



WHITE PAPER

Sponsored by **Rambus**

High Speed Memory Interface Chipsets Let Server Performance Fly

SUMMARY

The demands on server performance continue to increase at a tremendous pace. New requirements from large in-memory databases that are powering today's cloud services and advanced analytics tools are arriving just as the impact of Moore's Law is starting to slow. This is setting up a classic performance challenge that requires rethinking some of the core elements of today's server architectures, particularly when it comes to memory. One key new opportunity is for high-speed server memory interface chipsets, which enable high-speed memory performance without compromising on memory capacities. Companies looking to optimize their server memory architecture designs, and improve their overall server performance and reliability, should give serious consideration to optimized DDR4 memory interface chipsets, which enhance the performance of server memory modules.

"Server memory chipsets play a critical role in high-speed DDR4 designs. They allow server designers to maintain the high-speeds that DDR4 offers, while also enabling the higher capacity designs that today's applications require."—Bob O'Donnell, Chief Analyst

INTRODUCTION

In the world of high-performance automobiles from the likes of Porsche, Ferrari, and Lamborghini, much of the attention gets focused on the car's engine. Auto magazine reviews, online forums, and rabid fans regularly discuss and debate the performance specs of their favorite cars.

Sports car enthusiasts know, however, that there's a lot more to great performance than just an engine's horsepower rating. All the elements of the automobile's drivetrain—from its fuel-injection system to its tires and everything in between—need to work together to deliver the kind of jaw-dropping speed for which these cars are famous.

So it is in the world of today's servers. The high-performance CPUs at the heart of today's servers justifiably receive a great deal of the glory, but in truth, there are several key elements that keep the server operating at top performance. Most important among these elements is memory. High-speed memory and its connections to the CPU are like the fuel-injection system of a sports car, keeping the server engine running at its maximum potential and ensuring a smooth overall operation.

SERVER MEMORY ARCHITECTURES

Dynamic Random Access Memory (DRAM) sits at the heart of today's servers and continues to play an integral role in their operation. Applications and data are loaded from storage into DRAM and the CPU then acts on this data in order to perform the kinds of operations today's servers are expected to do.

In order to achieve the best possible performance, every single aspect of these elements and the connections between them need to be isolated, analyzed and optimized. The memory itself, commonly packaged in the form of DIMMs (Dual In-Line Memory Modules), for example, has seen numerous improvements in capacity, speed of internal operation, and types and speeds of connections to external devices with each generation.

Approximately 25% of today's servers are shipping with the latest generation DDR4 memory, which improves upon previous generations of memory technology by using lower power signaling to enable faster speeds. By 2017, that number is expected to hit 80%.

With the introduction of DDR4, server system designers can leverage DRAM that runs at speeds of 2,133 Mbps today, with future speeds running up to 3,200 Mbps. This performance boost also comes with real challenges, however, because the move to higher speeds degrades electrical signal integrity, especially with multiple modules added to system. In practical terms, this means it's becoming harder to achieve higher capacities at higher speeds.

In order to overcome this electrical limitation, memory designers use specialized clocks and dedicated memory buffer chips integrated onto the DIMMs. These server memory buffer chipsets play a critical role in high-speed DDR4 designs. They allow server designers to maintain the high-speeds that DDR4 offers, while also enabling the higher capacity designs that today's applications require.

As Figure 1 illustrates, there are two types of modern server DDR4 DIMMs. In a Registered DIMM (RDIMM) a Register Clock Driver (RCD) chip delivers a single load for the clock and command/address signals for the entire DIMM onto the data bus that connects between memory and the CPU. This enables a reduced impact on signal integrity versus an unbuffered DIMM, where all of the individual DRAM chips put multiple loads on clock and command signals. On a Load Reduced DIMM (LRDIMM), each individual DRAM chip has an associated Data Buffer (DB) chip—in addition to the RCD on the module—to reduce the effective load on the data bus, which enables higher capacity DRAMs to be used. The combination of the RCD and individual DBs constitute a complete server DIMM chipset.

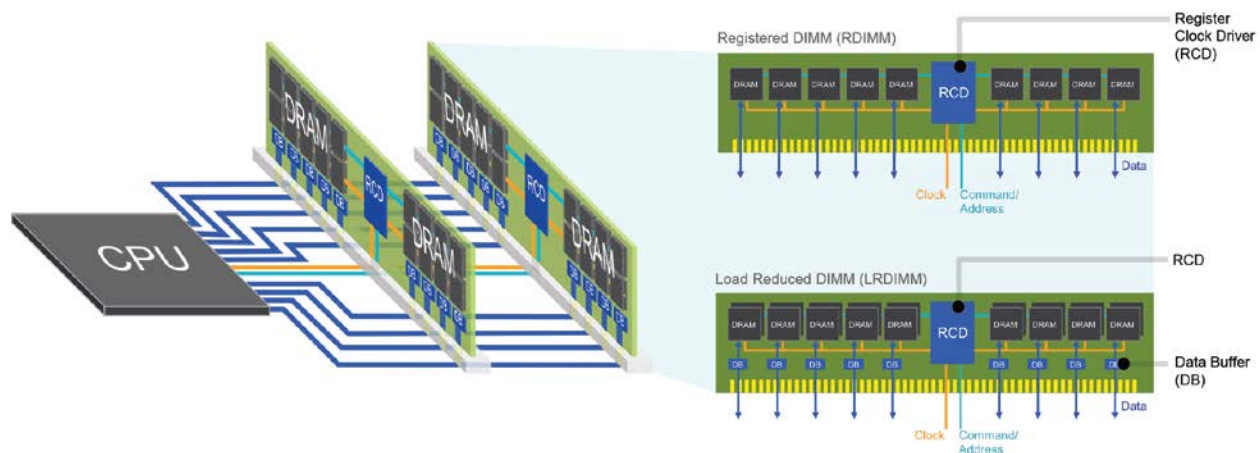


Figure 1

With the server DIMM chipset enabled, data is not actually sent straight to the CPU from memory, just as gasoline isn't sent straight to a car's engine from its fuel tank. A properly designed fuel injection system sends gas to the engine in exactly the right form, quantity and speed that it requires and, in an analogous way, memory buffers serve to regulate the delivery of raw data from memory into and out of the CPU.

The location of the data buffers on DDR4 LRDIMMs also play a vital role in helping them achieve better performance than DDR3 LRDIMMs. The big benefit is the reduced trace distance from each DRAM module to the memory bus and memory controller. While DDR3 LRDIMMs have a single centralized memory buffer that forces data to cross the distance of the DIMM module and back, DDR4 LRDIMMs have dedicated memory buffer chips located a very

short trace line away from the data bus. The real-world benefits are time savings that can be measured in nanoseconds, and improved signal integrity because of the shorter trace lines, both of which translate into better real-world performance for time-sensitive applications.

THE PERFORMANCE CHALLENGE

The reason why these improvements in server memory buffers matter is because of both the shifting nature of server workloads, as well as the slowing of semiconductor process improvements. Until recently, memory technologies have benefitted from the same type of Moore's Law improvements that have driven CPU makers to smaller process technologies and higher speeds. While Moore's Law continues to be a factor, it's becoming clear that the rate of change and the speed of process shrinks are slowing, especially for DRAM. This, in turn, is making wholesale improvements in DRAM performance more challenging—just as the performance demands of new applications are starting to ramp.

Today's cloud-based services, advanced analytics tools, and other big data applications are driving a higher set of expectations about server performance. Throw in the looming prospect (and opportunity) of the Internet of Things (IOT) and the stage is set for a very challenging environment in today's and tomorrow's data centers and enterprise servers.

Many of these new applications leverage very large in-memory databases to meet the performance expectations of today's increasingly connected, mobile world. To effectively use these databases, and other memory-intensive applications, improving the performance of moving data into and out of memory while increasing memory capacities is absolutely essential. Microseconds count when you need to provide real-time analytics on millions of financial transactions, or if you're trying to offer real-time language translation via a cloud-based service, for example.

The solution is a set of chips like the new Rambus R+ DDR4 Server DIMM chipset, which can reduce latencies in these time-sensitive applications, and ensure the best possible performance in delivering data to and from the CPU. This is particularly true for large multi-core CPUs, which can benefit from getting multiple dedicated lanes of memory bandwidth from a system architecture that uses these memory chipsets.

THE RAMBUS CONNECTION

Technology innovator Rambus is certainly no stranger to innovations in the memory space. Back in the 1990s, the company created RDRAM, a technology that offered breakthrough levels of performance to game consoles such as the Nintendo 64, x86 Pentium4 PCs, and several generations of the Sony PlayStation, which were starved for high-performance

memory. Since then, Rambus has continued to invest in leading-edge memory and high-speed serial link technologies, along with advancements in areas as far-ranging as cryptography, LED lighting, and lensless smart sensors for smart vision applications.

With the R+ DDR4 Server DIMM chipset, Rambus has chosen to enter the finished semiconductor market for the first time, offering their branded chips to DRAM and DIMM manufacturers such as Samsung, SK Hynix, Micron and more.

The new Rambus chips are DDR4 JEDEC-compliant, ensuring they will work with any standard server DDR4 DRAMs and function in any standard DDR4 server architecture. In fact, they surpass JEDEC's reliability requirements. They work at 2,666 Mbps and already include built-in support for 2,933 Mbps, making them well-prepared for future memory innovations.

In addition to high-performance, these new chips also include robust capabilities for debugging and service, which can be critical when designing new servers. They also offer frequency-based power optimization and work with the default BIOS—they will work out of the box.

CONCLUSION

The challenges of meeting today's server performance and capacity needs are very real and likely to get even tougher as time goes on. With the slowing of pure semiconductor process improvements, combined with the increasingly memory-hungry big data applications and workloads, server performance is coming to a serious crossroads. In order to meet those requirements, it's going to take more than just a faster engine—it's going to take a smarter overall system design.

Server memory chipsets and buffers such as the new technology offered by Rambus may not offer the same guttural satisfaction as the sonorous roar of a sports car engine, but for server enthusiasts who want the best possible performance for their data centers, these server memory chipsets can bring a similar sense of performance gratification.