Expect a Breakthrough Advantage in Next-Generation FPGAs

Stephen Lim, Product Marketing Manager, Components, Intel Corporation

WP-01199-1.2 White Paper

This white paper covers examples of why telecommunication bandwidth and the infrastructure behind it is driving FPGA capabilities, business challenges of ASICs and ASSPs, and how a tailored approach for programmable logic devices (PLDs) provide a leap in FPGA capabilities. This paper also outlines a portfolio of next-generation FPGAs and SoCs.

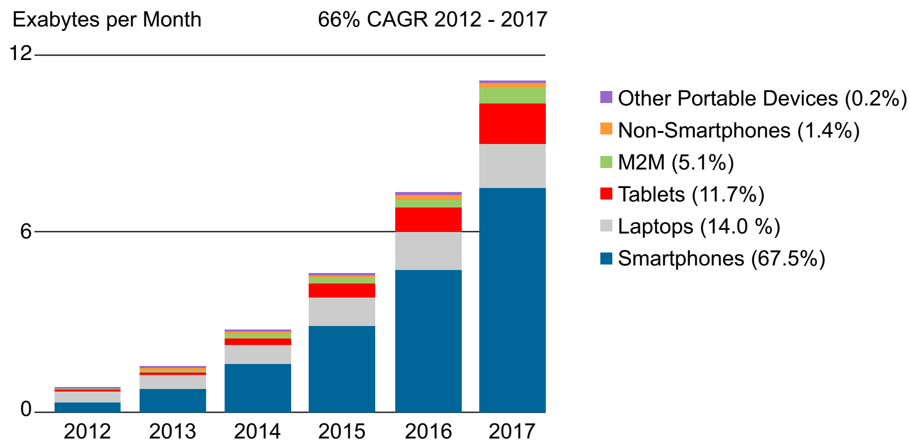
Introduction

Newly announced FPGAs will be one of the key enablers for hardware architects, software developers, and system designers to achieve their next-generation product goals. Exponentially increasing bandwidth requirements throughout the broad telecommunication infrastructures and the industries using this bandwidth are making it difficult for existing hardware and software solutions to offer the needed performance while still meeting their cost and power goals. ASICs, ASSPs, and standalone processors have growing limitations and inherent costs that PLD companies address. At the same time however, the broad scope of end applications that face these increased bandwidth challenges will require PLD companies to use different tools and options to meet a diverse set of needs. PLD companies that have access to these options and the ability to apply them effectively will give hardware and software developers breakthrough advantages and capabilities to build their next-generation products.

The Need for Ever-Increasing Bandwidth and Flexibility Drives the Need for a Breakthrough in Capability

The increased capabilities in smartphones and other portable devices are the reason for the dramatic leap in system performance that we will see in next-generation FPGAs. The explosion of mobility bandwidth requirements are putting a huge demand on the wireless, wired, and data center infrastructure capabilities. While the number of smartphones is growing at single digit percentage rates, the customers of these devices continue to drive more bandwidth with the ever-increasing smartphone capability. Much of this is due to the increased video content. In 2012, average smartphone data usage grew by 81 percent. Cisco expects mobile traffic to increase 66 percent per year through 2017 and two-thirds of all mobile traffic will be video content. At this time, mobile network speed is expected to increase by seven times and 4G networks to comprise 45 percent of all traffic [(1)] (see Figure 1).

Figure 1. Cisco Forecasts 11.2 Exabytes per Month of Mobile Traffic by 2017



Figures in legend refer to traffic share in 2017
Source: Cisco VNI Mobile Forecast, 2013

A brief overview of three infrastructure applications below are examples of why hardware and software developers are looking to FPGAs to address their next-generation products bandwidth, performance, power, and cost goals.

- Wireless remote radio units
- 400G wireline channel cards
- Data centers

Wireless Remote Radio Units

In the capital-intensive wireless infrastructure market, telecommunications operators desire to provide more bandwidth faster and cheaper. The faster these operators can do cost reductions, the more deployments they can do, the more area they can cover, and the faster they can serve customers—a huge advantage. The product strategy of these companies is to keep the datapath width the same and increase the clock frequency for as many generations as they can. Upcoming remote radio units will look for FPGAs to push close to 500 MHz of core performance for complex functions, such as implementing digital pre-distortion algorithms. This will preserve their investment in their radio architecture and allow them to cover a broader spectrum of radio frequency (RF) bandwidth. In doing so they look to have a better return on investment because less work needs to be done re-architecting a solution. Furthermore, their time-to-market advantage improves by getting these new products out faster. They must also lower their operating costs to drive cost per bit down because revenues per mobile subscriber grow at a far less rate than the data traffic per subscriber. Thus by not widening their datapath, and creating power efficient designs on smaller more power-efficient FPGAs, allows them to achieve this goal.

For more information, refer to the *Designing Polyphase DPD Solutions with 28 nm FPGAs* white paper.

400G Channel Cards

Another driving force in improving FPGA performance is the need to upgrade the network communications infrastructure. Next-generation 400G versus existing 100G channel cards will dramatically push system capabilities. The bandwidth jump of four times in the next-generation systems is much greater than in previous iterations. Because the market for this is still new, companies cannot risk building ASICs or ASSPs to achieve this goal. Integration of multiple 56 gigabits per second (Gbps) and 28 Gbps transceiver solutions to accommodate this level of bandwidth is needed, but only a part of the solution. More and faster logic to accommodate this higher bandwidth is also required. However since the dimensions of the chassis do not change, the power envelope is limited. The network infrastructure cannot tolerate solutions where power increases at a linear rate with bandwidth capability. For packet processing and traffic management applications at 400G bandwidth at 600 million packets per second, scaling the data path width and frequency can relieve the data path processing function but cannot scale for control path processing such as scheduling. Therefore high performance in all aspects of device capability is required: processing, memory interfacing, IO interfaces, and others. FPGAs remain the most attractive solution, but companies will need investments in higher performance per watt architectures, transceivers, and process technology to address this large leap in capabilities and challenges.

Data Centers

All the data and video that are being pushed and downloaded from these new wireless deployments and transported through the new 400G packet processing infrastructure also needs to be stored and processed. Computations per watt and computations per dollar is a key metric in data centers. FPGA's are increasingly used in the data center for data access, algorithm, and networking acceleration. Data center servers are bottlenecked getting access to data. The latest processors have more and more cores, but the bandwidth to external memory and data is not keeping pace with the increase in computing power. Many of these servers are running at average utilization rates and are well under peak processing power. These servers are good candidates for FPGA acceleration. Hardware acceleration through FPGAs becomes an attractive alternative to replacing these processors by focusing on the performance bottlenecks that software on processors cannot overcome.

Other applications are also looking to FPGAs to support their increased bandwidth requirements, such as video content providers moving to 4K video, cloud computing, and intelligence applications in defense. These applications face similar issues.

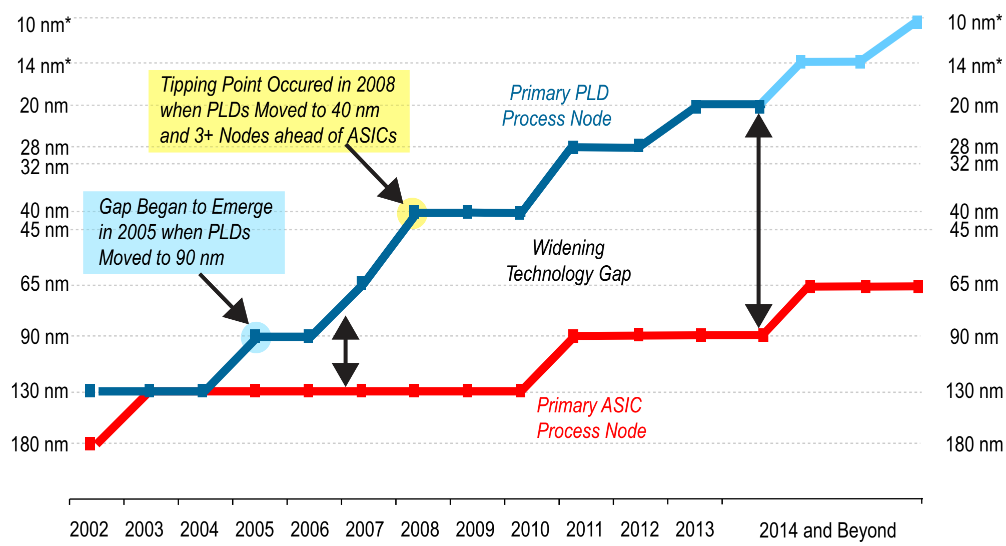
☞ For more information, refer to Microsoft Research: A Reconfigurable Fabric for Accelerating Large-Scale Datacenter Services

Increasing Business Challenges of Using ASICs and ASSPs

The longer time to market, higher upfront capital outlay, and high volumes required to justify designing an ASIC make it a very high-risk investment that fewer companies are venturing into. ASIC non-recurring engineering (NRE) costs for tooling masks and packages, intellectual property (IP) licensing, and physical design services can easily surpass $10 million for a 28 nm ASIC that in many cases can be addressed by a 20 nm or 14 nm FPGA. Although current generation FPGAs require a rigorous simulation verification methodology rivaling ASICs, the additional lab testing and ability to reprogram FPGAs save substantial manpower investment versus standard cell ASIC design. The overall cost of ownership must be considered when comparing an FPGA whose component price is higher than an ASIC of similar complexity. The break-even point justifying use of standard cell ASICs continues to move higher as leading-edge CMOS technology drives FPGA complexity, higher performance, and lower power not economically feasible for ASICs.

Using less expensive process nodes would put the ASIC at a disadvantage compared to FPGAs and ASSPs because these solutions can aggregate customers onto a more advanced process node and then compete on price and performance. Current generation FPGAs use 28 nm processes and are soon to be on 20 nm and smaller process technologies. However, most new ASIC design starts lag two to three nodes or more behind. The bigger the gap the more attractive FPGAs become in terms of price, performance, and level of integration. See Figure 2.

Figure 2. Programmable Logic vs. ASIC Primary Process Nodes for New Designs



Source: Altera; Data applies to new design starts.
** FinFET Technology. Timeframe for PLDs on 14 nm and 10 nm FinFET technology to be announced.*

Gartner expects the overall number of ASIC design starts to decline at a rate of 3.8 percent per year through 2016. In addition, with each passing year, each design start requires more volume to be profitable [(3)]. With only the largest of companies able to justify the cost of ASICs for their market, ASSPs and FPGAs become the only economically viable options for most companies.

However, the ASSP value proposition is also declining for several reasons:

- Challenges of processor performance scaling
- Increasing need for differentiation
- Increasing need to respond to markets (time-to-market)
- Limited flexibility for reconfigurability

Hardware architects were once able to rely upon increases in processor frequency and number of processor cores as a way to increase system performance in their next product. However, now hardware architects cannot rely on this method to increase system performance because processor frequency is not dramatically increasing over time, and parallelization by increasing the amount of processor cores might not address performance bottlenecks. Many hardware architects see the solution as creating specialized hardware to offload these software bottlenecks.

Creating specific hardened IP accessible by the processor alleviates some of these bottlenecks. However, all the additional hardware acceleration that makes the ASSP better than the previous generation is accessible for competing companies as well. Additionally, there may be unique software that has bottlenecks that cannot be sped up using an ASSP.

A key benefit of ASSPs is the fastest time-to-market, but not always. Smaller companies that need specific features in an ASSP not yet available have little leverage on the exact part they need or when they can have it delivered. Larger companies will also be reliant up on suppliers delivering exactly what they request. But once they get this part, it will also be available for other companies as well. FPGAs are an attractive solution to overcome these inherent challenges of ASICs and ASSPs and will be even more so with a dramatic increase in capabilities in the upcoming generation.

A Tailored Approach Provides Breakthrough Capabilities

To address the ever-increasing bandwidth and performance needs of communications, defense, broadcast, and storage, and still provide optimal solutions for extremely cost-sensitive and milli-watt power solutions in markets, such as factory automation, automotive, and consumer portables—a broad and deep set of expertise and tools are required. These include, but are not limited to:

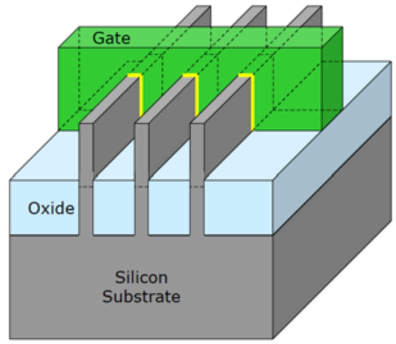
- Leading-edge manufacturing processes technology
- Investments in different architecture and IP
- High-performance integration of processors with programmable fabric

Leading-Edge Processes

Access to advanced process technologies is essential and a key advantage for semiconductor suppliers who invest at the leading edge. For example, new 3-D transistor technology also known as Tri-Gate or FinFET transistor technology is a breakthrough change in process technology (see Figure 3). It allows for a two times decreases in leakage current on transistors, which enables high performance or power capabilities.

☞ For more information, refer to *The Breakthrough Advantage for FPGAs with Tri-Gate Technology* white paper.

Figure 3. Tri-Gate Process Technology



Intel has shipped over 500 million units as of the third quarter of 2014, a testament to their process maturity and high experience working with FinFET based technology. Programmable solutions companies that can adopt this quickly and effectively will be able to provide significant performance gains. Furthermore, customers should look and ask for performance improvements not just in this 3-D transistor technology, but also in the process shrink as well. The recently announced 14 nm Tri-Gate process from Intel provides this process technology.

It is a known fact that no single process technology can meet the requirements of the diverse requirements in end equipment today—even if it is the smallest geometry or most "advanced" process. Suppliers of FPGAs and other programmable SoC products that solely rely on a one-size-fits-all approach, do a disservice to their customers. Factors such as time-to-market, cost, system integration with other components, and unit volumes may favor using other process technologies. For example, newer process nodes may not optimally support higher I/O voltages. Other types of process nodes have stronger advantages in cost per pin per I/O. So while 14 nm Tri-Gate process will be a foundation to provide the highest core performance at the lowest power, it may not be the optimal solution for all system applications. Other process technologies can complement Intel's 14 nm Tri-Gate process, such as TSMC's 20SoC and 55 EmbFlash, to meet a wide-range of system design goals.

For example, the 20SoC process from TSMC will allow customers to bring next-generation FPGAs into production even into high-volume bandwidth-intensive infrastructure markets ahead of 14 nm device availability. Customers will also see core performance improvements allowing systems to run in excess of 500 MHz and its ARM® processors up to 1.5 GHz while still lowering power by over 50 percent from similarly capable FPGAs that are in volume today. This 20 nm process will be one of the foundations to allow customers to meet key goals, such as cost per bit and performance per watt that telecommunications, data centers, and other applications require. Other processes such as embedded flash processes will allow system designers to have the lowest cost per I/O pin, enable milli-watt power solutions, and incorporate analog circuitry and non-volatile flash that are not economically viable on other processes.

☞ For more information, refer to the *Meeting the Performance and Power Imperative of the Zettabyte Era with Generation 10* white paper.

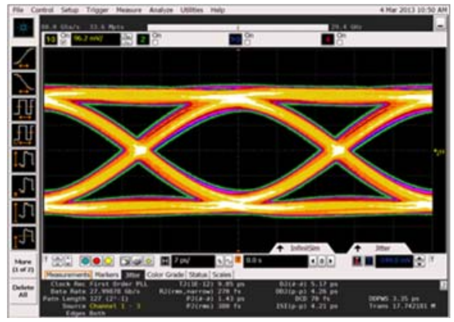
Architecture and IP

To meet the performance requirements of four times more bandwidth than today's applications requires more than advanced process technology. It will require new logic architectures, new IP, new serial interconnect, and so on.

Next-generation architectures will significantly improve core performance when combined with leading-edge process technology. For example, Altera recently announced the new high-performance architecture. When combined with Intel's 14 nm Tri-Gate process, it would allow for astonishing core speeds of up to 1 GHz. Included with this architecture is a dramatic improvement in digital signal processing (DSP) capability. Already an area where FPGAs excel, these DSP blocks will become much more efficient in floating-point operations. FPGAs enabled with them will allow for over 10 trillion floating-point operations per second (teraFLOPS) of performance. These will be one of the highest performance and most power-efficient solutions at 80 giga floating-point operations per second (GFLOPS) per watt, unachievable in existing DSPs or graphics processing units (GPUs). It would allow for breakthrough capabilities in high-performance computing data-intensive applications in financial, energy, cloud data analytics, and so on.

Serial bandwidth will dramatically improve as well by increasing the data rates, the number of channels, and by including more hardened features. FPGA companies have announced that their next-generation transceiver technology would run 56 Gbps data rates. Companies such as Altera currently offer FPGAs with monolithic transceivers at data rates up to 25.78 Gbps. The number of 25.78 Gbps channels alone will increase more than four times on next-generation FPGAs to implement multiple instances of next-generation 100G optical interfaces, such as CFP2, CFP4, and QSFP28. With enhanced signal conditioning techniques, such as adaptive decision feedback equalizers (DFE), transceivers are able to address high loss backplane applications even in electrically noisy environments. Furthermore, using techniques, such as hardened forward error correction (FEC), you can extend backplane reach well over 30 dB of channel loss to allow usage of lower-cost materials without sacrificing system bit error rate (BER) performance. Usability of transceivers is improving with the hardening of functions. For example, hard physical coding sub layer (PCS) blocks are available to handle multiple encoding schemes, such as 8b/10b and 64/66b along with key processing functions for Interlaken and 10 Gbps Ethernet (GbE) data streams. In addition, full protocol stacks are available for PCI Express® (PCIe®) Gen1, Gen2, or Gen3. Serial memory will also become widely deployed for usage with upcoming FPGAs. Serial memory interfaces leverage high-speed serial transceivers at 10-15 Gbps to overcome the bandwidth, latency, and power limitations of parallel memory interfaces. See Figure 4.

Figure 4. 25.78 Gbps Operation on 20 nm Process Technology From Altera



While the newest architecture, IP, and serial technology will be required for certain applications, such as 400G solutions, they would not be optimal for other applications and might adversely affect power and cost goals. Selectively applying these pieces of technology in different FPGAs targeted at different applications is essential.

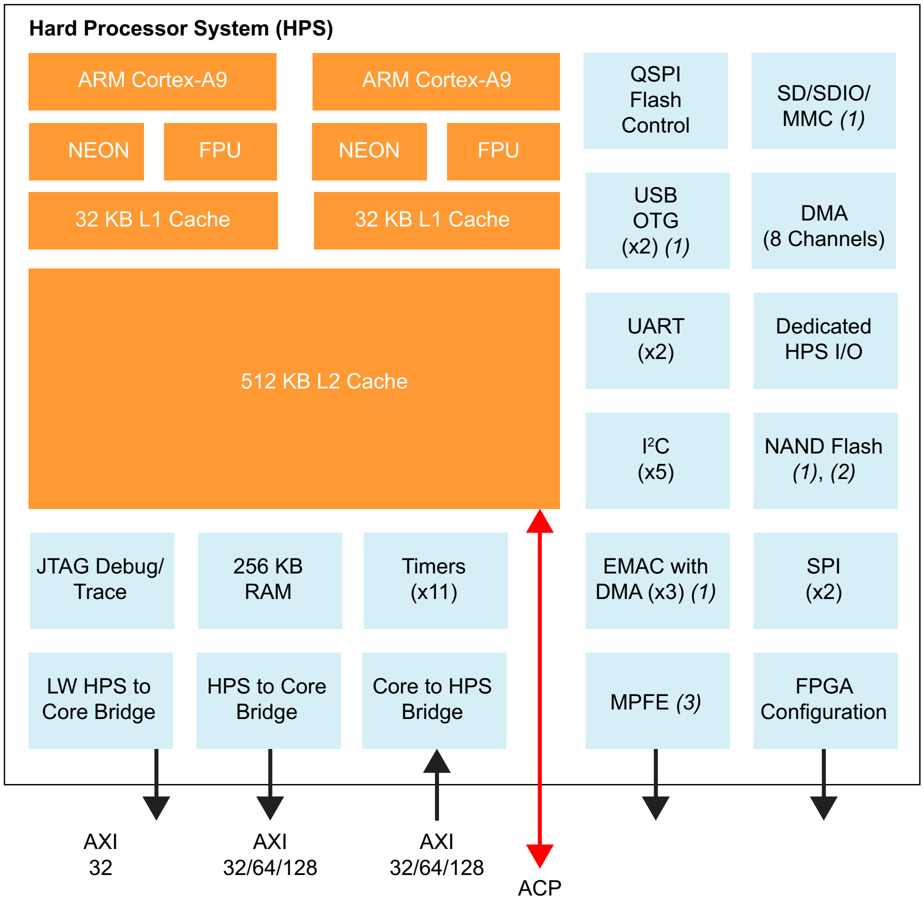
Processor Integration

While FPGAs capabilities have always sought to increase the level of integration of more components on a board, one of the most impactful has been the recent integration of ARM-based hard processor systems (HPS). The HPS integrate independent, but tightly integrated processors and hard peripherals to programmable logic to create system on a chip (SoC) solutions. While this integration began in the 28 nm programmable logic technology with ARM Cortex®-A9 processors, the greater proliferation of this processor architecture within FPGAs and a roadmap for these SoCs will have a positive reinforcement that a new ARM processor supplier has a long-term product roadmap. System architects will now have more options on having tighter integration to improve system performance, lower system costs, reduce system power, and reduce supply chain risk. System architects who have not looked at these programmable SoCs might be surprised to see:

- The numerous SoC offerings that exist across different types of device families
- The closely coupled integration between the programmable logic and processors to provide the high performance and low latency
- Engineers have access to this technology now with 28 nm SoCs, development kits, and tools.
- The level of ARM ecosystem support by select FPGA vendors

Figure 5 shows the second-generation HPS block with an ARM Cortex-A9 processor.

Figure 5. Second-Generation HPS Block with ARM Cortex-A9 Processor



Notes:
1. Integrated direct memory access (DMA).
2. Integrated ECC.
3. Multi-Port front-end interface to hard memory controller.

Next-Generation FPGAs and SoCs Are Coming

The first company to announce next-generation PLDs following the 28 nm process node is Altera with its Generation 10 portfolio. Altera uses the tailored approach, the most extensively of all the PLD providers by using different process technologies, different architectures and IP, and different methods of integration for its variety of low-cost, midrange, and high-end product families. The Generation 10 portfolio includes the Stratix® 10 and Arria® 10 FPGAs and SoCs to address applications that need a few moderate-speed transceivers all the way to applications that require multiple 28 and 56 Gbps transceivers. By applying the tailored approach to these two device families, they will have the largest leap in capabilities that hardware architects and system designers are yet to see thus far in an FPGA.

Altera's Generation 10 FPGAs and SoCs

Arria 10 FPGAs and SoCs

- Performance optimized for wireless, wireline, broadcast, and military applications
  - 1.6 times higher core performance versus previous generation midrange devices, and 15 percent faster than previous generation high-end FPGAs
  - Four times more bandwidth versus previous generation midrange devices, and two times more than previous generation high-end FPGAs, including 25.78 Gbps transceivers
  - Three times system performance (2,666 Mbps DDR4 SDRAM, Hybrid Memory Cube support, 1.5 GHz ARM processor)
- System integration for optimal capability or cost
  - Extensive ARM SoC options
  - Two times higher density with over one million logic elements (LEs)
- Power optimized to enable industry's highest midrange performance
  - 40 percent power improvement versus previous generation midrange devices, and 60 percent improvement versus previous generation high-end FPGAs

Stratix 10 FPGAs and SoCs

- Industry's first Gigahertz FPGAs and SoCs
  - Two times more core performance versus previous generation, four times transceiver bandwidth versus prior generation, with up to 144 channels, and a path to 56 Gbps transceivers
  - Up to 10 teraFLOPS of single-precision floating-point DSP performance
- Highest system integration possible in a single die
  - Industry's only major FPGA on Intel's 14 nm Tri-Gate process technology
  - Up to 5.5 million logic elements in a single die
  - 64 bit quad-core ARM Cortex-A53 processor
- Power optimized to enable industry's highest performance
  - Up to 70 percent lower total power versus previous generation

Hardware engineers using today's current generation of Altera® FPGAs will be in the best position to take advantage of these FPGAs by leveraging the same productivity tools, IP, and design migration capability. Software developers already have the ability to target the ARM HPS using Altera's SoC development kits and other tools. Furthermore, productivity with the design tool flow is expected to improve. Design creation time is expected to decrease with additional design tools and methodologies, such as Open Computing Language (OpenCL™), which allows HDL to be developed with C code. In addition, Altera also recognizes the need to improve compile times of two times per year to keep pace with these leaps in capability.

☞ OpenCL and the OpenCL logo are trademarks of Apple Inc., and used by permission by Khronos.

Conclusion

System architects in multiple markets are looking for alternatives to ASIC and ASSP solutions, and solutions that can address their bandwidth, performance, integration, and power requirements. Select FPGA companies will be set to deliver products that offer breakthrough advantages in FPGAs that has not been seen before. To deliver products that can address the customer's needs in as many end applications possible, such as 400G packet processing, wireless remote radios units, data centers, and high-performance computing, will require a variety of tools and options. A product strategy that uses a tailored approach drawing upon using different process technologies, architectures, and integration options targeted to different applications will give hardware architects the best possible choices and solutions. Altera's Generation 10 products provide breakthroughs in capabilities and advantages across a variety of different applications with tailored portfolio of FPGAs and SoCs.

References

1. Cisco Visual Networking Index (VNI): Global Mobile Data Traffic Forecast Update, 2012 – 2017:
   www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
2. White Paper: *Designing Polyphase DPD Solutions with 28 nm FPGAs*:
   www.altera.com/literature/wp/wp-01171-polyphase-dpd.pdf
3. Gartner report, Market Trends: Worldwide, ASIC and ASSP Design Starts Continue Declining Trend, 2012
4. Altera website:
   www.altera.com
5. Altera's corporate presentation
6. Forbes website:
   www.forbes.com/sites/greatspeculations/2013/01/22/intels-difficult-year-and-whats-ahead/

Further Information

- White Paper: *Meeting the Performance and Power Imperative of the Zettabyte Era with Generation 10*:
  www.altera.com/literature/wp/wp-01200-power-performance-zettabyte-generation-10.pdf
- White Paper: *The Breakthrough Advantage for FPGAs with Tri-Gate Technology*:
  www.altera.com/literature/wp/wp-01201-fpga-tri-gate-technology.pdf
- White Paper: *Designing Polyphase DPD Solutions with 28 nm FPGAs*:
  www.altera.com/literature/wp/wp-01171-polyphase-dpd.pdf
- Innovations at 14 nm and 20 nm – the Next-Generation Advantage web page:
  www.altera.com/technology/system-tech/next-gen-technologies.html

- Stratix 10 FPGAs web page:
  www.altera.com/devices/fpga/stratix-fpgas/stratix10/stx10-index.jsp
- Arria 10 FPGAs web page:
  www.altera.com/devices/fpga/arria-fpgas/arria10/arr10-index.jsp

Document Revision History

Table 1 shows the revision history for this document.

Table 1. Document Revision History

Date	Version	Changes
July 2016	1.2	Update transceiver speed from 28 Gbps to 25.78 Gbps.
June 2015	1.1	Minor modifications throughout.
June 2013	1.0	Initial release.