An In-Depth Look at Variable Stripe RAID (VSR)™

A white paper by Robbie Stevens, Senior Marketing/Technical Writer

Texas Memory Systems, Inc.

10777 Westheimer Rd.

Houston, TX 77042

www.RamSan.com

Scan this QR code with your smartphone for more information on Variable Stripe RAID.

Download a QR scanner here: http://2dscan.com

Table of Contents

Introduction

Variable Stripe RAID (VSR)™ is a new feature from Texas Memory Systems (TMS) that reduces the risk of uncorrectable errors occurring in user data on a NAND Flash chip while preserving as much storage space as possible. VSR maximizes chip storage space since it is not necessary to relocate an entire stripe following a failure.

TMS now offers products with VSR incorporated into their Flash controllers, allowing stripes containing bad planes to be seamlessly rebuilt and efficiently relocated. By dynamically altering the stripe size, VSR can drastically reduce wasted planes.

Background Information

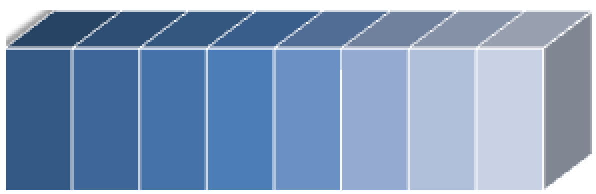
Flash Architecture

Flash is a type of non-volatile, solid state storage technology. In enterprise applications, multiple Flash chips are used together to produce solid state disks (SSDs) in the form of external rackmount systems or internal cards or drives.

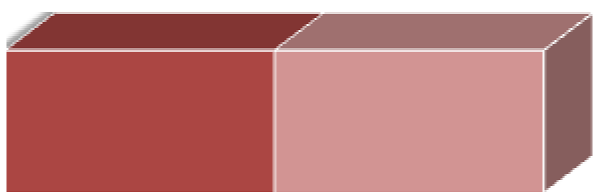
A Flash memory chip is divided into multiple nested entities. The 32nm Toshiba Flash chips used by the TMS Series-7 RamSan systems are divided as follows:

One chip:

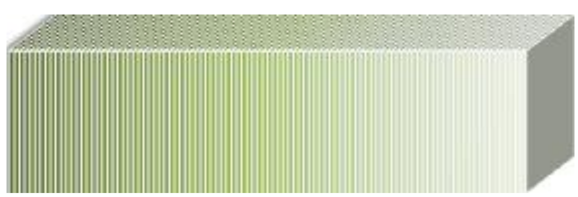
is divided into 8 dies:



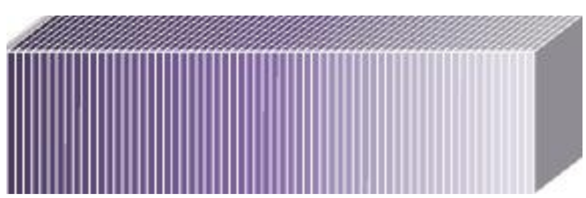
One die is divided into 2 planes:



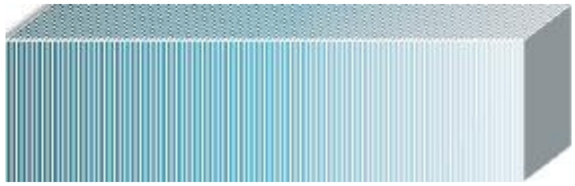
One plane is divided into 2,048 blocks:



One block is divided into 64 pages:



One page is divided into 8,192 bytes:



In total, each chip contains 8 dies, 16 planes, 32,768 blocks, 2,097,152 pages, and 17,179,869,184 bytes!

Bad Blocks Lead to Plane Failure

A block may fail for a number of reasons. While many of these errors are automatically corrected using an error-correcting code (ECC), sometimes errors occur that ECC cannot handle. For example, a block may be worn past its usable life or data may be unintentionally altered through instances called read-, write-, and erase-disturb errors.

If a block cannot be erased or written to, or if a read fails with uncorrectable ECC errors (that is, ECC cannot correct errors because too many bytes have failed), the PowerPC CPU in the Series-7 Flash Controller marks the block as bad and attempts to recover its data, if possible. Plane failure occurs if enough blocks within a plane go bad (256 for TMS systems). Failed planes can no longer be used to store data.

RAID Arrays

RAID (Redundant Array of Independent Disks) arrays are a popular way to mitigate the risks of losing data on an SSD in the event of a failed chip. In many RAID implementations, data is striped across all chips in an SSD, and if one chip fails the information on the remaining chips can be used to rebuild the missing data.

The types of RAID arrays used in RamSan systems stripe data across all chips. For example, a RamSan-70 uses 40 chips and four Flash controllers in its base configuration. Each RamSan-70 Flash controller connects to ten Flash chips and is initially set up to use a 9+1 rolling XOR RAID across all ten of its Flash chips. Nine data pages within each RAID stripe are protected by a tenth page of additional information that can be used to rebuild lost data in the event of uncorrectable data errors.

Prior RamSan RAID Implementation

All RamSan systems operate as RAID arrays (in most cases, they have used RAID-5). If a bad plane was detected, the controller automatically rebuilt the missing block and moved all data from that stripe to a new location. All blocks used by that stripe were then deactivated.

Many RAID-5 controllers automatically detect plane failure and rebuild the data elsewhere much like VSR does, increasing data stability. However, the reduction in usable capacity within one device limits the overall number of stripes that can be created within the entire group. If a single device loses one eighth of its available blocks due to a failure, then all other devices within the RAID group also see a capacity reduction of one eighth. Once all available capacity in the failing device has been utilized, then no more stripes may be created, even if unused capacity still remains within the functioning devices. In the case of a 9+1 stripe, one bad plane results in nine good planes going to waste. In an extreme case, an entire chip may fail. Such a condition prevents the creation of any new stripes, and renders the entire RAID group unusable. Over time, the number of good planes "trapped" in bad stripes can reduce storage capacity by a significant amount.

Variable Stripe RAID

When a bad plane occurs, VSR technology allows a plane to be removed from use without impacting the available capacity of other devices within the RAID stripe. Upon detection of a failure, the failing plane is removed from use (no further writes are allowed) and all used pages within the affected stripe are marked as "critical to move". Information from the affected stripe is then gradually relocated to known-good stripes, a process that is performed as a background task to minimize the required processing power.

Unlike traditional RAID, the loss of capacity in one device does not affect other devices. The functioning devices—those having more available capacity than that of the failing device—simply create stripes with fewer pages (such as 8 +1 instead of 9+1) to work around the failing plane. Even in the event of complete chip loss, new stripes may be created across the remaining good chips and the failing chip is simply taken out of circulation. With VSR technology, device failures result in far less capacity loss than traditional RAID.

Advantages vs. Traditional RAID

In a traditional RAID-5 array, a stripe can only survive the failure of one plane used by the stripe. If a second plane in a stripe fails and the data has not been relocated by the controller, the entire stripe is corrupted, resulting in data loss. In addition, if the stripe is relocated by the controller, the planes used by the original stripe are deactivated.

For example, suppose one plane in a 9+1 rolling XOR RAID-5 array fails (in this case, plane 6_1):

1_1 2_1 3_1 4_1 5_1 **6_1** 7_1 8_1 9_1 P_1

When the failed plane is detected, the data is rebuilt and moved to a new stripe, along with the other data in the stripe:

1_2 2_2 3_2 4_2 5_2 **6_2** 7_2 8_2 9_2 P_2

The original ten planes can then no longer be used:

~~1_1 2_1 3_1 4_1 5_1 6_1 7_1 8_1 9_1 P_1~~

In a VSR array, when a plane fails, the data from that stripe is rebuilt and moved to a new stripe in a similar fashion. However, the failed plane is removed from the original stripe, allowing the stripe to be used for new data. While the failed plane cannot be reused, the other nine planes are retained and additional data can be striped across them. For example, if the stripe was originally set up using a 9+1 rolling XOR RAID, once the failed plane is removed the stripe functions as an 8+1 rolling XOR stripe.

Suppose one plane in a 9+1 rolling XOR VSR RAID array fails (again, plane 6_1):

1_1 2_1 3_1 4_1 5_1 **6_1** 7_1 8_1 9_1 P_1

When the failed plane is detected, the data is rebuilt and moved (along with the other data in the stripe) to a new stripe, similar to RAID-5:

1_2 2_2 3_2 4_2 5_2 **6_2** 7_2 8_2 9_2 P_2

However, VSR striping allows the original stripe to be reused. Only the failed plane is deactivated, resulting in an 8+1 rolling XOR VSR RAID stripe for new data:

1_1 2_1 3_1 4_1 5_1——~~**6_1**~~——7_1 8_1 9_1 P_1

Furthermore, the new 8+1 RAID stripe can survive another bad plane. Suppose a plane fails in the new stripe:

1_1 **2_1** 3_1 4_1 5_1——~~6_1~~——7_1 8_1 9_1 P_1

The stripe is rebuilt and moved to a new location, and the remaining 7+1 stripe is made available for new data.

1_3——~~**2_3**~~——3_3 4_3 5_3——~~6_3~~——7_3 8_3 9_3 P_3

Since VSR salvages the good planes in a stripe containing a failed plane, much more storage space is retained.

Advantages vs. Competition

This technology is only available from Texas Memory Systems, a company committed to producing the fastest, most reliable enterprise storage available. TMS uses only single-level cell (SLC) Flash chips, which have a much longer (10x) expected life than the multi-level cell (MLC) chips used in competing products. By reducing space lost to failed planes, VSR technology is another way TMS extends the usable life and value of its products.

Conclusion

Variable Stripe RAID technology from Texas Memory Systems extends the usable life of its storage solutions, helping enterprise SSD users retain as much storage space as possible while protecting the integrity of their data. When VSR striping is employed, data from a failed plane is rebuilt and the plane is relocated. More importantly, the stripe containing the failed plane is not discarded; instead the failed plane is simply removed from the stripe, allowing new data to be written to the stripe. Since VSR technology automatically and dynamically adjusts stripe width following plane failure, this new RAID strategy can reduce wasted planes and efficiently protect valuable data.