Louisiana RII
CyberTools and Science Drivers Symposium

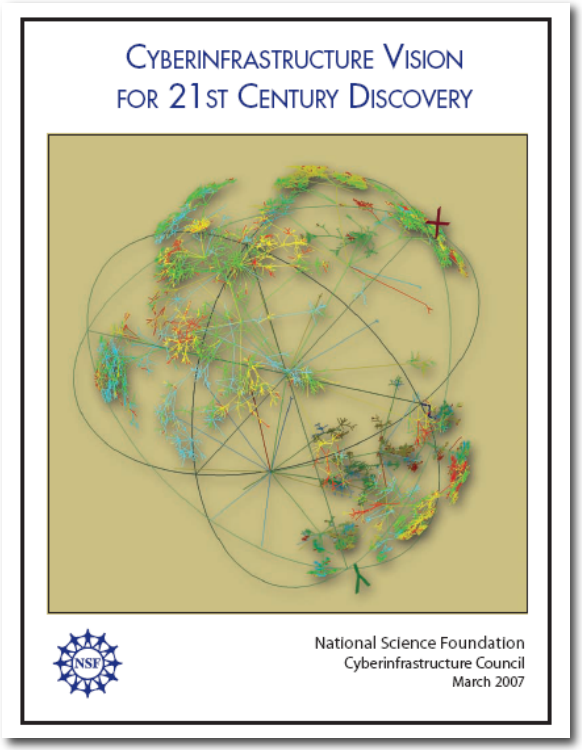
CyberTools Workpackages

Gabrielle Allen
Associate Professor, Computer Science
Louisiana State University

http://www.cybertools.loni.org

Goal of CyberTools

- Using motivating science drivers, research & develop technologies to further computational science in Louisiana and provide a sustainable, enabling cyberinfrastructure:
  - *Sharing of data at 10gbps end-to-end across LA*
  - *Scheduling all aspects of complex workflows*
  - *Data mining: experimental & simulation data*
  - *Science gateways: transparent user interfaces*
  - *Common data model; advanced, distributed viz*
  - *High level application toolkits and managers*



Bring users and tools to the national level (TeraGrid, Blue Waters, XD, DataNet…)

Louisiana EPSCoR

CyberTools Team

Louisiana
EPSCoR

The WorkPackages

- WP1: Data & Scheduling (Kosar)
  - **Tevfik Kosar**, Ismail, Mehmet, Ibrahim, Jack, Esma, Dengpan
  - **Shantenu Jha**, Zhifeng
  - **Box Leangsuksun**, Thanadach

- WP2 Information and Portals (Dua)
  - **Sumeet Dua**, Pradeep, Xian, Harpreet
  - **Gabrielle Allen**, Bharadhwaj, Archit
  - **Honggao Liu**, Sirish
  - **Sumanta Acharya**, Prasad
  - **Don Gaver**, Jerina
  - **Iyengar/Brener**, Swathi

- WP3 Visualization and Networks (Cruz-Neira, Ullmer)
  - **Brygg Ullmer**, Kexi, Cornelius
  - **Carolina Cruz-Neira**, Nikhil, Vignesh
  - **Amitava Jana**, Kazim, Sanjay
  - **Iyengar/Brener**, Bidur
  - **Ray Jindal**, Vinayak, Ninad
  - **Gabrielle**, Vinay, Farid, Werner, Andrei, Jinghua

- WP4 Application Toolkits (Jha)
  - **Shantenu Jha**, Abhinav, Ole, Nayong
  - **Mayank Tyagi**, Jeff
  - **Nat Brener**, Harsha
  - **Amitava Jana**, Sanjay
  - **Gabrielle Allen**, Lei, Frank, (Ashley)
  - **Erik Schnetter**, Oleg
  - **Don Gaver**, Kate, Jerina
  - **Sumanta Acharya**, Prasad, Somnath

Louisiana EPSCoR

CyberTools Team

- WP1: Data & Scheduling (Kosar)

3 GROUPS

- WP2 Information & Portals (Dua)

6 GROUPS

- WP3 Visualization and Networks (Cruz-Neira, Ullmer)

6 GROUPS

- WP4 Application Toolkits (Jha)

8 GROUPS

Louisiana EPSCoR

CyberTools Team

- WP1: Data & Scheduling (Kosar)

3 GROUPS

- WP2 Information & Portals (Dua)

6 GROUPS

- WP3 Visualization and Networks (Cruz-Neira, Ullmer)

6 GROUPS

- WP4 Application Toolkits (Jha)

8 GROUPS

Great opportunities and great challenges ☺

Louisiana EPSCoR

New Postdocs !!!

Dr Pradeep Chowriappa
WP2

Dr Xian Du
WP2

Dr Nayong Kim
WP4

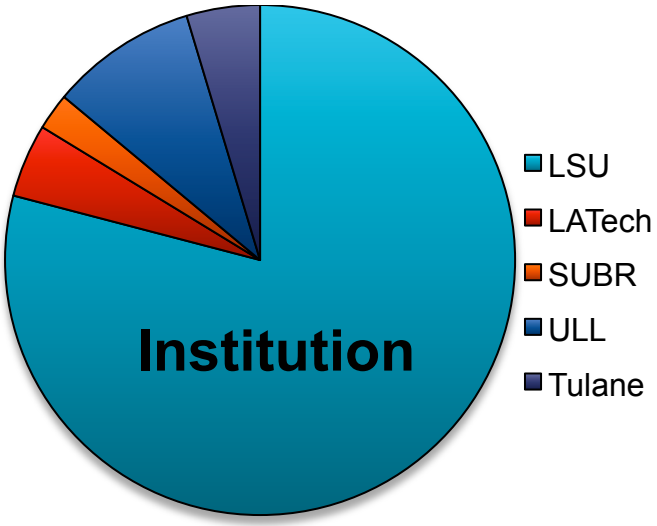
Dr Soon-Heum Ko
WP4

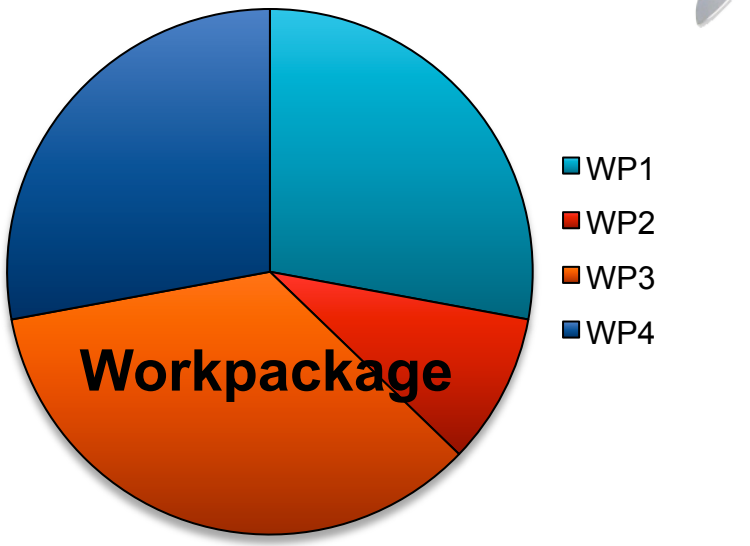
Dr Frank Loeffler
WP4

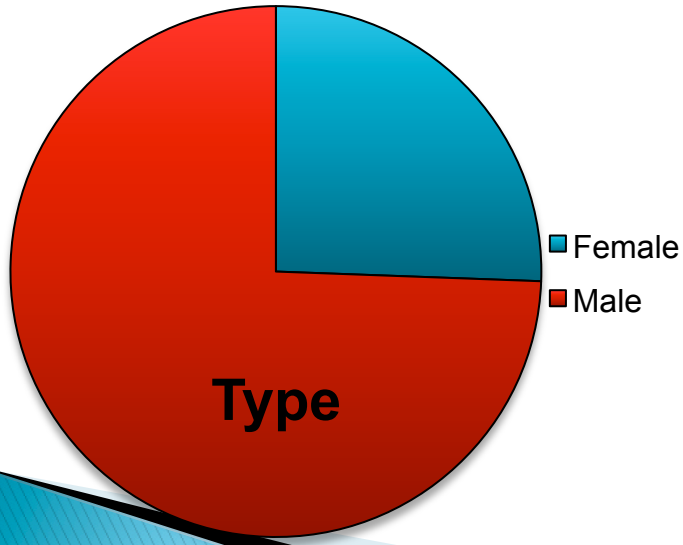
Dr Sanjay Kodiyalam
WP3/4

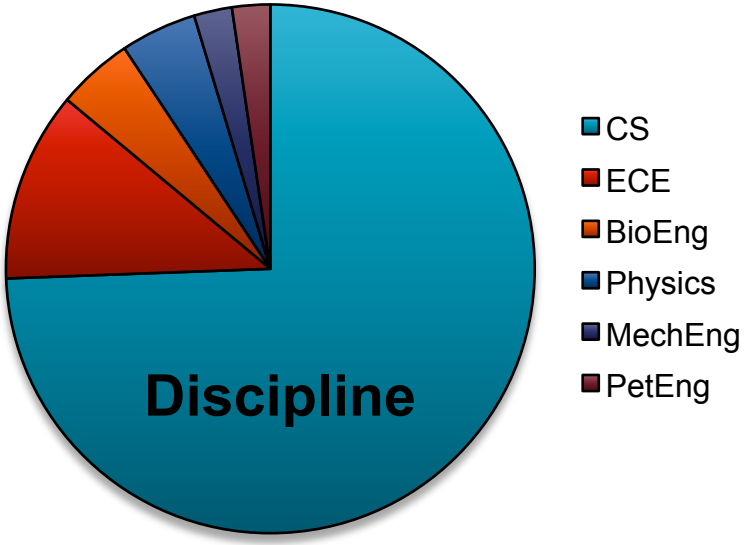
Louisiana
EPSCoR

43 Affiliated Graduate Students









Louisiana EPSCoR

Usual Computational Science

- Researchers each write their own codes (from scratch)
- Everyone has their own data format (often ASCII)
  - Not possible to share data, visualization tools
  - Hard to take advantage of e.g. parallel I/O
- Data files shared by email
- Codes coupled inefficiently via file exchange, wrappers
- Applications deployed manually, time consuming
- Etc.

- *Barrier to going beyond single code, single group*
  - *Hard to do interdisciplinary research !!!*
- *Scientific investigation is limited by our tools!*

Louisiana EPSCoR

WP1 Major Achievements

- Research:
  - A New Paradigm: Data-Aware Scheduling in Grid Computing (T. Kosar et al., FGCS)
  - Louisiana: A Model for Advancing Regional e-Research through Cyberinfrastructure (Phil. Trans. Royal Society)
  - Semantic Enabled Metadata Management in PetaShare (T. Kosar et al., Inter. Journal of Grid and Utility Computing)
- Funding:
  - NSF CAREER Award (Kosar)
- Development
  - PetaShare online at 7 sites, used by over 15 groups
  - Four different PetaShare interfaces for users
  - New release of Stork data scheduler (Stork 1.0

At least 2 book chapters, 3 journal publications, 6 peer-reviewed conference publications

Louisiana EPSCoR

WP1 Major Achievements

- Research:
  - A New Paradigm: Data-Aware Scheduling in Grid Computing (T. Kosar et al., FGCS)
  - Louisiana: A Model for Advancing Regional e-Research through Cyberinfrastructure (Phil. Trans. Royal Society)
  - Semantic Enabled Metadata Management in PetaShare (T. Kosar et al., Inter. Journal of Grid and Utility Computing)
- Funding:
  - NSF CAREER Award (Kosar)
- Development
  - **PetaShare online at 7 sites, used by over 15 groups**
  - **Four different PetaShare interfaces for users**
  - New release of Stork data scheduler (Stork 1.0

At least 2 book chapters, 3 journal publications, 6 peer-reviewed conference publications

Louisiana EPSCoR

WP1 Science Driver Interactions

- Biotransport:
  - Data storage & scheduling for CFD (LSU/Acharya)
- Bioinformatics/biostatistics:
  - End-to-end workflow management for protein folding & clustering (LSU/Bishop@Tulane)
  - Data storage for finite element modeling in bioinformatics (LSU/ Kodiyalam@SUBR)
- Other:
  - Data storage & scheduling for numerical relativity (LSU)
  - Distributed visualization (LSU)
  - End-to-end workflow management for reservoir uncertainty analysis (LSU, UCOMS) and coastal hazard prediction (LSU, SCOOP/COMI)

Louisiana EPSCoR

WP2 Major Achievements

- Research
  - Unique image classification algorithmic tool using associative patterns (S. Dua et al., ESA)
  - Data mining based information fusion schema with robustness to noise and anomalies (S. Dua et al., IJBA)
  - Feature ranking algorithm for supervised data mining methods (S. Dua et al., IEEE-CIBCB'08)
  - Wavelets-based dimensionality reduction algorithmic tool (S. Dua et al., BIOT-2008)
- Funding:
  - BOR Award (PI: Dua, Co-PI: Leangsuksun).
  - AFRL Funding for data fusion (Dua)
- Development
  - Initial application portals in place
  - Data mining tools identified and being turned into packages

At least 2 journal publications, 5 peer-reviewed conference publications.

Louisiana EPSCoR

WP2 Major Achievements

- Research
  - Unique image classification algorithmic tool using associative patterns (S. Dua et al., ESA)
  - **Data mining based information fusion schema with robustness to noise and anomalies (S. Dua et al., IJBA)**
  - Feature ranking algorithm for supervised data mining methods (S. Dua et al., IEEE-CIBCB'08)
  - Wavelets-based dimensionality reduction algorithmic tool (S. Dua et al., BIOT-2008)
- Funding:
  - BOR Award (PI: Dua, Co-PI: Leangsuksun).
  - AFRL Funding for data fusion (Dua)
- Development
  - **Initial application portals in place**
  - Data mining tools identified and being turned into packages

At least 2 journal publications, 5 peer-reviewed conference publications.

Louisiana EPSCoR

WP2 Science Driver Interactions

- Immunosensors:
  - Quantitative performance evaluation of micromixer (LATECH/Decoster, Senaka)
  - Quantitative characterization of cell growth (LATECH/DeCoster)
- Biotransport:
  - Dimensionality reduction methods (LATECH/Devireddy, Moldovan)
  - Developing CFD Portal (LSU/Acharya)
- Bioinformatics/biostatistics:
  - Data mining, information fusion methods (LATECH/Thompson, Hill)
  - Data clustering performance evaluation methods (LATECH/LSU)
  - Protein clustering by folding structures (LSU/Bishop, Blake, Ashbaugh)
- Other:
  - Astrophysics: LSU developing Black Hole K12 Portal

Louisiana EPSCoR

CFD Portal

GridSphere Portal

https://portal1.loni.org:8444/gridsphere/gridsphere?cid=CFD+Project

Apple Yahoo! Google Maps YouTube Wikipedia News (424) Popular How to Bull...m Doghouse TVNSports Using Boost ...s in Eclipse Member Benefits, Inc. Alex Orr's bl...hould use It

CFD Simulation Portlet

Enter Simulation Information here:

Number of Time Steps: 1000

Each Time Step Value: 0.00001

Input Grid File Path: Choose File no file selected

Use Restart File? Yes No

Input Restart File Path: Choose File no file selected

Customize Boundary Conditions? Yes No

Boundary Conditions:

Change Input Profile? Yes No

Profile File Path: Choose File no file selected

Profile Orientation Direction Coordinates

Either Three Real values for Direction Cosine

Boundary Condition Type: Wall, Neumann, Dirichlet, Convective

Physical Parameters:

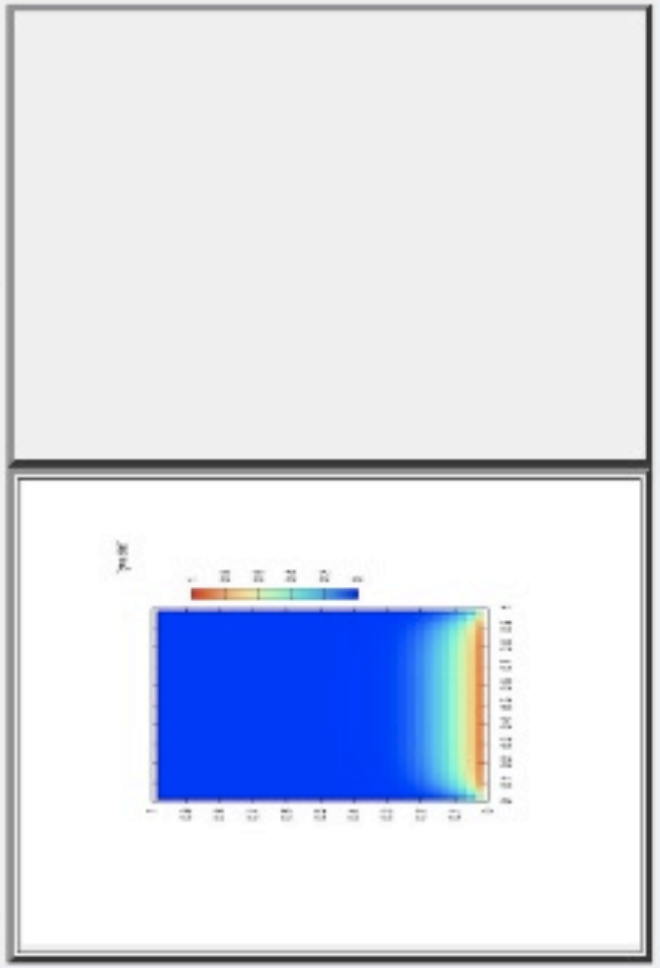
Thermal Conductivity: 1.0 (W/mK)

Density: 1.0 (Kg/m3)

Dynamic Viscosity: 0.001 (PaS)

Specific Heat: 1.0 (KJ/KgK)

Start the Simulation



WP3 Major Achievements

- Research
  - Common file format (F5) for datasets from different SDs (NJP'08)
  - Handling of curvilinear multiblock data & streamline computation
  - Optimized algs. for scalar, vector, tensor fields for interactive exploration
  - Viz Tangibles + Function Blade developments (TEI'08, TEI'09)
  - Alg for seed points, line integration in vector fields (Benger, Acharya et al WSCG'09)
  - Interactive Exploration Of Coastal Restoration Modeling In Virtual Environments. (EVR'09)
  - Virtual Hydrology Observatory: Immersive Visualization Of Hydrology Modeling (EVR'09)
  - Interactive Tracking of 3D Critical Points in Unsteady, Large-scale CFD (CGVR'08)
  - Selected for SCALE09 competition for large scale viz (Shanghai, May 09) (LSU)
  - Initial network model of LONI Southern Ring (ULL)
  - Integration with VISH + Tangibles in context of several SD applications
- Funding
  - Over three proposals submitted (NSF STCI, MRI, CPS)
- Planning and Development
  - Jindal/Park abstract
  - CAVE environment at SUBR
  - Two successful LONI/LSU networking meetings held.

Louisiana EPSCoR

WP3 Major Achievements

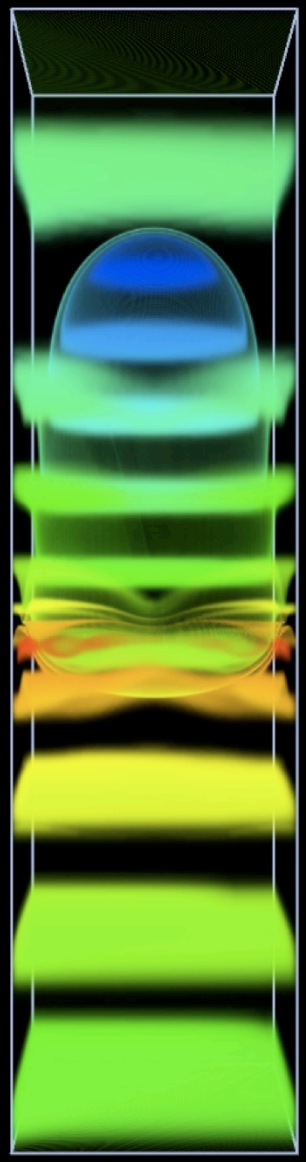
- Research
  - **Common file format (F5) for datasets from different SDs (NJP'08)**
  - **Handling of curvilinear multiblock data & streamline computation**
  - **Optimized algs. for scalar, vector, tensor fields for interactive exploration**
  - Viz Tangibles + Function Blade developments (TEI'08, TEI'09)
  - **Alg for seed points, line integration in vector fields (Benger, Acharya, WSCG'09)**
  - Interactive Exploration Of Coastal Restoration Modeling In Virtual Environments. (EVR'09)
  - Virtual Hydrology Observatory: Immersive Visualization Of Hydrology Modeling (EVR'09)
  - Interactive Tracking of 3D Critical Points in Unsteady, Large-scale CFD (CGVR'08)
  - Selected for SCALE09 competition for large scale viz (Shanghai, May 09) (LSU)
  - Initial network model of LONI Southern Ring (ULL)
  - **Integration with VISH + Tangibles in context of several SD applications**
- Funding
  - Over three proposals submitted (NSF STCI, MRI, CPS)
- Planning and Development
  - Jindal/Park abstract
  - CAVE environment at SUBR
  - Two successful LONI/LSU networking meetings held.

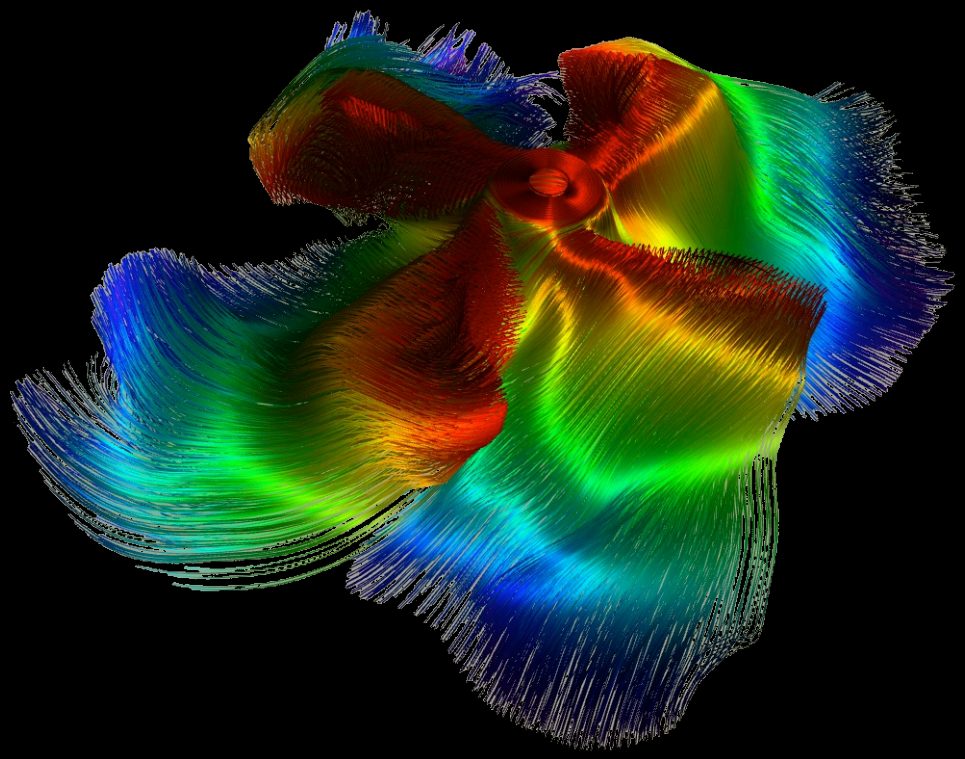
Louisiana EPSCoR

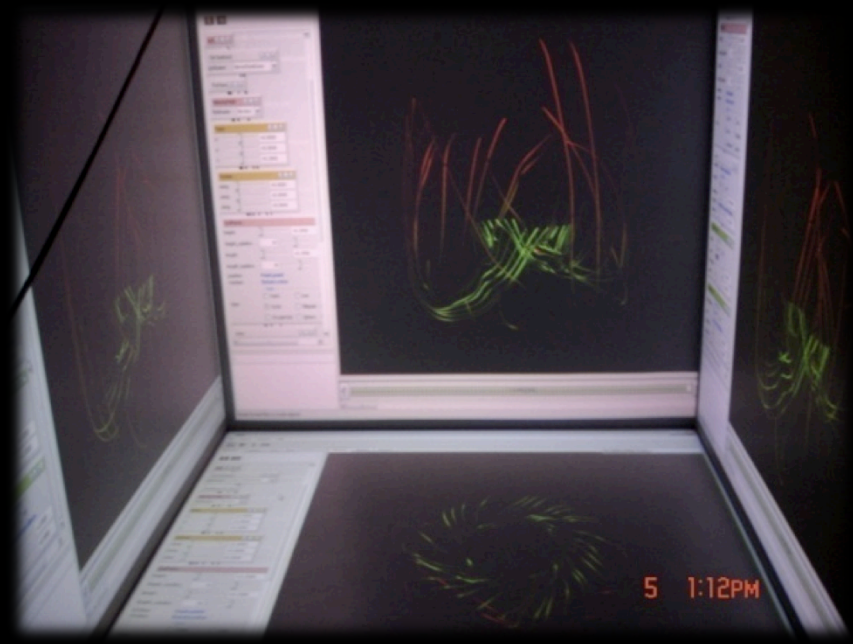
WP3: Science Driver Interactions

- Biotransport + Small Molecules
  - Visualization of pathlines of particles in fluid flow (Acharya)
  - CAVE visualization of fluid flows (Acharya, Nikotpoulos)
  - Tangibles integration via VISH
- Environmental Transport
  - ULL hydrology faculty in hydrology (Habib, Meselhe)
- Other:
  - Complex data from computational fluid dynamics **in a general framework, exploring new capabilities beyond existing ones** (Acharya, Nikitopoulos)
  - Schnetter et al, distributed visualization pipeline and common file format

Louisiana EPSCoR






5 1:12PM

WP4 Major Achievements

- Design of CFD-MD coupling scheme
- New design of CFD Toolkit
- NSF proposal “Community infrastructure for general relativistic MHD” (15 sites)
- NSF proposal “High Performance Computing Application Coupling Infrastructure” (PI: Erik, co-PI: Mayank)
- Prototype unstructured mesh toolkit (ICCS’09)
- Invited to take part in NSF Software Sustainability workshop (TR)

Louisiana EPSCoR

WP4 Major Achievements

- **Design of CFD-MD coupling scheme**
- **New design of CFD Toolkit**
- NSF proposal “Community infrastructure for general relativistic MHD” (15 sites)
- NSF proposal “High Performance Computing Application Coupling Infrastructure” (PI: Erik, co-PI: Mayank)
- Prototype unstructured mesh toolkit (ICCS’09)
- Invited to take part in NSF Software Sustainability workshop (TR)

Louisiana EPSCoR

WP4 Science Driver Interactions

- Immunosensors:
  - MD/CFD interface with biosensor science driver (LSU)
  - Integrating Tulane BEM code into Cactus (LSU/SUBR/ Tulane)
- Biotransport:
  - Integrating multi-block incompressible CFD code with Cactus Framework (LSU)
  - Design of CFD Toolkit (LSU/Tulane)
- Environmental Transport
  - Coastal modeling toolkit
  - Reservoir simulation toolkit
  - Astrophysics

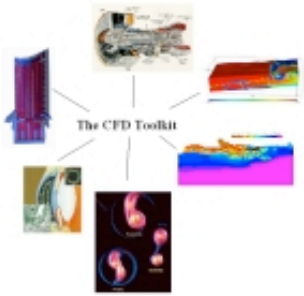
Louisiana EPSCoR

The CyberTools: LA Software

Data & Scheduling	Information & Portals	Visualization & Networks	Application Toolkits
PetaShare	AIMS	Core tangibles API	Cactus
petafs	ADAPT	Cross-viz platform API	Carpet
petashell	CLUSTER	EAVIV	CFD Toolkit
pcommands	P3MAP	F5 Library	SAGA
PetaShare Portal	PC4	InterViz API	SimFactory
HA-OSCAR	Black Hole Portal	Pathline module	
HARC	CFD Portal	Vish	

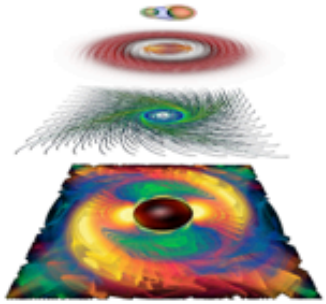
See http://www.cybertools.loni.org/sw.php

Louisiana EPSCoR



CFD Toolkit

The CFD Toolkit is an on-going research initiative at the Center for Computation & Technology (CCT), which is building a collaborative problem-solving environment for grand challenge fluid flow and transport problems. The CFD Toolkit is built using the Cactus framework and currently is developing modules to integrate a multiblock stirtank code with the Biotransport group and a boundary element method for complex flow with the Immunosensor group.



F5

F5 is a library that provides a simple C-API for writing HDF5 files using the concept of fiber bundles. F5 provides a common model to cover a wide range of data types used for scientific computing in a mathematically founded and systematic approach. The library includes file converters from common data output formats. In CyberTools the F5 file format is used as a common base for the varied data from the diverse science drivers, enabling applications to easily share data and visualization tools, and benefit from improvements in the I/O layers.

HARC

Highly-Available Resource Co-allocator. HARC is a system for reserving multiple resources in a coordinated fashion. HARC can handle multiple types of resource, and has been used to reserve time on supercomputers distributed across a US-wide testbed, together with dedicated 10 Gb/s lightpaths connecting the machines.

P3Maps

Protein Physico-chemical Property Maps. P3Maps provides an understanding of the sequence and physico-chemical relationship and paves the way for users to identify local sequence property modulations that impact protein function without changing the protein structure. It provides users a means to analyze multiple sequences (DNA or Protein) using existing tools of Clustal W and phylogenetic trees in high-performance computing environments. Its features provide users to analyze conserved domains using data mining techniques of prediction.

Louisiana
EPSCoR

Integrate & Leverage

Education

NSF CFD IGERT

NIH LBRN, others…

NSF ALPACA

DOE UCOMS

LONI INSTITUTE

Computer Science

Science

NSF Network MRI

NSF XiRel

DOD COMI

NSF TeraGrid

NSF VizTangibles MRI

NSF PetaShare MRI

NSF Blue Waters

Facilities/Support

Louisiana EPSCoR

Integrate & Leverage

Education

NSF CFD IGERT

NIH LBRN, others…

NSF ALPACA

DOE UCOMS

NSF PIF

LONI NSTITUTE

NSF PetaApps

Computer Science

Science

NSF Network

NSF STCI

DOD COMI

NSF Tera

NSF gibles RI

NSF STCI

NSF PetaShare MRI

NSF Blue Waters

Facilities/Support

Louisiana EPSCoR

Education and Training

- Actively tracking graduate students:
  - Progress in degree and graduate committee
  - Intersite visits
  - Publications, posters, conferences
  - Internships
  - Software tools
  - LONI accounts and training
- Interdisciplinary mentoring committees for all six postdocs

Louisiana EPSCoR

Research Modules (In Progress)

- Self contained tutorials, demos, services, presentations on every aspect of CyberTools for K-12 or undergraduate education
- Each faculty/staff responsible for one module, each module has consistent documentation.
- Examples:
  - K-12 Black Hole portal: produce waveforms on LONI
  - Vish Tutorial: Run Vish on PC to interact with various sample data sets
  - Cactus Tutorial: explore a live Cactus run and understand what we mean by a simulation

See http://www.cybertools.loni.org/rmodules.php

Louisiana EPSCoR

Research Modules

Protein Data Integration Tool Tutorial

Proteins are flexible macromolecules that contain backbones which can change from one specific folded conformation to another. Protein folding is frequently guided by local residue interactions that form clusters in the protein core. Evidence indicates that functionally important protein flexible residue interactions are governed by the hydrophobic propensities that they possess. This tool is designed to predict physico-chemical property-flexibility relationships that have been experimentally confirmed as functionally important. This tool leads students to apply graph theory and data mining technologies to extract and isolate protein structural features that sustain invariance in evolutionary related proteins, through the integrated analysis of different hydrophobicity scales over the 3D structure of proteins.

- Online Demo: Connect
- Tutorial (PDF)
- Hands-on exercises
- Software
- Level: Undegraduate

Computational Fluid Dynamics Toolkit Tutorial

Designing an efficient artificial heart requires understanding the blood flow in human heart just like improving the efficiency of a race car or Boeing 777 requires understanding their aerodynamics. With the advent of high performance computational platforms, computational fluid dynamics approach has made significant advances in the wide areas of engineering and sciences. This tutorial introduces the students with simple fluid dynamics concepts such as drag force and mixing by interactively changing the geometry and placement of obstacles (alphabet letters) in the flow path and visualizing these results. Several other examples such as pollutant dispersion in a model city are also used to explain the importance of such a simulation tool.

- Online Demo: Connect
- Tutorial (PDF)
- Software
- Level: K-12

Louisiana EPSCoR

Anchor Presentations

- Genosensor and Small Molecule Corner:
  - MD/CFD Coupling (Nayong)
  - Tangible interaction & Viz (Brygg, Kexi)
- Biotransport Corner:
  - CFD Portal and Workflow (Archit, Bharad)
  - CFD Toolkit (Jeff)
  - VISH Visualization (Werner, Farid)
- Computational Science Corner:
  - PetaShare (Ismail)
  - Data Mining (Sumeet Dua, Pradeep)
  - Cactus Tools and Visualization (Erik, Oleg)

Louisiana
EPSCoR