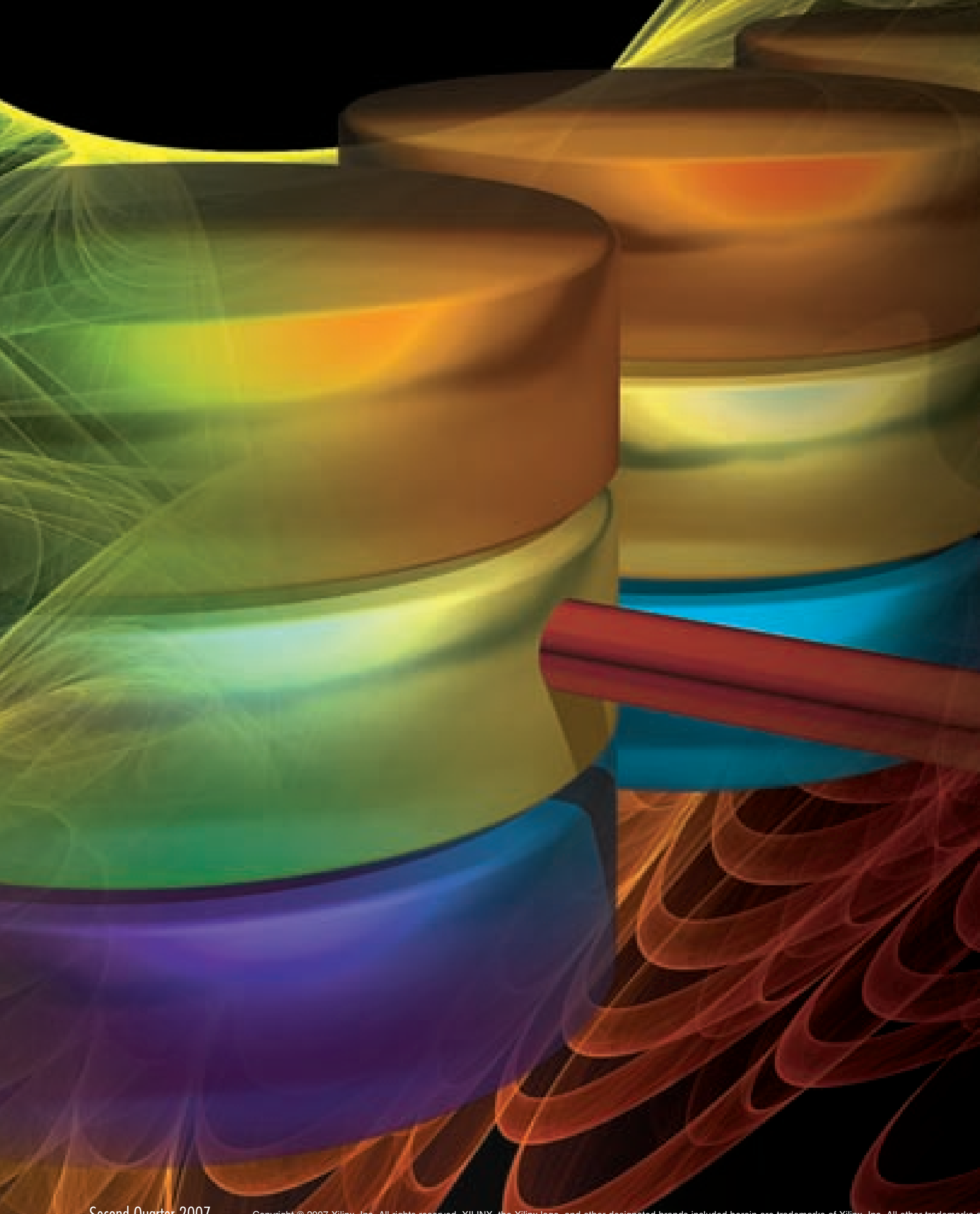
Building a 3-Gbps eSATA/SATA Hardware RAID 5 Solution

Xilinx FPGAs help RAID architect Accusys deliver innovative RAID storage.



by Paul Chu
Director of Product Marketing
Accusys, Inc.
paulchu@accusys.com.tw

Jason Lin
Hardware Design Manager
Accusys, Inc.
jasonl@accusys.com.tw

Cliff Tsai
Senior Field Application Engineer
Xilinx Taiwan Pte. Ltd.
cliff.tsai@xilinx.com

RAID (redundant array of independent disks) technology has been widely adopted in systems with hard disks for protecting data, ensuring system availability, and improving system performance. Accusys has delivered many forms of RAID products, including the low-cost, high-performance ATA hardware RAID 5 solution.

To build next-generation eSATA RAID products supporting the 3-Gbps SATA2 standard, Accusys used a Xilinx® Virtex™-4 FX FPGA to implement the eSATA host interface controller as well as the RAID engine. In this article, we'll explain the benefits of a 3-Gbps eSATA RAID solution and the role of the Virtex-4 FX FPGA.

Introduction to RAID

Different types of RAID algorithms offer different performance, reliability, and capacity. RAID 0 can improve I/O performance by concurrently distributing data over multiple disk drives, which also offers bigger space than the capacity of a single disk drive. However, if any of the disk drives in the RAID fails, the data will be lost. To protect the data, extra information associated with that data must be stored in case the original data has to be regenerated.

RAID 1 stores data to two disk drives simultaneously, so data is retained when one disk drive fails. But it is very costly because double storage capacity is required. RAID 5 offers the most cost-effective solution because only the capacity of one disk drive is needed for protecting data on any number of other disk drives. It also delivers superior performance because data is distributed to multiple disk drives, like RAID 0. As a result, RAID 5 is the most popular RAID level adopted in most of the storage systems deployed.

Advantages of Hardware RAID

RAID processing is performed either by software running on the host computer (software RAID) or by a dedicated RAID processor (hardware RAID). However, for most mission-critical systems or heavy-duty applications, a hardware RAID solution is preferred because it can deliver better performance with optimized hardware than software RAID, which consumes considerable resources on the host computer. Hardware RAID also provides superior reliability because it can regenerate the data of a disk drive in a shorter time; thus, the RAID can be recovered faster to a normal state. Therefore, mainstream RAID products are all implemented based on hardware RAID.

High-Speed and Low-Cost eSATA

One of the key features of a RAID solution is its host interface, by which the host computer accesses the storage presented by the RAID solution. There are two commonly used host interfaces: PCI-X/PCIe for RAID cards and SCSI family connections like Fibre Channel, or serial attached SCSI for RAID systems. Despite its high performance, using RAID solutions with these two interfaces is too troublesome for users without IT expertise because of the complicated installation procedure. On the other hand, the eSATA interface, which makes it easy to attach external storage devices, is gaining popularity in desktop computers.

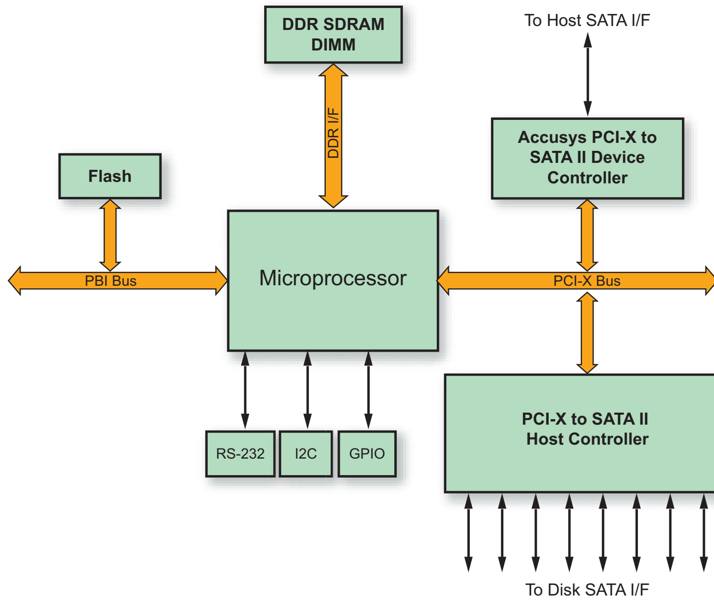


Figure 1 – Block diagram of SATA2-SATA2 RAID controller

The eSATA features a high-speed serial connection running at 3 Gbps and supports native command queuing (NCQ) for optimizing multi-tasking applications or multi-stream video processing. In addition, there is no need to install extra drivers or add-on cards for accessing eSATA devices, so its cost is lower and it is easier to use. As a result, RAID storage devices using eSATA as the host interface are good solutions in terms of performance and ease of use.

Prior Art

Most existing eSATA RAID products on the market do not support hardware RAID 5, which largely restricts their application fields. The RAID implementation of these products are derived from SATA port multipliers, offering only simple packet dispatching and supporting only basic RAID levels like RAID 0 or RAID 1. RAID 0 can serve high-performance applications like video editing, but users are bothered by the routine backup and unexpected data losses caused by bad sectors and drive crashes. RAID 1 provides data protection, but the capacity wasted is too great and there is no performance benefit.

Unlike other eSATA RAID solutions on the market, the Accusys ACS-76000 RAID system on chip offers hardware RAID 5. However, its performance is limited by the 1.5-Gbps SATA1 speed. As high-definition media content and other applications become popular, a higher performance solution will be necessary.

Implementation Overview

To achieve higher performance, the first requirement is to upgrade the storage interface from SATA1 to SATA2, so you can achieve 3 Gbps and support NCQ. The second requirement is to have a high-performance and standard bus for connecting to a processor. We chose the PCI-X bus because it is commonly used in high-performance embedded systems. Finally, to support hardware RAID 5 processing, an exclusive OR (XOR) engine is needed to calculate the parity and regenerate data. Figure 1 is a block diagram of the SATA2-SATA2 RAID controller board.

We selected the Xilinx Virtex-4 FX FPGA because it provides the multi-gigabit serial I/O for implementing the 3-Gbps SATA host interface and can support the SATA device-mode controller logic for connecting to the digital parallel bus interface of the MGT hard core. We also imple-

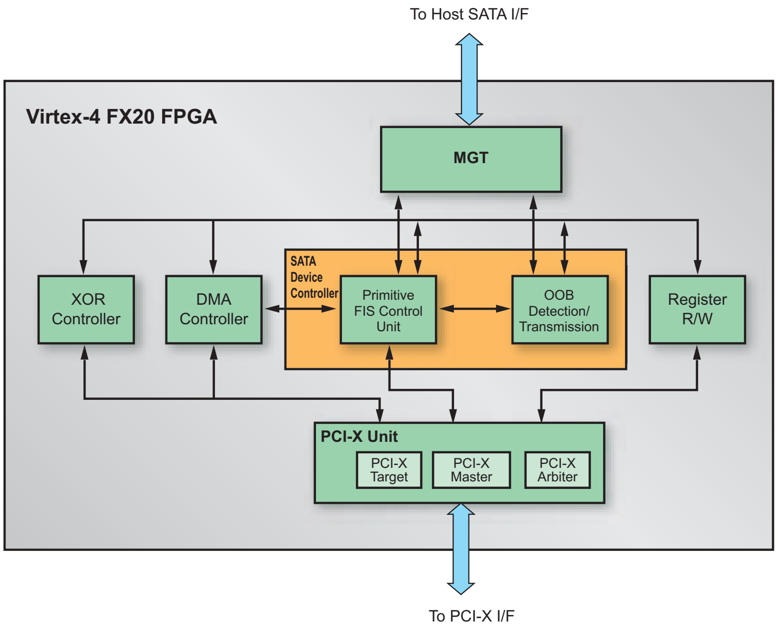


Figure 2 – Block diagram of 3-Gbps eSATA RAID engine

mented a 64-bit/133-MHz PCI-X controller and an XOR controller on the FPGA. A PCI-X-based external processor can parse commands from the SATA host and return status. In addition, the XOR controller can directly read the source memory data on the PCI-X bus and calculate the XOR data for the RAID 5 encoding. All of these tasks are done without the processor's intervention. Please refer to Figure 2 for a block diagram of the FPGA.

Implementing the 3-Gbps eSATA

Three different layers are defined in the SATA controller: the transport layer, link layer, and physical layer (PHY):

1. Transport layer. The SATA command execution and data transfer occurs by exchanging frame information structures (FISs), which contain ATA registers or disk data. The transport layer constructs FISs from the PCI-X bus for transmission, decomposes received FISs, and passes them to the PCI-X bus. In short, the transport layer is the interface between the link layer and the higher application layer, which in our design is the PCI-X bus.

2. Link layer. The link layer is responsible for packet framing, 8b/10b encoding and decoding, and generating and checking the CRC codes. The link layer also handles flow control, buffering data and exchanging primitives as needed to accommodate burst transfers. The Virtex-4 FX device's multi-gigabit transceiver (MGT) embeds the 8b/10b encoder and decoder, as well as a CRC-32 generation circuit, which provides tremendous convenience for building the SATA link layer.

3. Physical layer. The Virtex-4 FX FPGA's 6.5-Gbps RocketIO™ transceiver provides the basic building block of the SATA physical layer, responsible for deserialization of received data from the host PHY and for serialization of outbound 10b encoded data from the link layer. It also supports out-of-band (OOB) signals for initializing the connection between the SATA host and device. To support external 3-Gbps SATA2, Gen2m electrical specifications are required in the implementation of the SATA physical layer.

Implementing the PCI-X Bus

The PCI-X controller includes both the master and target function of a 64-bit/133-MHz PCI-X device. It is equipped with a buffer RAM for storing data from the SATA transport layer and from the PCI-X bus for processing the split transaction flow. In addition, an embedded PCI-X arbiter is implemented to provide more flexible control for the PCI-X master. The DMA controller plays an important role in this design because it can provide an efficient bi-directional data transfer between FISs (from the SATA transport layer) and the PCI-X bus. As a result, the performance of bulky data transfer significantly improves.

Implementing the XOR Controller

For applications like video editing or recording, efficient XOR calculation is mandatory to meet the demanding requirements for low I/O latency and high throughput. The XOR controller is designed with as many as 256 data sources and 128 chained commands. Each of these can calculate 64-KB parity data. With the XOR controller implemented on the FPGA, the processor and the memory bandwidth are freed from performing the lengthy and routine XOR calculation on 256 sources with 128 chained commands at 64 KB per entry.

Conclusion

The 3-Gbps eSATA hardware RAID offers high performance, low cost, and ease of use. As more computers and embedded systems are shipped with eSATA ports, the eSATA RAID product represents an important external RAID solution for customers requiring both high performance and high reliability. Leveraging the advanced features of Virtex-4 FX FPGAs such as its digital clock managers and easy-to-use building blocks like dual-port block RAM, we developed the key component for the 3-Gbps eSATA RAID system quickly and efficiently.

With our success delivering products based on Xilinx Virtex-4 FX FPGAs, Accusys is positioned to deliver eSATA RAID products for high-end applications like video editing and workgroup-shared storage, as well as entry-level applications like DVR backup or home storage. For more information about RAID and storage technologies, visit the Accusys website at *www.accusys.com.tw.*