WP380 (v1.0) October 27, 2010

Xilinx Stacked Silicon Interconnect Technology Delivers Breakthrough FPGA Capacity, Bandwidth, and Power Efficiency

By: Patrick Dorsey

The programmable imperative—the critical need to achieve more with less, to reduce risks wherever possible, and to quickly create differentiated products using programmable hardware design platforms—is driving the search for FPGA-based solutions that provide the capacity, lower power, and higher bandwidth with which users can create the system-level functionality currently delivered by ASICs and ASSPs.

Xilinx has developed an innovative approach for designing and manufacturing FPGAs that address two key requirements of the programmable imperative. Stacked silicon interconnect technology is the foundation of a new generation of FPGAs that breaks through the limitations of Moore's law and delivers the capabilities to satisfy the most demanding design requirements. It also enables Xilinx to reduce the time required to deliver the largest FPGAs in the quantities needed to satisfy end-customer volume production requirements. This white paper explores the technical and economic challenges that led Xilinx to develop stacked silicon interconnect technology and innovations that make it possible.

Introduction

As the role of the FPGA becomes more dominant in system design, the designs grow larger and more complex, demanding higher logic capacity and more on-chip resources. To date, FPGAs have depended predominantly on Moore's Law to respond to this need, delivering nearly twice the logic capacity with each new process generation. However, keeping pace with today's high-end market demands requires more than Moore's Law increases can provide.

The most aggressive adopters of FPGA technology are eager to employ the highest capacity, highest bandwidth devices of each new FPGA generation. However, the challenges of building large FPGAs early in the product life cycle can limit the ability to supply the volumes of devices these customers require for their production runs. That is because the circuitry overhead that enables reprogrammable technology negatively affects the manufacturability (and therefore, the supply) of the largest FPGAs. At the early stages of a new process node, when defect densities are high, die yield declines dramatically as die size increases. As the fabrication process matures, defect density falls and the manufacturability of large die increases significantly.

Thus, while the largest FPGAs are in short supply at product introduction, over time they eventually become available in quantities that support end-customer volume requirements. In response to the programmable imperative, a few leading-edge customers have challenged Xilinx to find a way to support their volume production requirements with the largest FPGAs as soon as possible after product introduction.

For example, the telecommunications market needs FPGAs that incorporate dozens of serial transceivers with increased interconnect logic and block RAM for advanced data processing and traffic management, while enabling use within current form factors and power footprints. To reap first-mover advantage, the equipment makers want to ramp up manufacturing of their new products as rapidly as possible.

Xilinx has responded to these requirements with an innovative approach for building FPGAs that offer bandwidth and capacity equaling or exceeding that of the largest possible FPGA die with the manufacturing and time-to-market advantages of smaller die to accelerate volume production. These benefits are enabled by stacked silicon interconnect technology, which uses silicon interposers with microbumps and through silicon vias (TSV) to combine multiple highly manufacturable FPGA die slices in a single package.

The Challenges of Interconnecting Multiple FPGAs

Stacked silicon interconnect technology solves the challenges that had previously obstructed attempts to combine the interconnect logic of two or more FPGAs to create a larger, "virtual FPGA" for implementing a complex design:

- The amount of available I/O is insufficient for connecting the complex networks of signals that must pass between FPGAs in a partitioned design and as well as connecting the FPGAs to the rest of the system.
- The latency of signals passing between FPGAs limits performance.
- Using standard device I/O to create logical connections between multiple FPGAs increases power consumption.

Key Challenge: Limited Connectivity and Bandwidth

System-on-chip (SoC) designs comprise millions of gates connected by complex networks of wires in the form of multiple buses, complicated clock distribution networks, and multitudes of control signals. Successfully partitioning an SoC design across multiple FPGAs requires an abundance of I/Os to implement the nets spanning the gap between FPGAs. With SoC designs including buses as wide as 1,024 bits, even when targeting the highest available pin count FPGA packages, engineers must use data buffering and other design optimizations that are less efficient for implementing the thousands of one-to-one connections needed for high-performance buses and other critical paths.

Packaging technology is one of the key factors to this I/O limitation. The most advanced packages currently offer approximately 1,200 I/O pins, far short of the total number of I/Os required.

At the die level, I/O technology presents another limitation because I/O resources do not scale at the same pace as interconnect logic resources with each new process node. When compared to transistors used to build the programmable logic resources in the heart of the FPGA, the transistors comprising device I/O structures must be much larger to deliver the currents and withstand the voltages required for chip-to-chip I/O standards. Thus, increasing the number of standard I/Os on a die is not a viable solution for providing the connections for combining multiple FPGA die.

Key Challenge: Excessive Latency

Increased latency is another challenge with the multiple FPGA approach. Standard device I/Os impose pin-to-pin delays that degrade the overall circuit performance for designs that span multiple FPGAs. Moreover, using time-domain multiplexing (TDM) on standard I/Os to increase the virtual pin count by running multiple signals on each I/O imposes even greater latencies that can slow I/O speeds down by a factor of 4X–32X or more. These reduced speeds are often acceptable for ASIC prototyping and emulation, but are often too slow for end-product application.

Key Challenge: Power Penalty

TDM approaches also result in higher power consumption. When used to drive hundreds of package-to-package connections across PCB traces between multiple FPGAs, standard device I/O pins carry a heavy power penalty compared to connecting logic nets on a monolithic die.

Similarly, multichip module (MCM) technology offers potential form-factor reduction benefits for integrating multiple FPGA die in a single package. The MCM approach, however, still suffers from the same restrictions of limited I/O count as well as undesirable latency and power consumption characteristics.

Xilinx Stacked Silicon Interconnect Technology

To overcome these limitations and roadblocks, Xilinx has developed a new approach for building production volumes of high-capacity FPGAs. The new solution enables high-bandwidth connectivity between multiple die by providing a much greater number of connections. It also imposes much lower latency and consumes dramatically lower power than the multiple FPGA approach, while enabling the integration of massive quantities of interconnect logic and on-chip resources within a single package.

Within the density range of an FPGA family, the medium-density devices represent the "sweet spot." That is, compared to the previous generation, they offer significantly greater capacity and bandwidth—on a die size that can be delivered earlier in the FPGA product life cycle than the largest devices in the same family. Thus, by combining several of these die in a single device, it is feasible to match or exceed the capacity and bandwidth offered by the largest monolithic devices, but with the manufacturing and time-to-volume advantages of smaller die.

Xilinx arrived at such a solution by applying several proven technologies in an innovative way. By combining through-silicon via (TSV) and microbump technology with its innovative ASMBL™ architecture, Xilinx is building a new class of FPGAs that delivers the capacity, performance, capabilities, and power characteristics required to address the programmable imperative. Figure 1 shows the top view of the die stack-up with four FPGA die slices, silicon interposer, and package substrate. Xilinx stacked silicon interconnect technology combines enhanced FPGA die slices and a passive silicon interposer to create a die stack that implements tens of thousands of die-to-die connections to provide ultra-high inter-die interconnect bandwidth with far lower power consumption and one fifth the latency of standard I/Os.

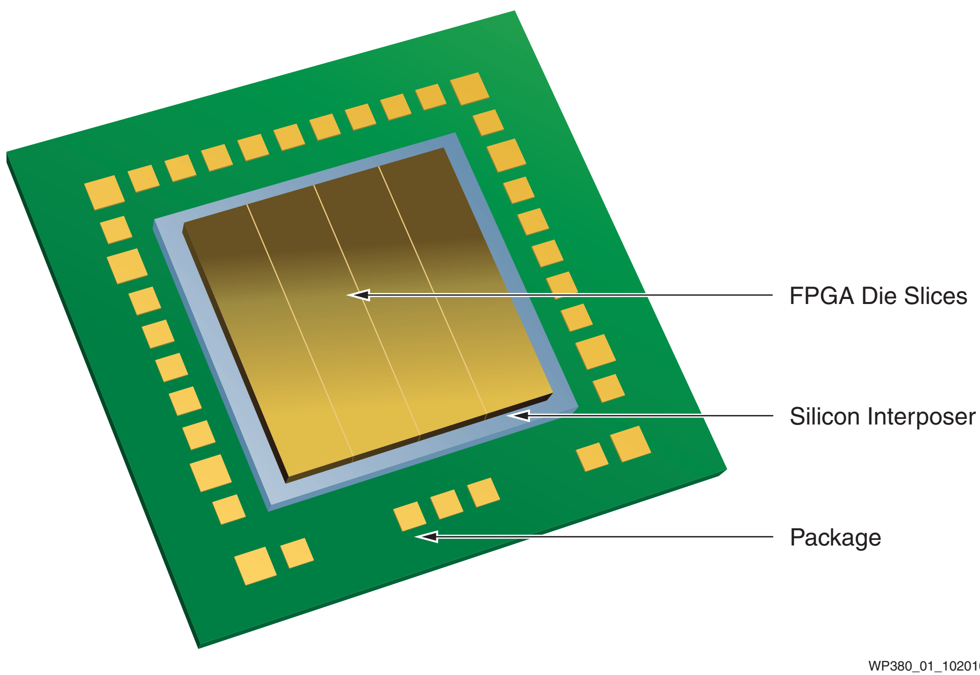
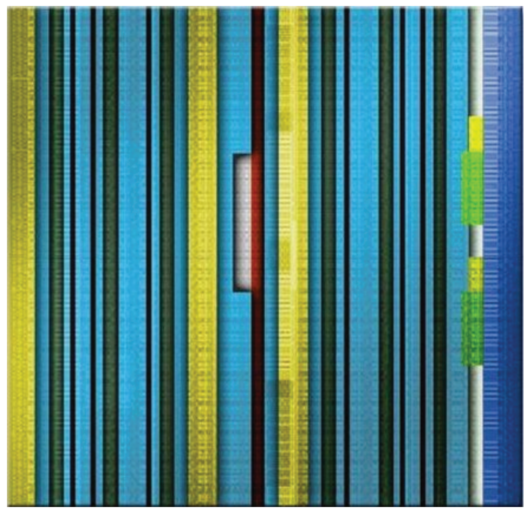


Figure 1: **Stacked Silicon Interconnect Technology Die Top View**

Originally developed for use in a variety of die-stacking design methodologies, silicon interposers provide modular design flexibility and high-performance integration suitable for a wide range of applications. The silicon interposer acts as a sort of micro-circuit board in silicon on which multiple die are set side by side and interconnected. Stacked silicon interconnect technology avoids the power and reliability issues that can result from stacking multiple FPGA dies on top of each other. Compared to organic or ceramic substrates, silicon interposers offer far finer interconnect geometries (approximately 20X denser wire pitch) to provide device-scale interconnect hierarchy that enables more than 10,000 die-to-die connections.

Creating FPGA Die Slices with Microbumps for Stacked Silicon Integration

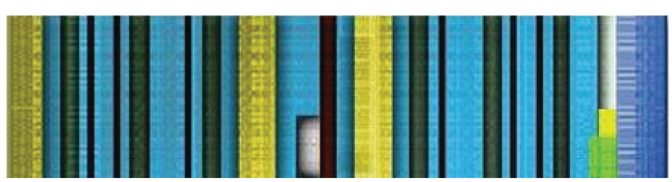
The foundation of Xilinx stacked silicon interconnect technology is the company's proprietary ASMBL architecture, a modular structure comprising Xilinx FPGA building blocks in the form of tiles that implement key functionality such as configurable logic blocks (CLBs), block RAM, DSP slices, SelectIO™ interfaces, and serial transceivers. Xilinx engineers organize the blocks in columns of each type of tile and then combine the columns to create an FPGA. By varying the height and arrangement of columns, the Xilinx engineers can create an assortment of FPGAs with different amounts and mixes of logic, memory, DSP, and I/O resources (Figure 2). The FPGA contains additional blocks for generating clock signals and for programming the SRAM cells with the bitstream data that configures the device to implement the end user's desired functionality.



WP380_02_102210

Figure 2: **Representation of an FPGA Built with ASMBL Architecture**

Starting with the basic ASMBL architectural construct, Xilinx has introduced three key modifications that enable stacked silicon integration (see Figure 3). First, each die slice receives its own clocking and configuration circuitry. Then the routing architecture is modified to enable direct connections through the passivation on the surface of the die to routing resources within the FPGA's logic array, bypassing the traditional parallel and serial I/O circuits. Finally, each slice undergoes additional processing steps to fabricate microbumps that attach the die to the silicon substrate. It is this innovation that enables connections in far greater numbers, with much lower latency, and much less power consumption than is possible using traditional I/Os (100X the die-to-die connectivity bandwidth per watt versus standard I/Os).



WP380_03_102010

Figure 3: **FPGA Die Slice Optimized for Stacked Silicon Integration**

Silicon Interposer with TSV

The passive silicon interposer interconnects the FPGA die. It is built on a low-risk, high-yield 65 nm process and provides four layers of metallization for building the tens of thousands of traces that connect the logic regions of multiple FPGA die (Figure 4).

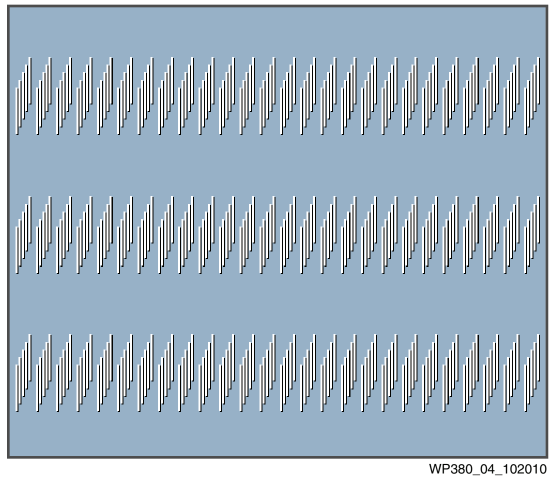


Figure 4: **Passive Silicon Interposer**

Figure 5 illustrates the concept of an "X-ray view" of the assembled die stack. It contains a stack-up of four FPGA die mounted side by side on a passive silicon interposer (bottom view). The interposer is shown as transparent to enable a view of the FPGA die slices connected by traces on the silicon interposer (not to scale).

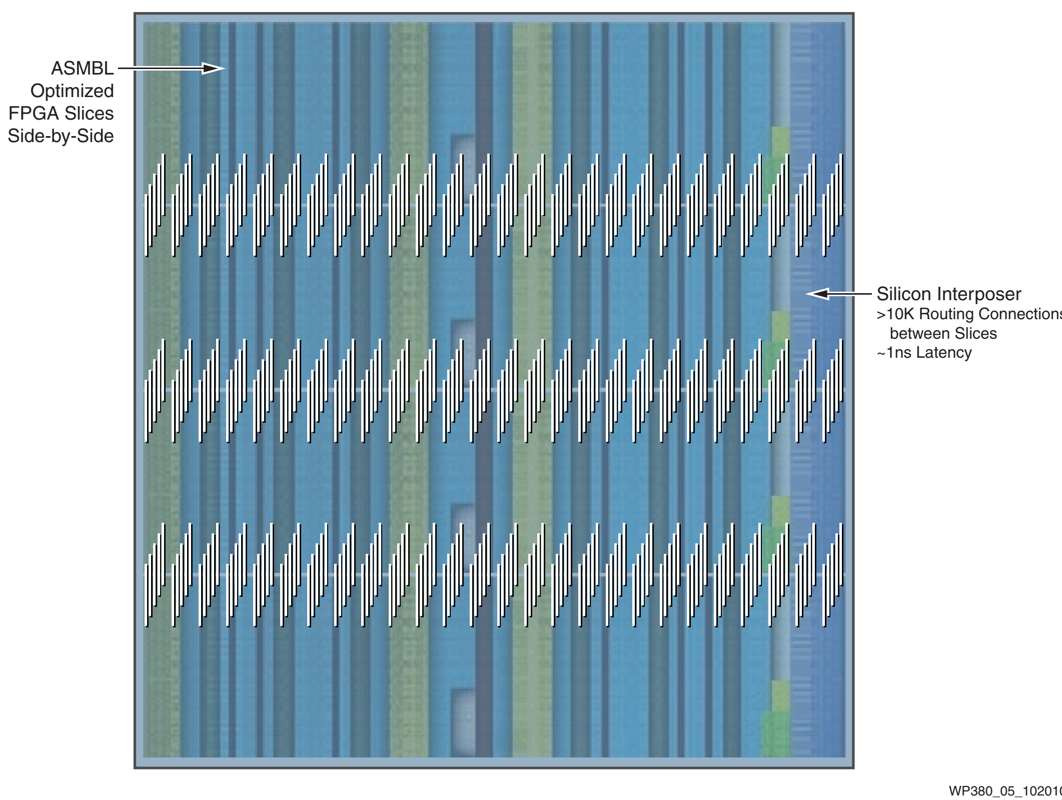


Figure 5: **"X-ray View" of the Assembled Die Stack**

The TSVs combined with controlled-collapse chip connection (C4) solder bumps enable Xilinx to mount the FPGA/interposer stack-up on a high-performance package substrate using flip-chip assembly techniques (Figure 6). The coarse-pitch TSVs provide the connections between the package and the FPGA for the parallel and serial I/O, power/ground, clocking, configuration signals, etc.

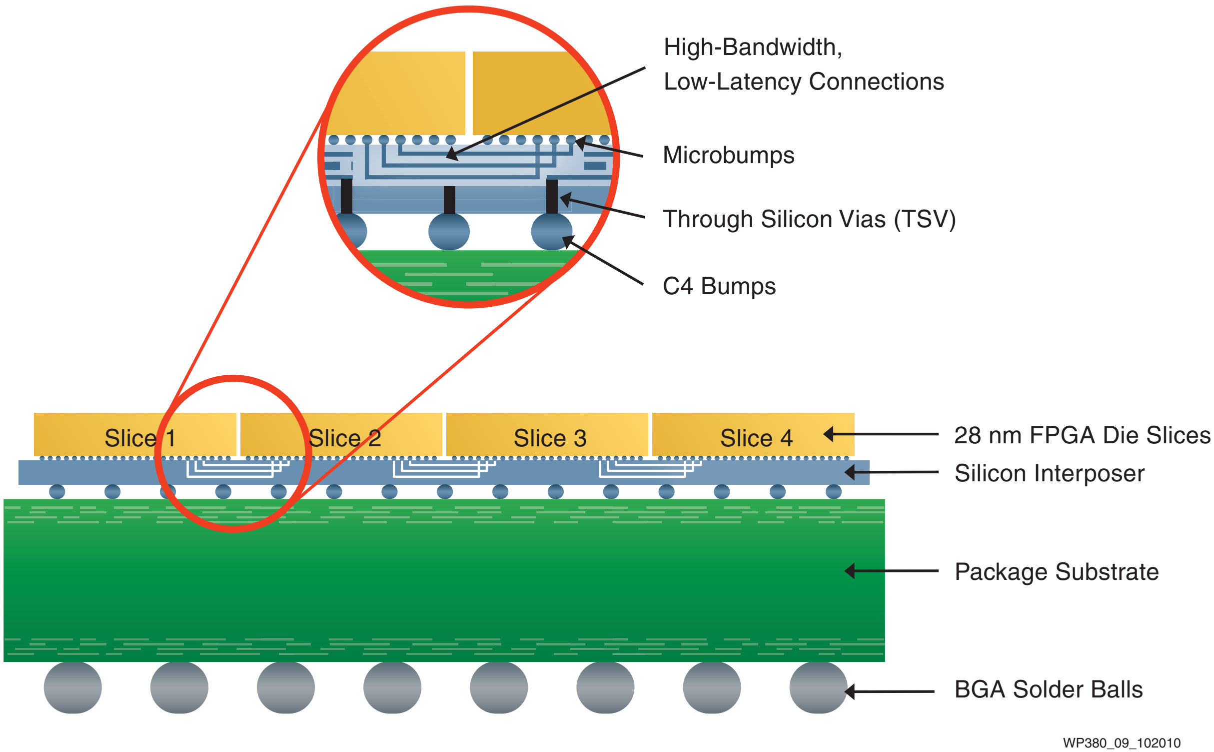


Figure 6: **Package Substrate**

Comprising numerous patent-pending innovations, this stacked silicon interconnect technology provides multi-Terabit-per-second die-to-die bandwidth through more than 10,000 device-scale connections—enough for the most complex multi-die designs. Xilinx is using this new technology to create the Virtex®-7 FPGA family, which offers unprecedented capabilities including up to: two million logic cells; 65 Mb of block RAM; 2,375 GMACS of DSP performance (4,750 GMACS for symmetric filters); 1,200 SelectIO pins supporting 1.6 Gb/s LVDS parallel interfaces; and 72 serial transceivers delivering 1,886 Gb/s aggregate bidirectional bandwidth.

Bringing Stacked Silicon Interconnect Technology to Production

The development strategy Xilinx has employed in the creation of the FPGA with stacked silicon interconnect technology begins with extensive modeling and the creation of a series of test devices, or test vehicles, used for design enablement and for manufacturability and reliability validation.

Stress simulation models show an additional advantage of stacked silicon technology. The silicon interposer functions as a buffer that reduces low-K dielectric stress and improves C4 bump reliability, compared to monolithic solutions.

Extensive simulations investigating the thermal impact of the die stack show that thermal performance of devices with stacked silicon interconnect technology is comparable to that of monolithic devices.

Xilinx is well on the way to volume production of the first FPGAs with stacked silicon interconnect technology, having completed over five years of research and development with industry-leading suppliers and extensive testing on series of multiple test vehicles. These test vehicles address process module development and integration, reliability assessment, supply-chain validation, design enablement, interposer known-good-die (KGD) methodology, and microbump electromigration (EM) rules.

Test vehicle-based reliability tests successfully completed to date include:

- 1,000 cycles of package and wafer level Temperature Cycle B evaluation of TSV, C4 balls, and interposer interconnects
- 1,000 hour high temperature storage evaluation of microbump joints
- 0.1% cumulative density function (CDF) for electromigration at the microbump joint

Xilinx already has a robust supply chain in place for the technologies required to build the industry's first FPGAs with stacked silicon interconnect technology. TSMC, Amkor, and Ibiden contribute their combined resources and expertise for fabricating 28 nm FPGAs and 65 nm silicon interposers, interconnect layers, microbumps, C4 balls, and package substrates as well as performing wafer thinning, die separation, chip-on-chip (CoC) attach, and package assembly.

FPGA Design with Stacked Silicon Interconnect Technology

One of the more substantial advantages afforded by Xilinx FPGAs with stacked silicon interconnect technology is the ability to treat it like a monolithic device. This is extremely important because partitioning a large design across multiple FPGAs presents a number of complicated design challenges that monolithic implementations avoid entirely.

The typical steps in a monolithic FPGA design flow include:

- Create a high-level description
- Synthesize into an RTL description that matches the hardware resources
- Perform physical place and route
- Estimate timing and adjust design for timing closure
- Generate a bitstream to program FPGAs

When working with multiple FPGAs, the designer (or design team) must partition the netlist across the FPGAs. Working with multiple netlists means opening and managing multiple projects, each with its own design file, IP libraries, constraint files, packaging information, etc.

Timing closure for multiple FPGA designs can also be extremely challenging. Calculating and accommodating propagation delays through the board to the other FPGAs poses new and complex problems. Likewise, debugging a design through multiple partial netlists in multiple FPGAs can be extremely complicated and difficult.

In contrast, when using FPGAs with stacked silicon interconnect technology, the designer creates and manages a single design project; stacked silicon interconnect technology routing is transparent to the user. The user performs a single design bring-up and debug with a standard timing closure flow.

Flexible Design Flows

The ISE® Design Suite supports the Virtex-7 family. Designers can choose from among multiple design flows for FPGAs with stacked silicon interconnect technology. Choices include a pushbutton flow and a block-based flow. The first flow focuses on ease of use and provides FPGA performance that is adequate for many designs. This flow automatically looks for ways to separate groups of logic with a minimum amount of interconnect (min-cut) so that each group of logic will route smoothly and efficiently.

The block-based flow facilitates hierarchical design methodologies to support team-based design, incremental builds, and additional performance tuning. The block-based flow also uses the PlanAhead™ design tool for optimal floorplanning.

Applications

Xilinx Virtex-7 FPGAs with stacked silicon interconnect technology break through the limitations of monolithic FPGAs, extending their value in some of the most demanding applications. For example, these devices are perfect for use in ASIC prototyping and can serve as pre-production and/or initial-production ASIC alternatives. In next-generation telecom systems, devices with dozens of serial transceivers enable flexible, single-FPGA solutions, such as a 300G protocol bridging, or a multiplexing transponder implementation that can replace multiple ASSPs, reduce cost by 60%, and reduce power by 50%. They also enable flexible, scalable, customized high performance computing solutions for scientific, oil and gas, financial, aerospace and defense, communications, networking, and life science applications. The parallelism inherent in the FPGA architecture is ideal for high-throughput processing and software acceleration. Support for a multitude of high-speed parallel and serial connectivity standards enables the convergence of compute and communications systems. In Aerospace and Defense, high transceiver count and thousands of DSP processing elements provided by FPGAs with stacked silicon interconnect technology enable advanced RADAR implementations.

Summary

As the only FPGA manufacturer to use stacked silicon interconnect technology to create super high capacity FPGAs with unmatched die-to-die bandwidth, Xilinx is breaking important new ground in the system-level integration arena. Stacked silicon interconnect technology will enable Xilinx to deliver the highest logic density, bandwidth, and on-chip resources with the fastest ramp to volume production at every process node.

Customers will find these FPGAs with stacked silicon interconnect technology significantly easier to design with than multiple FPGAs, with flexible tool flows that provide complete design tools for ease of use, yet allow designer interaction for achieving even higher performance.

Revision History

The following table shows the revision history for this document:

Date	Version	Description of Revisions
10/27/10	1.0	Initial Xilinx release.

Notice of Disclaimer