The Institute for Sustained Performance, Energy, and Resilience
University of North Carolina
Final Technical Report

Federal Agency: DOE/Office of Science: Advanced Scientific Computing Research

Cooperative Agreement Identification Number: DE-SC0006925

Project Title: The Institute for Sustained Performance, Energy, and Resilience (SUPER)

Principal Investigator:
Robert J. Fowler,
RENCI Director of High Performance Computing,
rjf@renci.org, rjf@unc.edu, 919 445 9670

Submission Date: 19 Jan, 2018

DUNS Number: 608195277

Recipient Organization:
University of North Carolina at Chapel Hill
Attn: Robin Cyr
Administrative Office Building, Suite 2200
104 Airport Drive CB# 1350
Chapel Hill, NC 27599-1350

Project Grant Period: 09/01/2011-08/31/2017

Reporting Period Date: 09/01/2011-08/31/2017

Reporting Frequency: Annual

1 Overview

In the Scientific Discovery through Advanced Computation program (SciDAC-3), 2011–2016, with no cost extensions into 2017, computational scientists working on behalf of the Department of Energy's Office of Science (DOE SC) exploited a new generation of petascale computing resources to make previously inaccessible discoveries in a broad range of disciplines including physics, chemistry, and materials science. The computing systems underpinning this work are increasing from tens to hundreds of peak PFlop/s, but have also evolved significantly from architectures in 2011, when SciDAC-3 began. Although Moore's law continues unabated, the end of Dennard scaling has caused a fundamental shift in computer architecture towards power efficiency [7]. To that end, processors are increasingly diverse as they strive to satisfy performance, productivity, reliability, and energy efficiency in the face of divergent computational requirements. The diversity among these machines makes it difficult to port today's scientific applications, much less achieving good performance. Thus, by the end of the project, vastly increased scale and heterogeneity have exacerbated performance optimization challenges while simultaneously promoting the issues of energy consumption and resilience to the forefront.

To promote successful use of emerging high performance computing (HPC) systems, by DOE's computational scientists, DOE SC's Advanced Scientific Computing Research Program (ASCR) funded the SciDAC-3 Institute for Sustained Performance, Energy, and Resilience (SUPER). This was a broadly-based project with expertise in compilers and system tools, performance engineering, energy management, and resilience. As in the previous SciDAC-2 Performance Engineering Research Institute (PERI) we pursued a strategy of leveraging the previous research investments by DOE and others. We have integrated the results to create new capabilities beyond the reach of any one group. To accomplish this, SUPER was organized to conduct the following research activities:

1. Performance portability: We extended performance measurement and autotuning technology to petascale and heterogeneous systems, thus permitting scientists to exploit a wide range of high-end systems from a common code base.
2. Energy efficiency: Relatively minor code changes can result in significant energy savings for some applications, with little or no impact on performance. We are investigating software energy efficiency techniques to help reduce DOE's energy costs.
3. Resilience: Petascale calculations are pressing the limits of reliability both in hardware and system software. We explored strategies to enable petascale applications to be resilient in the face of faults.
4. Optimization: We extended tools from the mathematical optimization community to develop strategies that collectively optimize performance, energy efficiency, and resilience.

SUPER researchers collaborated with the other SciDAC-3 institutes, with at least 13 of the 19 Scientific Computation Application Partnerships (SAP), and with members of the broader DOE computational sciences community. The collaborations focused on the real challenges facing petascale scientific computing as well as ensured broad and immediate impact from our research. Therefore, SUPER teams conducted the following engagement and integration activities:

1. Application engagement: We provided measurement and analysis tools as well as performance expertise to help application developers optimize performance of their codes, and to gauge whether closer collaboration in SUPER research activities would be useful. We worked with other SciDAC-3

institutes to analyze and to optimize their software for use in application codes.

2. Tool integration: We improved tool integration to enable end-to-end application optimization, including automated experiments for performance analysis and autotuning; power-aware instrumentation for energy evaluation, fault detection and vulnerability analysis; and scalable data management and presentation.

2 Activities at University of North Carolina

In this report, we describe activities on the SUPER project at RENCI at the University of North Carolina at Chapel Hill. While we focus particularly on UNC, we touch on project-wide activities as well as, on interactions with, and impacts on, other projects.

2.1 Architecture Awareness: RCRToolkit

Node-wide performance measurement and tuning have become critical, given current trends towards more cores per chip with less memory and interconnect bandwidth per core. Almost all applications are bound by either memory latency or memory bandwidth. Latency issues can be mitigated using code- and data-reorganization, whereas bandwidth problems require a more fundamental re-evaluation of the computational approach. Measuring bandwidth and other node-wide resources is therefore crucial.

UNC developed, then deployed, Resource Centric Reflection tools (RCRToolkit) to provide node-wide measurements and utilization models for shared resources (memory, interconnects, energy) on computer architectures of interest to DOE. Most of the previous performance-tool suites such as PAPI and the tools based upon it (HPCToolkit, TAU, etc.) historically presented a "first person" view that measures and accounts for performance based on performance-sensing from within a single application thread. These tools only measure events that occur while the thread is actually running. Collecting consistent information requires a node-wide "third-person" approach to performance monitoring. Based on our experience, TAU has incorporated aspects of the third-person approach based on interactions with RCRToolkit. PAPI has also been extended to collect some node-wide information.

RCRToolkit implements node-wide collection to support off-line analyses suitable for inclusion in the SUPER off-line autotuning framework, as well as on-line analyses to guide introspective adaptation at several levels: operating system kernel, user-level thread packages, and adaptive applications. An important goal is to couple the node-wide tools to existing "first person" performance tools to provide an integrated measurement and analysis framework. We modified the Rice University HPCToolkit to query the RCRToolkit blackboard data structure in shared memory and to use this information as a discriminating predicate to classify sampled events. We applied this to the USQCD Chroma code to distinguish L2 cache misses according to whether they occur when the memory system is fully utilized from those that occur when it is not [17]. We found that only one of the four major loops in the benchmark is bandwidth-limited; the others are, suprisingly, latency-bound.

We collaborated with SUPER researchers at the University of Oregon to couple RCRToolkit with TAU [13, 14] in the context of the XStack XPRESS project. The deployment of this tool on current leadership-class systems for applications of current interest is being funded through SUPER.

RCRToolkit currently runs on both stand-alone nodes and clusters based on multiple generations of Intel Xeon and Xeon Phi processor chips. The source has been placed in a Subversion repository that has been open to our research partners in academia and the National Laboratories since the end of April, 2014. We have also packaged a version of RCRToolkit as a SLURM plugin. This places RCRToolkit under the control of SLURM to start and stop the daemon as well as to manage data for offline analysis and to clean up when the job using RCRToolkit finishes.

We also couple RCRToolkit with cluster- and system-wide monitoring frameworks. On commodity systems, these frameworks include Nagios and PerfSonar to provide real-time introspection data driving system-wide adaptive management. A prototype of coupling arrangementg has been used for I/O performance studies [18, 19, 26, 27].

2.2 Application Awareness: Power and Energy

We previously redesigned and built a dozen units of our prototype PowerMon device to be suitable for wide-scale and flexible deployment, including in systems with accelerator devices. PowerMon2 [3] data is provided through a USB interface either to the monitored system or to an external system. Within the scope of SUPER, PowerMon2 has been used both by us and our collaborators [1, 2, 15, 29] to evaluate the power/energy behavior of systems and applications as well as to evaluate approaches to improving them. Led by researchers at Georgia Tech, we used PowerMon2 as the the basis for developing a roofline model for performance and energy tradeoffs [8]. PowerMon2 also has been used by staff at SNL as the inspiration for the PowerInsight device that they designed in conjunction with Penguin Computing and to help calibrate/validate the device [9, 16].

Using RCRToolkit as the common software infrastructure for monitoring power and energy, we are pursuing studies of introspective monitoring of energy consumption and of the adaptive chip management controllers of recent Intel processor chips. The tool has been used to study intrinsic temperature, energy, and performance variability in a nominally homogeneous cluster [22], to provide feedback for compiler-based autotuning[30], and to guide adaptive application-level scheduling [12, 22, 23]. RCRToolkit has also been used to characterize I/O performance [18]. It is being used currently as the sensing/measurement component of our research on dynamic runtime power/energy management.

Controlled environments for benchmarking and tuning of programs on large shared clusters are becomming more difficult to achieve. Experiments [22] with repeatedly running benchmarks as though in an auto-tuning context highlight substantial differences in energy consumption due to increases in parasitic currents between a relatively cool system and one that is running at full operating temperature. Furthermore, the temperature differences between chips due to intrinsic differences in fabrication, or on the cooling available at different physical locations, add sources of variation. Performance variablity from run to run or between chips is also measurable but is often smaller. If TurboBoost is enabled, the potential differences are much greater. Experimental frameworks for code optimization either need to control this variability or to compensate for it. This work highlights the difficulty of attempting to autotune using an ensemble computing approach on parallel systems.

We have continued to characterize the causes and magnitudes of these variations as well as to design adaptive run-time methods for detecting and compensating for node-to-node variation in MPI programs [14, 24, 25]

Sridutt Bhalachandra's Ph.D. dissertation research [4], supported by SUPER, focuses on transparent methods of improving the power and energy consumption for supercomputing applications with minimal impact on computational speed.

Recent processors equipped with on-board hardware counters allow real-time monitoring of operating conditions such as energy and temperature, in addition to performance measures such as retired instructions and memory accesses. Significantly, recent processors also provide the ability to control processor power utilization dynamically at the per-core level. Bhalachandra *et al.* [6] presented an experimental study using RCRToolkit on systems based on Intel Sandybridge and Haswell to identify bandwidth-constrained phases for each rank of MPI programs to then reduce or modulate CPU frequency transparently on a per-core basis

in those phases, saving energy with minimal performance impact. This was implemented using the pMPI interface and, thus, is transparent to the applications. The policy is evaluated when applied at coarse and fine grained levels on six MPI mini-applications. The best energy savings with the coarse and fine-grained application of the dynamic policy were 32.1% and 19.5% respectively with a 2% increase in execution time in both cases. On average, the fine grained dynamic policy led to a 1% speedup while the coarse grained dynamic policy caused a 3% slowdown. Energy savings through frequency reduction provide cost advantages, but they also reduce resource contention. They and create additional thermal headroom for non-throttled cores that can lead to improved performance.

A similar approach using runtime MPI instrumentation identifies MPI ranks that consistently arrive early at MPI collective operations [5]. Such ranks are clearly not on the critical path of the application so their clocks can be dynamically slowed or modulated to save power and energy without a performance penalty. When there are multiple ranks sharing a single multicore CPU, the reduced energy demand of the throttled cores frees up thermal headroom that can be used by the non-throttled cores on the critical path to increase their clock rates using turbo-boost.

Both of these dynamic policies are especially effective on systems to which an explicit power constraint has been applied. The more constrained the power, the more effective the methods become. This is expected to become more relevant for future exascale systems.

2.3 Architecture Awareness: End-to-end performance and resilience.

End-to-end management of ensemble and workflow computations affects all of SUPER's major research. Failures causing the re-running of ensemble and workflow elements manifest themselves as energy and performance problems. In addition to hardware faults that may corrupt results without bringing down an entire compute node, there are many sources of failure that must be dealt with at a higher level. Large-scale ensemble campaigns must be prepared to deal with compute-node failure, networking and I/O interruptions, system software problems, algorithmic bugs, and failure of the scientific models when presented with inputs outside their domain of applicability. SUPER researchers at UNC have pioneered autonomic computing methods that deal with failure of an ensemble member in any of these situations.

Depending on the source and other characteristics of the problem, the proper reaction may be to rerun the program on different hardware, to rerun it with an alternative input, or to alert the end user. In the exploration of high-dimension parameter spaces, such as the uncertainty quantification (UQ) campaigns run by researchers in QUEST, the successful completion of other calculations may render the failed run uninteresting, and, thus, not worth rerunning. Collaborations with QUEST to support resilient UQ campaigns was one of the anticipated targets for this work. In conjunction with this effort, UNC researchers also profiled parts of Dakota to identify and improve hot spots.

This work is continuing in coordination with the Pegasus group at ISI and with the RENCI networking research group. Versions have been demonstrated in the UNC/RENCI booth at SC11, SC12, and SC13. The demonstration involved a distributed ensemble running on NERSC Hopper, on RENCI's internal cluster, and on a cluster at Duke. The driving application is a density functional theory (DFT) computation used to screen a large number of crystalline materials for solar energy applications in support of the UNC Solar Fuels EFRC. In addition to reserving computational resources to run the ensemble rapidly, bandwidth was reserved on ESNET and Internet2 links between RENCI and NERSC. See Figure 2.1.

The resilient workflow-management framework has been deployed on top of the NSF ExoGENI experimental cloud architecture, one end-point of which is located at NERSC. A computational campaign reserves a "slice" of resources which includes computational nodes, storage, and communication bandwidth among

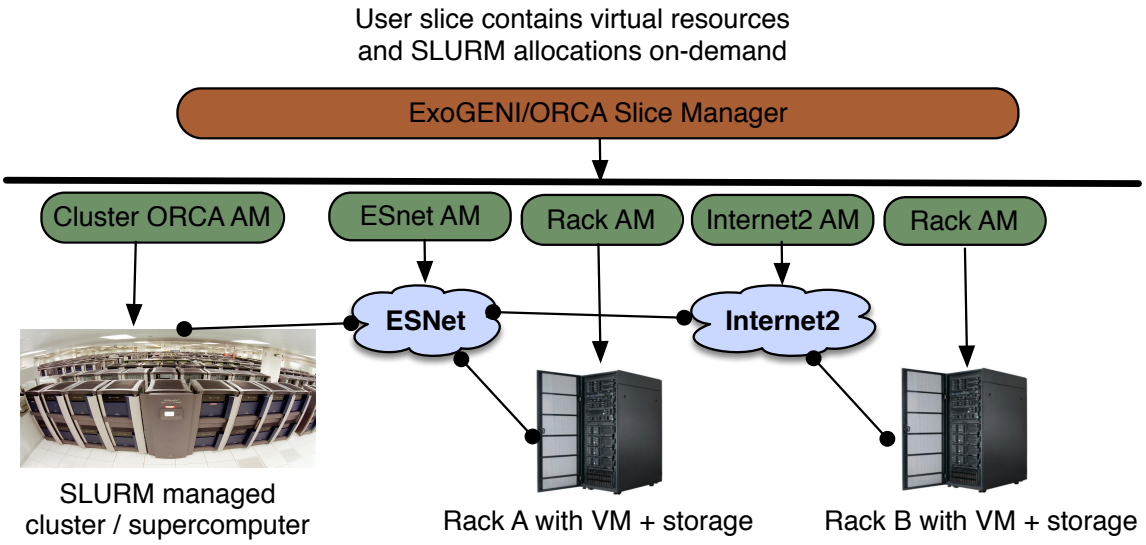

Figure 2.1: SLURM partition incorporated in an EXOGENI slice.

the distributed sites participating in the campaign. To extend the slice onto large clusters, we have prototyped a GlideIn manager and scheduler that will reside within a machine partition allocated by SLURM. The management layer uses SLURM system health monitoring facilities as well as self-reporting by the application to determine when it is necessary to react to existing or impending failures. The partition has a small number spare nodes that can be swapped in, and the design has hooks into SLURM to facilitate moving to an entirely new partition if the set of spares is inadequate. The prototype was demonstrated at SC13 and SC14.

Our goal has been to enable seamless integration of static infrastructure (like clusters and supercomputers) and dynamically provisioned NIaaS resources (slices). We are working to extend dynamically provisioned environments into and out of batch controlled static resources (in bursting) and (out bursting) , respectively, depending on the needs of workow applications. This requires integration of of data movement from slices to static infrastructure and vice versa. In the architecture, the Network Infrastructure as a Service (NIaaS) resource provisioning handler communicates with the underlying resource manager (SLURM, Torque/PBS etc.) to create allocations for workflows, provide semantics for glidein, and manage allocations for workow ensemble members, with particular emphasis on providing mechanisms for reliable resource control.

2.4 Application Awareness: Solar Materials

In the early stages of SUPER we continued work begun under PERI and a SciDAC-e project in support of the search for catalysts for solar fuels. This application drove system software debugging efforts, optimization exercises, and is serving as a driving application for improving the capabilities of performance tools.

The parallel spin-orbit configuration interaction (PSOCI) chemistry code developed in part under the SciDAC-e award is being applied to larger problems in electronic structure theory. The goal is a problem of size 500 million double group adapted functions, which is fairly large. Improvements to the code were performed on Hopper at NERSC and on Jaguar at ORNL (made possible through the PEAC INCITE project). To accomplish the chemical goals requires not only computational efficiency but faster communications. PSOCI uses Global Arrays (ga++, v5) for managing the data layer. The size of this problem led to the fixing

of several problems in the Global Arrays library. The science results from this effort are reported in [11, 28]. The use of PSOCI and DFT calculations to drive resilient end-to-end ensembles is reported in [20, 21, 26]

The work initiated under SUPER is continuing under the DOE PANORAMA project [10]. A current driving problem for this work is the analysis of data from the ORNL Spallation Neutron Source.

2.5 Application Awareness: Nuclear and High Energy Physics

SUPER tools were applied in our work as part of the *Computing Properties of Hadrons, Nuclei and Nuclear Matter from Quantum Chromodynamics* SciDAC project. We worked closely with USQCD physicists at JLab to improve the performance of the QDP++ library and the Chroma application. QDP++ is a domain-specific programming language for lattice computation implemented in C++ using template metaprogramming. it is a powerful environment for prototyping LQCD applications and augmented with carefully hand-tuned modules it is fast enough for production. Prototyping is crucial because while tuning existing code is important, the really big gains in scientific productivity in LQCD over the past years are due to new methods, both mathematical and in modeling and discretizing the underlying physics. Our goal was to lessen the dependence on hand-tuned code, thus reducing what's been called the "ninja programmer gap". Our approach is to use the expression template mechanism as a just-in-time (JIT) code generator to target architectures including Nvidia GPUs, Xeon PHI, and conventional multi-core systems. SUPER researchers are providing performance analysis as well as static compiler support to improve code generation and data movement between separately instantiated expressing templates.

A major part of the effort in 2015 focused on QPhiX, a framework compatible with QDP++ and Chroma for computing the Wilson Dslash and Dslash Clover operators and for embedding them in high-performance solvers. Originally written by JLab and Intel, QPhiX achieves nearly 500 GFLOPS on high-end Intel Phi (Knights Corner) accelerators. When backported to mainline Xeons, the package can obtain nearly 160 GFLOPS for Dslash, a factor of nearly three over the previous production code. UNC ported QPhiX to the BlueGene/Qs at ANL. This is currently the target of an optimization effort in that the prefetch mechanisms on the BG/Q require a different strategy than was used for the Intel Phis. A collaboration among JLab, Intel, and UNC worked on getting code ready for Knights Landing and Knights Hill systems. This will include refactoring the methods used in QPhiX to improve performance portability between systems, among applications, and to other phases of LQCD calculations. A driving use case was porting the methods to the MILC code base.

SUPER researchers are providing performance analysis as well as compiler support to improve data layout, code generation, and data movement between separately instantiated expression templates. These methods are also being applied to make the hand optimized code modular in support of performance portability and for reuse for new physics analyses.

GEANT4 simulates the interactions of physics particles with matter. It is based on the underlying principles of the physics involved with minimal approximations, so it is the standard against which other simulations are judged. At its peak, the physics community was reported to be spending one hundred million dollars a year worldwide on GEANT4 including acquiring and operating hardware and facilites, and storing the results. SUPER was asked to examine the code and to participate in a re-engineering effort in collaboration with researchers and developers at CERN, FermiLab, SLAC, and other facilities.

As part of the GEANT4 re-engineering effort, UNC researchers analyzed benchmark version of the Compact Muon Solenoid (CMS) simulator. The program executes approximately one instruction per clock cycle. This leaves considerable headroom for improvement, especially since much of this is overhead due to the highly modular structure of this C++ code. Memory cache analysis indicates acceptable miss rates for both

L1 and L2 caches for both data and instructions. While some data structures and regions of the code have relatively high miss rates, these make only a small contribution to overall performance.

Our analysis has led to several local improvements in the code resulting in performance improvements of several percent production code for the CMS experiment. There are, however, no glaring hotspots or points of obvious local inefficiencies. Instead, the execution time is diffused across many modules and is distributed along deep call chains to C++ methods.

Profiling using HPCToolkit helped to identify a call tree that uses between 12 and 22 percent of the wall clock time. This is rooted in the cross-section calculation part of the code. This module first calculates the probability of an interaction as a particle crosses a particular volume. Then, if a hadronic event actually occurs, it is necessary to identify exactly the kind of nucleus with which the particle interacts. These computations are based on the underlying isotopic populations for the material that occupies the volume. For instance, the stainless steel alloy used in the construction of the CMS detector at CERN is composed of at least four isotopes of iron, three of carbon, two of chromium, and one of manganese. The calculation for each potential interaction thus aggregates the results of ten or more separate calls that re-compute the cross section for each isotope. The composition of the material does not change during the simulation and the detailed per-isotope information is needed only after it has been determined that an interaction actually occured.

An existing global single-entry cross-section cache was ineffective, so we introduced a separate single-entry cache for each of the approximately 50 physics processes known to GEANT4. For a representative example, our improved cache is hit on 17% of the calls, accounting for 29% of the cycles used in this section. At least 10% of the total execution occurs here, so this reduces the production run time by at least 2.9%.

Of more than 18000 combinations, only nine particle/material pairs generated over 50 percent of the cross-section calls, and eighty pairs covered 90 percent. We therefore developed a fast surrogate model for common interactions between particles and composite materials. Each surrogate is built by querying the existing cross section functions for the set of relevant isotopes, and then creating a sparse, unbiased approximation for the composite material. The error tolerance for the approximation is set by the physicist sing the code. The slower calculation based on isotope populations is then done only when interaction actually occurs.

A version of the surrogate model was integrated with GEANT5 and has shown an improvement in run time over 8 percent for the CMS benchmark. Becuase the cross section calculations are now closely approximated, but are not exactly as they were previously, the results of individual Monte Carlo trials are now slightly different.

One physics process that cannot be dealt with this way is the calculation of neutron-absorbtion cross-sections at low energies. For example, the neutron absorbtion cross section of stainless steel exhibits well over 50 resolvable peaks at energies below 300 Mev. This level of complexity is amenable to our method, so the code excludes certain physics processes at certain energies from the fast path calculation.

2.6 Institute Awareness

We held extensive discussions with researchers in the SciDAC QUEST Institute. As discussed in the main body of this report, we identified five areas for potential collaboration. Due to constrained funding and higher priority problems, these activities could not be pursued beyond the working with QUEST researchers at Sandia National Laboratory to profile the Dakota code to identify hot spots and to suggest areas for improvement.

Our work on high-performance computational campaigns that are end-to-end relisilient continues to be the most promising area for substantial future collaboration. We are continuing our discussions regarding specific applications and computational campaigns to target for this work.

2.7 SUPER Education and Outreach at UNC

Robert Lewis, a UNC-CH graduate student joined SUPER in August 2013. He worked on RCR toolkit development with an emphasis on extensions to distributed memory systems, and the Intel Phi architecture. He graduated with an M.S. in May, 2015 and took a position with IBM.

Sarah Snoeyink worked as a research assistant for a year (Oct, 2013 to Sept, 2014) between graduating with a B.S. in Math and applying to graduate schools. She prototyped the off-line construction of surrogate models for hadronic cross-sections in GEANT4.

Diptorup Deb is a UNC CS graduate student research assistant working on a framework for building highly-optimized, statically-compiled domain specific languages with an emphasis on performance portability for LQCD codes. The portability goals include (1) optimization across whole applications, not just a key solver; (2) optimization across multiple LQCD models/discretizations; and (3) optimization across generations of multiple hardware architectures. Other valuable practical contributions have included the tuning of QPhiX, porting QPhiX to be able to use the GNU, IBM, and CLANG/LLVM compiler infrastructures, and the porting of QPhiX to the BlueGene/Q. His Ph.D. research is on compiler support to improve the performance portability of "ninja" codes such as QPhiX. He is being paid through the NP LQCD SAP and a follow-on SciDAC 4 project. Graduation is expected in Summer 2018

Sridutt Bhalachandra is a UC CS graduate research assistant working on energy efficiency. While his earlier work was supported by the XStack XPRESS project, the latter part of his dissertion research was supported by SUPER. He is graduating in Spring 2018 and has taken a postdoc position at Argonne Naltional Laboratory working on the exascale project.

3 Participants and Collaborating Organizations

The following individuals were funded to participate in the project. Note that the base funding for SUPER was augmented to support some participants to work on improving the Geant4 code for simulating the interactions between particles and materials, in particular to focus on improving its performance in simulating the large detector experiments at the Large Hadron Collider at CERN. The mechanism for this augmentation was to channel the funds through ANL, which then passed funds on to the university participants through subcontracts. The subcontract at UNC is handled as a distinct account with funding and reporting periods different from SUPER. For the purposes of this technical progress report, however, we are treating the two accounts as one project.

Name: Robert Fowler
Project Role: Principal Investigator.
Contribution to the Project: Leads the research.
augmentation.
Collaborated with individual in foreign country: No
Country(ies) of foreign collaborator: N/A
Travelled to foreign country: N/A
Duration of Stay: N/A

Name: Sridutt Bhalachandra
Project Role: Graduate Research Assistant
Contribution to Project: His Ph.D. research is on runtime software methods for improving power/energy consumption without adversely affecting computational performance. In some cases, better energy efficiency results in better computational speed. He also worked on the adverse impact of socket-to-socket energy and performance variability and autotuning methods. XPRESS project and internships at Sandia and Lawrence Livermore National Laboratories.
Collaborated with individual in foreign country: No
Country of foreign collaborator: N/A
Travelled to foreign country: No
Duration of Stay: N/A

Name: Diptorup Deb
Project Role: Graduate Research Assistant
Contribution to Project: He is an "in-house customer" for the SUPER performance and power-monitoring software. This is being applied to the analysis and optimization of Lattice QCD codes under the SciDAC Nuclear Physics SAP.
Collaborated with individual in foreign country: Yes
Country of foreign collaborator: United States person visiting Great Britain
Travelled to foreign country: No
Duration of Stay: N/A

Name: Robert Lewis
Project Role: Graduate Research Assistant.
Contribution to the Project: Development, porting, and maintenance of RCRToolkit. Power measurements

and tuning.
Collaborated with individual in foreign country: No
Country(ies) of foreign collaborator: N/A
Travelled to foreign country: N/A
Duration of Stay: N/A

Name: Anirban Mandal
Project Role: Staff Research Scientist.
Contribution to the Project: Performs performance measurement and analysis experiments. Leads our research effort on resilient adaptive workflows and ensembles. He is our principal liaison with QUEST.
Collaborated with individual in foreign country: No
Country(ies) of foreign collaborator: N/A
Travelled to foreign country: N/A
Duration of Stay: N/A

Name: Paul Ruth
Project Role: Staff Senior Research Scientist.
Contribution to the Project: Performance evaluation and tuning of GEANT4. Implementation of a fast-path method for computing interaction cross sections.
Collaborated with individual in foreign country: No
Country(ies) of foreign collaborator: N/A
Travelled to foreign country: N/A
Duration of Stay: N/A

Name: Sarah Snoeyink
Project Role: Non-student Research Assistant
Contribution to the Project: Prototyped the off-line construction for hadronic cross-sections in GEANT 4.
Collaborated with individual in foreign country: No
Country(ies) of foreign collaborator: N/A
Travelled to foreign country: N/A
Duration of Stay: N/A

Name: Jeffrey Tilson
Project Role: Staff Senior Research Scientist.
Contribution to the Project: He is a computational quantum chemist. He worked on the chemistry workflows for solar materials evaluation.
Collaborated with individual in foreign country: No
Country(ies) of foreign collaborator: N/A
Travelled to foreign country: N/A
Duration of Stay: N/A

Other organizations

SUPER is a collaboration involving 12 organizations as funded collaborators. They are: Argonne National Laboratory, Lawrence Berkeley National Lab (including NERSC), Lawrence Livermore National Laboratory, Oak Ridge National Laboratory, University of California, San Diego, University of Maryland, University of North Carolina, University of Oregon, University of Southern California, University of Tennessee, University of Texas, El Paso, and University of Utah.

In addition to the funded collaborators in SUPER, we are mandated to collaborate with other SciDAC-

funded projects, as well as with major users of DOE computing systems.

In addition to collaborations with U.S. universities and national laboratories, we have benefitted from cooperation with several U.S. Corporations.

Organization Name: Intel Corporation
Location of Organization: Santa Clara, CA and Beaverton, OR.
Partner's contribution: Intel has supported us through long-term loans of equipment for benchmarking, experimentation, and development. We have also had extensive discussions with Intel staff under NDA regarding instrumentation and control for building practical user-space performance and power tools.

Organization Name: Dell Corporation
Location of Organization: Round Rock, Texas and Raleigh, NC.
Partner's contribution: Dell has served as Intel's partner for arranging the loans of equipment.

Organization Name: Nvidia Corporation
Location of Organization: Santa Clara, CA, and Durham, NC.
Partner's Contribution: Loan of equipment and discussions with Nvidia staff regarding performance issues under NDA.

References

[1] P. Balaprakash, A. Tiwari, and S. M. Wild. Multi-objective optimization of HPC kernels for performance, power, and energy. In *Proceedings of the 4th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS13)*, November 2013. To appear, available at http://www.mcs.anl.gov/papers/P4069-0413.pdf.

[2] P. Balaprakash, A. Tiwari, and S. M. Wild. Framework for optimizing power, energy, and performance (poster). In *International Conference for High Performance Computing, Networking, Storage, and Analysis (SC13)*, November 2013. URL http://sc13.supercomputing.org/sites/default/files/PostersArchive/post154.html.

[3] D. Bedard, M. Y. Lim, R. Fowler, and A. Porterfield. PowerMon: Fine-grained and integrated power monitoring for commodity computer systems. In *Proceedings Southeastcon 2010*, Charlotte, NC, Mar. 2010. IEEE.

[4] S. Bhalachandra. *Runtime Methods to Improve Energy Efficiency in Supercomputing Applications*. PhD thesis, University of North Carolina at Chapel Hill, Jan. 2018. Defended Jan 2018, final copy not yet accepted.

[5] S. Bhalachandra, A. Porterfield, S. L. Olivier, and J. F. Prins. An adaptive core-specific runtime for energy efficiency. In *31st IEEE International Parallel & Distributed Processing Symposium IPDPS)*, page 947, May 2017.

[6] S. Bhalachandra, A. Porterfield, S. L. Olivier, J. F. Prins, and R. J. Fowler. Improving energy efficiency in memory-constrained applications using core-specific power control. In *5th International Workshop on Energy Efficient Supercomputing (E2SC)*, Denver, CO, Nov. 2017. ACM SIGHPC.

[7] S. Borkar and A. A. Chien. The future of microprocessors. *Communications of the ACM*, 54(5):67–77, May 2011.

[8] J. W. Choi, D. Bedard, R. Fowler, and R. Vuduc. A roofline model of energy. In *Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS13)*, Boston, MA, May 2013.

[9] K. P. David DeBonis, James Laros. Qualification for PowerInsight accuracy of power measurements. Technical Report SAND2013-9639, Sandia National Laboratories, Nov. 2013.

[10] E. Deelman, C. Carothers, A. Mandal, B. Tierney, J. Vetter, I. Baldin, C. Castillo, G. Juve, D. Krol, V. Lynch, B. Mayer, J. Meredith, T. Proffen, P. Ruth, and R. F. da Silva. Panorama: An approach to performance modeling and diagnosis of extreme scale workflows. *International Journal of High Performance Computing Applications*, to appear, 2015.

[11] W. Ermler, J. Tilson, and R. J. Fowler. Spin-orbit configuration interaction calculations of electronic spectra of RuO2+ and OsO2+ catalytic cores. In *Southwest Regional Meeting of the American Chemical Society (SWRMACS 2012)*, Baton Rouge, LA, 2012.

[12] R. Grant, S. Olivier, J. Laros, R. Brightwell, and A. Porterfield. Metrics for evaluating energy saving techniques for resilient HPC systems. In *Tenth Workshop on High-Performance, Power-Aware Computing (HP-PAC)*, Phoenix, Arizona, May 2014. To Appear.

[13] K. Huck, S. Shende, A. Malony, H. Kaiser, A. Porterfield, R. Fowler, and R. Brightwell. An early prototype of an autonomic performance environment for exascale. In *International Workshop on Runtime and Operating Systems at Scale (ROSS2013)*, Eugene, OR, June 2013.

[14] K. A. Huck, A. Porterfield, R. Fowler, N. Chaimov, H. Kaiser, A. D. Malony, and T. Sterling. An autonomic performance environment for exascale. In *Supercomputing Frontiers 2015*, Singapore, Mar. 2015.

[15] K. Kasichayanula, D. Terpstra, P. Luszczek, S. Tomov, S. Moore, and G. Peterson. Power Aware Computing on GPUs. In *Proceedings of SAAHPC '12*, July 2012. Recipient of Best Paper Award.

[16] J. Laros, P. Pokorny, and D. DeBonis. PowerInsight - a commodity power measurement capability. In *The Third International Workshop on Power Measurement and Profiling in conjunction with IEEE IGCC 2013*, Arlington, Va, June 2013.

[17] A. Mandal, R. Fowler, and A. Porterfield. System-wide introspection for accurate attribution of performance bottlenecks. In *Workshop on High-performance Infrastructure for Scalable Tools (WHIST)*, Venice, Italy, June 2012.

[18] A. Mandal, P. Ruth, I. Baldin, Y. Xin, C. Castillo, M. Rynge, and E. Deelman. Evaluating I/O aware network management for scientific workflows on networked clouds. In *Proceedings of the Third International Workshop on Network-Aware Data Management*, NDM '13, pages 2:1–2:10, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2522-6. doi: 10.1145/2534695.2534698. URL http://doi.acm.org/10.1145/2534695.2534698.

[19] A. Mandal, I. Baldine, Y. Xin, P. Ruth, and C. Heerman. Enabling Persistent Queries for Cross-aggregate Performance Monitoring. *IEEE Communications Magazine*, 52(5), May 2014.

[20] A. Mandal, P. Ruth, I. Baldin, Y. Xin, C. Castillo, F. Jiang, C. Schmitt, G. Juve, M. Rynge, E. Deelman, and J. Chase. Managing Dynamic Networked Cloud Infrastructure for Data Driven Scientific Workflows Using Proactive Introspection. In *Demonstration at Workshop on Innovating the Network for Data Intensive Science at SC 2014*, 2014. URL https://scinet.supercomputing.org/workshop/sites/all/themes/corporateclean-7.x-2.3/corporateclean/2014INDISAbstracts/demo103s3.pdf.

[21] A. Mandal, P. Ruth, I. Baldin, Y. Xin, C. Castillo, G. Juve, M. Rynge, E. Deelman, and J. Chase. Adapting scientific workflows on networked clouds using proactive introspection. Technical Report TR-15-01, Renaissance Computing Institute (RENCI), Chapel Hill, 2015. URL http://renci.org/technical-reports/tr-15-01-scientific-workflows-on-networked-clouds/.

[22] A. Porterfield, R. Fowler, S. Balachandra, and W. Wang. OpenMP and MPI application energy measurement variation. In *1st International Workshop on Energy Efficient SuperComputing (E2SC)*, Denver, CO, Nov. 2013.

[23] A. Porterfield, S. Olivier, S. Bhalachandra, and J. Prins. Power measurement and concurrency throttling for energy reduction in OpenMP programs. In *Ninth Workshop on High-Performance, Power-Aware Computing (HP-PAC)*, Boston, MA, May 2013.

[24] A. Porterfield, R. Fowler, S. Bhalachandra, B. Rountree, D. Deb, and R. Lewis. Application runtime variability and power optimization for exascale computers. In *International Workshop on Runtime and Operating Systems at Scale (ROSS2015)*, Portland, OR, June 2015.

[25] A. Porterfield, R. Fowler, S. Bhalachandra, B. Rountree, D. Deb, R. Lewis, and B. Blanton. Application runtime variability and power optimization for exascale computers. In *Proceedings of the 5th International Workshop on Runtime and Operating Systems for Supercomputers*, Portland Oregon, USA, 2015.

[26] P. Ruth, A. Mandal, C. Castillo, R. Fowler, J. Tilson, I. Baldin, and Y. Xin. Achieving performance isolation on multi-tenant networked clouds using advanced block storage mechanisms. In *6th Workshop on Scientific Cloud Computing (ScienceCloud'15)*, Portland, OR, June 2015. doi: 10.1145/2755644.2755649.

[27] P. Ruth, A. Mandal, C. Castillo, R. Fowler, J. Tilson, I. Baldin, and Y. Xin. Achieving performance isolation on multi-tenant networked clouds using advanced block storage mechanisms. In *ScienceCloud'15, Proceedings of the 2015 ACM HPDC International Workshop on Scientific Cloud Computing, Portland, OR, June 20115*, page to appear, 2015.

[28] J. L. Tilson, W. C. Ermler, and R. J. Fowler. Ci potential energy curves for three states of ruo2+. *Chemical Physics Letters*, 516(46):131 – 136, 2011. ISSN 0009-2614. doi: 10.1016/j.cplett.2011.09.075. URL http://www.sciencedirect.com/science/article/pii/S0009261411012176.

[29] A. Tiwari, M. Laurenzano, L. Carrington, and A. Snavely. Modeling Power and Energy Usage of HPC Kernels. In *Proceedings of the Eighth Workshop on High-Performance, Power-Aware Computing 2012*, HPPAC '12, 2012.

[30] A. P. Wei Wang, John Cavazos. Autotuning using polyhedral approach. In *Workshop on Polyhedral Compilation Techniques*, Vienna, Austria, Jan. 2014.