SUPER NEWSLETTER

March 2012

Upcoming Events:

• ESP March Workshop: "Code for Q"
http://www.alcf.anl.gov/esp-workshop-march-2012
March 19-21, 2012
Argonne National Laboratory
Shirley Moore, Sameer Shende

• SUPER All-hands meeting
March 29-30, 2012
UNC-Chapel Hill
All SUPER participants and invited guests

• Workshop on High-Performance, Power-Aware Computing (HPPAC 2012)
http:// http://hppac12.ac.upc.edu/
May 21, 2012
Shanghai, China (co-located with IPDPS 2012)
Bronis de Supinski, Workshop Co-chair

• International Workshop on Automatic Performance Tuning (IWAPT 2012)
http://iwapt.org/2012/
July 17, 2012
Paper submission deadline: March 26, 2012
Kobe, Japan (co-located with VECPAR 2012)

• Workshop on Productivity and Performance (PROPER 2012)
http://www.vi-hps.org/proper2012ws/
August 27/28, 2012
Paper submission deadline: June 4, 2012
Rhodes Island, Greece (co-located with Euro-Par 2012)
Shirley Moore, Steering Committee

From the Director

The overarching goal of SUPER is to ensure that DOE's computational scientists can successfully exploit the emerging generation of high performance computing (HPC) systems by providing strategies and tools to productively maximize performance, conserve energy, and attain resilience. We are currently six months into the project, and we have gotten off to a great start. Our management structure of breaking the project up into several thrust areas with different leaders is working well. We have a full conference call schedule, with one all-project call at the beginning of each month and a total of seven technical calls each month. I know from my participation in most of these. Our website is up and running, containing information about the project and the different thrust areas and featuring recent publications and accomplishments. We have an internal wiki with technical discussion areas for each of the thrust areas. Notes from all the conference calls are also kept in these discussion areas.

This month's newsletter features the energy thrust area, led by Laura Carrington at SDSC. This area is making excellent progress, as you can see from this month's articles on research in this area at SDSC, RENCI, and UTK. Managing the energy consumption of HPC systems is crucial in order to keep energy consumption from becoming the limiting factor in scaling scientific applications to massively parallel levels. The work that SUPER is doing in developing tools and techniques for measuring, modeling, and optimizing power consumption will play an important role in this effort. Furthermore, integration with the performance and overall optimization thrust areas of SUPER will enable energy consumption and performance to be optimized together and will help guide policies for achieving the best overall usage and scaling of future HPC systems.

- Bob Lucas

SUPER at SC'11

SUPER had a big presence at SC'11. We started off the week with a half-day all-hands meeting. During the week, SUPER researchers were busy with tutorial, paper, poster, and booth presentations. See www.super-scidac.org/sc11 for a list of SUPER-related events that took place at SC'11.

Bob Lucas leading the SUPER all-hands meeting

Allen Malony teaching a tutorial on Scalable Heterogeneous Computing on GPU Clusters

Jack Dongarra giving the traditional Top500 talk

Sameer Shende giving a TAU demonstration

SUPER Spotlight on Energy

written by Laura Carrington with help from Rob Fowler

The overall goal of the SUPER Energy Thrust is to develop tools to aid in the investigation of energy consumption of HPC applications and utilize these discoveries to develop tools and methodologies to reduce the energy consumption of HPC systems. To characterize and improve the energy costs of computing the Energy Thrust has three main focus areas: power measurement devices, instrumentation tools and models to connect the measurements to the applications, and the development of methods to reduce energy. RENCI has taken the lead in the area of power measurement devices with their PowerMon device[1]. Their devices are installed at two SUPER collaborating institutes (UCSD/SDSC and UTK). UTK has been working both with RENCI and SDSC to build a PAPI component to give users a familiar API to communicate with the PowerMon device. SDSC has set up a test system for the Energy Thrust with PowerMon and WattsUp devices attached. Soon a SDSC Intel Sandybridge test system will be available to the group as part of the *Gordon* supercomputer project; it will enable SUPER investigation into the efficacy of the power counters available on that processor to finely attribute power consumption. SDSC and UTK are focused on research and tool development in the area of reducing energy consumption via application-specific find-grained and coarse-grained Dynamic Voltage Frequency Scaling (DVFS) policies[2-4]. UTK is studying how slack times in DAG-based scheduling algorithms can be used to achieve power savings. In researching these policies, it is important to be able to attribute work done by the application to the power usage of various hardware components.

Along these lines one current area of focus at SDSC is in energy profiling. The goal of this work is to build a set of tools that are capable of attributing the energy used by an application to its lower-level code constructs (functions, loops, basic-blocks). Such a tool could be used the same way that data from a tool like gprof is used when examining performance – it could be used in guiding energy-related optimization efforts and also serve to better understand what factors affect application energy usage. The tool works by gathering the following two pieces of information during application execution:

Upcoming Events (cont):

- SC’12

http://sc12.supercomputing.org/
November 10-16, 2012
Paper abstracts due: April 20, 2012
Full papers due: April 27, 2012
Salt Lake City, Utah
David Bailey and Jeff Hollingsworth, Steering Committee

Super Spotlight on Energy (cont.)

Fine-grained (1000 per second) component-specific hardware information (processor, DIMMs, Motherboard, disks, GPUs) energy measurements using the PowerMon [1] device.

Counts of certain low-level program events associated with software including program control constructs (function and loop entry/exit, basic-block execution, etc.) that occur during each energy measurement period using SDSC's binary analysis and instrumentation package PEBIL [5].

Initial implementations of these measurement gathering tools have been completed, and we are currently evaluating some basic numerical techniques (direct and iterative linear solvers/ estimators) to determine the accuracy and stability of the tools and the resulting profiles that can be derived.

The second area of focus at SDSC has been in connecting the DVFS energy-efficiency policies with auto-tuning frameworks, which have traditionally focused on performance. In this work, power draw and execution time data obtained empirically for a small set of compute intensive kernel variants were used to develop CPU and DIMM energy consumption models. The models that we developed are very accurate; the maximum absolute average error between the measured and modeled values is less than 5.5% for three important kernels – matrix multiplication, stencil, and LU factorization.

These models were then used to rapidly explore the energy behavior of these kernels when subjected to various compiler-based optimizations and DVFS settings, automated through the use of a PERI (precursor to SUPER in Scidac2) developed auto-tuning framework (e.g. Active Harmony (UMD) and GCO (UTK)). The search uncovered kernel variants that had 7.7% more energy savings than the kernel variants originally used to construct/train the models. Such exploration without the models (i.e. actual execution and measurement of the kernel variants) would have taken approximately 42 days whereas the analysis done using the models took under a minute. This illustrates the potential of our model-based approach and one of several ways the methodology can be used in the exploration of reducing the ever growing energy costs of HPC.

1. Bedard D, Min Yeol L, Fowler R, Porterfield A: **PowerMon: Fine-grained and integrated power monitoring for commodity computer systems**. *IEEE SoutheastCon 2010 (SoutheastCon), Proceedings of the* 2010:479-484.
2. Laurenzano M, Meswani M, Carrington L, Snavely A: **Reducing Energy Usage with Memory and Computation-Aware Dynamic Frequency Scaling**. In: *Euro-Par 2011: 2011; France*; 2011.
3. Tiwari A, Laurenzano M, Carrington L, Snavely A: **Modeling Power and Energy Usage of HPC Kernels**. *in submission High Performance Power-Aware Computing (HPPAC12)* 2012.
4. Tiwari A, Laurenzano M, Carrington L, Snavely A: **Auto-tuning for Energy Usage in Scientific Applications**. In: *Proceedings of Workshop on Productivity and Performance (PROPER 2011): 2011; Bordeaux, France*; 2011.
5. Laurenzano M, Tikir M, Carrington L, Snavely A: **PEBIL: Efficient Static Binary Instrumentation for Linux.** In: *IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS): March 2010; White Plains, NY*; 2010.

Recent Software Releases :

TAU 2.21.1 release

December 14, 2011

(http://www.super-scidac.org/tau-2.21.1)

PAPI 4.2.1 release

February 14, 2012

(http://www.super-scidac.org/papi-4.2.1)

Component PAPI Moves Toward Energy Measurement

Written by Dan Terpstra, James Ralph and Vince Weaver

Traditionally PAPI, the Peformance API from the Innovative Computing Laboratory at the University of Tennessee, has provided fine-grained measurement of hardware performance counters found on cpu cores. With the introduction in 2010 of Component PAPI, it is now possible to extend these performance measurements to a variety of other domains, including such devices as GPUs and network interfaces. As part of the SUPER Energy Thrust, the PAPI team is working with researchers at RENCI and UCSD to extend the PAPI interface into the domain of energy measurement as well.

Work is underway to create PAPI components to interface to the PowerMon2 hardware from RENCI and the commercially available WattsUp?PRO power meter. With the addition of these power measurement components, users and tools will be able to collect simultaneous data on power and energy usage from multiple sources, along with performance data from CPUs, GPUs, IO devices and more.

The WattsUp?PRO component will allow measurement of AC wall power at a frequency of once a second. A USB connection is used to control and monitor the device. Given the external and asynchronous nature of this device, the PAPI interface is being stretched in novel directions. A worker thread will interface with the device itself, with PAPI collecting measurements from that thread. Energy calculations can be performed by the worker thread to provide both power and cumulative energy readings at fairly coarse granularity.

The PowerMon2 component is much more tightly coupled to the hardware. It provides measurements of DC voltage and current on 8 separate channels at sampling rates up to 3 kHz. In collaboration with RENCI, future versions of firmware will allow direct reading of energy from each channel as well. Like the WattsUp?PRO component, the PowerMon2 component is providing design challenges for the PAPI interface. Historically PAPI data was always monotonically increasing unsigned integer counts. With new measurement types comes a need for new data types as well, including instantaneous values such as power, arrays of values collected into a buffer with a constant time steps, and units such as joules and watts. A simple pseudocode example below demonstrates how the PowerMon2 PAPI component can be monitored in a user application:

```
PAPI_library_init( PAPI_VER_CURRENT );
PAPI_create_eventset( &EventSet );
PAPI_event_name_to_code( “sensor0.voltage”, &event);
PAPI_add_event( EventSet, event);
PAPI_event_name_to_code(“sensor0.current”, &event );
PAPI_add_event(EventSet, event);

PAPI_start( EventSet );
for ( ;; ) {
        < application inner loop >
        PAPI_read( &values );
        Printf(“instantaneous voltage,current (mV, mA )
                %lld, %lldt”, values[0], values[1] );
}
```

Featured SUPER Researcher: Ananta Tiwari

Where do you work and how are you involved with SUPER?

I am a Senior Computational Scientist at the San Diego Supercomputer Center (SDSC)/UCSD. At SDSC, I am affiliated with the Performance Modeling and Characterization (PMaC) lab. Dr. Allan Snavely and Dr. Laura Carrington lead the PMaC lab.

For the SUPER project I am a co-PI for UCSD, the lead institute of the energy-efficiency research which is the focus of my work.

Can you briefly summarize your educational and work background?

I received my PhD degree in Computer Science from the University of Maryland at College Park in January 2011. My PhD advisor was Dr. Jeffrey K. Hollingsworth. After receiving the degree, I joined SDSC.

Where are your from originally?

I am originally from Nepal.

What are your research areas of interest?

I am primarily interested in understanding and characterizing the factors that affect the energy usage and performance of current and future high performance computing applications and platforms. Given that energy-related constraints have emerged as one of the major design impediments for the next generation exascale systems, I believe that it is crucial to develop these characterizations and use them to build tools and methodologies to reduce energy requirements of HPC systems. Along these lines, my most recent work modeled the interactions between well-studied compiler optimization strategies from PERI auto-tuners and the energy consumed by CPUs and DIMMs. The models are trained using empirical data gathered for a very small portion of the points in the optimization parameter space. These models can provide valuable information related to the shape of the optimization parameter space to the heuristics-based auto-tuners.

What do you see yourself doing five years from now?

I see myself continuing my current energy efficiency and performance auto-tuning research that seeks to maximize the science that can be done per unit of energy spent to run the computing platform. The energy, concurrency, storage and resiliency challenges that have to be addressed on our way to the exascale era are daunting. That said, the way the HPC community has come together to tackle these challenges is very encouraging. Five years from now, I believe that we will have made significant progress towards deploying an energy-efficient exascale system.

What are some things you enjoy doing that don't involve computers?

I like spending time with my family and reading fiction novels. I am also an avid fan of the National Football League. During football season, I spend my Sundays watching the games.

Selected Recent Publications

- Teng Ma, George, Bosilca, Aurelien Bouteiller, and Jack Dongarra. "HierKNEM: An Adaptive Framework for Kernel-Assisted and Topology-Aware Collective Communications on Many-core Clusters," IPDPS 2012, Shanghai, China, May 2012, Best Paper (to appear).
- David Böhme, Bronis R. de Supinski, Markus Geimer, Martin Schulz, and Felix Wolf, "Scalable Critical-Path Based Performance Analysis," IPDPS 2012, Shanghai, China, May 2012 (to appear).
- Mitesh R. Meswani, Laura Carrington, Didem Unat, Allan Snavely, Scott Baden, Stephen Poole, "Modeling and predicting application performance on hardware accelerators," IEEE International Symposium on Workload Characterization, November 2011.
- Nick Rutar and Jeff Hollingsworth, "Data Centric Techniques for Mapping Performance Data to Program Variables," Parallel Computing38(1-2), Jan. 2012.
- Arthur A. Mirin and Patrick H. Worley, "Improving the performance scalability of the community atmosphere model," International Journal of High Performance Computing Applications 26:17-30, February 2012.
- K. Madduri, J, Su, S. Williams, L. Oliker, S. Ethier, K. Yelick, "Optimization of Parallel Particle-to-Grid Interpolation on Leading Multicore Platforms", IEEE Transactions of Parallel and Distributed Systems, accepted.
- Peng Du, Aurelien Bouteiller, George Bosilca, Jack J. Dongarra, and Thomas Herault, "Algorithm-Based Fault Tolerance for Dense Matrix Factorization," PPoPP 2012, New Orleans, LA, February 2012.
- Chun-Yi Su, Dong Li, Dimitrios Nikolopoulos, Mat Grove, Kirk Cameron and Bronis de Supinski. "Critical Path-Based Thread Placement for NUMA Systems", ACM SIGMETRICS Performance Evaluation Review 40(2), 2012.
- Dong Li, Bronis de Supinski, Martin Schulz, Dimitrios Nikolopoulos, and Kirk Cameron. "Strategies for Energy Efficient Resource Management of Hybrid Programming Models", IEEE Transactions on Parallel and Distributed Systems, 2012 (to appear).
- Dong Li, Jeffrey Vetter, Gabriel Marin, Collin McCurdy, Cristian Cira, Zhuo Liu, and Weikuan Yu, "An Analysis of Scientific Applications for Using Non-Volatile Memory in High Performance Computing," IPDPS 2012, Shanghai, China, May 2012 (to appear).
- Rountree, Barry, Dong H. Ahn, Bronis R. de Supinski, David K. Lowenthal and Martin Schulz, "Beyond DVFS: A First Look at Performance Under a Hardware-Enforced Power Bound," Eighth International Workshop on High Performance Power-Aware Computing (HPPAC 2012), Shanghai, China, May 21, 2012.

See the SUPER website for additional recent publications

SUPER website: http://www.super-scidac.org/
Contact: Bob Lucas, rflucas@isi.edu
Newsletter editor: Bonnie Browne, bonnie.e.browne@gmail.com