List of questions rose during the phone meeting on 06-20-2006

In *italics* are those questions that refer to TeraGrid.

**Authors:** S. Dong, I. Pivkin, and P.D. Richardson  
**Project URL:** [http://www.nibib.nih.gov/Research/MultiScaleModeling/Karniadakis](http://www.nibib.nih.gov/Research/MultiScaleModeling/Karniadakis)  
**Suggested questions and activities:** We suggest that questions that will bring the group to working together will involve sub-components of projects that have some degree of commonality and overlap. So, after looking over the projects, we suggest:

Q.1. How are we representing, for use at different scales, transport events and processes - applicable to various molecules such as lipids and proteins, mobile cells - and membrane-transmitted signals? Perhaps of interest e.g., to Lin (mucus formation, transport, activation of the flagellated mucus-transporting cells); Luebeck (mutation and migration); McCulloch (myocyte activation and transport of nutrients, metabolites, Purkinje fibers activation); Kunz (inhaled droplets, particulates, mucus formation and transport in airways, cough); Karniadakis (platelets, red cells, leukocytes, cell adhesion molecules including von Willebrand factor); Brasseur (villi and their motions, peristaltic and other motions of larger scale, selective transport at cell scale) ..

Q.2. How are we representing, for use at different scales, smooth-muscle-driven events? Without adding the process details, this could involve Brasseur, Bassingthwaighte, Lin, Kunz, Karniadakis, McCulloch, others.

Q.3. Most projects seem to involve either sequential computations, or parallel computations that may not be efficiently parallelized (involving a relatively small number of processors); collaborations on the parallel computing aspects could help many projects in the groups.

**Author:** Dan Haworth  
**Project URL:** [http://www.nibib.nih.gov/Research/MultiScaleModeling/Kunz](http://www.nibib.nih.gov/Research/MultiScaleModeling/Kunz)  
**Suggested Questions and Activities:** Things that would be of interest to me:

1) Learn more about TeraGrid computing (seconding what several others said).

2) Learn more about what NSF's next big high-performance computing initiative might look like, if such information is available.

3) Learn more about scripting tools/languages that facilitate code portability and interoperability (e.g., Python, others?)

**Author:** Bill Hazelton  
**Project URL:** [http://www.nibib.nih.gov/Research/MultiScaleModeling/Luebeck](http://www.nibib.nih.gov/Research/MultiScaleModeling/Luebeck)  
**Suggested Questions and Activities:** I would like to have regular conversations to hear how the different groups are progressing. I like your suggestion about learning more about the TeraGrid environment. Specific questions I have are: 1) Is a High Performance Fortran (HPF) or C compiler available on that platform? Our HPF license allows up to 64 processors, so would not make full use of what may be available. 2) What are the requirements/costs for making use of the TeraGrid system? 3) Is there a long queuing, or is processing time readily available?

**Author:** Jim Brasseur  
**Project URL:** [http://www.nibib.nih.gov/Research/MultiScaleModeling/Brasseur](http://www.nibib.nih.gov/Research/MultiScaleModeling/Brasseur)  
**Suggested Questions and Activities:** Two issues/questions that might be addressed within the group specific to nonlinear dynamical system calculations. I focus on nonlinear systems because that is where most of the issues lie.

1. General issues of what issues must be addressed in carrying our nonlinear calculations to scale up a computation to a massive parallel environment. Sub questions:
(a) How does a "TeraGrid" computer differ from what we traditionally called a "massive parallel" computer? Are these synonymous or is there some fundamental difference.

(b) How many practical types of cluster architectures and types should we seriously consider? Is it enough to consider only "Beowulf" clusters, which I define as Linux clusters built on open source operating systems with high-speed Ethernet interconnects in a hub-like arrangement?

(c) For a given cluster architecture (e.g., Beowulf as defined above) is there a fundamental "scale" or number of nodes/processors that makes the use of the cluster transition from "standard" to "massive" or "TeraGrid"? If so, what is the "scale" and how does it depend on the application itself (e.g., salability). Also if so, how does the use of the cluster (programming, suitability vs. non suitability of specific algorithms and numerical methods and dynamical systems, locality of information exchange, etc.) fundamentally change when once transitions from the "standard" scale to the "massive" or "tera" scale?

2. In context with issue 1, what are is the practical evolution in the coding of a dynamical system that one goes through to transition from "regular" parallelization to "massive" or "tera" parallelization?

Subquestions:

(a) What are the pros and cons of using High Performance Fortran, vs. MPI or OpenMPI?

(b) What diagnostics should one collect in the evolution process above to create optimal code for "regular" and "Massive" parallel computing?

(c) What dynamical systems can vs. cannot take advantage of "TeraGrid" computers? For example, Fourier spectral methods spend a large percentage of time in the FFT, which is notoriously non-local. Does this preclude these methods from being applied to "tera" or "massive" parallel environments?

Author: Michela Taufer
Project URL: http://www.nibib.nih.gov/Research/MultiScaleModeling/Taufer

Suggested Questions and Activates:

A.: What challenges are our projects facing in terms of computation, algorithms, and high performance computing? Can we define general criteria to analyze these challenges? Can we define general methods to address these challenges? Here with the term general, I mean criteria and methods that go behind single simulations or applications.

B.: One aspect that rose from some e-mails in the WG 5 is the urgency to move our simulations (that have been running so far on supercomputers) toward grid computing, i.e., TeraGrid. While the step toward cluster technology has been successful and relatively easy, this new steps towards grid computing seems more challenging. Can we overcome some barriers through a more in-depth knowledge of this technology? Can we define criteria to better understand whether our applications are indeed suitable and to what extent they are suitable? Can we measure a priory the amount of work is needed to port to and maintain on grid computing our simulations? Are there other issues such as loss in control on our simulations, security, and reliability of the results that bother us? With reference to this point, George Karniadakis published two papers on the challenges of moving applications to grid computing. I attach the papers of George to this e-mail.

C.: Computing environments are becoming larger and larger. Assuring their homogeneity in terms of e.g., computing resources and operating systems is becoming a real problem. Even in dedicated environments such as clusters, the upgrade of a compiler or libraries can have dramatic impacts on your simulations. Heterogeneity can result in divergence in MD and MC simulations and a consequent loss of trust in your simulation results. Can we build a pool of applications that are affected by these phenomena? Can we identify factors that affect divergences in simulations? Can we quantify the impact of this factor on divergence? Can we define common methodology that reduce or keep under control divergences in heterogeneous environments? Are these methodologies applicable to different applications and heterogeneous environments? With reference to this point, my project is doing some research. Also some papers of David Bailey (Lawrence Berkeley Lab) address aspects that are related.
D.: A key aspect of our projects is that they combine multi-scale models and simulation results at the different scales and these results need to be coupled. Simulations at different scales may be implemented with different codes in different languages. The human intervention may be a critical aspect to select, interpret, and redirect results from one simulation scale to another simulations scale. Do we have frameworks and tools today that facilitate the coupling of different scale simulations into a single simulation environment? Can we reduce the human intervention using adaptive methods, e.g., ML techniques? Do we still need tools to facilitate the monitoring and guide these multi-scale simulations?

Author: Wilfred Li
Project URL: http://www.nibib.nih.gov/Research/MultiScaleModeling/McCulloch
Suggested Activities:
I can help invite someone to talk about TeraGrid if necessary