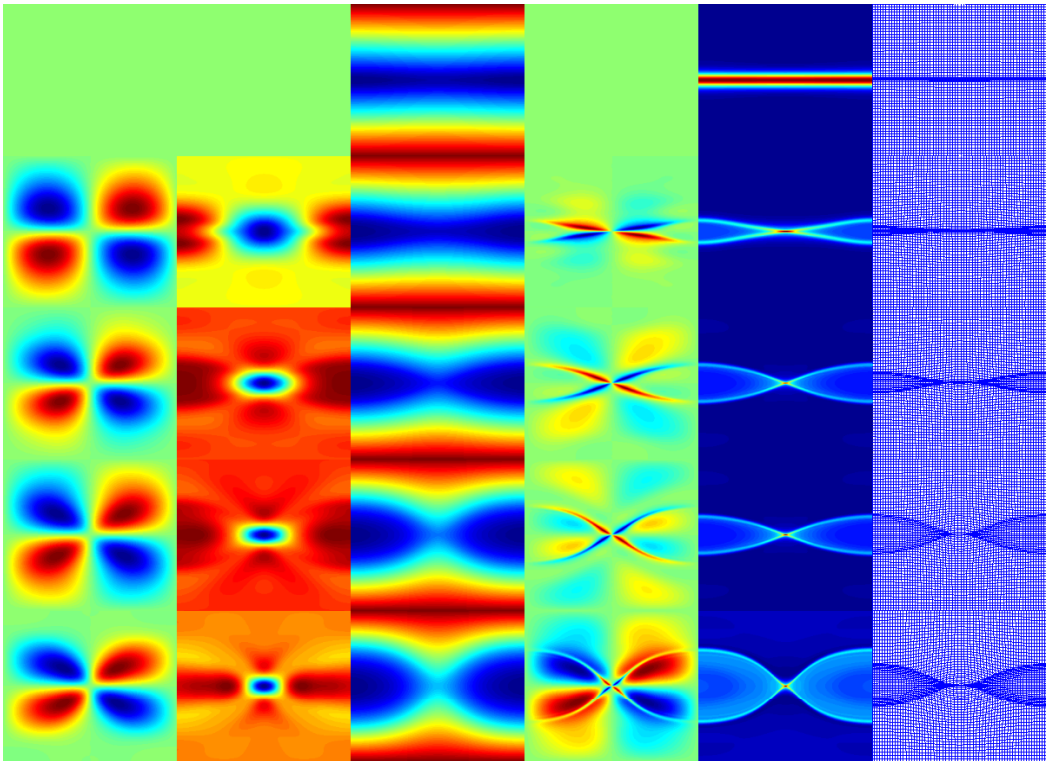
U.S. DEPARTMENT OF ENERGY | Office of Science

The Petascale Initiative in Computational Science and Engineering

NERSC, Lawrence Berkeley National Laboratory

Final Report

PI: Alice Koniges, Phone: 510-486-7481, Email: aekoniges@lbl.gov



October 2013

NERSC National Energy Research Scientific Computing Center

BERKELEY LAB Lawrence Berkeley National Laboratory

Contents

List of Figures

1 Introduction

The current computer architectural revolution provides opportunity for major increases in computational power over the next several years, if it can be harnessed. Effective use of multi-core, heterogenous, and GPU platforms is essential for science problems that require petascale computing, but is also necessary to continue the performance growth associated with Moore's Law for computing systems at all scales. This transition from simply increasing the single-processor and network performance to different architectural paradigms, forces application programmers to rethink the basic models of algorithm development and parallel programming from both the language and problem division standpoints. Additionally, advances in architecture and availability of resources are making the high performance computing (HPC) world more attractive to a variety of research areas that were not previously using advanced computing. Lawrence Berkeley National Laboratory (LBNL) is home to the NERSC Program, which serves the scientific computing needs of the Office of Science. As the Mission Computing Center for DOEs Office of Science, NERSC is tasked with helping over 4000 users to overcome the challenges of the architectural revolution through the use of new parallel constructs and languages. NERSC also enables a broader user community to take advantage of high performance computing.

The Computational Science and Engineering Petascale Initiative at LBNL identified key application areas with specific needs for advanced programming models, algorithms, and other support. The applications were chosen to be consistent with the current mission of the Department of Energy, with a particular focus on applications that benefit energy research, those supported by other ARRA (American Recovery and Reinvestment Act) funding and Energy Frontier Research Centers (EFRCs). The initiative paired post-doctoral researchers at NERSC with these high-impact projects. The post-doctoral research program was centered at the Oakland Scientific Facility (OSF) where NERSC resided. Each post-doctoral researcher was assigned to one or more projects and works directly with the project scientists. The post-doctoral researchers funded under this program generally spent a third to half of their time at the OSF facility and the remainder of their time embedded in the research project area. Alternatively, if the research project area scientists were not located in commuting distance to the OSF, the post-docs traveled for periods of time (1 - 3 weeks) to work directly at the research project facilities location. The post-docs were guided by NERSC staff experts in the computer science necessary for making these advances. The project started with funding at the beginning of the FY10. Nine post-docs fit into the funding and each post-doc completed an appointment of approximately two years.

Each section of this report covers the work of an individual post-doc. To address our milestones, we note that not only did we made significant advances to a variety of energy-related computational areas, we made a specific large contribution to an EFRC. In fact, EFRC director Professor Berend Smit describes the momentous improvement to his EFRC in the following statement, which describes how the calculations have gone from taking many years to a few weeks, and are thus now performable.

For Carbon Capture it is important to screen large sets of materials for their performance to separate CO_2/N_2. In the course of this project, we have developed a model that requires the calculation of the Henry coefficient of all materials. These Henry coefficients can be reliably estimated using molecular simulations. However, with the code that we have developed, the screening of our dataset of over 5 million materials would have taken us many years of CPU-time. The GPU code developed in the course of this project has reduced this to a few weeks of CPU-time. The importance of this development is that it allows us to screen the entire database. This allows us to identify classes of materials with optimal carbon capture performance.

The post-docs reported their research prolifically. In the publications section, we list the talks,

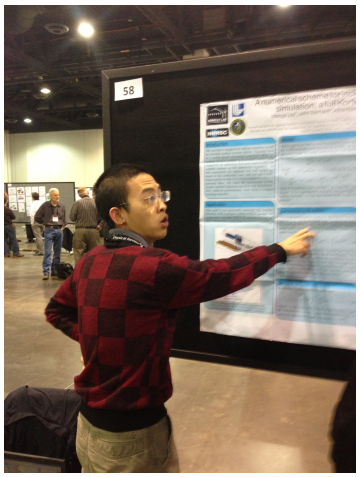


Figure 1: Post-doc Wangi Liu explaining poster at a conference in Providence, RI.

posters, and papers that the post-docs authored during their term in our program. These include one "Best Paper" associated with an International conference. An integral part of the program was to provide the post-doctoral researchers a productive environment for working with the application code teams. Despite the variety of application areas, there was considerable overlap in the methods and tools for achieving petascale application performance. Through interaction with the NERSC staff as well as the LBL computational science groups, the work produced a significant benefit to NERSC staff and users. For instance, the post-docs served as teaching assistants in NERSC sponsored classes and a NERSC-led tutorial at SC10, SC11, and SC12. One post-doc created and launched a scientific data bank for x-ray images. One worked on outreach to high school students and another helped foster diversity by working at the recruiting booth at the Grace Hopper conference.

2 Chemistry application codes for carbon capture

The Energy Frontier Research Centers (EFRCs) were established by the Department of Energy to conduct research related to energy issues. UC Berkeley Professor Berend Smit leads an EFRC that focuses on better ways to separate carbon dioxide from power plant flue gases. To reduce the costs associated with current carbon capture technologies, novel materials such as zeolites and metal-organic frameworks (MOFs) that are based on microporous networks are being studied and analyzed. However, the experimentally known framework structures make up a very small fraction of the more than millions of zeolites and MOFs that have different topologies. Given that it is expensive and time consuming to experimentally synthesize millions of different carbon capture material structures, the need to develop simulation codes that can characterize and ultimately screen a large database of carbon capture materials within a reasonable computation time is very important.

Jihan Kim worked with Professor Berend Smit on developing a high throughput GPU code to accelerate molecular simulations that can expedite the screening process. Kim developed a GPU code written in CUDA that can characterize and screen a large number of zeolite and MOFs to identify the most efficient structures for carbon capture. The code can calculate Henry coefficients and heats of adsorption, which characterize the selectivity of the materials at low pressure regimes.

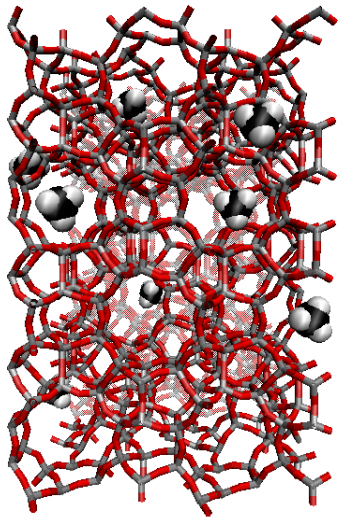


Figure 2: Graphical representation of a sample zeolite structure (MFI). The framework atoms consists of silicon and oxygen atoms (red-silver segments) and the carbon dioxide molecules (black-white spheres) are located in energetically favorable pores within the MFI structure.

The three major steps of the code are as follows: (1) energy grid construction - utilizing the 6-12 Lennard Jones potential and the Ewald summations for Coulomb interactions, an energy grid that is superimposed on top of the structure's unit cell is constructed. In order to expedite calculations, the code utilizes the GPU's fast constant memory to store the framework atoms and reduce the number of division operators and pre-compute shared terms amongst CUDA threads. (2) pocket blocking - in the CPU portion of the code, regions that are inaccessible to the gas molecules are automatically detected via a flood fill algorithm based on the energy landscape. (3) Monte Carlo Widom insertions - utilizing the energy grid and the pocket blocking methods from the previous routines, the Henry coefficients and the heats of adsorption are calculated in this routine by test inserting a single guest gas molecule inside the simulation box and collecting appropriate Boltzmann and energy terms. Overall, the GPU code provides over 50× speedup compared to the group's previous CPU code and accordingly was used to characterize over 100K diverse zeolite structures and 200K diverse zeolitic imidazolate framework (ZIF) structures on the Dirac GPU cluster at NERSC. The distributions of the carbon dioxide and the methane Henry coefficients for the experimentally known IZA structures are shown in Figure 2. This work was presented as a poster and a conference paper at 2011 SciDAC conference in Denver [1] and as a session talk in the 2012 GPU Technology conference [2].

To characterize selectivity at higher pressures, Kim implemented a GPU grand canonical Monte Carlo (GCMC) code, which provides the loading of gas molecules (e.g., carbon dioxide, methane, and nitrogen) as a function of pressure. The thousands of available CUDA threads were judiciously mapped to balance independent parallel simulations at different pressure values and parallelism within a single simulation to maximize performance. GCMC provides data that can be measured via experimental methods and provide a good check to verify that the simulations and the experiments indeed agree with one another. Kim analyzed a specific MOF structure called MOF-74 and using the GPU GCMC code to find the Lennard-Jones parameters that lead to matching between the simulation and the experimental data.

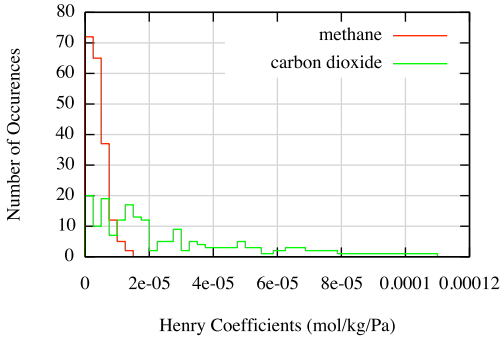


Figure 3: Histogram of CH_4 and CO_2 Henry coefficients for the 193 IZA zeolite structures.

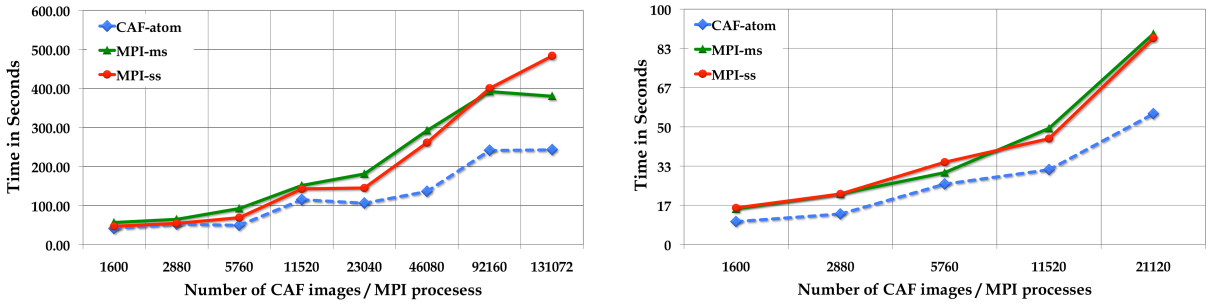
Kim also implemented a waste recycling Monte Carlo method on the GPUs where the Monte Carlo routine is parallelized by simultaneously executing multiple trial moves in a given Monte Carlo step. Kim conducted thorough studies on different kinds of parallelization methods in the GPU architecture and identifying each of the routine's strengths and weaknesses. For the waste recycling methods, Kim identified that in computing free energy profiles of particles in the caged zeolite structures, which is crucial for characterizing diffusive behaviors, the GPU waste recycling method performs better than other traditional GPU parallel algorithms. This work was published in the Journal of Chemical Theory and Computation [3].

Kim also worked with Professor Martin Head-Gordon from the UC Berkeley Chemistry department to accelerate the Q-Chem's RI-MP2 energy gradient routine [4]. The RI-MP2 routine exhibits fifth order scaling with respect to system size and the expensive DGEMM steps were converted to the CUBLAS DGEMM steps. To further accelerate the routine, the DGEMM routines were overlapped with FILE I/O reads and phtreads was used to parallelize independent loops. For the largest system simulated (i.e. twenty glycine molecules), a speedup of over 4.65 has been noted over an eight-core CPU code (run on Carver cluster at NERSC) in the computationally extensive step 4 routine which contains operations that scale by fifth order. This work was presented at the MQM conference in 2010 [5].

3 GTS — A massively parallel magnetic fusion application

The Gyrokinetic Tokamak Simulation (GTS) code — a global 3D Particle-In-Cell (PIC) code with MPI & OpenMP support — has been selected among many of today's fusion codes for the Petascale post-doc program. GTS is developed to study plasma microturbulence in toroidal, magnetic confinement devices called tokamaks. Microturbulence is a complex, nonlinear phenomenon that is believed to play a key role in the confinement of energy and particles in fusion plasmas, so understanding its characteristics is of utmost importance for the development of practical fusion energy.

In one of the levels implemented for parallelism, particles can move from one domain to another while they travel around the torus (the typical domain of magnetic fusion reactors). This shift phase is the one that has been studied the most so far in GTS since it represents the most communication intense routine in GTS and will gain in importance when scaling GTS to Petascale or even Exascale machines. At each time step, about 10% of the particles inside of a toroidal domain are communicated to adjacent toroidal neighbors, which translates to about 100GB of



(a) 1 OpenMP thread per instance

(b) 6 OpenMP threads per instance

Figure 4: Weak scaling studies of a Coarray shifter and two MPI particle shifter algorithms with no (a) and full (6 OpenMP threads per NUMA node) OpenMP support (b)

data having to be communicated each time step shift in a 1-billion particle simulation run. The original **MPI** particle shift algorithm implements a nearest neighbor communication pattern using MPI_Send_Recv functions. In the early phase of the project post-doctoral researcher Robert Preissl optimized the existing hybrid MPI/OpenMP particle shift implementations. This research resulted in a set of optimized single-threaded MPI algorithms for shift containing, among others, non-blocking MPI operations, various MPI communication techniques (e.g., buffered send), the usage of special MPI data types to eliminate particle packing overheads, etc. In addition — since GTS supports a hybrid MPI/OpenMP programming model — Preissl used OpenMP tasks to implement multithreaded MPI algorithms that exploited shared memory work-sharing constructs, as well as novel communication and computation overlapping techniques. However, all optimized MPI algorithms share the large bulk data transfers to exchange moving particles according to the rule that performance is optimized by sending fewer and larger messages.

This is in contrast to the strategy used for the new **Coarray** shift algorithms, which exploit the one-sided nature of the PGAS programming model. Preissl implemented novel Coarray algorithms — that cannot be expressed in a two-sided message passing scheme like MPI-1 — using more, but smaller messages with lower startup and completion costs. Building upon such lightweight one-sided communication techniques, communication can be efficiently spread out over a longer period of time, resulting into a reduction of bandwidth requirements and a more sustained communication and computation overlap. Moreover, the expression of the one-sided messaging semantics as language constructs improves the legibility of the code and allows the compiler to apply communication optimizations. Preissl also developed novel hybrid PGAS/OpenMP communication algorithms, which distribute the computational as well as the communication work load among OpenMP threads based on an advanced programming model, which extends the classical hybrid distributed/shared memory model as used in the MPI algorithms. It has to be taken into account that only the existing MPI communication kernel has been replaced by a new algorithm using Coarrays and the rest of the physics simulation code is left unchanged, which still has MPI function calls in it.

Figure 4 presents the wallclock runtime of a Coarray shifter implementation (*CAF-atom*) and of two MPI shift algorithms (*MPI-ms*, *MPI-ss*), running with no OpenMP support (Figure 4(a)) and full (i.e., 6 OpenMP threads per instance, on each NUMA node) OpenMP support (Figure 4(b)) on the Cray XE6 at NERSC. The algorithms are evaluated in a specially designed benchmark suite simulating GTS production run settings. Data for the singlethreaded experiments was collected for concurrencies ranging from 1600 up to 131072 processor cores. For the multithreaded runs we ran

on 9600 up to 126720 processor cores on the Cray XE6 machine. All runtime numbers presented in Figure 4 are based on weak scaling experiments. In both tests — for the singlethreaded and the multithreaded runs of the shifter benchmark suite — a steady increase in runtime for shifting particles in a torus with increasing concurrencies was observed, whereas one would expect a flat line along the x-axis for weak scaling experiments. This motivates optimization of the communication intensive GTS step to enable higher concurrencies in order to model future nuclear fusion devices. Figure 4 shows that in both cases (OpenMP turned on or off) the Coarray implementation substantially outperforms the best MPI implementations, despite the extensive work in profiling and optimization of the communication layer of the GTS code.

In addition, Preissl integrated the best multi-threaded PGAS and MPI algorithms into the GTS production code and correctly predicted the performance benefits of the particle shift phase for the full application code. The MPI optimizations were presented at the Cray User Group (CUG) meeting [4] and were published in the journal Scientific Programming [6]. The PGAS implementations were presented at the PGAS'11 [7] conference and at the SC11 conference [8].

4 Carbon Sequestration, Groundwater Flow and Porous Media

Accurate modeling of subsurface flow is of interest in several applications ranging from carbon sequestration to the flow of groundwater contaminants. The Porous Media Adaptive Mesh Refinement (PMAMR) code developed at LBNL has been successfully used to model the movement of both compressible fluids like CO_2, and incompressible fluids like contaminated water, through a variety of materials. The extension of this code to handle longer simulation times and thermal effects will broaden the spectrum of problems that can be investigated using PMAMR. In particular, larger time steps will enable the investigation of long-term behavior of carbon once it has been deposited in an aquifer. Carbon sequestration has been proposed as a way to slow the atmospheric and marine accumulation of greenhouse gases, but it is not clear what the long-term affects will be on the aquifers and their surrounding areas. The images in Figure 5, were generated by LBNL staff scientist George Pau, and highlight the interesting, detailed structures present in carbon sequestration. These features would be difficult to capture without the use of adaptive mesh refinement, an algorithm that allows the use of finer grids in select regions of the calculation.

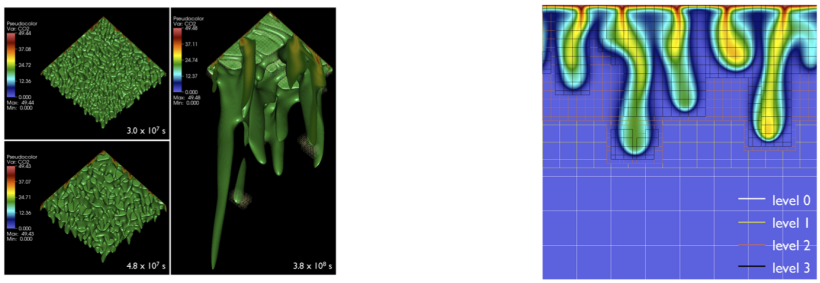


Figure 5: a) Isosurfaces of CO_2 density in the brine, this image highlights the structure of the 3D fingers. b) Block-structured adaptive mesh refinement in carbon sequestration calculation. Finest grids are concentrated near edges of the fingers, while coarse grids are used in the fluid, saving computational time. Courtesy of George Pau, LBNL.

Post-doc Kirsten Fagnan worked with John Bell in the Center for Computational Science and Engineering (CCSE). PMAMR, has been developed by George Pau, John Bell, Ann Almgren,

Marcus Day and Michael Lijewski in CCSE. In addition to modeling carbon sequestration, PMAMR is the structured component of the ASCEM project. ASCEM is an interdisciplinary, multi-lab effort to develop a new high-performance computing framework for simulating reactive flows in porous media. The DOE is interested in understanding the behavior of nuclear waste that is leaking from tanks in locations around the country. In order to understand when these slow-moving flows might reach water supplies, the calculations must be able to simulate 10's or 100's of years.

Fagnan's first work on the project involved the development of a more general Riemann solver for the multiphase, multicomponent equations being solved in PMAMR. This general solver was necessary in order to model problems with the more realistic van Genuchten permeability functions. This work was published as part of a paper on solvers for the compressible form of the multiphase, multicomponent equations [9].

Fagnan successfully added OpenMP threading to the chemistry solver that is called in the porous media code (PMAMR). The BoxLib libraries responsible for managing the adaptive mesh refinement routines were already threaded. Fagnan and collaborators worked out a methodology for distributing the grids based on the work required for the chemistry solve in a particular grid cell as opposed to distribution based on the volume of each grid cell. The volumetric approach works well when the chemistry is evenly distributed, but for typical groundwater flow problems, the chemistry was concentrated in a small region of the computational domain. Distributing based on the chemistry gives better performance in groundwater contamination calculations, especially as time progresses and more chemistry solves are required, see Figure 6. In order to obtain optimal performance out of PMAMR, Fagnan had to optimize the chemistry code (written by fellow LBNL researchers). Fagnan sped up the chemistry code by a factor almost 10 through some simple changes to how information was shared between different routines. This collaboration was important in helping to speed up a code that will be heavily used in the ASCEM project. She also performed both weak and strong scaling analysis of the code. In both cases, the hybrid programming model outperformed the MPI-only code. However, in the case of strong scaling, neither code performed well beyond around $6,000$ processors due to the bottleneck generated by the multigrid solver, see Figure 7 a). In the case of weak scaling, the code scaled to nearly $50,000$ processors, see Figure 7 b). It is also worth noting that the memory usage was significantly lower in the hybrid code than in its MPI-only counterpart (when using more than a few hundred cores), an important consideration when developing codes for the next generation of supercomputers. This work was published in the proceedings from the Cray User Group meeting in 2011 [10].

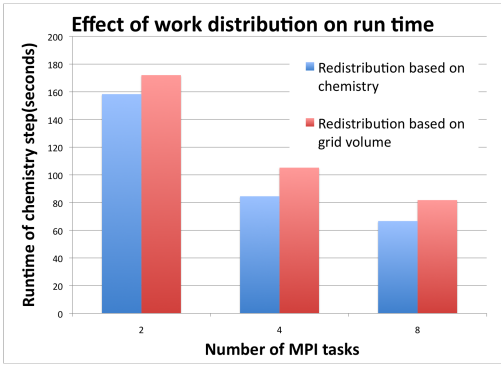


Figure 6: Improved calculation time when work load distribution is done based on the difficulty of calculating the solution to the chemistry problem.

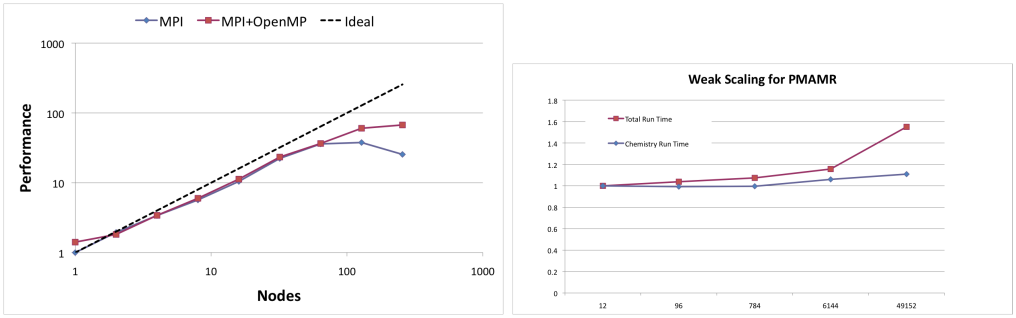


Figure 7: a) Comparison of MPI vs. Hybrid programming model for strong scaling performance of the PMAMR code. The hybrid-programming model outperforms MPI when using more than a few hundred cores, but neither methodology performs optimally due to the heavy reliance on Multigrid calculations in our solver. b) Weak scaling performance of hybrid programming model. Scaling is reasonable out to 49152 processors.

In addition to this, Fagnan worked on improving the time-stepping procedure in PMAMR. The flows present in the carbon sequestration and groundwater contaminant problems are slow moving and require simulation times on the order of hundreds of years. Since one is effectively modeling very small perturbations to a steady-state solution, we should be able to take large time steps in the calculation. PMAMR previously used a split method for solving the hyperbolic and parabolic components of the PDEs separately. While this splitting does preserve the steady-state present in the system, it leads to a severe time-step restriction due to large speeds generated during the explicit hyperbolic solve.

Fagnan overcame this time-step restriction by using a Jacobian-Free Newton-Krylov type method. Her approach uses a second or fourth order slope-limited reconstruction in a Godunov-type method to calculate the residual in the Newton step, so the method is higher-order than the existing implicit methods used in the ASCEM project. The solve is then accelerated through the use of a lower-order flux approximation as a preconditioner in the Krylov method. Fagnan investigated the accuracy and performance of this method on some toy problems. Estimates indicate that the time step can be increased by several orders of magnitude while maintaining high-order accuracy in the result.

5 Modeling a Next Generation Light Source

The Next Generation Light Source (NGLS) is a proposed X-Ray free electron laser (FEL) facility that will probe the nanoscale structure and atto- and femto-second dynamics of materials and molecular systems. Two essential components of the NGLS system are a linear accelerator, which produces bunches of high energy, relativistic electrons, and an FEL undulator, which produces light by coupling the energy of the electrons to the radiation field. The coherence and power of light produced by the FEL is acutely sensitive to the phase space of the electron bunches delivered by the accelerator. Meeting the stringent design goals of the NGLS will require improvements to each of these components. Simulation is essential to the development of these technologies because it is not feasible to construct a test device for each design change, and simple models do not provide sufficient information about the brightness and emittance of the electron and laser beams. Brian Austin worked with Jonathan Wurtele and Ji Qiang at LBNL's Center for Beam Physics to enhance

the simulation tools used to design the NGLS [11].

Austin worked primararily on the Impact-T accelerator modeling code. Impact-T uses particle-in-cell (PIC) and integrated Green's function methods to solve Poisson's equation. The domain decomposition technique used in Impact-T was parallelized using MPI and Impact-T simulations that scale up to thousands of processors and billions of particles were demonstrated prior to Austin's involvement. Poisson's equation must be solved at each time step during an Impact-T calculation, and the transpose steps of Impact-T's FFT-based solver can be a limiting factor for Impact-T's parallel scalability. Austin considered a number of approaches for improving the solver's communication patterns and ultimately identified a 'message consolidation' strategy that outperformed the others in all tests, with speedups that increase to as much as 40% at highest concurrencies [12]. Austin then added OpenMP directives to further increase the strong-scaling capacity of Impact-T. Austin also worked with members the LBL's visualization group to profile the performance of HDF-I/O functionality that was recently added to Impact-Z [13]. Austin ported Impact-Z's HDF-I/O routines to Impact-T and measured a speedup of about 30x over Impact-T's earlier ASCII I/O routines.

Optimization of the accelerator and FEL parameters is a primary need of the NGLS design project. Austin helped to develop a Matlab script that uses simplified scaling formulas to rapidly explore different engineering and physics trade-offs involved in the design of a next generation light source [14]. For more detailed design optimizations that require expensive simulations, Austin developed a parallel derivative free optimization algorithm that is used in conjunction with the Impact-T accelerator modeling code. He developed an optimization algorithm that combines the efficiency and robustness of trust-region approaches with the computational scalability of particle-swarm and Efficient Global Optimization methods.

Additionally, Austin remained active in the field of his graduate research and contributed to several recent papers and reviews on the topic of quantum Monte Carlo [15–17].

6 Frontiers in Soft X-Ray Coherent Scattering and Imaging

Ptychography is a new X-ray imaging technique which bypasses the need for a lens by recording diffraction patterns from overlapping regions of a sample and then computationally inverting. This allows it to reach higher resolutions than what is possible with X-ray microscopy, where the fabrication of high quality zone-plate lenses limits the resolution, or scanning transmission X-ray microscopy, where the size of the focal spot is the limiting factor. This new technique will allow not only the study the nanostructure of relatively thick materials such as polymers or cement, providing critical insights on how they can be improved, but also of biological samples such as bone, or frozen cells. Post-doc Filipe Maia worked with Stefano Marchesini from the Advanced Light Source (ALS) on optimizing algorithms for different X-ray diffraction imaging experiments including ptychography, nanocrystallography, phase contrast tomography and single particle diffraction.

The work on ptychographic image reconstruction is part of the Nanosurveyor instrument to be installed at at the COSMIC facility at the ALS. When completed, the Nanosurveyor instrument will be moved there, powered by the ptychographic reconstruction software.

This instrument will allow users to easily obtain nanometer resolution images of a large variety of samples, including both 2D and full 3D images. Keeping up with the data rate of the fastCCD of 200 frames per second while producing reliable reconstruction from the diffraction data, such that the user gets real-time feedback, is the main challenge of the project. This will make Nanosurveyor the first real-time ptychographic imaging instrument in the world.

In the early phase of the project Maia translated an existing Matlab implementation of the

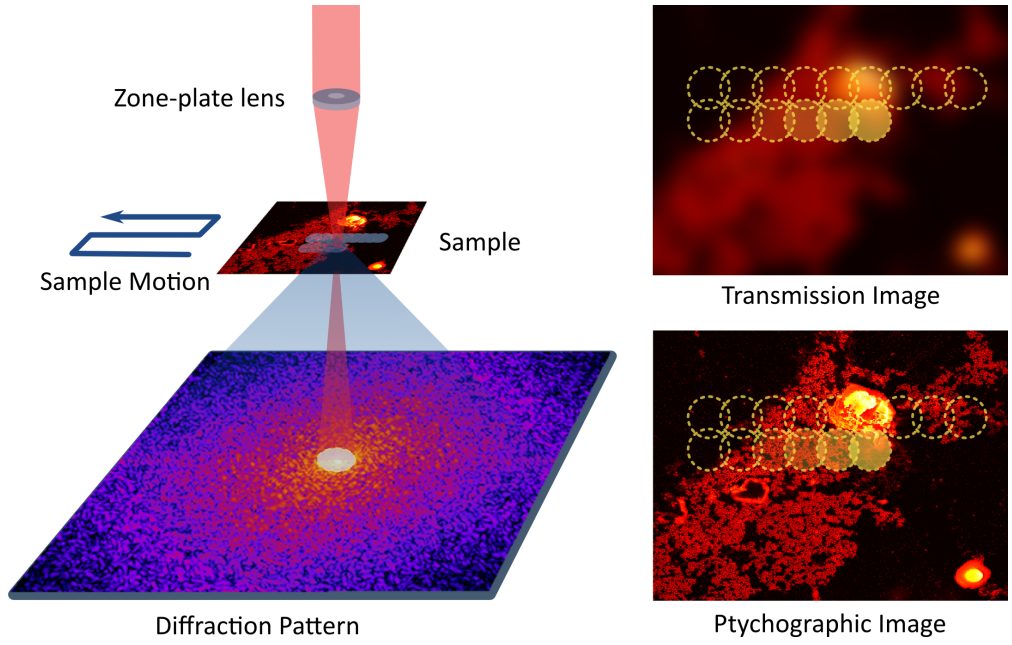


Figure 8: Schematic representation of a ptychographic experiments, and comparison with a traditional scanning transmission X-ray microscopy image, obtained from the same data.

reconstruction algorithm to C++ and CUDA. The initial code was serial and did not address several experimental difficulties, such as uncertainty in the beam position, and vibrations.

Maia devised a parallelization strategy for the algorithm, aiming to minimize communication while maximizing the convergence rate of the algorithm, which he then implemented [18]. The domain decomposition of the image introduced the problem of multiple parts of the image being offset by a random phase shift. This problem was corrected by the development of a patching code, which combines the different domains in a single coherent image. He then worked on adding the necessary techniques to correct for uncertainty in the beam position to the C++ code, based on Taylor expansion corrections first developed in Matlab [19].

Maia implemented a fast Radon transform using GPUs for phase contrast tomography [20], which when combined with fast CCD detectors can dramatically increase the throughput of tomographic beamlines at ALS. Maia also explored techniques for indexing diffraction patterns obtained from nanocrystallography experiments, a highly promising new technique [21].

Maia established a database for data from coherent X-ray imaging experiments, named CXIDB. It is available online at `http://cxidb.org`. Its goals are to broaden the availability of experimental data from new light sources such as the LCLS, to provide experimental datasets where algorithms could be tested and to help standardize data formats to facilitate exchange of data between different groups. The database currently contains about 3.5 million images.

7 Subsurface Geophysical Imaging

EMGeo is an advanced geophysical interpretation software for electromagnetic types of data. It leverages 21st century massively parallel computing resources and combines it with advanced electromagnetic measurement techniques, to provide a unique imaging capability, especially useful for hydrocarbon deposits.

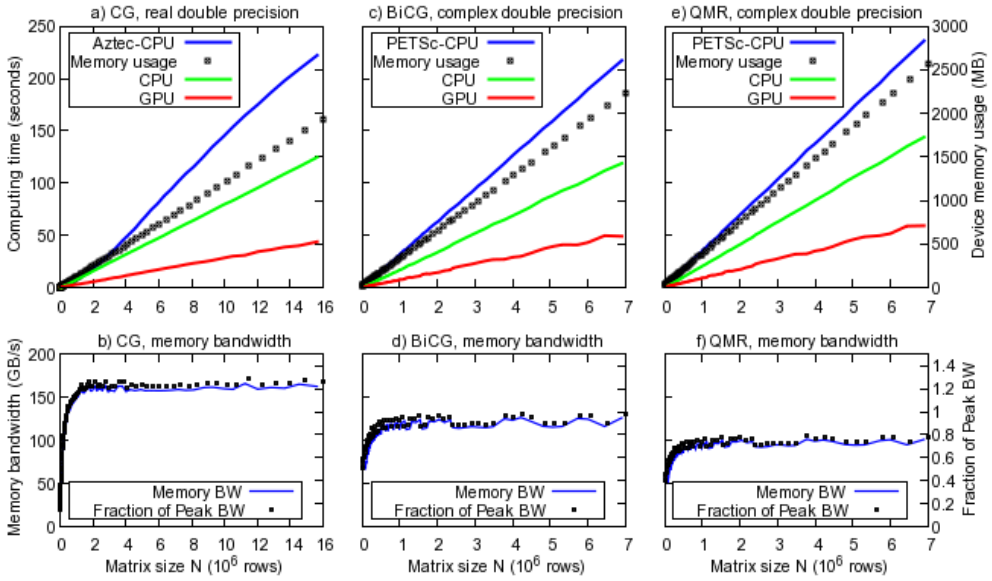


Figure 9: Solution times and memory bandwidth for the three different iterative Krylov solvers, CG (a,b), BiCG (c,d), and QMR (e,f). Each GPU implementation is compared against its original parallel FORTRAN90 version, as well as a parallel solver provided by an external library. All CPU solvers were run on eight processor cores.

Post-doc Maia devoted 20% of his time to work with Gregory Newman from the Earth Sciences Division on increasing the performance of Krylov solvers used in EMGeo, by taking advantage of the computing power of GPUs. The sparse matrices used in this project are relatively well structured, making them good candidates for the highly parallel nature of GPUs.

Maia started by studying which of the best matrix storage schemes would best suit the problem when it is run in GPUs, determining that the ELLPACK format is the most efficient. He then developed a single GPU implementation of the quasi minimal residual optimizer and BiCGStab, using efficient sparse-matrix vector multiplication routines.

Maia developed code to generate the sparse matrices directly in the GPU instead of copying them from main memory, thereby minimizing PCI-e bandwidth usage. He implemented a distributed version of the GPU Krylov solver using MPI by taking advantage of the existing communicating routines and changing the way data is passed to the GPU so that it is as close as possible to the CPU version. This reduces code duplication and improves the maintainability of the code. The communication routines were optimized by Maia to overlap the communication with the computation as much as possible decreasing the communication overhead [22].

8 Performance characterization for HPC applications

Over the last several decades, high performance computing endeavors have grown increasingly complex on both the hardware and the software fronts. With Moore's law still extant, we will see a 10–100× increase in on-chip parallelism in future manycore architectures. The upshot is that future applications are expected to scale in a commensurate manner in order to exploit such parallelism. Impediments to scalability in such an environment, as has been shown historically, are likely to come

from increased communication overhead, and from memory bandwidth limitations. Performance monitoring of HPC applications is therefore a critical element in the drive towards porting current applications to future architectures, particularly when the exascale era is ushered in.

Performance analysis aims to isolate bottlenecks to application performance. The insights given by the analysis can then be used for future optimization efforts. A tool based approach is usually taken, wherein one instruments the application under consideration with a performance monitoring tool, (e.g., IPM, CrayPAT, TAU, OpenSpeedShop) and then runs experiments with them to obtain various performance metrics that would help identify bottlenecks.

Post-doc Praveen Narayanan worked on several application in the areas of performance characterization and benchmarking. His initial focus was on fusion applications on the Cray XT/XE platform at NERSC followed by an examination of the user workload on the NERSC IBM iDataPlex platform in order to carry out a machine-wide performance analysis.

The performance characterization was primarily carried out using the profiling tools IPM and CrayPAT. IPM is a portable profiling infrastructure developed at NERSC for parallel codes. It provides a low-overhead performance profile of the performance aspects and resource utilization in a parallel program. Communication, computation, and IO are the primary focus. CrayPAT is a performance analysis tool offered by Cray for the XT/XE platform. It can be used to obtain various performance metrics such as communication, IO and computation, and may be used for profiling and tracing.

The fusion performance analysis was a collaborative effort with codes (see below) coming from several groups, containing a variety of underlying mathematical and computer programming models. The fusion codes considered for benchmarking may be classified as follows:

Particle-in cell approaches: GTS, VORPAL.

Discretized kinetic equations with implicit/explicit solvers: GYRO,NEO.

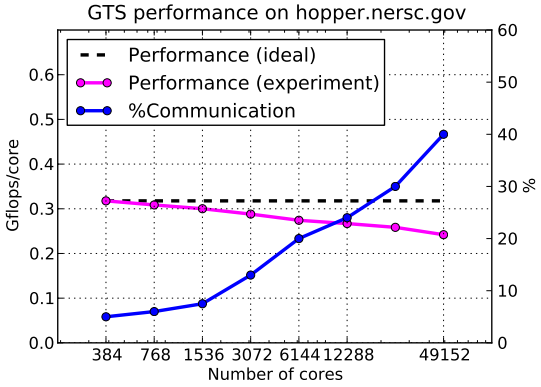
Discretized fluid equations requiring implicit solutions: BOUT++, M3D-C1,NIMROD.

Vignettes from the performance studies are presented. Figure 10 presents the flops performance and communication overhead for the four fusion codes GTS, GYRO, BOUT++ and VORPAL on the Cray XE6 platform, run at large concurrency. Performance degradation is seen as the concurrency is increased for all four codes. Also, by and large, the communication time as a percentage of the overall run time increases. As a general rule of thumb, it is noted that communication in excess of 50% of runtime implies significant performance degradation.

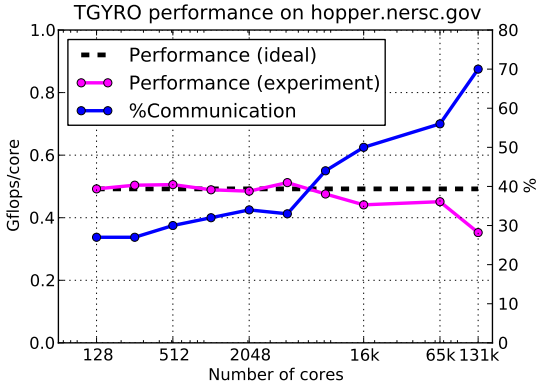
Having completed his initial profiling studies, Narayanan examined one of the component codes of the original study (BOUT++) in greater detail. This work was a collaboration with the authors of the PETSc applied mathematics library, in addition to the BOUT++ team. The profiling exercise brings to the fore the role of collectives in the time-stepping algorithms used by PETSc.

Together with his fusion profiling studies, Narayanan performed a workload analysis of runs on the carver-magellan supercomputing cluster at NERSC with a view to understanding how load imbalance affects runtime across a broad swath of runs from various users and at various concurrencies. A workload analysis such as this provides useful feedback on the computing infrastructure, for example, to help make decisions on future procurements based on the profile of the user workload. Logistically, this is carried out by writing parsers in Ruby and Perl, followed by writing the desired information into a database (MongoDB), which is queried and processed later.

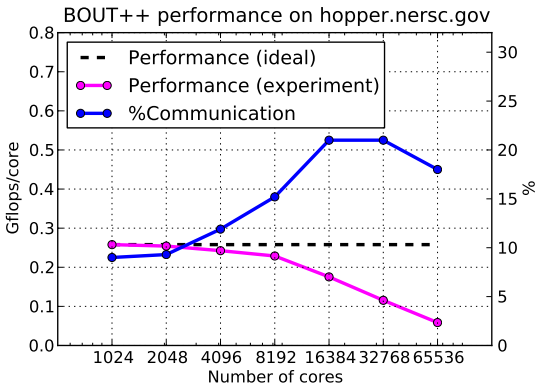
Narayanan's performance studies on the four fusion solvers GTS, GYRO, BOUT++ and VORPAL were presented at the CUG meeting in Fairbanks, Alaska in May, 2011 [23]. The follow



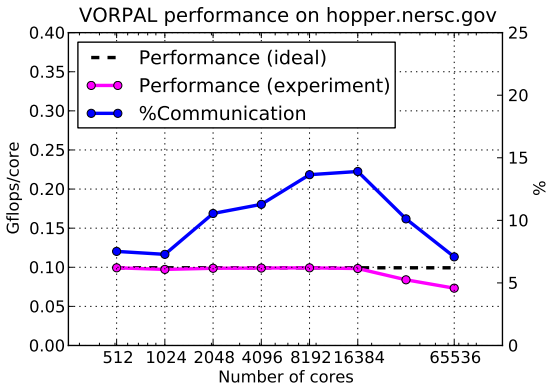
(a) GTS: weak scaling



(b) GYRO: weak scaling



(c) BOUT++: strong scaling



(d) VORPAL: weak scaling

Figure 10: Performance characteristics of fusion codes GTS, GYRO, BOUT++ and VORPAL on the Cray XE6. Scaling flops performance and communication as percentage of runtime for codes GTS (a), GYRO (b), BOUT++ (c) and VORPAL (d).

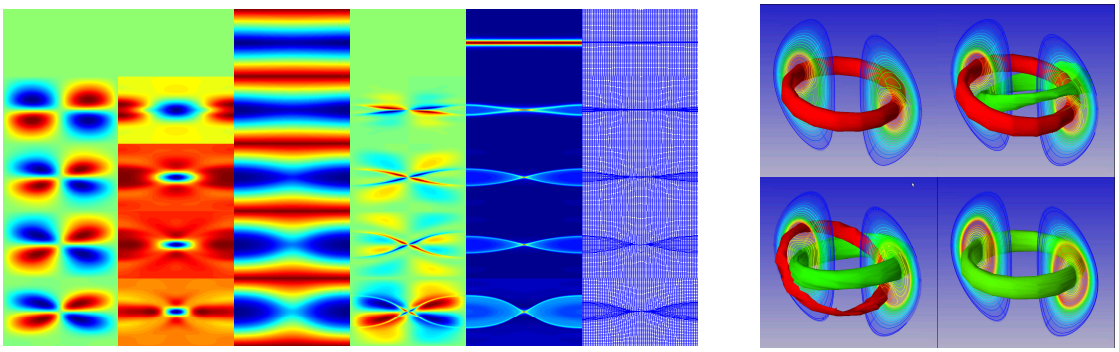
up performance study on the fusion solver BOUT++ (profiling the time-stepping algorithms implemented by PetSc) was presented at the International Conference on Numerical Simulation of Plasmas in September, 2011 [24].

9 Toward Practical Fusion Energy: Simulations of Wave-plasma Interactions

The next step toward fusion as a practical energy source is now on the way in the design and construction of ITER, in which the plasma will be heated by the high power radio frequency (RF) waves in the ion-cyclotron range of frequencies (ICRF). The SciDAC Center for Simulation of RF Wave Interactions with Magnetohydrodynamics (CSWIM) is developing an integrated modeling capability for plasma physics by focusing on the coupling RF waves with the MHD of the plasma as a prototype for comprehensive Fusion Simulation Project (FSP).

Post-doc Xuefei (Rebecca) Yuan worked with David E. Keyes from King Abdullah University of Science and Technology (KAUST) and Columbia University and Stephen C. Jardin from Princeton Plasma Physics Labratory (PPPL) on the numerical simulation of four-field extended MHD equations in dynamically adaptive curvilinear coordinates via Newton-Krylov-Schwarz (NKS). The MHD equations are transformed from Cartesian coordinates to solution-defined curvilinear coordinates and convergence and accuracy studies show that the curvilinear solution requires less computational effort due both to the more optimal placement of the grid points and to the improved convergence of the implicit solver. Because of the regularity of memory references that are inherited from the logically regular curvilinear coordinate grid, r-type refinement is certain to have advantanges over unstructured adaptive schemes in the multicore and SIMDized processor architectures of the future.

Figure 11 (a) presents a difficult case when the collisionless ion skin depth is large so that the Whistler waves, which cause fast reconnection, dominate the physics.



(a) adaptive grids (b) sawtooth cycle

Figure 11: Left: plots of the four fields (ϕ, V, ψ, B), the negative out-of-plane current density $-j$, and adaptive grids, where the collisionless ion skin depth $d_i = 1.0$; right: sawtooth cycle.

Yuan also worked with Sherry Li in the Computational Research Division (CRD) on a new algebraic hybrid linear solver, PDSLin (Parallel Domain decomposition Schur complement based Linear solver). PDSLin is based on a non-overlapping domain decomposition technique called

the Schur complement method, whereby subdomain problems can be solved by the direct solver SuperLU and the Schur complement system corresponding to the interface equations is solved using a preconditioned iterative solver. The enhanced scalability of PDSLin is attributed to its ability to employ hierarchical parallelism, namely, solving independent subdomain problems in parallel and using a subset of processors per subdomain. This requires only modest parallelism from SuperLU for each subdomain, and because the Schur system is often better condidioned than the original system, only a handful of iterations are needed for the Schur complement system,

The sawtooth cycle (shown in Figure 11 (b)) is an important dynamic that needs a better predictive model for ITER. The M3D code is a multilevel, 3D, parallel, plasma simulation code that is suitable for performing linear and nonlinear calculations of plasmas in toroidal topologies including tokamaks and stellarators. M3D-C^1 is an extension of M3D that uses higher-order finite elements and is fully implicit. For small system sizes, a direct solver such as SuperLU can be employed to obtain an accurate solution as long as the conditioner number is bounded by the reciprocal of the floating-point machine precision. However, SuperLU scales effectively only to 100s of processors or less. For larger system sizes, SuperLU can be used within a PETSc iterative solver as a block-Jacobi preconditioner to provide adequate solver performance, and it has also been shown that physics-based or other preconditioners can be applied to provide adequate solver performance. PDSLin has successfully been implemented in the M3D-C^1 code for the 2D planes, and the work for the 3D cases is ongoing.

Yuan's work on adaptive grids was presented at the 7th International Congress on Industrial and Applied Mathematics [25]. The work on linear solver was presented at the 22nd International Conference on Numerical Simulations of Plasmas [26] and published in the Computational Science and Discovery journal [27].

10 Modeling for the NDCX-II project

The Neutralized Drift Compression Experiment II (NDCX-II) is an ARRA-funded induction accelerator that started being commissioned in 2012. The final design calls for a 3 MeV, Li^+ ion beam, delivered in a bunch with characteristic pulse duration of 1 ns, and transverse dimension of order 1 mm when fully commissioned. The NDCX-II will be used in studies of material in the warm dense matter (WDM) regime, and ion beam/hydrodynamic coupling experiments relevant to heavy ion based inertial fusion energy.

ALE-AMR is one of the codes used for modeling NDCX-II. The code, which combines Arbitrary Lagrangian Eulerian (ALE) hydrodynamics with Adaptive Mesh Refinement (AMR), has physics models that include ion deposition, radiation hydrodynamics, thermal diffusion, anisotropic material strength with material time history, and advanced models for fragmentation. Some of the experiments at NDCX-II will explore the process of bubble and droplet formation (two-phase expansion) of superheated metal solids using ion beams. Thus, the physical model of surface tension needs to be incorporated into the code.

Post-doc Wangyi Liu improved the ALE-AMR code to model the NDCX-II experiment, especially the modeling of surface tension effect. Liu implemented two different surface tension models. The first one is a single-fluid diffuse-interface Korteweg-type model. It is implemented by adding a third-order space derivative to the stress. It is physically natural in that it does not make assumptions on the equation of state, but one cannot directly specify the surface width, and the implied surface width requires a too small mesh size.

The second model is a two-fluid volume-of-fluid model. This method calculates the curvature based on the height function within a 3*7 scheme. Although this is a two fluid model, Liu imple-

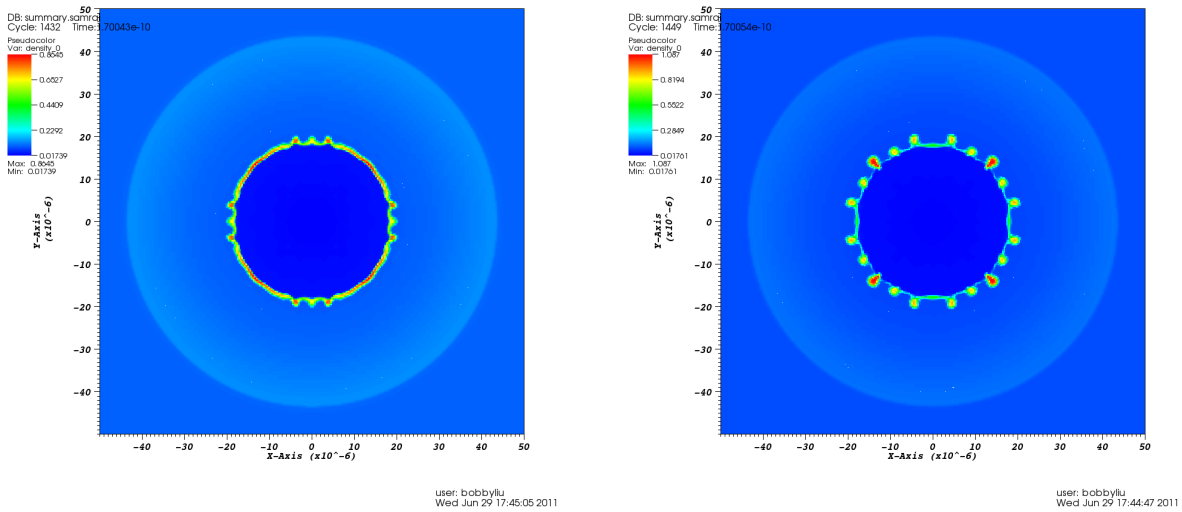


Figure 12: Results of ALE-AMR simulations of an expanding aluminum shell without the surface tension model (left) and with the surface tension model (right). The result with the surface tension model forms droplets, while the other does not.

mented a version that converts single material density into volume of fluid, and utilizes the same algorithm for a single-fluid version. In addition, this algorithm is implemented in 3D as well.

Both of these models were benchmarked to the classical test cases, e.g. the Laplacian equation, ellipsoid oscillation and the Rayleigh-Plateau instability. In addition, a test case was run where a heated aluminum shell expands due to the high pressure of the aluminum vapor inside; see figure 10. Both results show material breakup due to physical instability of the spinodal region, but the surface tension model clearly forms droplets, while the other one does not.

Apart from the potential NDCX-II simulation, this model is also used for an EUV lithography source simulation, in which a laser pulse of lower energy is used to hit a tin droplet. The droplet flattens, and surface tension is important in this situation.

Liu's work was presented in Scidac [28,29] and APS-DPP meetings [30,31]. Most recently Liu's work was included in two invited talks in Japan and China [32,33].

11 Algorithm Studies for Fluid Modeling on Structured Adaptive Grids

One of the main bottlenecks in the development of solution methods for the equations describing incompressible flow and other low-Mach number models for fluid dynamics that will scale to exascale systems is the ability to solve Poisson's equation $-\Delta\phi = \rho$. This is particularly true for Structured Adaptive Mesh Refinement (SAMR) grids, which are used extensively in a number of problems for DOE, including microscale modeling for carbon sequestration, as performed in the EFRC Center for Nanoscale Control of Geologic CO_2, led by LBNL.

The principal difficulty is that the solution to Poisson's equation is nonlocal, with the solution operator coupling all points in space with one another. The fact that the nonlocal coupling is smooth (in fact, real analytic) is the basis for fast algorithms for such problems, such as geometric multigrid. While these methods successfully exploit local regularity to minimize the number of floating point operations, they have been less successful in minimizing communications costs on high-end parallel

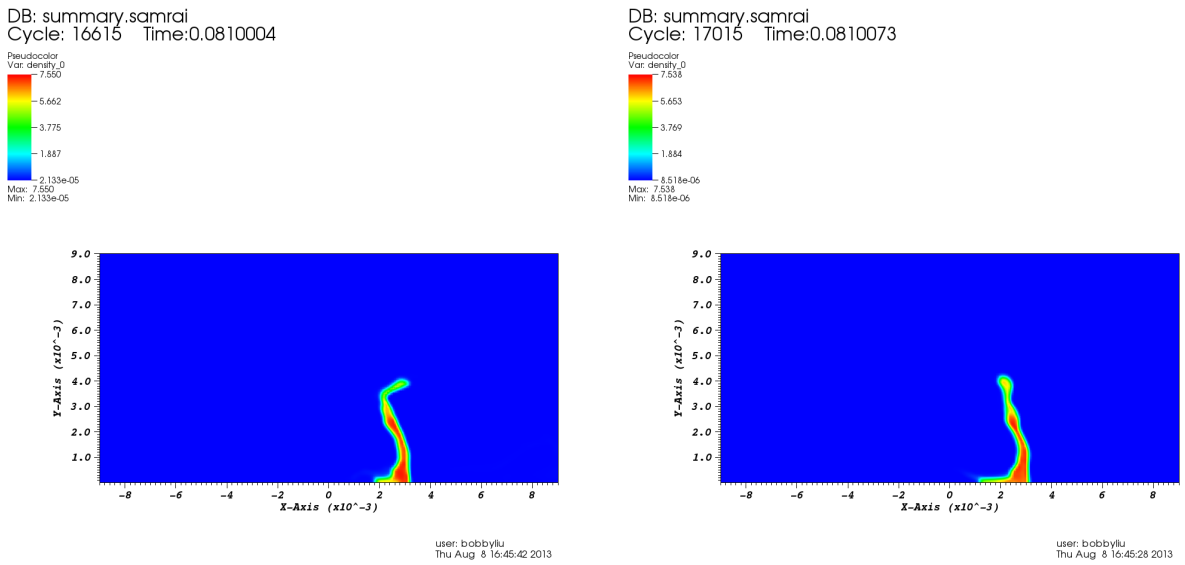


Figure 13: Results of ALE-AMR simulations of EUV heating on a tin droplet. Instead of a complete breakup, the material is flattened. Surface tension model (left) impede the break of the material.

systems, particularly in systems with high latency. In addition, the stencil operations that form the computational kernels for these problems will be difficult to optimize on multicore systems with large numbers of cores.

For these reasons, an alternative approach was pursued, based on computing solutions of Poisson's equation on a union of rectangular patches in terms of a local convolution with the charge on each patch, combined with a global solution for the right-hand side obtained from applying the coarse-grid Laplacian to the local solutions, and summing. The complete solution at any point is expressed as the sum of local solutions, plus a correction interpolated from the coarse grid from which the contributions from the local solutions have been subtracted. This Method of Local Corrections AMR (MLC-AMR) algorithm can be applied adaptively, with the patches comprising a fixed number of grid points, and recursively, with the same algorithm used to compute the solution at the next coarser level. MLC-AMR addresses the problems with conventional multigrid described above by reducing the communications costs by an order of magnitude, and by using as its compute kernel for a single patch, a combination of Fast Fourier Transforms (FFTs) and a simplified form of the fast multipole method, both of which can be optimized to obtain high performance on multicore nodes.

An initial version of this algorithm was developed at LBNL by P. McCorquodale and P. Colella, but its accuracy was limited by first-order accurate representation of far field nonlocal coupling, which was represented by communication of the monopole moments for each patch over longer ranges. A new version of the MLC that generalizes the original algorithm to allow for arbitrarily high-order accuracy for the nonlocal component of the representation is being developed. The key step is to perform the long-range calculation using the fields induced by the leading order terms in a Legendre expansion of the charge on a patch, of which the monopole term is the first term. If we use Legendre polynomials up to degree $P-1$, then we obtain a P^{th}-order accurate method. As in the monopole case, the communication of the long range fields is done by communicating the coefficients in the Legendre expansion of the charge, which is a relatively modest communication cost: for $P = 4$, for example, this corresponds to 20 coefficients per patch. Post-doc Christos

Kavouklis worked on this new approach with Dr. Phillip Colella, director of the Applied Numerical Algorithms Group at LBNL. Kavouklis derived and documented a detailed design for the algorithm that has been implemented in the Chombo SAMR Library.

Cover figure description: the physical solutions and the negative out-of-plane current density with a time-dependent sequence of curvilinear coordinates for the four-field extended MHD equations in magnetic reconnection simulation for the collsionless ion skin depth $d_i = 1.0$ at $t = 0.0, 10.0, 20.0, 30.0, 40.0$ (from top to bottom). From left to right: the stream function for the in-plane component of the ion velocity; the z component of the ion velocity; the stream function for the in-plane component of the magnetic field; the z component of the magnetic field; the negative out-of-plane current density; the adaptive grids. Two parameters in the density functions are $m = 1.0$, $r = 50.0$, and only $1/16$ grid points are plotted in each direction for a clear view of the adaptive grids.

Post-doc Publications, Posters and Talks

[1] J. Kim, A. Koniges, R. Martin, M. Haranczyk, J. Swisher, and B. Smit, "GPU Computational Screening of Carbon Capture Materials," *Poster and Proceedings Paper at: SciDAC 2011, Denver, CO*, 2011.

[2] J. Kim, A. Koniges, R. Martin, M. Haranczyk, J. Swisher, and B. Smit, "Computational Screening of Novel Carbon Capture Materials." Presented at GTC 2012, San Jose, CA.

[3] J. Kim, J. Rodgers, M. Athenes, and B. Smit, "Molecular Monte Carlo Simulations Using Graphics Processing Units: To Waste Recycling or Not?," *Journal of Chemical Theory and Computation*, vol. 7, p. 3208, 2011.

[4] A. Koniges, R. Preissl, J. Kim, D. Eder, A. Fisher, N. Masters, V. Mlaker, S. Ethier, W. Wang, M. Head-Gordon, and N. Wichmann, "Application Acceleration on Current and Future Cray Platforms," in *CUG*, 2010, **Best paper award winner**.

[5] J. Kim, A. Koniges, B. Smit, and M. Head-Gordon, "Calculation of RI-MP2 Gradient Using Fermi GPUs." Poster at: Molecular Quantum Mechanics 2010, Berkeley, CA.

[6] R. Preissl, A. Koniges, S. Ethier, W. Wang, and N. Wichmann, "Overlapping communication with computation using OpenMP tasks on the GTS magnetic fusion code," *Scientific Programming*, vol. 18, pp. 139–151, August 2010.

[7] R. Preissl, N. Wichmann, B. Long, J. Shalf, S. Ethier, and A. Koniges, "Advanced Communication Techniques for Gyrokinetic Fusion Applications on Ultra-Scale Platforms," in *PGAS*, 2011.

[8] R. Preissl, N. Wichmann, B. Long, J. Shalf, S. Ethier, and A. Koniges, "Multithreaded Global Address Space Communication Techniques for Gyrokinetic Fusion Applications on Ultra-Scale Platforms," in *SC'11*, 2011.

[9] G. S. Pau, J. B. Bell, A. S. Almgren, K. M. Fagnan, and M. J. Lijewski, "An Adaptive Mesh Refinement Algorithm for Compressible Two-Phase Flow In Porous Media," *Computational Geoscience*, vol. 16, p. 577, 2012.

[10] K. Fagnan, G. Pau, M. Lijewski, and N. Wright, "Acceleration of porous media simulations on the Cray XE6 platform," *Proceedings from the Cray User Group Meeting*, May 2011.

[11] J. Corlett, B. Austin, K. Baptiste, J. Byrd, P. Denes, R. Donahue, L. Doolittle, R. Falcone, D. Filippetto, S. Fournier, D. Li, H. Padmore, C. Papadopoulos, C. Pappas, G. Penn, M. Placidi, S. Prestemon, D. Prosnitz, J. Qiang, A. Ratti, M. Reinsch, F. Sannibale, R. Schlueter, R. Schoenlein, J. Staples, T. Vecchione, M. Venturini, R. Wells, R. Wilcox, J. Wurtele, A. Charman, E. Kur, and A. Zholents, " A Next Generation Light Source Facility at LBNL," *PAC 11 Conference Proceedings*, 2011.

[12] B. Austin, J. Qiang, J. Wurtele, and A. Koniges, "Influences of architecture and threading on the MPI communication strategies in an accelerator simulation code.," *Poster at: SciDAC 2011, Denver, CO*, 2011.

[13] J. Chou, K. Wu, M. Howison, Prabhat, O. Ruebel, B. Austin, E. W. Bethel, J. Qiang, R. D. Ryne, and A. Shoshani, "Parallel Index and Query for Large Scale Data Analysis," in *Proceedings of ACM/IEEE Supercomputing 2011 Conference (SC11)*, p. 1, November 2011.

[14] M. Reinsch, B. Austin, J. Corlett, L. Doolittle, G. Penn, D. Prosnitz, J. Qiang, A. Sessler, M. Menturini, and J. Wurtele, "System Trade Analysis for an FEL Facility," *Poster at: Free Electron Laser Conference FEL 2011, Shanghai, China*, 2011.

[15] J. Wang, D. Domin, B. Austin, D. Y. Zubarev, J. McClean, M. Frenklach, T. Cui, and W. A. Lester, Jr., "A Diffusion Monte Carlo Study of the O-H Bond Dissociation of Phenol," *J. Phys. Chem. A*, vol. 114, no. 36, p. 9832, 2010.

[16] B. Austin, D. Y. Zubarev, and W. A. Lester, Jr., " Quantum Monte Carlo and Related Approaches," *Chemical Reviews*, vol. 112, p. 263, 2012.

[17] D. Y. Zubarev, B. Austin, and W. A. Lester, Jr., " Quantum Monte Carlo Simulation of X-ray Absorption Spectrum of Pyrrol at the Nitrogen K-edge," *The Journal of Chemical Physics*, vol. 136, p. 144301, 2012.

[18] F. R. N. C. Maia, C. Yang, A. Schiroztek, and S. Marchesini, "Parallel real-time ptychographic image reconstruction using GPUs.," *Poster at: SciDAC 2011, Denver, CO*, 2011.

[19] C. Yang, J. Qian, A. Schirotzek, F. Maia, and S. Marchesini, "Iterative Algorithms for Ptychographic Phase Retrieval," *arXiv:1105.5628v1 [physics.optics]*, 2011.

[20] F. R. N. C. Maia, A. MacDowell, S. Marchesini, H. A. Padmore, D. Y. Parkinson, J. Pien, A. Schirotzek, and C. Yang, "Compressive phase contrast tomography," *Proc. SPIE*, vol. 7800, p. 78000F, 2010.

[21] F. R. N. C. Maia, C. Yang, and S. Marchesini, "Compressive auto-indexing in femtosecond nanocrystallography.," *Ultramicroscopy*, vol. 111, p. 807, 2010.

[22] M. Commer, F. R. N. C. Maia, and G. Newman, "Iterative Krylov solution methods for geophysical electromagnetic simulations on throughput-oriented processing units," *International Journal of High Performance Computing Applications*, vol. 26, p. 378, 2012.

[23] P. Narayanan, A. Koniges, L. Oliker, R. Preissl, S. Williams, N. Wright, M. Umansky, X. Xu, B. Dudson, S. Ethier, W. Wang, Maxim, J. Candy, and J. Cary, "Performance characterization for Fusion Co-design Applications." Presented at: CUG 2011, Fairbanks, AK, 2011.

[24] S. Farley, B. Dudson, P. Narayanan, L. C. McInnes, S. Balay, J. Cary, A. Koniges, M. Umansky, C. Woodward, X. Xu, and H. Zhang, "BOUT++: Performance Characterization and Recent Advances in Design," *Poster at: the 22nd International Conference on Numerical Simulation of Plasmas (ICNSP'2011), Long Branch, New Jersy, USA*, September 2011.

[25] X. Yuan, S. C. Jardin, D. E. Keyes, and A. E. Koniges, "Implicit coupling of Lagrangian grid and solution in plasma simulation with additive Schwarz preconditioned inexact Newton," *Poster at: the 7th International Congress on Industrial and Applied Mathematics (ICIAM'2011), Vancouver, BC, Canada*, July 2011.

[26] X. Yuan, X. S. Li, I. Yamazaki, S. C. Jardin, A. E. Koniges, and D. E. Keyes, "Application of PDSLin to the magnetic reconnection problem," *Poster at: the 22nd International Conference on Numerical Simulation of Plasmas (ICNSP'2011), Long Branch, New Jersey, USA*, September 2011.

[27] X. Yuan, X. S. Li, I. Yamazaki, S. C. Jardin, A. E. Koniges, and D. E. Keyes, "Application of PDSLin to the magnetic reconnection problem," *Computational Science and Discovery*, vol. 6, p. 014002, 2013.

[28] W. Liu, J. Barnard, A. Friedman, N. Masters, A. Fisher, V. Mlaker, A. Koniges, and D. Eder, "Modeling droplet breakup effects in warm dense matter experiments with diffuse interface methods in ALE-AMR code," *Proceedings of Scientific Discovery through Advanced Computing Program Conference 2011*, 2011.

[29] W. Liu, J. Barnard, A. Friedman, N. Masters, A. Fisher, V. Mlaker, A. Koniges, and D. Eder, "Modeling droplet breakup effects in warm dense matter experiments with diffuse interface methods in ale-amr code." Poster at: SciDAC conference, 2011.

[30] W. Liu, J. Barnard, A. Friedman, N. Masters, A. Fisher, A. Koniges, and D. Eder, "Modeling droplet breakup effects with applications in the warm dense matter ndcx experiment." Poster at: American Physics Society Department of Plasma Physics 2011.

[31] W. Liu, J. Barnard, A. Koniges, D. Eder, N. Masters, A. Fisher, and A. Friedman, "A numerical scheme for including surface tension effects in hydrodynamic simulation: a full korteweg type model without parasitic flows." Poster at: American Physics Society Division of Plasma Physics 2012.

[32] A. Koniges, W. Lui, S. Lidia, T. Schenkel, J. Barnard, A. Friedman, D. Eder, A. Fisher, and N. Masters, "Advanced Target Effects Modeling for Ion Accelerators and other High-Energy-Density Experiments," *Invited Talk at: 8th International Conference on Inertial Fusion Sciences and Applications, Nara, Japan, 2013*, September 2013.

[33] A. Koniges, W. Lui, Y. He, D. Eder, A. Fisher, N. Masters, and R. Anderson, "Multi-Material Plasma Modeling on HPC Platforms," *Invited Talk at: the 23nd International Conference on Numerical Simulation of Plasmas (ICNSP'2013), Beijing, China*, September 2013.