FPGA Performance Benchmarking Methodology

Introduction

This paper presents a rigorous methodology for benchmarking the capabilities of an FPGA family. The goal of benchmarking is to compare the results for one FPGA family versus another over a variety of metrics. Since the FPGA industry does not conform to a standard benchmarking methodology, this white paper describes in detail the methodology employed by Altera and shows how a poor methodology can skew results and lead to false conclusions.

The complexity of today's designs together with the wealth of FPGA and computer-aided design (CAD) tool features available make benchmarking a difficult and expensive task. To obtain meaningful benchmarking results, a detailed understanding of the designs used in comparisons as well as an intimate knowledge of FPGA device features and CAD tools is required. A poor benchmarking process can easily result in inconclusive and, even worse, incorrect results. Altera invests significant resources to ensure the accuracy of its benchmarking results.

Altera benchmarking methodology has been endorsed by industry experts as a meaningful and accurate way to measure FPGAs' performance:

To evaluate its FPGAs, Altera has created a benchmarking methodology that fairly considers the intricacies and optimizations of its competitors' tools and devices, as well as its own. Experiments that consider a variety of end-user operating conditions have been run on a suite of industrial benchmark circuits. The results of these experiments have been analyzed to accentuate experimental variations and to clearly identify result trends. I am convinced that the experimental methodology that has been used fairly characterizes appropriate user expectations for Altera's devices in terms of area utilization and speed versus their competitors.

■   Russell Tessier, Associate Professor at the University of Massachusetts, Amherst

Altera has created an impressive methodology for measuring and monitoring key performance characteristics of their and competitors' devices and tools. Designing a fair and accurate test methodology that yields reasonable metrics is a daunting task, and Altera has clearly made a considerable investment that has paid off in this area. Their system includes a comprehensive and representative test suite, automatic compensation for technology and IP differences between various FPGA families, and automatic generation of design constraints to get optimal results from synthesis and layout tools.

■   Kevin Morris, Editor, FPGA and Programmable Logic Journal

Benchmarking Methodology

For benchmarking, Altera advocates three comparisons: least effort, timing-constrained, and best effort. The effort level is a reflection of both the user's effort (how much work the user must do) and the tool's effort (how much total compile time will be required) to obtain the results. For the purpose of this white paper, the focus is on the maximum frequency of operation (f_{MAX}) of the designs.

■   The purpose of the least effort experiments is to get a good result with little user intervention and a relatively small amount of compilation time. This is useful because it is simple to do (using the recommended out-of-the-box settings) and gives an initial estimate of performance.
■   The purpose of the timing-constrained comparison is to use default CAD tool settings, but ensure that the CAD tool is optimizing for performance.
■   The purpose of the best effort comparison is to give an indication of the best possible result achievable. The experiments for this benchmarking comparison require longer individual compilation times than in a default push-button compile and more than one compile per design. These performance results can be significantly better than results from the least effort compilation and the timing-constrained comparison.

Many factors can affect benchmarking results including design selection, CAD tool settings, user constraints, and interpretation of results. The remainder of this section provides a detailed discussion of the key factors that affect the benchmarking results, and show how Altera's methodology addresses these factors including:

■ Benchmark design selection
■ Hardware description language (HDL) optimization
■ Software settings and constraints
■ Interpretation of results-timing analysis differences between software tools
■ Benchmark results reporting

Benchmark Design Selection

Altera exclusively uses customer designs for benchmarking and does not generate artificial circuits or use IP cores in isolation. The designs used are collected from a variety of market segments, such as networking, telecommunications, wireless and consumer products, and a variety of implementation technologies such as ASICs, gate arrays, and FPGAs from other vendors. By benchmarking a broad suite of customer designs, Altera ensures that the logic and connectivity present in our benchmark designs are accurate and representative of the complex interaction of large designs and FPGA CAD tools.

In order to use customer designs, Altera invests significant resources in converting designs to work with various synthesis tools and CAD vendors. Altera also ensures that functionality is preserved and appropriate code optimizations for the specific FPGA vendor are made (necessary because designs are often developed such that they are optimized for a specific FPGA). For Altera benchmarking comparisons, all designs in the high-performance segment used are synthesized with the same synthesis tool, Synplify. The low-cost FPGA comparison includes the use of both Synplify and integrated synthesis tools from each FPGA vendor.

Hardware Description Language Optimization

As mentioned in the earlier section, due to the use of real customer designs in benchmarking, each design must be converted to work with FPGA architectures under test. A team of Altera engineers is dedicated to the design conversion and optimization process. To ensure a fair comparison, for each design the conversion and performance optimization process takes weeks to be completed such that each design can take full advantage of the dedicated features present in both Altera® and Xilinx FPGA architectures.

Software Settings and Constraints

All CAD tools offer settings that provide a trade-off among design performance, logic resource consumption, compile time, and memory usage. The settings that produce the best results for one design are likely not the best for another. In addition, results can be improved by providing user constraints to guide the CAD tool. Even with a design set that is representative of customer designs, the outcome of benchmarking will vary significantly with software settings and the constraints applied.

For the least effort comparison presented in this paper, each design is compiled out-of-the-box with default settings and no constraints. For the timing-constrained comparisons, default CAD settings are used, but aggressive timing constraints are set. To determine aggressive timing constraints for each design, a frequency (f_{MAX}) constraint is applied to each clock in a given design such that the constraint is just beyond what is achievable for each clock. The best constraint is determined by increasing the constraint until it cannot be met.

Software optimization settings can also make a big difference. To determine the best settings, the Quartus® II development software provides a feature called Design Space Explorer (DSE), which automatically produces results for a variety of settings for a particular design. In addition to the benefits of automatically picking the best settings, the Quartus II software provides advanced algorithms for physical synthesis. These advanced features are essential to get the highest performance for designs automatically. Xilinx ISE software now also provides a similar tool, called Xplorer. For the best effort comparison, aggressive timing constraints are set and the best available CAD tool setting

is used to maximize performance. For Altera, DSE and physical synthesis are used for best effort benchmarking. For Xilinx, all available Xplorer options are attempted.

Device selection is also critical in ensuring fair and accurate benchmarking results. The smallest device that a design will fit into is selected based on the logic resources and dedicated features availability, such as memory and multipliers that a design needs.

Interpreting Results-Timing Analysis Differences Between Software Tools

Interpretation of results is as important as running the experiment correctly. Altera reports average performance results across all clock domains. However, blindly comparing the results from Quartus II and ISE software without paying detailed attention to the differences in timing analysis in each tool can easily produce misleading results.

The differences are in the assumptions made for timing analysis and in the interpretation of timing constraints. These aspects of the tools can have a significant impact on the reported timing numbers. By default, Quartus II timing analysis makes conservative assumptions and analyzes the design to give the user the most information. The ISE trace timing analyzer reports only what is constrained explicitly by the user; unconstrained paths are not reported. Therefore, without a good understanding of which paths the timing analysis tool is considering, one can arrive at a wrong conclusion.

Altera has invested significant resources to ensure that equivalent timing analysis is performed between the Quartus II and ISE tools in our benchmarking experiments. A summary of the differences between the tools is provided in Table 1.

> A detailed analysis, along with examples and how to configure the tools to perform equivalent timing analysis, can be found in the *Performing Equivalent Timing Analysis Between the Altera Quartus II Software and Xilinx ISE* white paper: **www.altera.com/literature/wp/wp_timingAnalysis.pdf**.

Table 1. Timing Analysis Issues in Xilinx ISE

| Design Structure	Xilinx ISE	Altera Quartus II Software
Registered Clock	The relationship between clocks is not inferred causing many legitimate paths to be ignored. If explicitly constrained, the output of a register will be treated as a separate clock domain and associated paths are then analyzed.	The relationship between clocks is inferred and the paths between them analyzed. The user can provide explicit constraints to ignore paths between clock domains if desired.
Derived (Gated) Clock	Analyzed only when explicitly constrained.	The relationship between clocks is inferred and the paths between them analyzed. Each derived clock can be treated as a separate clock domain with constraints.
Designs With Digital Clock Managers (DCMs) or Phase-Lock Loops (PLLs)	Analyzed if the constraint is applied to the input of a DCM.	Analyzed based on PLL clock settings made in the PLL MegaWizard® Plug-In.
I/Os: Setup and Clock-to-Out Times	Registered clock preceding the input or output register is not analyzed.	All worst-case structures analyzed.
Combinational Loop	Not analyzed. Warning reported.	Analyzed by default, but can be cut by the user if desired.

Benchmark Results Reporting

The data points used for benchmarking results reporting can significantly affect the conclusion. To illustrate this, the results of a benchmarking experiment are given in Figure 1. As the figure shows, it is typical to have a distribution of results due to differences in FPGA architectures and CAD tools. Some designs are implemented particularly well with a specific architecture and CAD tool while others are not. It is easy to see that by using a subset of designs, many different but incorrect conclusions can be produced.

Figure 1 also shows how a single example can be quite unreliable in representing overall performance advantage, as the results can range from a big win to a big loss. One must use a large number of designs to understand the real performance of a device and CAD tool. Given the amount of variation present in individual design results, an average

result over the whole design set, 73 designs, gives an indication of performance. Figure 1 illustrates how using just the top 50 designs can produce a different average result from the full set. For the equivalent Altera and Xilinx device comparison, the top 50 designs show an average 9.2 percent performance advantage, while the whole set yields a gain of 1.9 percent.

Figure 1. Cherry Picking Results is Dangerous
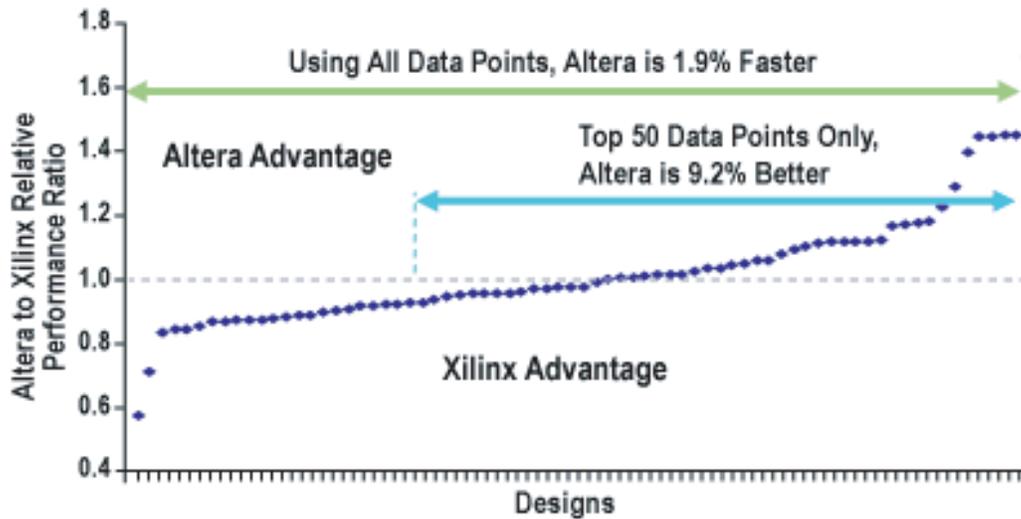
Note:
(1) The benchmarking data is based on comparing an Altera device to an equivalent Xilinx device using the latest available Quartus and ISE software at the time of analysis.

How can the user know which results are being shown and whether the results are relevant to a particular design situation? Unfortunately, the answer is that it is hard to know for two reasons. First, every FPGA vendor has its own unique set of designs for benchmarking. Second, disclosing information on designs is not possible because these designs are often real customer designs and are typically protected under nondisclosure agreements.

Conclusions

This paper demonstrates that FPGA performance comparisons depend strongly on the benchmarking methodology. Specifically, the factors that can have a large effect on results include the designs used, settings and constraints made, and differences in timing analysis. Altera Quartus II software provides a simple, automated mechanism to allow designers to obtain the best performance for their designs. In addition, Quartus II software performs detailed timing analysis by default to minimize the risk of design failures due to improper timing analysis.

When considering benchmarking results, designers should weigh all of the considerations described in this white paper as well as past experiences with tools and the reputation of the company presenting results. Our benchmark results show that for high-density FPGAs at both the 65-nm and 90-nm process nodes, Altera family FPGAs are *at least* one full speed grade faster than any competitive offering. The Altera Stratix® III family is a full speed grade faster than the Xilinx Virtex-5 family, and the Altera Stratix II family is a full speed grade faster than the Xilinx Virtex-4 family. For low-cost FPGAs, the Altera 90-nm Cyclone® II family provides an average 60 percent higher performance than the Xilinx 90-nm Spartan-3 family.

Altera's benchmarking methodology has been endorsed by third-party industry expert as a reliable and accurate way to determine FPGA performance from the same vendor and between competitive solutions.

Further Information

- *Stratix II Device Performance and Logic Efficiency Analysis:*
  www.altera.com/literature/wp/wpstxiiple.pdf
- *Stratix vs. Virtex-II Pro PFGA Performance Analysis:*
  www.altera.com/literature/wp/wpstxvrtxII.pdf
- Xilinx white paper, WP226, *Cyclone vs. Spartan-3 Performance Analysis*
- Xilinx white paper, WP206, *The 40% Performance Advantage of Virtex-II Pro FPGAs Over Competitive PLDs*, version 1.2, March 1, 2004