Virtual Experiments on the Neutron Science TeraGrid Gateway

V. E. Lynch1, J. W. Cobb1, E. Farhi2, S. D. Miller1, and M. Taylor2
Oak Ridge National Laboratory1 and Institut Laue-Langevin2

The TeraGrid’s outreach effort to the neutron science community is creating an environment that is encouraging the exploration of advanced cyberinfrastructure being incorporated into facility operations in a way that leverages facility operations to multiply the scientific output of its users, including many NSF supported scientists in many disciplines. The Neutron Science TeraGrid Gateway serves as an exploratory incubator for several TeraGrid projects. Virtual neutron scattering experiments from one exploratory project will be highlighted.

I. INTRODUCTION

The Neutron Science TeraGrid Gateway (NSTG) connects large neutron science instrument facilities with the cyberinfrastructure resources of the TeraGrid. Neutron science research enables developments in areas of materials research such as chemistry, complex fluids, crystalline materials, disordered materials, engineering, magnetism and superconductivity, polymers, and structural biology. Neutron scattering is a tool for researching the structure and dynamics of materials at the molecular level. The Spallation Neutron Source (SNS) at Oak Ridge National Laboratory and The Institut Laue-Langevin (ILL) in France are world centers for research in neutrons. They both have new capability for materials research with neutron scattering and need new cyberinfrastructure resources.

One example of how NSTG can enable neutron science is the recently demonstrated virtual neutron scattering experiments for several materials using the TeraGrid’s parallel computing resources. These virtual experiments have the ability to simulate a beam line from the moderator to detector. Theory driven simulation allows comparison at detector raw count level. This approach is new for the neutron community and opens up new vistas for experiment planning and analysis. However, the required computational time is much larger than neutron scientists have routinely used previously as an adjunct to experiments. We have demonstrated the ability of doing exactly this using a composite simulation tool that includes McStas[1-5], VASP[6-8], and nMoldyn[9-10] executing in a parallel high performance computing environment on the TeraGrid[11]. Virtual neutron scattering experiments require several steps. The VASP package is used for the time-consuming ab initio molecular dynamics calculations and the trajectory to $S(q, \omega)$ converter is the nMoldyn package. $S(q, \omega)$ is the neutron dynamic structure factor - or scattering law - which is a function of $q$, the vector momentum transfer, and $\omega$, the frequency. The virtual experiments are completed using the McStas package with $S(q, \omega)$ input. This is only one example of many potential future integrated uses of cyberinfrastructure with the conduct of experiments. The future integrated use of cyberinfrastructure is anticipated to grow rapidly.

The NSTG serves as an exploratory incubator for several TeraGrid projects that can eventually grow large enough to require their own allocation. NSTG is “jump starting” this process by incubating new projects and then supporting them as they mature within the TeraGrid. The NSTG functions both as a gateway for neutron community use and as outreach to a new user community. Present neutron science incubator projects include virtual neutron scattering experiments, interpretation of diffuse scattering data measured by single-crystal spallation neutron diffraction, e-Science applications, computer simulations in molecular biophysics combining neutron scattering and bioenergy, and parallel reduction of experimental data. The molecular biophysics incubation project recently matured and now receives its own NSF allocation.

As SNS continues to ramp up, we expect ramp up not only from successful outreach to the community but increases as the numbers of SNS users increase. With many proposals expected for the eighteen instruments of the SNS, the need for data reduction, data analysis, and simulations is expected to soon overwhelm local computing resources. Access to the resources of the TeraGrid is critical for these needs. With multiple members of each proposal, having accounts on each TeraGrid resource is not feasible, so NSTG accesses those resources using a community certificate from the gateway portal and then uses TeraGrid community accounts. We are following all TeraGrid requirements for these accounts and logging each user’s access through the portal. We use multiple TeraGrid sites for these calculations and use tools developed by the TeraGrid science gateways.

Closer integration of the SNS and the TeraGrid is the goal of this work. In the future, NSTG will flow streaming data from SNS for processing on the TeraGrid and move data from the SNS data repository for further processing and analysis on the TeraGrid. Also universities with SNS proposals will use the high bandwidth interconnect of the TeraGrid to support data movement between SNS and their institution. Real-time simulations of experiments will use TeraGrid
resources to predict experimental results as they are running or between experiments.

II. CYBERINFRASTRUCTURE

Creating a connection between discipline science and national scale Cyberinfrastructure requires an understanding of specific needs and how to meet those needs. The TeraGrid provides unprecedented and unique cyber resources. This includes creating and using advanced data storage, caching, and transport tools. It also includes execution orchestration of scientific tasks across TeraGrid resources and more generally across multiple distributed Cyberinfrastructures, remote and local.

Large national and international experimental user facilities have inherent cyberinfrastructure needs. In the past, many of those needs were met with internally provisioned cyberinfrastructure components. Sometimes this effort was executed excellently and has provided seminal leadership in the effort to define and build a national scale Cyberinfrastructure while at other facilities the required cyberinfrastructure was executed poorly or almost overlooked. An NSTG focus is to provide assistance to facilities in the planning and execution of the cyberinfrastructure needs. The NSTG is testing the hypothesis that a large user facility can rely on a national scale cyberinfrastructure, TeraGrid, to meet some or all of its needs and that it can be done in a more effective and cost efficient manner. The NSTG is working in close collaboration with the Spallation Neutron Source (SNS) at Oak Ridge as their principal facility partner. The SNS is a next-generation neutron source. It has completed construction at a cost of $1.4 billion and is ramping up operations. The SNS will provide an order of magnitude greater flux than any other facility in the world and will be available to all of the nation’s scientists, independent of funding source, on a reviewed basis. With this new capability, the neutron science community is facing order of magnitude larger data sets and new ways to manage and analyze data to optimize experimental results as they are running or between experiments.

As part of the SNS partnership, another focus area of the NSTG is to provide Cyberinfrastructure outreach, community development, and user support for the neutron science community. This includes not only SNS staff and users but extends to all six neutron scattering centers in North America and several dozen worldwide. There is also common interaction with light sources and other facilities as well as interest and interaction with instrument/cyberinfrastructure projects in general. The NSTG approaches these broad goals in two ways. First the staff that works for the NSTG effort predates the TeraGrids Science Gateway effort, but is now part of that effort. NSTG staff interacts with facilities (primarily the SNS) and with neutron science users. Second, the NSTG offers a modest local compute resource, the NSTG cluster. The NSTG cluster is a full TeraGrid computational resource. It deploys all required CTSS components and has an excellent uptime and availability record as evidenced from system logs and INCA reporters. The NSTG cluster operations are fully integrated into TeraGrid resource management including account creation and management, usage reporting, trouble ticket processing, and security coordination. The size of the NSTG cluster is, by design small. The budget for the NSTG effort is modest in TeraGrid terms (3% of total). While the NSTG would welcome the opportunity of providing a large computational resource for the TeraGrid, it is not possible to accomplish while maintaining the NSTG focus on integrating user facilities. Instead, the purpose of the NSTG cluster is to provide a local resource that is completely TeraGrid integrated. This makes the offer of TeraGrid integration easier in that the facility can have a certain level of comfort that their initial compute resource target is local. Once they become proficient in the utilization of TeraGrid, the move to TeraGrid at large is less daunting and is eased by the fact that NSTG cluster operations are a microcosm of all TeraGrid operations. In this was the NSTG cluster is a starter system for facility integration. The TeraGrid’s common software then allows NSTG users to utilize any TeraGrid resource. General trends are that the resource is highly available, well configured, well maintained and securely operated. Because of its small and underutilized status, the NSTG cluster has also been a good test-bed for TeraGrid technology rollout.

III. VIRTUAL EXPERIMENT WORKFLOW

A successful use of the NSTG cluster has been for neutron scattering virtual experiment calculations. The ability to construct and perform accurate virtual experiments is highly desirable for a number of reasons. Some examples are the following: To find the optimum conditions and parameters for performing real measurements, saving time and resources that would otherwise be spent searching manually for such conditions which could be used for a worthwhile experiment. To perform virtual experiments where it would otherwise be impossible or prohibitive in some way to perform a real experiment is another reason. To help in understanding an experimental measurement, e.g. separating the different contributions in the total signal (sample itself, container, sample environment, other background sources). In order to perform such experiments it is necessary to be able to simulate accurately scattering from all the parts of the experiment which are found in the neutron beam this includes containers, cryostats etc. as well as the chosen sample. As such it is...
necessary to be able to simulate accurately the materials which make up these components or are otherwise commonly found in the beam.

A number of different materials were investigated for simulation with differing success owing to constraints in time, resources and technical feasibility. The materials simulated are Aluminum, Copper, Vanadium, Indium, Silicon, Niobium, Alumina (Al2O3), Molybdenum, Cadmium, Liquid Helium, Silica (SiO2), and common gasses. Where the above materials are solid it was the polycrystalline or powder form which was considered.

VASP (Vienna Ab-initio Simulation Program) [6-8] is a package for performing ab-initio quantum-mechanical molecular dynamics using pseudopotentials and a plane wave basis set. This means that all calculations are performed from a purely quantum mechanical basis with no approximations, this should lead to good accuracy of calculation but is very computationally intensive, thus can only be performed for small numbers of atoms. Forces and stress can also be easily calculated with VASP and used to relax atoms into their instantaneous ground state. VASP was used to perform the bulk of the calculations, both optimization of initial structures and later molecular dynamics simulations, the results of which were output to nMOLDYN. Typical computations use twelve parallel processors for the order of weeks to months (in some cases) depending on the computer resources. Using more processors does not help since VASP has poor parallel computational efficiency, due to FFT operations on large matrices, which need lots of inter-node communications.

nMOLDYN [9-10] is an interactive analysis program for Molecular Dynamics simulations. It is especially designed for the computation and decomposition of neutron scattering spectra, but also computes other quantities. nMOLDYN produces an $S(q,ω)$ map from an input molecular dynamics simulation, thus turning what is just a description of how the atoms of a sample move over time into something which will immediately tell you how a sample will scatter neutrons. nMOLDYN can only sample the momentum values with a step higher than 0.1 Å$^{-1}$ as the code was designed for liquids and amorphous systems. This step is certainly too large to accurately describe the crystal structure of polycrystalline powders. However, an approximation of the diffraction pattern is obtained, together with the inelastic contributions. nMOLDYN’s serial computation time is typically of the order of an hour.

Phonon [12] is a program for calculating phonon dispersion curves, and phonon density spectra of crystals, crystals with defects, surfaces, adsorbed atoms on surfaces, etc. from either a set of force constants, or from a set of Hellmann-Feynman forces calculated within an ab initio program (in this case VASP). Phonon can also be used to create position files for VASP for crystals from their type and lattice constants. Phonon can also compute the coherent inelastic part of the neutron scattering signal, but this is not enough to model the complete scattering from samples, it can be used however to benchmark VASP simulations. Phonon’s serial computation time is typically of the order of a few hours in our case.

McStas [1-5] is a general tool for simulating neutron scattering instruments and experiments. When provided with $S(q,ω)$ maps for the coherent and incoherent scattering in a material and a list of simple constants relating to it (total coherent and incoherent cross sections for the material etc.) the differential cross section of the material can be calculated. This result can be used to simulate how an incident neutron of given energy will scatter off the material and thus allows virtual experiments to be performed. The instrument description used in the virtual experiments performed in the report is derived from the ILL TOF_Env example instrument description from the McStas package, which can simulate results for time-of-flight instruments (like the IN4 and IN5 instruments at the ILL from which experiment data was taken). Computation time is of order a few minutes to simulate a simple experiment on 32 processors, but it scales well to hundreds of processors for larger calculations. The McStas calculations are done in parallel on the TeraGrid using MPI. One possible McStas output format is NeXus. The NeXus files can be viewed using the same graphics software available in the Neutron Science Portal for the SNS experimental data.

Most of the materials considered were polycrystalline solids and thus most were simulated using the same procedure (unless otherwise stated), which is described here.

The initial construction of a single crystal cell was done in Phonon, using published literature values for the structures of the materials and other physical properties.

The meat of the calculations were done using VASP, which was run both on machines at the ILL using stock ILL executables, and on the NSTG TeraGrid cluster at ORNL in the United States (10,000 hours on grant TG-DMR060000T) using VASP executables compiled specifically for this purpose.

A description of the single cell created in Phonon was exported to and its geometry optimized in VASP, from which the cell parameters in Phonon were updated and used to create a super cell, of typically
order 100 atoms, which was again optimized using VASP.

The optimized geometry created was then checked by computing phonon dispersion curves in Phonon and comparing with curves taken from literature. The curves were computed by taking a file of atom displacements from Phonon, which was used together with the structure generated in VASP to create structures with displaced atoms. These were run through a single point energy calculation in VASP to give forces on the atoms. These in turn were used to create a file containing the Hellmann-Feynmann forces which was exported back into Phonon so that phonon dispersion curves could be computed, which could be used as an aid in determining if the geometry optimization yielded feasible results.

If the geometry optimization was deemed successful calculation of the molecular dynamics of the material at a chosen temperature could then be performed in VASP, first using an nvt ensemble to achieve a structure with some level of stability around the desired temperature and then using an nve ensemble calculation of around 30ps with 3fs timesteps. The results were input to nMOLDYN to generate coherent and incoherent structure factors. These structure factors were then converted into a format readable by McStas making use of a Matlab script for this purpose.

Once these structure factor matrices were obtained they could be used to perform virtual experiments which could then be compared with real experiments performed on instruments at the ILL. Where data of experiments for materials was readily available it was used to compare with simulation. However in certain cases experiments were performed to provide data for comparison and in others suitable data is not yet available. They were however compared with the coherent inelastic dynamics part of the simulations, in order to check their validity and accuracy.

**IV. VIRTUAL EXPERIMENT RESULTS**

To show the agreement between experimental data and virtual experiment results, the results of aluminum and copper virtual experiments will be presented.

The obtained value of the lattice constant for Aluminum differs only slightly from the measured value, this improves the likelihood of accurate final results especially for the structure i.e. position of Bragg peaks etc. The phonons correlated reasonably well with literature values (Figures 1,2), with just one missing mode in the (110) direction the only real difference between them.
Figure 4 – Bragg peaks for Aluminum at 300 K from virtual experiment at 1.1Å.

Figure 5 – Experimental TOF data measured using IN4 at the ILL for Aluminum can at 300K using 1.1Å beam (log scale)

Figure 6 – Simulated TOF data for Aluminum can using 1.1Å beam (log scale)

It can be seen that the main features of the real experiment (Figure 5) are reproduced well in the simulation (Figure 6). The simulation seems to give a fairly faithful representation of the scattering from Aluminum. A close look to the $S(q,\omega)$ map in Figure 7 shows energy stripes in log scale. This artifact is caused by a resonance between the phonon modes and the VASP simulation box size. While not ideal the effect is relatively small in this case and hence does not significantly alter the result, especially as the virtual experiment results are smoothed by the instrument resolution function.

Figure 7 – Aluminum coherent $S(q,\omega)$ map from nMOLDYN (log scale)

For the copper virtual experiment, the phonons correlated reasonably well with literature values (Figures 8,9), one missing mode in the (110) and slight discrepancy in one mode for the (100) direction, otherwise good agreement.

Figure 8 – Phonon dispersion curves for Copper in given crystal directions at 300K [14]

Figure 9 – Produced phonon dispersion curves for copper in directions (100) left, (110) middle, (111) right.
As can be seen for copper (Figure 10,11) the structure produced, namely the Bragg peaks, corresponds very well with what would be expected/measured. The positions of the Bragg peaks are accurately reproduced up to high angles (corresponding to maximum calculated values of $q$ at $10\,\text{Å}^{-1}$).

**Figure 10** – Theoretical Bragg peaks for Copper at 1.1Å incoming neutrons wavelength.

**Figure 11** – Bragg peaks for Copper at 300 K from virtual experiment at 1.1Å incoming neutrons wavelength.

**Figure 12** – Simulated TOF data for Copper plate using 1.1Å beam (log scale)

As can be seen the simulation (Figure 12) gives a very good reproduction of the experimental data (Figure 13), replicating all main features with good accuracy. As with Aluminum, there is some striping present in the Copper simulation (Figure 14), but it is not sufficient to seriously disrupt the results. The lack of intensity in Figure 12 at $90^\circ$ is due to the absorption in Copper for a perpendicular sample plate geometry.

**Figure 13** – Experimental TOF data for Copper plate measured using IN4 at the ILL at 175K using 1.1Å beam (log scale)

**Figure 14** – Coherent $S(q,\omega)$ map for Copper (log scale)

VI. CONCLUSIONS

This successful incubator project shows that the NSTG can be a valuable resource for neutron science instruments. In this project ab-initio MD and the $S(q,\omega)$ computations has been computed, both on the TeraGrid and at the ILL, mainly of powders. The TeraGrid NSTG computer has been a good resource for this work. The results have produced structure...
dynamic factors for common powders found in neutron scattering. These were simulated using VASP, then nMoldyn and McStas, and finally compared (whenever possible) to measurements. The 'computation side' is purely ab-initio i.e. there is no adjustable parameter, and the agreement with measurements is astonishing, especially for Al, In and Cu. In terms of 'large' computation, the MD trajectories have been obtained on about 30 ps, which is usually enough for a sub-meV energy resolution. McStas also enables us to input Monte Carlo N-Particle (MCNP) moderator simulation results to model the neutron source, still with a much lower statistical accuracy. This justifies even more large computing power, for an always more complex and detailed description of real neutron scattering beam lines and experiments.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under the following NSF programs: Partnerships for Advanced Computational Infrastructure, Distributed Terascale Facility (DTF) and Terascale Extensions: Enhancements to the Extensible Terascale Facility under current TeraGrid project allocation, TG-DMR060000T.

Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UT-Battelle, LLC for the U.S. Department of Energy under contract number DE-AC05-00OR22725.

Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting this article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

McStas is currently supported as a JRA of the NMI3 network by the European Commission under the 6th Framework Programme through the Key Action: Strengthening the European Research Area, Research Infrastructure Contract n: RII3-CT-2003-505925

REFERENCES

12) http://wolf.ifj.edu.pl/phonon/