Double Data Rate SDRAM: fast performance at an economical price

technology brief

Abstract

This paper describes Double Data Rate (DDR) SDRAM and the enhancements that enable it to double the peak bandwidth of its predecessor, SDRAM. This paper also explains the factors that make DDR SDRAM incompatible with SDRAM and the effect, if any, that it will have on customers' investment in SDRAM. It explains the reasons why DDR SDRAM and DDR-II are becoming more widely accepted than other, competing memory technologies in the server industry.

Introduction

Because system memory bandwidth has not kept pace with improvements in processor performance, a "performance gap" has developed between the processor and memory subsystems. Processor performance, which is often equated to the number of transistors in a chip, doubles every couple of years. In comparison, memory bandwidth doubles roughly every three years. Therefore, if processor and memory performance continue to increase at these rates, the performance gap between them will widen. The performance gap prevents many applications from effectively using the full computing power of modern processors because the processor is forced to idle while it waits for data from system memory.

Increasing the speed of the memory bus can improve overall system performance; however, technological challenges and higher manufacturing costs prevent system designers from speeding up the memory bus.

DDR SDRAM was defined in 1997 by the Joint Electronic Device Engineering Council (JEDEC), the semiconductor engineering standardization body of the Electronic Industries Alliance. Designed as an evolutionary advancement from PC100 and PC133 SDRAM, DDR SDRAM delivers twice the bandwidth yet follows the low cost legacy of SDRAM. DDR SDRAM and SDRAM share many features that allow DDR SDRAM to leverage SDRAM manufacturing and testing processes; however, the two technologies are incompatible.

This paper describes the enhancements that designers used to double the peak bandwidth of DDR SDRAM over SDRAM and the reasons why DDR SDRAM cannot be used in the same systems as SDRAM. Readers who are unfamiliar with SDRAM can refer to the paper titled "Memory technology evolution: an overview of system memory technologies" for background information.[1]

DDR SDRAM

The main factor that has prevented memory designers from further increasing the bandwidth of SDRAM is memory bus speed. Designers found that trying to speed up the memory bus by increasing the bus clock frequency negatively impacts signal integrity. Why? Because as the bus clock frequency increases, the signal timing becomes tighter. Tighter timing increases interference between signals. To avoid signal interference at higher speeds, circuits must be manufactured more precisely, which significantly increases hardware costs.

To develop DDR SDRAM, designers made enhancements to the SDRAM core to increase the data rate. These enhancements include:

- Prefetching
- Double transition clocking
- Strobe-based data bus
- SSTL_2 low voltage signaling

[1] This paper is available online at http://h18004.www1.hp.com/products/servers/technology/whitepapers/2.

Prefetching

In SDRAM (Figure 1), one bit per clock cycle is transferred from the memory cell array to the input/output (I/O) buffer or data queue (DQ). The I/O buffer releases one bit to the bus per pin and clock cycle (on the rising edge of the clock signal). To double the data rate, DDR SDRAM uses a technique called prefetching to transfer two bits from the memory cell array to the I/O buffer in two separate pipelines. Then the I/O buffer releases the bits in the order of the queue on the same output line. This is known as a 2n-prefetch architecture because the two data bits are fetched from the memory cell array before they are released to the bus in a time multiplexed manner.

Double transition clocking

Standard DRAM transfers one data bit to the bus on the rising edge of the bus clock signal, while DDR SDRAM uses both the rising and falling edges of the clock to trigger the data transfer to the bus (Figure 2). This technique, known as double transition clocking, delivers twice the bandwidth of SDRAM without increasing the clock frequency. DDR SDRAM has theoretical peak data transfer rates of 1.6 and 2.1 GB/s at clock frequencies of 100 MHz and 133 MHz, respectively.

Figure 1. Comparison of SDRAM architecture and DDR SDRAM 2n-prefetch architecture

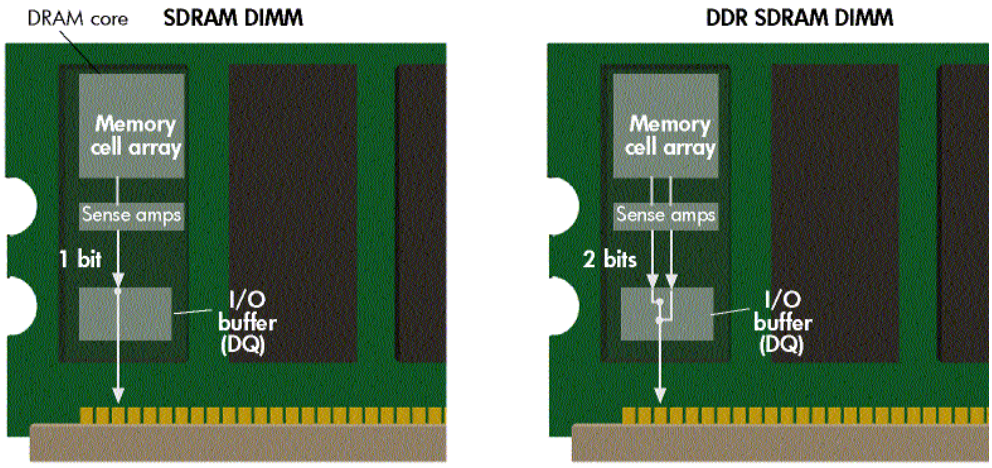
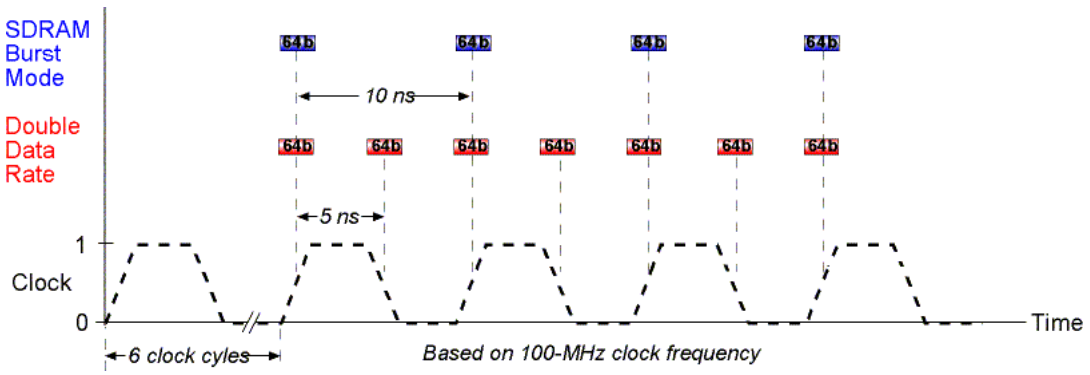


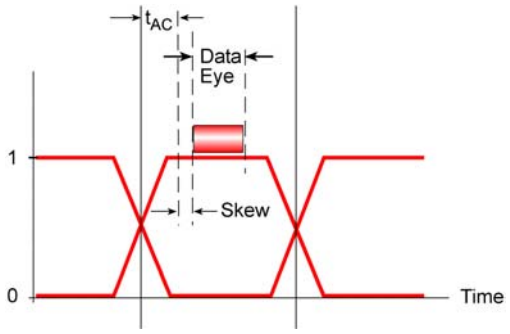
Figure 2. Data transfer rate comparison between SDRAM (with burst mode access) and DDR SDRAM



Strobe-based data bus

In a synchronous system, data output and capture are referenced to transitions in the memory bus clock. When the clock signal transitions, an operation is signaled to begin; however, a period of time must pass before the signal stabilizes. As shown in Figure 3, access time (t_{AC}) is the amount of time it takes to "open" the output line from the prior clock "tick." t_{AC} is specified as a maximum value because one DDR DRAM chip may have a t_{AC}=4 ns, while another chip may have a t_{AC}=6 ns. To be able to send data out on every clock cycle, t_{AC} must be fast enough to allow the signal to stabilize before beginning the actual output operation.

Figure 3. Representation of the data valid window, or data eye for DDR SDRAM



Because of possible distortion (skew) in the SDRAM chips and printed circuit board, as the bus frequency increases, it can become more difficult for components to capture data using a clock. This is especially true for DDR SDRAM because its data capture rate is twice the clock frequency. The issue is that the region where the data is valid (called a data eye) is so small that it is very difficult to meet setup and hold timing (the period of time that must pass to allow the signal to stabilize). For example, if the bus frequency is increased to 200 MHz, the clock cycle (transition) is only 5 ns. In an ideal world, each data word (64 bits) sent by the source device would arrive at the capturing device at exactly the same time so they could be captured every 5 ns. Because of real factors such as trace length differences and variations in temperature, voltage, and manufacturing processes, however, the data eye may move in relation to the fixed clock signal. As a result, the time the data is stable is 5 ns minus the extra delays caused by skew.

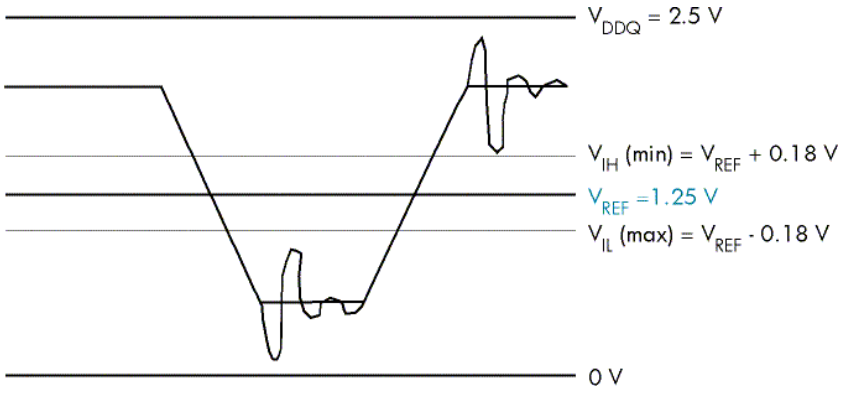
System designers devised a technique using bi-directional strobes to alleviate the tight system timing requirements on the data bus. The source device sends a reference signal called a strobe with the data. The data strobe signal helps the capturing device to locate data more accurately and resynchronize incoming data from different DIMMs. For example, during a write operation, the memory controller places the rising and falling edge of the strobe in the middle of the data eye so it can be used to capture the data at the SDRAM. For read accesses, SDRAM also sends data with a strobe, but the strobe edge is aligned with the data. The device receiving the read data must shift the strobe to the center of the data eye. Read and write accesses have different strobe alignment so that the delay circuitry can be centralized in one place (the controller) and does not have to be replicated in every DRAM device in the system. The command bus does not use a strobe, so it must still meet setup times to a synchronous clock.

SSTL_2 low-voltage signaling technology

Another difference between SDRAM and DDR SDRAM is the signaling technology. Instead of using a 3.3-V operating voltage, DDR SDRAM uses a 2.5-V signaling specification known as Stub Series-

Terminated Logic_2 (SSTL_2). This low-voltage signaling results in lower power consumption and improved heat dissipation. SSTL_2 references the incoming data to a reference voltage, V_{REF}, which is one half of the operating voltage, V_{DDQ} (Figure 4). The high and low states (V_{IH} and V_{IL}, respectively), of the input are determined as millivolts above and below V_{REF}.

Figure 4. SSTL_2 low voltage signaling



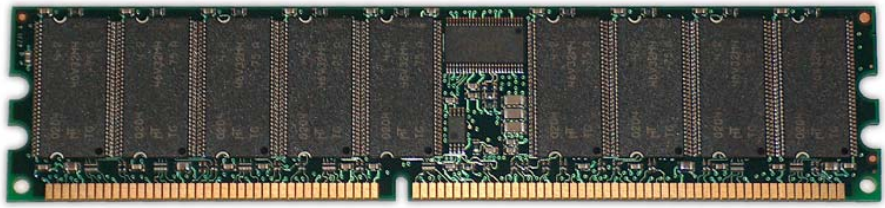
Compatibility

DDR DIMMs cannot be used in servers that support standard SDRAM DIMMs because of physical incompatibilities: DDR SDRAM DIMMs have 184 pins (Figure 5) compared to 168 pins used by standard SDRAM DIMMs. The DDR DIMMs have one notch while SDRAM DIMMs have two notches. Current users of PC100 and PC133 memory need not worry about DDR SDRAM module incompatibility because DDR SDRAM modules require DDR systems.

DDR SDRAM modules can be used in a DDR system with a slower bus speed; however, the modules will run only as fast as the bus. A DDR SDRAM module can be used in a system with slower DDR SDRAM modules, but it will run only as fast as the bus or the slower module, whichever is slowest.

Memory manufacturers provide unbuffered and registered versions of DDR SDRAM modules. Unbuffered DDR modules place the load of all the modules on the system memory bus, so they are typically used in systems that do not require high memory capacity. Registered DDR modules place only one load per module on the memory bus, regardless of how many DRAM chips are on the module. Therefore, registered DDR SDRAM modules are best suited for DDR systems that feature very high capacities.

Figure 5. 184-pin DRR SDRAM registered DIMM. The DDR SDRAM module has one notch.

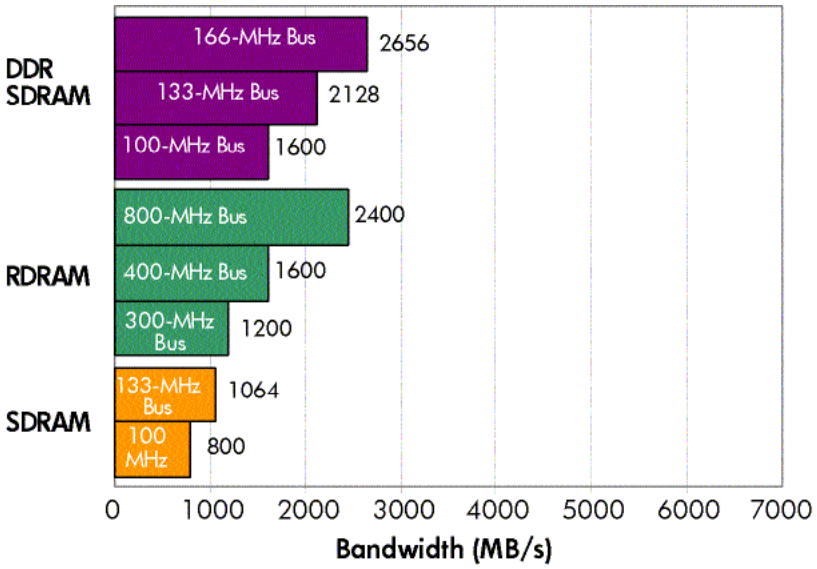


Performance

DDR SDRAM operates at data bus speeds (frequencies) of up to 400 MHz. At this frequency, 64-bit DDR SDRAM provides a theoretical peak bandwidth of 3.2 GB/s. The data bus can run at this frequency with the aid of strobes. However, the command bus does not use strobes and, therefore, must meet setup times to a synchronous clock. Thus, while the data bus operates at up to 400 MHz, the command bus can operate at a maximum of 200 MHz.

Rambus DRAM (RDRAM) is a system-wide design that allows data transfer through a bus operating in a higher frequency range than DDR SDRAM. In essence, RDRAM moves small amounts of data very fast, whereas DDR SDRAM moves large amounts of data slower. RDRAM is capable of operating at 800 MHz and providing a peak bandwidth of 2.4 GB/s.

Figure 6. Comparison of bandwidths for standard SDRAM, Rambus DRAM, and DDR SDRAM



Originally, the module naming convention for DDR-SDRAM was based on the effective clock rate of the data transfer. For example, DDR SDRAM that operates at 100 MHz or133 MHz was called PC200 or PC266, respectively. Then Rambus began using PC600 and PC800 to describe their 300-MHz and 400-MHz RDRAM modules. This falsely implied that Rambus PC600 and PC800 were faster than DDR PC200 and PC266. Therefore, the memory industry developed a numbering system based on the actual peak data transfer rate in MB/s. Thus, PC200 became PC1600 (64 bits÷8 bits per byte × 2 × 100 MHz = 1600 MB/s) and PC266 became PC 2100 (64 bits÷8 bits per byte × 2 × 133 MHz = 2133 MB/s).

Table 1 summarizes the naming conventions of various types of DDR SDRAM and their theoretical bandwidths.

Table 1. Naming conventions and theoretical bandwidths of DDR SDRAM

Component naming convention	Module naming convention	Bus speed	Peak bandwidth
DDR200	PC1600	100 MHz	1.6 GB/s
DDR266	PC2100	133 MHz	2.1 GB/s
DDR333	PC2700	166 MHz	2.7 GB/s
DDR400	PC3200	200 MHz	3.2 GB/s

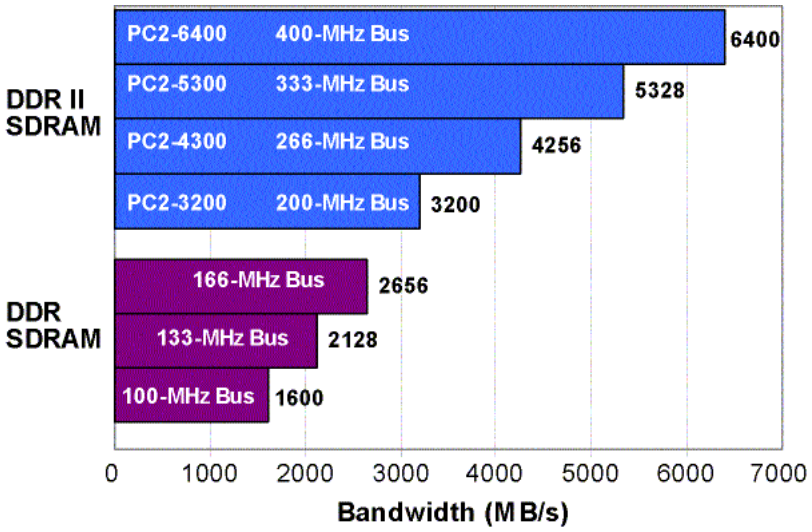
DDR-II SDRAM

DDR-II is the second generation of double data rate memory. It operates at a lower voltage and higher clock frequency than the original DDR SDRAM. DDR II uses Finepitch Ball Grid Array (FBGA) packaging, which not only reduces the packaging size, but also improves performance and heat dissipation.

Performance

DDR-II SDRAM improves bus utilization to deliver a maximum peak bandwidth of 6.4 GB/s (see Figure 7). This is achieved through an extra buffers, improved prefetch (4 bits), reduced electrical requirements (1.8 V), improved packaging and on-die termination. However, there is a tradeoff to achieve the higher bandwidth: higher latency. The higher latency is partially a result of the lower voltage, which decreases the speed of the memory cell array. Still, the lower power consumption of DDR-II makes it attractive for power-sensitive platforms such as notebook PCs.

Figure 7. Comparison of bandwidths for DDR SDRAM and DDR II SDRAM



DDR2-400 is the initial DDR-II entry into the market. DDR2-533 and DDR-667 are projected to have lower latencies than DDR2-400. Table 2 summarizes the naming conventions of various types of DDR-II SDRAM and their theoretical bandwidths.

Table 2. Naming conventions and theoretical peak bandwidths of DDR-II modules

Component naming convention	Module naming convention	Bus speed	Peak bandwidth
DDR2-400	PC2-3200	200 MHz	3.2 GB/s
DDR2-533	PC2-4300	266 MHz	4.3 GB/s
DDR2-667	PC2-5300	333 MHz	5.3 GB/s
DDR2-800	PC2- 6400	400 MHz	6.4 GB/s

Backward compatibility

DDR-II modules are not compatible with DDR modules. DDR-II modules operate at 1.8 V while DDR modules operate at 2.5 V. The DDR-II modules have 240 pins compared to 184 pins for DDR modules. Also, DDR modules do not support certain features necessary for DDR-II modules to function, such as On-Die Termination (ODT), Off-Chip Driver (OCD) calibration, Posted CAS and Additive Latency (AL), and variable write latency.

DDR-III SDRAM

JEDEC is currently developing the third-generation DDR SDRAM technology, DDR-III, which will continue to make improvements in bandwidth and power consumption. DDR-III is an evolution of DDR-II technology. For example, DDR-III will use 1.5-V signaling compared to 1.8 V for DDR-II and 2.5 V for DDR SDRAM.

Conclusion

DDR SDRAM continues to provide good performance in servers and workstations; but for memory subsystem performance to keep up with future processors, the memory bus bandwidth must continue to increase. Compared to revolutionary memory technologies, DDR SDRAM and DDR-II SDRAM deliver faster performance at lower cost, which is necessary for industry-wide adoption. While the basic principle of DDR SDRAM is simple, technical advances that enable it to achieve higher data rates also make it incompatible with standard SDRAM. DDR-II and DDR-III SDRAM promise to push the performance of the memory subsystem to even higher levels. DDR-III will offer even lower power consumption and increased bandwidth in comparison to DDR-II.

For more information

For additional information, refer to the resource listed below.

Resource description	Web address
"Memory technology evolution: an overview of system memory technologies" Technology Brief	http://h18004.www1.hp.com/products/servers/technology/whitepapers/adv-technology.html - 2

Call to action

To help us better understand and meet your needs for ISS technology information, please send comments about this paper to: TechCom@HP.com.

TC041004TB, 10/2004

invent