



WP454 (v1.1) March 23, 2015

High-Performance, Lower-Power Memory Interfaces with the UltraScale Architecture

By: *Tamara Schmitz*

Xilinx® UltraScale™ architecture-based devices, used in conjunction with DDR4 DRAMs, provide highly significant gains over previous generations in memory interface bandwidth, flexibility, and power use efficiency.

ABSTRACT

With bandwidth needs growing from one system generation to the next, the external memory interface must keep up with performance requirements while staying within the maximum power dissipation constraints inevitable in many applications.

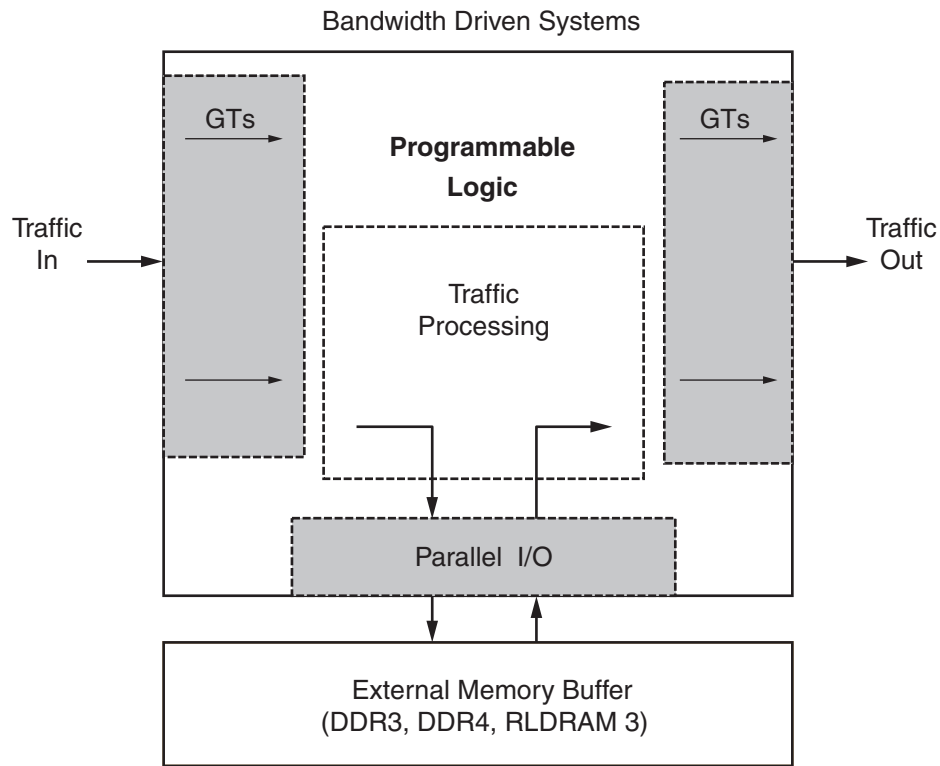
For higher-bandwidth systems, package size and pin count become limiting factors so I/Os must be used efficiently. All I/Os in the UltraScale architecture are memory interface capable, and the architecture allows I/O bank sharing of two interfaces. This provides the flexibility needed to simplify system design and maximize aggregate memory bandwidth per FPGA.

This white paper describes the types of innovative implementations available to designers using the UltraScale FPGA families. The UltraScale architecture enables building high-performance, lower-power DDR4 interfaces—at up to 2,666Mb/s—that are flexible enough to meet high-bandwidth, FPGA- and SoC-based system requirements.

Applications: Higher Bandwidth, Increased Flexibility, Lower Power

With next-generation systems, performance requirements are again increasing to fulfill the insatiable need for higher bandwidth. For example, wired communication systems are moving from 40Gb/s to 100Gb/s, 200Gb/s, or even higher rates.

In a typical system (see [Figure 1](#)), the incoming traffic from the gigabit transceivers needs to be processed in the FPGA fabric. Due to on-chip RAM size constraints, the traffic might have to be buffered externally in higher density memory devices. In many cases, multiple devices or interfaces (DDR4, DDR3, or RLDRAM3) need to be implemented to meet the external bandwidth needs. The FPGA processing rate and the external buffering rate need to match the traffic rate to avoid any stalls or system performance degradations. This means that the external memory bandwidth rate needs to exceed the traffic rate to compensate not only for read and write overhead, but also for the inevitable inefficiencies caused by various traffic patterns.



WP454_01_031315

Figure 1: Bandwidth-Driven Systems and the External Memory Buffer

The bandwidth of the external memory interface for an FPGA depends on several factors:

- **Number of interfaces**
(determined by the number of I/Os available in the package and their efficiency)
- **Data rate per bit**
- **Data bus width**
- **Data bus efficiency**
(percentage of time data is actually being transferred)

The total effective bandwidth can be expressed as a product function of all these factors as shown in [Equation 1](#):

$$TEB_{FPGA} = INTF \cdot DR_{BPS} \cdot DW_{BIT} \cdot DBE_{PERCENT} \quad \text{Equation 1}$$

where

TEB_{FPGA}	=	FPGA Total Effective Bandwidth
$INTF$	=	Number of Interfaces
DR_{BPS}	=	Data Rate (bits/sec)
DW_{BIT}	=	Data Width (bits)
$DBE_{PERCENT}$	=	Data Bus Efficiency (%)

Additionally, today's systems have become more compact, and thus tend to have constraints that limit power dissipated as heat by the FPGA package. Memory interfaces can be power-hungry, especially in the I/O domain where current-drawing termination schemes are used for signal integrity (SI) reasons. New design methods need to be devised that require less power, both through reduced voltage swings (like the 1.2V POD DDR4 I/O), as well as through circuit optimizations when possible.

Memory Trends and FPGA Solutions

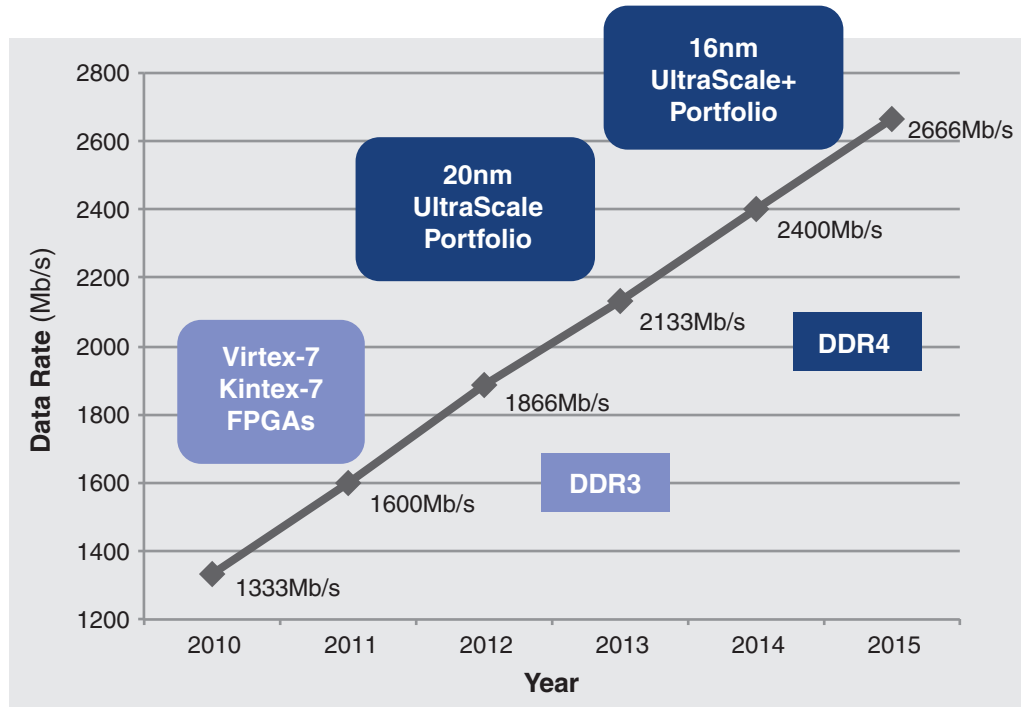
Traditionally, FPGA users have leveraged commodity DRAM for several generations of products, and are likely to continue to do so.

The DRAM market has typically followed PC market trends. Recently, however, the explosive growth of the mobile market (e.g., smart phones and tablets), coupled with data centers' needs for servers capable of both faster transmission rates and reduced power consumption, have become bigger drivers of DRAM demand than the legacy PC market.

DDR3 and DDR3L (the latter being the 1.35V I/O version of DDR3) have dominated the market, and are likely to continue to do so for the short term. However, the market is expected to gradually ramp up to DDR4 starting in 2014. The first DDR4s are most likely to be seen in server applications, leveraging the lower I/O voltage (1.2V) and power savings DDR4 offers, but at higher data rates than DDR3.

Users transitioning to the higher DDR4 data rates can take advantage of the speed grades that DDR4 offers: from 1600 to 2666Mb/s. Comparing maximum data transfer speeds, the mid- and high-speed grade UltraScale FPGAs offer 30% higher data rates—2400Mb/s with DDR4 DRAM versus 1866Mb/s for 7 series devices and DDR3. For an additional performance boost, 16nm

UltraScale+™ devices offer 2666Mb/s in mid- and high-speed grades, as shown in Figure 2.

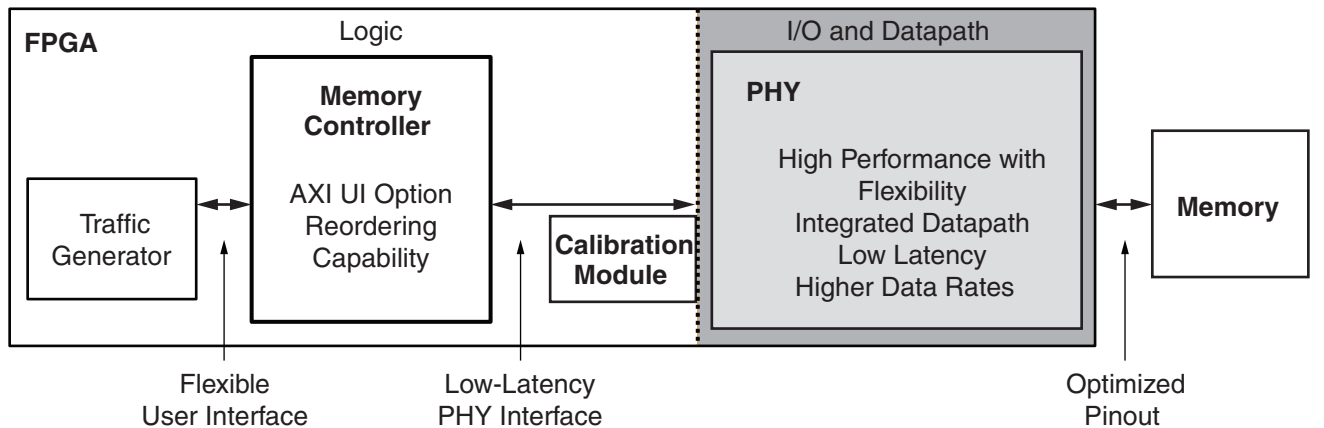


WP454_02_031315

Figure 2: DDR3 to DDR4 Transition and Data Rates Supported by UltraScale Architecture Portfolio

Flexible Architecture: Maximizing Memory Bandwidth

The UltraScale architecture-based memory interface solution has been completely re-architected with innovative silicon features optimized for higher performance with maximum flexibility at minimum power. Figure 3 outlines the major building blocks of the solution.



WP454_03_061714

Figure 3: Memory Interface Solution Optimized for Performance, Lower Power, and Flexibility

The optimized pinout maximizes the number of interfaces that can be implemented in a given FPGA package, and therefore the total bandwidth per device. The integrated PHY building blocks use dedicated paths to minimize latency and facilitate timing closure on high-frequency interfaces. The PHY is designed to be flexible to meet an FPGA's configurability needs, while at the same time provide a no-compromise solution for DDR4's higher data rate and SI needs.

The calibration logic is controlled by a MicroBlaze™ processor that provides added flexibility and ease of use. It allows the transfer of the calibration and timing margin data to the user for enhanced debugging capabilities. This is made possible with the System Debugger Tcl Mode (XSDB).

The memory controller features a reordering capability to maximize data bus efficiency and increase effective bandwidth. Additionally, it has the flexibility to connect to an AXI bus, providing an easy interface with other AXI IPs or to enable a multi-porting capability.

PHY Architecture for Increased Performance and Flexibility

The UltraScale architecture-based PHY solution consists of four byte-wide (13-bit) PHYs per I/O bank, along with two dedicated high-speed Tx PLLs and an MMCM for general clocking flexibility. The two Tx PLLs allow for two independent memory interfaces per bank. The four bytes support 52 I/Os per bank. Each 13-bit-wide byte is a high-speed digital PHY capable of low-latency transfer of data, address/command, and clocks to and from fabric and I/Os. The digital PHY is a fully integrated PHY built from the ground up as a no-compromise PHY solution for high-speed memory. *No compromise* means the PHY was built foremost to support DDR4 rates at lower power while still providing configurable flexibility that an FPGA requires. Features include ultra-low jitter due to its isolated supply, ultra-fine granularity de-skew and quarter delay shifting, and built-in self calibration against PVT variance with real-time VT tracking during live data. See [Figure 4](#).

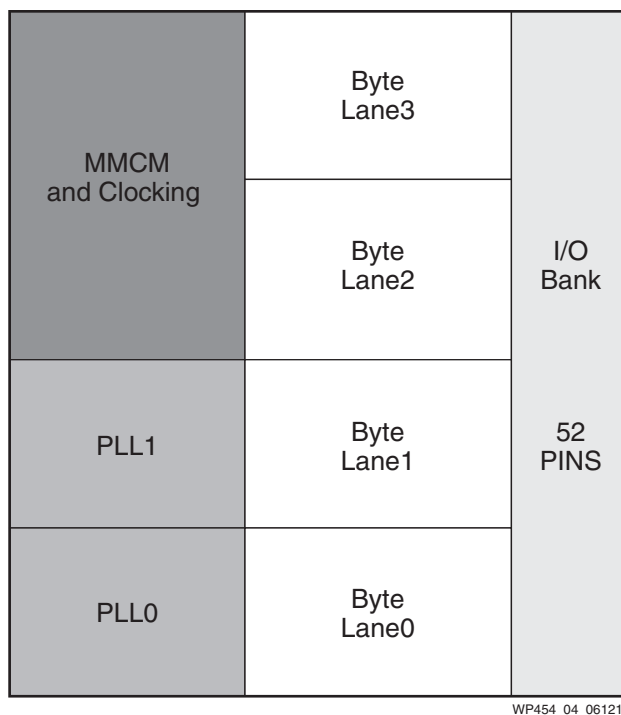
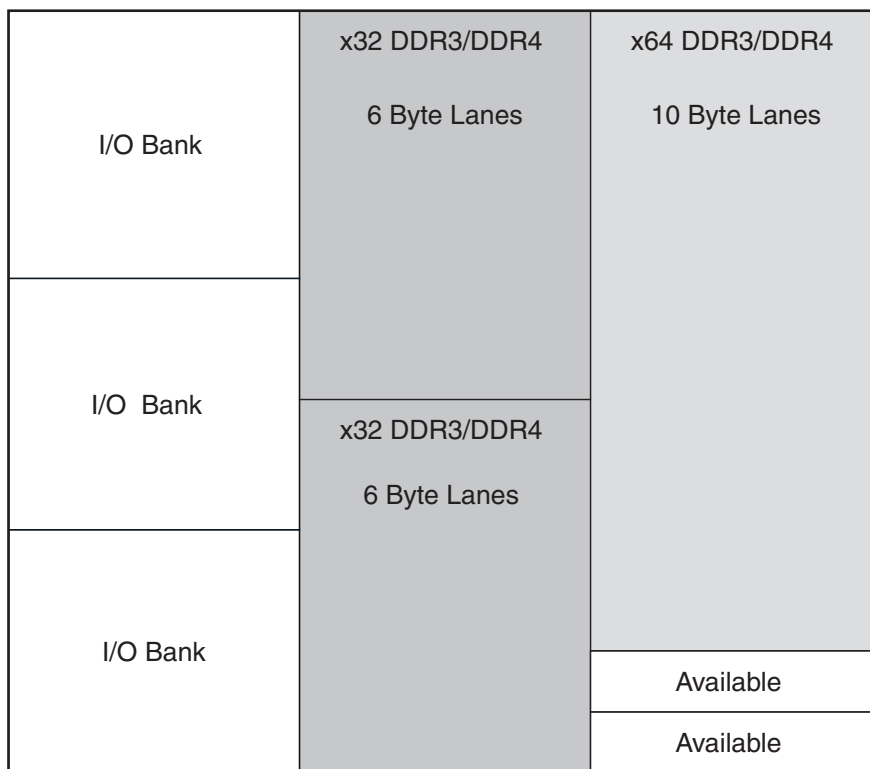


Figure 4: Bit Slice PHY Architecture for Increased Flexibility

DDR3 or DDR4 interfaces (32-bit or 64-bit) are very common. Figure 5 illustrates how the benefits of flexible architecture and I/O bank sharing between two interfaces can maximize pin usage.



WP454_05_061714

Figure 5: Sharing of I/O Banks between DDR3 or DDR4 Interfaces

An I/O bank can be split between two interfaces and rates. In addition, an I/O bank can be split at any byte-lane boundary. Two 32-bit interfaces can fit in three I/O banks, while a 64-bit interface can fit in two and a half banks, leaving two byte lanes available for other usage.

DDR4 Controller Options for Higher Bandwidth Efficiency

The DDR4 controller has optimized functionality compared to the previous-generation DDR3 controller. The new controller functionality leverages the bank group feature in the DDR4 architecture to improve data bus efficiency and lower access latency. Additionally, it features an optimized command queue structure that improves bandwidth efficiency by means of its capability to reorder commands and group reads and writes for faster bus turnaround.

Another notable feature of UltraScale architecture-based memory controller is its ability to improve command bandwidth utilization with improved internal clock timing, resulting in a shorter command-to-read data latency.

For improved flexibility, the soft controller provides users with options to customize the page management algorithm for maximum bandwidth. Users have the flexibility to set the page management option that best suits their application-specific command pattern.

UltraScale Architecture Benefits

Table 1 outlines the improvements in data rate (30–40% higher) and bandwidth (1.3X–3X higher, depending on device and package) that are available with UltraScale FPGAs. The new architecture improves latency of the PHY and the flexibility of I/O banks, while enhanced pin utilization improves the overall device bandwidth. The programmable I/O delays have finer resolution (5ps) and therefore allow for improved timing margins.

Table 1: Memory Interface Benefits: UltraScale Architecture vs. 7 Series FPGAs

Metric	Kintex/Virtex 7 Series FPGAs	UltraScale Devices	Benefit
Data Rate/Bandwidth	1866Mb/s	2400Mb/s (20nm) 2666Mb/s (16nm)	30–40% data rate increase; 1.3X to 3X bandwidth increase
PHY Latency	OK	Excellent	Faster command to data access
I/O Bank Flexibility	Good	Excellent	Improved pin utilization and device bandwidth
Number of I/Os per Bank	50	52 + 2 V_{REF}	Improved bank utilization and device bandwidth
Programmable I/O Delay	78ps	5ps	Improved timing margin
Advanced I/O Features	—	Pre-emphasis and Equalization	Better SI for higher DDR4 rates
DDR3/DDR4 Memory Depth Support	Number of ranks: 2 x8-based DIMMs	Number of ranks: 4 x8/x4-based DIMMs	Improved memory access depth
User Interface	AXI-4 Option for Multi-port	AXI-4 Option for Multi-port	Improved flexibility and ease of use with Vivado IPI software
DDR3/DDR4 Memory Controller	Good	Excellent	Improved efficiency and user flexibility
I/O Power	Good	Better with DDR4	Reduced total power

Advanced I/O features like pre-emphasis and equalization enable the use of higher data rates because of improved signal integrity. Pre-emphasis and de-emphasis have been used by Xilinx for many years in gigabit transceiver implementations to suppress low-frequency signal components at the transmitter. The same technique is used for the DDR4 I/O interface to improve the quality of the write channel. The de-emphasis implementation reduces the inter-signal interference (ISI) to ensure correct signal sampling at the memory device receiver. At the FPGA input, similar equalization techniques are used to boost the high-frequency components of the signals. Continuous-time linear equalization (CTLE) adds a programmable high-pass filter to the receiver to achieve a balance between the high-frequency and low-frequency components of the data stream.

The x4 bit (4-bit data width) device support and the multi-rank capability of the PHY enable improved memory depth. Using a x4 device rather than a x8 or x16 device for the same data bus width can provide two or four times the memory depth. Additionally, the multi-rank capability of the PHY enables multiple data or DIMM loads on the same address bus, increasing the memory depth of the system. Up to four ranks can be calibrated with the advanced PHY capabilities in the UltraScale architecture to ensure robust timing margin for quad-rank implementations. Finally, with DDR4 support, the I/O power is reduced, which benefits the overall system's power requirements.

Meeting Lower Power Requirements

Lower power requirements are necessary in today's systems, especially with memory interfaces that must transfer data to and from the FPGA at increasing rates. Transitioning to the UltraScale architecture with DDR4 DRAM can virtually double the total power benefit.

This is attributable to the lower 1.2V I/O voltage of the DDR4 interface as well as to several new power-saving features of the DDR4 architecture; as a result, power savings on the DRAM and I/O side have become significant. TSMC's 20-SoC process, which combines high performance and low power, and the UltraScale architecture's PHY design have also contributed to the substantial power savings.

In [Table 2](#), two comparisons are made (in the 2nd and 3rd rows) for a 32-bit, 20nm UltraScale FPGA memory controller/interface with DDR4 DRAM running at 1866Mb/s and 2400Mb/s, respectively, against a legacy 7 series FPGA and DDR3 DRAM running at 1866Mb/s, with all configurations executing 50% Read/Write transactions. The percent of power savings for each of the two UltraScale architecture/DDR4 configurations is given in the fourth column of the table:

Table 2: Power Savings with DDR4 and 20nm UltraScale FPGAs (Using XPE 2014.1)

Memory Interface, FPGA	Data Rate, Data Width	I/O and PHY Power Consumption	% Power Savings: DDR4 UltraScale FPGA vs. DDR3 7 Series FPGA
DDR3, 7 series FPGA	1866Mb/s, 32 bits wide	1.41W	—
DDR4, UltraScale FPGA	1866Mb/s, 32 bits wide	0.94W	36%
DDR4, UltraScale FPGA	2400Mb/s 32 bits wide	1.10W	15%

Ensuring Increased Productivity: Vivado Memory Interface Generator (MIG)

Designing a complete memory interface and controller for a custom configuration requires extensive modification of the base design to match the interface and FPGA device requirements for I/O placement and various memory device settings.

A complete memory controller and interface design can be generated with the Vivado® Design Suite's MIG GUI. It is available from Xilinx as part of the Vivado IP catalog. The benefit of the MIG GUI is that there is no need to generate the RTL code from scratch for the interface and controller or manually modify the existing example design code. The MIG GUI generates the RTL and constraints files based on user inputs. These files are based on a library of hardware-verified designs, with modifications coming from the user's inputs via the GUI.

The MIG GUI enables faster customization and implementation of the design. The MIG produces the customized design with the necessary constraints that ensure the desired performance. The designer has complete flexibility to further modify the RTL code or the constraints file. Unlike other solutions that offer black-box implementations, the MIG GUI outputs unencrypted code, to allow further customization of a design.

The MIG output files are categorized in modules that apply to different building blocks of the design—user interface, PHY, controller state machine, etc. The user can also optionally connect a different controller to the PHY generated by the MIG.

Additionally, the MIG generates a synthesizable testbench with memory checker capability. The testbench is a design example used in the functional simulation and hardware verification of the Xilinx base design. By issuing a series of Writes and Reads to the memory controller, the testbench can be used as a template to generate a *custom* testbench, which can estimate bandwidth efficiency and verify the expected performance of different memory access patterns.

For additional I/O placement flexibility, the MIG I/O bank selection can be customized at the I/O pin level by using the Vivado I/O pin planner. This feature provides a better match of board layout requirements with the MIG I/O guidelines. It also improves I/O utilization when multiple interfaces are implemented to maximize per-FPGA bandwidth.

Ensuring a Robust Design: Hardware Verification and Characterization

Hardware verification of memory interface and controller IP is an important step to ensure a robust, reliable, high-performance solution. For several generations of FPGA products, in fact, Xilinx has used a thorough test methodology to verify and characterize memory interface designs. This characterization process is based on a multitude of real system testing procedures—including the addition of system noise and strenuous PRBS data patterns—to ensure the functionality of process, voltage, and temperature (PVT) corner cases in a simulated system environment that can be made even more stressful than actual user systems in the field.

Several categories of tests are used in this characterization process:

- Voltage and temperature Shmoo plots
- Read/Write channel
- Calibration stability
- Long-term stability

These tests include:

- f_{MAX} testing from f_{MIN} to failure in 20MHz steps across PVT
- Similar large-sample testing of 250+ parts to check f_{MAX} and calibration adaptability
- Eye width measurements across PVT at the specified f_{MAX}
- Fabric noise generators used to emulate difficult system conditions
 - *Noise*: 80% additional utilization of FPGA flip-flops/block RAM/DSP, toggling at 33%
- Targeted system-level tests of clock jitter, SI/crosstalk effects, read eye size/shape, delay line monotonicity/jitter/steps, and calibration
- Functional pseudo-random testing through AXI interfaces to ensure functional correctness
- Targeted Read/Write channel margin tests

Additionally, Xilinx performs JEDEC compliance tests across PVT in partnership with industry leaders like Agilent to verify and demonstrate these capabilities.

For actual JEDEC compliance tests and demonstrations of the DDR4 interface and controller featuring a mid-speed-grade Kintex UltraScale device running at (and above) 2400Mb/s, visit www.xilinx.com/memory.

Conclusion

Achieving higher performance and lower power for memory interfaces is a design process that starts with the architecture definition. To enable both higher efficiency and higher sustainable data rates, new silicon features have been innovated to enable higher maximum data rates, efficient I/O utilization, and lower power requirements. With the UltraScale architecture portfolio, Xilinx has developed the highest performance memory interface solution in the industry by maximizing the DDR4 data rate capability to 2666Mb/s for mid-speed-grade devices, as well as improving the efficiency of the controller to sustain these high data rates in more demanding applications.

Xilinx is also continuing to provide easy-to-use software tools, like MIG, that enable faster customization of the core IP. Xilinx memory interface solutions are based on extensive hardware characterization to ensure that high performance is sustainable in typical systems under changing voltage and temperature conditions. Collectively, the solution not only maximizes memory interface capabilities and overall system performance but ensures rapid system bring-up to accelerate the user's design cycle.

Revision History

The following table shows the revision history for this document:

Date	Version	Description of Revisions
03/23/2015	1.1	Updated title and abstract; Figure 1 ; Memory Trends and FPGA Solutions ; Figure 2 ; UltraScale Architecture Benefits ; Table 1 ; and Conclusion .
06/30/2014	1.0	Initial Xilinx release.

Disclaimer

The information disclosed to you hereunder (the "Materials") is provided solely for the selection and use of Xilinx products. To the maximum extent permitted by applicable law: (1) Materials are made available "AS IS" and with all faults, Xilinx hereby DISCLAIMS ALL WARRANTIES AND CONDITIONS, EXPRESS, IMPLIED, OR STATUTORY, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, OR FITNESS FOR ANY PARTICULAR PURPOSE; and (2) Xilinx shall not be liable (whether in contract or tort, including negligence, or under any other theory of liability) for any loss or damage of any kind or nature related to, arising under, or in connection with, the Materials (including your use of the Materials), including for any direct, indirect, special, incidental, or consequential loss or damage (including loss of data, profits, goodwill, or any type of loss or damage suffered as a result of any action brought by a third party) even if such damage or loss was reasonably foreseeable or Xilinx had been advised of the possibility of the same. Xilinx assumes no obligation to correct any errors contained in the Materials or to notify you of updates to the Materials or to product specifications. You may not reproduce, modify, distribute, or publicly display the Materials without prior written consent. Certain products are subject to the terms and conditions of Xilinx's limited warranty, please refer to Xilinx's Terms of Sale which can be viewed at <http://www.xilinx.com/legal.htm#tos>; IP cores may be subject to warranty and support terms contained in a license issued to you by Xilinx. Xilinx products are not designed or intended to be fail-safe or for use in any application requiring fail-safe performance; you assume sole risk and liability for use of Xilinx products in such critical applications, please refer to Xilinx's Terms of Sale which can be viewed at <http://www.xilinx.com/legal.htm#tos>.

Automotive Applications Disclaimer

XILINX PRODUCTS ARE NOT DESIGNED OR INTENDED TO BE FAIL-SAFE, OR FOR USE IN ANY APPLICATION REQUIRING FAIL-SAFE PERFORMANCE, SUCH AS APPLICATIONS RELATED TO: (I) THE DEPLOYMENT OF AIRBAGS, (II) CONTROL OF A VEHICLE, UNLESS THERE IS A FAIL-SAFE OR REDUNDANCY FEATURE (WHICH DOES NOT INCLUDE USE OF SOFTWARE IN THE XILINX DEVICE TO IMPLEMENT THE REDUNDANCY) AND A WARNING SIGNAL UPON FAILURE TO THE OPERATOR, OR (III) USES THAT COULD LEAD TO DEATH OR PERSONAL INJURY. CUSTOMER ASSUMES THE SOLE RISK AND LIABILITY OF ANY USE OF XILINX PRODUCTS IN SUCH APPLICATIONS.