



Filling the Need for Speed

*Giang Le
Product Marketing Manager
DRC Computer Corp., Sunnyvale, CA
August, 2007*

The “I want it now” attitude permeates almost every aspect of everyday life and business, since time has become a precious commodity. In all aspects of business, from computing 3-D images of underground oil reserves from data collected by seismic testing, to searching huge databases, to rendering images in movies or medical scanners—doing more in less time has tremendous paybacks. Thus the search goes on to find ways to cut days, hours and even minutes from the time it takes to do the large number of computations needed for each application.

Today, most compute-acceleration solutions depend on increasing the clock speed or increasing the number of general-purpose CPUs used to perform the calculations. Servers or server blades, based on dual, quad, or more processors per server are set up in row after row of racks that fill large air conditioned rooms. However, the high power consumption of general-purpose processors such as the x86, Itanium, Opteron, Power 5, UltraSparc, etc.(often 60 to 120 Watts/chip), as well as the overhead of space needed to hold the systems and the air conditioning cost of keeping all the systems cool, can end up limiting the number of processors that can be assembled into a compute solution.

One alternative offered by some companies takes advantage of ASIC technology to craft custom chips that can accelerate the time-critical computations and thus deliver a 10X to 100X acceleration for specific algorithms. These ASICs can also keep the power consumption under control by replacing multiple servers or at least providing an alternative to adding more servers to the server farm.

However, such ASICs have limited flexibility since their architectures are “frozen” by the silicon implementation, and that can restrict the ability of the ASIC to accelerate updated versions of the algorithms or handle new algorithms. Additionally, it typically takes 12 to 18 months to design and start manufacturing an ASIC, as well as a significant investment in design tools and resources. Such ASICs are not commodity products and thus the cost per chip will be relatively expensive since the ASIC may only be used in the tens of thousands of units vs the million-plus quantities that CPUs are typically sold in.

A relatively new alternative to the custom ASIC solution leverages the latest multi-megagate field-programmable gate arrays (FPGAs). The FPGAs can be configured via downloadable bit streams and thus the compute-acceleration architecture can be optimized for each algorithm just by downloading a new bitstream. This gives users the best of both worlds – a flexible hardware acceleration solution using off-the-shelf chips that can be configured via software to deliver an optimized compute solution. The downloadable configuration pattern can also be updated, thus allowing the FPGA to “upgrade” its performance as algorithm enhancements improve the efficiency of the algorithm and new algorithms are added to handle new tasks.

Additionally, since the FPGAs are off-the-shelf components, development time is much shorter than that of an ASIC. Furthermore, the opening of the Opteron CPU interface by Advanced Micro Devices Inc. through its Torrenza platform further simplifies the integration of accelerators right onto a multi-CPU motherboard. By leveraging the high-speed HyperTransport channels and the Opteron processor interface, DRC Computer has created a reconfigurable processing unit (RPU) that drops right into an empty Opteron socket on the motherboard or blade server. The RPU then becomes a coprocessor to the remaining Opteron CPUs on the motherboard/blade.

Depending on the computation acceleration requirements, a motherboard can use one or more RPUs in conjunction with at least one Opteron. Thus on a four-way server with four CPU sockets, at least one socket will have an Opteron CPU. The remaining three sockets can then be filled with either CPUs or RPUs. Each RPU can accelerate the computations by anywhere from 5X to 100X, depending on the algorithm, thus eliminating the need to add more compute servers or purchase faster systems to replace the slower servers. Furthermore, the

FPGA solution is lower power than most of the CPUs and thus it can save a significant amount of power if the RPU's replace one or more CPUs on the server motherboards and blades.

The first RPU's from DRC are designed as coprocessors that drop into sockets designed for the AMD Opteron CPUs, and now that Intel has also opened up its CPU socket interface through its QuickAssist Technology Acceleration Abstraction Layer (AAL), an Intel compatible version will probably be available in 2008. The RPU110-L200 developed by DRC is a module that contains the largest available FPGA from Xilinx (the XC4VLX200), and that FPGA packs about 5 million gates of configurable logic, 96 DSP compute blocks, and a large internal memory of about 6 Mbits (not including the look-up table memory) and still other resources (Figure 1).

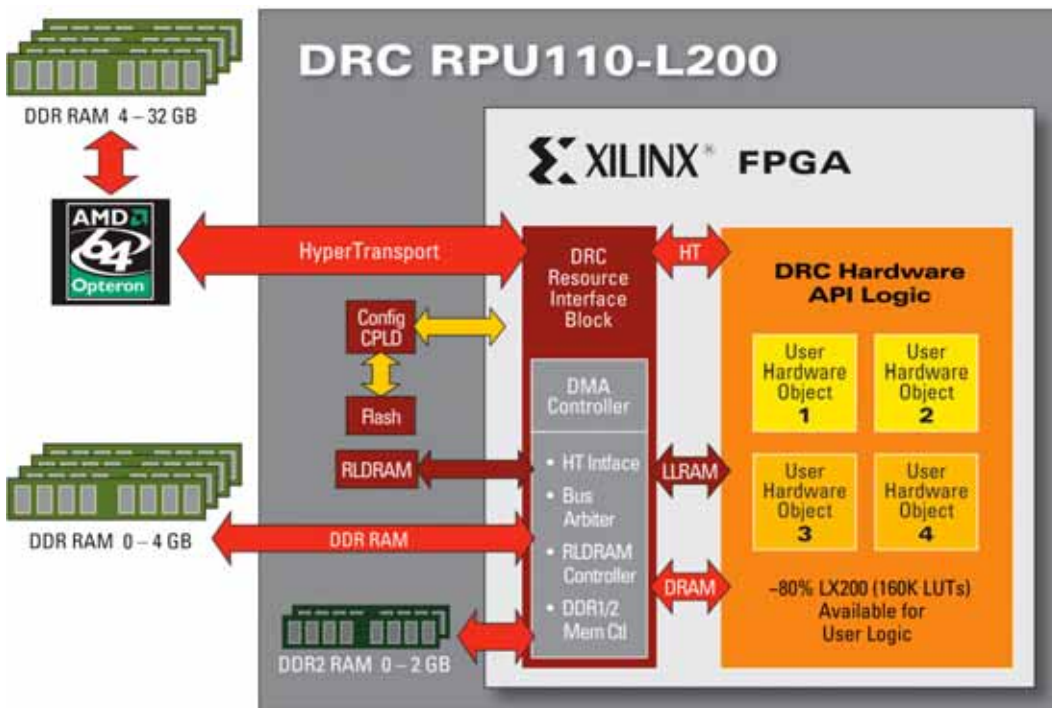


Figure – 1

Surrounding the FPGA on the module is a large local memory of up to 2 Gbytes of DDR2 DRAM), a separate DDR1 memory controller to access memory on the motherboard, and the multiple Hypertransport interfaces that will link to the other Opteron and RPU's on the system board. A wide, 288-bit bus connects to the local DDR2 DRAM to maximize the amount of data

that can be transferred on each bus cycle. The 144-bit wide Northbridge memory interface is the same as that of the Opteron to maintain socket compatibility.

In addition to the DDR1 and DDR2 memory interfaces, a third memory interface supports reduced-latency DRAMs for applications that require minimal latency when accessing data. The low-overhead HyperTransport links provide high-bandwidth low-latency data transfers into and out of the module (*Figure 2*). The RLD and motherboard memory interfaces run at 200 MHz, as does the DDR2 interface on the RPU. The wide interface of the local DDR2 memory on the RPU module reduces the time needed by the computation acceleration logic to move data, thus minimizing the chance that the coprocessor will be data-starved (it sits idling waiting for data to transfer, thus lowering the performance efficiency).



Figure - 2

The module drops right into the Opteron CPU socket and can not only leverage its local 2 Gbyte DDR2 memory, but it can access the host system's DDR1 memory as well (the 4 Gbytes that the CPU would typically talk to over its integrated north-bridge memory controller). A hardware application programming interface (API) layer developed by DRC provides a common programming interface between the dedicated memory interfaces and the internal configurable logic resources.

When a system boots the RPUs will load their initial configuration files from a flash memory in the module. That memory configures the memory interfaces and other system resources that allow the RPU to tie into the host system. These overhead functions consume about 20% of the FPGA's logic, leaving about 160k of the LX200's look-up tables in which the algorithm accelerators can be configured. As the operating system boots and the driver software is installed to make the system aware of the coprocessors, the system will download the

configuration patterns to the FPGA to set up the compute elements that will accelerate the algorithms. The API layer provides the interface between the host software and the configured compute accelerators that DRC refers to as User Hardware Objects.

By leveraging the Opteron interface the module looks like another CPU on the motherboard and the operating system and application software as they execute on the Opteron processor will execute standard calls to subroutines and returns from subroutines. If any of those subroutines are accelerated, the calls will link to the accelerated hardware subroutines, which will transfer the computations to the RPU over the HyperTransport links. When the RPU returns the results the application software treats the results as if they came from another CPU and continues executing.

To create the accelerated algorithms, DRC has also created a development system that consists of a dual-socket Opteron motherboard with either the RPU110-L200, or a smaller version, the RPU100-L60, which is based on the smaller Xilinx LX60 rather than the LX200 FPGA. This system can be used to try out various logic configurations and algorithms to determine the best mix of resources for the targeted application. It will help map the algorithm into the FPGA logic and define the logic resources needed to implement the algorithms in the FPGA fabric and then create the bit stream needed by the FPGA.

Captions:

1. The reconfigurable processing unit developed by DRC Computer includes a large FPGA, 2 Gbytes of DDR2 DRAM, as well as multiple low-latency HyperTransport links and an interface that allows it to drop into an Opteron CPU socket to serve as a coprocessor and accelerate computations.
2. A hardware application programming interface provides a consistent interface between the host software and the configurable hardware in the RPU. When inserted into one of the Opteron sockets in the motherboard, the RPU links to the other CPUs via its low-latency HyperTransport links.

Product trademarks belong to their respective manufacturers:

AMD – Opteron, HyperTransport

DRC – RPU

IBM – Power

Intel – Itanium

Sun – UltraSparc

About the Author:

Giang Le is a product marketing manager at DRC Computer Corporation. Prior to joining DRC, Giang held various marketing management positions at Sun Microsystems, Inc. He has a B.S.C.S and M.B.A.]

