WORKSHOP ON MODELING & SIMULATION OF SYSTEMS AND APPLICATIONS

August 12 – 14, 2015 – University of Washington, Seattle

Rich Carlson, *U.S. Department of Energy, Office of Science*
Laura Carrington, *San Diego Supercomputer Center*
Bruce Childers, *University of Pittsburgh*
Jon Hiller, *Science and Technology Associates*
Adolfy Hoisie, *Pacific Northwest National Laboratory*
Darren Kerbyson, *Pacific Northwest National Laboratory*
Martin Schulz, *Lawrence Livermore National Laboratory*
Dolores Shaffer, *Science and Technology Associates*
Ankur Srivastava, *University of Maryland*
Jeffrey Vetter, *Oak Ridge National Laboratory/Georgia Institute of Technology*
Noel Wheeler, *U.S. Department of Defense*
Sudhakar Yalamanchili, *Georgia Institute of Technology*

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.

(8/2010)

PNNL-25245

Workshop on Modeling & Simulation of Systems and Applications

Organizing Committee

Adolfy Hoisie (Chair)
Pacific Northwest National Laboratory
Rich Carlson
U.S. Department of Energy, Office of Science
Laura Carrington
San Diego Supercomputer Center
Bruce Childers
University of Pittsburgh
Jon Hiller
Science and Technology Associates
Darren Kerbyson
Pacific Northwest National Laboratory
Martin Schulz
Lawrence Livermore National Laboratory
Dolores Shaffer
Science and Technology Associates
Ankur Srivastava
University of Maryland
Jeffrey Vetter
Oak Ridge National Laboratory/ Georgia Institute of Technology
Noel Wheeler
U.S. Department of Defense
Sudhakar Yalamanchili
Georgia Institute of Technology

Steering Committee

Almadena Chtchelkanova
National Science Foundation
William Harrod
U.S. Department of Energy, Office of Advanced Scientific Computing Research
Anne Fitzpatrick
U.S. Department of Defense/HPC Modernization
Adolfy Hoisie
Pacific Northwest National Laboratory
Sue Kelly
U.S. Department of Energy/National Nuclear Security Administration
Dolores Shaffer
Science and Technology Associates
Noel Wheeler
U.S. Department of Defense

Workshop Date: August 12-14, 2015
University of Washington, Seattle

March 2016

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Executive Summary

The 2015 Workshop on Modeling & Simulation of Systems and Applications (hereafter ModSim 2015) marked the fourth annual gathering of computer scientists, hardware architects, application developers, and computer vendors, who spent more than two days addressing critical technology and research areas of emphasis for modeling and simulation (ModSim) activities.

As the expansion toward extreme-scale systems continues, ModSim's role in the hardware-software co-design process has become even more essential. ModSim promotes community-developed technology, techniques, and tools to support investigations relating to the combined power, performance, and reliability requirements, as well as the companion trade-offs, that affect all levels of the stack. Moreover, ModSim enables scientists and engineers to analyze future algorithms, applications, and computing systems well before they are actualized, affording the flexibility to make necessary design decisions and improvements before they can influence new computing devices; heterogeneous systems; complex workflows involving data instruments, networks, and computing; or increasingly sophisticated, data-intensive multiphysics applications.

Over the course of the workshop, participants examined the current state of the art in ModSim technologies, tools, and methods, including exploring various topics related to architectures and dealing with high-performance computing (HPC) on a large spectrum of scales (from embedded to extreme). Notably, this workshop again promoted why maintaining a commitment to ModSim research remains vital for complete computational science planning and execution because of how intricately it can affect effective HPC ecosystem development. Based on the presentations and discussions at every ModSim workshop since 2012, ModSim's applicability to the entire life cycle of systems and applications, from design to implementation to dynamic/online optimization, has been consistently reaffirmed.

ModSim 2015 culminated with topical discussions geared toward *measurement and benchmarking*, *best practices for model validation*, *predictive modeling of algorithms*, and *integration of thermal models into computer system modeling*. These four topics compose the core of this report. As in years past, they are presented via major research directions, gaps, cross-pollination, and path forward.

An event wrap-up along with the ModSim 2015 agenda, invited talks, and other presentations are available online at: http://hpc.pnl.gov/modsim/2015/index.shtml.

Acronyms and Abbreviations

AE	Artifact Evaluation
DOE	U.S. Department of Energy
HPC	high-performance computing
I/O	input/output
ModSim 2015	2015 Workshop on Modeling & Simulation of Systems and Applications
ModSim	modeling and simulation
PPR	power, performance, and reliability
SWAP	size, weight, area, and power

Acknowledgment

The Organizing Committee respectfully acknowledges and thanks the Steering Committee, as well as the agencies and institutions they represent, for their many contributions throughout the process of organizing and facilitating ModSim 2015. We also would like to thank our invited speakers and the many researchers who submitted position presentations and posters describing their novel research approaches for performance modeling and simulation at extreme scales. Special thanks to Charity Plata for her essential contributions to the ModSim web presence, for helping with the report writing, and for high-quality communications support throughout the process. We also acknowledge the effective administrative support for the committee meetings and logistics in Seattle from Jill Dornfeld.

Contents

Executive Summary ii
Acronyms and Abbreviations iii
Acknowledgment iv
1.0 Measurement and Benchmarking 1
1.1 Major Research Contributions 1
1.1.1 *Should the community have a standard set of benchmarks?* 1
1.1.2 *What are suitable abstractions that can be used?* 2
1.1.3 *Because the real goal is not 'lots of FLOPS' but getting science done, how are the 'science units' that a machine can do measured?* 2
1.2 Gaps 3
1.3 Cross-pollination 3
1.4 Path Forward 3
2.0 Best Practices for Model Validation 5
2.1 Major Research Contributions 5
2.2 Gaps 6
2.3 Cross-pollination 7
2.4 Path Forward 8
3.0 Predictive Modeling of Algorithms 9
3.1 Major Research Contributions 9
3.2 Gaps 10
3.3 Cross-pollination 10
3.4 Path Forward 10
4.0 Integration of Thermal Models into Computer System Modeling 12
4.1 Major Research Contributions 12
4.1.1 *Modeling granularity* 12
4.1.2 *Modeling interactions with performance and reliability* 13
4.1.3 *Trade-offs* 13
4.2 Gaps 13
4.3 Cross-pollination 14
4.4 Path Forward 14
5.0 Conclusion 16
6.0 References 17

1.0 Measurement and Benchmarking

Leads: Dolores Shaffer and Darren Kerbyson

The initial focus area of the 2015 Workshop on Modeling & Simulation of Systems and Applications (ModSim 2015), Measurement and Benchmarking, examined the current state of affairs as well as the needs and possible future directions for both the development and use of benchmarking. Much of the active and lively discussion centered on current high-profile benchmarking activities and their lack of actual technical use in guiding the direction of future computer system development. Many needs for benchmarks were discussed, including suitable abstractions and representation of applications, multiple metrics for evaluation, and the consideration of possible future requirements.

1.1 Major Research Contributions

The session began with opening remarks about how the headline benchmarks for Linpack and, more recently, the High Performance Conjugate Gradients (HPCG) project are wrong for high-performance computing (HPC). Although the provision of a single performance number affords comparison and to rank systems, its use does not extend any further, such as in assisting the optimization of current systems or guiding future system designs for the necessary range of requirements and workloads. In addition, it was observed that there is an important temporal element to what is being measured—workloads change over time, especially as systems scale and capabilities increase. Today’s workload should be viewed as a proxy for tomorrow’s work. As such, empirical measurements of what is running over time could help show trends in application requirements. Research contributions were characterized into the three main aspects:

1. Having a standard set of benchmarks
2. Providing suitable abstractions that represent workload
3. Understanding which metrics may be important in the future.

1.1.1 *Should the community have a standard set of benchmarks?*

The concept of a standard set of HPC benchmarks was deemed as lacking, yet it generally was acknowledged that such standards could provide good coverage of many workloads—as long as a suitable process for defining and evolving benchmark components in the set was available. Notably, there have been past attempts at generating benchmark suites, including in the mid-2000s by an interagency working group. Unfortunately, all have failed. A contributing factor was the disagreement on standard benchmarks within agencies. Hence, attempting to make the community agree on a fixed list remains unlikely to succeed. Certain benchmarks, such as the breadth-first search (used by Graph 500) or the High Productivity Computing Challenge (HPCC), attract organizations that need benchmarks because they are open, fairly easy to understand, and (at least for a time) relevant to HPC but have limited scope in workload coverage.

One problem in selecting a small set of applications to represent an entire workload stems from concerns in omitting or duplicating certain types of calculations. Within the Measurement and Benchmarking forum, several participants discussed problems their respective organizations have had in creating such sets because of concerns that if an application were dropped, a certain kind of calculation would not be represented. Therefore, vendors bidding on HPC development might produce systems that could not perform that calculation efficiently. Conversely, organizations often are concerned that if one challenging type of calculation occurs in more than one application in the set, vendors will place too much emphasis on that type in their designs.

1.1.2 *What are suitable abstractions that can be used?*

Various representations of applications have been suggested and implemented in the past. A recent example is the proxy applications from the U.S. Department of Energy (DOE) Co-Design centers. They were deemed “partially successful” in achieving their goal to represent key applications under development, but they fell short in fully representing the application for which they were designed. Instead, they were perceived to be more of a snapshot in time. Achieving suitable workload abstractions was deemed important for several reasons including: simplifying the amount of benchmarks that may be needed, targeting key characteristics of the applications and systems, and enabling the measurement process to be more tractable when transitioning into modeling and simulation (ModSim) environments to explore “what-if” analyses of design spaces.

Previously, considerable work focused on workload characterization, but it has lost momentum and is not a current “hot topic” research area. There was consideration regarding if workload characterization should be reinvigorated to assist in identifying and developing suitable abstractions. Along these lines, one suggestion was to think in terms of a limited set of parameters that could be used to describe a system and application. However, not many parameters actually affect performance, e.g., memory reuse is a big factor in performance. Later in the workshop, a presentation proposed that HPC performance space could be covered by roughly five total parameters. Though not discussed at length, there were initial concerns that setting parameters would not provide confidence that specific performance on a distinct application will be achieved.

There is a need for abstractions that can be boiled down to something to simulate. In reflecting if the community could create a tool to do this, the decision was a distinct “Maybe.” However, application implementations vary and may not be correctly reflected. One collective observation was that the community would refactor their code for new architectures. A further suggestion was that if both a full application and proxy were available, a vendor could be told to report on key pieces of the proxy. Then, the application owner could validate whether or not the proxy results are representative.

1.1.3 *Because the real goal is not ‘lots of FLOPS’ but getting science done, how are the ‘science units’ that a machine can do measured?*

The issue of metrics used in measurement and benchmarking remained the focal point during much of the group discussion. It generally was recognized that the ultimate goal was to achieve the most amount of science from a system—basically, achieving the fastest runtimes for a particular science run or highest throughput for a set of runs. Both performance metrics are important, yet they result in a slew of additional considerations: w*hat about power and energy consumption or reliability*? Even system thermal effects or numerical accuracy of calculations could be considered. These different metrics all were touched on as part of the discussions, but differing views were expressed, indicating there currently is no uniform approach.

With regard to power, it was observed that less energy is consumed if the application gets done as quickly as possible, indicating performance is the key metric. Yet, in contract to a facility standpoint, it was noted that power provisioning is more important than energy consumption. Moreover, the effective use of the power provision is made more complex as power swings, a noted “big problem.”

In regard to reliability, there were multiple viewpoints noting how important it was compared to performance and power. For example, silent data corruption was deemed insignificant by several participants yet important to others. Ultimately, thermal effects and numerical accuracy trade-offs were considered to be rather sporadic and in their infancy in terms of research.

1.2 Gaps

Several gaps in current research were identified:

- A need for a comprehensive benchmark suite, covering a large range of workloads, that should also be future proof (as much as possible) to enable a long-lasting view and comparison of system capabilities. Such a suite would have to be based on a thorough workload analysis, accounting for current applications and expected future requirements.
- A need for appropriate abstractions of the workload that can be used in ModSim activities. Such abstractions must represent the workload’s key capabilities while also being suitable for guiding future system and application designs.
- A need for uniformity in handling multiple metrics (currently not an established practice) that requires they be dealt with both in terms of measurement and use. For instance, knowing the relative importance of certain metrics for particular systems could assist in guiding application optimizations.

1.3 Cross-pollination

Given the integration of measurement needs for performance, power, reliability (PPR), and thermal, benchmarking activities also must become more integrated. The comparative interdependencies among the four dimensions (PPR and thermal) further reinforce the argument for integration and cross-pollination.

Benchmarking activities also relate to ModSim. Benchmarks offer the means for gathering input for models and simulations and help the validation process by offering convenient, narrow, and easy-to-quantify “probes” that can be both simulated and measured for ModSim validation and verification (for related observations, refer to Section 2.0).

1.4 Path Forward

There is a need to provide the community with a coherent and comprehensive strategy for measurement and benchmarking of their systems and workloads that will enable the optimization of current systems and guide future system designs. Based on discussions at ModSim 2015, new research should be directed at workload characterization to create community benchmarks (or “protocols”) for qualities that are less dependent on specific applications, e.g., fault tolerance, resilience, and perhaps even energy efficiency (although the latter may be difficult because energy efficiency often can be improved for specific applications).

To strengthen the development of future workload abstractions, representations of current workloads, such as proxy applications, should be investigated to examine their strengths and weaknesses. Other abstractions, including representing systems via a limited number of parameters, should be explored to examine their viability. One such model was presented in the mainstream of ModSim that used only five parameters to represent a system's performance.

.

2.0 Best Practices for Model Validation

Leads: Bruce Childers and Noel Wheeler

A major concern for model validation expressed during ModSim 2015 was determining how much accuracy is actually needed to model PPR sufficiently. In general, the answer seemed to depend on how much the problem under study needs to be characterized to answer a design question. Accuracy can vary from merely pointing to a direction (e.g., better or worse in some metric) to being highly accurate (e.g., specific power consumption or reliability rates). Moreover, the accuracy will vary tremendously depending on the stage of the design process and/or the questions being answered using a model.

2.1 Major Research Contributions

Given that different situations will require varying degrees of fidelity to the physical hardware (accuracy), it was suggested that models should be characterized in terms of their accuracy and how they were validated to achieve that degree of accuracy. That is, a statement, or "specification of accuracy and validation," should be made for a model to describe the validation approaches and its overall trustworthiness. To this end, new specification methods and languages likely will be necessary to capture both a description of the validation approach and the range of accuracy found in different metrics. A specification of accuracy and validation then can be attached to a model offering guidance about how and when a particular model might be used, for example, providing caution against employing a loosely validated model for trend analysis when determining absolute performance. It was considered that the specifications also could be automatically manipulated and used in model selection and composition. The specification also might prove useful for published results: it can be associated with a paper, offering readers insight into the experimental methodology and precision of the results.

Many factors may be included in the specification with accuracy clearly one aspect. The range of accuracy from absolute to trend should be identified. According to several discussion participants, a baseline level should be specified for the identified accuracy to be relevant. For example, a statement of accuracy and validation approach in reporting performance gains with a model should have a fully specified baseline for said gain (i.e., what is the gain being compared against that is important to be captured?). Generally, models also should predict the "direction" of the trend. Discrete data points were declared less useful as they may lack precision. If an analysis declares a "speedup," it also should assert which direction it is trending. Without the trend analysis, the speedup, by itself, may be meaningless, particularly when the underlying situation for the evaluation is changed (e.g., trying a new architecture widget in a different setting).

Another important factor in specifying accuracy and validation established during this best practices discussion was to provide qualitative justification for a trend with details on how results were developed. The specification clearly must explain what was done to facilitate the speedup. Indeed, published results often are lacking in this regard and do not sufficiently capture what was done, instead offering simplification mistakes, page limitations, or more emphasis on the novel widget rather than the methods used to validate models for the widget. To create the desired trend, the specification should capture an analysis of the model's facets that were adjusted and in which direction. This information is needed to specify accuracy and validation and allow other researchers to replicate the work, taking it in new, advancing directions.

A third factor of specification addressed at ModSim 2015 involved considering how one metric relates to another. Often, there are trade-offs between metrics, and a gain in one metric may come at the cost of a loss in another. Consequently, simply validating a model in one metric, without considering other metrics, may diminish the model's usefulness. The specification should capture how the validation considers the interplay of multiple metrics, i.e., PPR. Results reported with the model also should consider this interplay, reporting the gains and losses across all relevant metrics to understand gains in a specific context.

2.2 Gaps

The meeting participants readily determined there are many gaps in improving model validation and specifying the validation. First, when a model is evaluated, it requires a stimulus into the model. Understanding the stimulus is important to enable interpretation of the results. For example, consider a relatively simple stimulus with little variation is used to validate a model. Clearly, the model then will be validated only for that specific circumstance, and results could vary widely for a completely different stimulus with more variable behaviors. Thus, a validation is useful only after knowing how the validation was done with the stimulus.

Standard stimuli might be used to assist in the validation, and these need to be developed. For example, accepted input traffic patterns should yield known results with a model. Otherwise, the model is possibly flawed. This may argue for development of a suite of stimuli for validating models. The standard stimuli could take many forms, such as benchmark suites, traces, traffic patterns, proxy applications, and/or input data sets.

As part of the stimulus for a model, stress tests with microbenchmarks were considered useful to characterize models and understand limitations. The stress tests should be small enough to be well specified (as noted earlier) and cover a range of behaviors under which a widget will be used. Of course, full benchmarks also should be used to evaluate innovations in more realistic situations, but the stress test microbenchmarks can provide critical information for calibration. Both standardized stimuli and microbenchmarks are useful to validate models and capture their accuracy, as well as for comparing techniques across a standardized suite of behaviors. In today's research, we have, to a certain extent, moved away from microbenchmarks. These again must become part of the best practices employed for model validation and evaluations.

A second gap identified was the lack of infrastructure for validation, including critical facets for repeatability of experiments. Infrastructure is needed to encapsulate tools and build chains (e.g., archives and virtual machines) so others can accurately replicate the experimental process. Reconstituting the build process for experimental results should not be the difficult aspect of the process. Today, however, it is a major impediment to the interpretation of results and to understanding how those results might carry over to new situations.

A third gap was seen as the lack of a centralized forum for the community to access, leverage, and deposit stimuli/workloads, results, codes, etc. Simplifying procedures for providing benchmarks and stimuli is insufficient. A repository is necessary to store and share the stimuli, results, experimental setups, etc.

A fourth gap involves debugging tools. Per the discussion participants, variables surrounding model accuracy and validation should be removed or, at least, well understood. Valuable time can be consumed in debugging the experimental environment. Meanwhile, the question of validating the experiment becomes unclear: are you debugging the benchmark software or the simulation environment itself? Where is the bug, or, worse, how do we gain assurance that a "silent bug" is not influencing the results, especially when this bug may be in the model, experimental setup, or simulator?

A fifth gap in improving model validation originates in the reporting and specification of models (as noted in the previous section). Specifically, to gain deeper credibility in results for better interpretation, more understanding concerning what data surrounding the models and experiments actually must be provided is required. It was noted that some aspects of this already are being done, but what is not and what should be expected are unclear. Indeed, many papers feature results without basis or detail. Perhaps, some "standard" set of assumptions is needed for the results, along with the specification of accuracy and validation for the model, e.g., details about the parameters or configuration used. This information potentially could be provided in the same format used to run an experiment, i.e., the configuration files from the simulator would be captured and made available. Along with reporting the parameters and experimental setup, some indication of the magnitude range is needed. For example, are the reported results in the ballpark and correctly indicate the trends? While this can be partially satisfied by the specification of accuracy and validation for the model, the exact experimental conditions also are needed. With this reporting information, results will be more credible and likely to have impact. As noted, unreproducible experiments have questionable value overall and may be "one-of-a-kind" examples of results rather than impactful ones that can be directly built on. Along with specification, standardization of stimuli and practices, and better reporting procedures, a culture change needs to come about. The broader community must elevate its level of expectation to demand improved validation and reporting.

2.3 Cross-pollination

Opportunities for cross-pollinating ideas among communities were seen as a way to improve experimental methods, including validation. Several communities are pushing the frontier in this regard, and the ModSim community for HPC and computer architecture also can benefit from these ideas, which include publishing models/simulation software, patches, and improvements to simulation environments for new widgets before accepting papers. That is, any software changes to prototype a widget would be included and considered during the review process. Of course, this policy may impose a significant burden on evaluators, particularly during a large paper review for a conference or workshop. An alternative approach is Artifact Evaluation (AE), which reviews artifacts underlying experiments in an optional, post-acceptance process. AE provides a reward seal of "acceptance" that a paper has met some community standard for the software tools and experimental methodologies for the results. Specification of validation also could be included as part of this evaluation. Similar to an AE seal, rewards could be offered for the best contributions in highly visible conferences, such as the International Symposium on Computer Architecture (ISCA), International Conference for High-Performance Computer Architecture (HPCA), International Conference for High Performance Computing, Networking, Storage and Analysis (SC), and others. Funding agencies also could incentivize improved experimental methods and their related reporting by providing direct support to make benchmarks, models, and tools available via a community repository.

2.4 Path Forward

The path toward improved model validation involves advances in research, community building, and infrastructure. As noted, better ways to validate and understand model accuracy are required. Likewise, new ways to fully specify accuracy and validation methods are needed, possibly combined with ways to compose specifications of models. Investment in research definitely is required for these areas. A centralized exchange also needs to be established as a definitive rallying point for the community to contribute and share ideas, methods, and artifacts/experiments. Making the most of workshops, tutorials, competitions, awards, and other related mechanisms could be crucial to build awareness of and recognition for the community. ModSim researchers can set the example by exhibiting the extra effort to fully describe their experiments and validation approaches. By setting this good example, other research communities may be inspired to follow suit.

3.0 Predictive Modeling of Algorithms

Leads: Laura Carrington and Jeffrey Vetter

Upcoming systems will have significantly different designs for memory (e.g., reduced per core memory bandwidth and deeper hierarchies) and compute subsystems (e.g., many-core designs). Predictive models can inform PPR sensitivities of important HPC applications to these massive architectural changes by mapping applications properties to the underlying hardware resources. Often, predictive models are developed for key kernels of a given application, as well as respective implementation variants. Together, these models can aid developers in making algorithmic and implementation-level choices to prepare for upcoming architectures, as well as to influence future systems design. The discussions in this session focused on determining the key properties and capabilities of predictive models; how the ModSim community can determine their accuracy; and how the models can influence processes for co-design, procurements, and code optimization.

3.1 Major Research Contributions

For predictive models to assist in algorithmic and system design decisions, the models should be able to determine what additional performance benefits can be realized with incremental changes in hardware design or algorithmic changes. As long as they provide guidance about PPR sensitivity needed for the co-design process (e.g., whether a proposed architectural or algorithmic change has the potential to positively or negatively affect performance), these models do not need to be 100 percent correct. These sensitivity models can be analytical or statistical machine learning models that capture changes in an application's performance and/or power as a result of modifications in the hardware. Sensitivity models also can help point to specific hardware subcomponents that bottleneck performance (e.g., memory, interconnect, or input/output [I/O]), thereby providing an indication of what subcomponents need to be improved for better PPR responses. Session participants also suggested that historical evolution of technologies and their impact on performance and power also could be used as the basis for designing sensitivity models and determining the impact of major paradigm shifts in future hardware designs. For example, could the past 10-year history in the evolution of I/O subsystem and associated changes in the performance provide some hints on the impact of upcoming I/O designs? Early hardware prototypes cannot merely assist in projecting future performance, but they also could enable the development and validation of sensitivity models.

Identification of optimization strategies for current and future systems is another area where these sensitivity models can play a role. Sensitivity models can help identify and rank what characteristics of computations correlate to observed performance, thereby informing a set of optimization strategies for a system. The models also can offer an indication of what the performance improvement will be for different hardware changes (e.g., "what if" analysis of new memory hierarchy) and an application's scaling behavior. This information collected using models is critical for developers and vendors in deciding what software and hardware factors should be high priorities for the co-design process to be effective. Employing the current state of practice developers use for forming an abstract machine model to make projections about design and implementation choices could be made more formal and easily expressed to co-design stakeholders with predictive models.

Discussion throughout this session also involved new programming models (e.g., task-based models) and how they can be better understood with the help of application characterizations. For example, for a task-based model, characterizations could be used to ascertain the overhead of spawning and distributing tasks. Application developers tend to take a “wait-and-see” approach, meaning there currently is not enough modeling or measured evidence pointing to the benefit of emerging programming models for developers to adopt them. Taking an existing application and modeling its behavior as if it was redesigned as a task-based model is complex. Furthermore, such an effort is stymied by the lack of clarity in regard to what applications characteristics make them suitable for task-based execution. Task granularity and determining if there is some reasonable way of synchronizing tasks also are distinct considerations.

3.2 Gaps

The two main identified gaps are in the areas of 1) application characterization and 2) sensitivity model development. Understanding the detailed computational and communication characteristics of DOE workloads is vital in the co-design process. While some initial steps have been taken with tools such as Oxbow (Sreepathi 2014), there are no standard methodologies and metrics for in-depth application characterization that can feed into the co-design process. System designers must know what types of computations are required on the system, and application developers need to understand the application’s computational composition of their application and how that maps to the hardware. For both, a clear and standard set of metrics (that quantify the application’s interactions with major sub-components of the hardware) is needed to accurately express the computational properties of large-scale applications. Proxies could be used to simplify the workloads for architects, but whether a given set of proxies actually represents real application is a looming question. In addition to the gap in application characterization, a new breed of models, sensitivity models, will be critical moving forward in the co-design process. These sensitivity models need not be 100 percent accurate in PPR predictions. Rather, they would predict the application’s PPR trends as it reacts to changes in the hardware design. Understanding an application’s performance and power sensitivity to various design choices is critical for a co-design process.

3.3 Cross-pollination

Ideas for cross-pollination with other areas include the overlap between application and system characterization and benchmarking and measurement. Detailed characterization requires sophisticated measurement tools that can capture characterization metrics without perturbing application behavior. In addition, benchmarking often is used to validate the characterization measurements.

3.4 Path Forward

The path forward for improving application characterization and developing application sensitivity models includes the following aspects:

- Currently, there is a lack of methods and tools to construct good characterization of our applications portfolio. Even when characterization is available, there is no standard way to express it. Methodologies and tools required to characterize and express large-scale applications are essential to inform hardware designers about what is important from a PPR perspective. The ModSim community must decide on a set of application characteristics/metrics that can be used to accurately describe the

PPR aspects of applications. This also could include sensitivities to different settings/designs of hardware components.

- Invest in research and development for modeling new and emerging programming systems, such as task-based programming models.
- Invest in research and development to make models portable or easily adaptable to different system designs.

4.0 Integration of Thermal Models into Computer System Modeling

Leads: Sudhakar Yalamanchili and Ankur Srivastava

The group's focus was on assessing the needs, challenges, and potential solution directions confronting the integration of thermal models into models of single- and multi-node computing systems, which increasingly is becoming an important area to consider. For example, in military electronics, size, weight, area, and power (SWAP) are especially important as thermal management takes up to two-thirds of SWAP capacity to accommodate system needs, such as fans for cooling.

4.1 Major Research Contributions

Discussions around research threads and needs largely fell into the following categories:

4.1.1 *Modeling granularity*

Thermal behaviors are influenced by physical and architectural phenomena that occur across several orders of magnitude of timescale. This alone introduces several modeling challenges.

Accurate computation of the thermal field is a computationally expensive operation that can surpass even the demands of compute system modeling. This issue has spawned development of compact thermal models. The challenge here, especially when modeling traverses multiple layers of abstraction, is that the error accumulates across distinct physical models to the point where they may not be useful.

Because of the speed of computing, timescales also have changed over the decades. Temperature changes are experienced at the millisecond timescale, while power events (that lead to significant thermal responses) occur at nanosecond to microsecond timescales. This range diversity changes the way thermal elements are evaluated.

In addition, there is a need for the abstraction of models to higher levels of packaging using the composition of lower-level models, which can be challenging with respect to maintaining accuracy. Given the important role of packaging in determining thermal capacity, it is necessary to get the materials community involved in creating multiphysics models. These models then can be refined (validated) with measurement data.

In multiphysics modeling, small errors in each model can cascade to larger errors (especially when using compact models). One consideration is how accurate temperature must be to predict the occurrence of soft or hard reliability failures. Another issue involves the sensitivity of thermal models to errors in inputs such as power. System models use open-source models of power/energy. How sensitive these thermal models are to errors in the power models, as well as to the package models parameters, must be determined.

Fabrication variations affect leakage power and dynamic power, thereby affecting thermal behaviors. Models may need to incorporate this aspect at some level of granularity.

4.1.2 *Modeling interactions with performance and reliability*

With thermal capacity becoming the limiting factor in many designs, traditional performance and reliability metrics are tightly interwoven with thermal behaviors. Therefore, to optimize target systems, the community must understand, model, and traverse these relationships in simulations.

Recently, it has become increasingly important to understand how temperature interacts with performance and reliability, especially for HPC. For example, the vast majority of SWAP capacity in many U.S. Department of Defense systems stems from thermal management (Fleury 2015, O'Mara 2011).

In addition, there is relatively little understanding regarding how the interfaces between the many different packaging layers will fail. Furthermore, these behaviors are driven by stresses created by power management and application demands. This latter relationship also is not well understood.

The thermal management system's inherent reliability is a target of analysis, especially when considering fluidic cooling at both rack (e.g., chilled doors) and chip scales (e.g., microfluidics). Moreover, there are complex failure mechanisms that should be modeled, understood, and mitigated.

4.1.3 *Trade-offs*

Managing multiphysics-driven phenomena evokes the need for trading off on the impact of complex interactions. In particular, thermal effects now introduce the need for understanding and managing interactions among power (creating thermal fields/stresses), performance, and reliability.

Thermal management introduces cooling costs while also affecting the baseline power dissipation through temperature's nonlinear impact on leakage power, which can lead to thermal runaway (if not managed). The ability to trade off cooling cost and leakage power also must be incorporated into the system design.

There is a need for high-level models of efficiency for cooling technologies. These models will be used to determine when higher- versus lower-efficiency cooling is needed. Notably, in some instances, lower-efficiency cooling may be substantially cheaper to incorporate.

Trade-off analysis is facilitated with common abstractions at all levels and can be used as a basis for model composition. At each level, efficiency may be individually analyzed. There also is a need for iso-performance analysis across cooling solutions and thermal management techniques. Any analysis must be cognizant of the difference between power efficiency versus energy efficiency.

Physical constraints now play an increasingly important role. Computational density is an important parameter subject to trade offs between density and thermal cost/ability. The balance in these requirements remains an important consideration.

4.2 Gaps

There is a dearth of models that capture the interaction between energy consumption, temperature, and cooling knobs. This is due, in part, to a lack of understanding of the relationships between these physical phenomena. Core knowledge currently resides in different communities. To make progress, this gap must be bridged.

Moreover, there generally is insufficient cross-pollination between the thermal modeling community and other communities within HPC. Characterizing thermal management (along with thermal modeling) would be a useful step forward. For example, during the breakout sessions, Dan Ernst (from Cray Inc.) wanted to know if it is better for reduced leakage power to cool using liquid or to rely on traditional air-cooling.

The attendees acknowledged a significant gap in modeling at the rack level. They agreed the hot aisle versus cold aisle approach is not always the solution. The ability to link socket/board models to impact at the rack level in a computationally feasible manner is one area worth targeting.

4.3 Cross-pollination

Here, we identify efforts that span distinct disciplines and are necessary to move thermal modeling and integration challenges forward, including:

- Fast techniques for transient analysis are necessary.
- Understanding coupling between thermal fields with power delivery networks is necessary.
  - This clearly is an issue on chip and, perhaps, becomes a natural part of the thermal model. It does require an understanding of power conversion, regulation, and delivery techniques along with modeling compute and associated thermal phenomena.
  - It is not yet clear how much of an impact thermal fields have on power delivery at the rack level, another area requiring further research.
- It is important to understand thermal coupling with data center uninterruptible power source, which is now fully distributed through board, rack, etc.
- Thermal management is an important issue for the design and use of photonics. This creates a need for cross-pollination between the electrical, optical, and thermal communities.

4.4 Path Forward

There is a fundamental shift occurring in how we view and deal with thermal capacity. Historically, thermal capacity has been a constraint managed by packaging and hardware designers and out of the purview of resource managers (e.g., in runtimes and operating systems).

We note that thermal capacity now is a resource in the same sense as time, space, and power—one that impacts extreme-scale computing performance at a number of timescales and must be allocated and managed like any other resource. Notably, heterogeneous compute resources will consume this capacity at very different rates, for example, field-programmable gate array (FPGA) versus graphic processing unit (GPU) versus central processing unit (CPU). Therefore, performance interference between these elements will be similar to how they interfere at the memory system in the manner they consume memory bandwidth. If one element consumes thermal capacity much faster than others, peak temperatures will be reached quickly, and throttling will occur to avoid thermal events. However, reaching peak temperature at one point on a chip leaves much unused thermal capacity elsewhere on the chip, thereby reducing performance. Better-distributed consumption of the thermal capacity across all elements leads to more useful work that can be achieved in the same time interval.

With the preceding backdrop, the following steps will begin to address several of the important topics outlined within this chapter:

- Define metrics for performance-thermal coupling. This enables researchers to realize coordinated efforts by optimizing the same metrics.
- Create a limit study to translate thermal capacity to performance. This capacity already has been paid for, so the community needs to fully use it.
- Develop techniques for managing the temperature profile across the entire chip to maximize performance.
- Develop techniques for managing the temperature profile across a rack, or several racks, to maximize performance. The rack- and room-scale thermal capacity can be effectively used for load balancing to maximize performance for the hardware footprint.
- Begin to explore systems that employ various forms of three-dimensional (3D) and 2.5-dimensional packaging. Such packaging technologies will be a new driver for system design.
- New devices and materials that are inherently more energy efficient and have greater immunity to thermal issues also need to be investigated. For example, silicon carbide (SiC)-based electronics can withstand substantially higher temperatures, and nanomagnetics-based devices have zero leakage power.
- Create a modeling challenge that specifies the integration of thermal models with compute models to achieve a specific modeling goal. Identify metrics for researchers to try to optimize.
- Citing an effort at the University of Chicago, Hillery Hunter, a senior manager, computer architecture and memory strategist with IBM Corp.'s TJ Watson Research Center, suggested taking processor and dual in-line memory modules (DIMMs), putting thermocouples on the memory, creating a model, and validating said model with measurements. Researchers then could report their work with respect to the model.
- Dolores Shaffer, of Science and Technology Associates, suggested that now is a good time for the integration of thermal into computer system modeling because the interest in and early use of 3D packaging requires system architects to think more about cooling.
- Motivate and support the establishment of thermal measurement infrastructure at the chip, package, rack, and room scale to 1) validate multiscale thermal models and 2) develop performance-thermal co-design models, e.g., to support schedulers. Ideally, such an infrastructure would enable coordinated measurement across the full stack, chip- to rack-scale, through multiple software layers.
- Develop techniques for integrated modeling of performance, reliability, and temperature.

5.0 Conclusion

ModSim 2015 examined traditional modeling and simulations of power, performance, and reliability that impact computing architectures, applications, and system software. For the first time, we also included ModSim of the physical and architectural phenomena affecting thermal management in systems at all scales. Understanding thermal aspects helps in designing the best solutions for energy efficiency.

In addition, ModSim 2015 delved into improving predictive models to better inform algorithmic and system design decisions affecting PPR optimality. Such predictive models would capture changes in an application's performance and/or power stemming from hardware modifications or determine the integrated PPR impact of various architecture subsystems. Discussions on measurement and benchmarking informed the considerable benefits of a comprehensive strategy for applying these mechanisms to systems and workloads—both to optimize existing HPC systems and to actualize future system designs. Measurement and benchmarking serve a dual purpose with respect with ModSim. Benchmarks can be input to a multitude of ModSim classes, and they also help in verifying models to improve their accuracy and validate them for predictive use in design purposes.

As in past workshops, it generally was agreed that to enable a sustainable view and comparison of system capabilities that can evolve with growing and diverse HPC activities, ModSim will continue to be an essential driver in co-designing viable high-performance computing systems and applications. With each successive workshop, we have seen ModSim grow, gel, and become identifiable as its own distinct technical community.

This unique forum benefits application, architecture, and system software, and hence the HPC research constituency. It has helped the community chart a distinct course aimed at accelerating ModSim's progress and impact on important, practical challenges. It has inspired us to seek co-design methods that analyze future algorithms, applications, hardware, and software and can be engaged throughout system and application lifetimes. Such innovations are recognized as elemental to ModSim's full adoption in designing, analyzing, and optimizing future numerically and data-intensive systems. This aspect is even more conspicuous with the impending launch of the Exascale Computing Project and other programs stemming from the National Strategic Computing Initiative.

.

6.0 References

Fleury T. 2015. "Battling the Heat in Military CPUs." *Military Embedded Systems* 11(8):34-35. Available online at: http://mil-embedded.com/articles/battling-heat-military-cpus/.

O'Mara D. 2011. "Complex Systems Poses Tricky Thermal Management Challenges." *COTS Journal: The Journal of Military Electronics and Cooling*. 13(4):30-36. Available online at: http://www.cotsjournalonline.com/articles/view/101830.

Sreepathi S, ML Grodowitz, R Lim, P Taffet, PC Roth, J Meredith, S Lee, D Li, and J Vetter. 2014. "Application characterization using Oxbow toolkit and PADS infrastructure." In *Proceedings of the 1st International Workshop on Hardware-Software Co-Design for High Performance Computing* (*Co-HPC '14*), pp. 55-63. November 17, 2014, New Orleans, Louisiana. DOI: 10.1109/Co-HPC.2014.11.