Reconfigurable Computing: the Emerging Paradigm for High-Performance Computing

Alan D. George, Ph.D.
Director, NSF CHREC Center
Professor of ECE, University of Florida
(on behalf of faculty/staff of CHREC at Florida, GWU, BYU, and VT)
Computing Reformation

- End of wave (Moore’s Law) riding $f_{clk} +$ ILP (CPU)
  - Explicit parallelism & multicore the new wave
- Many promising technologies on new wave
  - Fixed & reconfigurable multicore device architectures
- Many R&D challenges lie on new wave
  - Tried & true methods no longer sufficient; complexity abounds
  - Semantic gap widening between applications & systems
    - e.g. App developers must now understand & exploit parallelism
- Inherent traits of fixed device architectures
  - App-specific: inflexible, expensive (e.g. ASIC)
  - App-generic: power, cooling, & speed challenges (e.g. Opteron)
  - Many niches between extremes (Cell, DSP, GPU, NP, etc.)
- Reconfigurable architectures promise best of both worlds
  - Speed, flexibility, low-power, adaptability, economy of scale, size
  - Bridging embedded & general-purpose computing, superset of fixed
What is a Reconfigurable Computer?

- System capable of changing hardware structure to address application demands
  - Static or dynamic reconfiguration
  - Reconfigurable computing, configurable computing, custom computing, adaptive computing, etc.
  - Often a mix of conventional fixed & reconfigurable devices (e.g. control-flow CPUs, data-flow FPLDs)

- Enabling technology?
  - Field-programmable multicore devices
  - FPGA et al. (broad & growing space)

- Applications?
  - Vast range – computing and embedded worlds
  - Faster, smaller, less power & heat, adaptable & versatile, selectable precision, high comp. density
Opportunities for RC?

From Satellites to Supercomputers!

10-100x speedups with 2-10x energy savings not uncommon
When and Where to Apply RC?

When do we need?
- When performance & versatility are critical
  - Hardware gates targeted to application-specific requirements
  - System mission or applications change over time
- When the environment is restrictive
  - Limited power, weight, area, volume, etc.
  - Limited communications bandwidth for work offload
- When autonomy and adaptivity are paramount

Where do we need?
- In conventional servers, clusters, and supercomputers (HPC)
  - Field-programmable hardware fits many demands
  - High DOP, finer grain, direct data-flow mapping, bit manipulation, selectable precision, direct control over H/W (e.g. perf. vs. power)
- In space, air, sea, undersea, and ground systems (HPEC)
  - Embedded & deployable systems can reap many advantages w/ RC
Multi-Core/Many-Core Taxonomy

Riding the new MC wave of Moore’s Law

- **Reconfigurable Architecture**
  - Homogeneous
  - Heterogeneous

- **Hybrid**
  - Heterogeneous

- **Fixed Architecture**
  - Homogeneous
  - Heterogeneous

**Devices with segregated RA & FA resources; can use either in stand-alone mode**

**Spectrum of Granularity In Each Class**

**Reconfigurability**

- Datapath
- Device Memory
- PE/Block
- Precision
- Interface
- Mode
- Power
- Interconnect
Reconfigurability Factors

**Datapath**
- Register
- Adder (+)
- Multiplier (×)
- Register

**Device Memory**
- 64 KB x 64
- PE (Processing Element)

**PE/Block**
- 8x8 MAC
- (Processing Element)

**Precision**
- 24-bit Multiply
  - (Processing Element)

**Interface**
- RC Device
- RLDRAM Memory Controller
- RLDRAM SDRAM

**Mode**
- PE1 Prg-A
- PE2 Prg-A
- PE3 Prg-A
- PE4 Prg-A

**Power**
- PE
- Performance
- Power

**Interconnect**
- PE
- MEM
- CHREC

**Byu University of Florida**
**RC Comes to Scientific Computing**

*Broad range of FPGA accelerator technologies*

- Altera & Xilinx FPGA devices
- Tightly coupled subsystems
  - Processor socket (AMD, Intel)
    - e.g. XDI, DRC/Cray, Nallatech
  - System interconnect slot
    - e.g. Celoxica (HTX), SGI (NUMAlink)
  - Memory slot
    - e.g. SRC (SNAP/MAP)
- Loosely coupled subsystems
  - PCIe and PCI-X peripheral cards
    - e.g. Nallatech, GiDEL, Alpha Data
  - Variety of board configurations
- Future?
  - New devices, tighter integration
  - Heterogeneous MC convergence
RC Comes to Scientific Computing

Broad range of FPGA application development tools

- Core libraries
  - Easily accessed as function calls from user apps
- Higher-level programming languages
  - Adaptations to common HLLs
    - C, Matlab, Simulink, Fortran, etc.
    - Design productivity with some loss of efficiency
    - Many vendors, tools, options
    - Domain of app scientists with basic H/W insight
- Lower-level programming languages
  - HDLs (hardware description languages)
    - VHDL, Verilog
    - Design efficiency with some loss of productivity
    - Domain of electrical/computer engineers
- Future?
  - Higher abstraction layers for app formulation
    - Focus on alg/arch exploration, mapping, prediction
    - Semi-automated bridge to code/core generation, libs
What is CHREC?

- NSF Center for High-Performance Reconfigurable Computing
  - First national research center in this field, established Jan’07
  - Leading research groups in RC/HPC/HPEC @ four major universities
    - Founding sites (2007-): Univ. of Florida (lead) and George Wash. Univ.
    - Expansion sites (2008-): Brigham Young Univ. and Virginia Tech

- Under auspices of I/UCRC Program at NSF
  - Industry/University Cooperative Research Center
    - CHREC is supported by both CISE & Engineering Directorates @ NSF
  - CHREC is both a National Center and a Research Consortium
    - University groups form the research base (faculty, students, staff)
    - Industry & government organizations are research partners, sponsors, collaborators, advisory board, & technology-transfer recipients
CHREC Members

1. AFRL Munitions Directorate
2. Altera
3. Arctic Region Supercomputing Center
4. Boeing [new]
5. Cadence
6. GE Aviation Systems
7. Harris Corp. [new]
8. Hewlett-Packard
9. Honeywell
10. IBM Research
11. Intel
12. L-3 Communications [new]
13. Los Alamos National Laboratory [new]
14. Luna Innovations [new]
15. NASA Goddard Space Flight Center
16. NASA Langley Research Center
17. NASA Marshall Space Flight Center
18. National Instruments [new]
19. National Reconnaissance Office
20. National Security Agency
21. Network Appliance [new]
22. Oak Ridge National Laboratory
23. Office of Naval Research
24. Raytheon
25. Rincon Research Corp. [new]
26. Rockwell Collins
27. Sandia National Laboratories

27 members with 37 memberships in 2008
CHREC features a strong team of ~40 graduate students spanning the four university sites.

CHREC Faculty (17 & growing)

- University of Florida (lead)
  - Dr. Alan D. George, Professor of ECE – Center Director
  - Dr. Herman Lam, Associate Professor of ECE
  - Dr. K. Clint Slatton, Assistant Professor of ECE and CCE
  - Dr. Greg Stitt, Assistant Professor of ECE
  - Dr. Ann Gordon-Ross, Assistant Professor of ECE
  - Dr. Saumil Merchant, Research Scientist in ECE

- George Washington University (partner)
  - Dr. Tarek El-Ghazawi, Professor of ECE – GWU Site Director
  - Dr. Ivan Gonzalez, Research Scientist in ECE
  - Dr. Sergio Lopez, Research Scientist in ECE

- Brigham Young University (partner)
  - Dr. Brent E. Nelson, Professor of ECE – BYU Site Director
  - Dr. Michael J. Wirthlin, Associate Professor of ECE
  - Dr. Michael Rice, Professor of ECE
  - Dr. Brad L. Hutchings, Professor of ECE

- Virginia Tech (partner)
  - Dr. Shawn A. Bohner, Associate Professor of CS – VT Site Director
  - Dr. Peter Athanas, Professor of ECE
  - Dr. Wu-Chun Feng, Associate Professor of CS and ECE
  - Dr. Francis K.H. Quek, Professor of CS
Elephant in Living Room

- Semantic gap between apps & architectures
  - Multicore world one of explicit parallelism
  - Yet, architectures increasingly complex to target

- How do we bridge this gap?
  - Holistic concepts & tools for app development
  - **FDTE model: F** abstraction bridges semantic gap
    - Formulation today is usually “seat of pants” within design
    - Poor method for algorithm, architecture, & mapping exploration
    - Lack of fundamental concepts & tools, complexity management
    - Common problem throughout multicore world (CPU, FPGA, Cell, etc.)
  - Formulation is missing link & potential salvation
    - Strategic design playground, abstraction, prediction
      - Irony: learning lesson prevalent in HPC science domains 😊
    - With transitions from **F** to **D** (automation, design patterns, etc.)
2008 CHREC Projects

Florida Site

- **F1-08: System-Level Formulation**
  - Abstraction layer, exploring complex alg. & arch. mappings
- **F2-08: Application Performance Analysis**
  - Run-time performance analysis for HLL-based RC apps
- **F3-08: Case Studies in Multi-FPGA Application Design**
  - Insight in multi-device apps, rapid prediction models, scalability
- **F4-08: Reconfigurable Fault Tolerance & Partial Reconfig.**
  - System-level FT, exploiting RTR and PR for dynamic response to hazards
- **F5-08: Device Characterization & Design Space Exploration**
  - Quantitative analysis of broad device space (FPGA, FPOA, TILE, ECA, FPCA)

George Washington Site

- **G5-08: Library Portability for HLL Acceleration Cores**
  - Provide HLL core portability via Portable Framework I/F (PFIF)
- **G6-08: Intelligent Deployment of IP Cores**
  - Identify HW tasks, deploy intelligently (grouping, IP interconnect)
- **G7-08: Partial Run-Time Reconfiguration for HPRC**
  - Explore PR for HPC, reduce reconfig. delay, HW virtualization
2008 CHREC Projects

BYU Site

- **B1-08: Core Library Framework for HPC/HPEC**
  - XML framework for encapsulating details of reusable circuit cores
- **B2-08: Heterogeneous Architectures for HPEC RC**
  - Device characterizations with RC/Fixed hybrids (FPGA, Cell, GPU)
- **B3-08: High-Reliability RC Design Tools and Techniques**
  - Device-level FT, auto. insertion of SEU mitigation, SEU estimation & detection
- **B4-08: Reliable RC DSP/Comm Systems**
  - Application-specific techniques for DSP/communications system design

Virginia Tech Site

- **V1-08: Model-Based Engineering Framework for HPRC Applications**
  - Explore concepts in MBE for HPRC, abstraction layer, app generation
- **V2-08: Process-to-Core Mapping for Advanced Architectures**
  - Study app hand-crafted mappings for new RC devices; decision framework
CHREC Studies with Science Apps

- Variety of single-FPGA apps studies conducted @ UF Site of CHREC in 2007
  - LIDAR processing, Molecular Dynamics simulation, PDF estimation, Multichannel DDC
- First, start with high-level formulation (“back of envelope”) & prediction
  - Using RAT = RC Amenability Test, developed by CHREC
    - Enter basic parms. of parallel alg. plus general platform data into RAT table
    - Outcome is predicted speedup of *that* algorithm on *that* platform
    - Iterative process with algorithm, precision, platform changes until satisfied
- Next, perform detailed design & coding of alg. in language of choice
  - LIDAR & DDC coded in AccelDSP, MD in Impulse-C, PDF in VHDL
- Translate & execute on platform of choice
  - Suitable platform already determined with aid of RAT
- Evaluate results (in wall-clock speedup vs. fast Xeon/Opteron on same platform)
  - **LIDAR**: predicted = 11.2, actual = 13.1 [on Cray XD1]
  - **MD**: predicted = 10.7, actual = 6.6 [on XDI XD1000]
  - **PDF**: predicted = 13.0, actual = 20.6 [on Cray XD1]
  - **DDC**: predicted = 26.1, actual = 22.6 [on Nallatech H101-PCIXM]
- Optionally, analyze execution, find bottlenecks, improve
  - Using PPW, toolset with RC extension developed by CHREC
  - e.g. in MD: quickly found inefficient buffer-size setting, speedup increased 16%
DARPA Studies @ CHREC

- Research roadmaps for app development on FPGA systems
  - Bridging app/arch semantic gap
    - Prevalent challenge of multi-core
  - RC to revolutionize DoD missions
- 2 DARPA studies by CHREC
  - One @ founding sites + Clemson
  - One @ expansion sites
- Focus areas
  - Study underlying tools limitations
    - Theory, practice, technologies
  - Formulate strategic research paths
    - Revolutionary, impactful
  - Craft research roadmaps
    - Highlight DARPA-hard challenges

- Exploration of a Research Roadmap for Application Development & Execution on FPGA-based Systems
- Future FPGA Design Methodologies and Tool Flows

<table>
<thead>
<tr>
<th>I. Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Algorithm design exploration</td>
</tr>
<tr>
<td>(b) Architecture design exploration</td>
</tr>
<tr>
<td>(c) Performance prediction (speed, area, etc.)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Linguistic design semantics and syntax</td>
</tr>
<tr>
<td>(b) Graphical design semantics and syntax</td>
</tr>
<tr>
<td>(c) Hardware/software codesign</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Compilation</td>
</tr>
<tr>
<td>(b) Libraries and linkage</td>
</tr>
<tr>
<td>(c) Technology mapping (synthesis, place &amp; route)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Test, debug, and verification</td>
</tr>
<tr>
<td>(b) Performance analysis and optimization</td>
</tr>
<tr>
<td>(c) Run-time services</td>
</tr>
</tbody>
</table>
Conclusions

- Growing impact of RC in scientific computing
  - **HPC** and **HPEC**; from satellites to supercomputers!
  - Best of both worlds (speed & power of ASIC, versatility of GPP)

- Broad range of acceleration platforms & tools
  - Device, tool, and system technologies evolving to meet science needs

- Research & technology challenges abound
  - All phases of **FDTE** model, device/system archs., etc.
    - Similar to challenges throughout multicore Moore’s wave
  - **CHREC** sites and partners leading key R&D projects

- Industry/university collaboration is critical to meet challenges
  - Incremental, evolutionary advances by vendors *not* sufficient
  - **CHREC** research collaborations addressing tough problems
  - Industry & government as partners, catalysts, tech-transfer recipients
Thanks for Listening! 😊

- For more info:
  - www.chrec.org
  - george@chrec.org

- Questions?