CyberInfrastructure/ CyberScience: *What does it mean for Materials Research*

Thomas A. Weber
Director, Materials Research

June 19, 2004
Urbana, IL

What's next?

- **Cyberinfrastructure [CI]**
  - **Vague:** ***A shared integrated system of interconnected computation, communication, and other information technology that supports a range of activities in a research community.***
  - **Elements:** ***Advanced computing hardware, networks, software, data storage, data management, digital libraries ...***
  - **What does CI mean**
    - **for the computational materials research community?**
    - **for the broader materials research community?**

What's next?

- **Cyberinfrastructure [CI] from the ACP**

  Historically, infrastructure was viewed largely as raw resources like compute cycles or communication bandwidth. As illustrated by many activities in the current PACI centers and by the recent NSF middleware program, the scope of infrastructure is expanding dramatically beyond this narrow definition. ***For purposes of the ACP, infrastructure will comprise of a diverse set of technologies, facilities, and services and intangibles like design processes and best practices and shared knowledge. A major technological component is software that participates directly in applications and software tools that aid in the development and management of applications. A critical non-technological element is people and organizations that develop and maintain software, operate equipment and software as it is used, and directly assist end-users in the development and use of applications.***

  The ACP seeks to bring about dramatic and beneficial change in the conduct of science and engineering research. Applications will greatly expand their role and become increasingly integral to the conduct of science and engineering research.

What's next?

- **MPS view: Science drives Cyberinfrastructure**
  - **Workshop at NSF in April – MPS wide: Common themes? Unique needs?**
  - **Our concern here:**
- **Our concern here:**

  Computational Materials Research Community (Condensed Matter Physics, Materials Science, Solid State Chemistry, & Polymers)
  - **What is the science that the computational materials research community aspires to do? ("Cyberscience")**
  - ***What are the cyberinfrastructure needs of the computational materials research community to do the science?***

What are the priorities ?

CYBERSCIENCE WORKSHOP

Identifying Major Scientific Problems in the Mathematical and Physical Sciences and Their CyberInfrastructure Needs

WEDNESDAY, APRIL 21, 2004
8:00 AM – ROOM 110
National Science Foundation

INVITED SPEAKERS

Dr. Brent Fultz, California Institute of Technology

Dr. David Keyes, Columbia University

Dr. Vijay Pande, Stanford University

Dr. Dan Reed, University of North Carolina

Dr. Larry Smarr, University of California, San Diego

Dr. Alex Szalay, Johns Hopkins University

AFTERNOON BREAKOUT SESSIONS

I. Algorithms and Software; II. Software Infrastructure; III. Hardware, Facilities;
IV. Network Infrastructure; V. Data Management and Analysis

Information
Morris L. Aizenman, Senior Science Associate, MPS Directorate, 703-292-8807

Credit © Eric J. Heller, Harvard University

Cyberscience Workshop

To identify needs for cyberscience, ***defined as the science that cannot be done without the advanced capability of cyberinfrastructure.***

- representatives from all divisions in MPS
- Dan Reed, NCS
- Larry Smarr, UCSD
- Alex Szalay, JHU
- Brent Fultz, Caltech
- Vijay Pande, Stanford
- David Keyes, Columbia

Breakout sessions

- Algorithms and Software
- Software Infrastructure
- Hardware and Facilities
- Network Infrastructure
- Data Management and Infrastructure

Dan Reed

- “the purpose of computing is insight, not numbers” Hamming
- “the purpose of cyberinfrastructure is science, not geek toys” Reed
- Computing for science vs computing as science

Larry Smarr

- Cosmic scale applications for cyberinfrastructure (science applications)
- More powerful supercomputers and software is the key
- Learn from past experiences
- *Formation of the first galaxies--new instrumentation
- Run faster -2 black hole collisions, just need more computing tine as corrections, data increases

Alex Szalay

- Discoveries are made at the edges and boundaries of science
- Utility of computer networks grow as the number of possible connections
- Internet and grid tools are converging
- Virtual observatory allows to look at astronomical questions in real time -software is the link that needs to be worked on.
- Optimization of searching is needed (software and algorithms)
- Key, looking for one small thing in a haystack. How do you find it?
- Data exploration has no owner.

David Keyes

- Can simulation produce more than "insight"?
- "The computer literally is providing a new window through which we can observe the natural world in exquisite detail." J. S. Langer
- Orbach says ITER design of plasma reactor would be capable of achieving fusion based on a simulation

Brent Fultz (neutron scattering)

- Need to build an interesting software systems
  - Reductions of the data
  - Direct comparison to simulations of detector in real time-- “smart experiments”
  - Direct comparison to physical fit in real time
  - Direct visualization of vibrations or structures in nearly real time (viz the lipid!)
  - Data archiving and metadata

Vijay Pande

- Protein folding as self-assembly dynamics
- Coupling theory/simulation/experimental
- "if you cannot predict what I can measure, then why should I believe you"
- *ie:* must provide insight and not just reproduce experiments
- Timescales of molecular motions are FAST (fsec) so long timescales takes a long time
- Uses distributed computing on public machines!

Main derived topics (am)

Data mining of large data sets--software for searching and optimization

Data archiving-who is responsible, who has access, who pays?

Peak performance vs complexity of simulations and calculations

Computation for science vs computation as science

Visualization of data (viz wall at LANL)

Smart experiments-analyze and learn as you go, adapt

Large-scale simulation, higher resolution, more DOF, more parameters

Science drivers (mini review)

- Basic predictions using models and simulations (also of experiments)
- Simulations of events that are not practical (supernovae, nonlinear fluid dynamics no nuclear testing, etc)
- Huge cosmological problems, the details of particles
- Predictions of biological and chemical assembly and processes
- Predictions of new energy sources (magnetic fusion energy) CFD
- Biochemical physical questions-genomics, networks,motors to cilia, hydrodynamics, rheology
- Micro to macro…materials, complexity of scale and project management

Closing session (summary)

- Hi-end ($$, supercomputers, centers) vs low-end (small clusters, low$) investments
- Mid range >$2M funding is missing “some opportunities are missed”
- One size does not fit all
- International and Interagency approach
- “Darwinian” selection
- Natural link to educational activities
- Reliable,robust, maintianable integrity vs
  dynamic, evolving, adaptive

Recommendations

- Development of tools for cyberscience
  - [SBIR-type process: Phase I–>Phase II]
  - Support science research
  - Sits on the cyberinfrastructure (CISE)
- Reallocation a portion of the budget to support cyberscience award

  -supplements to proposals with cyberscience tool components

  -CFP verbiage and program officer expectations
- MPS add cyberscience component to web page
  - Communication issues
  - visibility
- Coordination of cyberscience and cyberinfrastructure must be addressed “up front”

Other things to think about

- Long term support for people in infrastructure is not MPS role

Where do we want to go?

- **CyberScience must drive CyberInfrastructure**
  - **What science *will* we engage at our frontiers?**
    - **5 years? 10 years?**
  - **What CI *will* we need to make advances?**
    - **5 years? 10 years?**
- **What is high priority? What is lower priority but still important?**

For more information ...

- **Atkins report**

  http://www.cise.nsf.gov/sci/reports/toc.cfm