Are You Losing (Out On) Your Memory?

Why Memory and Storage Have Reached Their Flashpoint

Vishal S. Kapoor, VP Marketing, Cadence Design Systems

Mobile devices are the computers of this generation—yes Mr. Steve Jobs, we accept that we're in the post-PC era! These devices, and the enterprise infrastructure that supports them, are all about content and access to that content. Efficiently managing data in and out of enterprise and consumer systems—the data centers, cloud applications, smart phones, and tablets—is a challenge and an opportunity. Memory and storage access to manage data affect the architecture, cost, and success of today's systems.

Overview

In recent years, compute infrastructure has gone from performing powerful, large calculations to managing huge numbers of transactions on large databases as quickly as possible. Much of this transition is driven by the need for immediate access to information on mobile devices, which look more like desktops in both their daily application and implementation technology (multi-core, multi-level cache, etc.) Such transaction processing systems leverage an enormous number of parallel compute and memory/storage components.

The challenge is keeping these systems fed with data, and this challenge continues with storage. The single biggest disruption in the storage industry has been the adoption of the solid-state storage device, mainly NAND Flash. Energy efficiency, instant-on, and an ever-smaller form factor drive NAND Flash access in mobile devices; for enterprise, NAND Flash adoption is driven by data performance, power distribution, and reliability.

While the Apollo mission landed men on the moon with 152Kbytes of onboard memory, today's cell phones require Gbytes of memory to deliver GPS-enabled maps and capture live video. The need for memory and storage is clearly rising. And the focus is no longer on how fast you can compute, but how fast you can access your information. Where we once cared about millions of instructions per second (MIPS), we must now focus on input/output operations per second (IOPS). If you're in the electronics industry, you should become familiar with this term.

With consumerization driving tight and unforgiving market windows, today's engineering projects require rapid integration of hardware (both digital and custom) with software—and software demands are going up. These engineering constraints must be balanced with stringent time-to-market and ubiquitous cost requirements. Most important is getting to market fast, preferably first and with a differentiated solution. Today, you just can't build a successful piece of electronics without intelligently managing your memory and storage.

Universal Memory: Utopia vs. Reality

With an infinite amount of fast memory, all on-chip processors and processing elements would have their collective memory and storage requests satisfied as soon as they were posted. An unconstrained memory resource requires neither memory management nor storage. But cost constraints prohibit adding sufficient memory to meet unconstrained memory capacity and the bandwidth requirements of today's SoCs.

Reality looks something like this: memory is expensive, storage access is slow, different types of memory are better suited to meet different needs, and solutions for one application are not applicable to others. System development teams must balance component costs with bandwidth requirements. Many memory and storage problems can be solved by adding more memory in parallel, by increasing frequency, or by selecting higher-performance memory—but this can lead to prohibitively high costs and power consumption. However, by increasing the intelligence of

the memory and storage access, design teams can achieve the highest bandwidth and the most efficient storage capacity while lowering costs and power consumption.

How simple would life be if we had memory that was dense, fast, cheap, and non-volatile! But reality comes with challenges like latency, cost, capacity, and performance. It's a tradeoff among technology, business, and time-to-market goals that requires extensive, sophisticated memory and storage management to make it all work on the SoC. Adding intelligence to the memory and storage subsystem is the closest we can get to utopia.

The Overall Design Context

Systems are changing fast. An SoC was once defined as a chip with an embedded processor that accessed a couple of memory banks. Cutting-edge SoCs today comprise multiple processing engines (CPUs, GPUs, NPUs, DSPs) that generate and consume data; the interconnect to transport data between engines; and a memory subsystem that serves the bandwidth and latency needs of all data generators and consumers in the system. And each component must be optimized for what it does best.

Outside the chip, disruptions are beginning to show. Each new DDR DRAM generation calls for an end to the high-speed parallel interface on memory, and we're already seeing the mobile memory subsystem moving to serial interfaces or Wide I/O approaches that utilize through-silicon-via (TSV) 3D-IC connections between SoCs and memory. However, designers still need to manage the physical design of the DRAM PHY and the signal integrity requirements of the package and board when dealing with fast parallel interconnect between SoC and DDR memory.

A balanced storage solution for servers is different for each application and system. Servers today are moving to a tiered strategy, using NAND Flash as a virtual system memory in place of raided HDDs. This new class of memory has high IOPS and the capacity to handle many different random operations, gaining 100x the performance over older architectures. Making these solutions OS-aware is a challenge; linking connected data through multiple tiers is tricky given different protocols, address translation layers, and metadata. In the mobile market, the capacity growth is staggering, from an average of 2GB just 2 years ago to more than 8GB today (with the high end moving to 64GB of on-board storage). This higher performing NAND requires new interfaces like MIPI M from the MIPI and JEDEC groups, MSATA from JEDEC for a protocol, and smaller connector and form factors.

The Solution

Only a few companies in the world have all the resources to address the entire SoC — processing engines, interconnect, memory subsystem, physical design, package, and board. The answer calls for powerful, wide-ranging, and yet focused solutions in each area in the form of reusable IP. The required attributes for each kind of IP are:

- Performance: intelligent and "full custom," not "general purpose"

- Features: all the necessary features for any of the SoCs that choose to use it

- Flexibility: fast integration into any SoC, with a minimum of re-engineering required

- Comprehensiveness: input from a collection of customers, suppliers, and standards bodies

- PHY layers: compatibility with package and board design needs

Memory and storage also have specific requirements. Memory controller intelligence should include support for heterogeneous multi-processing with different masters and latency requirements while considering the demands of high-speed DDR3 and DDR4. Storage controllers should be designed to support multiple Flash interface standards, optimize the traffic for each, and have a simple, high-performance connection to the SoC operating system.

Summary

There are many sophisticated management techniques to maximize the effective bandwidth of DDR DRAMs and NAND Flash Memory ICs. To ensure data reliability when using NAND Flash devices, you can ask an internal development team to learn the details and incorporate them into custom memory and storage controllers. Or, you can take the faster path to a competitive, working SoC by leveraging the considerable expertise embodied in configurable memory and storage management IP. Focus on the differentiating elements of your design; don't spend time creating your own IP.

The selection of IP for memory and storage management is no longer a simple matter. Choosing IP to perform these tasks now requires as much care as selecting the processing elements that make use of the attached memory and storage devices. One way to ensure SoC design success is to start in the right direction with your memory and storage management. It's not optional; it's a necessity—one that's reached its flashpoint.