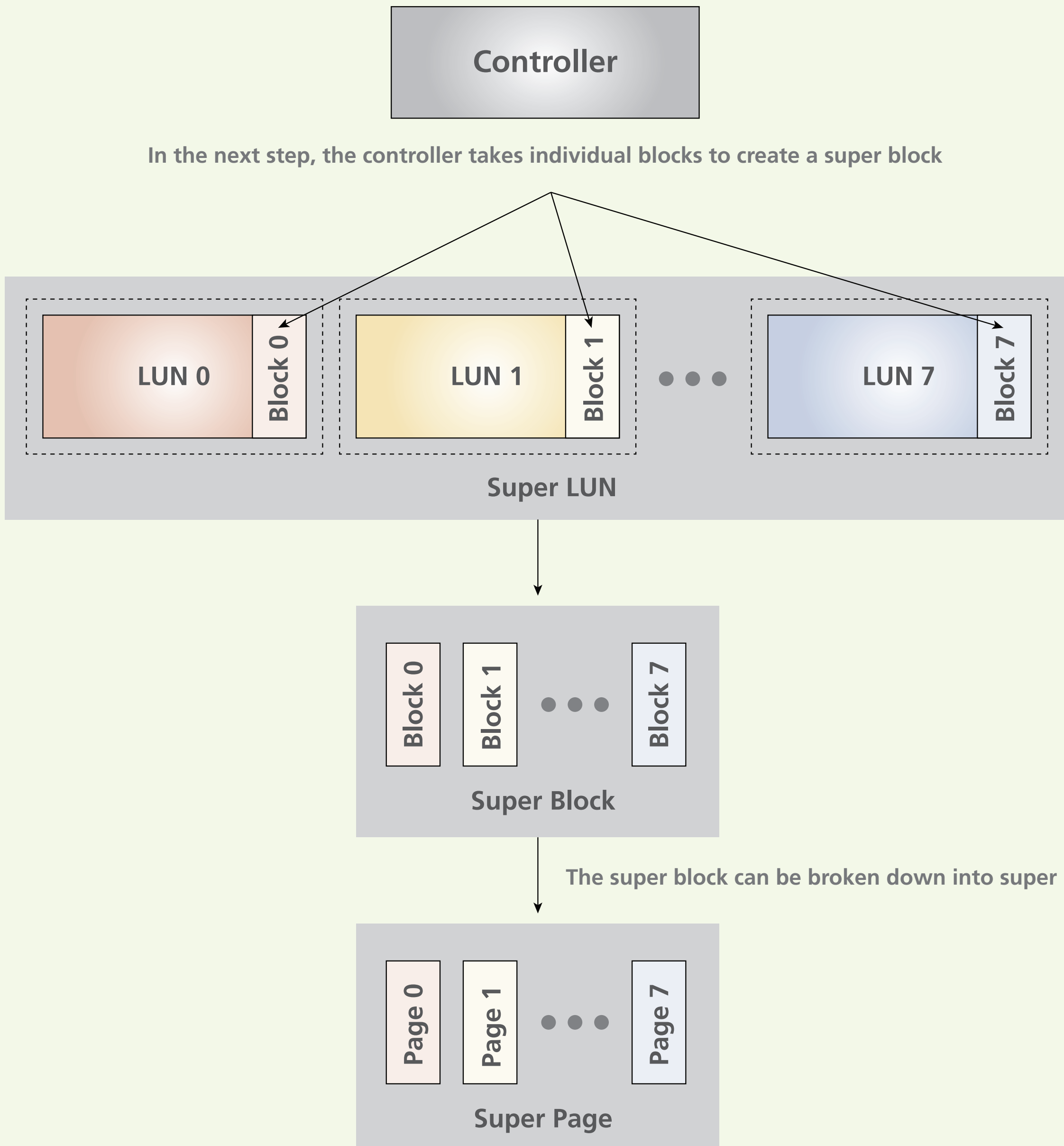

Figure 5: Controller creates super blocks, then super pages

How Does RAIN Work in Micron RealSSDs?

RAIN technology is a method of adding protection beyond the ECC of today and also taking into account the best way to manage the NAND devices. With striping or parallelism already in place within the SSD, taking advantage of a RAID 5-like architecture seems to be the most logical data-protection choice. However, there are many possible ways to implement this parity- protection scheme. There are trade-offs that must be considered when deciding the best way to use these architectures.

Figure 6 shows a hypothetical example of how implementing RAIN can impact the end-user capacity of the drive—the amount of NAND available to the user can be affected by the addition of these protection schemes. As in the HDD environment, where this is already standard practice, having NAND capacity consumed by a protection scheme is an acceptable trade-off for SSD end users in order to have access to these features.

In this example, assume the following:

- RAW NAND capacity of the SSD = 512GB = “X”
- RAIN design = 1 parity element for each 7 storage elements = “Y” = 7/8 (a 1:7 ratio) = 0.875
- Over-provisioning level = “Z” = 0.78 ((100-22)/100)

We then would have the following:
User-available capacity = 512GB x 0.875 x 0.78 = 350GB

Now let’s look at the path to creating a proper RAIN recovery from NAND as used in Micron RealSSD products like the P320h PCIe drive.

The structure of the data is shown in Figure 7—there is a planned or fixed architecture per-channel of the SSD to ensure proper data management. The “P” represents the parity page being used.

In the following sections, we will use this representation to show how a data failure can take place, be recovered, and how the new data structure will be intact and can be carried forward through the life of the SSD. This discussion will focus on a 7+1 RAID 5 architecture.

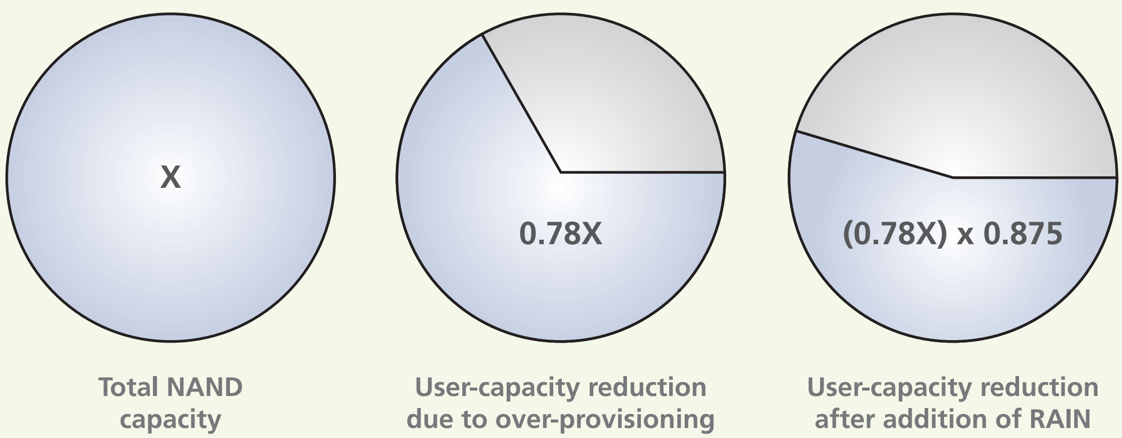

Figure 6: How implementing RAIN can affect the end-user capacity of the drive

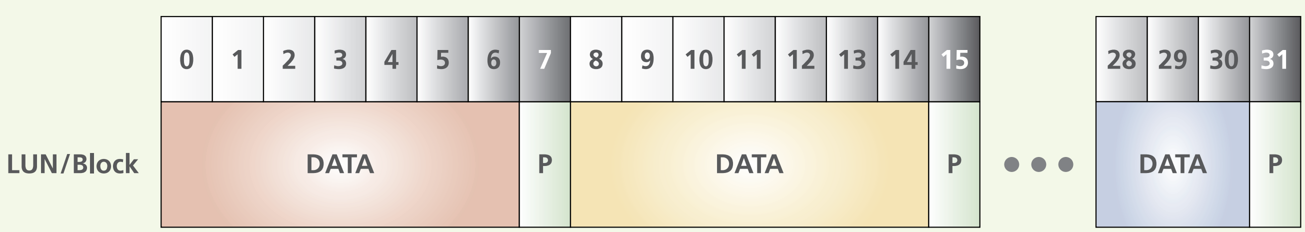

Figure 7: Fixed architecture per channel of the SSD to ensure proper data management

Many different parity stripes can be considered and implemented, with each having a unique impact on the overall design, capacity, and possibly the performance of the drive. In the case of Micron SSD firmware development, the differences in the performance of these data stripes were carefully characterized to determine the best performance of the SSD and the best protection method, as required by the media being used. Table 1 represents how the data stripes are handled by the RealSSD P320h PCIe drive.

Attribute	**256GB Raw Capacity SSD**			
	No Parity	*7 + 1P*	*3 + 1P*	*1 + 1P*
Sequential Read @ 64KB transfer size	1.5 GB/s	1.5 GB/s	1.5 GB/s	1.2 GB/s
Sequential Write @ 64KB transfer size	1.5 GB/s	1.5 GB/s	1.5 GB/s	1.2 GB/s
Random Read @ 4KB transfer size	375k	375k	375k	300k
Random Write @ 4KB transfer size (full drive)	171k	171k	143k	94k

Table 1: How data stripes are handled by the RealSSD P320h PCIe solution

As Table 1 shows, only the RAID 1 version of the parity architecture has an impact on performance. Because this method of protection consumes 50% of the NAND in parity, it does not provide a viable choice for SSD design.

The deciding factor for the remaining methods is which one is required by the NAND being used. The larger the data stripe (number before parity), the greater the possibility that a data failure may still take place. In the Micron SSD case, there is no loss of performance at a 7+1 level, and this still provides enough additional protection for the NAND in the current solutions. As the media changes, and there is a need for higher levels of data protection, this choice will be re-evaluated.

Now that we've established how RAID works in HDD and why ECC is not sufficient for proper data recovery in NAND-based SSDs, next we'll discuss how the complex algorithms of the Micron RealSSD implementation of RAIN actually work to recover data. RAIN is a redundant array of independent NAND, and taking this definition literally is the most efficient way to discern what is occurring in the SSD. While most simple SSD designs are focused on parallelism to the point of failure, Micron has developed a solution that ensures independence as well as parallelism. To illustrate, we will take an example of the P320h drive from Table 1 and walk through a failure and recovery.

An Example of Failure Recovery with RAIN

For the sake of clarity, the use of a smaller stripe of data with only seven pages will be used, or, in this case, a simple RAIN super block. (Additional details and unique features of RAIN and the SSD firmware management are not being discussed here due to the deep technical aspects and Micron-specific IP. These details are available through more direct discussions with Micron.)

In the following example, if we refer to each page number shown as an individual data point, we will examine whether the loss is focused on these super pages. In actual drive applications, the level of data recovery can also be on either a smaller, individual NAND page or a much larger failure recovery.

While data is being moved around within the super pages and super blocks through either host-initiated data writes or through drive level wear leveling and garbage-collection activities, RAIN will successfully recover data, as shown in Figure 7.

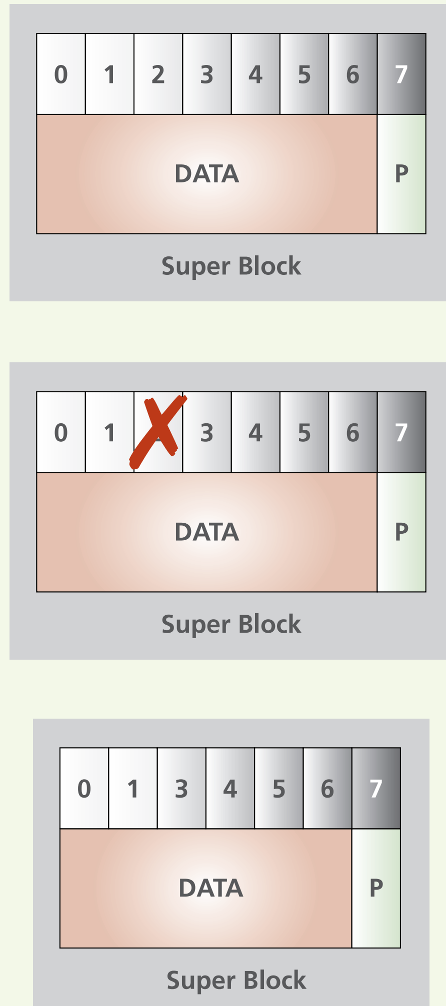

Figure 8: RAIN data recovery during data moves

1. As data being written to the drive goes through the process of being programmed, data is arranged so that parity can be created and stored in the eighth page of the super block. Over time, as ECC is correcting data, there is no need for RAIN to be applied for any form of data recovery. RAIN is focused on defects that occur due to unexpected NAND issues, or as the device wear grows, and ECC limits are reached.

2. As data is read from the drive, it is processed through the RAIN engine. In the event that any portion of the super block is flagged as an error, the primary RAIN algorithm kicks in and recovers the data via the parity process of using an XOR algorithm on the failed data. In the event that this data failure is a READ process only, the newly recovered data is then stored in a new super block location on the drive, with other data being moved around in the drive as part of the normal wear-leveling exercise. Instead of simply moving existing data and replacing with new write data, the newly recovered data is instantaneously moved into that wear-algorithm recovery.

3. Note that the existing super block where the failure had occurred is not instantaneously moved around within the drive. That type of recovery is used in competitive solutions, and, although effective, it creates unnecessary wear and write amplification on an already-wear-restricted product. The Micron implementation of RAIN has the ability to leave the existing super block intact until it is absolutely required to reprogram the entire block. This now smaller super block has no impact on performance or drive data management, a key feature of the independent nature of the solution.

4. Data is then gathered by the controller to indicate that this recovery has occurred. It is tracked in the internal logs of the drive, and in some cases it can be displayed in SMART attributes, depending on the drive and level of implementation needed.

Conclusions

Data integrity is a key attribute that enterprise customers require, and potential customers prefer the assurance that nothing will be lost in their data stores.

Current solutions are focused heavily on data reliability. With HDD solutions, RAID is not optional, but *required*, to ensure this reliability. These solutions focus on ways to store more data in smaller, redundant locations. The continued use of tape-based storage design is another example of the need for secure data storage.

As SSDs have moved into enterprise data environments, they no longer fit the definition of RAID where the "I" stands for inexpensive. Therefore, implementers need solutions that do not require them to create data sets that need redundant "safety" drives. Finding ways to solve this unusual situation can best be addressed by looking *inside* the SSD, rather than at ways to add layers on top of the use case.

Micron

The creation and focus of Micron's RAIN is to provide this internal solution to customers. SSDs are certainly more robust than HDD solutions currently in use. But, as with all technology, there is always a small chance to experience issues. RAIN is that next layer within the SSD that provides customers the peace of mind to implement solutions with fewer SSDs, yet still have existing HDD RAID protection levels.

Technology will continue to advance inside the SSD and outside in the end-storage solutions. Micron's goal and focus is to ensure that features like RAIN are always one step ahead of customers' growing needs for cost-effective data storage solutions with high reliability and endurance.