Emergence of Segment-Specific DDRn Memory Controller and PHY IP Solution

By Eric Esteve (PhD)

Analyst

July 2016

IPnest

Emergence of Segment-Specific DDRn Memory Controller IP Solution

By Eric Esteve (PhD) Analyst, Owner IPnest

This White Paper describes a family of DDRn and LPDDRn memory controller IP solutions specifically configured for different market segments: mobile, high-end consumer and infrastructure. Each option provides the best compromise in terms of power consumption, performance and area (PPA), by varying the optimized feature set, and even the technology node selection between 16nm and 28nm. Such a diversified offering of the same functionality addresses the emergence of new market needs, adapted to specific segments. This paper was prepared by IPnest and sponsored by Cadence, but the opinions and analysis are those of the author.

Memory Controller IP Is Clearly the SoC Design IP Masterpiece

The definition of a system-on-chip (SoC) may sound vague, but it becomes more precise when we consider that the chip integrates a processor. It can be one or multiple CPUs, GPUs or DSP cores, or a combination of those, but the SoC is clearly defined by integrating the processing unit(s), allowing an autonomous system to be built. This processing unit needs to access a large amount of DRAM, implemented outside the chip as SDRAM devices. For many years, the industry has used double data rate (DDR) SDRAM. After the introduction of DDR came the definition of DDR2 protocol, but very soon came the need for a low-power variant, or LPDDR, to support the wireless mobile industry. The semiconductor industry is served today by memory devices supporting various protocols, like DDR4, DDR3, LPDDR4, LPDDR3, GDDR5, HBM, HMC, etc. The trend is clearly to define application specific memory-protocols. But developing many, and different, memory controller IP is resource and time consuming and not the best option for a vendor.
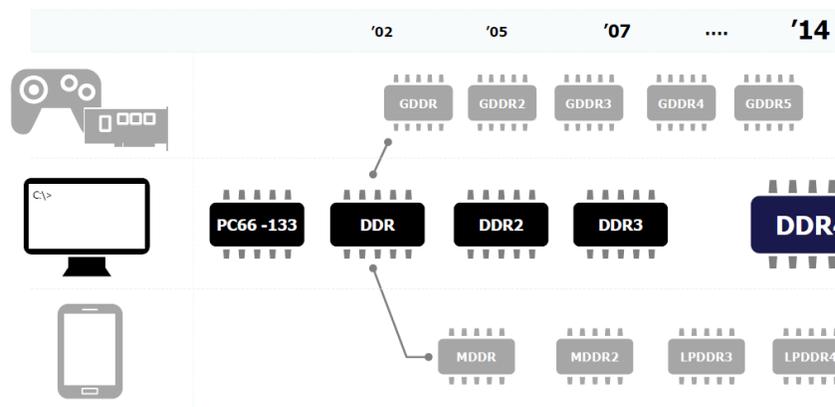

Figure 1: The Various DRAM Protocols

Considering the various protocol standards in competition, it would be a wise marketing decision to define a subset of memory controllers supporting a majority of high-volume segments in which applications are both price sensitive and time-to-market driven. Is it possible to define one unique memory interface IP that is able to support various applications? A soft controller IP is by its nature configurable, but the hard PHY IP must be designed for each application. Such a hard macro should be configurable to offer the necessary differentiation that these high-volume segments require. It is interesting to point out the major benefit offered by this unique memory interface IP: robustness and higher reliability. In fact, this unique IP will be integrated by many more chip makers than other IP and the quality will reach maturity level more quickly. Both the SoC integrators and the IP vendor will benefit from an IP concentrating more development and maintenance resource, but which type of memory PHY IP can offer enough configurability?

Let's focus on DDR3/4 and LPDDR3/4 protocols and explore how the related memory PHY IP can be optimized to best support various applications like high-end consumer, mobile and infrastructure. The goal is to rely on one standard hard macro only, which can be fine-tuned to become application specific. The memory PHY IP is the most crucial piece of design in a SoC. If this IP fails, the SoC is simply unusable. That's why the SoC chipmaker will take advantage of a unique memory controller design that is more robust, stable and easier to maintain by the IP vendor versus use of a variety of IP. But chip makers are right when expecting this unique hard macro to be optimized for their application, whether high-end consumer, mobile or infrastructure.

Target Applications and Specific Requirements

Let's review those applications and related market segments where the performance of the integrated memory controller has a direct impact on the system latency or speed, communication bandwidth or power consumption.

Infrastructure

Figure 2: Infrastructure: Data Center

The main challenge for the Infrastructure segment, like in data center and networking applications, can be summarized by this data: Applications require 60% more bandwidth every year. SoCs developed for networking applications require the delivery of high bandwidth while running high performance computing and achieving large memory capacity. High bandwidth, along with large memory capacity and performance requirements, directly translate to the memory controller IP requirements, with the controller being expected to deliver more data per cycle at the highest possible frequency. Fulfilling these two requirements, high bandwidth and high performance, is known to generate high power consumption and possible heat dissipation issues. Finally, it's important to remind SoC architects targeting infrastructure segments that these solutions have to provide a rich set of enterprise-class RAS (reliability, availability, and serviceability) features. Reliability is linked to the mean time between failure (MTBF) which means the average time before the system produces incorrect results. Availability is strongly linked with reliability as it's the amount of time a device is actually operating as the percentage of total time it should be operating. Because no system can be 100% perfect, serviceability defines the simplicity and speed with which a system can be repaired.

The right memory controller IP solution will have to fulfill these performance and RAS requirements to best support the infrastructure segment.

Mobile

Figure 3: Our World Is Going Mobile

The smartphone segment is certainly the most competitive for chip makers today, and the application processor SoC is developed on leading-edge technologies. End users expect to benefit from higher image definition, better sound quality, and ever faster and more complex applications pushing the limits of the application processor performance in terms of frequency (higher) and power consumption (the lower, the better). On top of this performance efficiency goal, SoCs developed for mobile are also very cost sensitive. Because the production volume can reach several dozens of millions of units, optimizing the silicon area is a must, even by a fraction of a square millimeter. Keeping in mind these considerations, it's likely that the memory controller IP should also allow using the same SoC to support both the current memory standard and the yet to come new release. This flexibility allows integrating the cheapest DRAM at the time the mobile phone is sold, but without redesigning the SoC.

We can summarize the most important features that the memory controller IP will have to support for mobile applications: high performance, very low power and flexibility.

High End Consumer

Figure 4: High End Consumer Example (Home Cinema)

Defining high-end consumer segments is not as straight forward as it is for a mobile phone or a data-center. In the consumer segment, we expect to list cost-sensitive applications, ranging from automotive entertainment or ADAS to set-top-box or PC peripheral devices. Time to market is important, but not as important as it is for mobile phone application processors. The end user expects to use a system characterized by high data efficiency more than raw performance. Because the final cost has to be optimized, architects have to design a system in such a way as to minimize the number of components and to select as inexpensive a device as possible, while delivering the most attractive user experience. Good time to market, high data efficiency, high system integration, and high cost sensitivity are the requested features. These high-end consumer segments are varied, but their common need is for the best possible flexibility, allowing the end user to always be in a position to select the optimum solution, specifically in terms of cost.

This need for high flexibility is the specification guidance for the memory controller IP serving the high-end consumer segments.

Memory Controller IP Solution Specifications by Segment

Cadence has been the leader in memory controller IP solutions since the Denali acquisition in 2011. The company delivers memory controllers that are a combination of soft (RTL) IP and the PHY or hard IP, supporting various foundries. For example, with TSMC, the company targets 16nm to support mobile and infrastructure applications and 28nm for consumer applications. We will illustrate how Cadence has deployed the concept of a unique but flexible memory controller and PHY hard IP offering that is able to support infrastructure, mobile and high-end consumer segments by means of configurability.
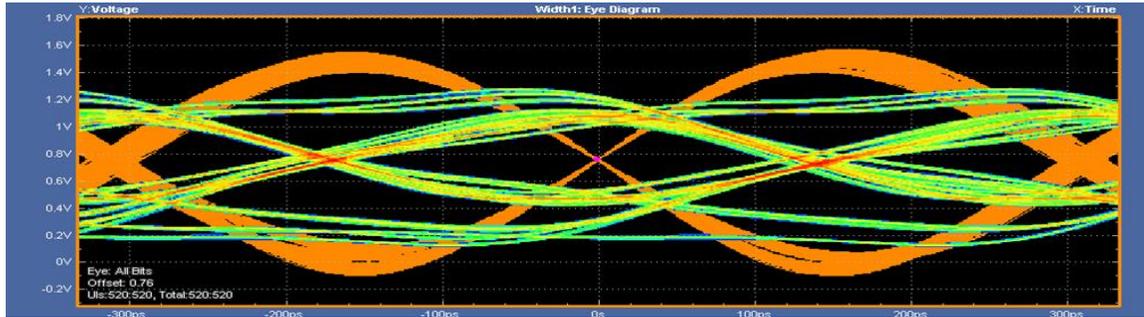
Infrastructure

Figure 5: Demonstrated DDR4 3200Mbps in TSMC 16 FF+ LL

In the data center and networking segments, the memory controller should support the largest possible bandwidth and memory address and offer the best possible reliability (RAS), all while staying power conscious. The protocols supported include DDR3, DDR3L, and DDR4, as there is no need for LPDDRn support. The maximum PHY clock frequency is 1600MHz and the maximum data rate is 3200Mbps. Note that overclocking is not supported in order to maintain maximum data integrity and reliability. The data buses can be 16, 32, 40, 64 or 72 bits wide, with 72 bits being the default. The address bus can be configured in the 13-18 bits range, 18 bits by default, allowing the user to access the largest possible memory space.

The DQ-to-DQS ratio is kept as low as 4:1 (compared to 8:1 in the other two configurations) to minimize the maximum skew between clock (DQS) and data (DQ). The memory controller supports per-rank-leveling (PRL) as well as write leveling for x4 DRAM. These specifications are intended to maximize data integrity and optimize system reliability.

The DFI (interface between the soft and PHY IP) training mode can be set as PHY independent (PI) or PHY evaluation. During PI training, all the training is done in the PHY. The DFI interface allows the PHY to request memory to be placed into a specific memory state and then take ownership of the DFI (keeping the bus idle). The PHY completes all training operations and then releases the bus. Even if supporting PI requires more functionality in the PHY, it helps to avoid any compatibility issues.

In the infrastructure segments, the CPU must use as much memory space as possible, and using Dual Inline Memory Modules (DIMM) is a must-have feature. The Cadence memory controller IP solution supports registered DIMMs (RDIMM), unregistered DIMMs (UDIMM) and load reduced DIMMs (LRDIMM). Note that this feature is infrastructure specific and these DIMMs are not supported in the other segments.

This memory controller supports 3DS specification, with two, four or eight devices stacked and a default value of four devices. 3DS definition enables DRAM stacking with better electrical characteristics, as the memory devices are connected using a Through-Silicon-Via (TSV) instead of wire bonding. The memory controller will only communicate with the master (DRAM) chip, minimizing signal load.

A final point to consider is that the infrastructure-oriented memory controller does not support power gating, unlike the memory controller targeting high-end consumer or mobile segments.
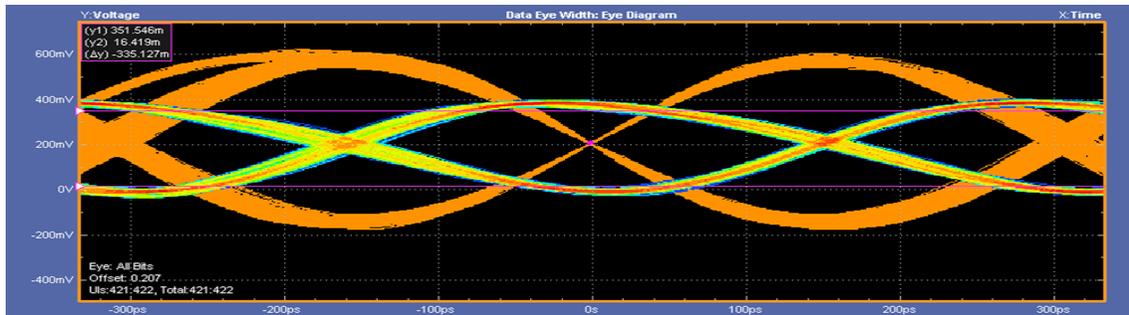
Mobile

Figure 6: Demonstrated LPDDR4 3200Mbps in TSMC 16 FF+ LL

The protocols supported in the mobile segment include LPDDR3, and LPDDR4, as DDRn support would lead to extra power consumption. The maximum PHY clock frequency is 1600 MHz and the maximum data rate 4266Mbps, indicating that overclocking is supported. It may look strange to see a higher data rate for mobile than for infrastructure - until you realize that the end user expects to benefit from good entertainment experience, and is not necessarily expect the maximum possible data integrity. When watching a video, this high data rate guarantees good image quality, but when/if a wrong pixel (due to one data bit in error) is inserted in the movie or video, it doesn't impact the user experience.

The data buses can be 16-, 32, 40, 64, 72 bits wide, with 32 bits being the default. The address bus can be configured at 10 bits (for LPDDR3 only) or 12 bits, far lower values than for the infrastructure segment. In fact, the capacity of the DRAM integrated in a mobile phone is limited to a few GBytes and there is no need to support a wider address bus.

The DQ-to-DQS ratio is 8:1, a relaxed specification compared to the previous configuration (infrastructure), and the maximum skew between clock (DQS) and data (DQ) is probably higher. The memory controller only supports PRL, there is no support for write leveling for x4 DRAM. The DFI training mode is set only to PI. In the mobile segment, the memory controller specification must be kept as simple as possible to allow the fastest possible design: rather than RAS features, the main concern is time to market!

There is no 3DS support in this memory controller option, as this choice aligns with the mobile industry status. No complex and expensive TSV implementation for DDR4 is needed.

The last feature, power gating, is really associated with the mobile industry needs. For the end user, the capability of the smartphone to be used as long as possible before having to be charged can be the main factor influencing a purchase decision; it's certainly within the top three concerns when buying a system. That's why the memory controller IP targeted for the mobile industry should provide a power-gating feature.

High End Consumer

If we remember that the consumer segments are varied, but all cost sensitive, we understand that the system designers need flexibility first. This flexibility allows the selection of the cheapest DRAM device at the time of system release as well as the ability to change the supported protocol for opportunistic reasons. That's why this memory controller IP supports all of the following protocols: DDR3, DDR3L, DDR4, LPDDR4, LPDDR3, and even LPDDR2.

Moreover, the SoC architect should have an extended choice for the technology node. Selecting FinFET technology like 16nm may be an option too expansive in term of non-recurring engineering costs (NRE), as it is not optimized in terms of time to market for a SoC targeting high-end consumer applications, especially if the SoC doesn't need the highest performance. Cadence has developed the same memory controller and PHY IP solution for 28nm to offer another option for the high-end consumer segment.

For this high-end consumer-oriented memory controller, the only difference between 16nm and 28nm is the maximum data rate: the memory controller supports a data rate up to 4266Mbps in 16nm, but only 3200Mbps in 28nm. All the other features are identical.

The high-end consumer-specific memory controller IP is very similar to the memory controller supporting mobile applications and we will only highlight the few differences in term of supported features, as the other features are equivalent to the mobile-specific memory controller. The number of address bits is 13 to 17 (by default 17) for the consumer option, 10 to 12 for the mobile option. In fact, the high-end consumer segment application may integrate larger memory space than mobile. The number of Clock Enable signals is 1, 2, 4 or 8 (4 by default) for the consumer option, when it's only 1, 2 or 4 (2 by default) for the mobile.

The memory controller IP configurations dedicated to the high-end consumer applications are characterized by a much greater flexibility than for mobile, as a consumer system may support all the possible protocols, and the PHY can adapt to a SoC in 28nm in addition to 16nm.

Conclusion

The evolution of Moore's law leads us to re-consider technology choices with respect to the targeted market segment, as selection of the most advanced technology node is no longer the only available option. It may be wiser to select 28nm instead of 16nm for segments like high-end consumer, for example. These considerations are pushing memory controller IP vendors to offer differentiated products in each targeted segment. The differentiation is also made through configuration of various features, like power gating, 3D stacking support, number of address bits, overclocking, and more. We have described three configurations, where the memory controller is optimized to best support the specific requirements of high-end consumer, mobile, and infrastructure applications.