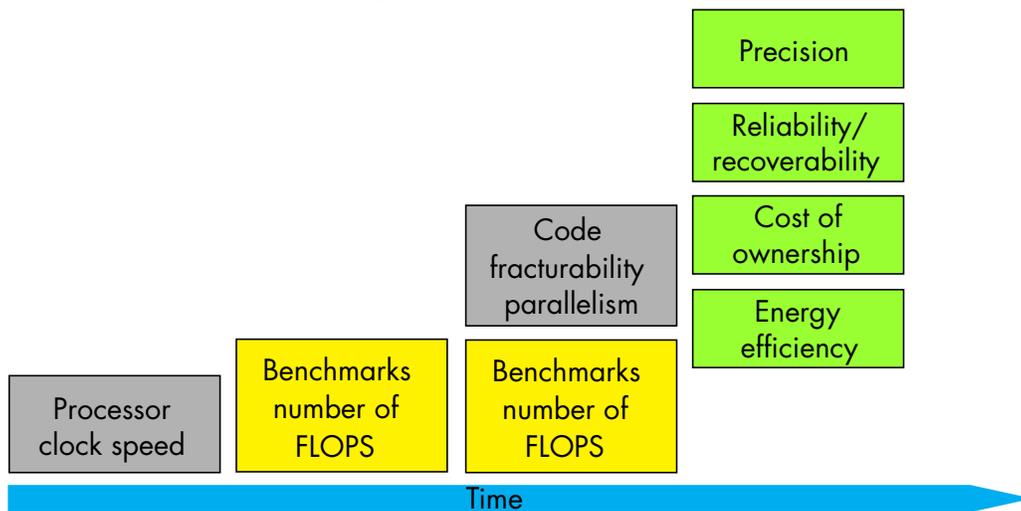


Assessing FPGA DSP Benchmarks at 40 nm

Introduction

Benchmarking the performance of algorithms, devices, and programming methodologies is a well-worn topic among developers and research of high-performance computing appliances. Independent and partisan benchmark results are a difficult thing to navigate and understand. As shown in [Figure 1](#), benchmarking has evolved from looking just at clock speed, to encompassing operations per second, and now including total cost of ownership and energy efficiency. Because each methodology is based on a set of assumptions, designers and developers are forced to critique each with a focus on the first two stages of scientific inquiry. Understanding and using reliable device benchmarks for system decisions is essential for scientific and military applications. While individual benchmark results may be insufficient for making device decisions on a standalone basis, it is important to understand the trends and factors that make up the results of benchmarking studies.

Figure 1. Brief Evolution of Benchmarking Considerations



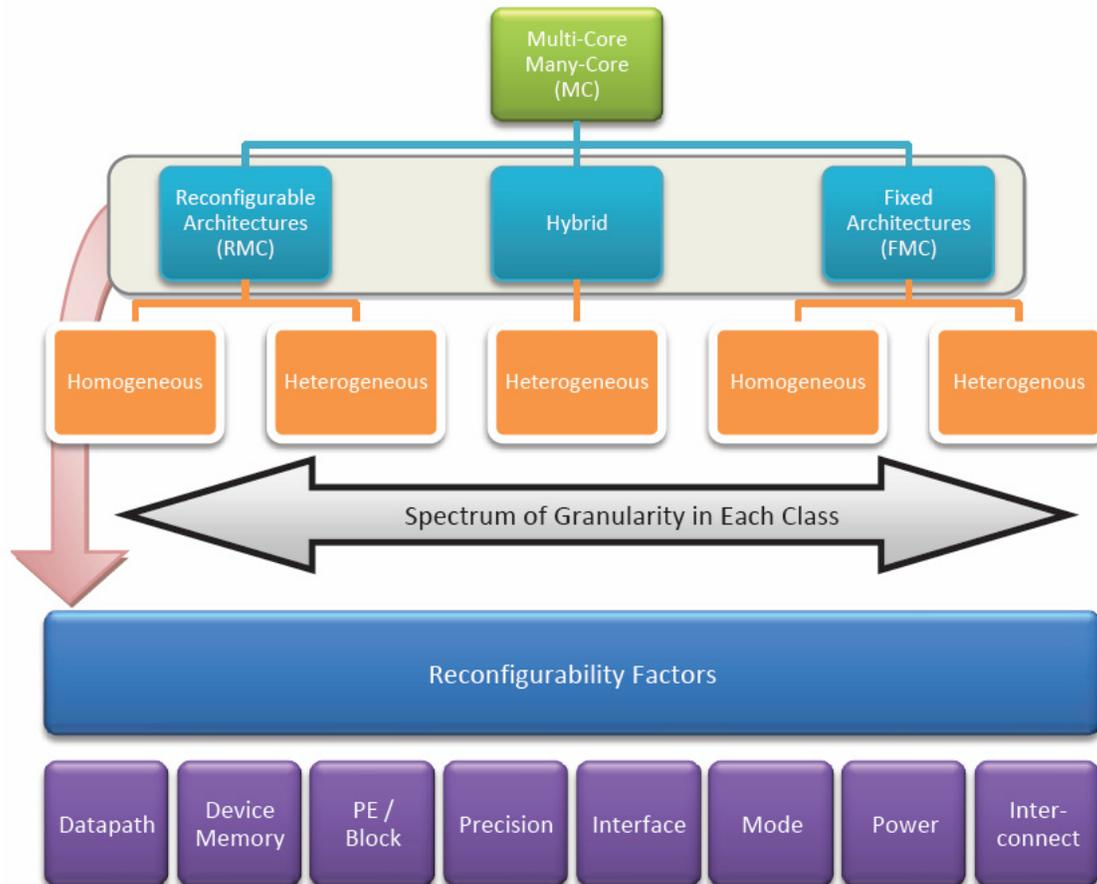
This white paper takes a brief look at a set of benchmarks for high-performance computing devices, focusing primarily on a study developed by the National Science Foundation (NSF)'s Center for High-Performance Reconfigurable Computing (CHREC). The object of examining these benchmarks is to help the designer decide which conditions or benchmark figures are most important for the design. They can also assist in determining whether Altera® 40-nm Stratix® IV FPGAs are appropriate engines for a high-performance computing solution, such as military sensor or molecular dynamics model.

Design flexibility, while not a performance benchmark, not surprisingly is a key characteristic that propels FPGA solutions. The usefulness of design flexibility for designers of high-performance computing applications lies not only in algorithm and memory utilization, but also the order in which parts of the computing devices are “powered up” in each stage of a high-throughput application.

Solution Space for High-Performance Computing

As the CHREC study explains, solutions for high-performance applications can be roughly divided into fixed multicore (FMC) and reconfigurable multicore (RMC) devices ([Figure 2](#)).

Figure 2. CHREC Taxonomy of Multicore Devices for Benchmarking Categories



FMC devices are a host of solutions in which the computing cores are fixed in the device and the programming model revolves around these processing units. Solutions in the FMC space include multicore Opteron, Athlon, Atom and Xeon processors, multicore systems such as Cell Broadband engine, Clearspeed devices, Freescale Power PCs, and graphical processors like the NVidia Tesla.

RMC devices are a class of systems where the distribution of resources is algorithm dependent. Solutions in the RMC space include field-programmable gate arrays (FPGAs), field-programmable object arrays (FPOAs), and reconfigurable processors such as TILE64 and Raytheon's Monarch chip.

Results for Reconfigurable Computing Devices

The CHREC study breaks out the benchmarks into RMC versus FMC devices, and offers an overall summary of computational density per watt. It then breaks down measurements based on bit width, where there is the highest variation in performance based on device architecture. For FMC devices, this architecture difference is based on the data pathways, while for RMC devices it is based on the hard DSP blocks. The DSP/memory ratio is also taken into account. The primary result of this study is the considerable difference in computational density per watt between FMC devices (not covered in this white paper) and RMC devices (FMC results are not covered in this white paper). Reconfigurable devices show nearly an order of magnitude more capability when factoring energy efficiency as a core benchmark variable.

Within the category of reconfigurable devices, FPGAs consistently show the best results using the CHREC benchmarks. A variety of devices from both Altera and Xilinx are assessed, with the overall parallel operation advantage going to the newest 40-nm FPGAs.

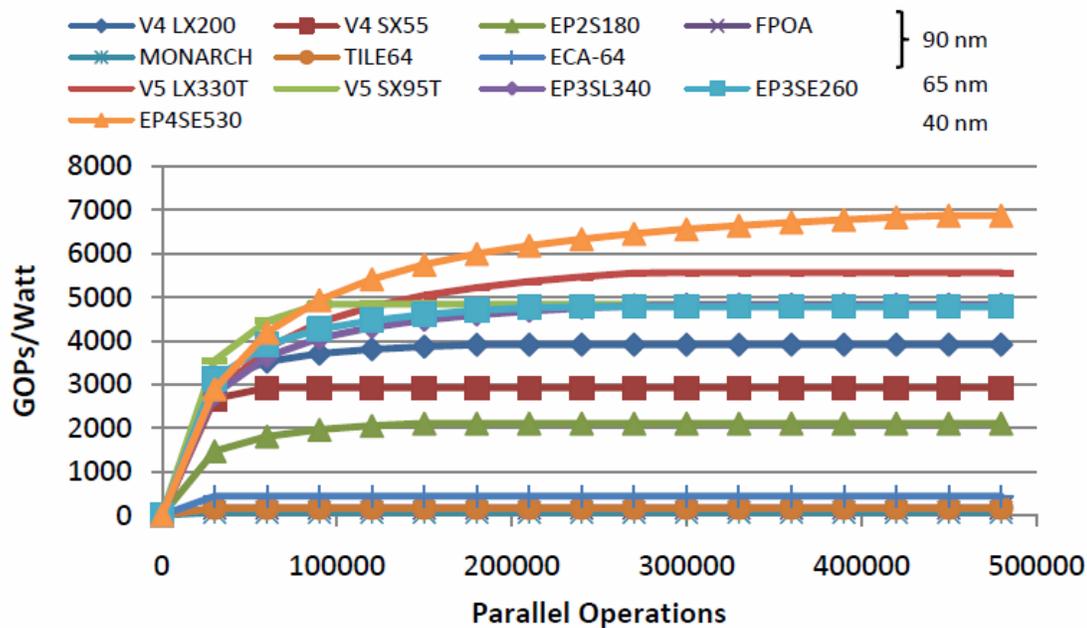
Accounting for Power in Designs and Benchmarks

Brian Halla, CEO of National Semiconductor, is the most recent industry powerbroker to remind us that “Moore’s Law never mentioned anything about power efficiency.” Yet a great deal of emphasis is placed on power efficiency in nearly all benchmarks available today. One very obvious reason for this is a general marketing drive towards “green” computing and energy cost-efficiency. However, there are several other drivers that influence this category in the high-performance market.

Smaller form factors are key in developing scalable applications and mobile supercomputing. Power and heat dissipation are often the primary drivers in developing these smaller form factors. In addition, the precision and clock stability of a high-performance computing application is affected by temperature, with less power-efficient architectures giving less reliable results and a greater number of failure conditions.

A key result in most benchmarks favoring RMC applications is power efficiency. This may seem counterintuitive, as circuits that are not optimized for an algorithm rarely perform as well as customized hardware. However, a key conclusion in the CHREC study is that today’s RMC technology allows the user to power-down unused portions of logic or reconfigure power biasing based on the characteristics of an algorithm. Figure 3 depicts the ratio of gigaoperations (GOPs) per second to power for different reconfigurable devices.

Figure 3. Computational Density Per Watt (GigaOperations), Reconfigurable Devices

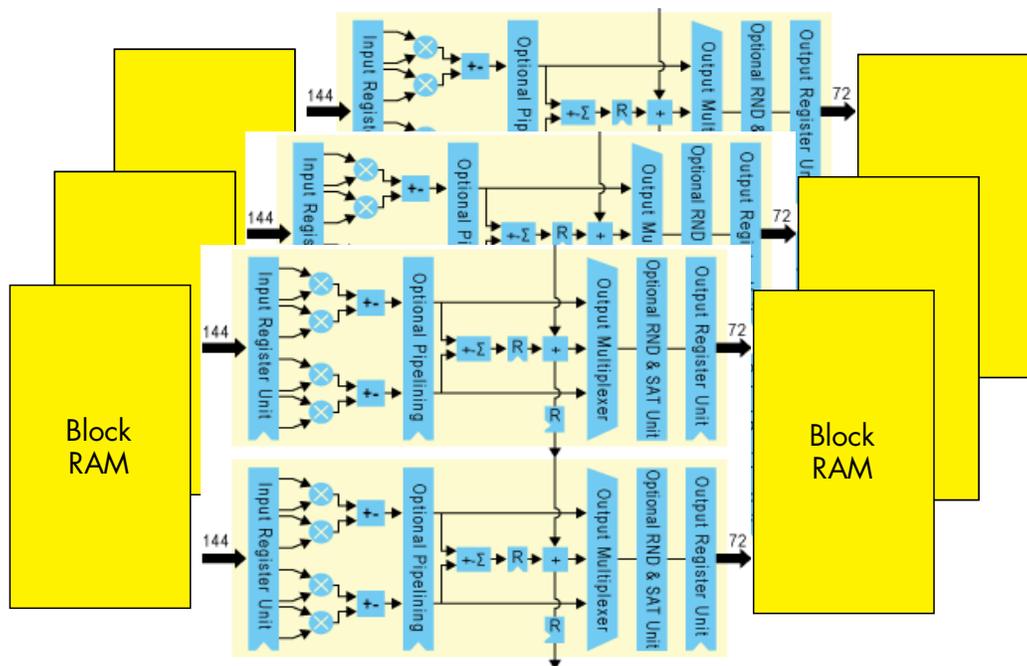


An example of this is Altera’s Programmable Power Technology, which allows users to restrict power to or limit the bias power on circuits that are unused or not in the critical performance path of the algorithm. These adjustments are made automatically in Altera’s Quartus® II design software, with additional power optimizations possible through guided application notes.

Chip Memory Bottlenecks

Another important result from these benchmarks is identifying the architecture-based bottlenecks from various high-performance devices. In FPGAs, these bottlenecks include “memory sustainability,” where the number of parallel operations is limited by the number of on-chip memories required to access the inputs and outputs of the DSP blocks. When there is not enough memory on the device to utilize its full DSP capability, then it cannot be used to its full potential. Figure 4 shows that a lot of block memory bandwidth is required to use many parallel double-precision floating-point operations.

Figure 4. FPGA Block RAM Availability Is a Large Factor in Highly Parallel DSP Operations—DSP Needs a Large Number of Block RAMs



CHREC noted that Altera devices fared better in many categories because:

“Altera FPGAs show very good performance since they are not as limited by memory-sustainability issues as other FPGA vendor devices... The Altera FPGAs have a better balance between the number of parallel operations and the number of memory locations, and thus they do not exhibit this limitation.”

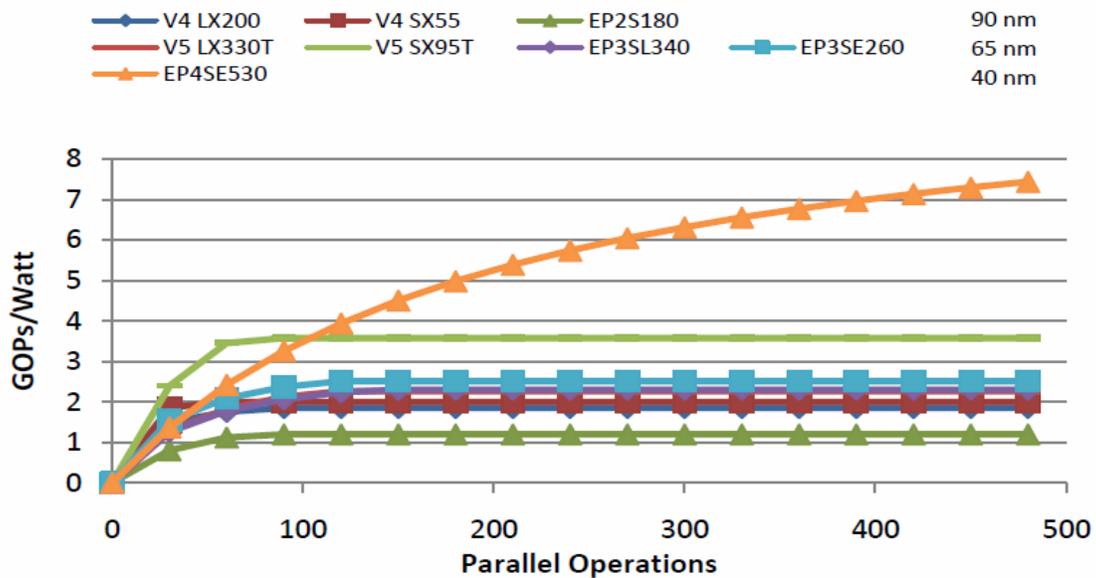
Floating-Point Operation Results and Military Applications

Military radar applications increasingly require algorithms with floating point precision. A multi-element radar may require waveform generation, filtering, matrix operations, and signal correlations where some floating point is required to achieve the best resolution. Synthetic aperture radar in airborne applications can also require floating point resolution in the algorithms which combine the various observations. Various military applications require single or double floating-point precision in finite impulse response (FIR), fast Fourier transform (FFT), and matrix math operations.

Extending the bit-width requirements of DSP calculations becomes a guiding requirement on selection of FMC hardware, and can lead to highly variable results for RMC devices. In general, FPGAs have highly desirable features for use in mixed use and many floating-point operations. Only FPGAs can implement mixed datapaths so efficiently, especially for late design changes in bit-width partitioning.

As mentioned previously, differing RMC architectures lead to differing results at different bit widths. FPGAs did consistently well across 16- and 32-bit integer, as well as single-precision and double-precision floating-point operations. Altera and Xilinx FPGAs alternately showed superiority depending on the operation bit width and the efficiency of that operation within its DSP block architecture. Figure 5 shows the results for the most computationally difficult category, double-precision floating point parallel operations per watt, which was dominated by Altera’s 40-nm Stratix FPGAs.

Figure 5. Double-Precision Floating-Point Computational Density Per Watt (GigaOperations), Reconfigurable Devices



Altera has made a Floating Point Compiler tool available to CHREC for their on-going benchmarking studies and activities. More information about this compiler, and how it can improve the benchmark performance of a floating-point application, is described in Altera's white paper, *Floating Point Compiler: Increasing Performance With Fewer Resources*.

Summary

Every high-performance application necessarily has its own set of benchmarks, to include cost and the ability to design, maintain, upgrade, and scale. The flexibility of FPGAs gives them a decisive advantage in most of these categories, leaving computational speed and density as the primary design criteria. Independent studies, such as this effort by CHREC, increasingly are finding high-density FPGAs to be credible and intelligent solutions for reconfigurable high-performance computing, with further accelerations possible with iterative design improvements. The extension of these solutions into mixed use and double-precision floating-point operations may occur very rapidly over the next three years.

Further Information

- J. Williams, A. George, J. Richardson, K. Gosrani, and S. Suresh, “Characterization of Fixed and Reconfigurable Multi-Core Devices for Application Acceleration,” *ACM Transactions on Computational Logic*, National Science Foundation Center for High-Performance Reconfigurable Computing (CHREC), University of Florida
- Altera’s Floating Point Megafunctions:
www.altera.com/products/ip/dsp/arithmetic/m-alt-float-point.html
- *Floating Point Compiler: Increasing Performance With Fewer Resources:*
www.altera.com/literature/wp/wp-01050-Floating-Point-Compiler-Increasing-Performance-With-Fewer-Resources.pdf
- *Using Altera FPGAs for Double Precision Floating Point:*
www.altera.com/literature/wp/wp-01028.pdf
- *FPGA Vendor Selection Criteria for Military Programs:*
www.altera.com/literature/wp/wp-01094-select-military-vendor.pdf
- Benchmarking Methodology:
www.altera.com/products/devices/performance/benchmark/per-benchmarkmeth.html
- *FPGA Performance Benchmarking Methodology:*
www.altera.com/literature/wp/wpfpgapbm.pdf
- *Guidance for Accurately Benchmarking FPGAs:*
www.altera.com/literature/wp/wp-01040.pdf
- Open Cores Organization IP Benchmarking Tools:
www.opencores.org

Acknowledgements

- Michael Strickland, Sr. Manager, High-Performance Computing Business Unit, Altera Corporation



101 Innovation Drive
San Jose, CA 95134
www.altera.com

Copyright © 2009 Altera Corporation. All rights reserved. Altera, The Programmable Solutions Company, the stylized Altera logo, specific device designations, and all other words and logos that are identified as trademarks and/or service marks are, unless noted otherwise, the trademarks and service marks of Altera Corporation in the U.S. and other countries. All other product or service names are the property of their respective holders. Altera products are protected under numerous U.S. and foreign patents and pending applications, maskwork rights, and copyrights. Altera warrants performance of its semiconductor products to current specifications in accordance with Altera's standard warranty, but reserves the right to make changes to any products and services at any time without notice. Altera assumes no responsibility or liability arising out of the application or use of any information, product, or service described herein except as expressly agreed to in writing by Altera Corporation. Altera customers are advised to obtain the latest version of device specifications before relying on any published information and before placing orders for products or services.