Computational Needs for RHIC Theory

RHIC Physics and Computational Nuclear Theory

The study of the properties of hot and dense QCD matter is one of the four main areas of nuclear physics research described in the 2007 NSAC Long Range Plan and reaffirmed in the 2013 Long Range Plan Implementation Report. Data taken at the Relativistic Heavy Ion Collider (RHIC) since the year 2000 has shown conclusively that the high temperature phase of QCD matter is a quark-gluon plasma with the characteristics of a "nearly perfect" liquid. This conclusion has been confirmed by measurements in Pb+Pb collisions at much higher energy at the CERN Large Hadron Collider (LHC). The RHIC and LHC experiments have not only shown that the quark-gluon plasma behaves as a nearly inviscid liquid, but also that it is highly opaque to energetic partons, resulting in strong jet quenching.

The physics goal for the next decade is to characterize the properties of this quark-gluon plasma liquid by quantitative extraction of important medium parameters from precision measurements of sensitive observables, including hadron spectra, angular distributions and correlations, jet observables, and electromagnetic probes. The DOE Performance Measures for High Temperature, High Density Hadronic Matter require: "By 2015, create brief, tiny samples of hot, dense nuclear matter to search for the quark-gluon plasma and characterize its properties." The criterion for the grade "Excellent" demands that "its properties such as temperature history, equation of state, energy and color transport (via jets), and screening (via heavy quarkonium production) are characterized."[1]

To achieve these goals, detailed comparisons of theoretical calculations with a variety of experimental observables are necessary. These theoretical calculations all require significant computational resources and are absolutely essential for the success of the overall RHIC program. To perform meaningful comparisons, for many observables event-by-event calculations are required [1]. For example, the azimuthal anisotropy in the produced particle distributions is highly sensitive to event-by-event fluctuations. Furthermore, the details of fluctuations need to be under control when studying observables sensitive to the details of the phase transformation and the presence of a critical point.

Direct comparison of theoretical calculations with experimental data thus requires the implementation of various important sources of fluctuations. These include initial state geometry fluctuations, hydrodynamic fluctuations, and fluctuations in the freeze-out mechanism. Naturally, this demands significant computational resources, because the various stages of the calculation need to be evaluated thousands of times.

Experimental data sets for a single beam energy and projectile combination tend to surpass several petabytes. The modeling community must not only reproduce results extracted from these data sets for numerous classes of observables covering numerous collisions, but must also investigate a high-dimension (of order two dozen) parameter space. In the last few years modelers have taken on the effect of fluctuations of the initial state, which requires analyzing hundreds of initial state configurations for a single impact parameters. Additionally, as the field begins to analyze and interpret data from the beam energy scan, modelers must forego the two-dimensional descriptions applied at the highest energies and consider approaches that model the dynamical

[1] *Report to NSAC of the Subcommittee on Performance Measures*, August 2008.

evolution in three spatial dimensions, without the simplification of boost-invariance along the beam direction. Combined with the increased data number and size of the experimental data sets resulting from measurements at many different beam energies and from an increased number of collision systems, the numerical demands facing the modeling community will grow by 3 to 4 orders of magnitude. This will require additional resources, for running at first-tier facilities and for coping with the additional demands associated with the increased sophistication of the modeling. Tailoring the numerical codes for running on specialized machines will additionally require access to small local CPU or GPU clusters for testing purposes.

Generally, for typical projects addressing the main physics questions that we seek to answer, the following aspects are characteristic of the computing needs:

- A large majority of computational problems contain Monte-Carlo components that can be trivially parallelized on multiple CPU.
- Commodity clusters provide the best and most cost-effective approach to fulfill a majority of these computing needs.
- Depending on the scale of the required calculations, these resources can either be sourced locally or via national grid computing resources.
- most of the community's needs for large scale computing resources are readily available via NERSC allocations or the Open Science Grid (OSG). The OSG provides up to 100K CPU-hours per day availability for a single group (OSG is jointly funded by DOE and NSF); access to the OSG is readily and informally available to all DOE & NSF funded research groups, with near immediate access.
- For projects requiring highly parallelized large scale computing, allocation of resources on NERSC are the best option.
- Small scale computing resources are best locally sited and only require a modest investment/upkeep of about $10K per year per group.
- Currently the largest challenge is not CPU, but data storage capability. If data storage is initially buffered at local (university) facilities, then the required bandwidth for data transfer to national facilities becomes a significant challenge. Further effort is required for an optimum tie-in of high performance computing (HPC) storage facilities with grid computing setups such as the OSG.
- None of the above is particularly specific to RHIC physics; the usage pattern described above can be regarded as typical for many computational nuclear physics groups.

As examples of typical computational needs, we describe in the following two case studies of research projects to be conducted at Brookhaven National Lab and at Duke University.

Case Study I: Brookhaven Theory Group

The main computational efforts within the Brookhaven Theory Group are the study of the non-equilibrium early-time dynamics in heavy-ion collisions, and the event-by-event simulation of their complete space-time evolution, including sophisticated initial state models [2], 3+1 dimensional viscous fluid dynamics [3], and microscopic hadronic cascades. The requirements for these projects are detailed below:

- 2 senior scientists, 3 postdocs, 1 graduate student
- research focus: ab-initio calculation of early-time non-equilibrium heavy-ion collision evolution, event-by-event 3+1 dimensional viscous fluid dynamics, detailed description of initial states including JIMWLK evolution
- For the study of some observables on the order of 2000 events are necessary per centrality class, e.g. v_n fluctuations or other fluctuation studies. Studying 5 centrality classes this makes 10000 events, for a single set of initial conditions and parameters. Pure 2+1D hydro calculations would require 25,000 core hours for one set of parameters, 3+1D hydro would need $\sim 160,000$ core hours. Using the hadronic afterburner would add $\sim 150,000$ core hours. Studying various collision systems, energies, different temperature dependent η/s, different equations of state and freeze out temperatures, leads to an estimate of at least 20 different configurations. This leads to 0.5 million core hours for the 2+1D simulations. Studying 20 variations in 3+1D hydro needs 3.2 million core hours, and including the hadronic afterburner in some studies (10) adds 1.5 million core hours.

  We thus estimate a minimal need of 5.2 million core hours for the study of various collision systems over a range of energies using event-by-event simulations. Required data storage for this project is on the order of 10-20 TB.
- A major uncertainty in the calculations discussed in the previous point stems from the description of the very early stages immediately after the collision. Initially the produced matter in a heavy-ion collision is in a state far from thermal equilibrium. Important theoretical questions concern the details of the thermalization mechanisms and the onset of hydrodynamic behavior as well as when and to what extent the formation of a thermalized Quark Gluon Plasma is achieved [4]. On a qualitative level, important insight on the thermalization process can already be gained from 3+1 dimensional simulations for the SU(2) gauge group and without quarks [5, 6]. In this setup, the computational cost of simulating the dynamics of a single field configuration is typically between 1000 - 10000 core hours depending on the lattice size and the number of observables. With realistic initial conditions, an average 50-100 independent field configurations are needed to study the thermalization process in a single heavy-ion collision. Including a systematic study of the parameter and discretization dependence one needs to perform simulation for at least 4-8 different parameter sets and 4-8 different lattice discretizations for a fixed parameter set. This leads to a total estimate of at least 400k core hours.

  The computational cost to perform the same analysis for the SU(3) gauge group is about a factor of 10 higher, due to the more complicated color algebra. However, because of the larger color algebra one may also expect a factor of 1/2 reduction in the statistics needed to compute observables in a single event. We thus estimate the over-all cost to perform a more quantitative analysis for the SU(3) gauge group to be at least 2 million core hours based on the above estimates for the SU(2) case.

- Including dynamical quarks into non-equilibrium simulations of the early stages of high-energy collisions will provide an important step towards a complete theoretical description and is expected to yield qualitatively new insights into the real-time dynamics of quark production and the chemical equilibration of the Quark Gluon Plasma. By use of the stochastic procedure of dynamical "low-cost" fermions [7], the computational cost to evolve the fermion fields can be significantly reduced and exhibits the same scaling in the lattice size (N^3) as the gauge field evolution. However, present simulations e.g. in 1+1 dimensional QED need about a factor of 100 larger sample sizes than in the absence of dynamical fermions [8], making QCD simulations feasible only on somewhat smaller lattices, e.g. covering only a sub-volume of the plasma. Taking this into account, we estimate that a series of qualitative studies of quark production and chemical equilibration can be performed with a dedicated computational budget of about 2 million core hours. The study of more complicated observables such as e.g. quark flavor correlations would further increase the computational requirements by at least a factor of two.

- We estimate the computational requirements for non-equilibrium simulations of the early stages of high-energy heavy-ion collisions to be at least 4.4 million core hours on a time scale of about two years.

Case Study II: QCD Group at Duke University

- 2 senior scientists, 2 postdocs, 3 graduate students

- Computational focus is on modeling of relativistic heavy-ion collisions and using those models for knowledge extraction via a comprehensive model to data comparison; this type of comparison requires the mapping of the model parameter space with sample sizes per parameter set that allow for a statistically meaningful comparison to the recorded RHIC & LHC data (i.e. similar level of uncertainty on the data and on the model side).

- Running one particular model design over its full parameter space: 2.5 hours per 2+1D event (including viscous relativistic fluid dynamic ("vRFD") and microscopic cascade ("micro") stages), 1000 events per parameter set, 6 parameters varied simultaneously, 10 values per parameter, 10 centrality bins 1,500,000 CPU hours on the OSG (can be completed in approx. 2 weeks).

- To extract science, multiple model designs need to be tested and run; annual CPU requirements approx. 10-20 million CPU hours. Making the transition from 2+1D vRFD to 3+1D vRFD (necessary for the RHIC beam energy scan program) will increase the runtime per event by approx. a factor of 10.

- Required data storage 10+ TB per month; currently CPU is being traded against storage, i.e. data are discarded after initial analysis, to be regenerated later if again required.

- Local computational resources needed to run the above computational setup: approx. 50 nodes and a 100 TB storage array. Annual upkeep/renewal cost about $10K.

There are a significant number of U.S. nuclear theory groups engaged in RHIC physics. All of them rely on computational techniques, and their needs range from modest desktop computing to significant resource computing or even large allocations at leadership class computing facilities. The most prominent RHIC theory groups are: Brookhaven, Columbia, Duke, ECU, ISU, Kent State, LBNL, Minnesota, NCSU, MIT, MSU, OSU, Purdue, Stonybrook, TAMU, UIC and Wayne State. In particular the groups at OSU and MSU are presently involved in research activities that mirror the computational demands listed in the above case studies. One can easily anticipate that the recent development of jet shower Monte-Carlo codes for the study of jet quenching and other medium effects on hard and/or heavy partons created in relativistic heavy-ion collisions will generate similar-sized computational needs by several other groups, including LBNL, Wayne State, and TAMU.

Suggested Resource Recommendations and Policies:

Large investments by the US and international science communities have created a relativistic heavy-ion collision program that is producing data of unprecedented precision and information content. Computational nuclear theory is essential for an adequate exploitation of these data and the full realization of their potential to provide a quantitative understanding of the hottest and densest forms of matter ever created by mankind. This requires significant computational resources at all scales: leadership, capacity and desktop computing. The following recommendations aim to ensure that the community's computational needs are met:

- Continued support of the Open Science Grid in addition to NERSC for capacity computing, with provisions to expand these resources as the demand rises.
- Development and deployment of distributed and/or highly networked data storage capabilities that interface with the OSG and other distributed computing resources (local/campus grids); this could go hand in hand with or be facilitated by an improved tie-in of large scale storage of national HPC resources with grid computing setups.
- A change of policy that allows (contrary to present practice) university groups to re-allocate, in consultation with their program officer, modest funds ($10K-20K annually) from their base grants towards small-scale local computing resources.

Note that these recommendations are complementary to the leadership class needs that are essential for lattice QCD type calculations. They impact a community of equal size and importance to the RHIC program. Only the combined efforts of both the lattice QCD and RHIC theory communities will ensure the overall success of the heavy-ion programs at RHIC and LHC and of the full return on capital for the multi-billion dollar investments made by DOE and NSF in these scientific endeavors.

References

[1] Gale C, Jeon S and Schenke B 2013 *Int.J.Mod.Phys.* **A28** 1340011

[2] Schenke B, Tribedy P and Venugopalan R 2012 *Phys. Rev. Lett.* **108** 252301

[3] Gale C, Jeon S, Schenke B, Tribedy P and Venugopalan R 2013 *Phys.Rev.Lett.* **110** 012302

[4] Berges J, Blaizot J P and Gelis F 2012 *J.Phys.* **G39** 085115 (*Preprint* 1203.2042)

[5] Berges J, Boguslavski K, Schlichting S and Venugopalan R 2014 *Phys.Rev.* **D89** 074011 (*Preprint* 1303.5650)

[6] Epelbaum T and Gelis F 2013 *Phys.Rev.Lett.* **111** 232301 (*Preprint* 1307.2214)

[7] Borsanyi S and Hindmarsh M 2009 *Phys.Rev.* **D79** 065010 (*Preprint* 0809.4711)

[8] Hebenstreit F, Berges J and Gelfand D 2013 *Phys.Rev.* **D87** 105006 (*Preprint* 1302.5537)