U.S. DEPARTMENT OF ENERGY | Office of Science

Potpourri: Tools, SW, testbeds, V&V, proxy apps, system access

Jeffrey Vetter

Presented to
DOE Modeling and Simulation Workshop
Seattle
9 August 2012

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

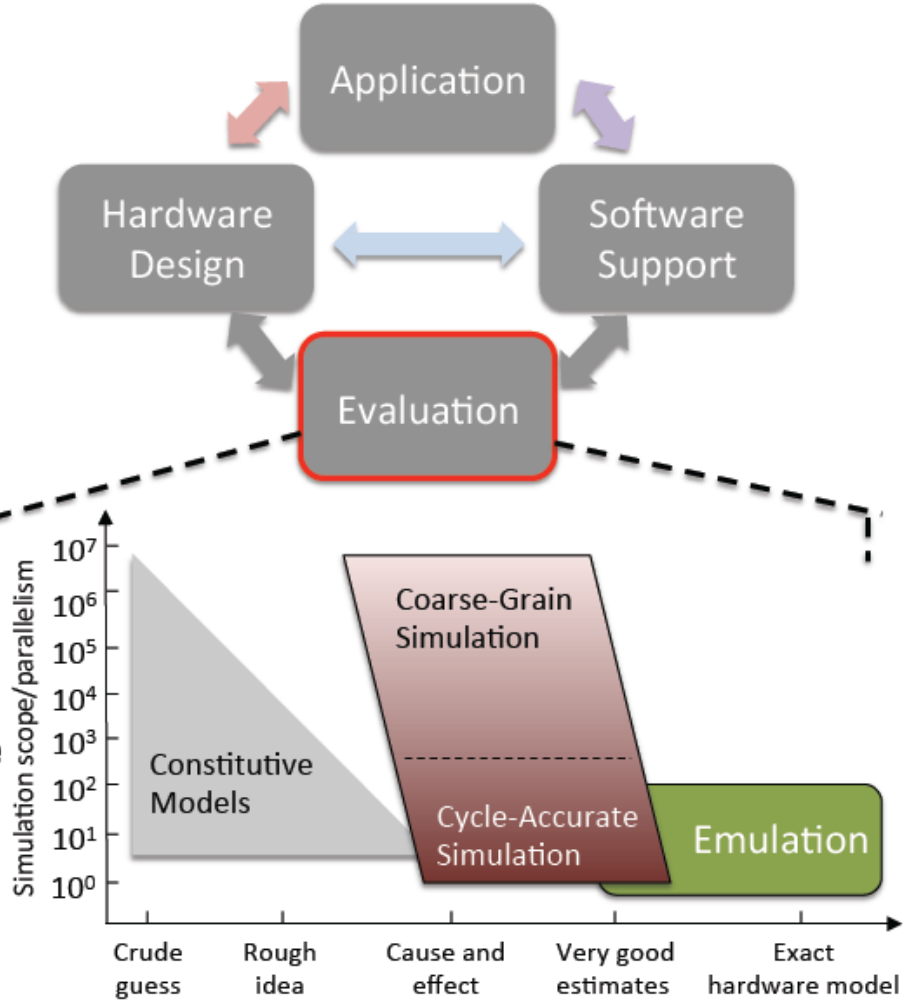
Georgia Tech College of Computing
Computational Science and Engineering

http://ft.ornl.gov ♦ vetter@computer.org

Modeling and Simulation as a Co-design Tool

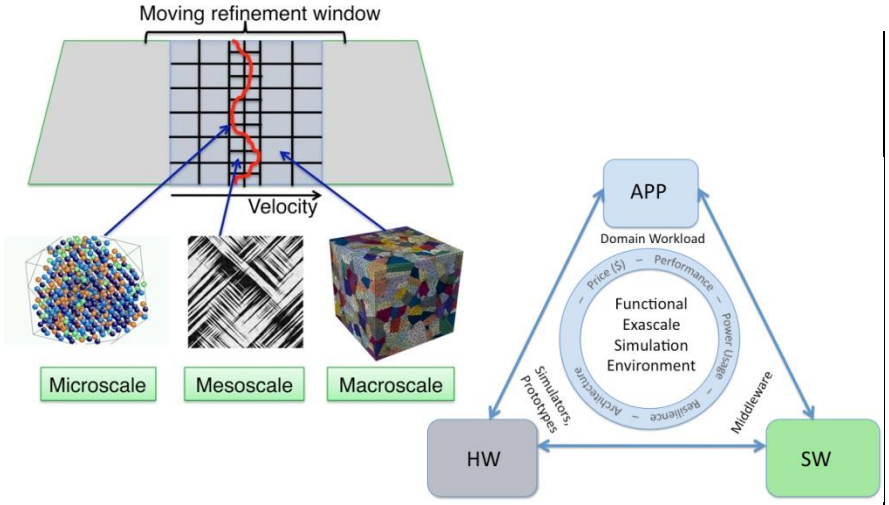
- **Ultimate Question:**
  - Do my applications run well on the machine?
- **Intermediate Questions:**
  - Is the application programmed in the best way?
  - Is there a good mapping of hardware support for software?

- **----------Evaluation ----------**
- **Constitutive Models – can be powerful, but hard to investigate new concepts and complex interactions**
- **Coarse-Grained Simulation – accurate, predicts trends, can scale**
- **Cycle-Accurate Simulation – highly accurate, but can only scale so far**
- **Emulation – essentially exact and fast, but expensive**



Source: Harrod, Lucas et al. 2012

Exascale Co-Design Center for Materials in Extreme Environments



Novel Ideas

Embedded Scale-Bridging Materials Science
- Adaptive physics refinement
- Asynchronous task-based approach

Agile Development of Proxy Application Suite
- Single-scale apps target node-level issues
- Scale-bridging apps target system-level issues

Co-optimization for P^3R: Price, Performance, Power, and Resiliency
- ASPEN, SST models & simulators
- GREMLIN emulator for stress-testing

Impact and Champions

IMPACT. Our goal is to establish the interrelationship between hardware, middleware (software stack), programming models, and algorithms required to enable a productive exascale environment for multiphysics simulations of materials in extreme mechanical and radiation environments.

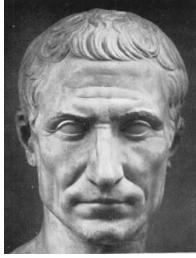
The design and development of extreme environment tolerant advanced materials by manipulating microstructure and interfaces, at the grain scale, depends on such predictive capabilities.

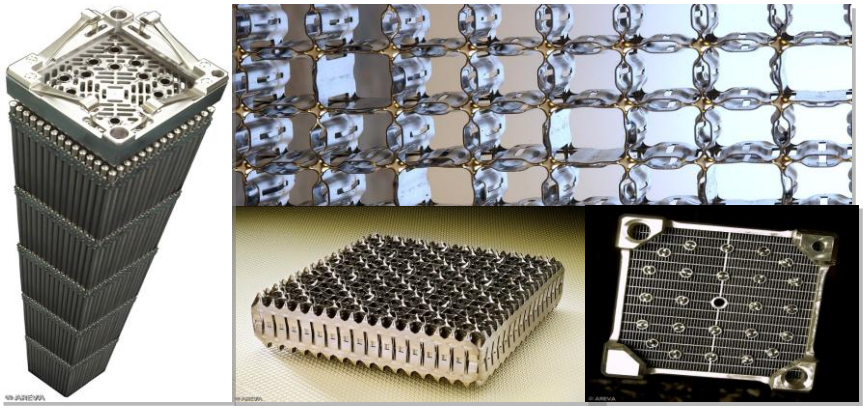
***Director*: Tim Germann (LANL)**
***Deputy Director*: Jim Belak (LLNL)**

Milestones/Dates/Status

	Scheduled	Actual
• Kickoff Workshop	AUG 2011	AUG 2011
• Initial molecular dynamics (MD) SPMD proxy app(s)	DEC 2011	DEC 2011
• Initial scale-bridging MPMD proxy app(s)	MAY 2012	
• Prototype MD DSL	SEP 2012	
• Assessment of data/resource sharing requirements, both for scale-bridging and *in situ* visualization/analysis	2013	
• Demonstrate scale-bridging on 10+ PF-class platform	2015	

CESAR – Center for Exascale Simulation of Advanced Reactors





Novel Ideas

- Develop innovative, scalable algorithms for neutronics and thermo-hydraulics computations suitable for exascale computers
- Couple high-fidelity thermo-hydraulics and neutronics codes for challenging multi-scale, multi-physics computations
- Drive design decisions for next-generation programming models and computer architectures at the exascale

Impact and Champions

Simulating a complete nuclear power system in fine detail will fundamentally change the paradigm of how advanced nuclear reactors are designed, built, tested and operated.

- Every step of the nuclear regulatory timeline can be compressed by guiding expensive experiment efforts.
- New designs can be rapidly prototyped, accident scenarios can be studied in detail, material properties can be discovered, and design margins can be dramatically improved.
- Scientists can analyze problems for a wide range of novel reactor systems.

Milestones/Dates/Status

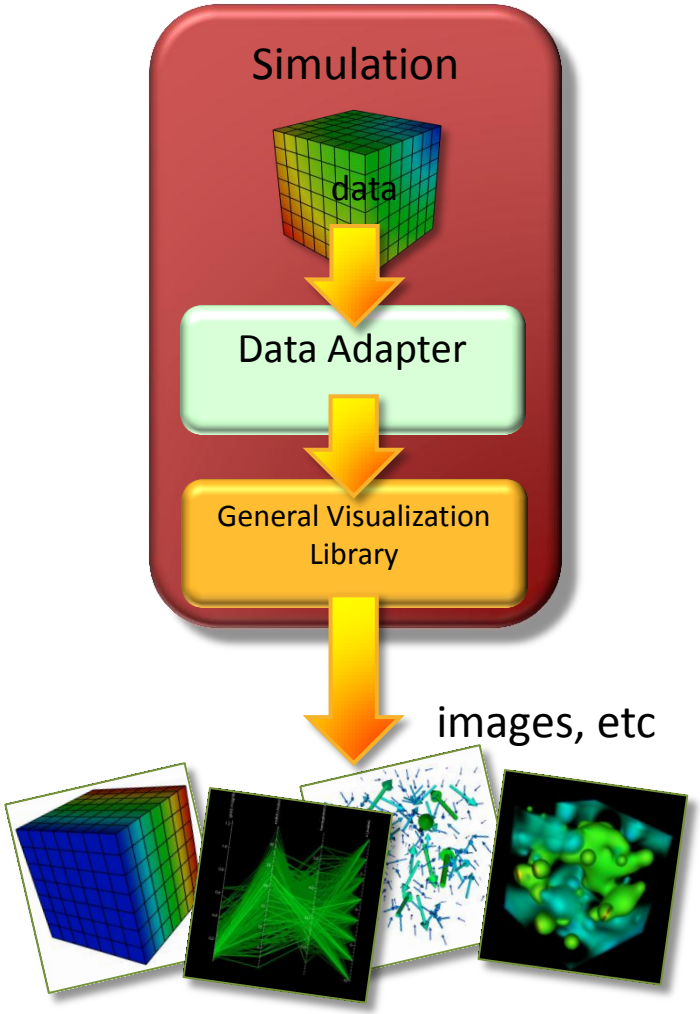
	Scheduled	Actual
• Kernels, initial codes in repository	1/12	12/11
• Formulation of 1st-year calculation	1/12	1/12
• NEK data structures in MOAB	1/12	1/12
• Initial performance model for NEK	7/12	-
• Initial performance analysis for UNIC	7/12	-
• Initial uncertainty quant. runs	7/12	-
• Complete pin bundle calculations	10/12	-
• Custom viz design for NEK/UNIC output	12/12	-

Principal Investigator: Robert Rosner, ANL

March 1, 2012

New I/O models: Tightly Coupled General In Situ Processing

- Simulation uses data adapter layer to make data suitable for general purpose visualization library
- Rich feature set can be called by the simulation
- Operate directly on the simulation's data arrays when possible
- Write once, use many times



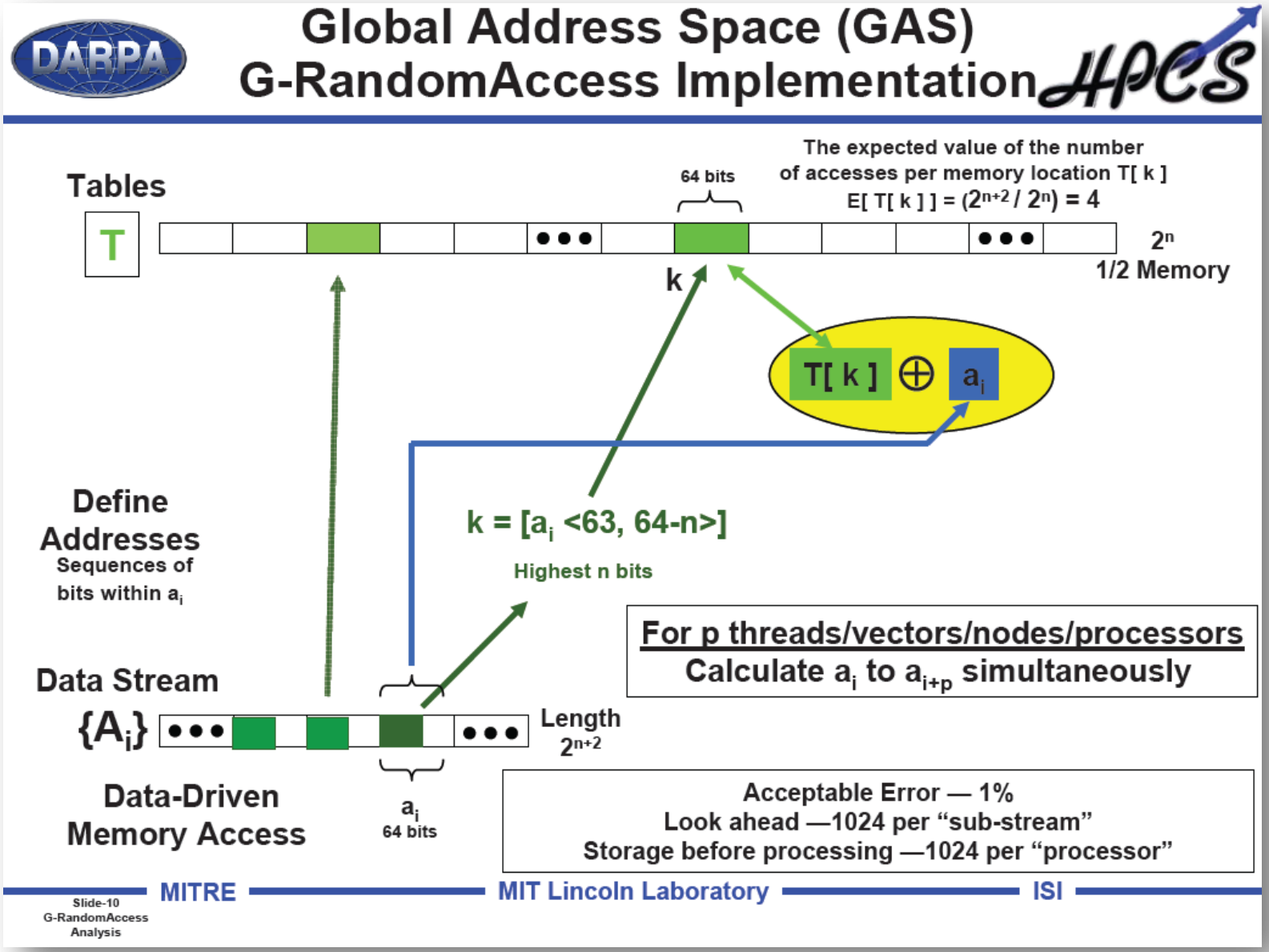
B. Whitlock, J. Favre, and J.S. Meredith, "Parallel In Situ Coupling of a Simulation with a Fully Featured Visualization System," in *Eurographics Symposium on Parallel Graphics and Visualization (EGPGV) in association with Eurographics, 2011*

Tentative Ranking of Predictive Techniques

	Speed	Ease	Flexibility	Accuracy	Scalability
Ad-hoc Analytical Models	1	3	2	4	1
Structured Analytical Models	1	2	1	4	1
Simulation – Functional	3	2	2	3	3
Simulation – Cycle Accurate	4	2	2	2	4
Hardware Emulation (FPGA)	3	3	3	2	3
Similar hardware measurement	2	1	4	2	2
Node Prototype	2	1	4	1	4
Prototype at Scale	2	1	4	1	2
Final System	-	-	-	-	-

Integrated approaches can combine aspects of all of these techniques.

Ad-hoc models

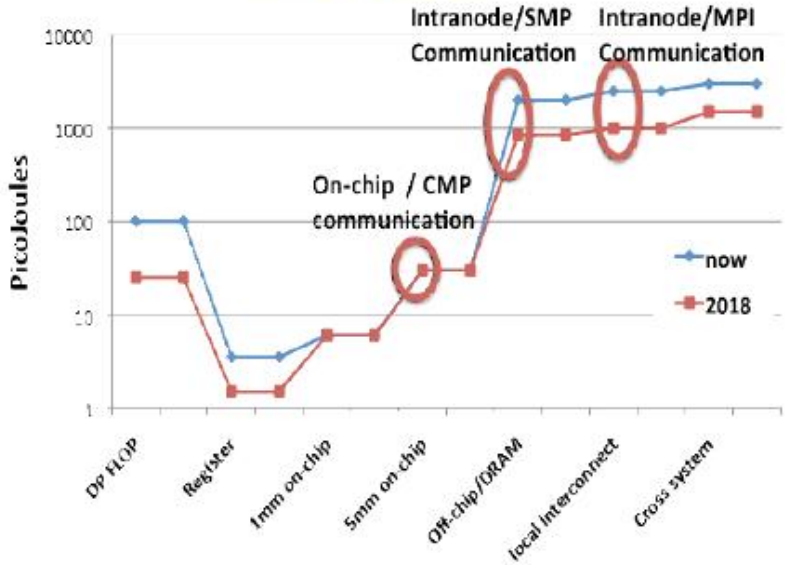


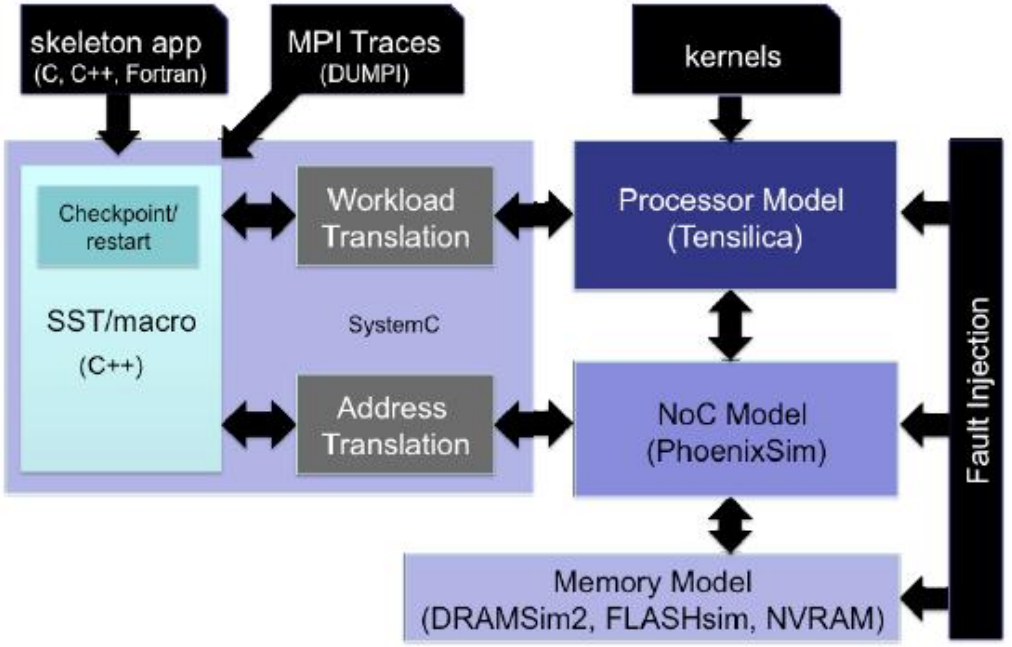
Source: CoDeX team

Mixed Model Simulation

Create flexible, modular, interoperable simulation environment using System-C Industry standard

- *More agile environment to rapidly configure experiments to answer questions posed by vendors and CoDesign centers*
- *Enables accurate multiscale evaulation of energy costs for data movement*





U.S. DEPARTMENT OF ENERGY | Office of Science

NERSC | LLNL | Sandia National Laboratories | Berkeley Lab

Aspen: A Domain Specific Language for Performance Modeling

PI: Jeffrey S. Vetter, ORNL
Kyle Spafford, ORNL

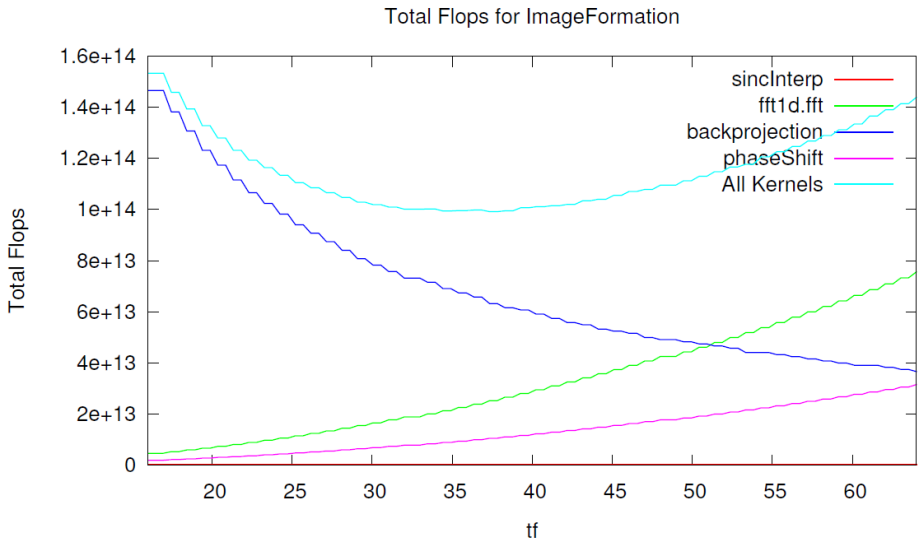
Objectives

- *Design and implement a new language for analytical performance modeling*
- *Use the language to create machine-independent models for important applications and kernels*
- *Develop a suite of analysis tools which operate on the models and produce key performance metrics like available parallelism, arithmetic intensity, and message volume*

Accomplishments

- *Developed a new language, compiler, and set of analysis tools*
- *Constructed models for important apps and mini-apps: MD, UHPC CP 1, Lulesh, 3D FFT*

K. Spafford and J.S. Vetter, "Aspen: A Domain Specific Language for Performance Modeling" To appear in the Proceedings of the ACM\IEEE Conference on High Performance Computing, Networking, Storage, and Analysis. (SC 12).



Example: Studying how the floating point requirements changed based on TF, an application-specific tiling factor in UHPC CP#1

Impact and Champions

- *Increase understanding of application performance requirements*
- *Facilitate early-stage performance planning*
- *Sponsored by DoE – ExMatEx CoDesign Center, DARPA UHPC Echelon Team*

```
1 kernel localFFT {
2   exposes parallelism [n^2]
3   requires flops [5 * n * log2(n)] as dp,
        complex, simd
4   requires loads [a * n * max(1, log(n)/
        log(Z)) * wordSize] from fftVolume
5 }
```

Listing 2. Aspen statements for the local 1D FFTs

CSMD

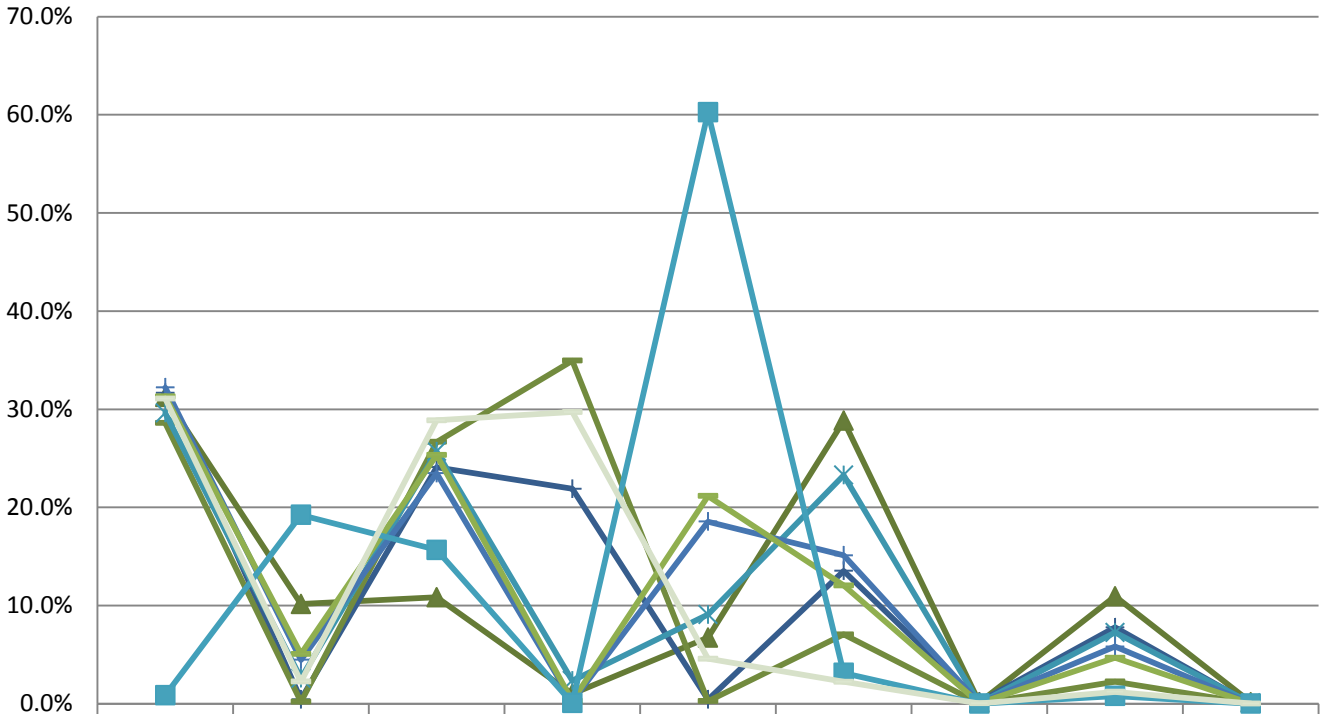
Q&A

- **Co-design applications teams**
  - How do we provide feedback to algorithm and app designers at a coarse yet useful resolution of resource information?
- **Software**
  - How do we improve the integration of the software stack into the modeling and simulation process?
  - How do we improve the evaluation of new operating systems, programming models, etc on simulators and emulators?
  - How do we use modsim at runtime and in programming environments?

Proxy Apps

Calibrating co-design (and other) codes

- **ExMatEx**
  - Lulesh
  - CoMD
  - Spasm
  - Ddcmd
  - M-tree
  - VPFFT
- **CESAR**
  - NEK, NEKBONE
  - OpenMC
  - UNIQ, MOCFE



	MemOps%	Mem SIMD%	Moves%	FlOps%	Fp SIMD%	IntOps%	Int SIMD%	BrOps%	Misc%
MOCFE Whole Program	31.2%	10.1%	10.8%	1.0%	6.7%	28.8%	0.1%	10.9%	0.1%
SPASM Whole Program	31.7%	0.4%	24.1%	21.9%	0.4%	13.5%	0.2%	7.8%	0.0%
DDCMD Whole Program	28.6%	0.2%	26.7%	34.9%	0.3%	7.1%	0.0%	2.3%	0.0%
NEK5000 (MHD) Whole Program	29.6%	2.6%	25.7%	2.4%	9.1%	23.3%	0.1%	7.2%	0.1%
NEKBONE(4096Strong) Whole Program	32.2%	4.5%	23.5%	0.3%	18.6%	15.1%	0.1%	5.8%	0.0%
NEKBONE(1024Weak) Whole Program	31.3%	5.1%	25.3%	0.3%	21.2%	12.1%	0.1%	4.7%	0.0%
HPCC: HPL Whole Program	0.9%	19.2%	15.7%	0.1%	60.2%	3.1%	0.0%	0.8%	0.0%
LULESH Whole Program	31.1%	2.2%	28.9%	29.7%	4.6%	2.2%	0.0%	1.2%	0.0%

Mantevo miniapps project

Sandia National Laboratories

- Enable rapid exploration in application space context.
  - Target key performance issues.
  - Developed/owned by application team.
  - Enables meaningful conversation across different communities.
  - ASC L2 Milestone validating connection just completed.

Miniapp	Capability
miniMD	Lennard-Jones MD
miniFE	Implicit Finite Element (FEM)
miniGhost	Eulerian boundary exchange
miniXyce	Electronic device simulation
miniITCFE	Implicit Thermal Conduction FEM
miniETCFE	Explicit Dynamics FEM
PhD mesh	Explicit FEM

Source: A. Rodriguez

The Scalable HeterOgeneous Computing (SHOC) Benchmark Suite

http://bit.ly/shocmarx

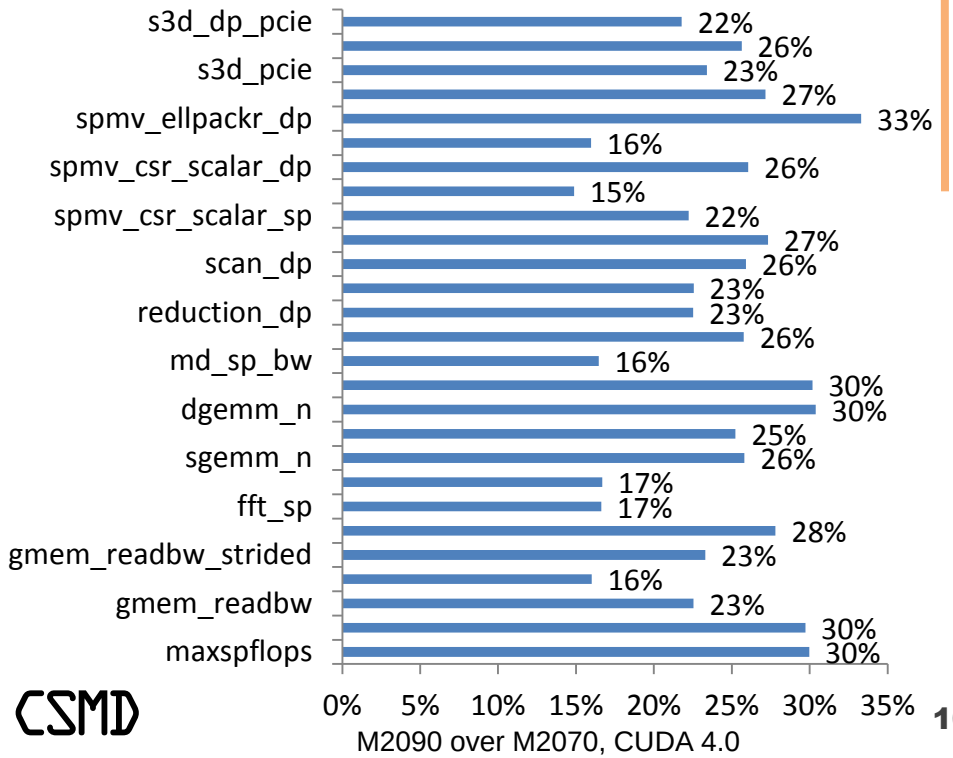
PI: Jeffrey S. Vetter, ORNL
Future Technologies Group

Objectives

- *Design and implement a set of performance and stability tests for HPC systems with **heterogeneous** architectures*
- *Implemented each test in MPI, OpenCL, CUDA to*
  - *Evaluate the differences in these emerging programming models*
  - *Across diverse range of architectures (e.g., NVIDIA, AMD, ARM)*
- *Open Source for easy use, porting, contributions*

Accomplishments

- *Consistent open source software releases*
  - *Over 10000 downloads internationally since 2010*
  - *Used in multiple procurements worldwide*
  - *Used by vendors and researchers for testing, understanding*
- *Overview published at 3rd Workshop General-Purpose Computation on Graphics Processing Units (GPGPU '10)*



Impact and Champions

- *Increase understanding of how important applications will map to emerging architectures*
- *Provide a standardized test suite for architecture evaluations, procurements, and acceptance tests*
- *Entice contributions from HPC community*
- *Sponsored by NSF, DOE*

A. Danalis, G. Marin, C. McCurdy, J. Meredith, P.C. Roth, K. Spafford, V. Tipparaju, and J.S. Vetter, "The Scalable HeterOgeneous Computing (SHOC) Benchmark Suite," in Third Workshop on General-Purpose Computation on Graphics Processors (GPGPU 2010)`. Pittsburgh, 2010

Q&A

- **Proxy Apps**
  - Today's apps are a good start but how do we represent future (10yr) apps?
  - How do we identify the metrics that proxy apps should represent and then design calibrated proxy apps?
  - What are the important features for simulation and modeling?
- **Caveat: Blackcomb analysis of memory access patterns**
  - Most mini-apps remove any 'interesting' data structures (and, hence, memory access patterns)

Testbeds

Technologies Assessment (~2007-2009)

	Availability	Productivity	Reliability	Performance
Intel MIC	No	?	No	Not yet
Clearspeed	No	No	Yes	Yes
Cell	No	No	Yes	Yes
Cyclops64	No	No	Yes	Partial
FPGAs	Yes	Some cases	Partial	Not on FP
AMD GPUs	Yes	OpenCL/FSA	No	Yes
NVIDIA Fermi	Yes	CUDA/OpenCL	Yes	Yes
NVIDIA Kepler	No	CUDA/OpenCL	Yes	Yes
SGI Atom	No	-	-	-
Cray XMT	Yes	Yes	Yes	For graphs
Sun Rock	No	-	-	-

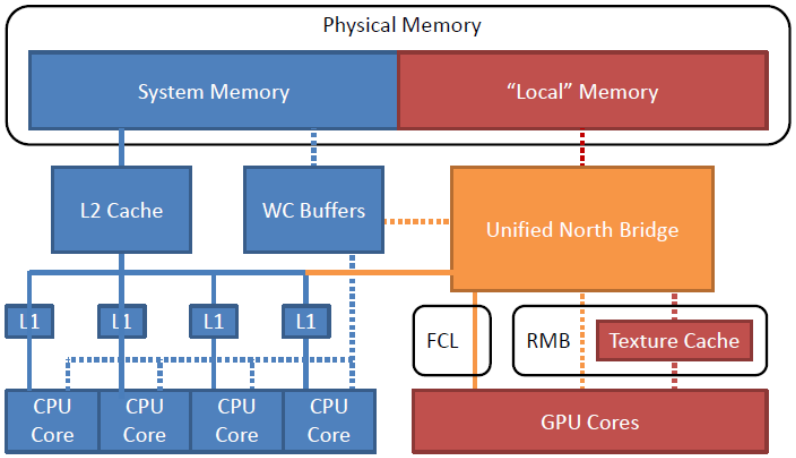
Testbeds

Sandia National Laboratories

- AMD Fusion
  - 104 nodes, Llano Fusion APU: K10 (4x2.9 GHz) + GPU (400x600 MHz) with common address space; Qlogic IB.
  - Upgrading all nodes to Trinity Fusion APU (August/Sept).
- Cray XK6
  - 52 nodes, AMD Interlagos (8/16 @2.1 GHz) + Nvidia Fermi GPU (16x32).
- Intel MIC
  - 42 nodes, Xeon Westmere (2x6 @3.46 GHz, 24GB) + Knights Ferry (KNF: 30/32 cores @1.05 GHz, 2GB); Mellanox IB.
  - Knights Corner (KNC) node.
  - Intel Sandy Bridge cluster: 42 nodes x 2 x 8, toward KNC.
- Tilera TILE-Gx36 processors
  - 4 x 36 cores@1.2 GHz.
- Convey HC-1ex
  - Xeon Nehalem (4 @2.13 GHz), 4 FPGA Co-processor, 8 FPGA “personalities”.
- Calxeda/ARM
  - 1.1 Ghz 4-8 nodes, 4 cores per node.
- Cray XE6
  - 20 nodes AMD 2x8 Magny-Cours + Gemini interconnect.
- Nvidia: 8 Fermi GPUs

Source: A. Rodriguez

AMD Llano's fused memory hierarchy



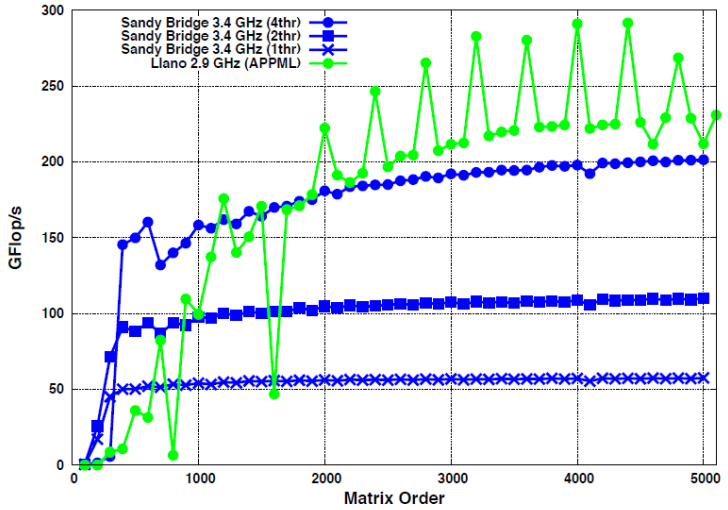
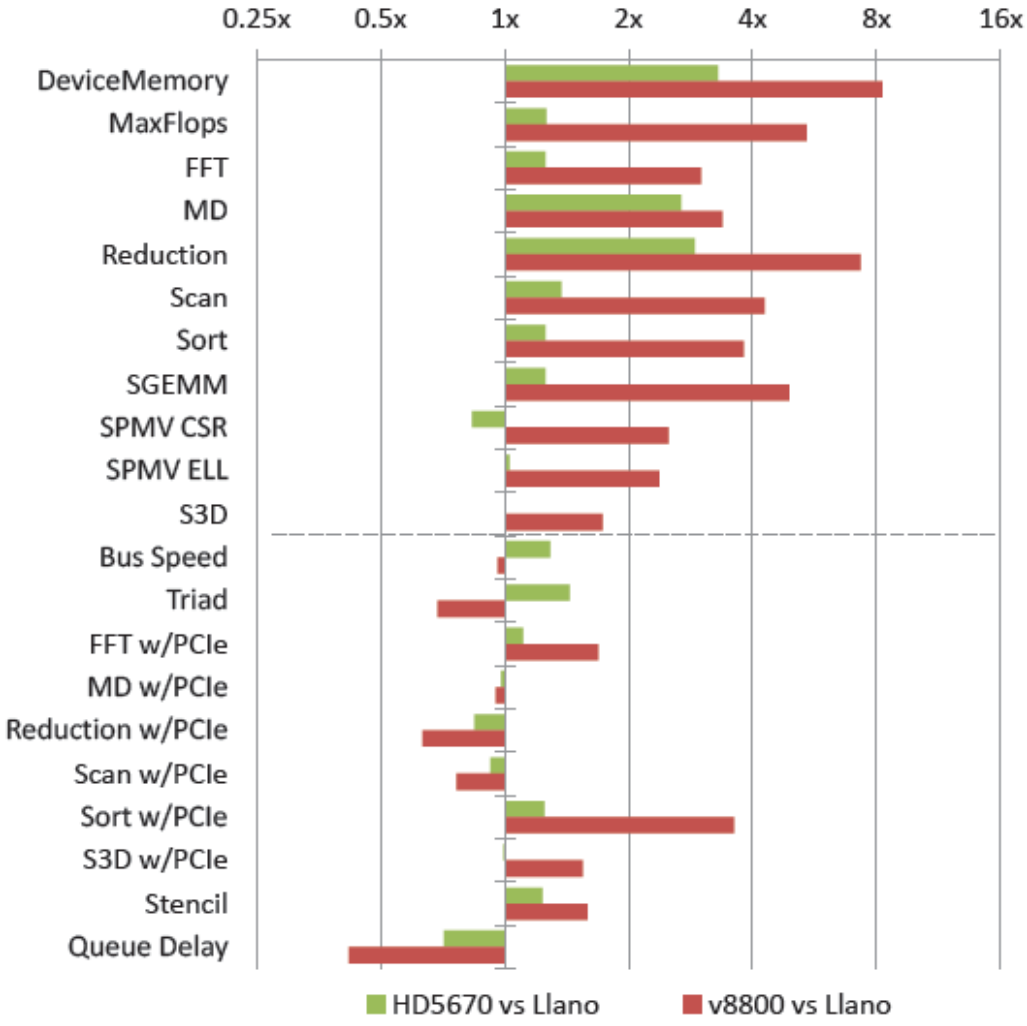


Figure 3: SGEMM Performance (one, two, and four CPU threads for Sandy Bridge and the OpenCL-based AMD APPML for Llano's fGPU)



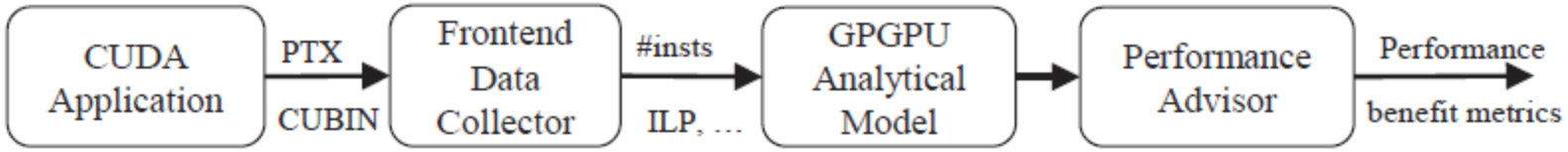
K. Spafford, J.S. Meredith, S. Lee, D. Li, P.C. Roth, and J.S. Vetter, "The Tradeoffs of Fused Memory Hierarchies in Heterogeneous Architectures," in ACM Computing Frontiers (CF). Cagliari, Italy: ACM, 2012.
Note: Both SB and Llano are consumer parts, not server parts.

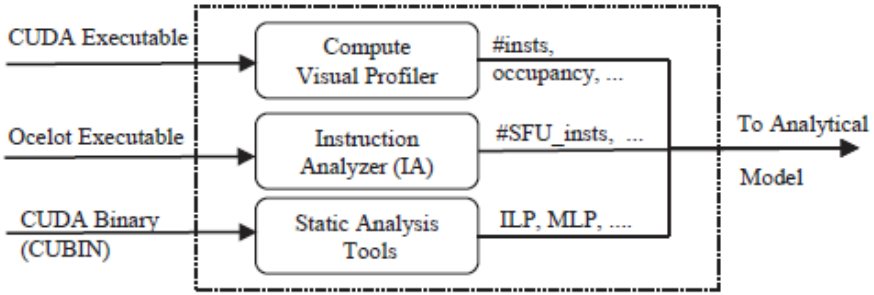
Q&A

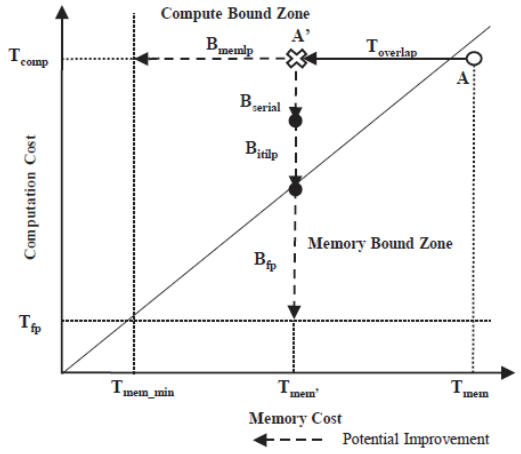
- **Testbeds**
  - We will have diverse architectures possible for exascale
    - Heterogeneous nodes
    - Different memory models
    - Multi-mode memory devices (NVRAM, non-ECC)
  - How do we select our testbeds?
  - Do we have corresponding simulators?

Other ModSim Tools and Uses

Porting applications to new architectures: identifying concurrency and possible benefits on a GPU







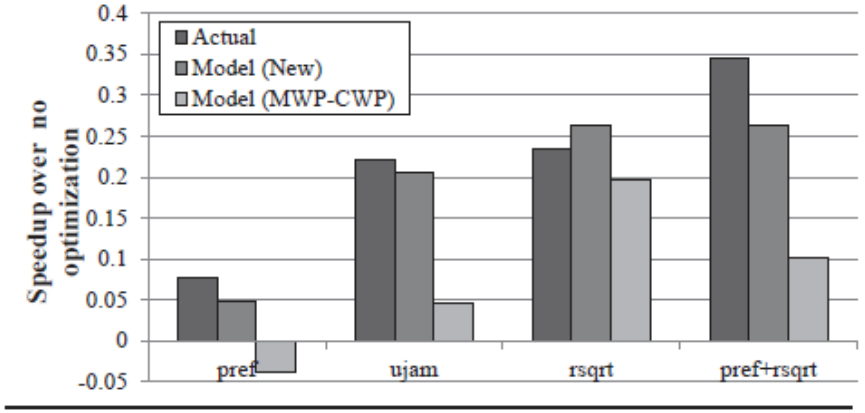


Figure 9. The performance improvement prediction over the baseline FMM_U of the MWP-CWP and our models.

Courtesy Rich Vuduc, Georgia Tech, DOE Vancouver Project: https://ft.ornl.gov/trac/vancouver

Embedding performance models in applications

- Specify a Performance Expectation using a 'model'
  - `$ipc_peak * 0.15 < $ipc`
- If sustained IPC drops below 15% of peak
  - Take an action
- Measurement and data collection left with PA runtime
- Unnecessary to store raw data – lower overhead
- PA enabled compiler could optimize instrumentation

```
pa_start (&pa, '$ipc_peak*0.15<$ipc');
for (j = 1; j <= lastrow - firstrow + 1; j++)
{
  sum = 0.0;
  for (k = rowstr[j]; k < rowstr[j + 1]; k++)
      sum = sum + a[k] * p[colidx[k]];
  w[j] = sum;
}
pa_end(pa);
```

Acknowledgements

- **Contributors and Sponsors**
  - Future Technologies Group: http://ft.ornl.gov
  - US National Science Foundation Keeneland Project: http://keeneland.gatech.edu
  - US Department of Energy Office of Science
    - DOE Vancouver Project: https://ft.ornl.gov/trac/vancouver
    - DOE Blackcomb Project: https://ft.ornl.gov/trac/blackcomb
    - DOE ExMatEx Codesign Center: http://codesign.lanl.gov
    - DOE Cesar Codesign Center: http://cesar.mcs.anl.gov/
    - DOE Exascale Efforts: http://science.energy.gov/ascr/research/computer-science/
  - Scalable Heterogeneous Computing Benchmark team: http://bit.ly/shocmarx
  - US DARPA NVIDIA Echelon
  - International Exascale Software Project: http://www.exascale.org/iesp/Main_Page
  - NVIDIA CUDA Center of Excellence