El-Ghazawi

Reference [EL-GH08]

RESEARCH FEATURE

The Promise of High-Performance Reconfigurable Computing

Tarek El-Ghazawi, Esam El-Araby, and Miaoqing Huang, George Washington University
Kris Gaj, George Mason University
Volodymyr Kindratenko, University of Illinois at Urbana-Champaign
Duncan Buell, University of South Carolina

Several high-performance computers now use field-programmable gate arrays as reconfigurable coprocessors. The authors describe the two major contemporary HPRC architectures and explore the pros and cons of each using representative applications from remote sensing, molecular dynamics, bioinformatics, and cryptanalysis.

In the past few years, high-performance computing vendors have introduced many systems containing both microprocessors and field-programmable gate arrays. Three such systems—the Cray XD1, the SRC-6, and the SGI Altix/RASC—are parallel computers that resemble modern HPC architectures, with added FPGA chips. Two of these machines, the Cray XD1 and SGI Altix, also function as traditional HPCs without the reconfigurable chips. In addition, several Beowulf cluster installations contain one or more FPGA cards per node, such as HPTi's reconfigurable cluster from the Air Force Research Laboratory.

In all of these architectures, the FPGAs serve as coprocessors to the microprocessors. The main application executes on the microprocessors, while the FPGAs handle kernels that have a long execution time but lend themselves to hardware implementations. Such kernels are typically data-parallel overlapped computations that can be efficiently implemented as fine-grained architectures, such as single-instruction, multiple-data (SIMD) engines, pipelines, or systolic arrays, to name a few.

Figure 1 shows that a transfer of control can occur during execution of the application on the microprocessor, in which case the system invokes an appropriate architecture in a reconfigurable processor to execute the target operation. To do so, the reconfigurable processor can configure or reconfigure the FPGA "on the fly," while the system's other processors perform computations. This feature is usually referred to as runtime reconfiguration.[1]

From an application development perspective, developers can create the hardware kernel using hardware description languages such as VHDL and Verilog. Other systems allow the use of high-level languages such as SRC Computers' Carte C and Carte Fortran, Impulse Accelerated Technologies' Impulse C, Mitrion C from Mitrionics, and Celoxica's Handel-C. There are also high-level graphical programming development tools such as Annapolis Micro Systems' CoreFire, Starbridge Systems' Viva, Xilinx System Generator, and DSPlogic's Reconfigurable Computing Toolbox.

Readers should consult *Computer*'s March 2007 special issue on high-performance reconfigurable computing for a good overview of modern HPRC systems, application-development tools and frameworks, and applications.

HPRC ARCHITECTURAL TAXONOMY

Many early HPRC systems, such as the SRC-6E and the Starbridge Hypercomputer, can be seen as attached processors. These systems were designed around one node of microprocessors and another of FPGAs. The

0018-9162/08/$25.00 © 2008 IEEE

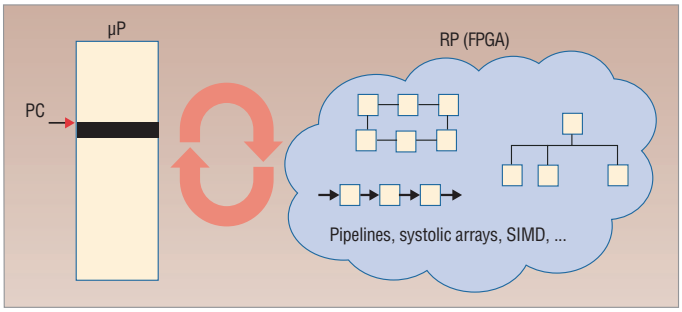

Figure 1. In high-performance reconfigurable computers, field-programmable gate arrays serve as coprocessors to the microprocessors. During execution of the application on the microprocessor, the system invokes an appropriate architecture in the FPGA to execute the target operation.

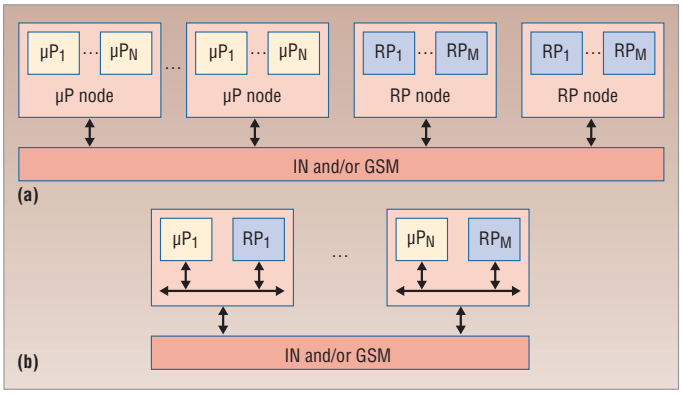

Figure 2. Modern HPRCs can be grouped into two major classes: (a) uniform node nonuniform systems (UNNSs) and (b) nonuniform node uniform systems (NNUSs).

two nodes were connected directly, without a scalable interconnection mechanism.

Here we do not address these early attached processor systems but focus instead on scalable parallel systems such as the Cray XD1, SRC-6, and SGI Altix/RASC as well as reconfigurable Beowulf clusters. These architectures can generally be distinguished by whether each node in the system is homogeneous (uniform) or heterogeneous (nonuniform).[2] A uniform node in this context contains one type of processing element—for example, only microprocessors or FPGAs. Based on this distinction, modern HPRCs can be grouped into two major classes: uniform node nonuniform systems and nonuniform node uniform systems.

Uniform node nonuniform systems

In UNNSs, shown in Figure 2a, nodes strictly have either FPGAs or microprocessors and are linked via an interconnection network to globally shared memory (GSM). Examples of such systems include the SRC-6 and the Altix/RASC. The major advantage of UNNSs is that vendors can vary the ratio of reconfigurable nodes to microprocessor nodes to meet the different demands of customers' applications. This is highly desirable from an economic perspective given the cost difference between FPGAs and microprocessors, and it is particularly suitable for special-purpose systems.

On the downside, having the reconfigurable node and the microprocessor node interact over the shared interconnection network makes them compete for overall bandwidth, and it also increases the latency between the nodes. In addition, code portability could become an issue even within the same type of machine if there is a change in the ratio between the microprocessor nodes and the FPGA nodes.

A representative example of the UNNS is the SRC-6/SRC-7, which consists of one or more general-purpose microprocessor subsystems, one or more MAP reconfigurable subsystems, and global common memory (GCM) nodes of shared memory space. These subsystems are interconnected through a Hi-Bar switch communication layer. The microprocessor boards each include two 2.8-GHz Intel Xeon microprocessors and are connected to the Hi-Bar switch through a SNAP interface. The SNAP card plugs into the dual in-line memory module slot on the microprocessor motherboard to provide higher data transfer rates between the boards than the less efficient but common peripheral component interconnect (PCI) solution. The sustained transfer rate between a microprocessor board and the MAP processors is 1,400 Mbytes per second.

The MAP Series C processor consists of one control FPGA and two user FPGAs, all Xilinx Virtex II-6000-4s. Additionally, each MAP unit contains six interleaved banks of onboard memory (OBM) with a total capacity of 24 Mbytes. The maximum aggregate data transfer rate among all FPGAs and OBM is 4,800 MBps. The user FPGAs are configured such that one is in master mode and the other is in slave mode. A bridge port directly connects a MAP's two FPGAs. Further, MAP processors can be connected via a chain port to create an FPGA array.

Nonuniform node uniform systems

NNUSs, shown in Figure 2b, use only one type of node, thus the system level is uniform. However, each node contains both types of resources, and the FPGAs are connected directly to the microprocessors inside the node.

Examples of such systems are the Cray XD1 and reconfigurable clusters. NNUSs' main drawback is their fixed ratio of FPGAs to microprocessors, which might not suit the traditional vendor-buyer economic model. However, they cater in a straightforward way to the single-program, multiple-data (SPMD) model that most parallel programming paradigms embrace. Further, the latency between the microprocessor and its FPGA coprocessor can be low, and the bandwidth between them will be dedicated—this can mean high performance for many data-intensive applications.

A representative example of the NNUS is the Cray XD1, whose direct-connected processor (DCP) architecture harnesses multiple processors into a single, unified system. The base unit is a chassis, with up to 12 chassis per cabinet. One chassis houses six compute cards, each of which contains two 2.4-GHz AMD Opteron microprocessors and one or two RapidArray Processors (RAPs) that handle communication. The two Opteron microprocessors are connected via AMD's HyperTransport technology with a bandwidth of 3.2 GBps forming a two-way symmetric multiprocessing (SMP) cluster. Each XD1 chassis can be configured with six application-acceleration processors based on Xilinx Virtex-II Pro or Virtex-4 FPGAs. With two RAPs per board, a bandwidth of 8 GBps (4 GBps bidirectional) between boards is available via a RapidArray switch. Half of this switch's 48 links connect to the RAPs on the compute boards within the chassis, while the others can connect to other chassis.

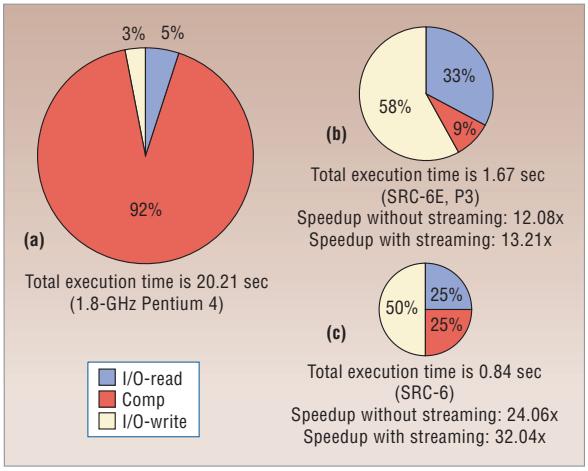

Figure 3. Execution profiles of hyperspectral dimension reduction. (a) Total execution time on 1.8-GHz Pentium 4 microprocessor. (b) Total execution time on SRC-6E. (c) Total execution time on SRC-6.

NODE-LEVEL ISSUES

We have used the SRC-6E and SRC-6 systems to investigate node-level performance of HPRC architectures in processing remote sensing[3] and molecular dynamics[4] applications. These studies included the use of optimization techniques such as pipelining and data transfer overlapping with computation to exploit the inherent temporal and spatial parallelism of such applications.

Remote sensing

Hyperspectral dimension reduction[3] is representative of remote sensing applications with respect to node performance. With FPGAs as coprocessors for the microprocessor, substantial data in this data-intensive application must move back and forth between the microprocessor memory and the FPGA onboard memory. While the bandwidth for such transfers is on the order of GBps, the transfers are an added overhead and represent a challenge on the SRC-6 given the finite size of its OBM.

This overhead can be avoided altogether through the sharing of memory banks, or the bandwidth can be increased to take advantage of FPGAs' outstanding processing speed. Overlapping memory transfers—that is, streaming—between these two processing elements and the computations also can help. As Figure 3a shows, such transfers (I/O read and write operations) take only 8 percent of the application execution time on a 1.8-GHz Pentium 4 microprocessor, while the remaining 92 percent is spent on computations.

As Figure 3b shows, the first-generation SRC-6E achieves a significant speedup over the microprocessor: 12.08× without streaming and 13.21× with streaming. However, the computation time is now only 9 percent of the overall execution time. In the follow-up SRC-6, the bandwidth between the microprocessor and FPGA increases from 380 MBps (sustained) to 1.4 GBps (sustained). As Figure 3c shows, this system achieves a 24.06× speedup (without streaming) and a 32.04× speedup (with streaming) over the microprocessor.

These results clearly demonstrate that bandwidth between the microprocessor and the FPGA must be increased to support more data-intensive applications—an area the third-generation SRC-7 is likely to address. It should be noted, however, that in most HPRCs today, transfers between the microprocessor and FPGA are explicit, further complicating programming models. These two memory subsystems should either be fused into one or integrated into a hierarchy with the objective of reducing or eliminating this overhead and making the transfers transparent.

Molecular dynamics

Nanoscale molecular dynamics (NAMD)[4] is representative of floating-point applications with respect to node performance. A recent case study revealed that when porting such highly optimized code, a sensible approach is to use several design iterations, starting with

the simplest, most straightforward implementation and gradually adding to it until achieving the best solution or running out of FPGA resources.[5]

The study's final dual-FPGA-based implementation was only three times faster than the original code execution. These results, however, are data dependent. For a larger cutoff radius, the original CPU code executes in more than 800 seconds while the FPGA execution time is unchanged, which would constitute a 260× speedup. The need to translate data between the C++ data storage mechanisms and the system-defined MAP/FPGA data storage architecture required considerable development effort. When creating code from scratch to run on an FPGA architecture, a programmer would implement the data storage mechanisms compatible between the CPU and FPGA from the beginning, but this is rarely the case for existing code and adds to the amount of work required to port the code.

Although the "official benchmark" kernel employs double-precision floating-point arithmetic, the NAMD researchers applied algorithmic optimization techniques and implemented their kernel using single-precision floating-point arithmetic for atom locations and 32-bit integer arithmetic for forces. Consequently, the final design occupies most available slices (97 percent), yet utilization of on-chip memory banks (40 percent) and hardware multipliers (28 percent) is low. The fact that the slice limit was reached before any other resource limits suggests that it might be necessary to restructure code to better utilize other available resources. One possible solution is to overlap calculations with data transfer for the next data set to use more available on-chip memory.

Despite the relatively modest speedup achieved, the NAMD study clearly illustrates the potential of HPRC technology. FPGA code development traditionally begins with writing code that implements a textbook algorithm, with little or no optimization. When porting such unoptimized code to an HPRC platform and taking care to optimize the FPGA design, it is easy to obtain a 10×-100× speedup. In contrast, we began with decade-old code optimized to run on the CPU-based platform; such code successfully competes with its FPGA-ported counterpart. It is important to keep in mind that the study's 100-MHz FPGA achieved a 3× application performance improvement over a 2.8-GHz CPU, and FPGAs are on a faster technology growth curve than CPUs.[6]

Lessons learned

Optimization techniques such as overlapping data transfers between the microprocessors and FPGAs with computations are useful for data-intensive, memory-bound applications. However, such applications, including hyperspectral dimension reduction and NAMD, can only achieve good performance when the underlying HPRC architecture supports features such as streaming or overlapping. Streaming can be enabled by architectures that are characterized by high I/O bandwidth and/or tight coupling of FPGAs with associated microprocessors. New promising examples of these are DCP architectures such as AMD's Torrenza initiative for HyperTransport links as well as Intel's QuickAssist technology supporting front side bus (FSB) systems. Large enough memory bandwidth is another equally important feature.

By memory bandwidth we mean that the memory system has sufficient multiplicity as well as speed, width, or depth/size. In other words, because FPGAs can produce and consume data at a high degree of parallelism, the associated memory system should also have an equal degree of multiplicity. Simply put, a large multiple of memory banks with narrow word length of local FPGA memory can be more useful to memory-bound applications on HPRCs than larger and wider memories with fewer parallel banks.

Vendor-provided transparent transfers can enhance performance by guaranteeing the most efficient transfer modes for the underlying platform.

In addition, further node architecture developments are clearly necessary to support programming models with transparent transfers of data between FPGAs and microprocessors by integrating the microprocessor memory and the FPGA memory into the same hierarchy. Vendor-provided transparent transfers can enhance performance by guaranteeing the most efficient transfer modes for the underlying platform. This will let the user focus on algorithmic optimizations that can benefit the application under investigation rather than data transfers or distribution. It also can improve productivity.

SYSTEM-LEVEL ISSUES

We have used the SRC-6 and Cray XD1 systems to investigate system-level performance of HPRC architectures in bioinformatics[7] and cryptanalysis[8-10] applications. These applications provide a near-practical upper bound on HPRC potential performance as well as insight into system-level programmability and performance issues apart from those associated with general high-performance computers. They use integer arithmetic, an area where HPRCs excel, are compute-intensive with lots of computations and not much data transfer between the FPGAs and microprocessors, and inherit both spatial and temporal parallelism.

We distributed the workload of both types of applications over all nodes using the message passing interface (MPI). In the case of DNA and protein analysis, we broadcast a database of reference sequences and scatter sequence queries. The application identified matching scores locally and then gathered them together. Each

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.