

A New FPGA Architecture and Leading-Edge FinFET Process Technology Promise to Meet Next-Generation System Requirements

The Intel® HyperFlex™ FPGA Architecture meets the performance requirements of next-generation systems.

Author Introduction

Allan Davidson

Senior Product Marketing Manager
Intel Programmable Solutions Group

The Intel® HyperFlex™ FPGA Architecture and Intel's 14 nm Tri-Gate process technology enable Intel Stratix® 10 FPGAs and SoCs to deliver levels of performance and power efficiency that were unimaginable in previous-generation high-performance FPGAs. These devices deliver:

- 2X the core performance and 5X the density compared to previous-generation Stratix V FPGAs
- Up to 70 percent lower power than Stratix V FPGAs for equivalent performance
- Logic, internal memory, and DSP blocks capable of 1 GHz operation
- Embedded quad-core 64 bit ARM* Cortex*-A53 hard processor system (in SoC variants)
- Familiar FPGA design techniques supported by the proven Intel Quartus® Prime software

Industry challenges

In all major industries, electronic system developers are being challenged to pick up the pace of the breakneck speed increases they have delivered over the past several decades. Not only do customers in markets as diverse as military communications and computer storage environments want faster systems, they also want smaller hardware that consumes less electricity.

The need for speed is not new. What is changing is that requirements in wireless, wireline, military, broadcast, compute and storage are becoming even more demanding. In many of these fields, demands are rising in high double-digit growth rates reminiscent of boom markets.

Table of Contents

Introduction	1
Industry Challenges	1
Delivering the Unimaginable	2
The Intel HyperFlex FPGA Architecture	4
The Hyper-Aware Design Flow ...	5
Conclusion.....	5
References.....	6
Where to Get More Information ..	6

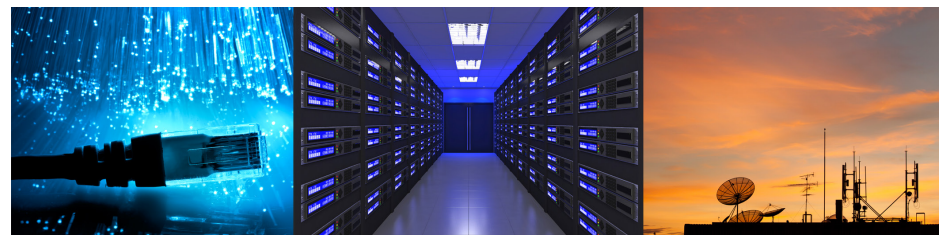


Figure 1. Customer Demands Are Increasing in Applications such as Wireline, Data Center, and Wireless

The information and communications technology sector typifies the customer demands that are challenging the capabilities of system designers and the semiconductor suppliers who support them. Globally, bandwidth consumption is doubling every two to three years as staggering amounts of data traverse global networks. In 2016, the gigabyte equivalent of all movies ever made will crisscross these networks every three minutes. Industry researchers at TeleGeography⁽¹⁾ noted that Internet bandwidth more than doubled between 2010 and 2012, soaring to 77 Terabits per second.

Communication requirements will soar as everything from automotive to industrial equipment to consumer products like refrigerators connect to the Internet. Gartner predicts that the Internet of Things will hit 26 billion units in 2020, almost 30 times more than 0.9 billion in 2009.⁽²⁾

Many trends contribute to the never-ending increase in bandwidth. For example, 100 Gigabits per second (Gbps) Ethernet is just beginning to displace the popular 40 Gbps version, yet the IEEE recently established a task force that will pursue a 400 Gbps standard.⁽³⁾

Wired networks still carry the largest volumes of data, but wireless markets are poised to reverse that. In 2011, wired devices accounted for nearly 55 percent of IP traffic, according to a Cisco report. The explosive growth of smart mobile devices makes it easy to predict that wireless devices will soon generate the majority. Cisco predicts that mobile data traffic will skyrocket from 1.6 exabytes in 2013 to 11.2 exabytes in 2017.⁽⁴⁾

Hefty double digit growth rates in other fields further highlight the need for faster data handling equipment. Satellite communications, which are still driven by military usage, are soaring as drones and satellites generate more data that is used by far more personnel than in the past. That is prompting surging requirements at land-based backhaul stations.

Northern Sky Research (NSR) forecasts greater than 50 percent growth in the global installed base of satellite backhaul sites between 2012 and 2022.⁽⁵⁾ This growth is driven by the need to serve 3G/4G/LTS backhaul requirements cost effectively for mobile operator clients. NSR projects that combined high throughput satellite capacity demand will grow by 133.5 Gbps by 2022 for backhaul services alone. NSR forecasts the global satellite broadband access market will add over 4.3 million net new subscribers in the coming ten years with North America adding nearly 2.4 million new subscribers by 2022.

Back on earth, mobile communications are driving the demand for faster chips and systems. A Cisco forecast predicts mobile traffic growth of 78 per cent from 2011 to 2016. Much of that will be driven by video, which should grow 90 per cent annually between 2011 and 2016. Cisco predicts that by 2016 mobile video will account for over 70 per cent of the mobile data traffic.⁽⁶⁾

The public's demand for video is also fueling strong growth in the broadcast industry. By the end of this decade, most countries are expected to complete the transition to digital TV. The growth of HDTV and the emergence of Ultra High Definition technology are expected to drive the need for faster editing and transmission systems.

Reducing power and heat

In all these fields, simply designing faster equipment is no longer enough. Power consumption has become a major issue, driven by environmental concerns as well as by the cost savings that come with power reductions. Chipmakers and system designers alike have made power reduction a central element in most projects. System design teams also appreciate the decrease in heat that comes with reduced power consumption, which lets them spend less time on heat removal.

Data centers are the poster child for power and heat reduction. Between 2011 and 2012, Global data center power requirements grew by 63 percent, rising to 38 gigawatts (GW) in 2012, up from 24 GW in 2011, the DatacenterDynamics 2012 Global Census said.⁽⁷⁾ In the U.S., many statisticians believe that data centers consume around 2 percent of the country's electricity usage, citing research by Jonathan Koomey.⁽⁸⁾

The need for a new approach

Regardless of the industry, next-generation systems require ever-increasing data throughput and higher clock frequency performance. Faced with this reality and the need to get new products to market quickly, many companies are now using FPGAs as key components in their system designs. The data throughput of these FPGAs is often a critical factor in determining the overall system performance.

To improve the data throughput in an FPGA, the most commonly used technique is to make on-chip buses wider and wider. It is common to use 512 bit, 1,024 bit, or even wider buses in FPGAs. These wide buses require costly FPGA resource utilization and power dissipation. Moreover, it is difficult to perform high-speed logic functions like comparators or checksums across every bit of the bus.

In addition to wider buses, system designers extensively pipeline data paths, increasing the clock frequency. However, pipelining a wide bus requires that each bit of the bus consume additional FPGA resources, which again is expensive. It is not practical to continue to make wider and wider buses.

Moving to the next technology node improves performance. However, as process geometries continue to shrink, the interconnect delays between the logic blocks increasingly dominate the FPGA's total delay. Evolving existing FPGA architectures with the next technology node does not address this concern. A better solution is needed to address these increasingly significant interconnect delays.

Delivering the unimaginable

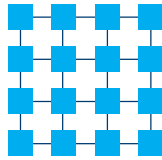
The new Intel HyperFlex FPGA Architecture in Intel Stratix 10 devices is an innovative approach that addresses these concerns. It provides performance and power efficiency that is simply not possible with conventional FPGA architectures. Using the new Intel HyperFlex FPGA Architecture combined with Intel's 14 nm Tri-Gate process technology, designers can achieve 2X the core performance in Intel Stratix 10 FPGAs and SoCs compared to previous-generation high-performance FPGAs.

The HyperFlex advantage

The key innovations that contribute to the HyperFlex advantage are:

Registers everywhere

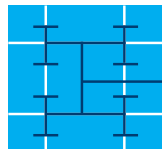
The “registers everywhere” in the interconnect routing, called Hyper-Registers, are distinct from the conventional registers that are contained within the adaptive logic modules (ALMs). A Hyper-Register is associated with each individual routing segment in the device; Hyper-Registers are also available at the inputs of all functional blocks such as ALMs, embedded memory (M20K) blocks, and digital signal processing (DSP) blocks. The Hyper-Registers are bypassable, allowing the design tools to select the optimal register location automatically, after place-and-route, to maximize core performance.



Having Hyper-Registers throughout the interconnect means that performance tuning does not require additional ALM resources (unlike conventional architectures) and does not require additional changes or added complexity to the design's place-and-route. Additionally, having Hyper-Registers built into the interconnect helps to reduce routing congestion.

Enhanced core clocking

The programmable clock tree synthesis allows system designers to create localized clock trees, reducing skew and timing uncertainty to obtain maximum core clocking performance. This capability is a key feature that allows the Intel HyperFlex FPGA Architecture to reach 2X performance. In addition, the core clocking uses intelligent branch-enables to reduce the dynamic power dissipation in the clock networks.



Hyper-Aware design flow

The Hyper-Aware design flow includes three new improvements:



- A Fast Forward Compile tool that allows performance exploration and guides the user to maximum design performance.
- A Hyper-Retimer step that supports performance optimization after place-and-route.
- Enhanced synthesis and place-and-route algorithms that use the Hyper-Registers.

The benefits of high performance—beyond high performance

The Intel HyperFlex FPGA Architecture's increased core performance offers several benefits to the system designer; benefits that go beyond the obvious one of simply running the core faster:

Higher core performance makes timing closure easier and faster, improving the design team's productivity and shortening the product's time-to-market.

Higher core performance allows designers to use a slower speed grade device while still exceeding performance requirements, reducing the cost of the solution.

Higher core performance that allows a design to run 2X faster can be implemented at half the original internal bus width, shrinking the total design size. Therefore, the design fits in a much smaller device, reducing the cost of the solution.

The Intel advantage

In February of 2013, Altera announced that Intel's 14 nm Tri-Gate (FinFET) process technology would be used to fabricate the next-generation Intel Stratix 10 FPGAs and SoCs. This technology provides breakthrough levels of density, performance, and power efficiency. It is based on 3D FinFET (Tri-Gate) transistors, which are replacing conventional 2D planar MOSFET transistors as geometries shrink below 20 nm. All major silicon foundries have announced their intention to move towards the 3D FinFET transistors. In December 2015, Intel completed the acquisition of Altera. With Intel as the foundry for Intel Stratix 10 devices, Intel® FPGA customers have access to a number of unique benefits, provided by “The Intel Advantage.” These benefits make the Intel 14 nm Tri-Gate technology the ideal process for implementing the new Intel HyperFlex FPGA Architecture.

The top five benefits that Intel FPGA customers obtain are:

- **Exclusivity**—Intel is the only major FPGA vendor that has uses 14 nm Tri-Gate technology. Only Intel FPGA customers have access to this industry leading process technology.
- **Production Capability**—Other major semiconductor foundries have announced plans to develop new processes based on FinFET transistors. However, there is a steep learning curve when moving FinFET technology from the research labs into production. So far, only Intel has made the transition into production—it has already shipped over 500 million FinFET transistor devices.
- **A Node Ahead**—Intel debuted its Tri-Gate process at 22 nm, over three years ago. This technology has shrunk to 14 nm, which is the technology used in Intel Stratix 10 FPGAs and SoCs. The other semiconductor foundries are developing FinFET processes that will start out using existing 20 nm design rules. They are not employing the same shrink as Intel, effectively, leaving Intel a node ahead, resulting in sizable performance, power efficiency, and density advantages.
- **Maturity**—Intel is utilizing second-generation 14 nm Tri-Gate technology. None of the other foundries have publically stated when they will start building chips with first-generation FinFET processes. Intel Stratix 10 FPGAs and SoCs benefit from the maturity of the Intel 14 nm Tri-Gate process technology.
- **Design Expertise**—Intel has proven its ability to design and produce high-speed logic, analog, digital, and mixed-signal circuits using FinFET transistors. This wealth of design expertise, ensures that Intel Stratix 10 FPGAs and SoCs make the best use of the capabilities of Intel's 14 nm Tri-Gate process technology.

Intel also offers the only major, high-performance FPGAs and SoCs with US-based manufacturing. It provides access to world-class package and assembly capability; and it

enables Intel to develop heterogeneous multi-die devices that integrate 14 nm Intel Stratix 10 FPGAs and SoCs with other advanced components—which may include SRAM, DRAM, ASICs, processors, and analog components—in a single package. These benefits form “The Intel Advantage,” exclusively available to Intel Stratix 10 FPGA and SoC customers.

The Intel HyperFlex FPGA Architecture

The centerpiece of the new Intel HyperFlex FPGA Architecture is its innovative “registers everywhere” design that adds bypassable Hyper-Registers to every routing segment in the FPGA core and at all functional block inputs. Figure 2 shows a bypassable Hyper-Register where the routing signal can bypass the register and go straight to the multiplexer, or go through the register first. The multiplexer is controlled by one bit of the FPGA configuration memory (CRAM).

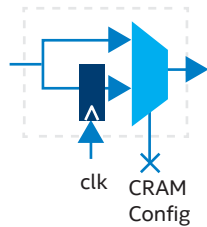


Figure 2. Bypassable Hyper-Register

Figure 3 shows a small section of the FPGA fabric with nine ALMs and the interconnect routing that connects them. The Hyper-Register location is indicated by the squares at the intersection of each horizontal and vertical routing segment.

To maximize the performance of a design using the Intel HyperFlex FPGA Architecture, designers use a three-step process that is based on familiar design techniques: register retiming, pipelining, and design optimization. The Hyper-Registers allow designers to use familiar design techniques to increase the performance of the design well beyond what is possible in conventional FPGA architectures. When these common techniques are implemented using the Hyper-Registers instead of the registers in the ALMs, the techniques are renamed as Hyper-Retiming, Hyper-Pipelining, and Hyper-Optimization. Table 1 summarizes the performance gains achieved in each step.

As process geometries shrink, the interconnect delays between the ALMs are becoming dominant and are limiting performance. Locating the Hyper-Registers in the interconnect routing—where they can best address this issue—is one of the key innovations of the Intel HyperFlex FPGA Architecture.

Hyper-Retiming

The design is retimed using the Hyper-Registers in the interconnect routing. This process requires little to no user effort yet it results in an average performance gain of 1.5X for Intel Stratix 10 devices compared to previous generation high-performance FPGAs. Hyper-Retiming

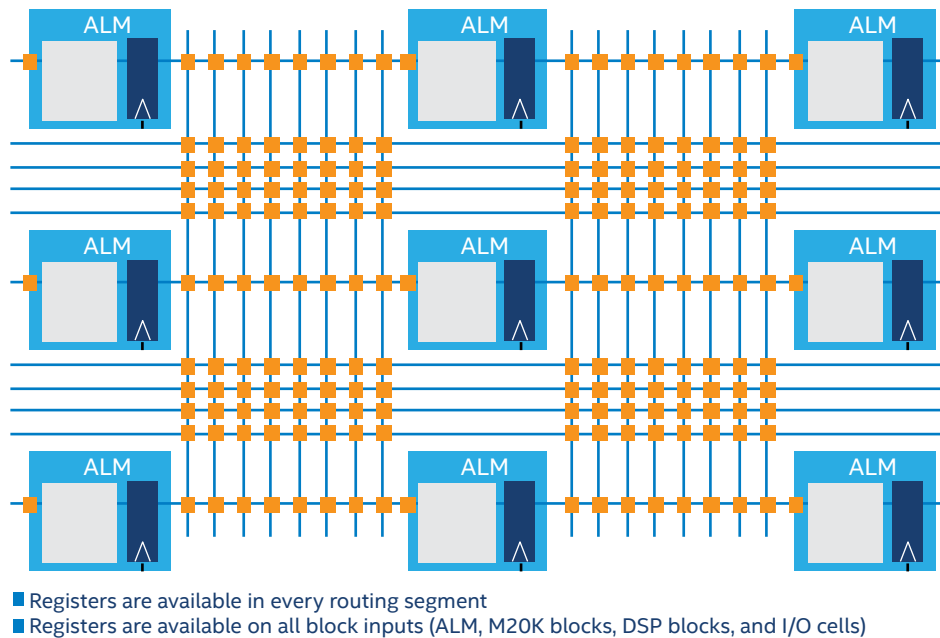


Figure 3. “Registers Everywhere” Intel HyperFlex FPGA Architecture

Step	Architecture Advantage	Effort Required	Core Performance (vs. Previous-Generation High-Performance FPGA)
1	Hyper-Retiming	No change, or minor RTL changes	1.5X
2	Hyper-Pipelining	Added pipelining	1.65X
3	Hyper-Optimization	Design dependent	2X or more

Table 1. Three-Step Process to Maximize Performance Using the Intel HyperFlex FPGA Architecture

eliminates critical paths by moving registers out of the ALMs and into the interconnect, balancing register-to-register delays and allowing the design to run at a faster clock frequency. Because there are Hyper-Registers throughout the interconnect, the register location is fine-grained. Conventional retiming requires additional FPGA logic and routing resources and requires the design to be recompiled, refitted, and rerouted. In contrast, Hyper-Retiming does not use any additional FPGA resources and is performed after place-and-route, providing a significant core performance boost with little or no designer effort.

Hyper-Pipelining

The design is pipelined and retimed using the Hyper-Registers. This technique requires minor user effort and results in an average performance gain of 1.65X for Intel Stratix 10 devices compared to previous generation high-performance FPGAs. Hyper-Pipelining eliminates long routing delays by adding additional pipeline stages in the interconnect between the ALMs, allowing the design to run at a faster clock frequency. Again, the Hyper-Registers located throughout the interconnect allow a fine-grained selection of the register location. As with Hyper-Retiming, Hyper-Pipelining does not use additional FPGA logic and routing resources, and it is done after place-and-route.

Hyper-Optimization

After accelerating data paths with Hyper-Retiming and Hyper-Pipelining, some designs are limited by control logic such as long feedback loops and state machines. To achieve higher performance, it is necessary to restructure these logic sections to use functionally equivalent feed-forward or pre-compute paths instead of long combinatorial feedback paths. This method requires a bit more effort, depending on the design; however, it results in average performance gains of 2X or more in Intel Stratix 10 devices compared to previous generation high-performance FPGAs. In a conventional architecture, this process is called design optimization. In the Intel HyperFlex FPGA Architecture, this process is called Hyper-Optimization because the Hyper-Registers apply the benefits of Hyper-Retiming and Hyper-Pipelining to the feed-forward or pre-compute paths.

The Hyper-Aware design flow

Intel has developed a powerful set of new tools, integrated into the Intel Quartus Prime design software, that help system designers take full advantage of the Intel HyperFlex FPGA Architecture and maximize the developer's design



Figure 4. The Hyper-Aware Design Flow

productivity. Figure 4 shows the Intel Quartus Prime Hyper-Aware design flow.

Fast Forward Compile

This new tool guides the user through the performance optimization process by identifying performance limiting areas of the design, identifying where and how many pipelines could be used to boost performance, and highlighting critical control-path bottlenecks (such as long feedback loops). The tool also allows designers to predict the performance of their existing design if it were implemented in a Intel Stratix 10 device, enabling optimal use of the new Intel HyperFlex FPGA Architecture.

Hyper-Retimer

The Hyper-Retimer step occurs near the end of design compilation. It performs post place-and-route performance optimization using the Hyper-Registers for optimal fine-grained Hyper-Retiming. This step also allows the user to implement Hyper-Pipelining much more easily than conventional pipelining. The Fast Forward Compile report identifies which clock domains can benefit from pipeline stages and how many pipeline stages are needed. After the designer modifies the RTL and places the prescribed number of pipeline stages at the boundaries of each clock domain, the Hyper-Retimer automatically places the registers within the clock domain at the optimal locations to maximize the performance. This auto-placement along with the Fast Forward Compile report makes pipelining easier than ever.

Hyper-Aware algorithms

Hyper-Aware algorithms used during synthesis and place-and-route allow the tool to reduce logic resources by predicting which registers can be moved out of ALMs and into Hyper-Registers in the interconnect routing.

Conclusion

The combination of the new Intel HyperFlex FPGA Architecture and the Intel 14 nm Tri-Gate process technology enables Intel Stratix 10 FPGAs and SoCs to deliver previously unimaginable levels of performance, density, and power efficiency in a programmable logic device. Intel Stratix 10 devices offer:

- 2X the core performance and 5X the density compared to previous-generation Stratix V FPGAs
- Up to 70 percent lower power than Stratix V FPGAs for equivalent performance
- Logic, internal memory, and DSP blocks capable of 1 GHz operation
- Embedded quad-core 64 bit ARM Cortex-A53 hard processor system (in SoC variants)
- Hyper-Aware design flow
- Familiar FPGA design techniques supported by the proven Intel Quartus Prime software

References

- ¹ <http://www.telegeography.com/press/press-releases/2012/09/06/global-internet-capacity-reaches-77-tbps-despite-slowdown/index.html>
- ² <http://www.gartner.com/newsroom/id/2636073>
- ³ <http://www.ieee802.org/3/400GSG/>
- ⁴ http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html
- ⁵ <http://www.nsr.com/news-resources/the-bottom-line/satellite-backhaul-trunking-are-capacity-driven-markets/>
- ⁶ http://www.ciscoknowledgenetwork.com/files/222_03-27-2012-CKN_Cisco_Mobile-VNI-Forecast_2012_CKN_Deck.pdf
- ⁷ <http://www.computerweekly.com/news/2240164589/Datacentre-power-demand-grew-63-in-2012-Global-datacentre-census>
- ⁸ <http://www.koomey.com/post/8323374335>

Where to Get More Information

For more information about Intel and Intel Stratix 10 FPGAs, visit <https://www.altera.com/products/fpga/stratix-series/stratix-10/overview.html>

- ¹ <http://www.altera.com/literature/wp/wp-01201-fpga-tri-gate-technology.pdf>
- ² <http://www.altera.com/devices/fpga/stratix-fpgas/stratix10/stx10-index.jsp>

