Runnemede: an Architecture for Ubiquitous High-Performance Computing

Legal Notices

- INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL’S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL® PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. INTEL PRODUCTS ARE NOT INTENDED FOR USE IN MEDICAL, LIFE SAVING, OR LIFE SUSTAINING APPLICATIONS.

- Intel may make changes to specifications and product descriptions at any time, without notice.
- All products, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.
- Intel, processors, chipsets, and desktop boards may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.
- Code names featured are used internally within Intel to identify products that are in development and not yet publicly announced for release. Customers, licensees and other third parties are not authorized by Intel to use code names in advertising, promotion or marketing of any product or services and any such use of Intel’s internal code names is at the sole risk of the user.
- Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.
- Intel, Intel Inside, and the Intel logo are trademarks of Intel Corporation in the United States and other countries.
- *Other names and brands may be claimed as the property of others.
- Copyright © 2009 Intel Corporation.
- This research was, in part, funded by the U.S. Government. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.
DARPA UHPC Program

Runnemede: Intel’s UHPC research architecture

50 GOPS/Watt

Ubiquitous
Heterogeneous Cores

Control Engine (CE)

- Reg. File
- Scratchpad Memory
- SW-Managed Cache
- I/O HW
- Netwk. Int.

eXecution Engine (XE)

- Large Register File
- Scratchpad Memory
- SW-Managed Cache
- Netwk. Int.
Heterogeneous Cores

Control Engine (CE)
- ALU
- Reg. File
- Conventional Core
- SW-Managed Cache
- HW
- Software-controlled local memories

eXecution Engine (XE)
- Large Register File
- ALU
- Reduce memory BW
- Optimized for app. kernels
- SW-Managed Cache
- SW-Managed Cache
- Int.
Heterogeneous Cores

Control Engine (CE)

- Register File
- ALU
- Operating System Code
- Software-controlled local memories

eXecution Engine (XE)

- Large Register File
- Reduced memory BW
- Application Code

Conventional Core

Optimized for a pp. kernels
Reduce mem ory BW
Software-controlled local memories

Operating System Code

Application Code
Blocks: Cores Grouped for Locality

CE

XE XE XE XE

L1 Data Network

L1 Barrier Network

Net. Int.

XE XE XE XE

L2 Scratch-pad
Blocks: Cores Grouped for Locality

- Distributes work to XEs
- Remote memory references
- L1 Data Network
- L1 Barrier Network
- Barriers, multicast, reductions
- L2 Scratch-pad
Case Studies

• Co-design for Synthetic Aperture Radar
• Scratchpads vs. caches
• Network analysis (in paper)

Energy unit: double-precision floating-point multiply (FM64)
Co-Design for Synthetic Aperture Radar

SAR: UHPC “challenge problem”

HW, SW co-designed for energy efficiency
Codesign for Synthetic Aperture Radar

- Initial Version
- w/ SIN/COS Inst.: 86% Reduction
- w/ Improved Alg.: 45% Reduction
- w/ Re-Use in L1: 29% Reduction
- w/ Compiler Opt.: 47% Reduction

Energy Consumed (FM64)
Codesign for Synthetic Aperture Radar

- **Initial Version**
- **w/ SIN/COS Inst.**
- **w/ Improved Alg.**
- **w/ Re-Use in L1**
- **w/ Compiler Opt.**

**Energy Consumed (FM64)**

- **Compute**
- **Memory**
- **Network**

Reduction:
- **75% Reduction**
- **96% Reduction**
Comparing Scratchpads and Caches

Questions:

• Energy?
• Programming Effort?
Matrix Multiplication

- Network Energy
- L1 Energy
- L2 Energy
- DRAM Energy

Memory Energy (FM64)

- Naïve Cache Parallelization: 5.0E+12
- Best Compiled Cache
- 1-Level Scratchpad (Copy Loops): 2.5E+11
- 1-Level Scratchpad (Block Transfers)
- 2-Level Scratchpad (Copy Loops)
- 2-Level Scratchpad (Block Transfers)

51% Less
Givens QR Decomposition

Memory Energy (FM64)

- Naïve Cache Parallelization: 5.7E+11
- Best Compiled Cache: 3.6E+10, 20% Less
- 1-Level Scratchpad (Copy Loops)
- 1-Level Scratchpad (Block Transfers)
- 2-Level Scratchpad (Copy Loops)
- 2-Level Scratchpad (Block Transfers)

Network Energy
L1 Energy
L2 Energy
DRAM Energy

February 25, 2013
Public
Intel Labs
Conclusion

Runnemede is an energy-optimized research architecture

- NTV circuits, power/clock gating, co-design, SW-managed memory
- Co-design: 4x energy improvement
- SW-managed memory: 2-4x memory energy improvement