WP435 (v1.1) May 13, 2014

Xilinx UltraScale: The Next-Generation Architecture for Your Next-Generation Architecture

By: Steve Leibson and Nick Mehta

The Xilinx® UltraScale™ architecture delivers unprecedented levels of integration and capability with ASIC-class system- level performance for the most demanding applications.

The UltraScale architecture is the industry's first application of leading-edge ASIC architectural enhancements in an All Programmable architecture that scales from 20 nm planar through 16 nm FinFET technologies and beyond, in addition to scaling from monolithic through 3D ICs. Through analytical co-optimization with the Xilinx Vivado® Design Suite, the UltraScale architecture provides massive routing capacity while intelligently resolving typical bottlenecks in ways never before possible. This design synergy achieves greater than 90% utilization with no performance degradation.

Some of the UltraScale architecture breakthroughs include:

- Strategic placement (virtually anywhere on the die) of ASIC-like system clocks, reducing clock skew by up to 50%
- Latency-producing pipelining is virtually unnecessary in systems with massively parallel bus architecture, increasing system speed and capability
- Potential timing-closure problems and interconnect bottlenecks are eliminated, even in systems requiring 90% or more resource utilization
- 3D IC integration makes it possible to build larger devices one process generation ahead of the current industry standard
- Greatly increased system performance, including multi-gigabit serial transceivers, I/O, and memory bandwidth is available within even smaller system power budgets
- Greatly enhanced DSP and packet handling

The Xilinx UltraScale architecture opens up whole new dimensions for designers of ultra-high-capacity solutions.

More is Better

Since the introduction of all things digital, the fundamental and undeniable trend for digital systems across all markets is that "more is better." This expectation has become the basic driver behind systems requiring higher resolution, more bandwidth, and more storage. The "more" mind-set also logically leads to several truisms:

- More devices are generating more data.
- More data means that data must flow faster.
- More fast-flowing data demands more computations per second.
- More applications need quicker access to more data.
- More data integrity is required as the amount of data grows and data rates increase.

This rapid growth in data creation and in data-transmission rates is occurring across almost every market and amplifies the need for new device architectures to address the challenges associated with:

- Massive data flow and routing with ASIC-like clocking
- Massive I/O and memory bandwidth
- Faster DSP and packet processing
- Power management
- Multi-level security

UltraScale Architecture: The Next-Generation All Programmable Architecture from Xilinx

To address system performance in the range of multi-hundreds of gigabits per second with smart processing at full line rate, scaling to terabits and teraflops, a new architectural approach is required. The mandate is not simply to increase the performance of each transistor or system block, or to scale the number of blocks in the system. Rather, the idea is to fundamentally improve the communication, clocking, critical paths, and interconnect to address the massive data flow and real-time packet and image processing.

The UltraScale architecture delivers unprecedented levels of integration and capability with ASIC-class, system-level performance for the most demanding applications, which require massive I/O and memory bandwidth, massive data flow, and superior DSP and packet-processing performance. Tuned to provide massive routing capacity and analytically co-optimized with the Vivado design tools, the UltraScale architecture delivers unprecedented levels of utilization —greater than 90%— without degradation in performance.

The UltraScale architecture is the industry's first application of leading-edge ASIC architectural enhancements in an All Programmable architecture that scales from 20 nm planar through 16 nm FinFET technologies and beyond, in addition to scaling from monolithic through 3D ICs. The UltraScale architecture not only addresses the limitations to scalability of total system throughput and latency, but directly addresses interconnect—the No. 1 bottleneck limiting system performance at advanced nodes.

The Xilinx UltraScale architecture is designed to address next-generation system-level performance requirements associated with next-generation systems (see Figure 1).

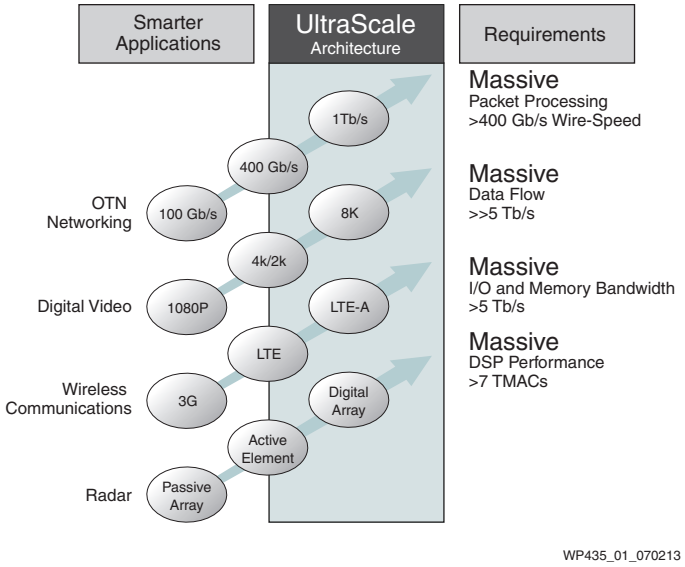

Figure 1: **Next-Generation High-Performance Target Applications Examples**

Hundreds of design enhancements went into the UltraScale architecture. These enhancements synergistically combine to enable design teams to create systems that have greater functionality, run faster, and deliver greater performance per Watt than ever before. See Figure 2.

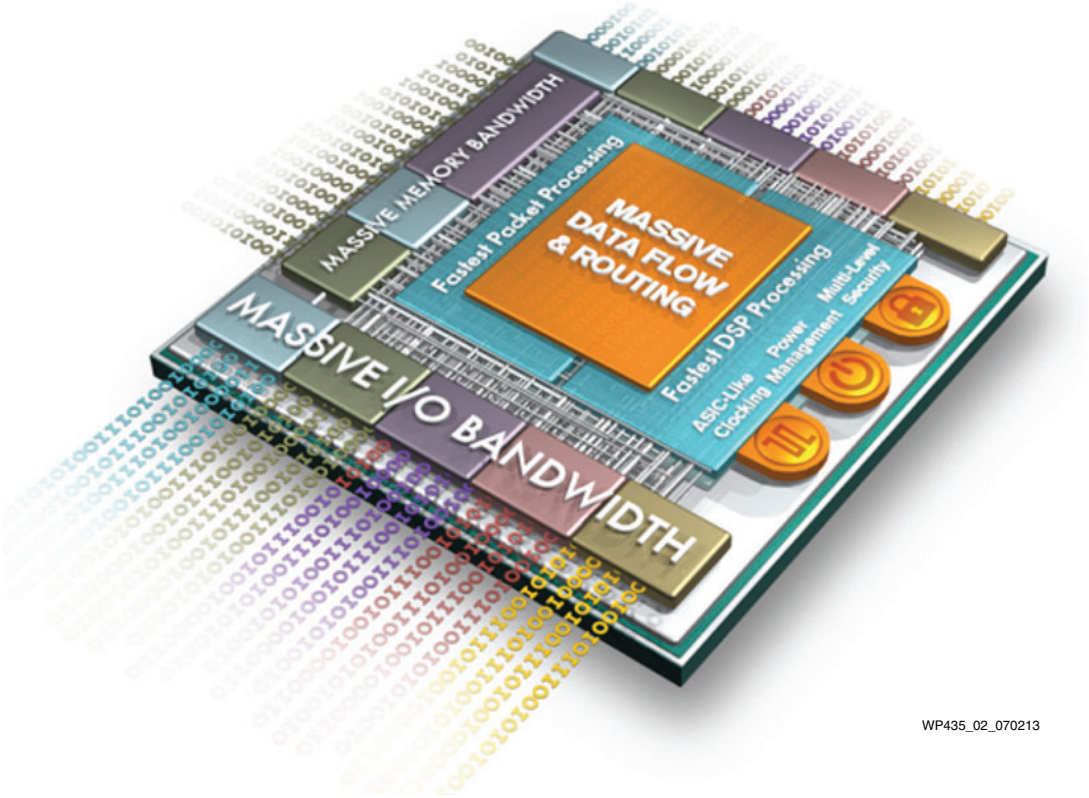

Figure 2: **The Xilinx UltraScale Architecture**

The UltraScale architecture in conjunction with the Vivado Design Suite delivers these next-generation system-level capabilities to system design teams:

- Massive data flow optimized for wide buses supporting multi-terabit throughput with lowest latency
- Highly optimized critical paths and built-in, high-speed memory, cascading to remove bottlenecks in DSP and packet processing
- Enhanced DSP slices, incorporating 27x18-bit multipliers and dual adders that enable a massive jump in fixed-point and IEEE Std 754 floating-point arithmetic performance and efficiency
- Step function in inter-die bandwidth for 2nd-generation 3D IC systems integration and new 3D IC wide-memory optimized interface
- Massive I/O and memory bandwidth, including support for next-generation memory interfacing with dramatic reduction in latency, optimized with multiple hardened, ASIC-class 100G Ethernet, Interlaken, and PCIe® IP cores
- Multi-region ASIC-like clocking, delivering low-power clock networks with extremely low clock skew and high-performance scalability
- Power management with significant static- and dynamic-power gating across a wide range of functional elements, yielding significant power savings
- Next-generation security with advanced approaches to AES bitstream decryption and authentication, key-obfuscation, and secure device programming
- Elimination of routing congestion through co-optimization with Vivado tools for >90% device utilization without degradation in performance or latency

System designers can pool these system-level capabilities in multiple combinations to solve a variety of problems, best illustrated by this generalized block diagram of a wide datapath design. See Figure 3.

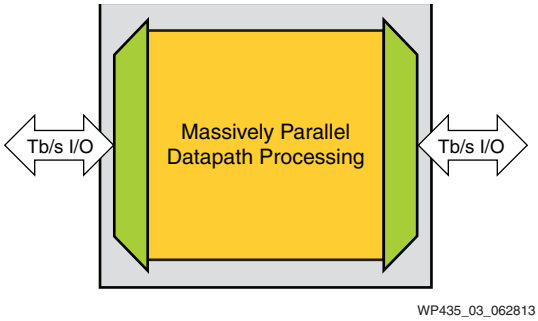

Figure 3: **Terabit I/O Requires Massively Parallel Datapaths**

Here, data streams with data rates on the order of terabits per second enter and exit from the left and the right. The system must convey these streams between the left and right I/O ports while performing the requisite processing. The I/O transmission is through high-speed serial transceivers, operating in the multi-Gb/s range. As soon as the multi-Gb/s serial streams enter the device, they must fan out to match the data flow, routing, and processing capabilities of the on-chip resources.

The Challenges of Designing Terabit Systems: Clock Skew and Massive Data Flow

For a real-world example, assume that the port bandwidth for the left and right I/O ports is 100 Gb/s. This means that the on-chip resources must also handle at least 100 Gb/s traffic. Designers typically employ a wide bus or datapath, ranging anywhere in size from 512 bits to 1,024 bits to handle the associated data throughput, yielding a system clock that matches the capabilities of the on-chip resources. For even higher line rates extending to 500 Gb/s, bus widths on the order of 1,024 and 2,048 bits are not uncommon.

Now consider the clocking requirements for these types of buses. Before the advent of the UltraScale architecture, operation at the high end of system clock frequencies could lead to worst-case clock skew across these massive datapaths, approaching up to 50% of the total system clock period. With almost half the clock period consumed by clock skew, designs would need to rely on heavy pipelining to even have a chance of achieving the target system performance. With only 50% of the clock period left for computation, the chances are low that the resulting solution would prove to be viable. Beyond consuming large amounts of register resources, extensive pipelining has a significant impact on overall system latency — which again proves to be unacceptable in today's high-performance systems.

UltraScale Architecture's ASIC-Like Clocking

Thanks to the UltraScale architecture’s multi-region ASIC-like clocking, designers can now place system-level clocks at the most optimal location—virtually anywhere on the die—leading to a reduction in system-level clock skew by as much 50%. Placing the clock-driving node in the geometric center of a functional block and balancing the skew across leaf clock cells addresses one of the critical bottlenecks that stands in the way of multi-terabit system-level performance. Reducing the overall system clock skew also eliminates the need for extensive pipelining and the associated latency that comes with it. The UltraScale architecture’s ASIC-like clocking not only removes any restrictions around clock placement, it also allows for a large number of independent, high-performance, low-skew clock sources to be realized in the system design. This is a radical and powerful departure from clocking schemes found in previous generations of programmable logic devices. From the system designer's perspective, this solution to clock skew simply eliminates the problem.

Taming the Challenges of Massive Data Flow

Traditionally, very high-performance applications employ wide buses or wide datapaths as a means to match the data flow routing to the processing capabilities of the on-chip resources. However, scaling performance through wide buses comes with its own set of challenges beyond the need to simply tame clock skew. Competitive architectures have notoriously proven to be severely starved in the quantity and flexibility of routing resources suitable for high-performance designs. Addressing applications with 100 Gb/s throughput using FPGAs with lower-performance interconnect architectures can lead to the need for data buses on the order of 1,536 to 2,048 bits wide.

While the wider bus implementation might lead to the need for a lower system clock frequency, significant timing-closure challenges now arise due to a lack of routing resources required to support systems with wide buses. The situation is further aggravated by the fact that some FPGA vendors use antiquated place-and-route algorithms based on simulated annealing, which are blind to global design metrics, such as the level of congestion or the total wire length. Thus, designers are forced to consider trade-offs that require lowering the performance of the system (typically not an option), extensive pipelining at the expense of latency, or gross underutilization of the available device resources. In all cases, these solutions prove to be inferior or inadequate. More importantly, the fundamental limitation in routing resources required to address applications on the order of 100 Gb/s found in traditional FPGAs all but guarantees that addressing next-generation multi-terabit applications will be out of the realm of possibility, or will come at the expense of very poor device utilization or latency.

Further complicating matters, scaling performance through massively wide data buses holds the added expense of significant growth in overhead logic circuitry to support the implementation of these wide buses, compounding the challenges of achieving timing closure.

An example based on Ethernet packet sizes best illustrates the situation. Ethernet has a minimum packet size of 64 bytes (512 bits). Assuming a 2,048-bit-wide bus is used to implement a 400G system, up to four packets can fit within this bus.

Large amounts of logic are required to handle the various scenarios and combinations of packets — four complete packets, or one, two, or three complete or partial packets — can exist across the 2,048-bit-wide bus. It takes large amounts of complex replicated logic to address these possible combinations. Additionally, it can be necessary to speed up (or extend the performance of) some sections of logic to handle the case when the bus requires four packets to be simultaneously processed and written to memory. Design considerations ranging from speeding up the logic to using four independent duplicate memory controllers to process multiple packets in tandem further stress routing resources, driving the need for architectures with even greater amounts of high-performance, low-skew routing resources. See Figure 4.

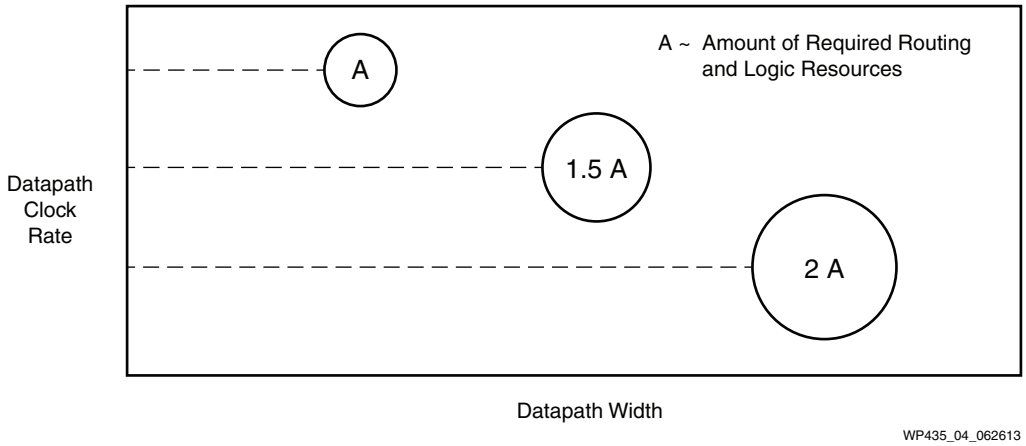

Figure 4: **Increased Datapath Clock Width and Clock Rates Require More Logic and Routing Resources**

Semiconductor Process Scaling Impacts Interconnect Technologies

As the industry pushes to 20 nm semiconductor process technologies and beyond, a new challenge arises in the RC delay associated with the copper interconnect, which inhibits the level of performance scaling that can be achieved by migrating to the next node. This increase in transistor interconnect delay has a direct effect on the overall system performance that can be achieved, and reinforces the need for routing architectures that can deliver the level of performance required for next-generation applications. The UltraScale routing architecture was developed with a deep understanding of next-generation process technologies and was expressly designed to mitigate the effects of copper interconnect, which if not properly addressed can also become a system performance bottleneck.

UltraScale Interconnect Architecture: Optimized for Massive Data Flow

The UltraScale next-generation interconnect architecture represents a true breakthrough in programmable-logic routing. Xilinx placed a critical focus on addressing next-generation applications, which must support massive data flow, ranging from multi-gigabit smart packet-processing applications through multi-terabit datapath applications. Historically, routing or interconnect congestion has been a significant limiting factor in achieving timing closure and quality of results when implementing wide logic blocks—extending to bus widths of 512 bits, 1,024 bits, and beyond. Highly congested logic designs often could not be routed in earlier device architectures; if the tools do manage to route a congested design, the resulting design frequently runs at a lower-than-desired clock rate. The UltraScale routing architecture essentially removes routing congestion completely. The result is simple: if the design fits, it routes.

Consider this analogy: Think of a busy intersection in the center of a city. Traffic is moving from north to south, south to north, east to west, and west to east. Some vehicles are attempting to make turns. All of this traffic tries to move simultaneously. Usually, you get a large traffic snarl. Now consider the same sort of intersection on a modern, well-designed high-speed freeway or Autobahn. Road designers create dedicated ramps—fast tracks—to take traffic smoothly from one part of a major freeway intersection to another. Traffic moves from one part of the freeway to another at full speed. There are no snarls.

Xilinx has added the same sort of fast tracks to the UltraScale architecture. These additional fast tracks carry data between nearby logic elements that are not necessarily adjacent but are still logically connected by a particular design. The result is an exponential increase in the amount of data the UltraScale architecture can manage, as shown in Figure 5.

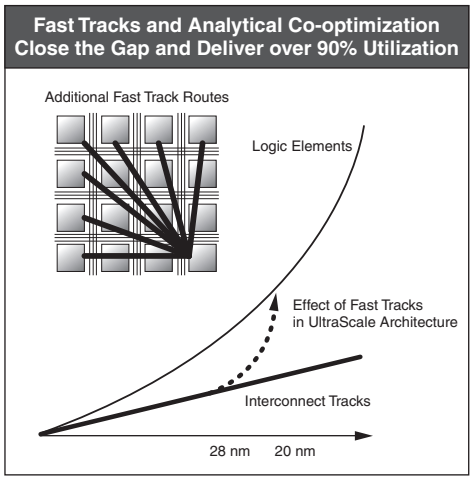

Figure 5: **Increases in Real and Effective Routing Tracks Help Keep Pace with Growing Complexity**

UltraScale Architecture Stacked Silicon Interconnect Technology Enhances Everything

Few technological developments have had the tremendous impact on device capacity and performance as the integration of Stacked Silicon Interconnect (SSI) technology, proven by the first generation of Xilinx's 3D ICs based on the 7 series All Programmable devices. The SSI technology integration makes it possible to build larger devices one process generation ahead of the industry benchmark. This continues to be the case for the Xilinx second-generation UltraScale architecture-based 3D ICs.

Because the silicon die in a 3D IC can communicate with one another through connections that are denser and faster than is possible when die are individually packaged, it takes less power for this inter-die communication (assuming the die do not need to drive the added impedance of die-to-package and board-level interconnections). So, the net result of the SSI technology integration is to greatly expand capacity and performance while reducing power consumption relative to individually packaged die. In addition, system security is enhanced because the die-to-die communications are no longer easily accessible at the board level.

The Virtex® UltraScale and Kintex® UltraScale family members contain a step-function increase in both the number of connectivity resources and the associated inter-die bandwidth in this second-generation 3D IC architecture. The big increase in routing resources and inter-die bandwidth ensures that next-generation applications can achieve their target performance and achieve timing closure at extreme levels of utilization.

THE CHALLENGES OF SMART, FAST PROCESSING

Whether the goal is increased packet throughput, more DSP GMACs, or more megapixels/second displayed on a screen, the technical challenges are the same for any high-performance system, as illustrated in Figure 6.

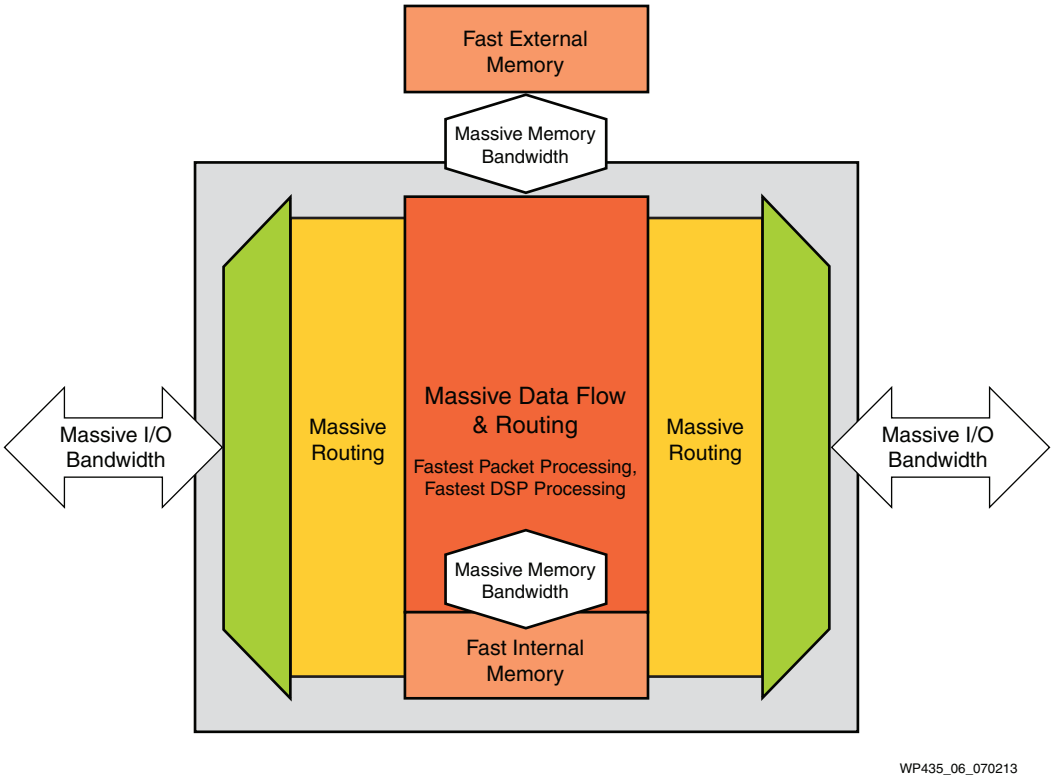

Figure 6: **High-Performance Systems Require Massive Bandwidth**

No matter the application, the problem is fairly simple: A large amount of data enters the system over multiple high-speed serial ports, ranging from tens to hundreds of gigabits per port. This high-speed data needs to be routed to the processing logic and processed in real time, a task that demands speed (typically DSP or packet processing) to handle the high data rates. Incoming data and intermediate results must be stored either within the system, close to the processing elements, or in fast bulk memory located adjacent to the system. After the data has been processed, it must be routed to the high-speed output transceivers to be passed along. Figure 6 illustrates:

- System data input and output over these high-speed serial lines require massive I/O bandwidth through rock-solid multi-gigabit serial transceivers. These serial transceivers must be reliable and have a very low bit-error rate.
- The massively parallel routing fans out from the multi-gigabit serial transceivers to the massively wide function-processing block, which requires wide fanout capability with low clock skew. Routing these massively parallel buses is a challenge.
- Massive data flow processing requires high-throughput logic and DSP blocks coupled with very fast internal and external memory access using memory interfaces with massive memory bandwidth. This type of processing severely stresses the data- and clock-routing capabilities of any architecture.

All of these performance goals must be met within certain power limits. Systems must operate within available power and cooling limits, as shown in this conceptual graph in Figure 7.

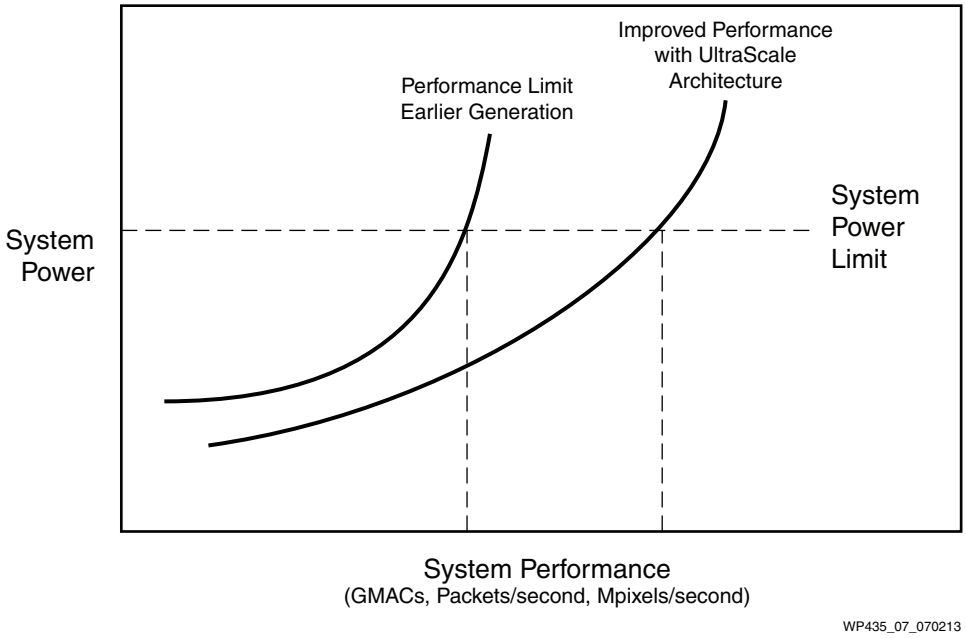

Figure 7: **UltraScale Architecture Transcends Earlier Power and Performance Limits**

The components that make up the UltraScale architecture are tuned to all of the many complex requirements of next-generation processing systems.

Delivers Massive I/O Bandwidth

The UltraScale architecture simultaneously offers a significant increase in performance and a reduction in power consumption with respect to high-speed serial transceivers. Virtex UltraScale devices provide next-generation serial transceivers that are capable of supporting serial system bandwidth in excess of 5 Tb/s.

The UltraScale architecture-based GTY and GTH serial transceivers incorporate internal gearbox logic that translates the multi-Gb/s serial data line rates into the wider data buses — operating at multiple hundreds of megahertz — that are required to match the on-chip logic and memory speeds. The transceiver gearboxes eliminate the cost of external gearbox chips in system designs. Similarly, an integrated fractional phase-locked loop (PLL) for the UltraScale architecture-based GTY serial transceivers converts one reference clock into multiple line rates, eliminating the need for external voltage-controlled crystal oscillators (VCXOs). This feature alone can save literally dozens of discrete devices and hundreds of dollars in system designs that employ many high-speed serial ports running at dissimilar line rates.

The UltraScale architecture-based ASIC-class serial transceivers have greater flexibility than the transceivers in earlier device generations while retaining the bulletproof auto-adaptive equalization features (automatic gain control, continuous-time linear equalization, decision feedback equalization (DFE)) of the Xilinx 7 series All Programmable devices. Xilinx's auto-adaptive equalization maintains bit-error rates at undetectable levels (e.g., $<10^{-17}$) and permits the UltraScale architecture-based transceivers to directly drive high-speed, multi-gigahertz backplanes.

Delivers Massive External and Internal Memory Bandwidth

The UltraScale architecture takes memory interfacing to a new level by enabling multiple DDR3/4-capable SDRAM memory controllers and including hardened DDR physical-layer (PHY) blocks on chip. Compared to earlier device generations, UltraScale architecture-based devices have:

- More SDRAM controllers
- Wider SDRAM ports
- Faster memory ports

As a result, UltraScale architecture-based All Programmable devices can provide more than 1 Tb/s of DDR SDRAM memory bandwidth to handle the massive data flow, fast processing, and enormous memory requirements of leading-edge, next-generation system designs. The hardened SDRAM PHY blocks reduce read latency by 30% compared with soft-core PHYs, while the ability to control DDR4 SDRAMs reduces the power needed for external memory by more than 20%.

On-chip block RAM performance can often be a critical factor that affects a system's maximum clock rate. Xilinx has rearchitected the block RAM in the UltraScale architecture-based All Programmable device's to match the performance of the other programmable blocks in the system while decreasing power consumption. The new block RAM architecture supports high-speed memory cascading, removing bottlenecks in DSP and packet processing. This new architectural feature employs output multiplexers in a novel way so that the Vivado Design Suite tools can also efficiently create large, fast RAM arrays and FIFOs without using additional on-chip routing or logic resources.

Xilinx has also enhanced UltraScale architecture-based block RAM FIFO configurations to permit input and output ports of different widths on the same FIFO. This feature is helpful when the FIFO is used to cross from one system clock domain to another — and the UltraScale architecture now supports many more clock domains.

Enabling Fast, Smart Processing

The performance requirements for DSP and packet-processing systems grow over time due to end-customer demand, as illustrated by this graph in Figure 8:

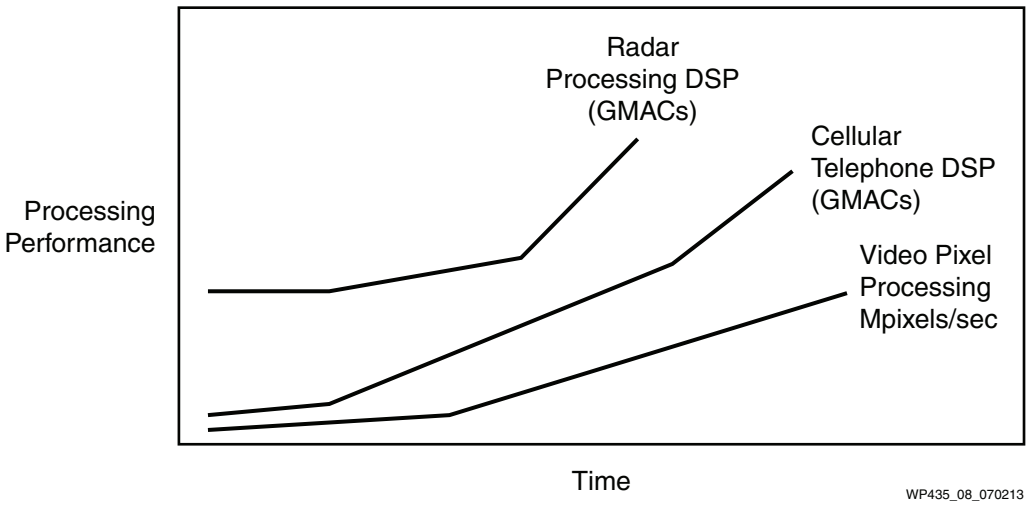

Figure 8: **System Performance Requirements Grow over Time**

The need to extract greater signal information from the noise, the desire to create ever-more life-like images, or the quest to handle the insatiable growth in packet traffic all drive these growing performance requirements. Yet there is always the need to handle these performance demands within an economic envelope that sets the practical limits of what can be accomplished, as shown in Figure 9, depicting the performance and cost trends over time for LTE and LTE Advanced (LTE-A) base stations:

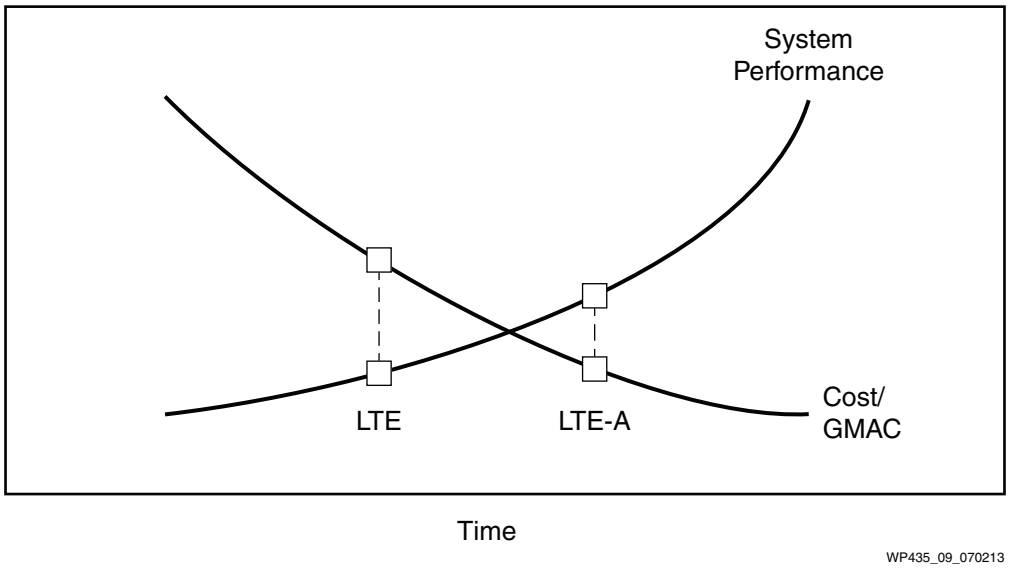

Figure 9: **LTE Performance and Cost Trends**

In short, what customers want is more system performance at less cost — a perpetual trend in most electronics industries. This is exactly what the UltraScale architecture delivers.

Enhanced DSP Improves on a Performance Leader

Xilinx has significantly enhanced the Virtex-7 FPGA's DSP48E1 DSP slice, already the industry's performance leader, for the UltraScale architecture, to permit faster digital signal processing while consuming fewer routing or logic resources outside of the DSP block. A number of innovations applied to the DSP slice improve multiplication and MACC operations, enhancing functional performance and reducing power consumption.

The UltraScale architecture-based DSP48E2 DSP slice incorporates 27x18-bit multipliers that permit the mapping of larger functions into fewer DSP slices. For example, the DSP48E2 block, with its wider 27x18-bit multipliers, can implement IEEE Std 754 double-precision arithmetic using two-thirds fewer DSP blocks compared with the same function implemented with the DSP48E1 blocks in Xilinx 7 series All Programmable devices.

Inclusion of wide-MUX and wide-XOR functions in the DSP48E2 slice allows non-DSP computations, such as error correction and control (ECC), cyclic redundancy checking (CRC), and forward error correction (FEC), to use the DSP slices as wide, high-speed, hardened logic blocks. These enhancements increase performance, lower power consumption, and reduce the usage of configurable logic blocks (CLBs), leaving more CLBs free for the implementation of other functions. It is through these innovations to the DSP block and others that the UltraScale architecture is able

to simultaneously meet the next-generation application demands of increased processing requirements and lower cost points.

Extended Performance for Smart Packet Processing

The insatiable need for bandwidth continues to drive investments in the upgrading of the network communications infrastructure. Massive data traffic resulting from the transport of digital video has led to the accelerated maturity of 100 Gb/s based networking equipment and the increased demand for 400G solutions—and beyond. Packet processing can present significant performance challenges to even the most advanced architectures at the current industry rates of hundreds of gigabits per second. Fundamental packet-processing functions, including checksum calculation and bridging functions performed at line speed, can have a significant impact on performance and resource utilization.

Beyond addressing the massive data flow associated with high-performance packet processing, the UltraScale architecture contains many innovations tailored expressly for packet processing. Among them are modifications to the DSP48 block to support CRC 32 checksum calculations at wire speed, and the inclusion of hardened Gigabit Ethernet MACs and Interlaken chip-to-chip interfaces to support breakthrough performance and new levels of integration for smart packet processing.

UltraScale Architecture Addresses System-Level Power Requirements for Next-Generation Systems

While system-level performance continues to scale with each successive product generation, the expectation (and system requirement) is that power consumption will either remain constant or continue to decrease. For example, for wired communications infrastructure equipment, next-generation line cards supporting higher bandwidth or compute performance must achieve these gains without any change in the form factor or power envelope of the line card. And while it could be said that this trend is at odds with the very nature of increased system performance (traditionally at the expense of increased power consumption), improvements in system-level power savings continue to be derived through integration, power-management strategies, and advanced process technologies.

The UltraScale architecture builds upon a history of delivering unrivaled system-level power reduction with each successive generation of All Programmable logic families. Low-power semiconductor processing coupled with significant static- and dynamic-power gating enabled through silicon and software techniques results in up to 50% overall system power savings over the Xilinx 7 series devices—already the lowest-power programmable logic device leader.

Power savings translates to one of two things for the designer: lower power-budget and thermal-management requirements, or increased speed—very important levers to address the increasing requirements of next-generation applications.

UltraScale Architecture Security Protects IP, Prevents Tampering

The use of Xilinx All Programmable FPGAs continues to expand across almost every market to the point where these devices are now becoming the core of many new systems.

The increasing ubiquity of Xilinx All Programmable devices makes protecting the IP contained within them as important as protecting the data that the devices process. As awareness of security threats has grown, the security community has responded with a set of policies and standards that are often the driving force behind design security. The range of security threats or potential weaknesses that designers must consider to deploy products that are deemed secure can be quite extensive. An abbreviated list of vulnerabilities includes complacency, incomplete security measures, back doors, design defects, device defects, single-event upsets, bitstream decoding, spoofing, Trojan horses, readback, side channels, and fault insertion.

Based on a heritage of more than five generations of security solutions and innovations, the UltraScale architecture extends Xilinx leadership in delivering secure solutions through the incorporation of multiple enhanced security features that further protect IP loaded into the device and prevent tampering. The UltraScale architecture's security improvements include stronger, more advanced approaches to AES bitstream decryption and authentication with further extended key-obfuscation features. The result is a robust, industry-leading solution that addresses the continuously changing landscape of next-generation security requirements.

Analytical Co-optimization = Predictable Success

Delivering the unprecedented levels of integration, capability, and ASIC-class system-level performance for the most demanding applications on its own would be challenging enough. The UltraScale architecture scales from 20 nm planar through 16 nm FinFET technologies and beyond, in addition to scaling from monolithic through 3D ICs while delivering unprecedented levels of utilization — upwards of 90% — without degradation in performance. The only way to meet these extreme objectives was by the co-optimization of the UltraScale architecture in conjunction with the Vivado Design Suite.

First introduced for the Xilinx 7 series device families, the Vivado Design Suite is an SoC-strength design environment built from the ground up for the next decade of All Programmable devices, including the UltraScale architecture. The Vivado Design Suite attacks the key design bottlenecks in programmable systems integration and implementation to enable up to a fourfold productivity advantage over competing development environments.

Achieving the extreme performance, integration, and quality-of-results objectives for next-generation design requires entirely new approaches to device placement and routing. Traditional FPGA place-and-route tools have relied upon simulated annealing as the primary placement optimization algorithm. This algorithm starts with a random placement and then works to optimize a given cost function. Because of the random nature of the initial placement and the subsequent moves, traditional simulated-annealing algorithms do not scale to million-LUT designs for multiple reasons. First, the time associated with the iterative rip-up, reroute, and evaluation of alternative placements becomes unwieldy, even when the design has conservative performance or device-utilization constraints. More importantly, because simulated-annealing algorithms are based upon local moves, they are blind to global design metrics, such as the level of congestion or the total wire length. For designs requiring multi-terabit performance employing wide buses with

requisite "zero" clock skew, employing a place-and-route algorithm that is ignorant of total wire length and congestion level becomes untenable.

An optimal placement solution depends upon multiple dimensions, such as timing, wire length, and congestion metrics.

The Vivado Design Suite uses a multi-variable cost function to find the optimal placement, allowing the designer to quickly find a routeable solution even at device utilizations of greater than 90%, without degradation in performance. Run times are consistently faster than with alternative solutions, while the variance in results is much tighter, enabling design closure with fewer iterations and at performance levels and device utilization that are unprecedented in the industry.

UltraScale Architecture and Process Technology

Process technology is an important consideration in any device architecture, and the UltraScale architecture is designed to span multiple process technologies. Xilinx and TSMC collaborated on the 28 nm HPL (low-power, high-performance) process technology that has been a major factor in the tremendous success of the Xilinx 7 series All Programmable devices. Based on the companies' experience in that partnership, Xilinx and TSMC developed the 20 nm 20SoC planar process to enable the first generation of UltraScale architecture-based All Programmable devices, which yielded first silicon in 2013.

However, Xilinx designed the UltraScale architecture to also be able to exploit the additional performance, capabilities, and power savings of the process node that follows 20SoC — namely, 16 nm FinFET. Here again, the Xilinx UltraScale architecture and Vivado tool suite were co-optimized with the TSMC 16 nm FinFET process technology through the unique Xilinx "FinFast" development program, which drew on significant engineering talent at both Xilinx and TSMC. As a result, Xilinx and TSMC will be producing first silicon for second-generation UltraScale architecture-based All Programmable devices in 2014.

Conclusion

To address levels of system performance in the multiple hundreds of gigabits per second with smart processing at full line rate, scaling to terabits and teraflops, a new architectural approach is required. Xilinx has developed the next-generation UltraScale architecture and Vivado Design Suite with the needs of next-generation, high-performance systems in mind. The UltraScale architecture delivers ASIC-class, system-level performance for the most demanding next-generation applications — those that require massive I/O and memory bandwidth, massive data flow, and massive DSP and packet-processing performance at unprecedented levels of utilization, over 90%, without degradation in performance.

The UltraScale architecture is the industry's first application of leading-edge ASIC architectural enhancements in an All Programmable architecture that scales from 20 nm planar through 16 nm FinFET technologies and beyond, in addition to scaling from monolithic through 3D ICs. Through a combination of TSMC's leading-edge technology and co-optimization with the next-generation Vivado Design Suite, Xilinx is a year ahead in delivering 1.5x to 2x realizable system-level performance and integration. This achievement is the equivalent of being a generation ahead of our competition.

To find out how the UltraScale architecture can enable you to achieve the objectives of your next-generation design or to learn more about the UltraScale architecture-based All Programmable FPGA families, contact your local Xilinx sales office or go to www.Xilinx.com.

Revision History

The following table shows the revision history for this document:

Date	Version	Description of Revisions
05/13/2014	1.1	Updated The Challenges of Designing Terabit Systems: Clock Skew and Massive Data Flow and Extended Performance for Smart Packet Processing.
07/08/2013	1.0	Initial Xilinx release.

Disclaimer

Automotive Applications Disclaimer