



## SAS Raid-on-Motherboard: Affordable, High Performance RAID

SCSI Trade Association White Paper  
By Paul Griffith, Broadcom

Full-featured RAID data protection is becoming a standard feature in businesses of almost any size, thanks to the increasing affordability of implementing RAID technology. Cost-effective RAID-on-motherboard (ROMB) solutions, using integrated RAID-on-Chip (RoC) devices, enable system integrators to implement robust hardware RAID solutions while optimizing their server motherboard investments. ROMB offers higher performance and stability at a lower cost than host bus adapter (HBA) RAID, Zero-Channel-RAID, and modular RAID-on-motherboard solutions. As a result, the RAID-on-motherboard system gives system integrators an ideal hardware RAID solution for servers, requiring the optimum combination of reliability, availability and affordability.

As the successor to parallel SCSI, Serial Attached SCSI (SAS) has emerged to deliver higher levels of reliability than previous generations for mission-critical transactional applications that require 24/7 online access with no data loss. The highly scalable and flexible architecture of SAS enables RAID topologies to support multi-node clustering for high availability failover or load balancing. In one of its most significant advances, the SAS interface is also compatible with lower-cost Serial ATA (SATA) drives. This enables system builders the flexibility of integrating either SAS or SATA devices while dramatically reducing the costs associated with supporting two separate interfaces.

With the union of high-security, cost-effective ROMB and the built-in reliability and availability features included with SAS, system IT managers can now meet the data security requirements of tomorrow on the restrictive budgets of today.

### Typical RAID Components

To appreciate the differences and benefits of the various available RAID system implementations, one must first understand the major components that comprise a RAID subsystem.

- **I/O Processor**

The main component of the RAID subsystem is the I/O processor (or IOP). RAID software is executed within this processor to manage such tasks as disk virtualization, cache processing and logical volume configuration. With this dedicated RAID functionality, the architectural design of the IOP typically provides a bigger impact on the performance of the RAID code, which is designed specifically for the IOP rather than incorporating a higher clock frequency. The IOP frees the host CPU from interim interrupts generated by I/O requests to member disks that are configured in storage arrays. It only interrupts the host CPU one time per I/O request, regardless of the total number of disks that reside within the storage array.

Typically, the IOP is the only component in the RAID subsystem known to the host, as all other RAID components are kept hidden. When the host identifies a RAID

subsystem, it only needs to recognize and communicate with the I/O processor. This level of abstraction effectively hides RAID traffic above the IOP by not presenting it to the system or host.

- **Dedicated XOR Engine**

A single disk can protect data on any number of associated disks by performing a simple Boolean XOR operation. For example, under RAID-5, the XOR engine creates parity data that is the result of an XOR operation on all other data elements within the same stripe. The XOR function can be implemented in dedicated hardware such as an XOR ASIC engine or as part of the I/O processor that may have integrated XOR functionality. This dedicated function within the hardware greatly increases the throughput of data requiring this operation.

- **I/O Controller**

The I/O controller (or IOC) communicates directly with individual disks. When the IOP makes a request for data, it directs the IOC to retrieve the data from a physical disk and return the data to either the IOP or host, depending on direction from the application. The IOP never communicates directly with the disks, leaving that task to the IOC, which can support one of several disk interface technologies such as Fibre Channel, Serial ATA (SATA) and SAS.

- **Cache Memory**

Fast cache memory is also used for host communications, RAID algorithms and temporary data storage between the host and disk drives that is governed strictly by the IOP. Various cache size allocations can positively (and negatively) affect the performance and stability of the application. Since RAID code is executed within cache memory, the code is protected from outside device drivers or applications that run within the host memory and operating system.

- **Flash Memory**

RAID code uses Flash memory for storage as it is preserved even when the power is off. When the system boots up, the IOP retrieves RAID firmware from the flash memory so that it can execute IOP tasks in faster cache memory

- **Battery Backed Cache**

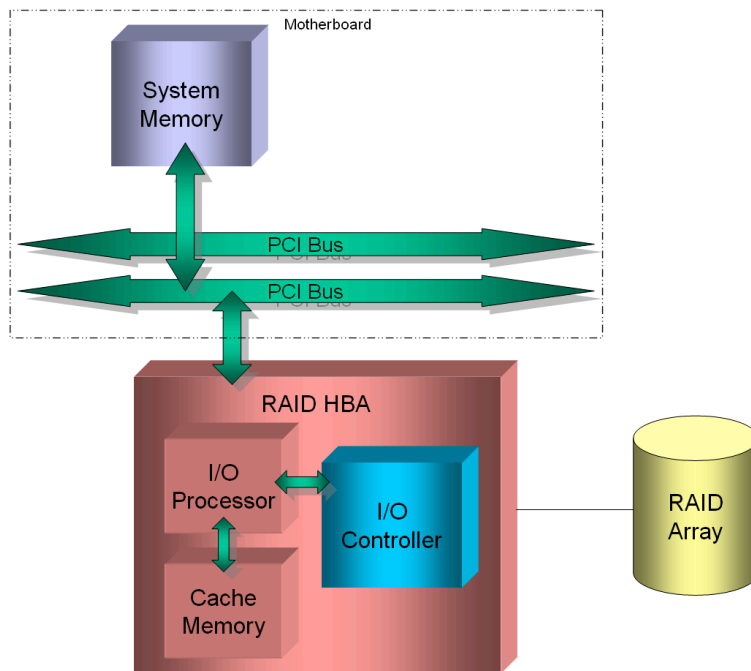
Once a data write is completed, the write cache must be able to protect the data in the event of a power loss before the write data is moved to the drives. If external power to the system is interrupted, if the host computer fails, or if the RAID controller fails, battery power will maintain data in the cache long enough for the user to recover the data and preserve the integrity of business-critical information.

## Comparing RAID Implementations

As intelligent RAID technology continues to grow in popularity, several different types of RAID implementations must be considered to achieve the greatest balance of performance, flexibility and reliability, all at an affordable cost. At first glance, the idea of implementing RAID directly on the motherboard of a server may seem to have some drawbacks; however, by comparing this with other typical implementations, ROMB designs are clearly at the forefront of achieving high performance, reliable RAID at a comparatively lower cost. The following describes the three most popular RAID implementations that are currently available and includes benefit comparisons for each.

### Host Bus Adapter Cards

Host-bus adapter expansion cards that plug into servers through a PCI-X® or PCI Express® connector are the most common way to implement RAID in a server system. The HBA method offers at least one advantage that the integral ROMB does not. That advantage is flexibility. To upgrade a RAID controller, the HBA can be removed by simply replacing the adapter card. Though various RAID vendor choices are available, most RAID applications from different vendors are not currently interchangeable, although an effort is underway to try and alleviate this. When HBAs are incompatible, administrators are forced to back up all of their data, replace the HBA, re-create the previous RAID volumes, and then restore all of the data and applications.



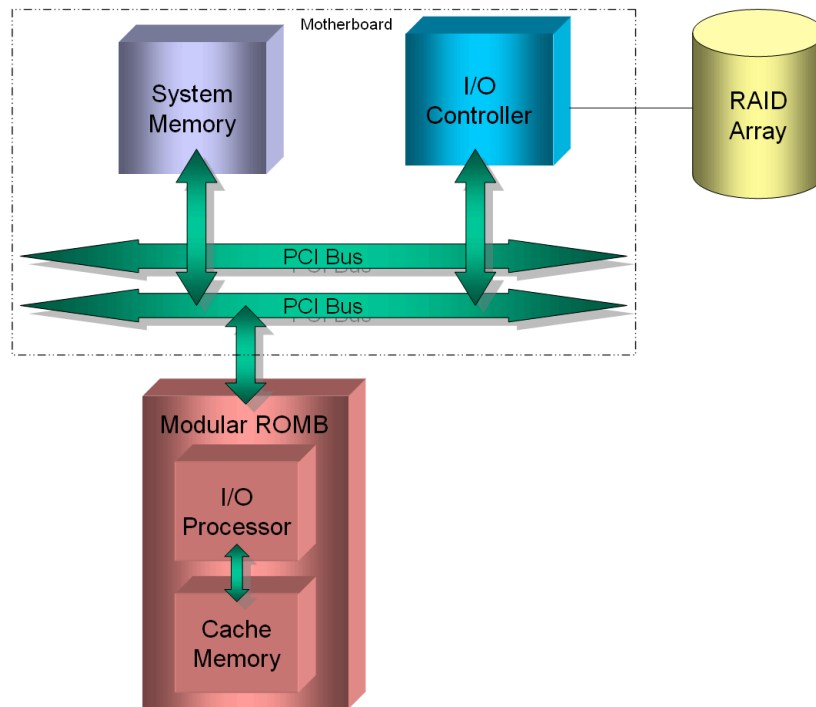
**Figure 1: Typical RAID HBA Implementation**

Performance can also be affected when inserting an HBA into a PCI slot that is shared with other components within the system. Although most servers have multiple PCI buses, administrators may have difficulty determining which components are sharing buses with other slots or devices. Two devices sharing the same bus can easily create a bottleneck in the system and slow down performance. HBA add-in cards are typically the most expensive option when implementing RAID into a server platform.

### Zero Channel RAID (ZCR) or Modular ROMB

ZCR addresses some of the drawbacks of using an HBA card. Essentially, modular ROMB implementations still use all of the same components of an HBA card, however they are located differently. With modular ROMB, the IOC is not located on the PCI add-in card; but instead, is embedded directly on the motherboard. A modular RAID HBA is still required to complete the

RAID subsystem. However, since the IOC is already available on the motherboard, no IOC is necessary on the HBA. Special additional logic is still required on the motherboard to allow for proper interaction between the IOP and IOC when the modular HBA is present.



**Figure 2: Typical ZCR / Modular ROMB Implementation**

Modular ROMB was created to provide a less expensive RAID alternative to the HBA card. Modular ROMB costs less by taking full advantage of the IOC that already exists on the motherboard so no IOC is needed on the HBA. Unfortunately, modular ROMB can suffer from other drawbacks including compatibility problems, loss of the I/O controller, flexibility and performance.

Some previous ZCR systems have failed to completely hide the IOC from the host, instead, leaving it up to the end-user to disable certain BIOS settings and avoid loading certain device drivers. Otherwise, the IOC could be the source of conflicts between the ZCR card and a host application.

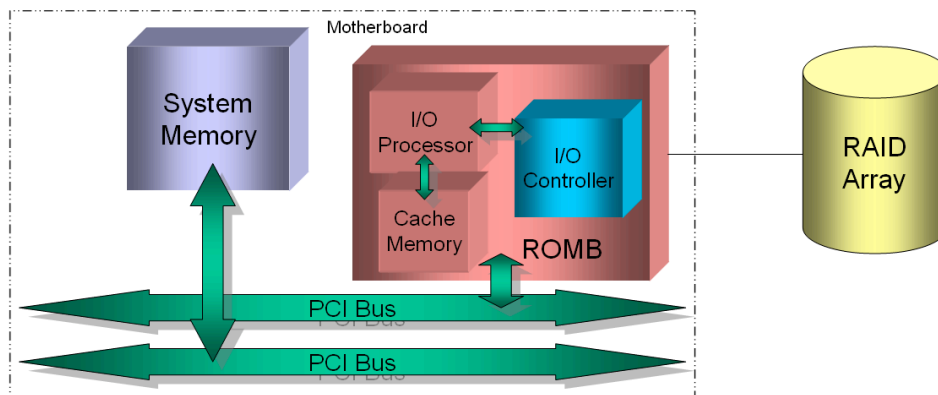
When a modular ROMB card is inserted into the system, the on-board I/O controller can no longer be used for simple drive connectivity, the likely reason it was originally there. If administrators require the embedded IOC for other purposes, they may not have the ability to upgrade the system to modular ROMB.

Modular ROMB cards require special slots designed for the particular card. They cannot be added to any slot in the system. These special slots allow the cards to correctly interact with the system and use the on-board IOC.

Finally, modular ROMB performance can suffer because of the IOC placement on the motherboard. This configuration presents a data flow pattern that can double the amount of traffic on the PCI bus compared to a RAID HBA card or embedded RAID chip on the motherboard. Although modular ROMB systems are designed knowing that this limitation exists, the data flow may still create a system bottleneck.

### **RAID-on-motherboard (ROMB)**

ROMB places every piece of the RAID subsystem directly on the motherboard with most components available in a single ASIC. This approach provides high performance and stability. The primary disadvantage of ROMB is the lack of hardware upgradeability, although features and performance can generally be upgraded in firmware. The flexibility of HBA-based RAID has long been challenged by the simple price advantage of ROMB. By embedding the HBA directly on the motherboard, the overall solution price can be substantially less than what it would cost to buy components separately.



**Figure 3: Typical ROMB Implementation**

System designers can optimize the motherboard layout for the RAID subsystem since the components are fixed with regard to each other, the PCI buses and other devices. As a result, designers can create a ROMB system in which other PCI devices cannot interfere with RAID traffic, optimizing throughput and eliminating performance bottlenecks.

Because the RAID subsystem is designed directly into the motherboard, it provides greater stability. Typically, system designers rely on industry standard specifications and validation to ensure interoperability. In addition to the benefits of open standards, ROMB subsystems can also avoid interoperability problems by fixing the location and selection of such major components as the IOP and IOC. Fixing the location and components reduces interoperability issues and helps ensure the maximum amount of validation time since the entire configuration is validated as part of the motherboard. Furthermore, because these components are embedded on the motherboard, they typically receive better airflow and cooling, which also contributes to improved system reliability.

### **Dominance of RAID-5**

There are three main reasons why companies typically choose to implement a RAID solution: (1) fault tolerance/data protection; (2) increased system performance; and (3) increased data capacity. Equally important is that the solution should be scalable so it can grow with the

business as its needs change. High-performance I/O controllers with RAID functionality distribute data across multiple disk drives in a manner that speeds access, improves reliability and protects data integrity through fault tolerance.

There are several different RAID levels or configurations that offer varying degrees of data protection and performance. The simplest RAID configurations either "stripe" data across two disk drives to increase data transfer speed (RAID-0), but offer no data protection; or "mirror" redundant data onto a second drive, without increasing performance (RAID-1). More advanced configurations involve three or more disk drives, and simultaneously provide fault tolerance, increased performance, and the ability to "recreate" information onto a spare disk drive, should a disk drive failure occur (RAID-5). These more advanced RAID configurations are preferred in server environments where maximum data availability and performance is critical.

Arguably RAID-5 has emerged as the most implemented variant because of its efficient use of disk resources and its high level of parity consistency that it provides to stored data. RAID-5 uses XOR operations to validate data writes, to create parity consistency checks, and to rebuild data if disk drives fail or data stored in the array becomes otherwise corrupted. With RAID-5, this activity produces disk striping with rotating parity, whereby the parity information is rotated across all disks in the array, rather than being targeted to a dedicated parity disk. Striping, with rotating parity, maximizes the performance of the storage array while affording excellent data protection at a comparatively low cost for most applications.

A typical RAID-5 write sequence (usually referred to as a read-modify-write) requires the completion of up to eight discrete operations, including a computational operation. While it is possible to perform all of these steps as a function of the RAID software, the processing burden created by a write-intensive transaction database application would soon create significant latency in both server and storage operations that would result in reduced application performance. In addition to the I/O burden on the CPU, another issue deals with the significant traffic across the server bus that potentially starves other applications of their resource requirements.

### **Maximizing RAID availability with SAS**

Serial Attached SCSI, the successor technology to the parallel SCSI interface, leverages proven SCSI functionality and promises to greatly build upon the existing capabilities of the enterprise storage connection. SAS offers many features not found in today's mainstream storage solutions. These include drive addressability of up to 16,384 devices and reliable point-to-point serial connections at speeds of up to 3Gbps (Gigabits per second).

SAS has emerged to deliver higher levels of reliability than previous generations of SCSI for mission-critical transactional applications that must be online around-the-clock with no data loss. When combining SAS with RAID, companies can now use multiple disk drives together, with fault tolerant designs, to create highly reliable, high-performance subsystems that support any type of e-business.

A key advantage of SAS is that its backplane design and protocol interface allow both SAS and SATA drives to be used in the same system. Though each drive type is typically used for different applications, most enterprise users have needs for both technologies. The ability to mix and match these drives is a powerful benefit for designers and users.

SATA drives are designed primarily for low cost bulk storage where transaction rates are low and data availability is not mission-critical. As a result, SATA drives feature lower spindle speeds (typically 7,200 rpm) and a lower mean-time-between-failures (MTBF) versus SAS drives.

SAS drives, on the other hand, are built for high-performance, high-availability use. SAS drives will operate at higher spindle speeds (10,000 to 15,000 rpm) with compensation for rotational vibration to assure data integrity, and are built for higher reliability. SAS drives are designed for environments where data transactions are high and data availability is essential.

RAID arrays attached to a SAS controller can be created using either SATA or SAS disks, or a combination of both. However, the performance and reliability of the resulting RAID array is in large part a function of the quality of the RAID technology and underlying RAID implementation that is used.

Depending on the application, a SATA disk RAID array can be used as an excellent data repository to provide more dependability through intelligent RAID. Similarly, a SAS disk RAID array will provide a more durable, resilient and higher performance option that meets greater business demands for enhanced data availability and protection.

### **Summary**

The demand for denser, less-expensive, higher performing storage solutions is leading to the deployment of ROMB technology on more and more motherboards. ROMB solutions are more cost-effective and provide more reliable alternatives to add-in RAID solutions such as Zero-Channel-RAID and host bus adapter cards because they don't duplicate components that already exist on the motherboard and are, by design, made to completely interoperate with other components located on the motherboard as well.

By combining the cost, reliability and performance benefits of ROMB, with the flexibility and availability features inherent in Serial Attached SCSI, these ROMB storage solutions can be cost-effectively included in mainstream server, storage and application markets, increasing the choices and value propositions to end users and IT professionals.

*Broadcom® and the Broadcom logo are among the trademarks of Broadcom Corporation and/or its affiliates in the United States, certain other countries and/or the EU. PCI X® and PCI Express® are trademarks of PCI-SIG Corporation. Any other trademarks or trade names mentioned are the property of their respective owners.*

*Paul Griffith is a Strategic Marketing Manager at [Broadcom](http://www.broadcom.com) Corporation and currently serves as Secretary on the SCSI Trade Association (STA) Board of Directors.*

*For more information on Broadcom, visit <http://www.broadcom.com>.*

*For more information on the SCSI Trade Association, visit: <http://www.scsita.org>*