CSC 7700: Scientific Computing

Introduction

Dr Gabrielle Allen

Logistics

- Web: https://wiki.cct.lsu.edu/sci-comp/
- Class will be held in 338 Johnston Hall (CCT)
  - One exception
- Lectures:
  - Tuesday 12.10-1.30pm
  - Thursday 12.10-1.30pm
- Schedule of classes
  - https://wiki.cct.lsu.edu/sci-comp/Schedule

Office Hours

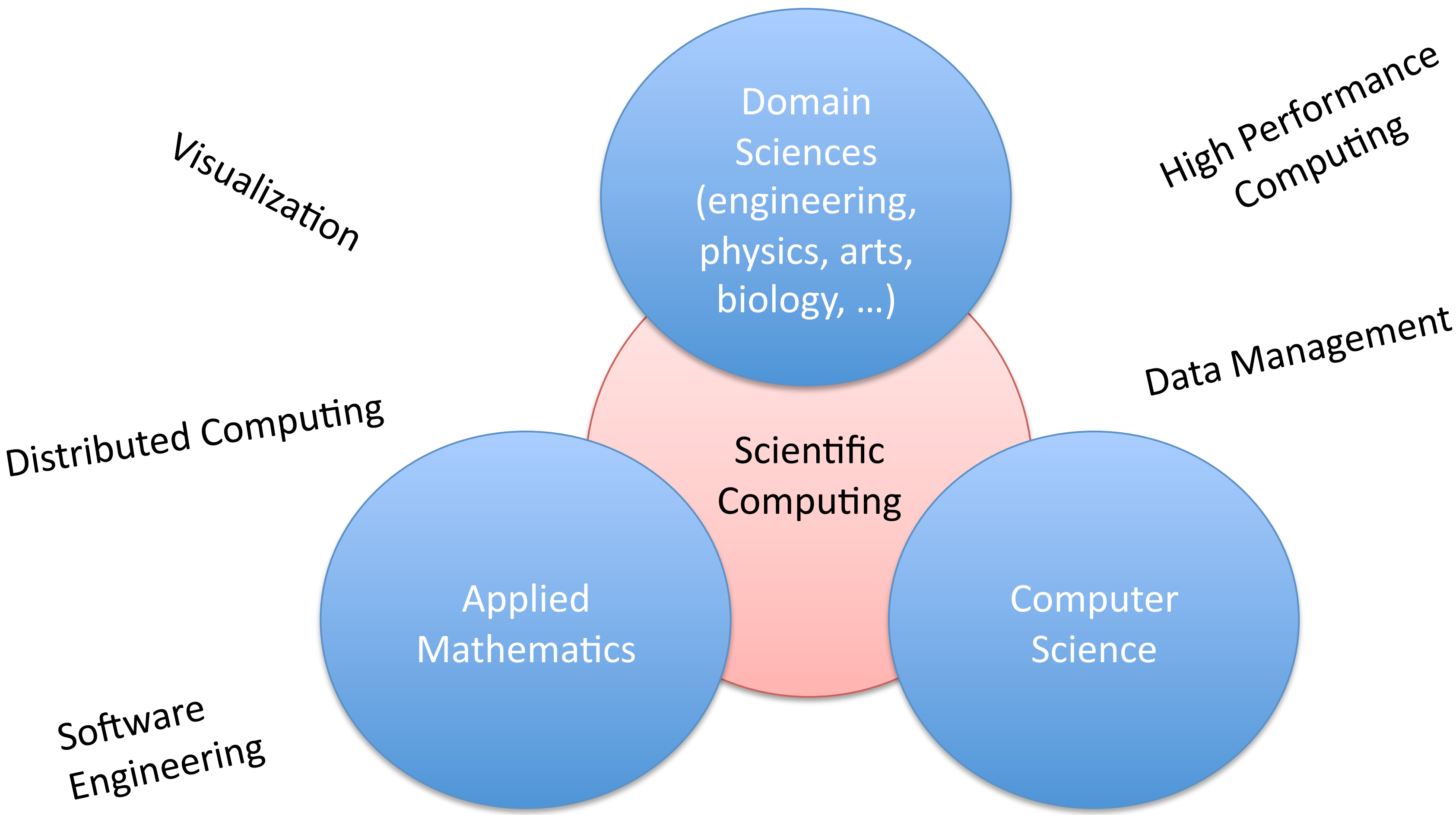
- Dr Allen
  - Wednesday 10am to noon, Johnston Office
  - Also use AIM: gridrebel or email
- Other instructors
  - Posted on web pages

Overview of Course

- Motivation and Scientific Computing
- Methodology and Modules
- Teaching Faculty
- Grading
- Cyberinfrastructure Resources

Motivation and Scientific Computing

Scientific Computing

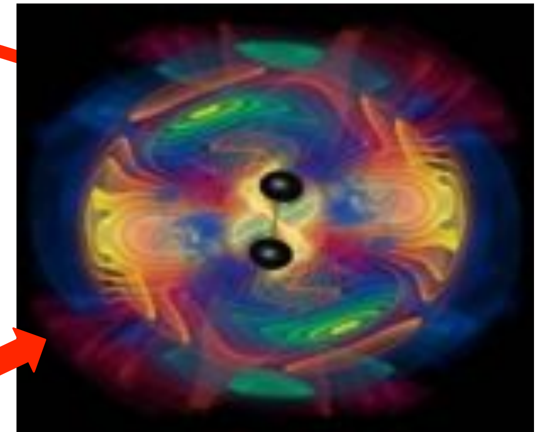


Scientific Computing (from wikipedia)

- Computational science (or scientific computing) is the field of study concerned with constructing mathematical models and quantitative analysis techniques and using computers to analyse and solve scientific problems. In practical use, it is typically the application of computer simulation and other forms of computation to problems in various scientific disciplines.
- The field is distinct from computer science (the study of computation, computers and information processing). It is also different from theory and experiment which are the traditional forms of science and engineering. The scientific computing approach is to gain understanding, mainly through the analysis of mathematical models implemented on computers.
- Scientists and engineers develop computer programs, application software, that model systems being studied and run these programs with various sets of input parameters. Typically, these models require massive amounts of calculations (usually floating-point) and are often executed on supercomputers or distributed computing platforms.

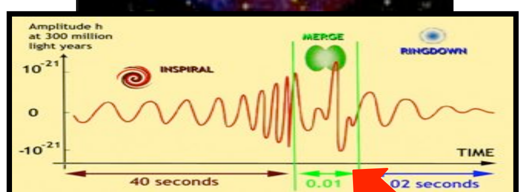
Gravitational Wave Physics

Models & Simulation



Observations

Scientific Discovery!



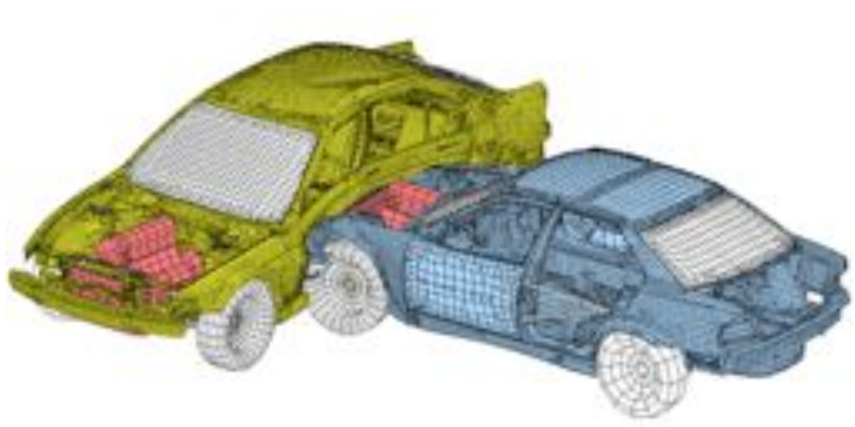
Theory

$$\partial_t \alpha - \beta^i \partial_i \alpha = -\alpha^2 f(\alpha)\ \mathrm{tr}K$$

$$\tilde{\Gamma}^i = \tilde{\gamma}^{jk} \tilde{\Gamma}^i_{jk}$$

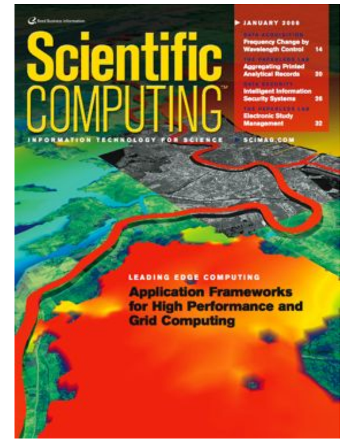
$$G_{\mu\nu} = 8\pi T_{\mu\nu}$$

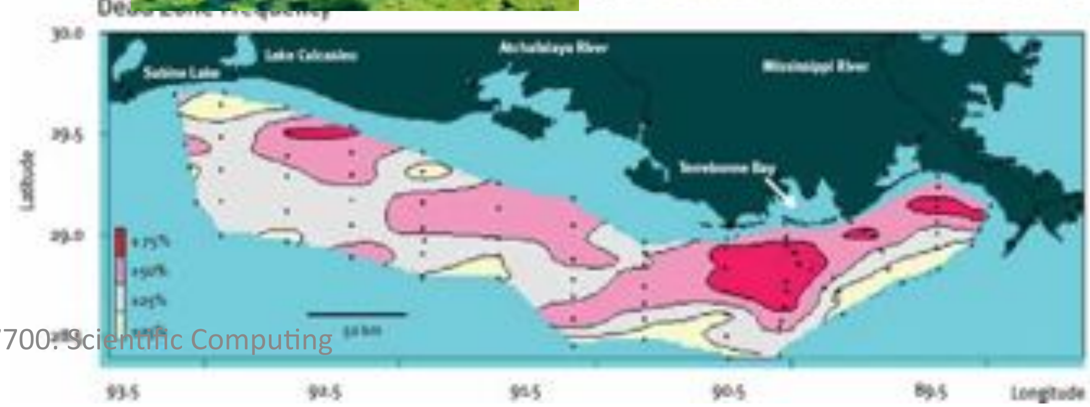
Crash Simulation



Modeling Louisiana Coastal Area

- Rich dynamic environment for modeling: coupled models, multi-scale, real-time data (sensors, satellites)
  - Hurricane forecasts
  - Emergency preparedness
  - Wetland reconstruction
  - Ecological studies and fish populations
  - Oil spill mitigation
  - Levee design
  - Sea rescue
  - Hypoxia “Dead Zone”
  - Algae blooms



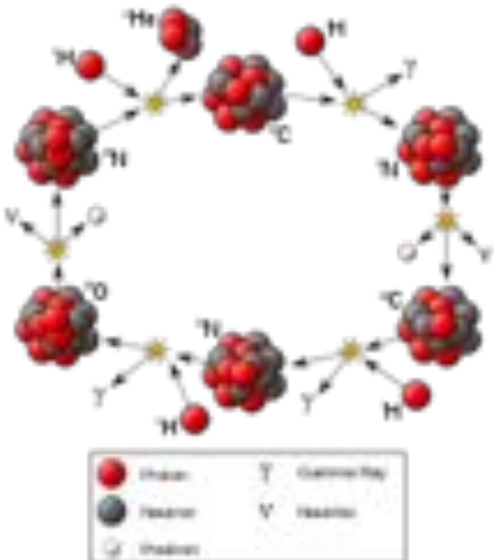


Hurricane Forecasting

- 5 days from landfall advisories from NHC provides best guess of track and intensity
- Surge/wave models prepared
- Wind forcing either from models or other methods
- Run large ensemble of models
- Archive model results
- Calculate products,
  - e.g. MOM, MOF, MEOW
  - compare with observations

Evacuation decisions 24hrs before landfall

Nuclear Fusion

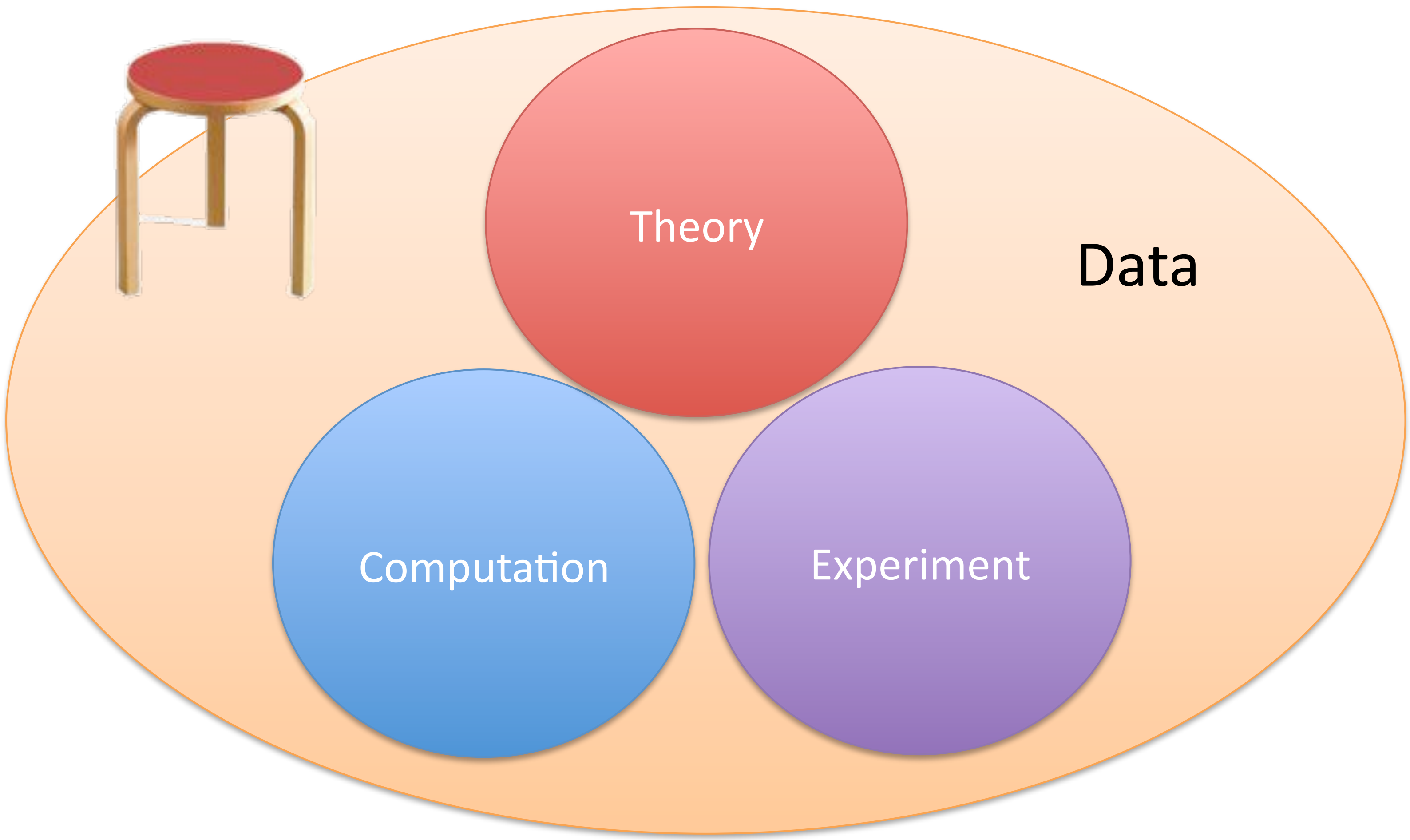


Tokamak reactor

- Aim: Efficient Power generation: no emissions, no radioactive by-products
- Models investigate properties of tokamak reactor: how to build to withstand fusion environment
- Validate models against experiments

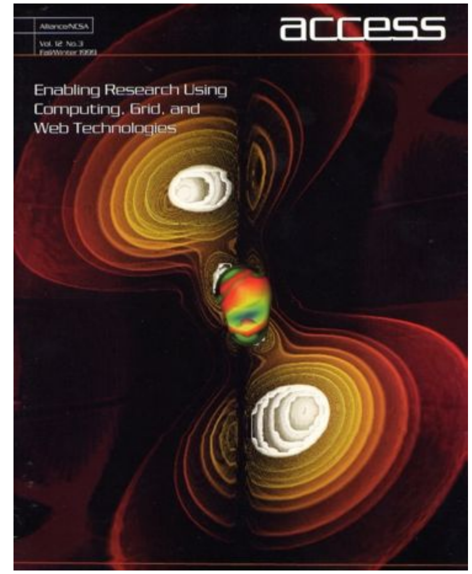
Industrial Modeling

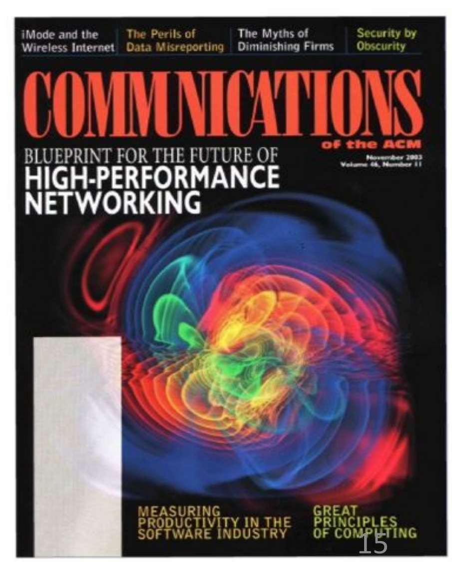
Modern Research: All Disciplines!!



Computational Science Needs

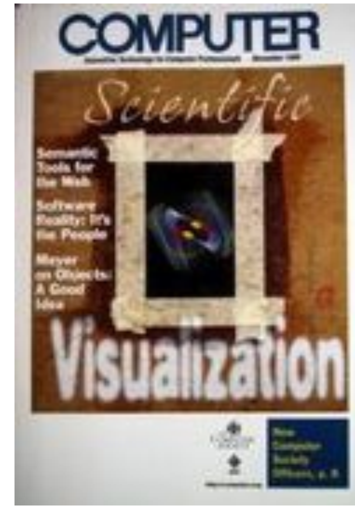
- Requires large mix of technologies & expertise!
- Collaboration across physics, engineering, computer science, applied mathematics, …
- numerical algorithms
  - Finite difference, spectral, monte carlo, elliptic
  - Multiscale, adaptive mesh refinement
- Different computational components
  - Large scale parallellization, new architectures
  - I/O, visualization
- New challenges
  - Petascale, data, complexity
- Complex infrastructure!
  - LONI, TeraGrid, DEISA, PRACE, Blue Waters, FutureGrid, Open Science Grid, …
- How to achieve all this? How to train students?

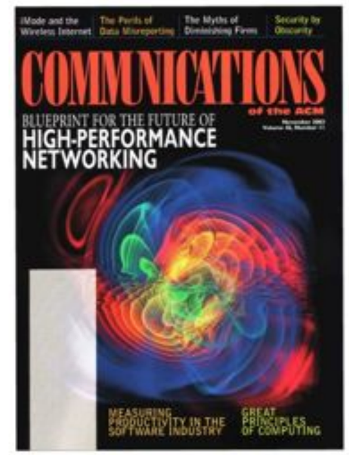


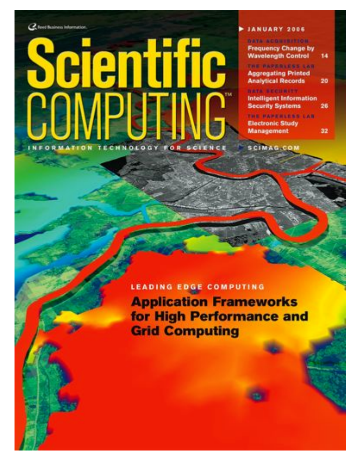


Research Issues

- Highly scalable algorithms (Petascale, exascale computing, heterogeneous architectures)
- Storage & networks: provenance, metadata
- Visualization (Large data, interactive, remote, AMR)
- Software engineering (code generation, verification & validation, interfaces, interoperability, etc)
- Workflows, Grid & distributed computing
- Education



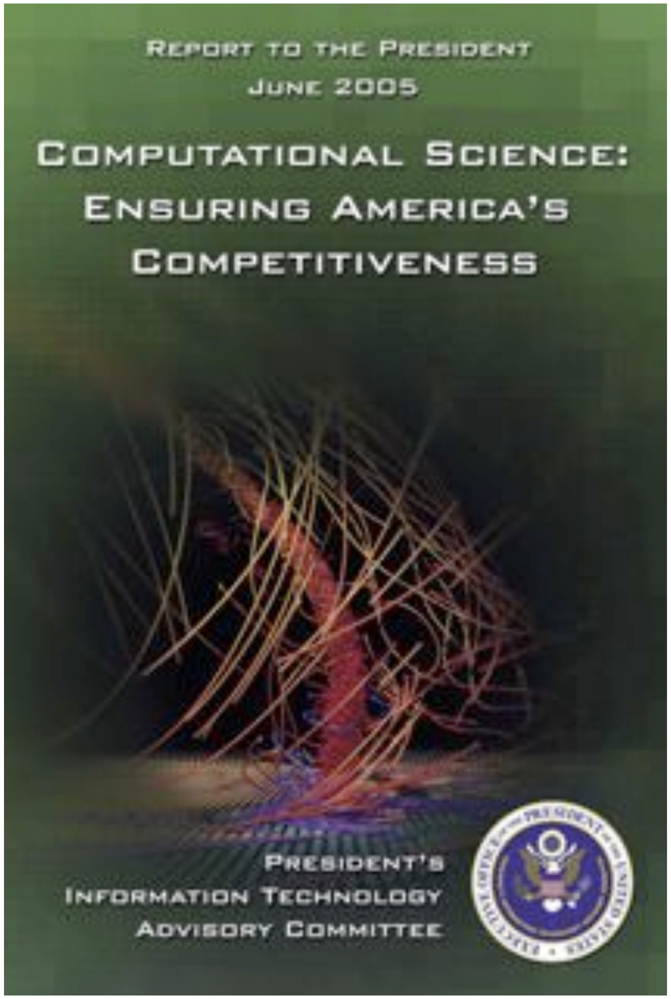




Nationally Relevant:
e.g. PITAC Report Summary:

"Computational science -- the use of advanced computing capabilities to understand and solve complex problems -- is critical to scientific leadership, economic competitiveness, and national security. It is one of the most important technical fields of the 21st century because it is essential to advances throughout society."

"Universities must significantly change organizational structures: multidisciplinary & collaborative research are needed [for US] to remain competitive in global science"



Collaborations for Complex problems:
Innovations will occur at boundaries

CSC 7700 Motivation

- Generally recognized that we do not have curricula in place to education students in computational science
  - Universities are changing: e.g. MS at LSU
- Lack of “real world” education puts our students at a disadvantage for modern research projects in scientific computing
- Experiences of the teaching faculty for this course!

Methodology and Modules

CSC 7700 Approach

- Design a course that prepares students for working in scientific computing research projects at CCT
  - For CS/EE students: understand issues of scientific computing, current technologies, get real experience
  - For domain students: understand CS/EE research directions, current technologies, get real experience
  - Learn to collaborate
- Teaching faculty already working together (e.g. NSF CyberTools)
  - Regular meetings over summer to discuss curricula and approach

Modules

- Each focused on one specific area. Emphasis on
  - Broad overview relevant to modern scientific computing, more depth into selected specific technologies
  - Real world use of current technologies
- Curricula, lectures, coursework, exam by leader in the field
  - Overseen by Drs. Allen/Jha
- New course and curricula developed
  - Teaching team have weekly meetings to ensure coordination between modules

Modules

- A: Basic Skills (Loeffler)
- B: Networks and Data (Hutanu)
- C: Simulations and Application Frameworks (Schnetter)
- D: Scientific Visualization (Benger)
- E: Distributed Scientific Computing (Jha)

A: Basic Skills

- This module will review of basic knowledge and prerequisites for scientific computing from a computer science and cyberinfrastructure perspective.
- Topics that are likely to be covered include:batch systems; unix, shells and ssh; parallelization; compilation and make; performance analysis and strong/weak scaling; verification and validation; sources of errors and numerical precision; overview of numerical methods and partial differential equations.

B: Networks and Data

- This module covers network basics, advanced network topics, middleware and distributed computing basics, data-intensive computing and grid-based visualization.
- Practical experience: iperf, GridFTP, high-speed networks, web services.

C: Simulation and Application Frameworks

- This module covers: What is simulation (focus on IVPs), e.g. basic schedule, components involves. High level overview of PDEs and numerical approximation. Design of applications, focusing on application frameworks. Case study for Cactus, including interfaces, capabilities, complexities involved in cutting edge applications, current needs.
- Practical work will include: Compile and run Cactus applications, understand Cactus configuration files, write simple Cactus component including verification.

D: Scientific Visualization

- Covers: Scientific Visualization vs. Visualization - scope and concepts; the visualization pipeline; differential geometry and differentiable manifolds; vector calculus; geometric algebra; discretized manifolds; topology; object-oriented and generic programming in C++; rendering: raytracing vs. rasterization; OpenGL; vertex buffer objects; vertex, geometry and fragment shaders; shading models; data types for scientific visualization; file formats and I/O; HDF5; modular programming; software components
- Practical experience: Able to use Vish to visualize data, ability to modify Vish components for visualization. Simple data reader? Something with data description?

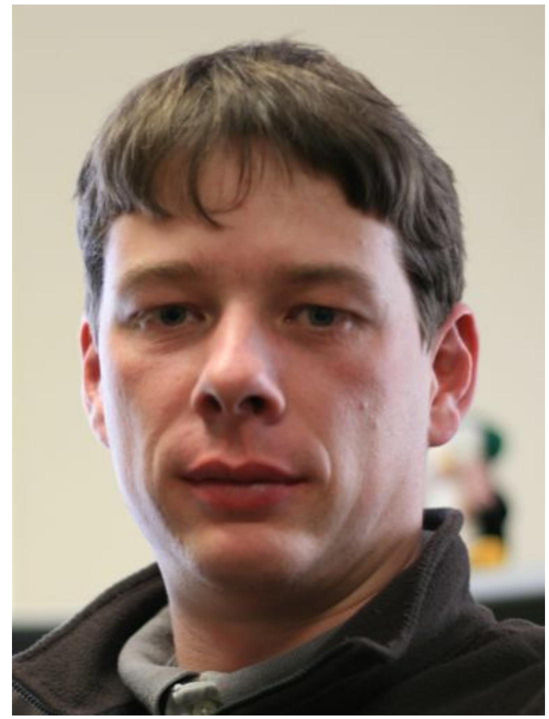
E: Distributed Scientific Computing

- This module will cover the theory and practice of Distributed Computing as applied to Scientific Applications. We will cover traditional Grids -- high-throughput as well as high-performance, as well as emerging infrastructure such as Clouds (eg Amazon EC2, Azure). We will use FutureGrid -- NSF's Track-2d Experimental Grid System as the test-bed.

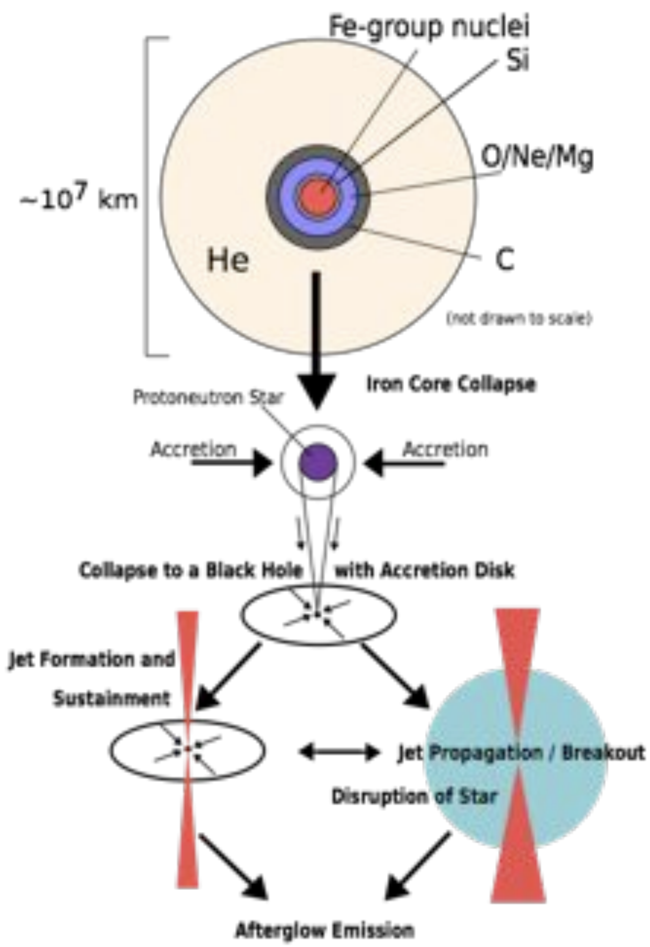
Teaching Faculty

Dr Frank Loeffler

- Module A: Basic Skills
- Postdoc, Center for Computation & Technology
- Relativistic astrophysics and high performance computing, lead developer for Einstein Toolkit Consortium



Gamma-Ray Burst Grand Challenge

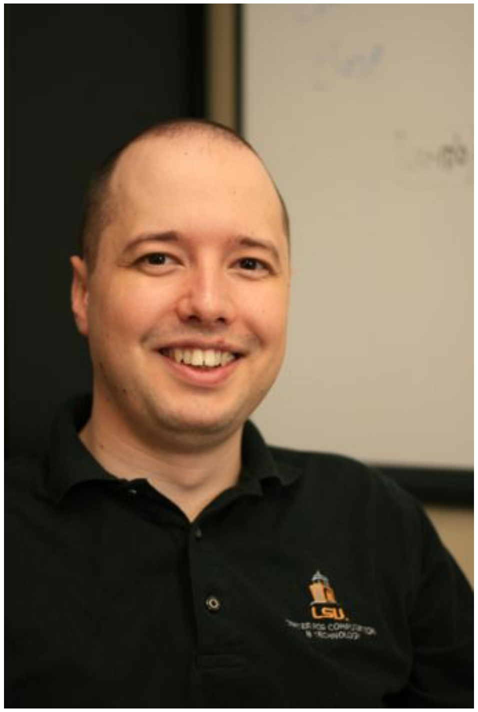


- Most energetic events in the universe
- Mechanism still a riddle; grand challenge in astrophysics
- Modeling requires expertise in many fields of physics (general relativity, magneto-hydrodynamics, neutrinos, ...)
- Requires petascale computing and collaboration

LSU

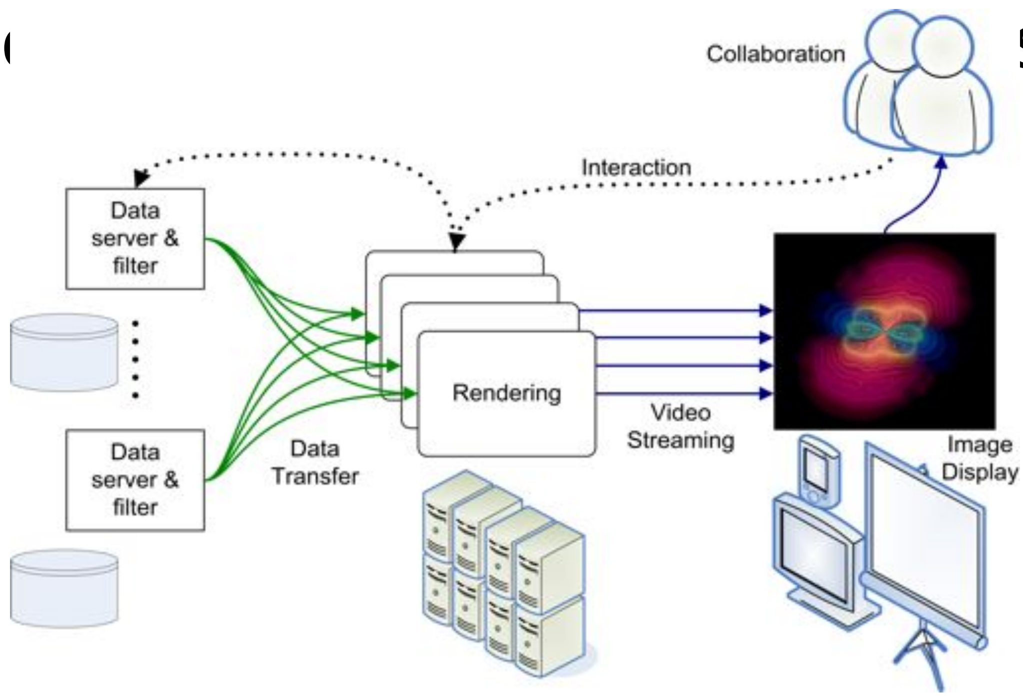
Dr Andrei Hutanu

- Module B: Networks & Data
- Research Scientist, Center for Computation & Technology
- Applications of high speed networks, network testbeds, lead of eaviv project
- Lead of $300K NSF EAGER project



Remote Interactive Viz: eaviv

- NSF Project (LSU/NCSA): Use new optical networks and services for dynamic configuration (Internet2 DCN)
- Target: Truly interactive visualization of larg s

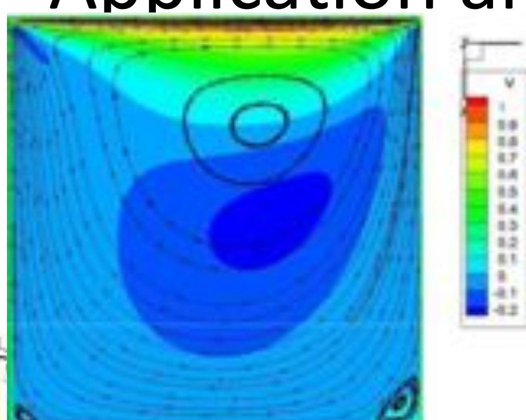
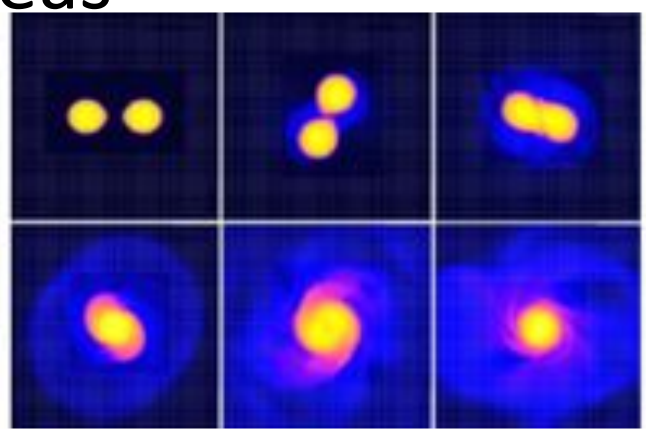
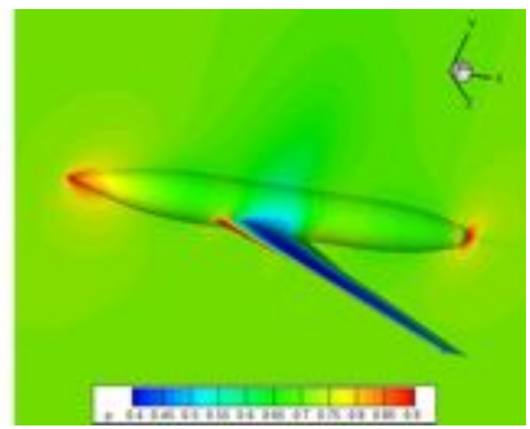


Dr Erik Schnetter

- Module C: Simulations & Application Frameworks
- Assistant Research Professor, Physics & Astronomy
- Relativistic astrophysics, Cactus Framework lead architect, originator of Carpet AMR infrastructure
- PI of $1.5M NSF PetaApps project

Cactus

- Open source component framework for HPC
- Modular system with high level abstractions
  - Components “thorns” defined by their parameters, variables, methods
  - Cactus “flesh” binds everything together
  - Cactus Computational Toolkit
- Application areas

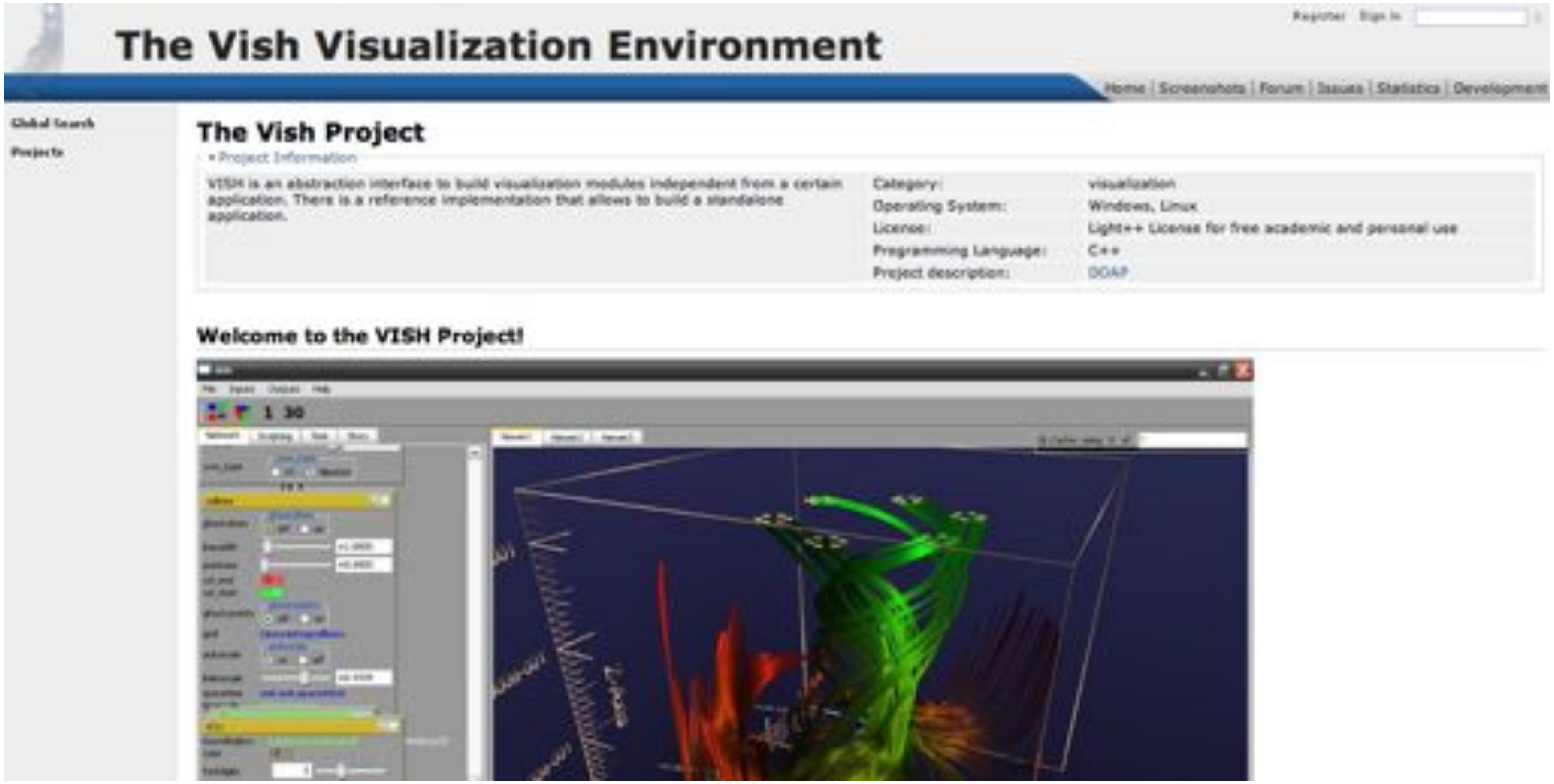




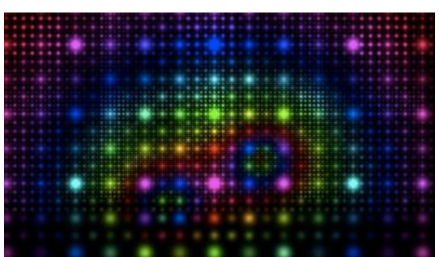
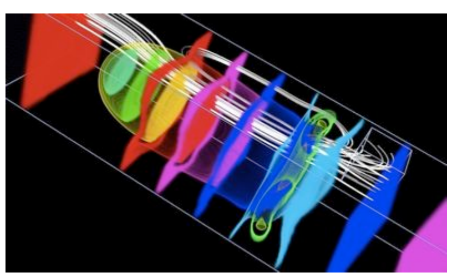
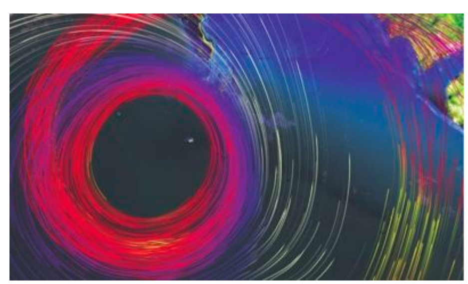
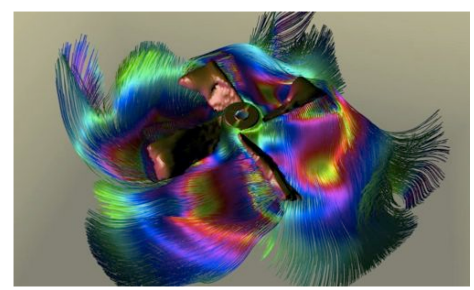
Cactus Framework

Dr Werner Benger

- Module D: Advanced Visualization
- Research Scientist, Center for Computation & Technology
- Originator of the Vish visualization framework

Vish

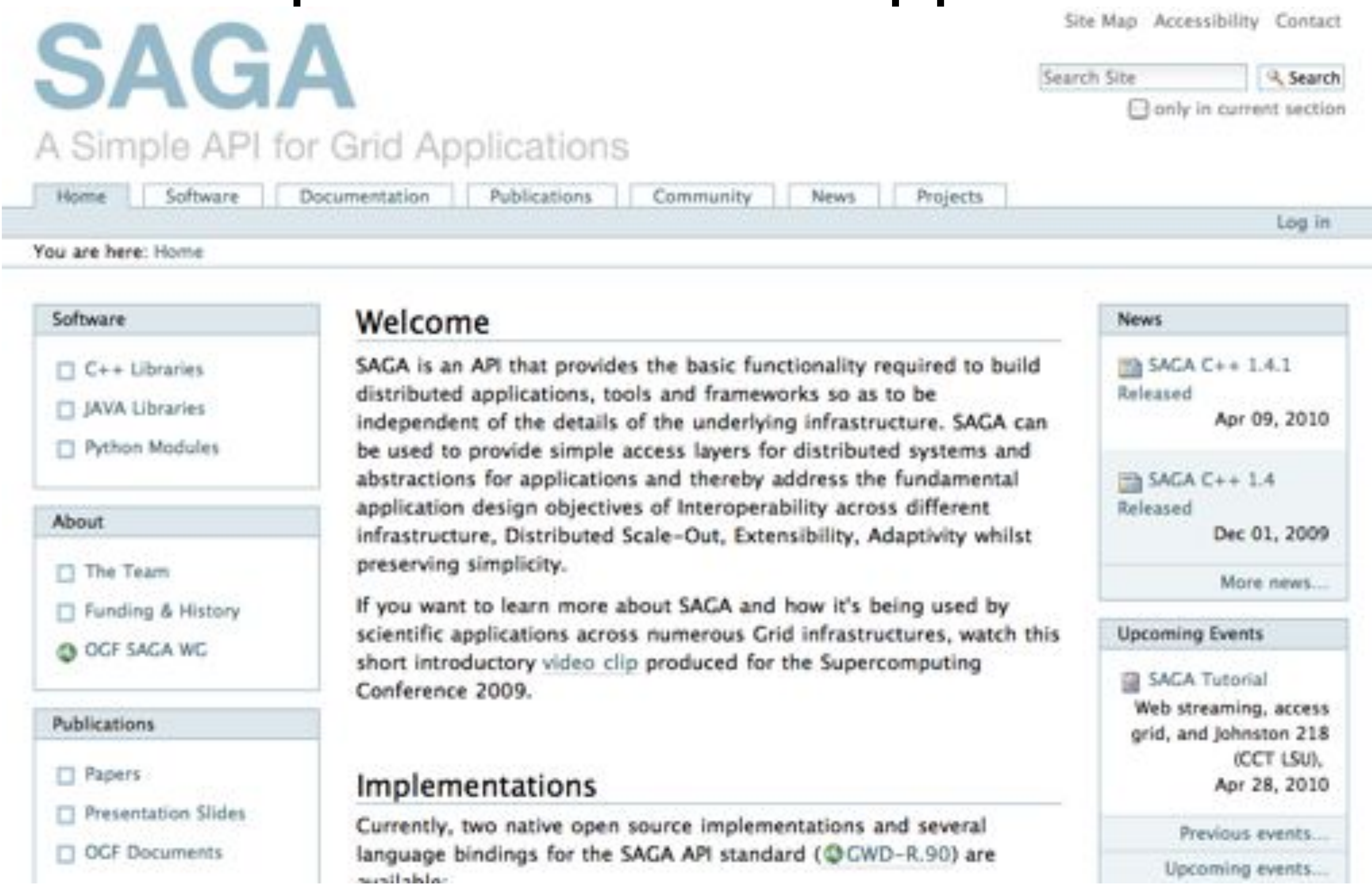

The Vish Visualization Environment
Register Sign in
Home | Screenshots | Forum | Issues | Statistics | Development
Global Search
Projects
The Vish Project
Project Information
VISH is an abstraction interface to build visualization modules independent from a certain application. There is a reference implementation that allows to build a standalone application.
Category: visualization
Operating System: Windows, Linux
License: Light++ License for free academic and personal use
Programming Language: C++
Project description: DOAP
Welcome to the VISH Project!

Dr Shantenu Jha

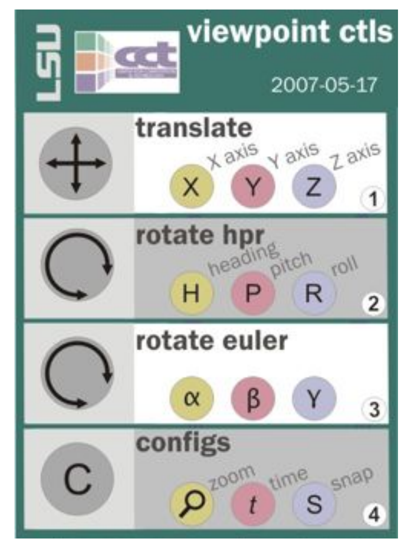
- Module E: Distributed Scientific Computing
- Assistant Research Professor, Computer Science, Director for Cyberinfrastructure Development CCT
- Lead of the SAGA project

Simple API for Grid Applications



Viz Tangibles

Above left: co-located and distributed users collaboratively manipulate a 3D visualization in an AccessGrid meeting using viz tangibles interaction devices
Right: an "interaction tray" is used together with RFID-tagged "tangible menus" to access and manipulate scientific visualizations



Credit: Brygg Ulmer

Sponsored by:

NSF

PetaShare

A Distributed Data Archival, Analysis and Visualization Cyberinfrastructure for Data-intensive Collaborative Research.

Announcements:

PetaShare Storage is Online! April 21, 2008

PetaShare storage is now online and accepting allocation proposals. [Read More]

Overview

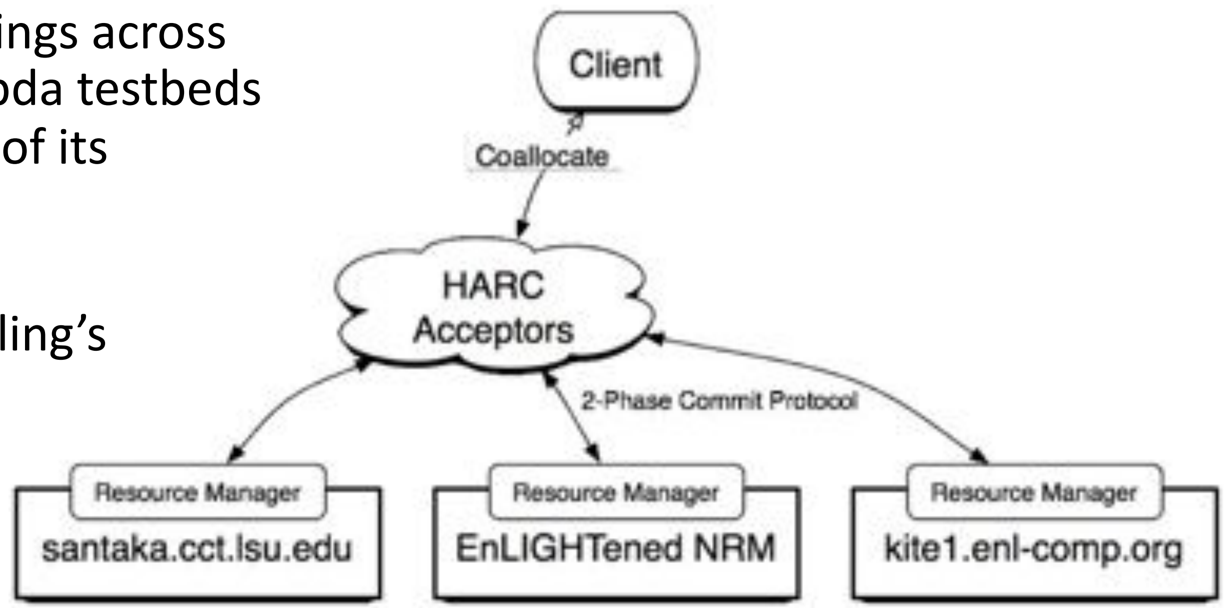
PetaShare is an NSF sponsored project which responds to the urgent need of scientists who work with large-scale data generation, sharing and collaboration requirements. PetaShare aims to enable domain scientists to focus on their primary research problem, assured that the underlying infrastructure will manage the low-level data handling issues.

PetaShare employs a very novel approach to solve the distributed data sharing and management problem. Unlike existing approaches, PetaShare treats data storage resources and the tasks related to data access as first class entities just like computational resources and compute tasks, and not simply the side effect of computation. The key technologies that are being developed in this project include data-aware storage systems, data-aware schedulers, and cross-domain metadata scheme which take the responsibility of managing data resources and scheduling data tasks from the user and perform these tasks transparently.

An initial prototype of PetaShare is deployed at seven Louisiana campuses: Louisiana State

HARC: Highly Available Resource Co-scheduler

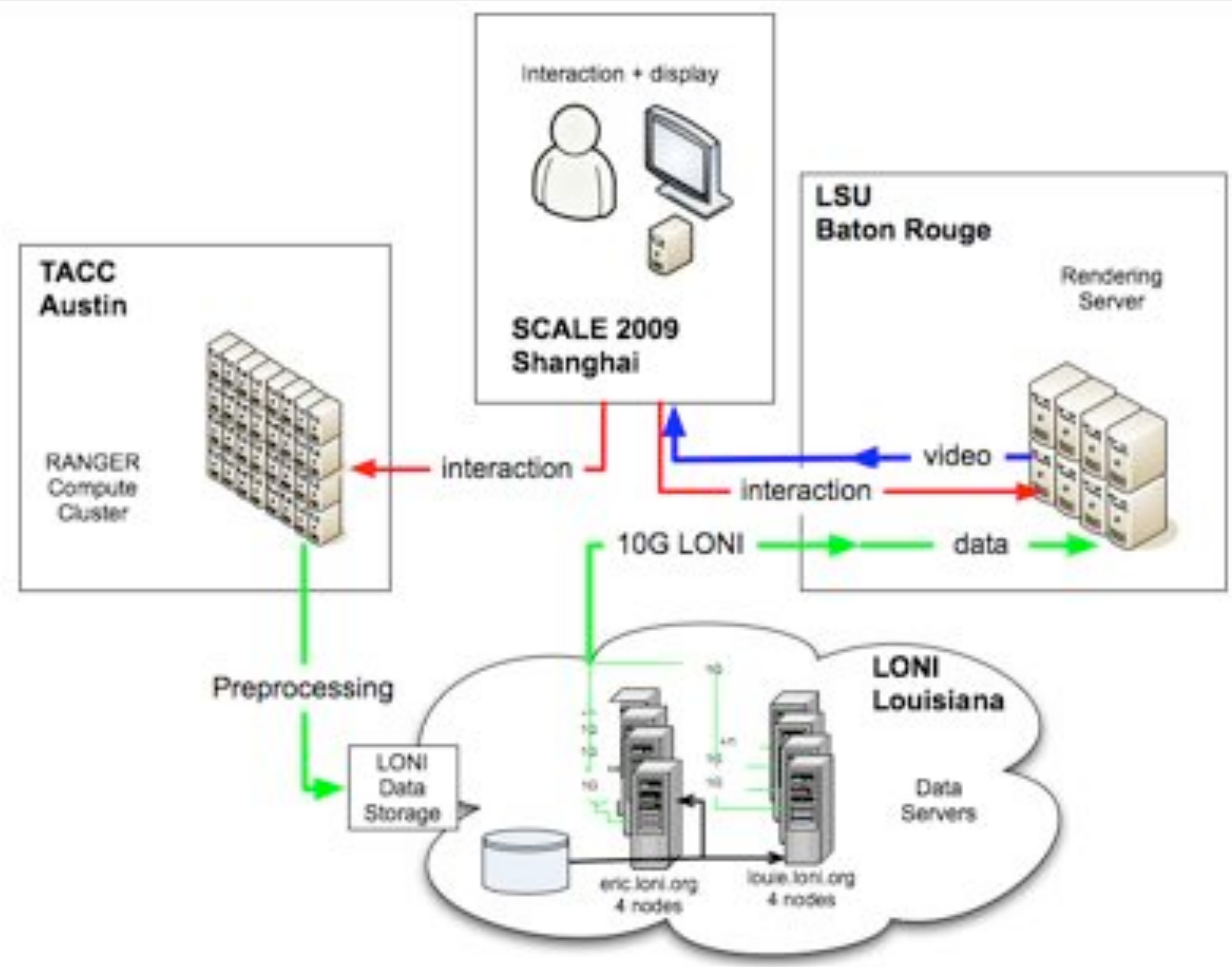
- Extensible, open-sourced co-allocation system
- Can already reserve:
  - Time on supercomputers (advance reservation), and
  - Dedicated paths on GMPLS-based networks with simple topologies
- Uses Paxos Commit to atomically reserve multiple resources, while providing a highly-available service
- Used to coordinate bookings across EnLIGHTened and G-lambda testbeds in largest demonstration of its kind to date
- Used for setting up the network for Thomas Sterling's HPC Class which goes out live in HD



Credit: Jon MacLaren

Winners IEEE SCALE09!

(Hutanu, Schnetter et al)



- 2048 core simulation code
- 8 node renderer at LSU
- Remote LONI nodes as data servers
- 1024^3 spatial resolution (1GByte/ timestep)
- 20 timesteps cached remotely
- Scalable Approach!!!

Grading

CSC 7700: Grading

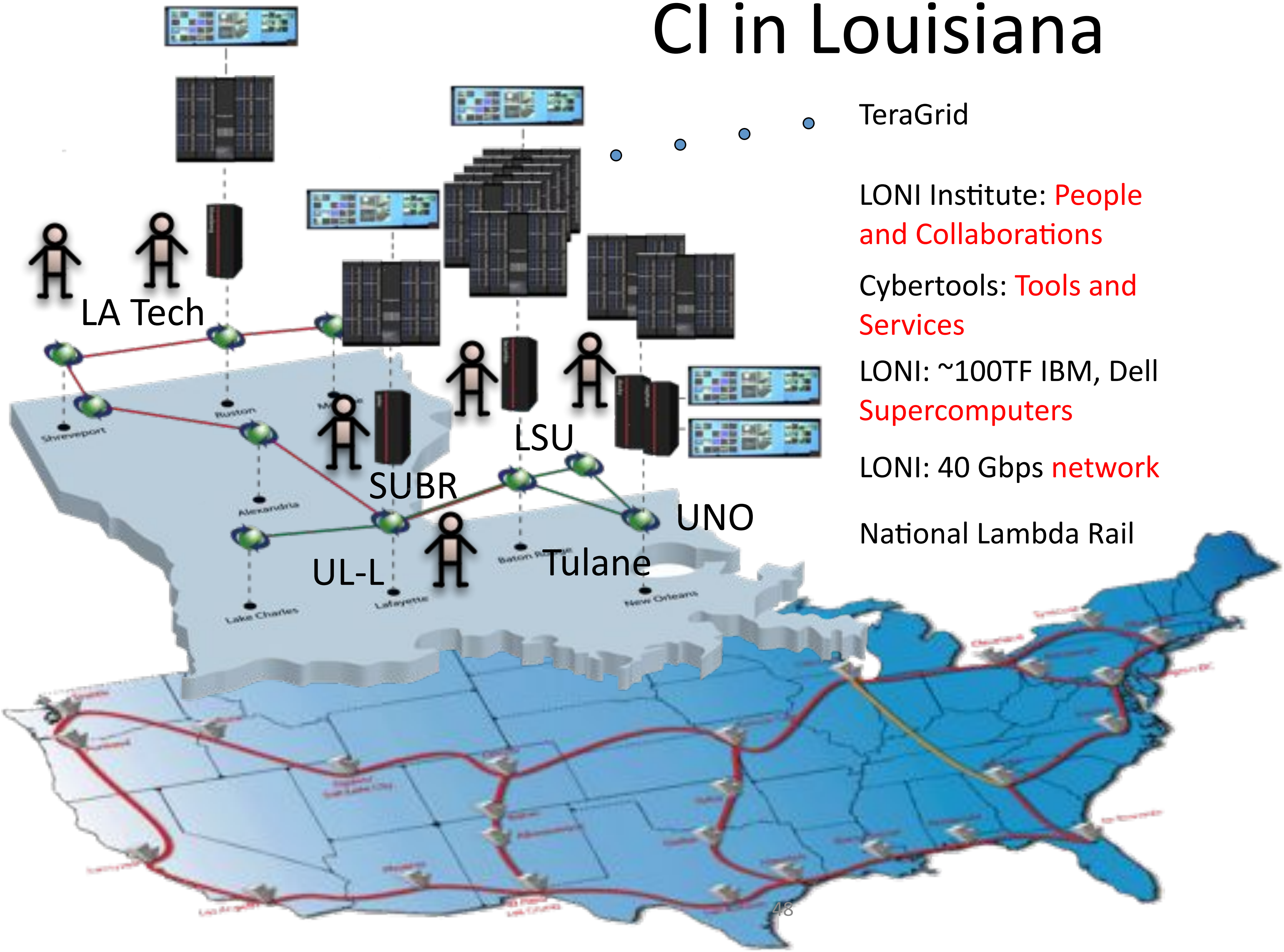
- Emphasis is on practical work and experience
- Aim is not to have tricky exam questions, but to convince us that you have acquired real experience
- This is the real world! Don't leave coursework till the last night, you will need to put in real hours and solve issues yourself or with colleagues

CSC 7700: Grading

- The class will be graded by module with a final exam which will cover all modules.
- The breakdown of grades between modules and the final exam is
  - 20% Final exam
  - 20% Module B: Networks and Data
  - 20% Module C: Simulations and Application Frameworks
  - 20% Module D: Scientific Visualization
  - 20% Module E: Distributed Scientific Computing
  - Within each module, half of the grades will be for coursework, with the other half awarded either by project, essay, or exam depending on the module.

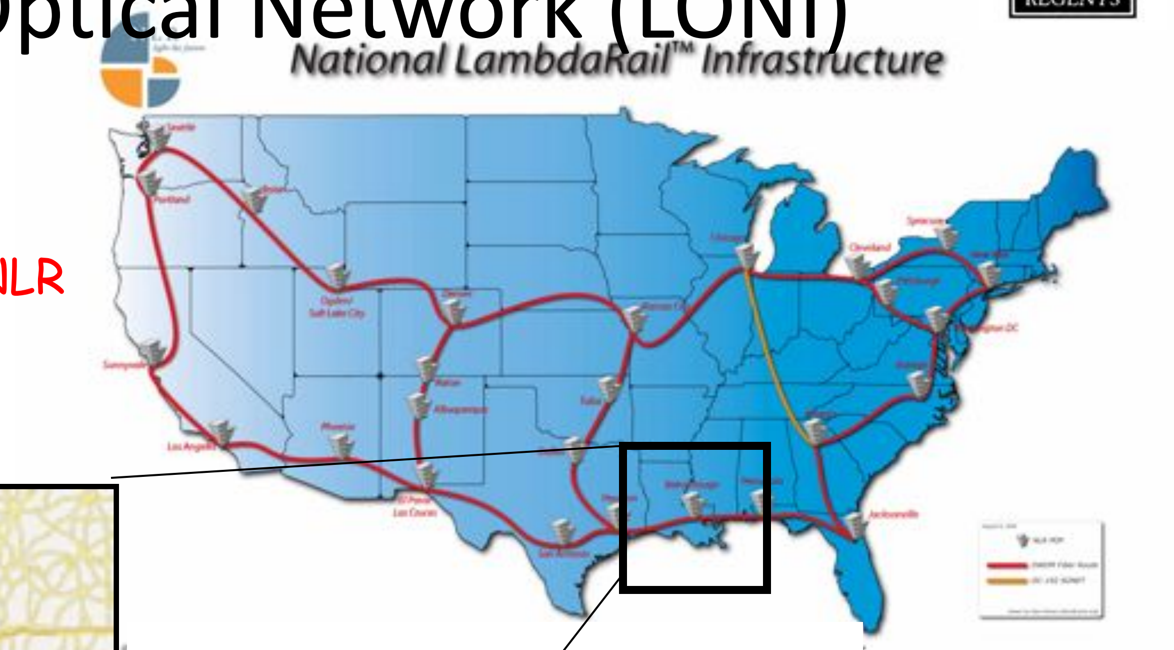
Cyberinfrastructure Resources

CI in Louisiana



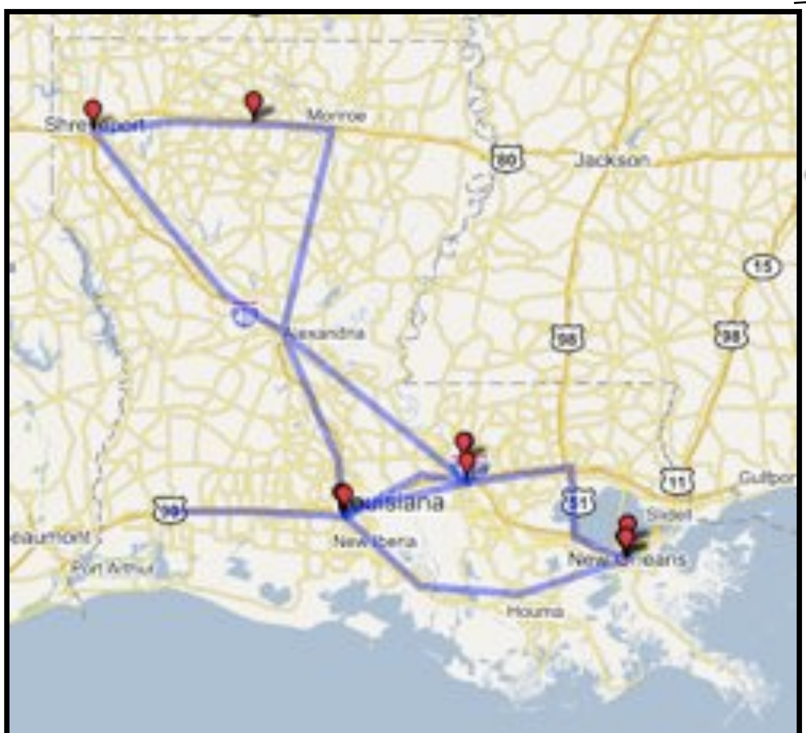
- TeraGrid
- LONI Institute: People and Collaborations
- Cybertools: Tools and Services
- LONI: ~100TF IBM, Dell Supercomputers
- LONI: 40 Gbps network
- National Lambda Rail

Louisiana Optical Network (LONI)



State initiative ($50M) to support research (2004):

- 40 Gbps optical network + NLR
- Connects state universities, health science centers



- Compute resources: ~100 Tflops across state
- Data resources ~500TB with NSF PetaShare
- LONI customers: MS universities, K12, hospitals, LPB

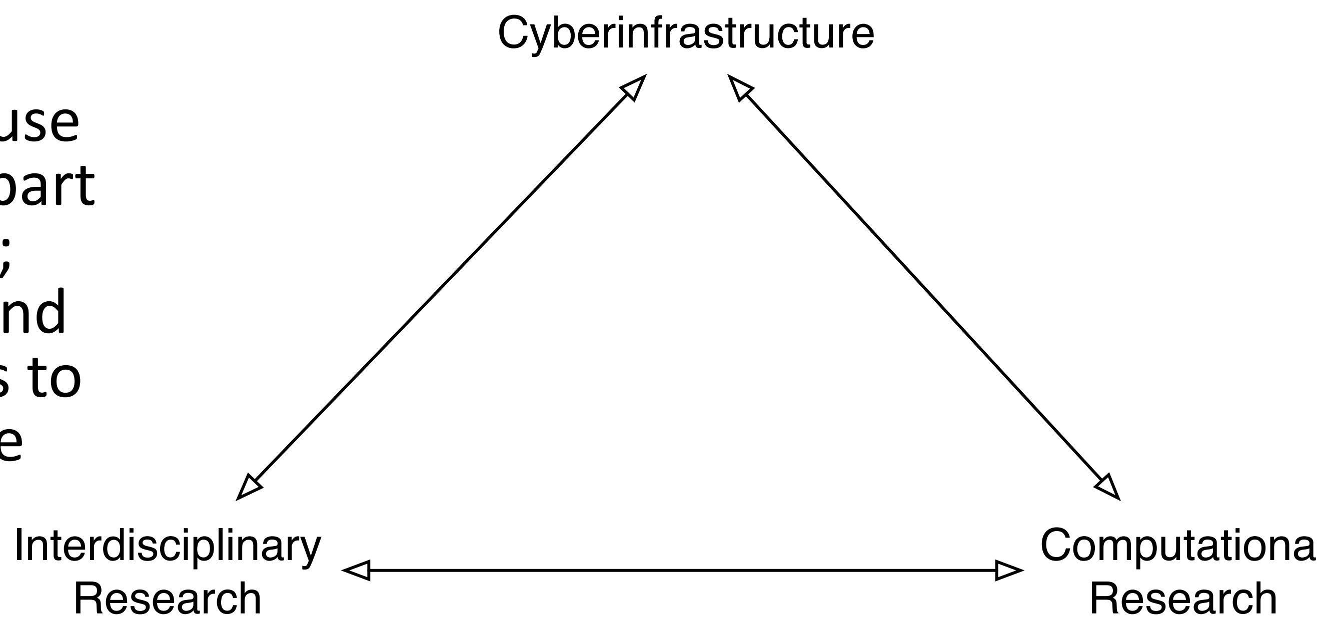
NSF TeraGrid

Center for Computation & Technology

- **Mission**: The CCT at LSU is an innovative and interdisciplinary research environment, advancing computational sciences, technologies, and the disciplines they touch. The CCT serves Louisiana through international collaboration, leading progress through revolutionary advancement in academia and industry.
- **Strategic Plan:**
  - Covers: Research; Education; Service; Infrastructure; Economic Development
- **https://cct.lsu.edu/uploads/CCTStrategicPlan20062010.pdf**

CCT Plan: Designing a Computational Science Initiative at LSU

- **Cyberinfrastructure**: access to local/national compute, storage, network, visualization resources; support; end-to-end integration.
- **Interdisciplinary research**: university policies on joint appointments, university-wide curricula, appreciation of computational science outcomes (e.g. software, data)
- **Computational research**: faculty who develop or use computation as part of their research; undergraduate and graduate courses to train and educate students



Louisiana Response: Center for Computation & Technology (CCT)

- State commitment for IT in 2002
  - $25M/year for Vision 20/20, $9M to LSU
  - University commitment to build new programs
  - Opportunity to build new world class program in interdisciplinary research & education
- Ed Seidel recruited to LSU, created & implemented vision for state-wide collaboration
  - Center for Computation & Technology (2003)
  - Louisiana Optical Network Initiative (2005)
  - CyberTools (2007), LONI Institute (2007)
  - Multidisciplinary Hiring Initiatives (2007)

This Class is for You!

- Ask questions, give us suggestions, let us know what you don't know or understand
- Let us know right away about any problems particularly as we start the practical work
- Plan to request your participation in initial and final survey so we can analyze the effect of this class.