Creating Infrastructure for Research Collaboration

Arun K. Somani, Anson Marston Distinguished Professor, Electrical and Computer Engineering, Iowa State University

In the world of today, information is key to success. Plenty of research done today requires and/or uses exa-scale computers. High performance computing (HPC) is essential for advanced research in virtually all disciplines of science and engineering, and in particular in the fields of bio-, materials-, and information-technologies, all identified as strategic priorities of Iowa State University. The remarkable increase in computational capability (more broadly cyber infrastructure) over the last 40 years has changed societies, disciplines, and governments. Availability of contemporary HPC services is instrumental in researchers' ability to improve research quality and competitiveness. They are also essential in attracting and retaining top faculty working in vital disciplines that rely on the computational sciences, build strong inter-institutional research collaborations, and lead to new discoveries by enabling these researchers to scale up models beyond the known edges of prior work. A considerable investment in new HPC is thus crucial.

With these thoughts in mind, how can Iowa State University best position itself to optimize the use and development of cutting-edge HPC and to lead the changes? A University HPC Committee (U-HPC-C) was formed to address the need of HPC users at ISU campus. Our model to address the said needs included multi-faceted aim to achieve economies of scale in space utilization, to maximize the use of physical facilities, to leverage available funds, and to move from an ad hoc to a planned approach to support most medium to large scale users. **We believed that a strong case can be made for the acquisition of a new HPC platform that would satisfy the needs of the projects described herein, and provide sufficient capacity to meet the needs of a broad number of important research groups on campus.** The merit of the new HPC platform includes the far-reaching and impactful research thrusts that it enables, by accelerating knowledge discovery, spanning areas as diverse as animal sciences, plant genomics, climate modeling, and wind power generation.

ISU's computer users can be separated into two communities: i) those that do science and HPC is an instrument for them, and ii) those that do science to advance the state-of-the-art in HPC by developing new algorithms, architectures, data storage techniques, and management of computation. These two communities and their HPC needs are superficially different, but deep down both communities benefit and have benefitted from the strong interaction between them. *A guiding principle of our goal is to ensure that the first community benefits*

from the new instrument to the fullest possible extent while facilitating time for HPC innovation by the second community. We also believed that the new HPC platform required 100-200 TFlops compute capability, large memory (up to 16GB per core) and 1,000+ TB storage, and strong support for parallel I/O. The new HPC platform would handle the storage needs of various applications.

Science and Its Need

We identified the following major projects that would benefit from such a facility.

Project 1. Biosciences. ISU has large, well established interdisciplinary research and education programs at the forefront of biological sciences. These research efforts span microbial, plant and animal species, and are fully integrated with teams of computational scientists due to the transformation of biology as a data-driven science. In the microbial arena, ISU researchers are engineering microbial organisms for biorenewable production of chemicals and important energy-related compounds such as hydrocarbons. ISU has a preeminent program in plant sciences research with particular emphasis on biotechnology of important cereal crops such as maize, barley and soybean. Our comprehensive plant sciences research spans food, feed, fiber and fuel through eight research centers, along with environmental research through the Center for Carbon Capturing Crops. ISU animal and veterinary science researchers are engaged in genomics and systems biology of livestock for effective breeding, improving quality and nutrition of food, and study of diseases that affect livestock.

Though these applications are broad and diverse, fundamental advances in genomics and the common genetic mechanisms underpinning all life forms provide many cross-cutting synergies among these seemingly disparate fields of research. In particular, data-intensive experimental equipment is commonly used, including sequencers, microarrays, and mass spectrometers to measure metabolic fluxes. Thus, the computational and bioinformatics tools needed to drive such research share common methodologies and computing needs. The emergence of high-throughput DNA sequencing technologies and their rapid proliferation and throughput gains during the past five years has created a compelling need for the HPC equipment.

Sequencing individual plants and animals[1] from previously sequenced species (i.e., a template genome exists) enables determining structural variation (indel, SNP, CNV, re-arrangements) along with annotation. The animal sciences group is planning tens to hundreds of individual sequences for the bull, cow, pig, and chicken genomes, roughly the same size as the ~3 billion bp human genome. As part of the ISU Beef Research herd, researchers are sequencing 100 plus bulls in the next year followed by 5-10 bulls/year thereafter. For a nominal 50X sequencing coverage through short reads and 100 individuals, the data size is approximately 15 TB. Similarly, the plant sciences group is engaged in the sequencing of multiple varieties of crop plants. In addition to

storage and computational needs, genomic rearrangements are common in plants due to selective breeding, which creates short evolutionary history.

Genome assembly. Genome assembly[2] is the problem of inferring the unknown genome of an organism from a high coverage sampling of short reads from it obtained from a high-throughput sequencer. Assembly is needed when sequencing a new species, or when sequencing a new individual of a known species where genomic modifications make the use of a template genome unviable. ISU researchers are working on engineering *E. coli* for production of hydrocarbons.

Genome wide association studies enable identification of Single Nucleotide Polymorphisms (SNPs), which are single base differences between genomes of individuals from the same species. One use would be to link groups of SNPs to a phenotype such as a particular disease. Modern genome wide association studies are conducted using SNP chips, which consist of hundreds of thousands of high density probes allowing many SNPs to be interrogated simultaneously. For example, the animal sciences group is currently in the process of migrating from 50K SNP chips to 500K SNP chips as they are becoming available for cattle, swine and chickens. It is envisioned that whole genome data (3 million plus markers per individual) will be available in the near future (see Genome Re-sequencing section above). The goal is to perform interaction studies using all of these genes, which requires fitting a Markov chain of typically 50,000 cycles. Computationally, this results in a linear system of 500,000 equations that need to be repeatedly fit 50,000 times in order to compute the relevant posterior distributions as per the Bayesian approach.

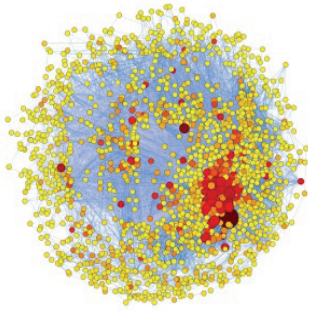


Figure 1. Backbone of Arabidopsis genome network containing 1,556 genes and 22,073 interactions. Display represents a union of the shortest paths between each pair of the top 5% of the hubs in the entire genome network. The size of a node is proportional to its degree and the intensity of its color is proportional to its betweenness centrality.
Picture Courtesy S. Aluru.

Biological Network Inference and Analysis. Biological networks[3] (see Figure 1) represent interactions between genes, RNAs, proteins and metabolites whose collective system-level behavior determines the biological function. Inference and analysis of networks depicting flow of activity and regulation are fundamental to understanding biological processes. Such knowledge is vital to the pursuit of engineering efforts to achieve a desirable outcome or to understand or treat disease (how does a malfunctioning gene alter the pathway?). ISU researchers have developed methods for gene network inference and analysis at the whole-genome scale, and protein interaction networks at the genome-scale. A typical data set for gene networks involves thousands of experiments, each providing expression values of every gene in the organism (~20,000 – 50,000 for plants), resulting in as many as 100 million values or more. The data

can be from microarray experiments or the newer RNA-seq experiments via next-gen sequencing. Inference is carried out via parallel methods that use sophisticated non-linear approaches.

Metabolomics and *in Silico* Modeling. Developing a comprehensive understanding of a biological system also requires the processing and integration of a large number of datasets from a variety of high throughput instrumentation, beyond just sequencers. ISU is home of several large databases funded by the NSF and USDA that analyze and make these datatypes available to the research community (e.g., PLEXdb, MaizeGDB, PlantGDB, PlantMetabolomics.org, and animalgenome.org). Several ISU researchers are engaged in metabolic flux analysis, a powerful diagnostic tool enabling quantification of all steady state intracellular fluxes in a metabolic network. The resulting maps provide a measure of the extent of contribution of various pathways in cellular metabolism. Another major research effort is to build genome-scale *in silico* models of organisms to be able to conduct flux analysis and ask "what if" questions to study the effect of genetic manipulations on desired end goals such as increasing the production of a metabolite.

Project 2: Multiscale Methods for Grand Challenge Problems

Methods that can accurately and efficiently span multiple length and time scales are required to address many "grand challenge" problems, such as design of new materials for specific applications, capture of solar energy, study of heterogeneous catalysis, simulation of the processes for biomass conversion to usable energy, simulation of atmospheric phenomena such as aerosol formation and reactivity, and analysis of enzyme catalysis. ISU researchers' goal is to develop methods that are capable of providing both accuracy and computational efficiency. Developing such methodology starts with high-level quantum mechanics (QM) methods that are computationally expensive and mapping the high-level potential onto a potential that is much simpler and much less computationally demanding, with minimal loss of accuracy. This process is called coarse graining. It divides a large molecular system into smaller, more computationally tractable fragments so that the properties of each fragment can be computed on a separate node, while the accuracy is not significantly compromised. This fragmentation scheme[4] greatly reduces the cost of a QM calculation and facilitates multi-level parallelism. The advantage of such a method and some remaining challenges can be illustrated by considering a large cluster of water molecules. Because water is arguably the most important liquid and solvent, to perform a realistic molecular dynamics (MD) simulation on water, one needs to start with ~1,024 waters molecules. Using 131,000 BG/P cores, a single energy + gradient calculation on 1,024 molecules requires 1.2 minutes. A realistic MD simulation with 1 million time steps, this time translates into 1,200,000 minutes = 2.28 years! One way to address the challenge presented by the steep scaling of QM methods is to map (coarse grain) the FMO potential onto a much simpler potential. Then one can employ graphical processing unit (GPU) accelerators.

Formation and Reactions of Atmospheric Aerosols. An understanding of the effect aerosols[5] have on the climate (and global climate warming) has become increasingly important over the last several decades. Primary and secondary aerosols affect Earth's radiative balance by scattering and absorbing light directly, and act indirectly as cloud droplets to influence the distribution, duration, precipitation processes, and radiative properties of clouds. It is also recognized that the spatial and temporal distributions of aerosols due to industrial activities are as important in determining overall climate changes as is the influence of greenhouse gases. Understanding the nucleation, growth and evaporation rates of aerosols as well as their chemical properties is essential to improve climate models and overall global climate prediction as well as the general chemical ecosystem in the atmosphere. Gaining this understanding is extremely challenging both from a scientific and a computational science point of view.

Design of Dendrimers. Dendrimers are a class of polymer with regularly branching repeat units emanating from a central core (Figure 2). They are synthesized using a series of controlled reaction steps, which endows these high molecular weight molecules with the structural precision of a small organic molecule. This attribute gives dendrimers an advantage over other polymers in biomedical applications, where strict regulatory requirements are imposed on polymer-based materials for use in humans. Applications for dendrimers include microbicides, drug and gene delivery, tissue engineering, imaging, and water and soil remediation. Designing dendrimers with specific materials properties requires a thorough understanding of how changes in the size, shape, and surface chemistry of a dendrimer affect its interactions with target species, such as low molecular weight organic drug molecules or industrial pollutants. Molecular simulation with atomistic resolution is an invaluable counterpart to experimental observations, provided that realistic and accurate molecular models are available.

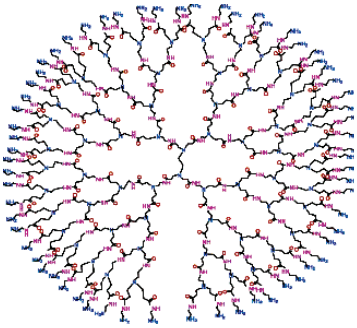


Figure 2. Structure of dendrimer showing core-shell structure.

Project 3: Computational Fluid Dynamics (CFD) Modeling

Introduction. The CFD modeling group at ISU makes extensive use of high-performance computing to carry out cutting-edge research using simulation methods in fluid mechanics and multiphase flows[6]. These simulations will fundamentally advance the century-old challenge: the direct computation of turbulent flows at flight conditions. Each simulation requires about 40 processors with a memory of 1GB per processor and the run time of a typical simulation is about 10 days. Particle-resolved DNS simulations of large risers in three dimensions have not been performed previously. Algorithmic developments made in designing optimal parallelization strat-

egies for particle-resolved DNS with fluid-particle coupling have broader applications in sprays, bubbly flows and device-scale simulations of gas-solid flow applications that employ discrete element methods to treat the solid phase.

Fundamental Physics of Multiphase Reactive Flows. Multiphase reactive flows are encountered in energy generation processes using fossil or bio-based fuels and in emerging technologies to capture the resulting CO_2 emissions. Technologies such as chemical-looping combustion and capture of CO_2 using dry sorbents promise to reduce greenhouse gas emissions. CFD plays an important role in the design and scale-up of these devices. The multiphase flow in these industrial devices is characterized by complex hydrodynamics, heat transfer and both exothermic and endothermic chemical reactions. A fundamental understanding of the interaction between hydrodynamics, heat transfer and chemistry over a wide range of physical parameters (such as solid phase volume fraction, Reynolds number) is crucial for successful device-scale CFD models that can be used to design systems for these emerging technologies. ISU researchers have developed a method that represents the contact mechanics of multiparticle collisions between particles up to the close-packing limit of solid volume fraction (Figure 3), requiring use of high-performance computing.

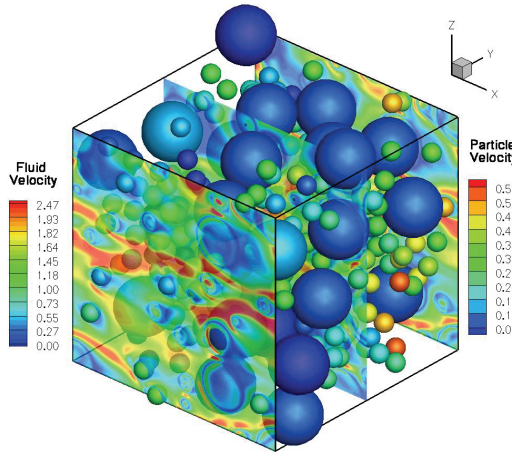
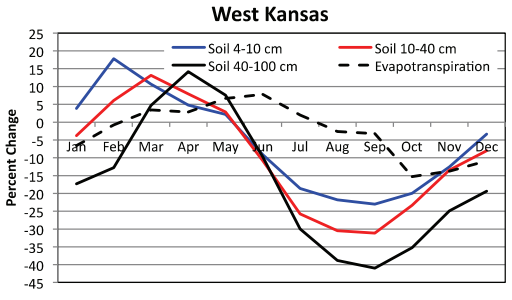


Figure 3. Contours of fluid velocity in a freely evolving polydisperse gas-solid flow.

Project 4: Coupled Dynamics of Land Use Change and Regional Climate Extremes

The ISU climate simulation group is developing first-of-its kind agricultural policy-climate projection systems to address food security and climate change[7,8]. The goal is to perform iterative simulations that are both unprecedented in model coupling and in capability for addressing uncertainty growth in these projection systems. Decision makers rely on projections of climate, land use, and agricultural productivity to develop policies that promote increased production and acceleration of efficient agricultural practices in developing countries. Current projection systems are not dynami-

Figure 4. Percent change of soil moisture and evapotranspiration across western Kansas predicted by regional climate simulations. Two 25-yr simulations were generated with identical weather conditions but differing land use. (Change is current land use minus POLYSYS projection).



cally coupled, so that feedbacks between land use and climate change are not considered. We are integrating projections of climate change and policy-driven agricultural land use change through our novel design of climate and agricultural projection systems. For example, projected land use change to increase switchgrass production in the Great Plains creates a soil water deficit that reduces plant transpiration during summer and fall (Figure 4), and, by extension, biomass productivity. This model-based result, in combination with similar findings from hay cropping systems, strongly argues that water resources are insufficient to attain the policy target.

Project 5: Other Application Areas

Prediction and Discovery of Materials: Reverse Engineering of Crystal Structures. The prediction of crystal structures from their chemical composition has long been recognized as one of the outstanding challenges in theoretical solid state physics[9] [10]. From the materials design perspective, it is desirable to have a method that requires no prior knowledge or assumptions about the system. Our aim is to develop reliable computational tools that can predict material structures from given chemical compositions in the emerging new area of computational discovery of materials with optimal properties. Algorithms proposed to tackle this challenging problem include simulated annealing, genetic algorithms (GA), and basin or minima hopping. The GA has proved to be a powerful approach to predict material structures using first principles calculations and the knowledge of the chemical composition. It also scales well in terms of throughput achievable by increasing the number of computing nodes.

Atom Probe Tomography. HPC also is critical to establish a new computational paradigm and infrastructure that will enhance 3-D atomic reconstruction for an emerging new instrumentation technology, atom probe tomography (APT). APT is a powerful microscopy tool that enables spatial resolution of hundreds of millions of atoms at the sub-nanoscale. APT is the only instrument capable of mapping the 3-D composition of a metal, semiconductor, or an insulator with atomic resolution. The 3-D reconstruction of this direct space information provides unprecedented capabilities for characterizing materials at the atomic level.

Design of Large Energy System Design Optimization. With increasing demand of energy, power system expansion must be planned while addressing the integration of various different renewal energy sources, satisfying the policy requirements, and accounting for interdependencies among the energy sources and their transformation. We at ISU are addressing these problems under a NSF supported EFRI project[11] [12]. In capacity planning problems, when total demand in the system is greater than the total supply, we need to include commodity generation and capacity expansion capabilities in the model. We have developed a transformation methodology to transform capacity expansion problems into linear network flow problem, which allows solution of such problems using multiple computers.

Education. ISU has developed the Nation's largest bioinformatics and computational biology graduate program and is recognized as a major provider of a trained workforce in these critical areas. The program is supported by two NSF IGERT grants and a long-term REU, and serves as a common educational platform for students pursuing research work in biosciences through our various departments, centers and institutes. ISU also received a new IGERT grant in the wind energy area. We will incorporate the algorithms and methods to carry out large scale computations in two advanced classes. These courses will provide high-level training for students and prepare them for research in high performance computation. The state-of-the-art HPC cluster, coupled with our research programs focused on real world problems, will position ISU to better compete for women and underrepresented minority graduate students.

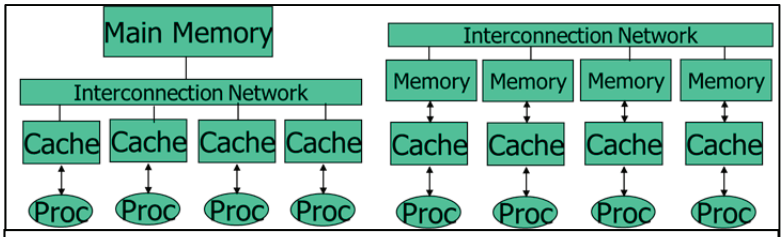


Figure 5: Shared and Distributed Memory Machines

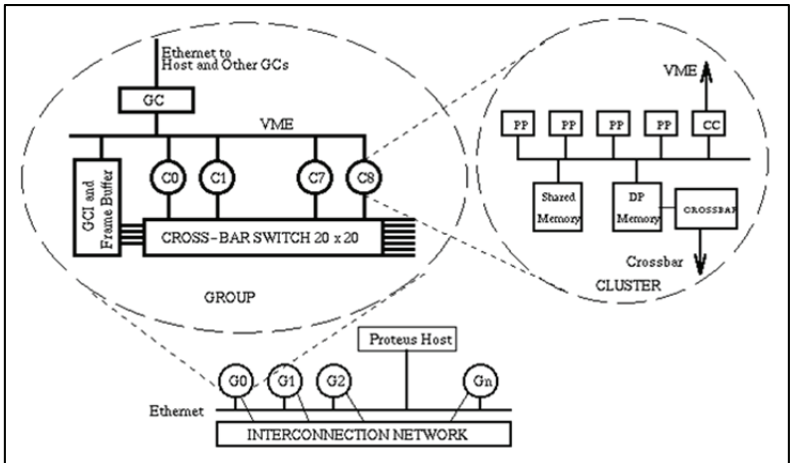


Figure 6: Proteus Distributed Memory Machine with an Efficient and Easy to Use Hierarchical Network

HPC Infrastructure

For the desired goals for science, it is clear that a new HPC platform is essential to continue the track record of the extraordinary advances made possible by use of our previous generation machines, the BlueGene/L and Sun cluster systems supporting interdisciplinary and multi-disciplinary research. *Each research project above explicitly defines its need for the new HPC cluster.* Our applications are large and data-driven, and require more memory per node for processing. They generate large volumes of data that must be stored for reuse and sharing. Therefore, we decide to approach the NSF MRI program to propose a cluster configuration that represents a balanced machine and maximizes the use of processing cores, memory, and storage. It was a group effort built upon our culture of sharing resources and our tradition of developing new HPC applications in collaboration with domain experts to solve leading-edge problems in science and engineering.

Earlier Machines. In the past (1970s, 1980s, 1990s) when research needed computing, specific machines were designed to fulfill the need[13]. Examples include Maspar, Thinking machine, Intel Paragon, IBM SP1, SP2 etc. They were based on two dominant models (as shown in Figure 5): i) Shared Memory machines where memory is shared by all processors; and ii) Distributed Memory machines where data is exchanged through messages. It is easy to program under the first paradigm, but the machines do not scale. It is bit harder to program under the second paradigm, but machines scale better. In both cases, the main important features for high performance included: i) fast multiple processing elements connected with high-speed networks; ii) efficient partitioning of problem; iii) efficient algorithms using large chunks partitioning and using coarse-grain messages with low overhead per message; and iv) low network latency which is tolerant to communication latency and allowed computation/communication overlap. One example of such a machine is the Proteus machine designed and built by the author at the University of Washington as shown in Figure 6. This machine was used for coarse grain image processing elements and included an efficient computing node and a coarse grain message passing system for large data processing.

Newer Paradigms: Earlier machines were difficult to manage and program. Since then much progress has been made in their design and programming. They use commodity processor and networking systems, and are easy to manage/program providing cost to computational efficiency. A new computing paradigm has emerged that includes the following:

1. Infrastructure as a service where rather than buying a machine now, one can rent them from Amazon, GoGrid, AppNexus, HP, and many others. These vendors create a cloud that delivers computing and storage as a service.
2. Platform as a service where users simply require a nice API (application programming interface) and platform designer takes care of the rest of implementation, such as a database, web server, and development tools.
3. Software as a service where users just run their applications like Google email or virtual desktop.
4. A shared infrastructure like cloud that is shared by a large number of users, which is how many universities are structuring their computing infrastructure. These are cost effective if sharing can work.

HPC Machine: We adopted the last paradigm for our effort. We started to develop a collaborative HPC infrastructure model with a goal to position ISU strategically for advancement in research. The goals set for our effort included the following:

1. Identify and address the needs of HPC users at ISU campus;
2. Support most users from medium to large scale;
3. Utilize and leverage resources;
4. Achieve economies of scale in space utilization;

5. Maximize the use of physical facilities-power/cooling/racks /cable;
6. Leverage available funds;
7. Longer commitments and unchanged vision;
8. Move from an ad hoc to a planned approach; and
9. Sustainable over long term!

System Configuration: When we discussed the needs with our user groups, the following configuration emerged. Our HPC cluster consists of the following types of nodes to meet the anticipated demands.

Head Node. The Head Node is critical to the operation of the entire cluster. Home directories are stored on the head node. It uses hot swappable redundant power supplies, large storage for user directories, and includes a dual port 10Gb NIC with Fiber connections to connect to the campus network.

Three types of Compute Nodes: GPU, Fat, and Thin. Four "Fat" Compute Nodes will have 16GB memory per core and the rest of the Compute Nodes will be "Thin" nodes with 8 GB memory per core. 32 GPU Compute Nodes include dual GPU system.

Interconnects: Compute, Head, and Storage Nodes are interconnected via QDR InfiniBand switch, a Gigabit Ethernet switch and an Ethernet switch for IPMI.

Storage: The storage is designed to be two tiered: 150 TB of fast, reliable storage for scratch space using the Lustre file system plus about 850 TB of raw storage using NFS file system that is scalable, highly reliable, expandable and fault tolerant storage with no single points of failure.

System Software: The system software includes (i) The Red Hat Enterprise Operating System; (ii) The Lustre parallel file system; (iii) Intel's Fortran, C and C++ Compilers; (iv) Intel's Cluster Toolkit; (iv) Intel's MPI; (v) PGI's Fortran, C and C++ Compilers; (vi) PGI's CUDA Fortran; (vii) GNU's Fortran, C and C++ Compilers; (viii) Allinea's DDT and OPT; and (ix) The MOAB's Cluster Suite for the GPUs.

Racks are cooled so that little or no heat is put into the machine room.

Realization: A group of faculty members was identified to develop external funding for this proposal, specifically targeting the NSF MRI (major research instrumentation) program. A successful $2.6M proposal for a large heterogeneous machine was developed that met the needs of a plurality of researchers. It is hard to bring people together, in particular from multiple disciplines, but not impossible. But once done, it is worth the effort. We at ISU are very proud of our success.

Acknowledgements: The project description is based on details provided by various researchers including Drs. Srinivas Aluru, Patrick Schnabel, Jim Reecy, Robert Jernigan, M. Gordon, Theresa Windus, Monica Lamm, Rodney Fox, Z. J. Wang, Baskar Ganapathy Subramanian, Shankar Subramaniam, Eugene Takle, Chistopher Anderson, Bruce Harmon, Krishna Rajan, Jim McCalley, and Glenn Luecke.

References

1. S. Gnerre, I. MacCallum, D. Przybylski, F.J. Riberiro *et al*. "High-quality draft assemblies of mammalian genomes from massively

parallel sequence data", *Proceedings of the National Academy of Sciences USA*, 108, 1513 (2011).
2. J. Zola, M. Aluru, A. Sarje and S. Aluru, "Parallel information theory based construction of gene regulatory networks," *IEEE Transactions on Parallel and Distributed Systems, Vol.* 21 (12), 1721 (2010).
3. Y. Tamada, S. Imoto, H. Araki, M. Nagasaki *et al*. "Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers", *ACM/IEEE Transactions on Computational Biology and Bioinformatics*, 8, No. 3, pp. 683 (2011).
4. G. Pranami, L. Slipchenko, M. H. Lamm and M S. Gordon, "Coarse-grained intermolecular potentials derived from the effective fragment potential: Application to water, benzene, and carbon tetrachloride," in Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications, D.M. York and T.-S. Lee, Eds., Springer-Verlag, 2009.
5. Asadchev, V. Allada, J. Felder, B. Bode, M.S. Gordon, and T.L. Windus, "Uncontracted two-electron repulsion integral implementation on multicores and GPUs using the Rys quadrature method", *J. Chem. Theory Comput.* 6, 696 (2010)
6. S. Tenneti, R. Garg, C. M. Hrenya, R. O. Fox and S. Subramaniam, "Direct numerical simulation of gas-solid suspensions at moderate Reynolds number: Quantifying the coupling between hydrodynamic forces and particle velocity fluctuations," *Powder Technology* 203, 57 (2010).
7. Georgescu, M., D. B. Lobell, and C. B. Field "Direct climate effects of perennial bioenergy crops in the United States," *Pro. Nat. Acad. Sci.*, 108, 4307 (2011).
8. G. Hoogenboom, "Decision support system to study climate change impacts on crop production. Special publication," American Society of Agronomy ASA Special Publication, 2003. Also see http://www.icasa.net/dssat/.
9. J. Maddox, "Crystals from first principles," *Nature*, 335, 201 (1988).
10. K.D.M. Harris, R.L. Johnston, B.M. Kariuki, "The genetic algorithm: Foundations and applications in structure solution from powder diffraction data," *Acta Crystallogr.* A54, 632 (1998)
11. E. Ibanez, and J. McCalley, "Multiobjective evolutionary algorithm for long-term planning of the national energy and transportation systems," Energy Systems Journal, Vol. 2(2), pp. 151-169 (2011).
12. A.K. Somani and Jinxu Ding, "Parallel Computing Solution for Capacity Expansion Network Flow Optimization Problems," Journal of Computing, (July 2012).
13. A.K. Somani, C. Wittenbrink, R. M. Haralick, L. G. Shapiro, J. N. Hwang, C. Chen, R. Johnson, and K. Cooper, "Proteus System Architecture and Organization," in the Proc. of 5th International Parallel Processing Symposium, June 1991, pp. 287-294.