

Design Support for Reconfigurable Accelerators on future Cray Systems

Clay Marr, *DRC Computer*

ABSTRACT: *Cray and DRC announced a relationship several months ago in which DRC would be providing FPGA acceleration technology for some of Cray's future systems. DRC is now providing a development platform to allow Cray's customers to write or port applications to run on these accelerators and test them on this relatively inexpensive workstation.*

The DS/XT, development system is designed to emulate the acceleration node in a Cray system as closely as possible. This system will enable Cray's end customers to begin development early and fully debug accelerated algorithms or applications without consuming cycles of the main system once installed.

This presentation will discuss some of these details as well as describe the balance of the tools and options that provide a complete stand-alone development environment.

KEYWORDS: DS/XT, Accelerators, Reconfigurable Co-processors, RPU

1. Introduction

DRC

DRC Computer Corp. is a 3 year old company, founded to provide reconfigurable computing platforms. With the advent of larger FPGAs and an open CPU-to-CPU bus standard like AMD's HyperTransport™ bus, the notion of a tightly-coupled co-processor using reconfigurable FPGAs became feasible.

After raising some seed funding, DRC developed a prototype RPU (Re-configurable Processing Unit) that plugs directly into an Opteron socket on a 2-way server motherboard. Interfacing directly to the HT bus, the RPU takes advantage of all the motherboard resources (memory, CPU-CPU communication, I/O etc.) in exactly the same tightly coupled way as a second Opteron processor would providing the minimum latency, maximum bus and memory bandwidth and memory size possible.

With some subsequent venture funding, the company began shipping its first commercial products in October of 2006. These products included both the RPU and design systems for developing applications to run on the RPU.

Cray and DRC Agreement

In May of 2006, Cray and DRC entered into an agreement to adapt DRC's RPUs to a new family of Cray Supercomputers. DRC was chosen for advantages cited above as well as the ease with which the DRC RPU could be integrated into the new Cray architecture, which is based on AMD's Opteron™ processors and Hyper Transport bus.

Under the agreement, DRC will provide RPUs on an OEM basis to Cray. Cray will integrate the RPUs in systems according to customer demand. DRC will, in addition, provide design environments for the end-customer to enable the development of applications without the need to access the main supercomputer platform.

2. Acceleration potential: Commercial application examples

Seismic 3D imaging

In this real example, creating the 3D image in Figure 1 was created using standard microprocessors with a couple of hundred nodes, each consisting of a 2-way server with dual or quad-core Opterons plus memory and disk. Generating the final image requires processing,

analysing and iterating over terabytes of data dozens of times. Figure 1 took several weeks to generate on such a setup.

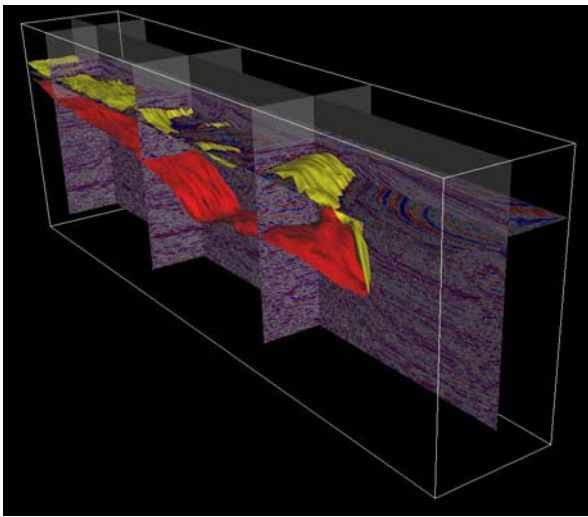


Figure 1. 3D image of sub-terrainian structures

There are several algorithms used in seismic imaging. One such algorithm used by a few DRC customers is a “star” finite differences algorithm. In this algorithm, a “star” matrix representing either a 2-D or 3-D set of input vectors is convolved with a constants vector, generating a results matrix. This results matrix is stored to disk for further image processing. Figure 2 shows an example of the 3D “star” finite differences stencil used in this application.

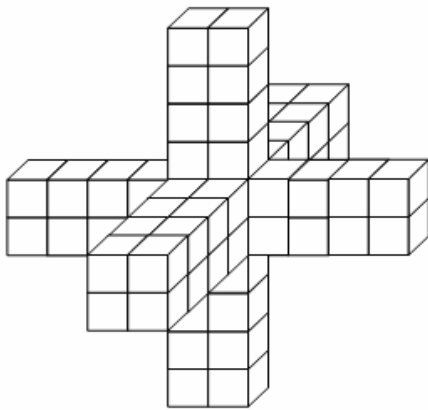


Figure 2. 3D “star” finite differences stencil.

In this application, the core floating point logic implemented in the RPU is capable of processing data well beyond the dataflow rate of other accelerator solutions. Using the largest FPGA from Xilinx and multiple memory controllers, the RPU co-processor is

capable of simultaneous access to both the motherboard and on-board RPU memories at data rates of up to 14.4GB/s. DRC and its development partners have been able to demonstrate acceleration factors of 9-14 times when compared to the performance of the core algorithm running on dual or quad-core Opteron.

In addition to the core algorithm, other application tasks continue to run on the Opteron on each node. The ultimate measure of success is measured by wall-clock time comparisons between the accelerated and non-accelerated application execution times. In these examples, wall-clock time from beginning to end of the application run is 4-5 times faster with the RPU co-processor.

Another way to realize the value of application acceleration is to reduce the cost of generating the images. In this example, the RPU system results in:

- ▶ 5X node count reduction,
- ▶ 85% saving in utility bill, and
- ▶ Less than half the cost.

Other algorithms

Other algorithms exhibit significant acceleration as well. Without going into technical detail, here are a few of the results:

Smith Waterman:

50-60X (compared to single core), (ref. to paper by O. Storaasli, W. Yu, J. Maltby, and D. Strenski)

Euro-Option pricing model using Monte Carlo:

41X (compared to single core)

(Ref: Design Space Exploration of the European Option Benchmark using Hyperstreams
G. Morris and M. Aubury)

3. Standard Technology

Maximum Leverage/Minimum risk

Cray and DRC

One of Cray’s key business requirements was to find a reconfigurable technology that leverages the investment being made by the industry, and maintains the strategic technology, architecture and implementation of its own system design.

An accelerator device that fits into an Opteron socket allows for the following kind of architecture:

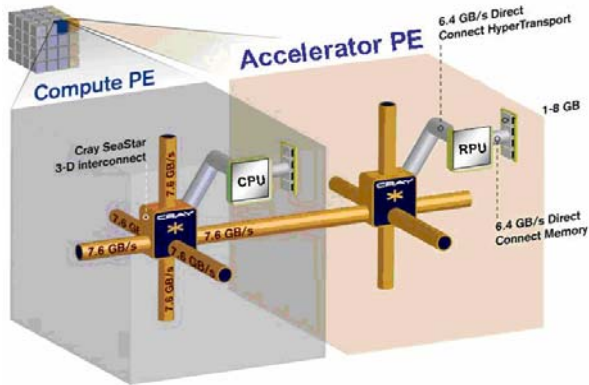


Figure 3. CPUs and RPUs - peers on a fast interconnect

The architecture in Figure 3 shows the use of different types of processing engines configured to provide the types of processing required using Cray's high-speed interconnect to maximum advantage. The result is high bus bandwidth as well as lowest possible latency.

In Figure 4, you can see how the Cray XT architecture allows for a heterogeneous mix of Compute Nodes, Service Nodes and Accelerator Nodes in a 3D array. This allow applications to be written and executed in the most optimum way to achieve the highest possible performance by taking advantage of serial and highly parallel processing paths as needed.

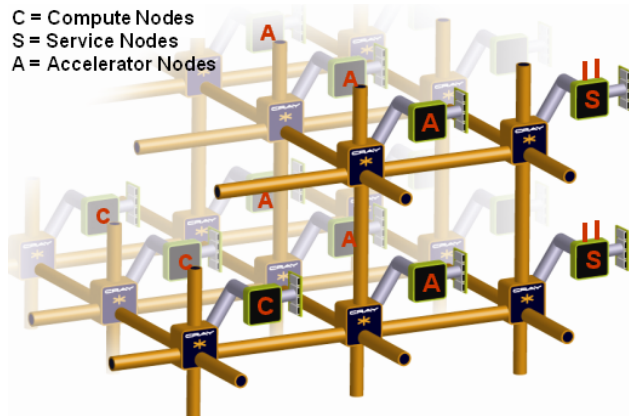


Figure 4. XT System Configuration

In the next figure, we drill down further and look at the internal architecture of the Accelerator Processing Engine.

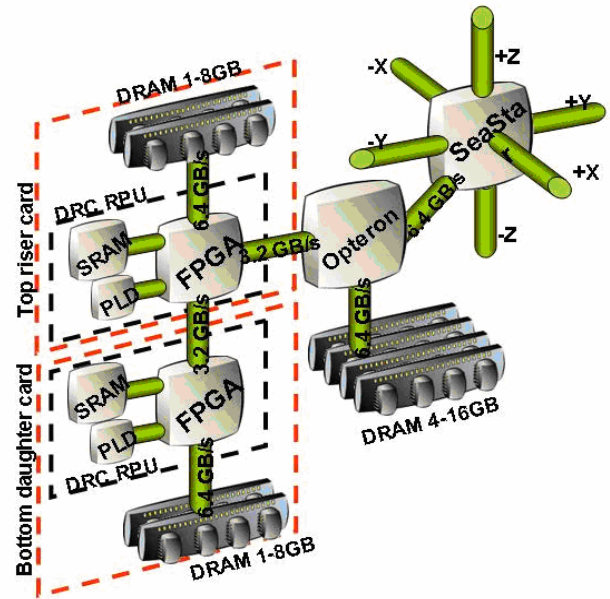


Figure 5. Accelerator Node Detail

At this level, the Reconfigurable Computing Node (also called the Accelerator PE) looks like an industry standard Opteron motherboard with a connection to the SeaStar fabric. The Opteron handles MPI traffic across the fabric as well as tasks assigned to it within the reconfigurable node.

Standard HT buses are used to interconnect the Opteron to the SeaStar and the Opteron to the first FPGAs and between the first and second FPGA. The sockets are all industry standard as are the memories and memory interfaces.

The significance of this approach lies in the leverage achieved from the use of industry standards:

- ▶ Provides a standard programming environment
- ▶ Leverages world-class processors, memories, and motherboard technology
- ▶ Uses world-class FPGAs
- ▶ Enables quick response to standards changes:
 - Upgradeable software and flexible hardware
 - High performance
 - Fast time to market

4. Design Systems

One critical factor in using reconfigurable computing nodes within a Cray system is the ability to develop and test the software design to run on the accelerator engines. In most cases, the software development is a separate task and is performed asynchronously or in parallel to the overall application development.

Hardware configurations are identical

DRC has a system to support the development of code for the accelerator.



Figure 6. DS/XT Development Systems

The DS/XT development system is designed to emulate the reconfigurable computing node in a Cray system as closely as possible. It provides the same architecture in hardware and software, the same CPU/RPU communication and node-to-node communication environment as the eventual system implementation.

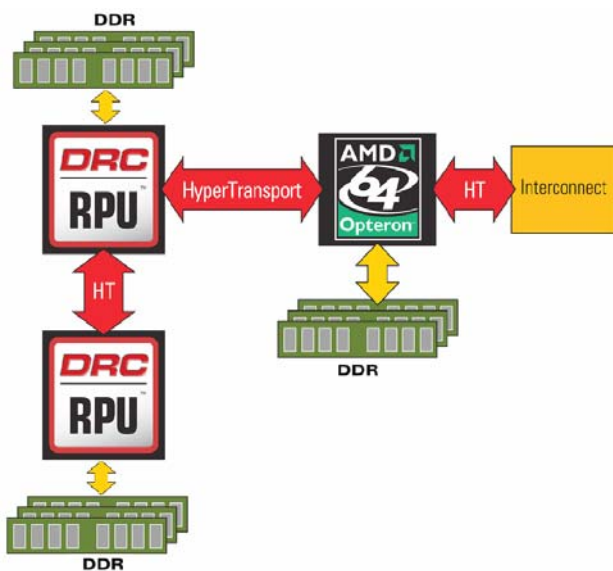


Figure 7. DS/XT Acceleration Node Detail

Notice, for example, the HT interconnects between the Opteron and the RPU and again between the 2 RPUs is an exact copy of the Cray Accelerator node. Memories of the same type and size are attached to the Opteron as well as the RPUs in both configurations in exactly the same way.

The Interconnect block on the DS/XT can link to another DS/XT via 1GB Ethernet or Infiniband and represent the same connectivity as the SeaStar fabric. For software development, the Opteron would use MPI to connect nodes in very much the same way.

Software and service eases design effort

DRC's RPSysCore

DRC provides a programming interface within the RPU to make development and reconfiguration simple for the user-defined logic. The RPSysCore programming interface, abstracts many of the most difficult issues in programming an FPGA. The RPSysCore API allows user logic to make simple read/write commands to any given memory location and not worry about any of the physical I/O pins, placement or timing for any external interface.

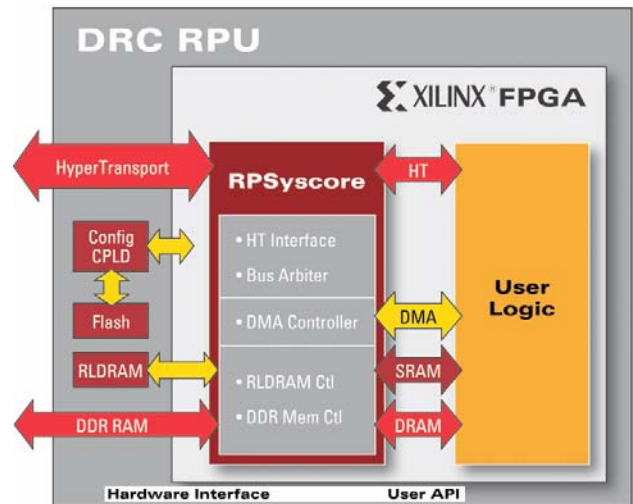


Figure 8. RPSysCore programming interface
Abstracts many of the most difficult programming issues

Software Compilers

The Xilinx ISE tool set is included with the DS/XT. For experienced hardware or FPGA designers, this can represent the entire tool set and design flow for application development. It contains synthesis, mapping, place and route, and bit generation tools, in addition to an HDL simulator. For in-circuit debugging, ChipScope Pro

is available from Xilinx, providing experienced hardware designers with a virtual logic analyser inside of the RPU. For improved performance and efficiency, Synplicity offers a high performance synthesis engine (Synplify Pro) and HDL back-annotated debugging tool (Identify) for the RPU.

For developers starting with C or other higher level software tools, a variety of software compilation tools are also available. For example, compilers for converting C to RTL are currently available from these vendors:

- Celoxica www.celoxica.com
- DSPlogic www.dsplogic.com
- Impulse C www.impulsec.com
- Mitrion www.mitrionics.com

Each of these tools has its own set of features and advantages. Visit the websites listed above for more information.

Optional limited licenses for these software tools are available for introductory prices, providing users an inexpensive way to try them. They can be ordered with the design system or separately from the 3rd party.

Software Development Services

All of the vendors mentioned above also provide consulting services to assist end-users develop their applications, algorithms and port those to the RPU platform. In addition to these vendors, there are additional service providers, such as:

Synective (Sweden)
XISS (Houston)
XLBiosim (Switzerland)

Typical Project Flow

A number of design projects have been completed or are in progress using the DRC RPU and development tools and systems. In general, there are 3 phases to those projects.

1. Analysis and design consists primarily of architecting the application for fine-grained parallelism. There are few tools today which do this automatically. It can take 2-3 months to get this done, unless the application was previously architected for multi-threading. The remainder of this phase generally involves choosing an implantation strategy, choosing 3rd party service providers, tools etc. and putting a project plan and budget in place.

2. Prototyping consists of simulating the functions for size and performance, designing the firmware and bringing up a prototype system. This phase can take 2-4 months and 3-6 man-months or more depending on the complexity and amount of optimization desired.
3. Production is a final phase of validation and installation of the application on the main computer system. This can take an additional 1-2 months and 1-4 man months of effort.

5. Usability

Documentation

Documentation of the DS/XT includes:

- Quick Start Guide
- Descriptions of the
 - RPU
 - RPSysCore API (H/W)
 - RPU API (S/W)
- Tutorial Example
- Glossary of terms
- CD containing:
 - Latest Release
 - RPSysCore
 - RPU API
 - Example

Password protected website

- For registered users
- Application examples

Vendor support

The DS/XT comes with a 90 day warranty and one year optional maintenance for both hardware and software. The maintenance consists of remote support for technical problems by phone or email, hardware replacements next day as required and software updates and fixes as released. Major releases that are separately priced are not included. The maintenance package also includes up to 30 hours of application support, again by phone or email. Phone support is available 8:00 am to 5:00 pm Pacific Time, Monday through Friday.

Support for tools will be provided directly by the tools vendor once a software license is registered with them.

7. Conclusion

Several customers have told DRC they need two things.

1. A design environment dedicated to the design and test of applications and algorithms that will run on a Cray Supercomputer. The design system should be complete and stand-alone from the main computer. The developers are often reluctant to put their work-in-process on the main system until it is reasonably complete and debugged.
2. The design environment needs to support the best available tools and methodologies for creating the algorithms in hardware (FPGAs).

The DRC DS/XT is available now - allowing Cray's end customers to begin development early. The system is expandable to support continuing development as the applications grow and proliferate.

DRC's RPSysCore is a major step in simplifying the reconfigurable co-processing design process. The tools partners that support the DS/XT are the leaders in the industry, giving the end-customer his choice of the best tools available.

Acknowledgments

DRC would like to thank Cray and DRC's vendors and partners for their support and participation in putting this complete environment together, keeping the user experience foremost in their consideration.

About the Author

Clay Marr is the VP of Sales and Marketing for DRC Computer. He has been involved in high-performance computing solutions dating from the mid-1970s with IBM plug-compatible mainframes with a division of Philips. He has also been involved in reconfigurable computing from the mid-1980s when he co-founded Plus Logic, an early player in CPLDs (eventually purchased by Xilinx).