Computational Challenges for Galaxy Formation

Thomas Quinn
University of Washington, Astronomy

January 18, 2015

1 Computational Resource Challenges

Galaxies are made of stars. This obvious statement points out the great resolution challenges of understanding galaxy formation: star formation (itself a computational challenge) occurs on sub-parsec scales while the environment for galaxy formation varies on scales greater than 100 megaparsecs. Hence, a straightforward estimate gives dynamic range requirements of 10^8 in length and 10^{24} in mass. This number should also be multiplied by the number of timesteps needed to follow galaxy evolution for the age of the Universe: roughly 10^7.

Significant progress has been made by separating out the physics on different scales, and the introduction of subgrid models. Galaxy morphologies have been successfully modeled with the use of subgrid star-formation, and the clustering of galaxies on gigaparsec scales has been modeled using semi-analytic (phenomenological) models for the star formation history of individual galaxies. Nevertheless, these methods have limitations because of the couplings between the scales. Stellar feedback influences the dark matter distribution within galaxies; stellar photons contribute to the reionization of the intergalactic medium, and the large scale environment determines the gas supply that galaxies use to form stars. Hence, simulations that couple these various scales are required.

Current simulations using the capability of petascale systems like Bluewaters reach from 100 parsecs to 25 megaparsecs, but to properly model current future large galaxy surveys or to draw conclusions about the nature of dark matter from galaxy morphologies, several orders of magnitude more computational capability will be required.

Accomplishing the science goals of computational galaxy formation will continue to require the highest available computational capability.

2 Algorithmic Challenges

Gravity is one of the more computationally difficult aspects of galaxy formation, particularly since the solution of Poisson's equation at any point formally depends on the mass density at every other point. This issue becomes exacerbated as computing becomes more distributed. The FFT, a traditionally efficient method for solving Poisson's equation, becomes increasingly difficult to scale well has node and core counts increase (e.g [1]). Hence development and testing of other algorithms becomes necessary for computational galaxy formation to effectively use future capability computing.

As the physical modeling in simulations gets more detailed, radiative transfer will also be required. This is also an "all-to-all" calculation, and algorithms for efficiently and accurately performing this calculation are being actively developed. Scaling these algorithms to large distributed systems are likely to be more challenging than scaling gravity algorithms.

3 Software Development Challenges

The previous two sections lead to the software development challenge. For computational cosmology, we need to both develop sophisticated algorithms, and scale them to the largest machines available. Furthermore, we want this software to be portable to new architectures as they come along. Clearly we want to abstract away the details of the machine as much as possible just as we do for traditional programming. Unfortunately the currently most portable parallel programming model, MPI, in the first instance, forces the programmer to think of the machine in a very detailed manner. Furthermore, hybrid architectures (multicore, SIMD, GPU) usually require some hybrid of programming paradigms with MPI. High level languages, (e.g. Chapel, CAF) on the other hand, work well if the abstractions they provide match the operations that the computation requires (e.g., array operations.) However, complex algorithms (adaptive, divide-and-conquer, etc.) may very difficult to express in these languages. A middle possibility is for the programmer to do the division of the algorithm into parallel components, and have a runtime system map those components on to the processors (Charm++, ParallelX, SWIFT). This strategy as so far been successful in getting good scaling to .5 million cores [4], and promises to go further. Nevertheless, it is not clear that this strategy will work for the more fine grain parallelism for GPUs, although there has been some preliminary success [2].

Even if computational cosmologists, with heroic effort, succeed in getting getting their codes to scale to the largest available resources, the software development challenge is not over. First, the hallmark of science is reproducibility. Given the complexity of the dynamics, results of the simulations should not be trusted until they are verified by independent codes. Code comparison projects [3] are necessary to make real scientific progress, and this means that more than one code needs to be able to run at scale. Secondly, the simulations produce large datasets which will need parallel tools to analyze. This obviously involves development of the analysis code that also needs to run at scale.

In summary, **the computational needs of galaxy formation also requires significant software development effort of algorithms that can run at scale. This effort can only reasonably go forward with the help of parallel languages and runtime systems to help abstract away the complexities of future capability machines.**

References

[1] T. Ishiyama, K. Nitadori, and J. Makino. 4.45 Pflops Astrophysical N-Body Simulation on K computer – The Gravitational Trillion-Body Problem. *ArXiv e-prints*, November 2012.

[2] Pritish Jetley, Lukasz Wesolowski, Filippo Gioachin, Laxmikant V. Kalé, and Thomas R. Quinn. Scaling hierarchical n-body simulations on gpu clusters. In *Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '10, Washington, DC, USA, 2010. IEEE Computer Society.

[3] J.-h. Kim, T. Abel, O. Agertz, G. L. Bryan, D. Ceverino, C. Christensen, C. Conroy, A. Dekel, N. Y. Gnedin, N. J. Goldbaum, J. Guedes, O. Hahn, A. Hobbs, P. F. Hopkins, C. B. Hummels, F. Iannuzzi, D. Keres, A. Klypin, A. V. Kravtsov, M. R. Krumholz, M. Kuhlen, S. N. Leitner, P. Madau, L. Mayer, C. E. Moody, K. Nagamine, M. L. Norman, J. Onorbe, B. W. O'Shea, A. Pillepich, J. R. Primack, T. Quinn, J. I. Read, B. E. Robertson, M. Rocha, D. H. Rudd, S. Shen, B. D. Smith, A. S. Szalay, R. Teyssier, R. Thompson, K. Todoroki, M. J. Turk, J. W. Wadsley, J. H. Wise, A. Zolotov, and f. t. AGORA Collaboration29. The AGORA High-resolution Galaxy Simulations Comparison Project. *Astroph. J. Supp.*, 210:14, January 2014.

[4] H. Menon, L. Wesolowski, G. Zheng, P. Jetley, L. Kale, T. Quinn, and F. Governato. Adaptive Techniques for Clustered N-Body Cosmological Simulations. *ArXiv e-prints*, September 2014.