SAND: 2012-4683C

AMD Fusion12 DEVELOPER SUMMIT

EARLY EXPERIENCES WITH HETEROGENEOUS COMPUTE

Using heterogeneous compute for HPC

S.D. Hammond, J.A. Ang, R.F. Barrett, J.H. Laros, D. Doerfler, M.A. Heroux, C.R. Trott, P. Crozier
Center for Computing Research
Sandia National Laboratories, NM, USA

Thanks to Ted Barragy, Bill Brantley and Greg Branch at AMD

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

NNSA National Nuclear Security Administration

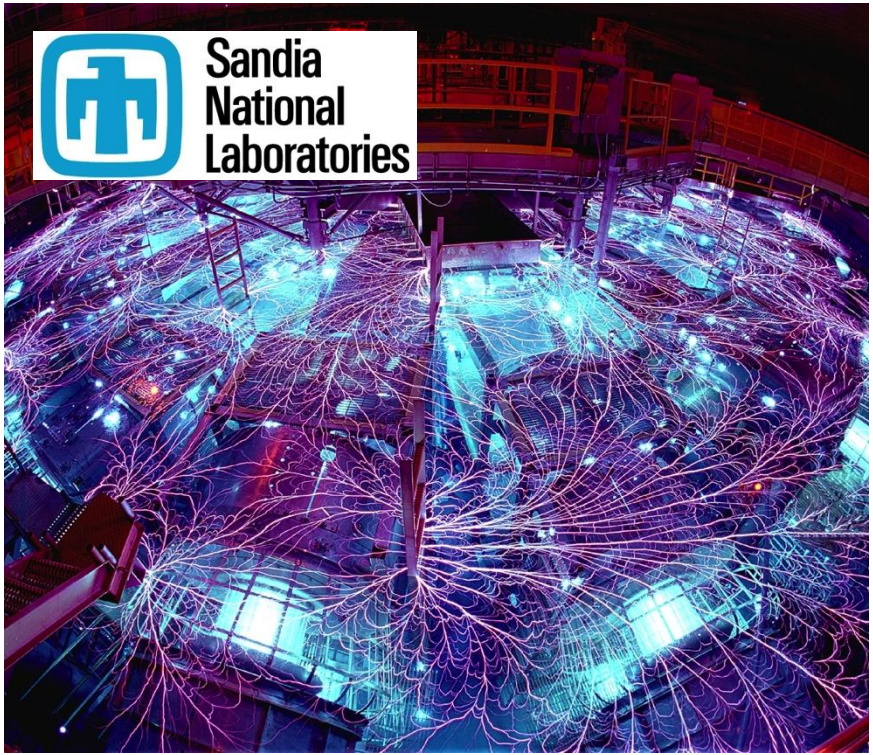
U.S. DEPARTMENT OF ENERGY

WHAT'S IN THIS TALK?

- What is Sandia and how to we fit into the bigger computing picture?
- Why are we interested in heterogeneous compute and what can it offer us?
- What are our experiences with getting started?
- Where are we going from here?

SANDIA AND OUR COMPUTING CHALLENGES

SANDIA NATIONAL LABORATORIES



- Sandia National Laboratories
  - Nuclear Weapons and Safety Systems
  - National & Homeland Security
  - Non-Proliferation Assessments
  - Defense Systems and Assessments
  - Energy, Climate and National Infrastructure
  - Science, Technology and Engineering

- Based in Albuquerque, NM and Livermore, CA
- Projects under the NNSA, DoE Office of Science, DoD, NSF, external industry *etc*

COMPUTING AT SANDIA

- Computing at Sandia is challenging:
  - Multiple scientific and engineering domains, diverse portfolio of needs
  - Multiple users and uses (even within single projects)
  - Multiple technologies, programming languages, libraries, compilers, toolkits, *etc*
  - Some of the largest scales of computing being used in production today (100,000+ processor cores)
  - Need to provide robust, dependable solutions with very high levels of efficiency

- The future:
  - More of the above, bigger scales, more complexity, growing demand
  - Changing technology – contemporary processors and the way we do things now may just not continue to be available or scale to our needs
  - Limited by power – *need* to be more energy efficient, energy limits our maximum scale

THE REAL CHALLENGE

Runs on O(100K) cores

- O(1M) lines of application code
- O(1M)+ lines of library code
- Legacy code (up to 20-30 years)
- Multiple implementation languages (Fortran, C, C++ etc)
- Multiple Terabytes of memory for working sets
- Explicit Message Passing between nodes

Runs on O(1M-1B) threads

- O(1M) lines of application code
- Multiple Petabytes of memory for working sets
- Energy efficient?
- Reliable/Redundant?
- Programming model?

Today → Exascale

$1e10^{15}$ → $1e10^{18}$

THE REAL CHALLENGE

Runs on O(10K) cores

- O(1M) lines of application code
- O(1M)+ lines of library code
- Legacy code (up to 20-30 years)
- Multiple implementation languages (Fortran, C, C++ etc)
- Multiple Terabytes of memory for working sets
- Explicit Message Passing between nodes

What is the 'best' way to port existing applications to new systems?

Runs on O(100K-1B) threads

O(1M) lines of application code
Multiple Petabytes of memory for working sets

- Energy efficient?
- Reliable/Redundant?
- Programming model?

Today → Exascale

$1e10^{15}$ → $1e10^{18}$

THE REAL CHALLENGE

We see two potential paths – *evolutionary*, and *revolutionary*

Runs on O(10K) cores

- O(1M) lines of application code
- O(1M)+ lines of library code
- Legacy code (up to 20-30 years)
- Multiple implementation languages (Fortran, C, C++ etc)
- Multiple Terabytes of memory for working sets
- Explicit Message Passing between nodes

Evolutionary
Gradual modification of code over time

Revolutionary
Rewrites of code using new technologies

Runs on O(100K-1B) threads

- O(1M) lines of application code
- Multiple Petabytes of memory for working sets
- Energy efficient?
- Reliable/Redundant?
- Programming model?

Today → Exascale

$1e10^{15}$ → $1e10^{18}$

WHY ARE WE INTERESTED IN HETEROGENEOUS COMPUTE?

OUR INTEREST IN APU TECHNOLOGY

- APUs are interesting because of the potential options it gives us moving forward
  - **Evolutionary path**: run existing code unmodified on x86 cores, gradually port performance critical sections to GPU. Potential of lower total performance but significant saving in time and cost
  - **Revolutionary path:** completely rewrite codes for the GPU portion, likely to give best performance but potential to be expensive and lengthy development period

- Codes at Sandia also come in many flavors
  - Highly parallel: benefit significantly from GPU
  - Difficult to parallelize (“high Amdahl fraction”) – need a good CPU core
  - Mixed codes: selectively choose where compute kernels run for best performance

- Flexibility is key to having a general platform for our workloads (not just specific applications)

OUR INTEREST IN APU TECHNOLOGY

- Potential for saving in energy
  - Use what you need, turn off what you don't need, efficient mapping of compute to appropriate resource
  - Reduced need to transfer data from GPU to CPU if they are integrated
  - Potential for greatly reduced energy / latency for data movement within an APU when CPU-GPU transfers are required

- Heterogeneous/Fusion-style devices feel like a good option for the future (GPUs alone probably won't get *all* of our codes to Exascale, CPUs alone unlikely to get *many* of our codes to Exascale)

WHAT ARE WE DOING WITH HETEROGENEOUS COMPUTE?

HOW ARE WE REACTING TO THE FUTURE COMPUTING CHALLENGE

- Sandia is either leading or participating in several NNSA/Office of Science projects to understand platforms, gain experience and identify issues/opportunities for future computing systems
  - NNSA/ASC and SC/ASCR Experimental Architectures program
  - Structural Simulation Toolkit (SST) for Architectural Modelling
  - Mantevo “mini-apps” Project
  - Office of Science Co-Design Centers (Materials Modeling and Combusion)
  - Programing Models Projects including the qthread runtime *etc*
  - System software expertise in kernel design, message passing optimization *etc*
  - And lots more…

HOW ARE WE REACTING TO THE FUTURE COMPUTING CHALLENGE

- Sandia is either leading or participating in several NNSA/Office of Science projects to understand platforms, gain experience and identify issues/opportunities for future computing systems
  - **NNSA/ASC and SC/ASCR Experimental Architectures program**
  - Structural Simulation Toolkit (SST) for Architectural Modelling
  - Mantevo “mini-apps” Project
  - Office of Science Co-Design Centers (Materials Modeling and Combusion)
  - Programing Models Projects including the qthread runtime *etc*
  - System software expertise in kernel design, message passing optimization *etc*
  - And lots more…

AMD FUSION APU CLUSTER AT SANDIA

- **Teller Cluster at Sandia**
  - AMD Llano Fusion APUs (2.9GHz)
  - Integrated by Penguin
  - 104 Nodes
    - Some nodes 8GB of faster 1866MHz memory
    - Some nodes 16GB of slower 1600MHz
  - InfiniBand interconnect
  - SSDs on-node for checkpoint research
- Sandia and Penguin working to upgrade to AMD Trinity Fusion before end of FY12

HOW ARE WE REACTING TO THE FUTURE COMPUTING CHALLENGE

- Sandia is either leading or participating in several NNSA/Office of Science projects to understand platforms, gain experience and identify issues/opportunities for future computing systems
  - NNSA/ASC and SC/ASCR Experimental Architectures program
  - Structural Simulation Toolkit (SST) for Architectural Modelling
  - **Mantevo "mini-apps" Project**
  - Office of Science Co-Design Centers (Materials Modeling and Combusion)
  - Programing Models Projects including the qthread runtime *etc*
  - System software expertise in kernel design, message passing optimization *etc*
  - And lots more…

MANTEVO MINI-APPLICATIONS PROJECT

- Our production codes are:
  - Big: O(1M lines of source) + O(50 libraries) = O(>1M lines of code)
  - Multiple implementation languages
  - Complex multi-physics, export controlled, possibly classified, input dependent *(add problem here)*

- Mantevo (http://www.mantevo.org)
  - Mini-applications – condensed, O(1K) lines of source, few (if any) dependencies
  - Enable exploration of areas of algorithmic interest (e.g. CG solve, Molecular dynamics force *etc*)
  - Open source (LGPL), freely distributable
  - Implemented using variety of programming models
  - Lots of academic and lab research into understanding code, performance, porting issues *etc*

WHERE ARE WE WITH HETEROGENEOUS COMPUTE?

WHERE ARE WE CURRENTLY WITH HETEROGENEOUS COMPUTE?

- Mantevo mini-applications implemented in OpenCL
  - MiniFE – Finite-element engineering code, CG solve phase in OpenCL, finite assembly in development
  - MiniMD – Molecular-dynamics, Lennard-Jones problems running in OpenCL
  - (in development) MiniGhost – stencil compute code mixed with complex communication behaviors

- Working with several other NNSA/DoE labs to bring OpenCL and heterogeneous compute to wide community
  - Los Alamos is using OpenCL for several mini-apps and system benchmarks (including Molecular Dynamics EAM potentials for materials modeling)
  - Stanford (SLAC) – investigating use of AMD math libraries to accelerate workloads for complex simulations (potentially huge savings for FFT calculations)

FINITE ELEMENT USING OPENCL ON THE APU/CPU…

Medium "Typical" Problem (200 x 200 x 100)

	WAXPBY	DOT	MATVEC	CG
MPI+GCC	3.93 GF/s	2.14 GF/s	3.18 GF/s	3.16 GF/s
OpenCL (CPU)	3.67 GF/s	4.24 GF/s	3.01 GF/s	3.13 GF/s

Large Problem (250 x 250 x 250)

	WAXPBY	DOT	MATVEC	CG
MPI+GCC	3.82 GF/s	3.41 GF/s	3.01 GF/s	3.17 GF/s
OpenCL (CPU)	3.75 GF/s	4.45 GF/s	3.00 GF/s	3.10 GF/s

Engineering Sample Trinity APU

FINITE ELEMENT USING OPENCL ON THE APU/CPU…

Medium "Typical" Problem (200 x 200 x 100)

	WAXPBY	DOT	MATVEC	CG
MPI+GCC	3.93 GF/s	2.14 GF/s	3.18 GF/s	3.16 GF/s
OpenCL (CPU)	3.67 GF/s	4.24 GF/s	3.01 GF/s	3.13 GF/s

Large Problem (250 x 250 x 250)

	WAXPBY	DOT	MATVEC	CG
				GF/s
				GF/s

OpenCL code about as fast but… is more portable since we can run on GPUs, accelerators and CPUs

Demonstrates how well AMD is doing with OpenCL compiler, important to not lose performance in move to OpenCL code

mple Trinity APU

FINITE ELEMENT USING OPENCL ON THE APU/GPU…

Medium "Typical" Problem (100 x 100 x 100)

	WAXPBY	DOT	MATVEC	CG
MPI+GCC CPU	3.8 GF/s	4.08 GF/s	2.70 GF/s	2.85 GF/s
OpenCL (APU)	2.9 GF/s	0.30 GF/s	1.24 GF/s	1.17 GF/s
OpenCL (Discrete)	7.4 GF/s	0.43 GF/s*	7.25 GF/s	5.06 GF/s

Engineering Sample Trinity APU

FINITE ELEMENT USING OPENCL ON THE APU/GPU…

Medium "Typical" Problem (100 x 100 x 100)

	WAXPBY	DOT	MATVEC	CG
MPI+GCC CPU	3.8 GF/s	4.08 GF/s	2.70 GF/s	2.85 GF/s
OpenCL (APU)	2.9 GF/s	0.30 GF/s	1.24 GF/s	1.17 GF/s
OpenCL (Discrete)	7.4 GF/s	0.43 GF/s*	7.25 GF/s	5.06 GF/s

We are excited about the prospects. These speeds *include* the cost of data movement.

Efforts underway to change matrix formats to be GPU friendly (currently CSR is not efficient but its what our applications use, ELL should be faster)

OUR EXPERIENCES

- Early days, this is not a perfect world
  - It took us 20 years to get parallel MPI applications ported and optimized
  - 12 months into a Fusion system we have codes running (in parallel) and some aspects optimized
  - GPU memory aperture size under Linux is a serious limit for large data sets
- Lack of zero-copy features in Linux holds us back (significantly) but AMD is working to help us
- APU chips are performing well, Llano is a good start, Trinity looks interesting
- OpenCL compilers are good, debuggers and profilers coming along
  - We recognize that our demands push tools to the limit, big data sets, complex software
- OpenACC would be welcomed by our developers and user community (more portable?)

WHY IS HETEROGENEOUS COMPUTING IMPORTANT TO US (AND YOU?)

THE IMPORTANCE OF HETEROGENEOUS COMPUTING

- Sandia and other labs are responsible for a wide variety of missions including developing solutions to improve the competitiveness of US industry and business as well as national security

- We have:
  - Complex mission drivers which support diverse users
  - Increasing sophistication of models, science and problems
  - Pressure to deliver solutions in shorter time frames

- Heterogeneous computing allows us to partition work to the “right” type of processor, to get work done faster and to meet these demands
  - GPUs alone are not enough, CPUs alone use too much power

THE IMPORTANCE OF HETEROGENEOUS COMPUTING

- Sandia is working with AMD and the wider HPC community to:
  - Get tools to work with OpenCL including profilers, debuggers, MPI etc
  - Acting as a test center for how OpenCL and heterogeneous computing can be used in production HPC
  - Giving feedback on tools, hardware, drivers *etc* when really pushed to the absolute limit
  - Want to look at OpenACC and grow community expertise

- Important to Sandia but also important to you
  - Similar algorithms?
  - Similar experiences?
  - Ported libraries and optimized best-practice studies
  - We *all* need tools and we *all* want performance

SOME THOUGHTS AND CONCLUSIONS

SOME THOUGHTS MOVING FORWARD

- High performance computing faces some huge challenges moving forward
  - Energy usage
  - Portability and maintaining the performance of portable code
  - Balance of parallel algorithms and serial code (not *everything* can be made parallel)

- Heterogeneous computing is not easy but it looks like a promising path

- Sandia is working hard to ensure heterogeneous computing in the form of Fusion APUs can work for us in the future
  - Porting mini-applications, investigating tools, compilers, optimizations *etc*

- Exciting times, great challenges, early days but just the beginning

http://www.sandia.gov / sdhammo <at> sandia.gov

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Disclaimer & Attribution

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors.

The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. There is no obligation to update or otherwise correct or revise this information. However, we reserve the right to revise this information and to make changes from time to time to the content hereof without obligation to notify any person of such revisions or changes.

NO REPRESENTATIONS OR WARRANTIES ARE MADE WITH RESPECT TO THE CONTENTS HEREOF AND NO RESPONSIBILITY IS ASSUMED FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION.

ALL IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE ARE EXPRESSLY DISCLAIMED. IN NO EVENT WILL ANY LIABILITY TO ANY PERSON BE INCURRED FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

AMD, the AMD arrow logo, and combinations thereof are trademarks of Advanced Micro Devices, Inc. All other names used in this presentation are for informational purposes only and may be trademarks of their respective owners.

The contents of this presentation were provided by the individual(s) and/or company listed on the title page. The information and opinions presented in this presentation may not represent AMD's positions, strategies or opinions. Unless explicitly stated, AMD is not responsible for the content herein and no endorsements are implied.