Increasing Efficiency with Hard Memory Controllers in Low-Cost 28 nm FPGAs

WP-01188-1.1 White Paper

With the hard memory controller (HMC) in Altera® Cyclone® V FPGAs, designers can maximize efficiency and flexibility, thereby achieving low power and low cost for their systems and applications.

Introduction

The efficiency of the memory controller is an increasingly critical component in determining the actual bandwidth of a system. This effective bandwidth is a critical factor in determining the actual performance of a system. With the HMC in Cyclone V FPGAs, you can maximize the efficiency and flexibility, and achieve the low power and low cost for your applications and system.

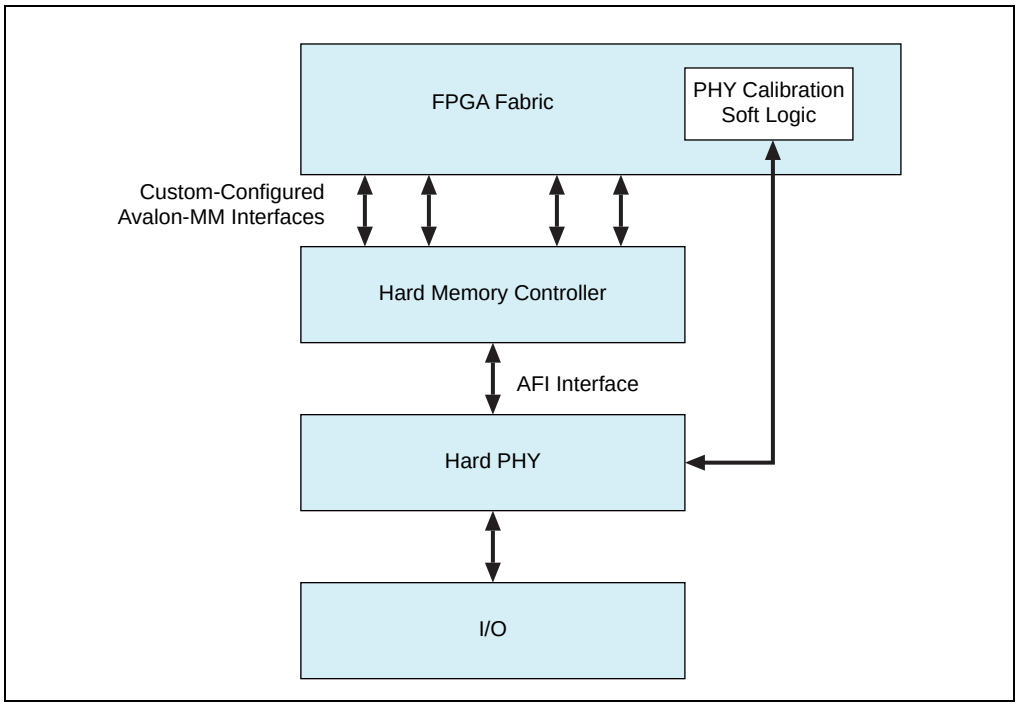
As the low-cost family member of Altera's 28 nm product portfolio, Cyclone V FPGAs provide the industry's lowest system cost and power, along with high functionality and performance levels that make the device family ideal for differentiating designers' high-volume applications. The HMC is a component that enables easy-to-use high-speed DRAM interfacing for Cyclone V FPGAs.

101 Innovation Drive
San Jose, CA 95134
www.altera.com

Memory Interface Architecture

The HMC is composed of three primary blocks: the memory controller, the memory PHY, and the I/O structure, as illustrated in Figure 1.

Figure 1. Cyclone V HMC Architecture



The memory controller implements an interface to the FPGA fabric, accepts bus commands from user logic, translates them into the correct commands for the DRAM, and handles the details of the DRAM access. In addition, the Cyclone V HMC manages multiple bus transactions. By doing so, one memory controller can be shared between multiple processes, eliminating the need for the designer to develop custom logic to manage these details.

The memory PHY used in the HMC uses a combination of hard and soft logic to implement the timing synchronization and calibration, the RAM chip initialization, and the data rate conversion.

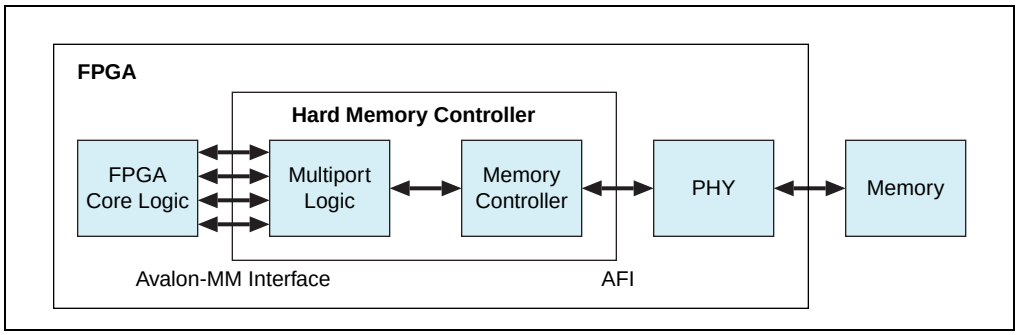
The I/O structure works closely with the PHY to implement the correct signaling as specified by the JEDEC standard for DRAM. The Cyclone V family supports DDR3, DDR2, and LPDDR2 standards.

For additional flexibility, the memory controller or the memory controller and the PHY can be bypassed, allowing the designer to use a customized memory controller or PHY instead.

The MPFE

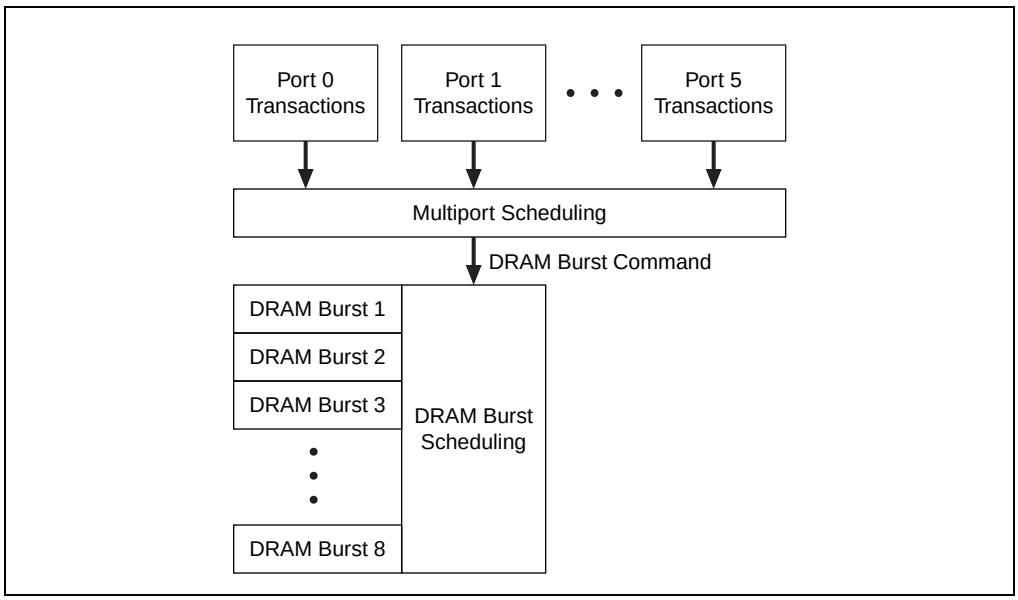
The hard multiport front-end engine (MPFE) is a new feature in Cyclone V FPGAs. It allows multiple internal processes in the core logic to access the same HMC. As shown in Figure 2, the MPFE is one of the hard blocks in the HMC, connecting the memory controller and FPGA core logic. The interface between core logic and the HMC is the Avalon® Memory-Mapped (Avalon-MM) system bus. The HMC interfaces the PHY via the Altera PHY Interface (AFI).

Figure 2. Simplified Diagram of the Cyclone V HMC



The multiport logic allows up to six local interfaces from the core logic to access a controller. The MPFE acts as the scheduler and chooses which bus transaction should be served next. As shown in Figure 3, the complete memory controller provides two levels of scheduling for the six addresses and command ports. Multiport scheduling takes place in the MPFE and DRAM burst scheduling takes place in the high-performance memory controller.

Figure 3. Cyclone V HMC Scheduling



The multiport scheduler in the MPFE decides which of the active addresses and command ports to service next. Bus transactions longer than a DRAM burst are scheduled as a series of DRAM bursts, with each DRAM burst being separately arbitrated. This approach ensures that high-priority traffic is served as soon as possible and that long bursts and short bursts can be effectively interleaved on ports. These bursts then are serviced based on a deficit round robin (DRR) arbitration

algorithm. The algorithm uses two components: the absolute priority for a transaction and the weight of a port. Absolute priority is intended for applications where one master should always get priority above or below others. Port weight is used to express relative priority between ports at the same absolute priority level. These two components can be updated dynamically, and allow for temporary over and under service so the controller can smooth out memory traffic.

In addition to bringing high efficiency and better performance to a system, a MPFE can also be used as a good approach to save I/O count in many applications. Suppose that we have an application where two external memory interfaces (each 16 bits wide) are needed, as shown in Figure 4. The MPFE provides the designer with the option of integrating these two memory interfaces into a shared memory configuration, thereby reducing component count, saving board space, and maximizing device I/Os. For example, Figure 5 shows a solution where those two agents (or core logic applications) share one HMC interface with 32 bit-wide scheduling by the MPFE. In this case, the I/Os in the bottom banks are available for other system requirements.

Figure 4. Two HMCs Consume I/Os

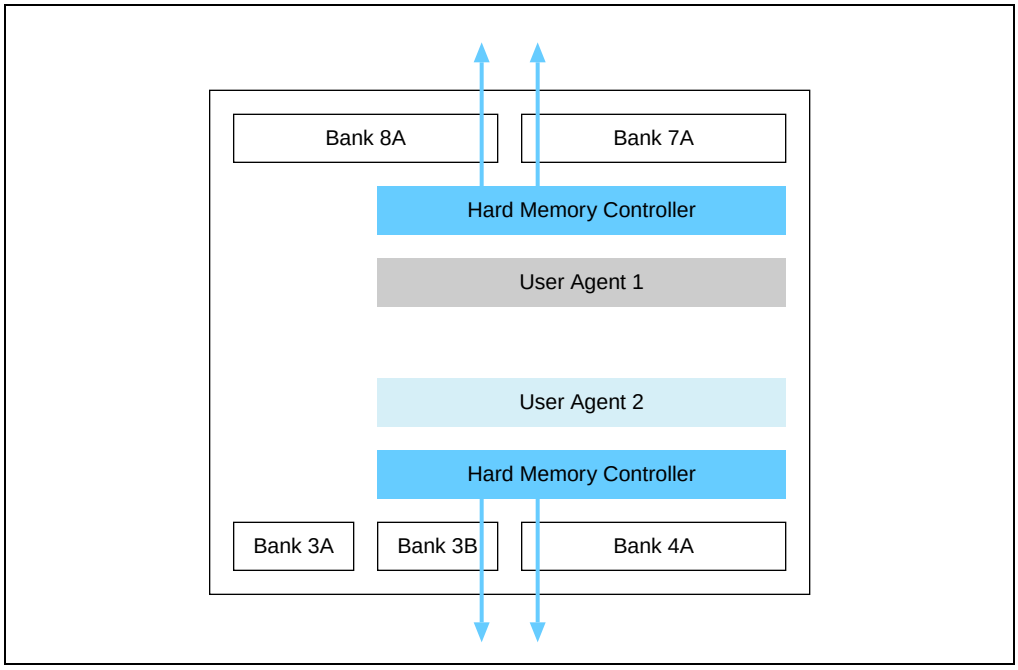
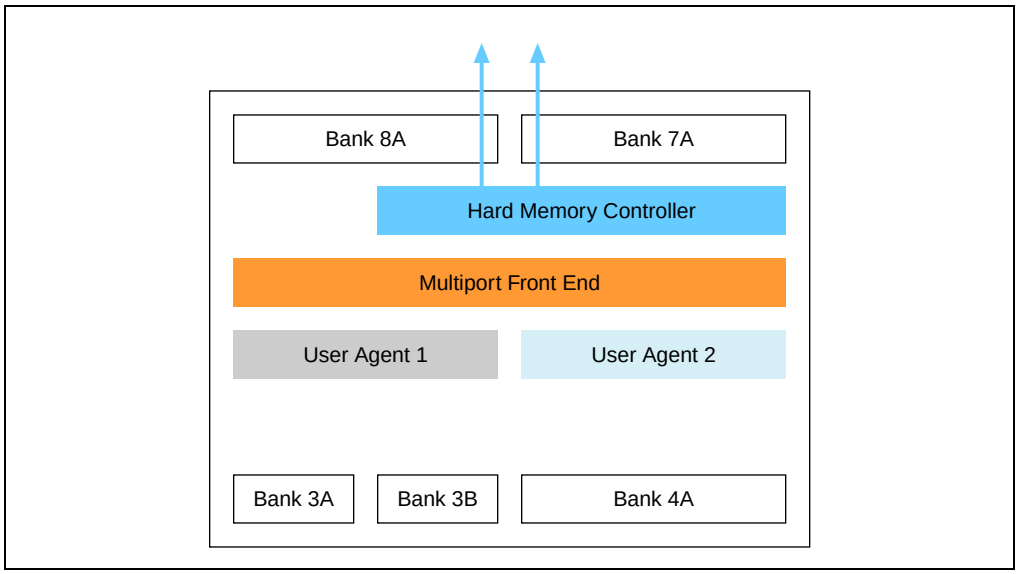


Figure 5. Consolidation of Two HMCs



The Controller

Investing in higher speed memory interfaces means an investment in memory bandwidth. Bandwidth depends on the efficiency of the memory controller controlling the data transfer to and from the memory device, as illustrated in Equation 1.

Equation 1. Bandwidth Calculation

$$\text{Bandwidth} = \text{Data Width (bits)} \times \text{Data Rate} \times \text{Efficiency}$$

Example 1 shows a 32 bit interface that has 70% efficiency and runs at 400 MHz frequency has the bandwidth of 17.92 Gbps.

Example 1. Calculating Bandwith

$$\text{Bandwidth} = 32 \text{ bits} \times 2 \text{ Clock Edges} \times 400 \text{ MHz} \times 70\% = 17.92 \text{ Gbps}$$

While DRAM typically has an efficiency of around 70%, the Altera memory controller increases the efficiency up to 92%.

The efficiency of the controller at managing the required DRAM commands is critical in achieving maximum bandwidth through the interface. As shown in Equation 2, efficiency is the amount of occupied DQ cycles (non-idle) that occur on the bus divided by the total number of cycles on the bus.

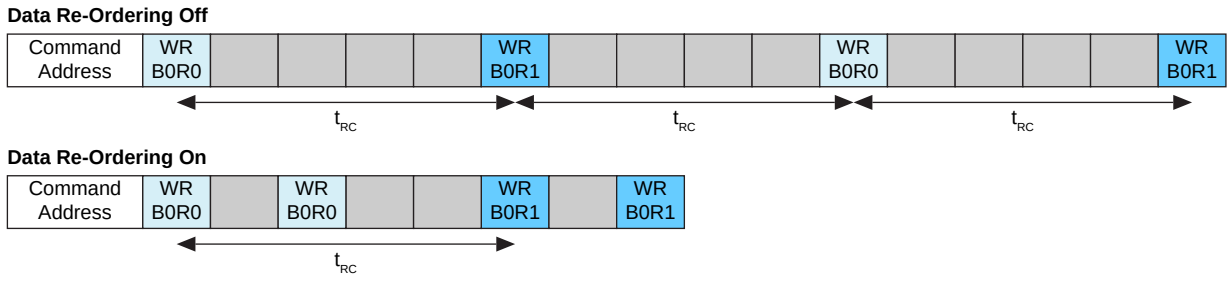
Equation 2. Efficiency Calculation

$$\text{Efficiency} = \frac{\text{Number of Clock Cycles That DQ Bus Is Occupied}}{\text{Number of Clock Cycles in the Period}}$$

Within the controller, DRAM burst scheduling chooses the best command to execute among the pending DRAM bursts to increase efficiency. DRAM memories are most efficient when making consecutive reads and writes, especially within the same row (intra-row operations). When the DRAM operations cross row or bank boundaries, the idle time (t_{RC}) increases, reducing efficiency. The core of the high-efficiency HMC is an inter-row data reordering and command reordering engine.

The inter-row data reordering serves to minimize tRC by reordering commands going to different bank and row address; commands going to the same bank and row address are not ordered. Figure 6 illustrates how inter-row data reordering works.

Figure 6. Data Reordering for Minimum t_{RC} *(1)*



Note:

(1) t_{RC} is the activate-to-activate (same bank) timing requirement.

DDR protocols are naturally inefficient, because commands are fetched and processed sequentially. The waste of potential cycles means that the DDRx command and DQ bus are not fully utilized, resulting in decreased efficiency. Shown in Figure 7 and Table 1, the command reordering, or advance bank management, feature allows banks for future transactions to be opened before the current transaction finishes data.

Figure 7. Advance Bank Management Waveform

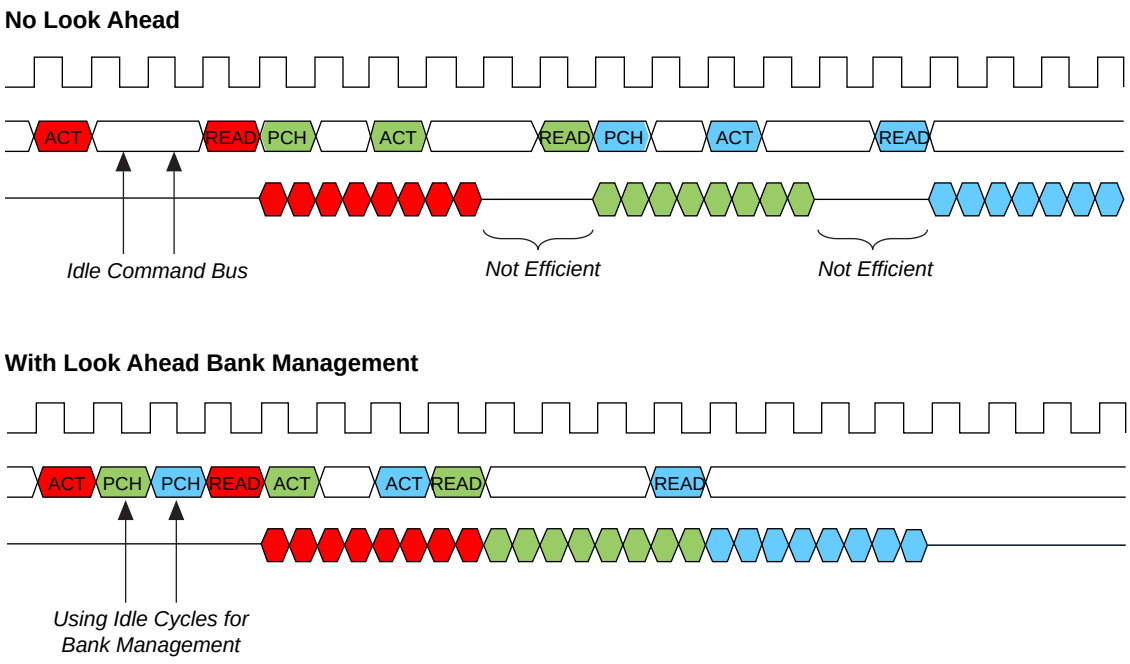


Table 1. Advance Bank Management

Command	Address	Condition
Read	Bank 0	Activate Required
Read	Bank 1	Pre-Charge Required
Read	Bank 2	Pre-Charge Required

The Cyclone V HMC supports memory error correction codes (ECCs), both calculated by the controller and by the user. ECCs can correct for some data corruption errors, increasing data integrity and DRAM efficiency. For user ECCs, the controller does not distinguish between data and ECC bits. In user ECC mode, all bits are treated as data bits and written to and read out of memory. User ECCs can be used to implement non-standard memory widths (e.g., 24 and 40 bits) where ECCs are not required. Controller ECCs are supported for 16 and 32 bit widths, each case requiring eight additional bits of memory for an actual memory width of 24 and 40 bits.

The PHY

The Cyclone V device family supports hard and soft memory interfaces. Hard memory interfaces use the hard memory controllers (with hard multiport front-end blocks embedded) and hard UniPHY blocks in the device. Soft memory interfaces use soft UniPHY intellectual property (IP). You can also build a custom PHY, a custom controller, or both, as desired.

The hard UniPHY is instantiated together with the hard memory controller. In addition to the PHY data path that uses the hard IP blocks in the devices, the hard UniPHY also uses the dedicated hardware circuitries for certain component management, including the read-write (RW) and PHY managers, which saves logic element (LE) resources, and to allow better performance and lower latency.

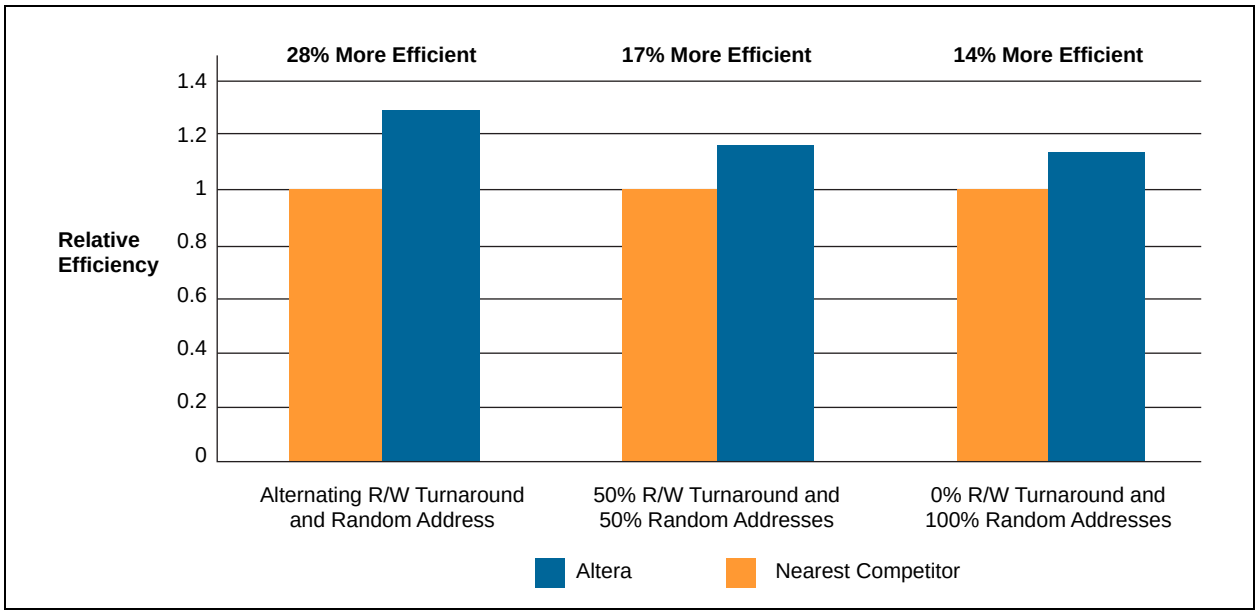
The Cyclone V FPGA has a fixed number of hard PHYs. Dedicated I/O pins with specific functions for data, strobe, address, command, control, and clock must be used together with each hard PHY.

Comparing to the hard memory PHY, the soft memory PHY gives you the flexibility to choose the pins to be used for the memory interface. The soft memory PHY also supports wider interfaces as compared to the hard memory PHY.

To achieve better performance including lower latency and higher efficiency, both the soft and hard UniPHY IP are optimized to take advantage of the Cyclone V FPGA's I/O structure and the Quartus® II software's TimeQuest Timing Analyzer. Using the HMC with UniPHY IP ensures the highest reliable frequency of operation across process, voltage, and temperature (PVT) variations.

Cyclone V FPGAs have built-in circuitry in the I/O element to convert data from full rate (the I/O frequency) to half rate (the controller frequency) and vice versa. Memory controllers with UniPHY IP and the Altera memory controller MegaCore® functions can run at the half the frequency of the I/O interface of the memory devices to allow better timing management in high-speed memory interfaces, as shown in Figure 8. The UniPHY IP can also dynamically choose the number of DQS delay chains that are required for the system.

Figure 8. Memory Efficiency of DDR3 SDRAM Memory Controllers with UniPHY



Low Power

Hardening the memory controller moves this functionality from soft logic to hard logic gates. This implementation is more power efficient. By reducing signal path length, the HMC circuit has less capacitance, thereby reducing the circuits overall dynamic power consumption.

In addition to the DDR2 and DDR3 memory interfaces, the Cyclone V HMC also supports LPDDR2 which reduces system power with low voltage and multiple power-saving features in the DRAM, including:

- Self-refresh provides the capability to disable I/O switching and clock after a JEDEC-specified time. The memory controller also supports partial array self-refresh and automatic refresh frequency adjustment for temperature-sensitive DRAMs depending on the RAM device support.
- Power-down and deep power-down provide the capability to disable the clock enable pin. You can program the exit timer to set the exit to be fast or slow.

By requesting these technologies, the HMC can hit the lowest power modes and the quick entry/exit.

Low Cost

The Cyclone V HMC contributes to lower the total system cost in terms of:

- Up to 2 HMCs, each of which saves as many as 11K LEs and 11 M10K RAM blocks
- Wide interface (up to 40 bit) reduces command and address pin count
- 8 bit and 16 bit wide interface designed for cost-optimized systems
- Bonding the interface to support wider data bits
- MPFE allows multiple users functions to efficiently share one memory interface
- Soft controller available for ultimate flexibility

With those advantages, the designer can maximize the productivity with lower cost and shorter time to market.

Conclusion

Cyclone V FPGAs provide you the most efficient and lowest latency HMCs, along with the high flexibility, low power, and low cost. All of those allow you easily interface with today's higher speed memories in a more cost- and power-effective way. Altera is committed to continuously help designers create fast and robust memory interfaces, to not just make those designs successful but to ensure that implementation is both fast and easy.

Further Information

- Cyclone V FPGAs: Lowest System Cost and Power: www.altera.com/devices/fpga/cyclone-v-fpgas/cyv-index.jsp
- External Memory Interface Spec Estimator: www.altera.com/technology/memory/estimator/mem-emif-index.html
- *External Memory Interfaces in Cyclone V Devices*: www.altera.com/literature/hb/cyclone-v/cv_52006.pdf
- *External Memory Interface Handbook*: www.altera.com/literature/lit-external-memory-interface.jsp

Acknowledgements

- Jay Lu, Product Marketing Manager, Asia Pacific, Altera Corporation

Document Revision History

Table 2 shows the revision history for this document.

Table 2. Document Revision History

Date	Version	Changes
November 2012	1.1	Minor text edits.
November 2012	1.0	Initial release.