The Crimson Grid Initiative @ Harvard

Transforming Research Computing: A Strategic Roadmap for DEAS IT (*and perhaps, Harvard as a whole*)

A project to implement a novel technology & organization model for interdisciplinary & collaborative computing to support science & engineering research

Project Motivation

The legacy of operating technology-silos or ‘stovepipes’ no longer consistent with current and future directions in Harvard Scholarship: interdisciplinary collaboration

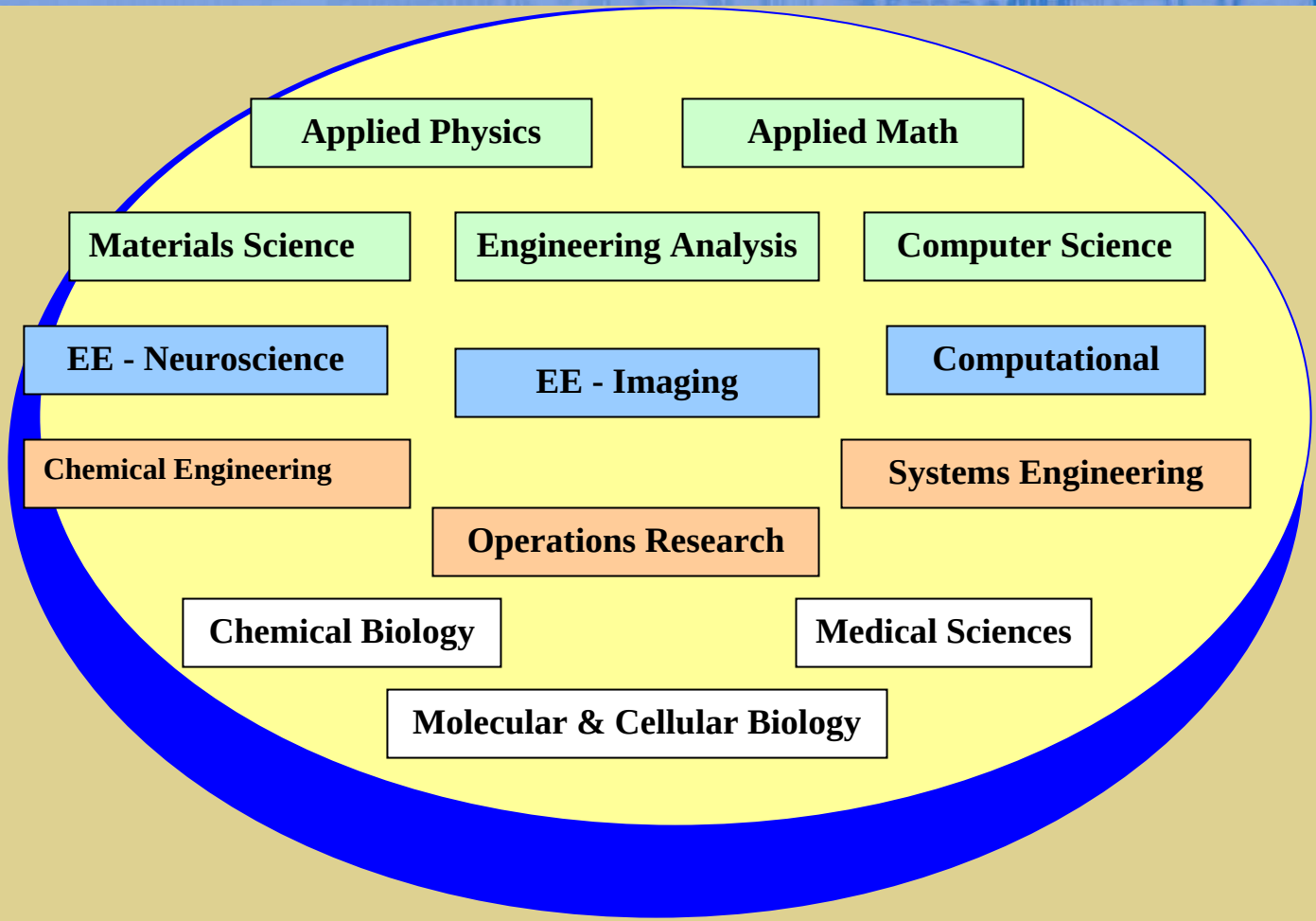
Why Reorganize?

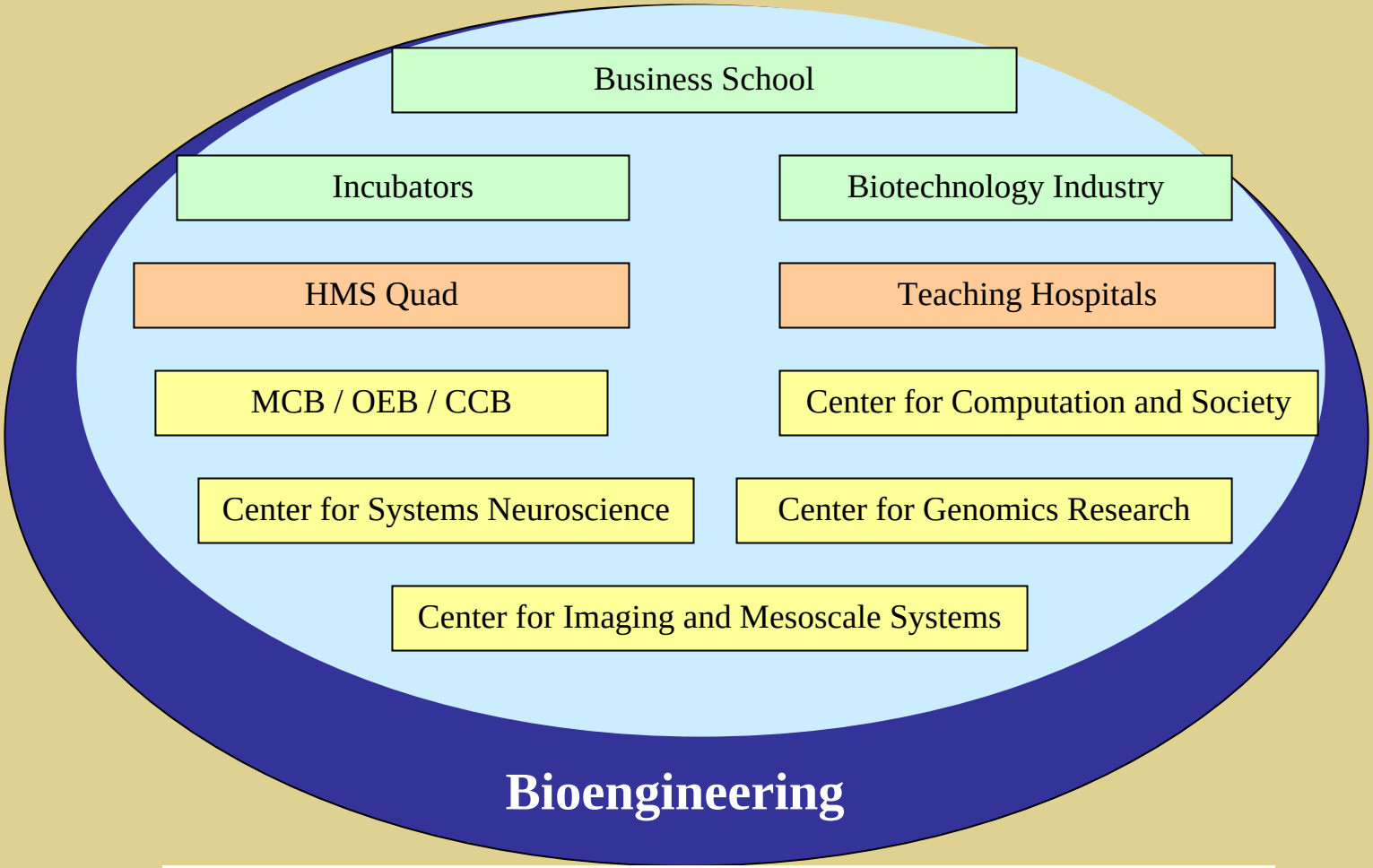
- Existing organization structures and roles inadequate to meet expectations of the science & engineering research community
- Many universities struggle with balancing the inadequacies between central-decentralized models- most tailor a mixed approach; needs a new paradigm
- Investments in IT by labs/units using current models or experimental ones tried elsewhere cannot assure payoffs to research computing at Harvard

Why Is a New Model Important for Harvard ?

- Cyber Infrastructure development is a key priority for national and international sponsors of research- e.g., NSF, LHC-Cern, DOE, NIH: ***Future of Harvard's research vision is now intimately connected to Harvard's cyber infrastructure!***
- Interdisciplinary Faculty collaborations are a high priority, and ***IT support MUST align itself to meet the new needs*** - e.g., DEAS
- Research Computing cannot be separated from support for "personal productivity" tools and services- ***need for a convergence of scientific, instructional, administrative and network systems***

NO CURRENT ORGANIZATION MODEL CAN PROVIDE IT ALL


Applied Physics
Applied Math
Materials Science
Engineering Analysis
Computer Science
EE - Neuroscience
EE - Imaging
Computational
Chemical Engineering
Systems Engineering
Operations Research
Chemical Biology
Medical Sciences
Molecular & Cellular Biology



DEAS is a connecting hub with FAS Science initiatives, HST, HMS, the School of Public Health, and the school of Dental Medicine. It also links with HBS, biotechnology industry, and startups.

Why Grid ?

Convergence of switched arch & web Services creates new opportunities

- ✓ Better exploit /manage workload across resources- higher speed, lower costs
- ✓ Efficiencies in collaboration
- ✓ Secure federated data access in a large distributed environment

The Grid is a fabric that:

- coordinates resources not subject to centralized control
- uses standard, open, general-purpose protocols and interfaces
- Delivers non-trivial qualities of service

Crimson Grid Overarching Goals

- Test & evaluate grid enabling technologies
  - In the context of interdisciplinary, collaborative research
  - Hardware/Software/Business Models/Delivery&Support models
- Test, integrate and establish a working Grid Middleware solution
  - GRSI Grid Reference System Implementation at DEAS
- Develop & Test Grid application services
  - Science and engineering research in higher education

Crimson Grid Technical Objectives:

- Understand the grid ecology
- Extend and characterize the immediate/long-term benefits to campus
- Evaluate issues that surround the evolution of a ‘campus grid’
- Identify long-term strategic benefits

Assumptions

- Understanding of ‘grid’ not complete
- What is a ‘grid’ ? No universally shared ontology, (yet)
- Grids have a ‘ecology’
- Benefits to Science not yet fully characterized; Catch 22?
- Campus specific strategic issues- no one solution?
- We can perhaps take intelligent peeks into the future

Crimson Grid "Test Bed"

- Grid technology chosen as the vehicle for creating a cost-effective computing and data infrastructure
- Leverages the "Bio" + "Science" community presence in the Boston/Massachusetts area among early adopters
- In addition support the needs of Nano, Info, and Eco technologies

The Crimson Grid Test bed Phases

1. Connectivity

The first phase of the Crimson Grid is connectivity. This is being facilitated by FAS on-campus network and UIS/Northern Crossroads (NOX) Internet2 - Boston Metro Ring access.

2. Reference System

The second phase of the Crimson Grid will be to establish a working grid Reference System Implementation. IBM will be working with Harvard University to develop a GRSI that can be shared with future partner institutions both within and outside of Harvard.

3. Pilot Grid Applications

With the GRSI deployed at DEAS it will be the objective of the Crimson Grid to launch and pilot several grid applications. These applications will emphasize both compute and data grid facilities and represent applications from Life Sciences, Bio-technology, Applied Sciences, Engineering, and High Performance Computing.

4. Grid Expansion

Other grid nodes will be fostered under the umbrella of Crimson Grid initiatives.

Approach:

- Work at the interface of innovation and production
- Build an ecosystem
- Establish role of Faculty as stakeholders
- Build Roles for Industry
- Serve as ‘sandbox’ for campus technology test beds– zero penalty for failure!

The Crimson Grid 'Thingy'

- 'Crimson Grid' is a 'gateway'
- Grid of grids (federated resources at the level of labs, depts, schools)
- GT4 (Globus release); skipped NMI
- Personalized Portals for every grid (Gridsphere-GT4)
- Mix of platform archs - 32 bit, 64 bit, SMP
- Infiniband - MPI
- Distributed data storage sandboxes

Example Use Cases

- MacroInformatics
- NanoInformatics
- Health & BioInformatics
- EcoInformatics

Macro- Collaboration between DEAS, Physics, Chemistry,Applied Mathematics, Applied Mechanics in areas of Material Sciences at Macro Scales

Nano- DEAS, Cornell, Gatech, Howard, NC State, Penn Stae, Stanford, UCSB, Umich, Uminn, UNM, UT-Austin, UWash

H&Bio- Collaboration with Martinos Center on cancer research (DEAS, MGH, MIT-HST)

Eco- Atmospheric Science, Ocean Modelling, Earthquake Engineering

Some local details

- Implemented GT4 WS-GRAM and gridftp capabilities in the GridSphere portal system (MDS4 planned soon, and feasibility investigation of RLS at portal level).

- Developed a standard Portal system template that can be customized and deployed for each research group.

- Refactored the GAMA server / portlet functionality to allow for a single CA, but multiple O/OU strings; this enables:

- The policy work of distributing and managing the credentials for different projects at Harvard to be delegated person-wise to administrators with the projects

- The technical work of managing the real credentials to be centralized to CrimsonGrid services;
  users access via MyProxy or CAS.

- Ported the GAMA services to GT4 to enable project based authentication using GSS (each project is given a "credential manager for this project" certificate, and allowed to generate and sign only certain DN certificates, appropriate for that project).

Some local details

- **Are implementing a centralized job submission "Nexus"**
  - **With priority based on the user's DN, schedule jobs on grid nodes throughout Harvard, with baseline resource scavenging available**
    (all nodes allow for low priority use by non-group users)
  - **At the request of resource owners, can be used at the submission phase to aggregate or segregate resources (combine a few grid nodes into a tighter VO at the request of the owners, or prevent certain users from ever running on a certain grid nodes, etc.)**
  - **Allows the CrimsonGrid staff to implement technically the policies decided on and between resource owners.**
- **Will develop a standard CrimsonGrid customized GT4 install for grid node owners (for now, CrimsonGrid staff do this by hand), unless…**

Example Use Cases

- MacroInformatics
- NanoInformatics
- Health & BioInformatics
- EcoInformatics

Macro- Collaboration between DEAS, Physics, Chemistry,Applied Mathematics, Applied Mechanics in areas of Material Sciences at Macro Scales

Nano- DEAS, Cornell, Gatech, Howard, NC State, Penn Stae, Stanford, UCSB, Umich, Uminn, UNM, UT-Austin, UWash

H&Bio- Collaboration with Martinos Center on cancer research (DEAS, MGH, MIT-HST)

Eco- Atmospheric Science, Ocean Modelling, Earthquake Engineering

Some Focus Use Cases

- **Soft Tissue Modeling for Surgical Simulation**
  Group: Harvard Biorobotics Laboratory
- **Cancer Tumor Modeling**
  Group: Complex Biosystems Modeling Laboratory, MGH
- **Cardiac Tissue Engineering/Excitation-Contraction Coupling**
  Group: Disease Biophysics Group
- **Understanding and Predicting El Nino**
  Group: Climate Dynamics Group.
- **COBRA project data analyses**
  Group: Atmospheric Sciences
- **Deformation transients in major fault zones and relation to earthquake generation.**
  Group: Solid mechanics and Geophysics
- **Project ATLAS**
  Group: High Energy Particle Physics
- **Grid Resource Brokering**
  Group: Systems Research

About NNIN

This section contains:

The members of NNIN include the following major nanotechnology user facilities:

- The Cornell Nanoscale Facility at Cornell University
- The Stanford Nanofabrication Facility at Stanford University
- The Michigan Nanofabrication Facility
- The Microelectronics Research Center at the Georgia Institute of Technology
- The Center for Nanotechnology at the University of Washington
- The Penn State Nanofabrication Facility at the Pennsylvania State University
- Nanotech at the University of California at Santa Barbara
- The Microtechnology Lab at the University of Minnesota
- The Nanoscience at the University of New Mexico
- The Microelectronics Research Center at University of Texas at Austin
- The Center for Imaging and Mesoscale Structures at Harvard University
- The Materials Science Research Center of Excellence at Howard University
- The Triangle National Lithography Center at NCSU (DUV lithography only) (Affiliate)

Current industry partners

IBM

Intel

Microsoft

Network Appliance

SGI

Immediate Benefits

The convergence of switched architectures and web services has created a new opportunity for changing the paradigm of computing to a "grid" where the functional utilities of client-server architectures are maintained yet providing a seamless way for:

- **Sharing resources**
- **Cyber-collaboration**
- **Convergence of a diverse range of tools and services**
- **Large scale data collaboration**
- **New research tools integrated to research workflow**
- **Leveraging core funded and research funded assets and facilities**
- **Distributing and retaining control of administrative policies on systems resources and IT assets including facilities and services where it belongs**
- **Establishing new models for re-engineering campus support organizations**
- **Scaling assets and distributing management to appropriate campus loci**

Some Early Lessons

- A reference system is a moving target; lots of local variations
- Lots of stuff out there, don't waste time/$
- If you need to sell, then don't sell- time sink
- One cannot point to large grid projects to justify campus investments
- Middleware vs. middleware

Campus Grid Issues

- Technology fabric (HEAS)
- TCO (HUIS)
- Business model (HBS)
- Social model (FAS)
- Public Policy impacts (KSG)
- Value to Industry

Next Phase

- Build collaborations with other campus grids
- Work to develop a national dialog on campus grids
- Seek out funding partners
- Work to better understand “edge services”