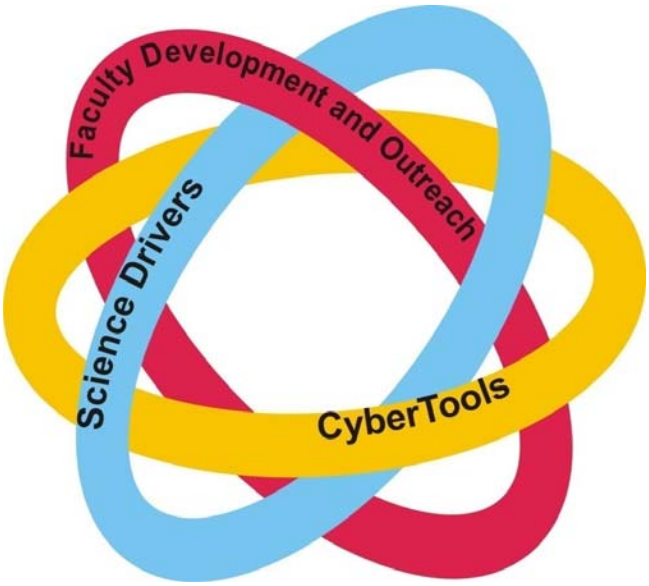
Louisiana Cyberinfrastructure and Science Drivers Symposium



Friday, August 22, 2008

Thomas Jefferson Room, 1-136A, Claiborne Building
1201 N. Third Street, Baton Rouge

Morning Agenda

8:00 – 8:50 a.m.	Coffee & Donuts/Poster setup
8:50 – 9:10 a.m.	Welcoming Remarks & Introductions
9:10 – 9:30 a.m	Overview of Project (Khonsari)

Science Driver (SD) Presentations (including time for Q&A)

9:30 – 10:00 a.m.	Geno/ Small Molecule Sensors (Soper, Carroll, Chen, Emory, Moldovan, Nikitopoulos)
10:00 – 10:30 a.m.	Immunosensors (Cortez, Ali, Bishop, Hamlington, Pillert, Agarwal)
10:30 – 10:50 a.m.	Break/Networking
10:50 – 11:10 a.m.	Biotransport Computations (Acharya, Moldovan)
11:10 – 12:10 p.m.	Poster Presentations
12:10 – 1:00 p.m.	Working Lunch
1:00 – 1:10 p.m.	Announcement of Poster Competition Award(s)

Afternoon Agenda

CyberTools WorkPackage (WP) Presentations and Demonstrations (including time for Q&A)

1:10 – 1:30 p.m.	Overview (Seidel)
1:30 – 2:10 p.m.	Applications and Application Toolkits, WP4 demonstrations (Jha, Kim, Schnetter, Tyagi) *Demo 1: Fluid Flow using Multipatch Solvers in Cactus* *Demo 2: Visualizing the BEM Code* *Demo 3: Distributed Replica-Exchange Using SAGA*
2:10 – 2:40 p.m.	Scheduling and Data Services, WP1 demonstrations (Kosar, Bishop, Bahsi, Wang, Dua, Brener) *Demo 1: End-to-end Workflow Management* *Demo 2: Distributed Data and Retrieval*
2:40 – 3:10 p.m.	Visualization, WP3 demonstrations (Ullmer, Hutanu, Amatya, Ge, Iyengar) *Demo 1: Remote Visualization*
3:10 – 3:50 p.m.	Break
3:50 – 4:15 p.m.	Outreach (Soper, Cortez, Allen, Ullmer)
4:15 – 4:30 p.m.	Evaluation/Assessment (Ramsey, Magee-Brown)
4:30 – 5:00 p.m.	Final Discussion (ERB)

External Review Board Members

Mr. James Hoehn, Executive Director, EPSCoR/IDeA Foundation, Washington, DC

Dr. Ron Hutchins, Associate Vice-Provost for Research and Technology and Chief Technology Officer, Georgia Institute of Technology

Dr. Vince McKoy, Professor of Chemistry at the California Institute of Technology

Dr. Harold Silverman, Senior Vice Provost, State University of New York System

Dr. Valerie Taylor, Professor and Department Head, Computer Science, Texas A&M University.

Graduate Student Posters

Louisiana State University

João Abecasis "Enabling Distributed Applications with SAGA"

Raghava Alapati "Application of Molecular Dynamic Simulation for Transport across Cellular Interfaces"

Vinay C. Amatya "Distributed Visualization"

Alborz Amirsadeghi "Simulation and experimental study on demolding for nanoimprint lithography"

Emir Mahmut Bahsi "Workflow enabling Large-scale scientific applications"

Mehmet Balman "Enhancements in Stork Data Placement Scheduler"

Hua Cao, Asim Shrestha, Rathika Natarajan, Gaurav Khanduja, Dipesh Bhattarai "Data Fusion"

Promita Chakraborty "Design and Performance Analysis of a Distributed HPC Molecular Dynamics Code on Distributed Resources"

Pin-Chuan Chen "Toward a Modular High Throughput CFPCR Array from a Single Nanoliter CFPCR"

Junseo Choi "Low Cost Fabrication of Micro- and Nanopores in Fress-Standing Polymer Membranes for Study Lipid Adsorption"

Thomas Dufaud "Hybrid Coupled Continuum-Molecular Dynamics Simulation Tool for the CFDToolkit"

Anvar Gilmanov "A Novel Flow-Structure Interaction Methodology for Biological Systems"

Tim Gilmanov, Anvar Gilmanov "Development and Application of a Material Point Method for Structure Calculations in Biological Systems"

Prasad Kalghatgi "Development of CFD Modules for CFDTOOLKIT"

Namwon Kim "Multi-Phase Flow in Polymer Microfluidic Systems"

Oleg Korobkin "Coupling an Einstein and an Euler code via the Cactus framework"

Archit Kulshrestha, Harsha Bhagawata "Towards Cyber Infrastructure for Dynamic Storm Surge Predictions"

Benoit Laveau “Experimental Study of droplet motion on a ratcheted surface”

Tae Yoon Lee “A polymer modular system for mutation detection”

C. Nancy Lekpeli “Transport of Molecular Clusters through Nano-scale Channels”

Samuel Njoroge “An Automated Genosensor System Using Modular Microfluidics “

Paul Okagbare “Small Molecule Sensor: HTS for Drug Discovery”

Daniel Park “Design and fabrication of small continuous flow PCR devices for a multi-well CFPCR platform"

Celine Ramet “Design Optimization and Realization of an Electrophoretron Cycler”

Sudheer D. Rani “Numerical Simulations of Misalignment Effects in Pressure Driven Flows for Micro-Fluidic Interconnects”

Asim Shrestha, Dimple Juneja “Data Mining”

Katerina Stamou “Real-time Information Services for Scientific Applications”

Ibrahim H Suslu, Mehmet Balman, Ismail Akturk, Xingqi Wang “Distributed Data Management in CyberTools”

Cornelius Toole, Jr. “LIGO Outreach Tangibles: Integration of Tangible Interaction and Visual Computing as Gateways to Science and Cyberinfrastructure”

Eamonn D. Walker “Numerical Simulations of Micro-Scale Segmented Two-Phase Flows for Bio-Analytical Chip Applications”

Esma Yildirim, Dengpan Yin “Predicting Optimal Level of Parallelism in Wide Area Data Transfers”

Byoung Hee You “Assembly tolerance for injection molded modular, polymer microfluidic devices”

Louisiana Tech University

Pradeep Chowriappa “An Algorithmic Tool for Protein Structure Classification based on Conserved Hydrophobic Residues”

Senaka Kanakamedala, Mangilal Agarwal “Evaluation of Microsensor and Micro-Mixer for Biosensor Applications”

Harpreet Singh “Medical Image Classification using Weighted Association Rules based Classifier”

Tulane University

Mehnaaz Ali, Amit Jain "Coupling Antibody Binding to Enzyme Activation in Miniaturized Immunosensor Devices"

Kate Hamlington, Jerina Pillert "Computational Model of a Microfluidic Mixing Chamber for Miniaturized Immunosensor Devices"

Symposium Attendees

Sumanta Acharya	Louisiana State University
Mangilal Agarwal	Louisiana Tech University
Gabrielle Allen	Louisiana State University
Tom Bishop	Tulane University
Nate Brener	Louisiana State University
Marion Carroll	Xavier University
Pin-Chuan Chen	Louisiana State University
David Claypool	Louisiana State University
Ricardo Cortez	Tulane University
Rachel Cruthirds	Louisiana Board of Regents
Carolina Cruz-Neira	University of Louisiana at Lafayette
Mark DeCoster	Louisiana Tech University
Thomas Dufaud	Louisiana State University
Sumeet Dua	Louisiana Tech University
Hideki Fujioka	Tulane University
Don Gaver	Tulane University
Jinghua Ge	Louisiana State University
Jim Gershey	Louisiana Board of Regents
Anvar Gilmanov	Louisiana State University
Raju Gottumukkala	University of Louisiana at Lafayette
Cindy Greer	Louisiana Board of Regents
Les Guice	Louisiana Tech University
Jim Hoehn	EPSCoR/IdEA Foundation
Ron Hutchins	Georgia Institute of Technology
S.S. Iyengar	Louisiana State University
Amitava Jana	Southern University
Susan Jernigan	Louisiana Board of Regents
Shantenu Jha	Louisiana State University
Ray Jindal	University of Louisiana at Lafayette
Daniel S. Katz	Louisiana State University
Joohyun Kim	Louisiana State University
Sanjay Kodiyalam	Louisiana State University
Tevfik Kosar	Louisiana State University
Michael Khonsari	Louisiana Board of Regents
Benoit Laveau	Louisiana State University
C. Nancy Lekpeli	Louisiana State University
Honggao Liu	Louisiana State University
Frank Löffler	Louisiana State University
Mary Jo McGee-Brown	Qualitative Research & Evaluation for Action, Inc.
Vince McKoy	California Institute of Technology
Marsha J. Miller	University of Louisiana at Lafayette
Dorel Moldovan	Louisiana State University

Michael Murphy	Lousiana State University
Jarek Nabrzyski	Louisiana State University
Daniel Park	Louisiana State University
Karthik Poobalasubramanian	Louisiana Board of Regents
Linda Ramsey	Louisiana Tech University
Bety Rodriguez-Milla	Louisiana State University
Erik Schnetter	Louisiana State University
Ed Seidel	Louisiana State University
Harold Silverman	The State University of New York System
Steve Soper	Louisiana State University
Balamurugan Subramanian	Louisiana State University
Jennifer Tate	Louisiana State University
Valerie Taylor	Texas A&M University
Hilary Thompson	LSU Health Sciences Center
Isaac Traxler	Louisiana State University
Kathy Traxler	Louisiana State University
Mayank Tyagi	Louisiana State University
Brygg Ullmer	Louisiana State University
Sam White	Lousiana State University
Maggie A. Witek	Louisiana State University
Shizhong Yang	Southern University
Denpan Yin	Louisiana State University
Byoung Hee You	Louisiana State University
Zhiyu Zhao	University of New Orleans

Graduate Student Attendees

João Abecasis	Louisiana State University
Ismail Akturk	Louisiana State University
Raghava Alapati	Louisiana State University
Mehnaaz Ali	Tulane University Health Sciences Center
Vinay C. Amatya	Louisiana State University
Alborz Amirsadeghi	Louisiana State University
Emir Mahmut Bahsi	Louisiana State University
Harsha Bhagawaty	Louisiana State University
Junseo Choi	Louisiana State University
Pradeep Chowriappa	Louisiana Tech University
Chris Clayton	Southern University
Swathi Laxmi Dubbaka	Louisiana State University
Tim Gilmanov	Louisiana State University
Kate Hamlington	Tulane University
Andrei Hutanu	Louisiana State University
Amit S. Jain	Tulane University
Lei Jiang	Louisiana State University
Dimple Juneja	Louisiana State University
Prasad Kalghatgi	Louisiana State University
Senaka Kanakamedala	Louisiana Tech University
Namwon Kim	Louisiana State University
Oleg Korobkin	Lousiana State University
Archit Kulshrestha	Louisiana State University
Tae Yoon Lee	Louisiana State University
Vignesh Natesan	University of Louisiana at Lafayette
Samuel Njoroge	Louisiana State University
Paul Okagbare	Louisiana State University
Taehyun Park	Louisiana State University
Jerina Pillert	Tulane University
Emma Pineda	Tulane University
Celine Ramet	Louisiana State University
Sudheer D. Rani	Louisiana State University
Harpreet Singh	Louisiana Tech University
Nikhil Shetty	University of Louisiana at Lafayette
Katerina Stamou	Louisiana State University
Ibrahim H Suslu	Louisiana State University
Cornelius Toole, Jr.	Louisiana State University
Sirish Tummala	Louisiana State University
Eamonn D. Walker	Louisiana State University
Xinqi Wang	Louisiana State University
Esma Yildirim	Louisiana State University
Zhifeng Yun	Louisiana State University

Louisiana Cyberinfrastructure and Science Drivers Symposium

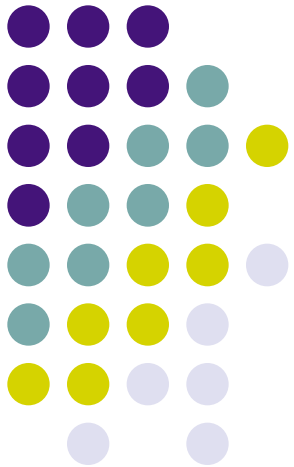
External Review Board Meeting

Michael Khonsari

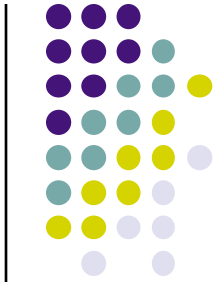
Louisiana EPSCoR Project Director

Associate Commissioner for Sponsored Programs Research and Development

Louisiana Board of Regents



EPSCoR External Review Board

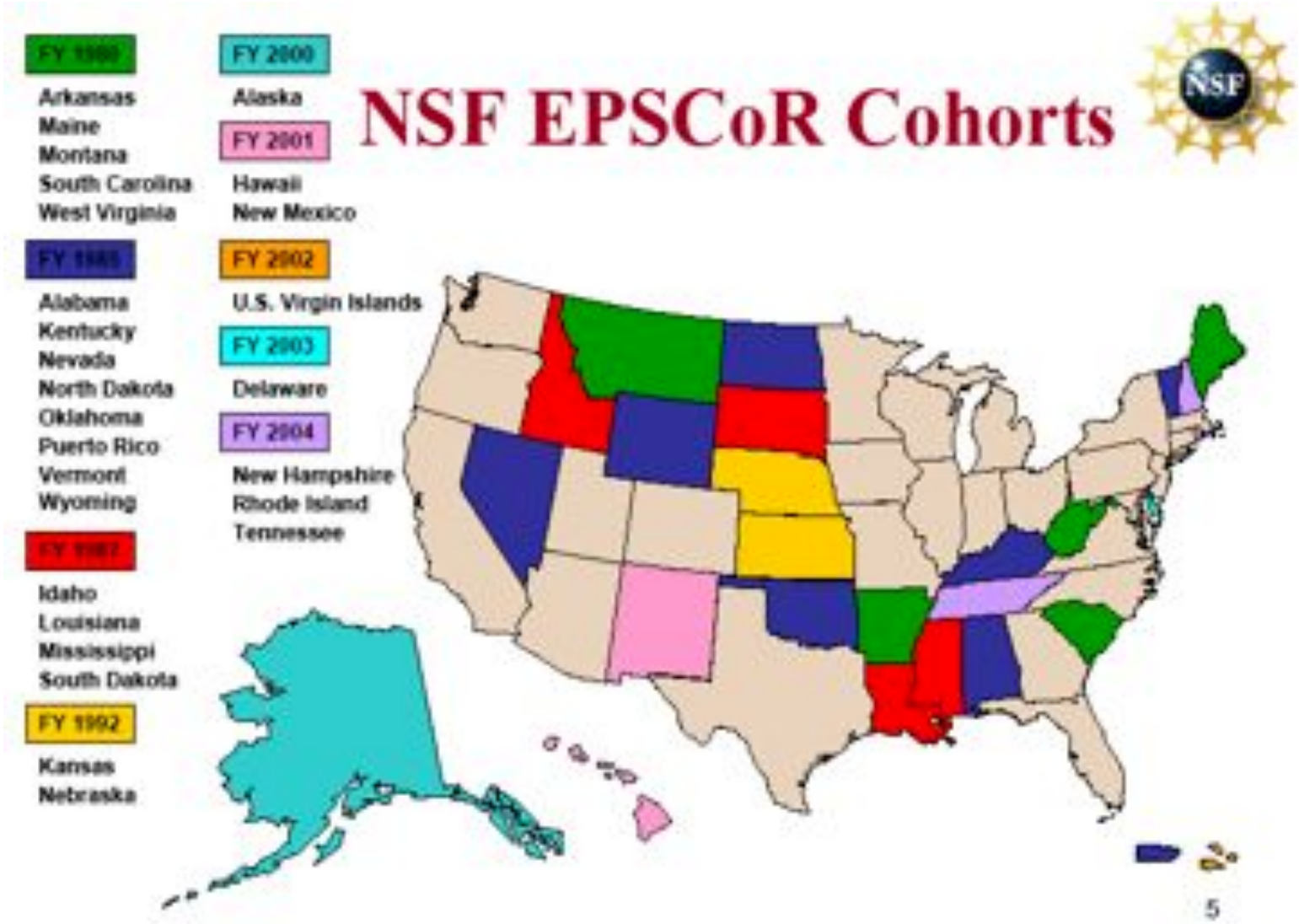


- Mr. James Hoehn, Executive Director, EPSCoR/IDeA Foundation, Washington, DC
- Dr. Ron Hutchins, Associate Vice-Provost for Research and Technology and Chief Technology Officer, Georgia Institute of Technology
- Dr. Vince McKoy, Professor of Chemistry at the California Institute of Technology
- Dr. Harold Silverman, Senior Vice Provost, State University of New York System
- Dr. Valerie Taylor, Professor and Department Head, Computer Science, Texas A&M University.

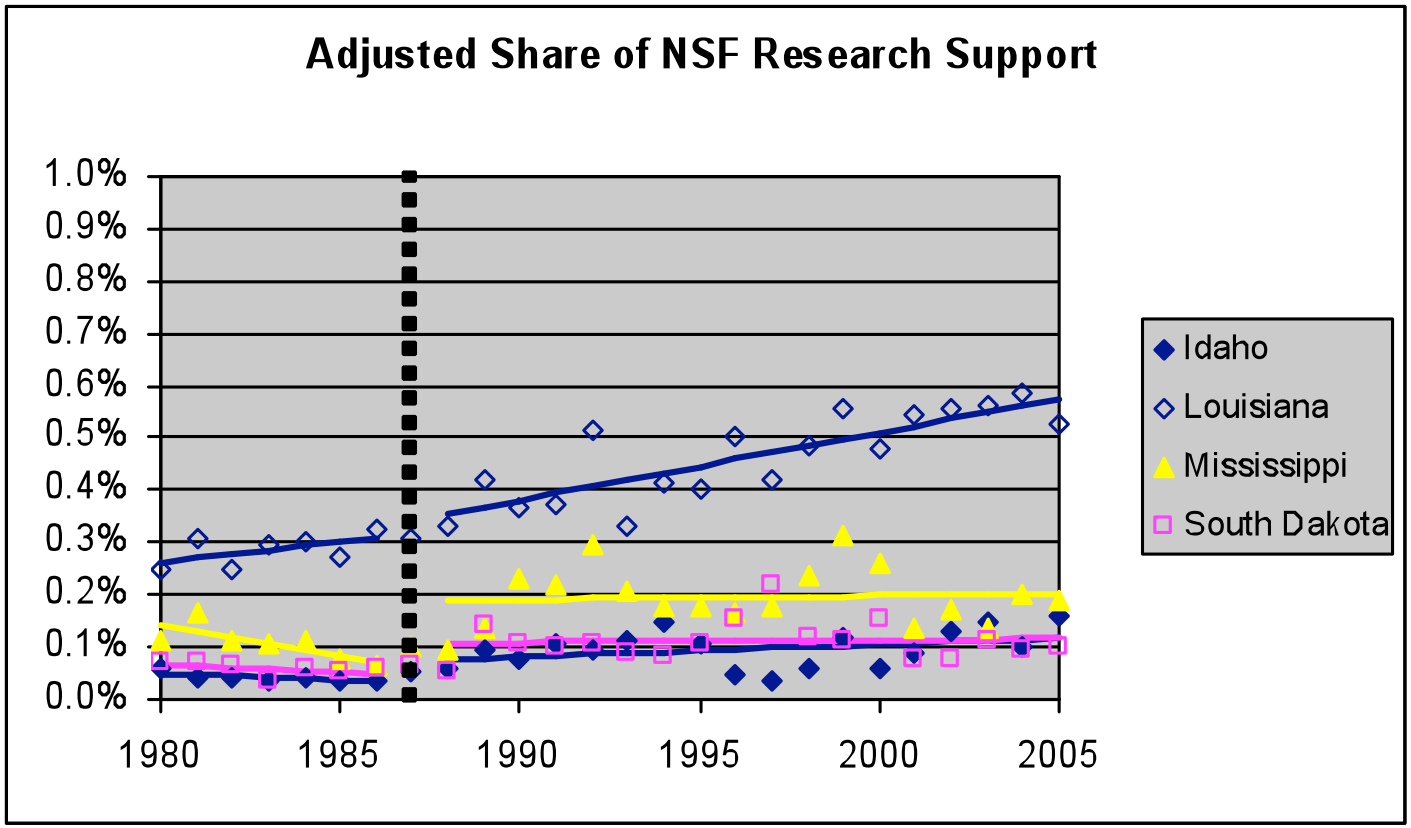
Louisiana EPSCoR

Louisiana EPSCoR is housed and integrated within the Louisiana Board of Regents, the coordinating body for public higher education in the State.

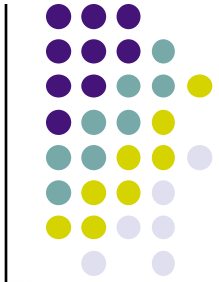
NSF EPSCoR Cohorts



Comparison to Cohorts



History of EPSCoR Funds Received



Agency	Federal	BoR Support Fund
NSF	$ 41,967,836	$21,142,036
NASA	8,745,236	7,811,560
NIH	90,856,739	0
DOE	8,551,388	6,639,590
DOD	8,737,013	2,639,949
EPA	1,023,649	994,542
DOC	250,000	300,000
TOTAL	$160,131,861	$39,527,677

NSF- EPSCoR Co-Funding

Co-Funding History for Louisiana since 2000

Year	Total	EPSCoR	# Awards
FY 00	$ 2,204,762	$ 1,059,992	9
FY 01	4,330,652	2,128,636	13
FY 02	5,823,318	2,922,531	26
FY 03	8,822,162	2,995,274	18
FY 04	14,186,798	5,961,313	28
FY 05	8,134,360	2,740,529	24
FY 06	8,396,096	3,058,041	26
FY 07	10,363,434	4,691,849	16
Total	**$ 62,261,582**	**$ 25,558,165**	**160**

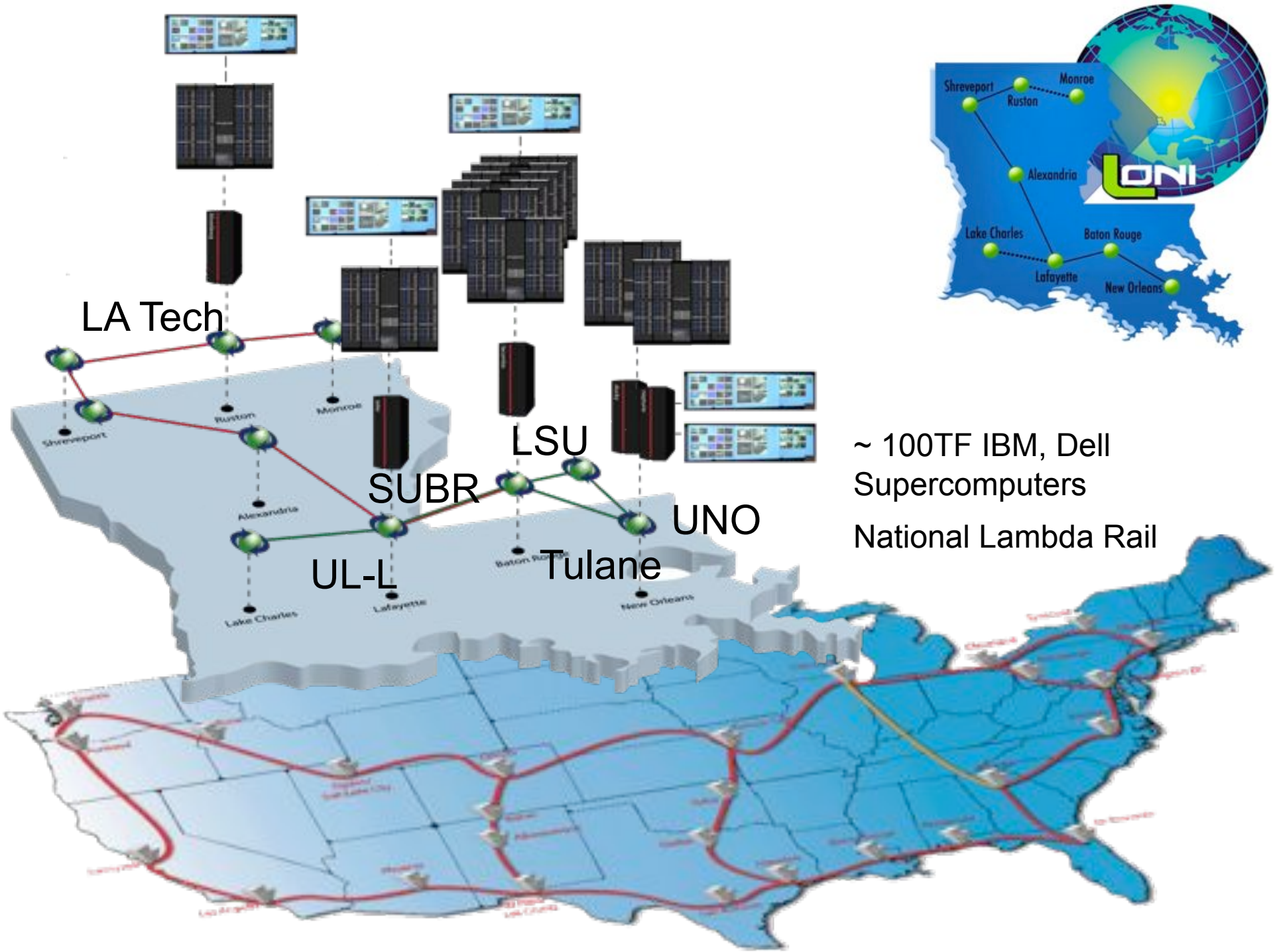
State's Strategic Investments

- *Louisiana: Vision 2020*
- IT drives Economic development:
  - Gov. Foster *Vision 2020* initiative committed $25M annually to five campuses in 2001
    - LSU's Center for Computation and Technology (CCT)
    - ULL's Louisiana Immersive Technologies Enterprise (LITE)
  - Gov. Blanco committed $40M over ten years to create Louisiana Optical Network Initiative (LONI) in 2004

Louisiana: Vision 2020

Louisiana Optical Network Initiative (LONI)

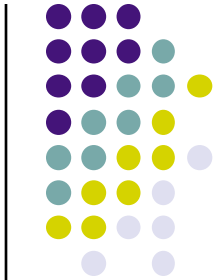
- LONI connects Louisiana to the National LambdaRail, a high-powered nationwide fiber-optic network.
- LONI connects the State's major research institutions with optical fiber delivering up to 40 Gigabits per second.



~ 100TF IBM, Dell Supercomputers

National Lambda Rail

Catalyst for Collaboration



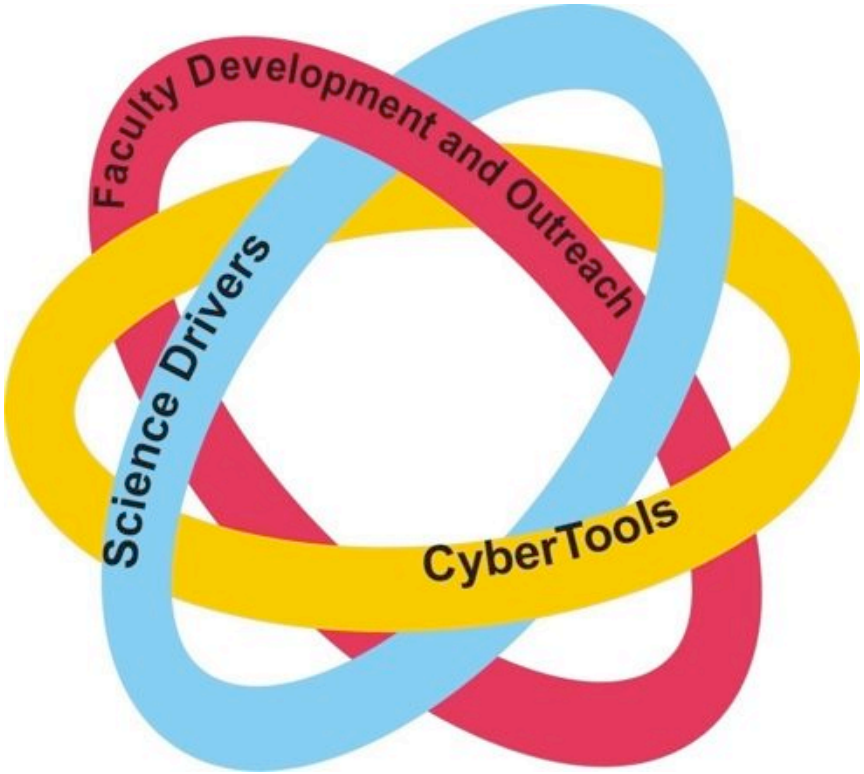
- Prior to EPSCoR, Louisiana institutions competed against each other for research funding.
- LA EPSCoR has been a catalyst in achieving increased statewide collaboration and national competitiveness.
- LA EPSCoR broke down boundaries among campuses.

RII 2007-2010
Research Theme

The focus of the Louisiana EPSCoR project is the development of a multi-functional cyberinfrastructure to broadly enable significant advances in modern science and engineering.

Louisiana's Research Infrastructure Improvement Strategy (RII 2007-2010)

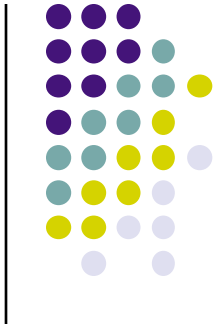
- This is by far the most comprehensive proposal LA EPSCoR submitted



Institutions Involved

- Louisiana State University
- Tulane University
- Louisiana State University Health Sciences Center
- Louisiana Tech University
- Tulane University Health Sciences Center
- Southern University
- University of Louisiana at Lafayette
- University of New Orleans
- Xavier University

Proposal Development



- Review Panel Selected
  - Program Director (San Diego Supercomputer Center)
  - Co-Director (Pittsburgh Supercomputing Center)
  - Associate Vice Provost for Research & Technology and Chief Technical Officer (Georgia Tech)
  - Emeritus Vice President for Strategic Initiatives and former Director, ERC for Computational Field Simulations (Mississippi State University)
  - Vice Chancellor for Research (North Carolina A&T) and former Director of the Computational and Information Sciences Directorate (CISD) with the Army Research Laboratory

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

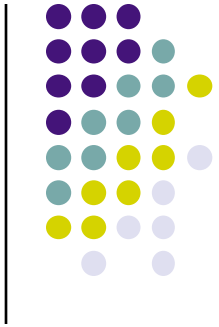
PITTSBURGH SUPERCOMPUTING CENTER

Georgia Institute of Technology

Mississippi State University

- Proposals reviewed, ranked, and top four were interviewed

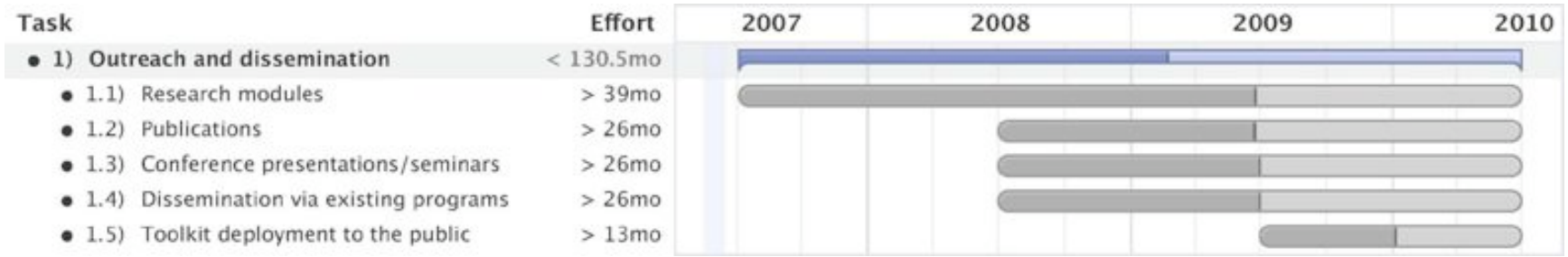
LA EPSCoR RII Proposal



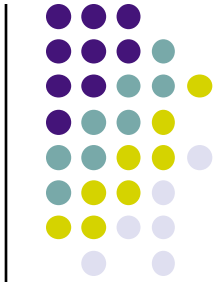
- Submitted October 2006
- Negotiation process between Team Leaders, BoR, and NSF took nearly a year
- Several hundred + page responses
- Visit by NSF Program Directors

LA EPSCoR RII Proposal

- Strategic Plan developed including:
  - Milestones & Deliverables
  - Gantt Charts
  - Metrics
  - Organizational Charts
  - Management Plan
  - Budget



EPSCoR External Review Board



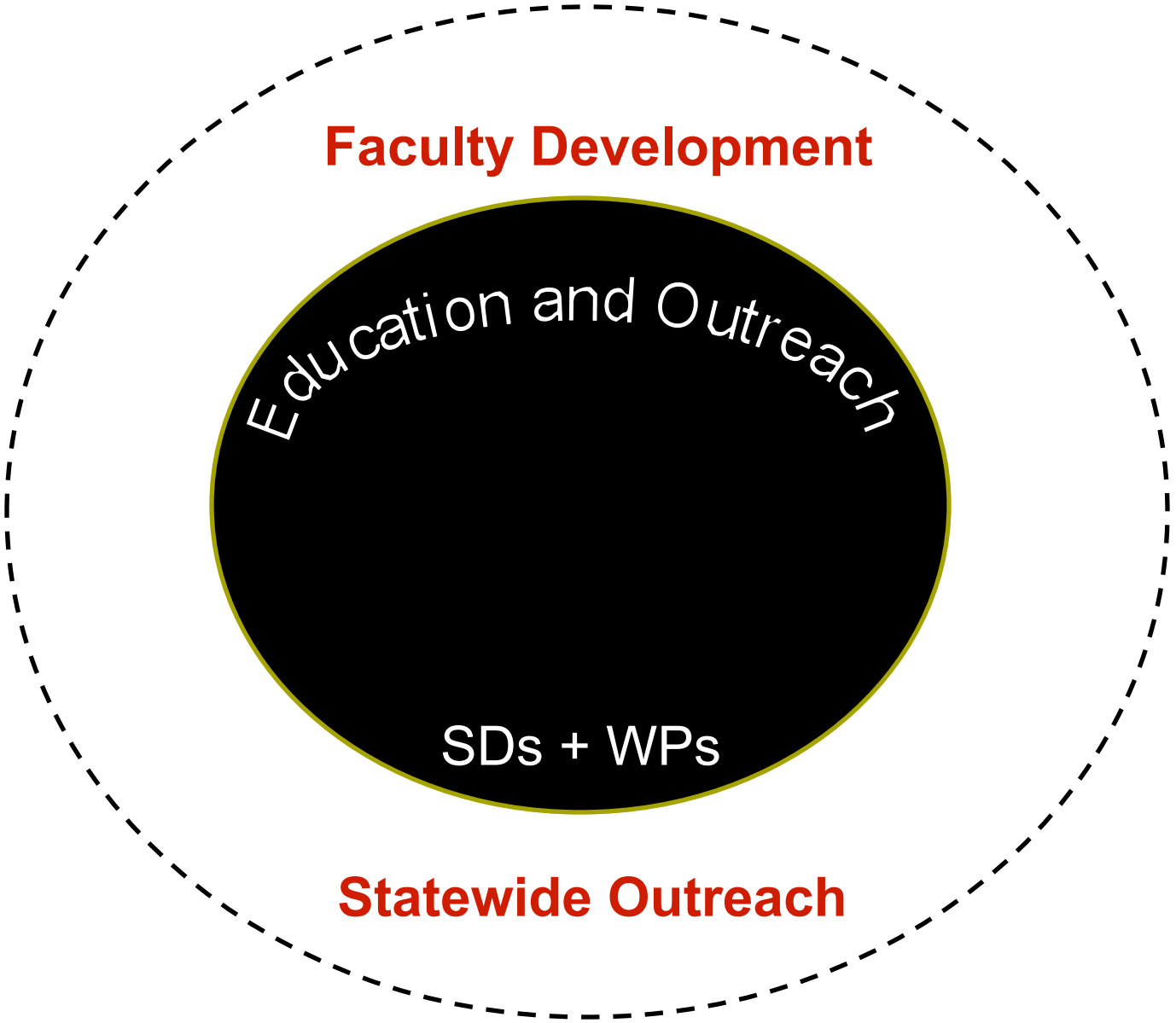
- Mr. James Hoehn, Executive Director, EPSCoR/IDeA Foundation, Washington, DC
- Dr. Ron Hutchins, Associate Vice-Provost for Research and Technology and Chief Technology Officer, Georgia Institute of Technology
- Dr. Vince McKoy, Professor of Chemistry at the California Institute of Technology
- Dr. Harold Silverman, Senior Vice Provost, State University of New York System
- Dr. Valerie Taylor, Professor and Department Head, Computer Science, Texas A&M University.

Cooperative Agreement

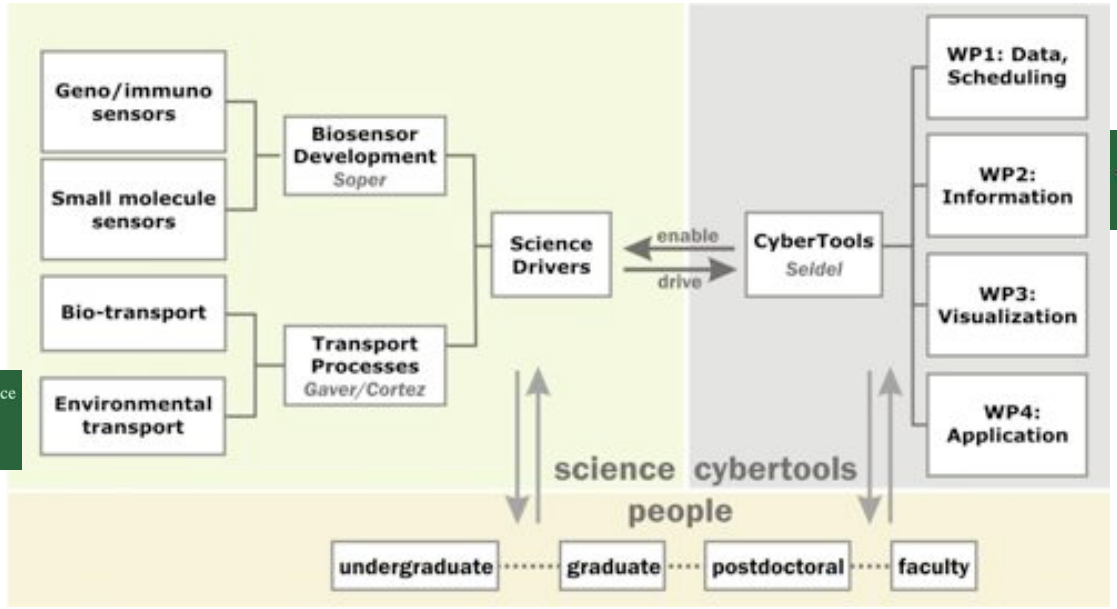
- Issued in October 2007
- Special Requirements include:
  - Evaluation & Assessment Plan

  (submitted January 2008)
  - Six-month Interim Reports (first project and expenditure reports submitted March 2008)
- Annual Report submitted July 2008

All reports and plans have been reviewed and approved by NSF


Faculty Development
Education and Outreach
SDs + WPs
Statewide Outreach

Management Structure until September 1, 2008

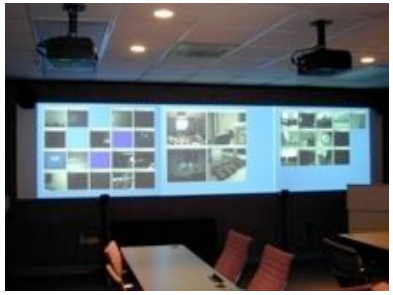


Meetings and Coordination

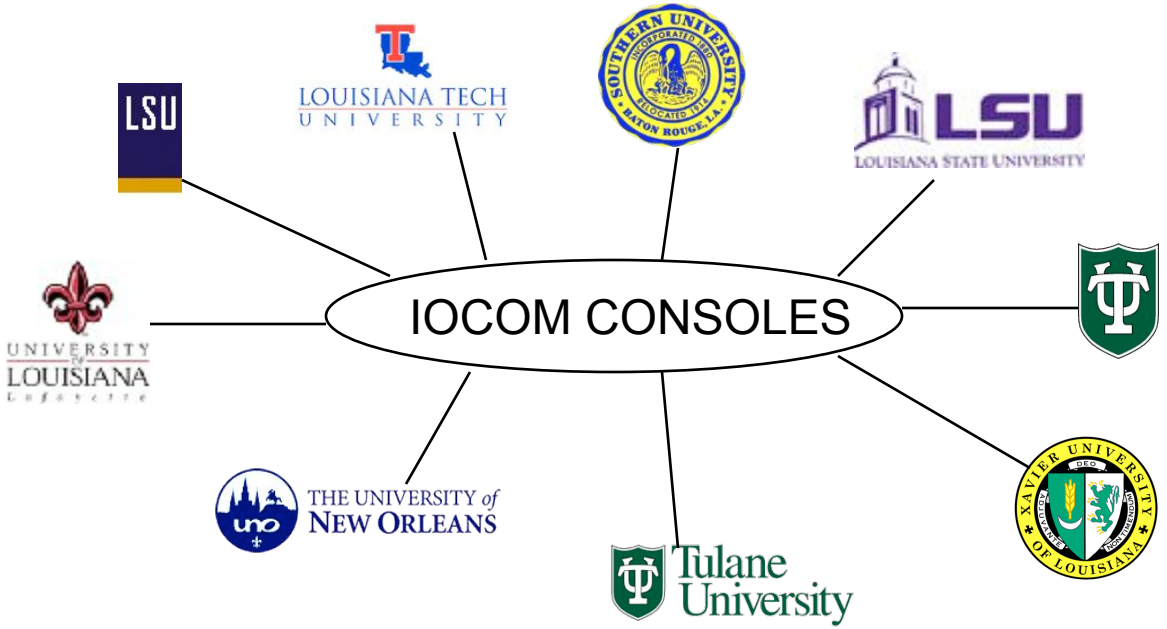
- Kick-off Meeting – October 18, 2007 (included State EPSCoR Committee meeting & External Review Board meeting)
- Six Science Executive Committee (SEC) meetings
- Two All Hands Meetings held at Louisiana State University and the Louisiana Board of Regents.
- All Hands Meeting for project review, including a poster competition – August 22-23, 2008 (attended by the ERB)

Facilitating Collaboration

- IOCOM System
  - Consoles
  - Desktop licenses, cameras, microphones,etc.



Access Grid Compatible



Faculty Development Programs

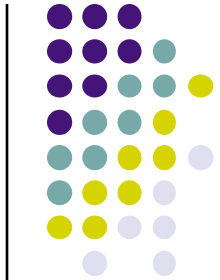
- Research development
  - Grant writing seminars
  - Travel grants for emerging faculty
  - Planning grants for major initiatives
  - SBIR/STTR phase zero grants
  - Pilot funding for new research
  - Links with industry, research centers, and national labs

- Outreach and Human Resource Development
  - State/regional conferences and seminars
  - Displays at State Capitol
  - Speaking of Science (SoS) speakers bureau
- Technology
  - Faculty expertise database and solicitation search
  - Online proposal submission and reporting

Education & Outreach

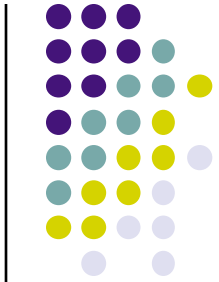
- The multi-functional *CyberTools* are being developed in association with the Science Driver projects, while the education and outreach activities are highly integrated into this process and into each Science Driver and *CyberTools* component.

Agenda



- Organization of the meeting
- Poster competition
- Saturday meeting

Agenda



8:00 – 8:50 a.m. Coffee & Donuts/Poster setup

8:50 – 9:10 a.m. Welcoming Remarks & Introd

9:10 – 9:30 a.m Overview of Project

Science Driver (SD) Presentations

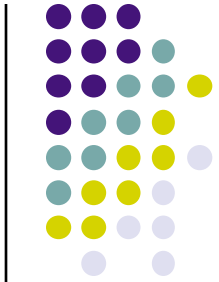
9:30 – 10:00 a.m. Geno/ Small Molecule Sensors

10:00 – 10:30 a.m. Immunosensors

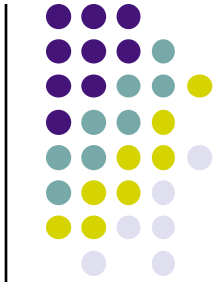
10:30 – 10:50 a.m. Break/Networking

10:50 – 11:10 a.m. Biotransport Computations

POSTER PRESENTATIONS



11:10 – 12:10 p.m.	Poster Presentations
12:10 – 1:00 p.m.	Working Lunch
1:00 – 1:10 p.m.	Poster Competition Award(s)



CyberTools WorkPackage (WP) Presentations and Demonstrations

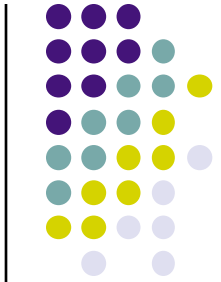
1:10 – 1:30 p.m. Overview

1:30 – 2:10 p.m. Applications and Application Toolkits, WP4 demonstrations

Demo 1: Fluid Flow using Multipatch Solvers in Cactus

Demo 2: Visualizing the BEM Code

Demo 3: Distributed Replica-Exchange Using SAGA



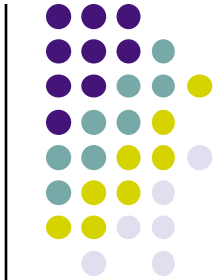
2:10 – 2:40 p.m. Scheduling and Data Services, WP1 demonstrations

Demo 1: End-to-end Workflow Management

Demo 2: Distributed Data and Retrieval

2:40 – 3:10 p.m. Visualization, WP3 demonstrations

Demo 1: Remote Visualization

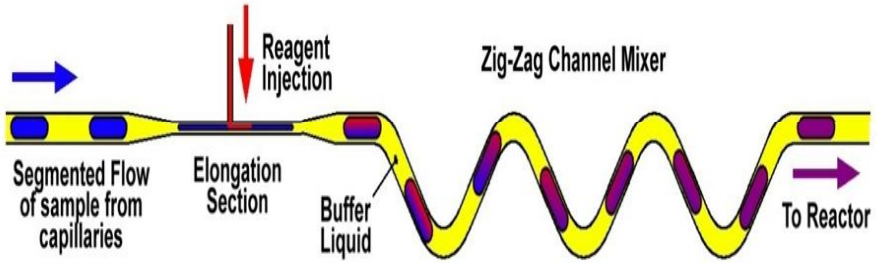


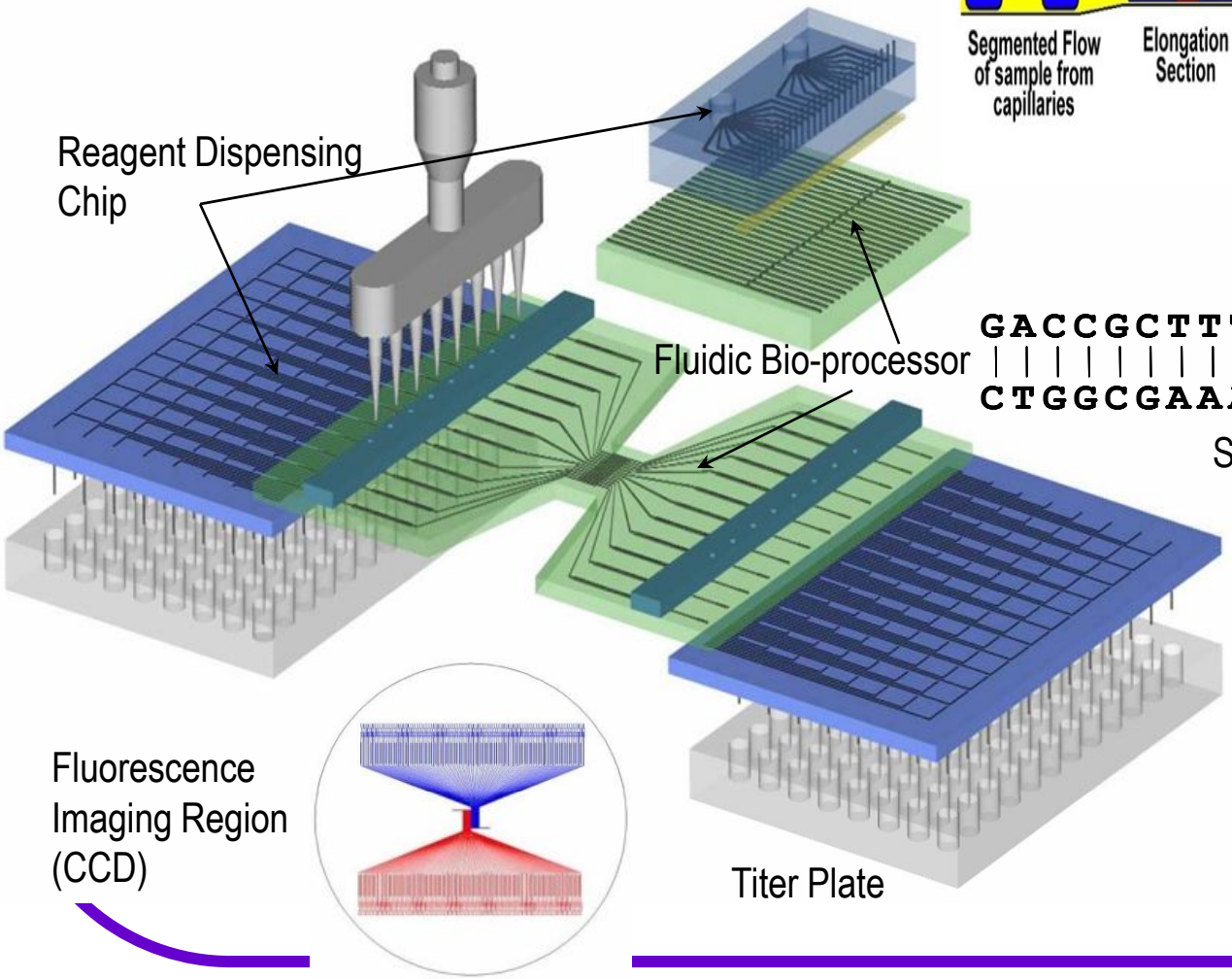
3:10 – 3:50 p.m.	Break
3:50 – 4:15 p.m.	Outreach
4:15 – 4:30 p.m.	Evaluation/Assessment
4:30 – 5:00 p.m.	Final Discussion (ERB)

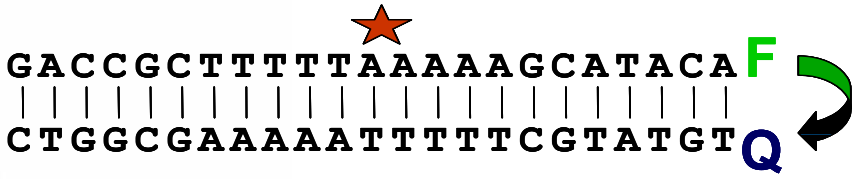
Posters by Okagbare, Kim, Rani, You, Shrestha/Juneja, Walker

Small Molecule Sensor (Steve Soper, LSU)

Screen combinatorial library for inhibitors of L1-EN – Genome instability







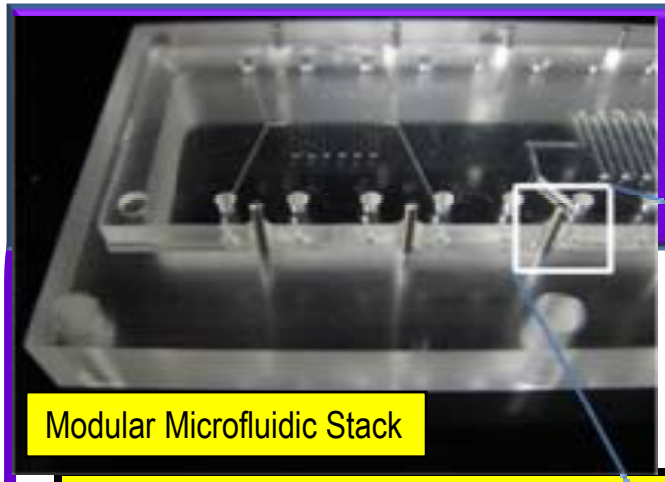
Substrate for L1-EN

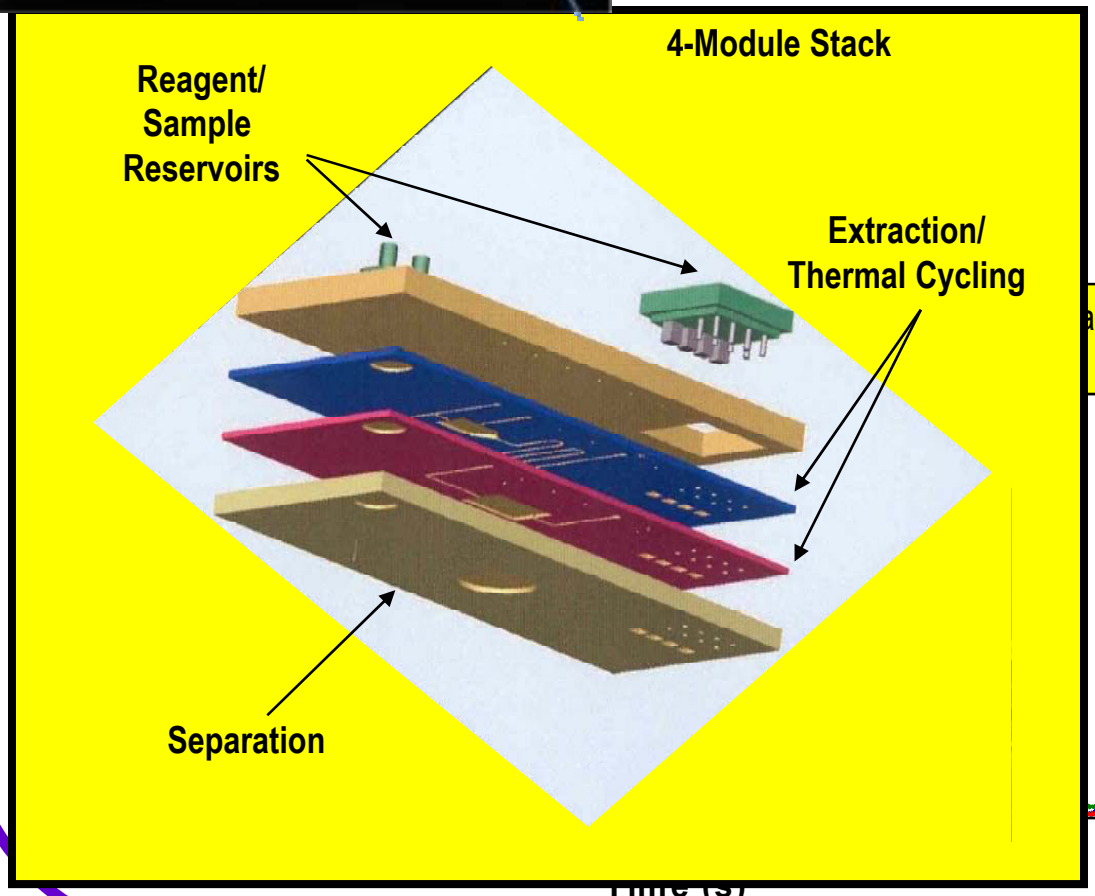
VALUE ADDED

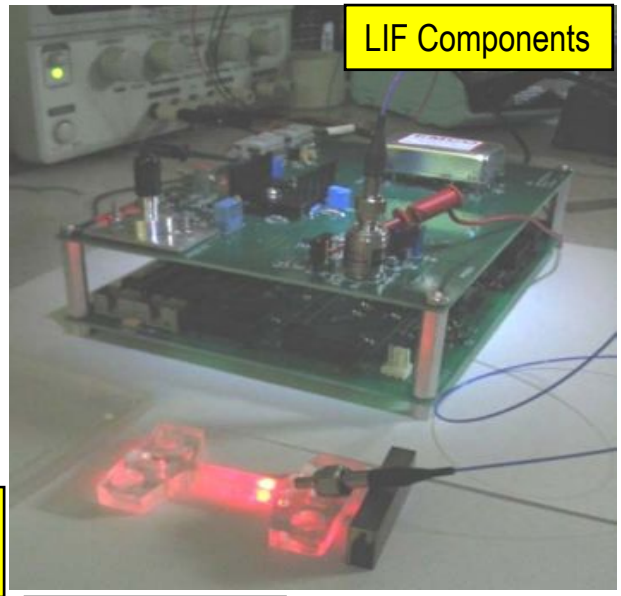
- Current state-of-the-art instruments (Evotec);
  - process 140,000 samples day^{-1}
  - Robotic fluid handling
  - Uses 1-5 µL of reagents

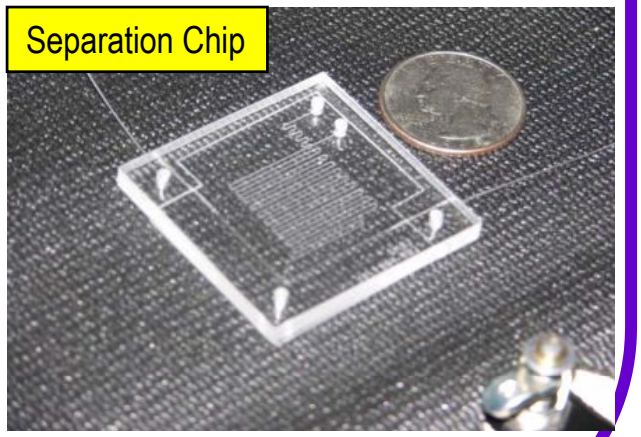
  Small Molecule Sensor System;
  - Process ~10^9 samples day^{-1}
  - Full automation affected by microfluidics
  - Imaging readout with high sensitivity
  - Uses 1-5 pL of reagents
- Interdisciplinary project (synthetic, analytical, material chemists; mechanical engineers; molecular biologists; pharmaceutical industry)
- Experimental chemists/engineers become familiar with HPC (WP3, WP4)
- CHALLENGE – How to mine and organize the data generated (WP1)

Genosensor System (Steve Soper, LSU)









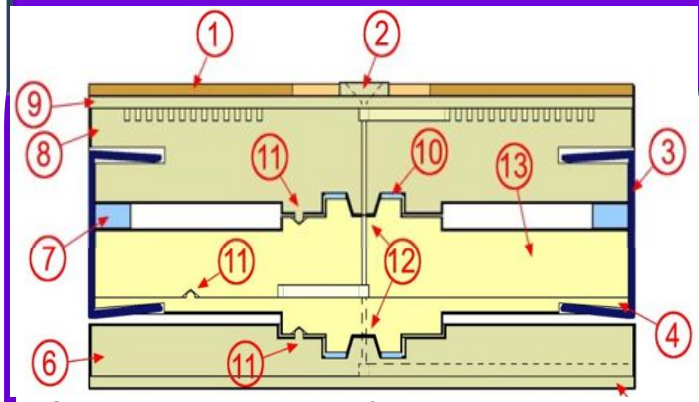
VALUE ADDED

- Current state-of-the-art instruments (ABI);
  - Multiple instruments for processing genetic samples
  - Large footprint and not field deployable
  - Requires specialized technicians to affect assay
  - Long assay turn-around time (6-8 h)

  Genosensor System;
  - Full automation affected by microfluidics and process integration
  - Short assay turn-around-time (30 min)
  - Field deployable without sacrificing assay performance
- Interdisciplinary project (synthetic, analytical, material chemists; mechanical engineers; molecular biologists; computer scientists)
- Reduce design/development time using system-level modeling
- CHALLENGE – Fabricate integrated system with multiple processing steps (WP3; WP4)

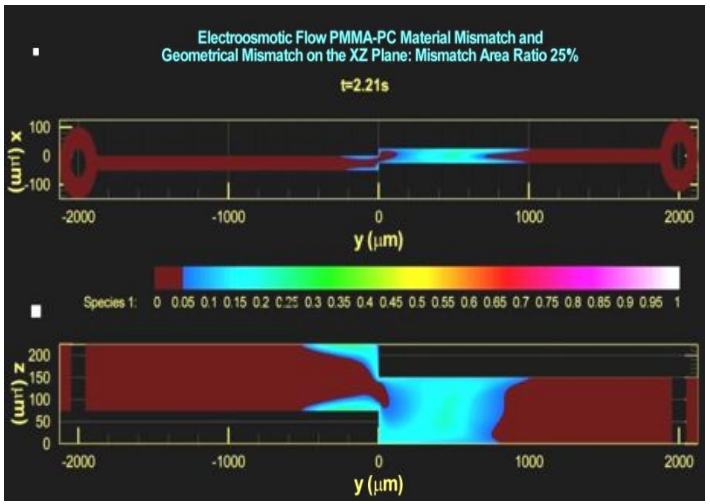
Posters by Chen, You, Njoroge, Rani, Park, Kalghatgi

Genosensor System (Steve Soper, LSU)

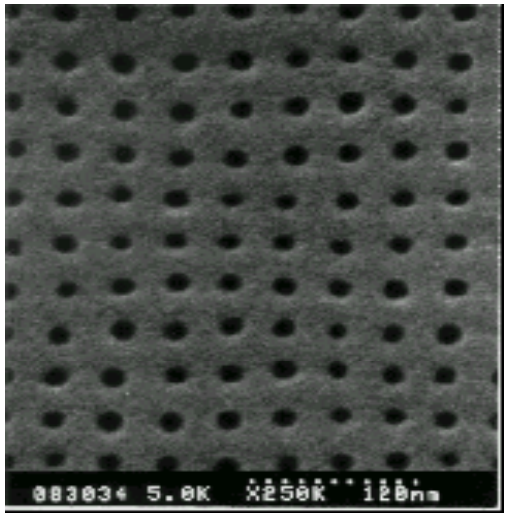


Conceptual drawing of super hydrophobic interconnect.

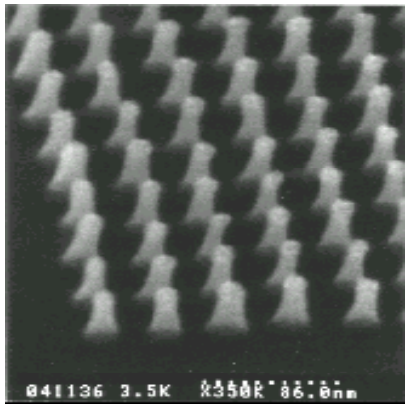
CFD simulation for module mis-alignment.



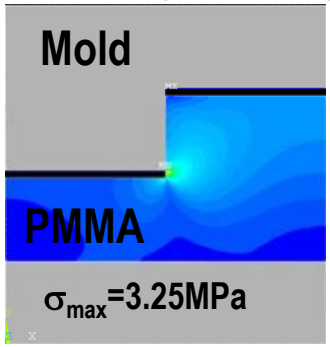
Interconnects: Designing modularity across different materials and scales.
Computational Needs – CFD simulations of Newtonian fluids across mixed-scale materials (nano-to-microchannel transport.

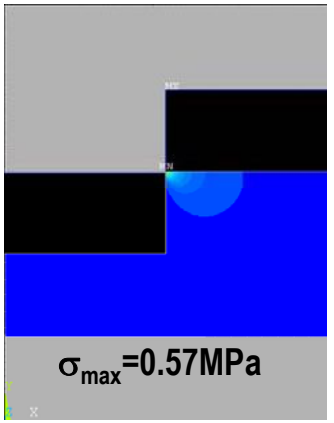


NIL



Stress during demolding



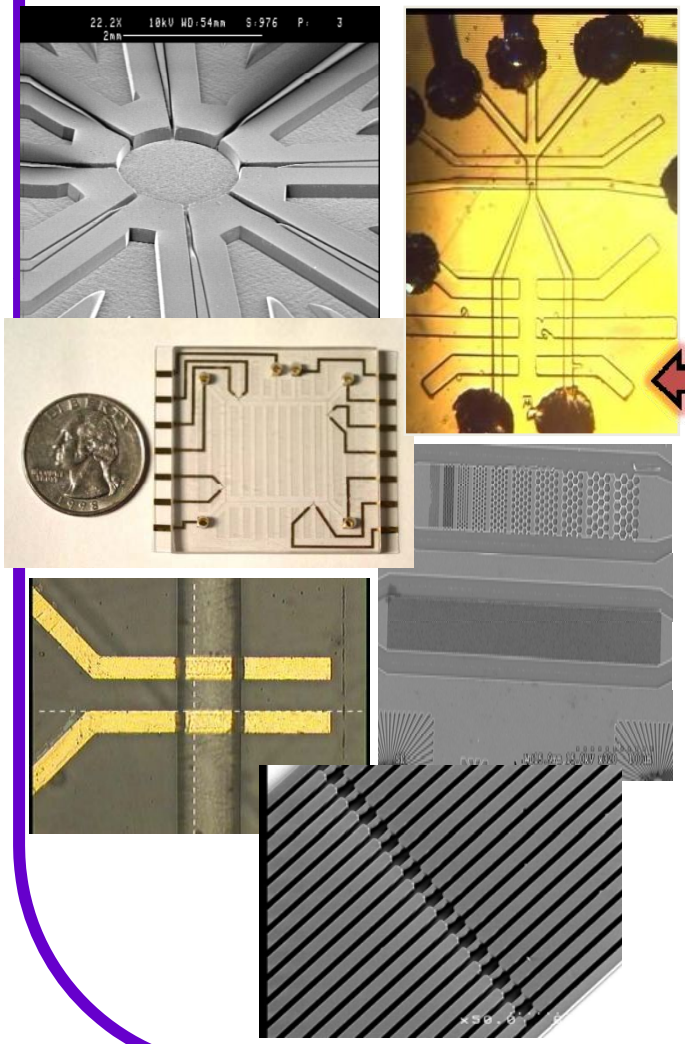


Nanofabrication : Nanoimprint lithography (NIL) to build nanostructure domains (extraction, extension).
Computational Needs – Modeling Non-Newtonian Fluids during mixed-scale replication.

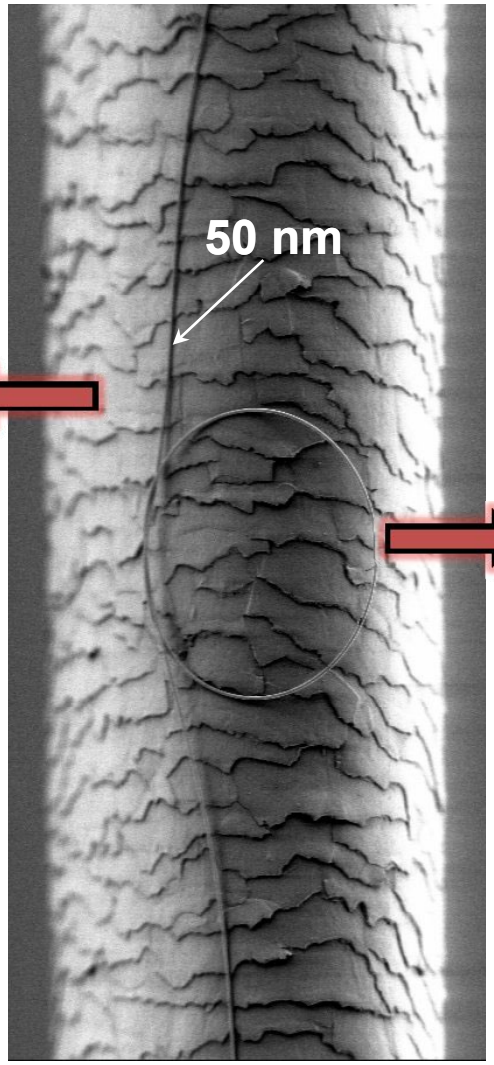
Posters by Dufaud, Lekpeli

Nanoscale Sensors: Rethinking the Molecular Processing Paradigm (Steve Soper, LSU)

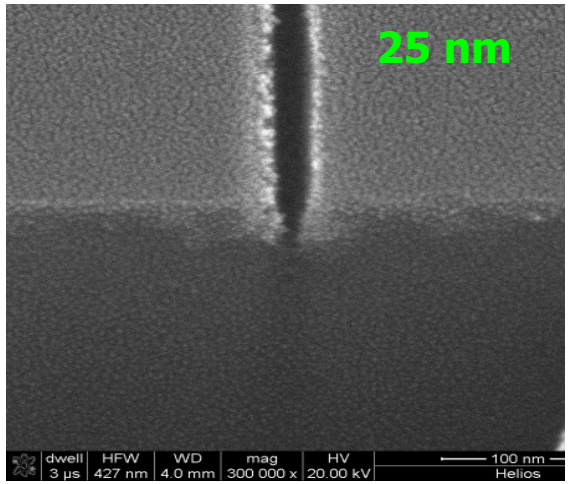
Microtechnology

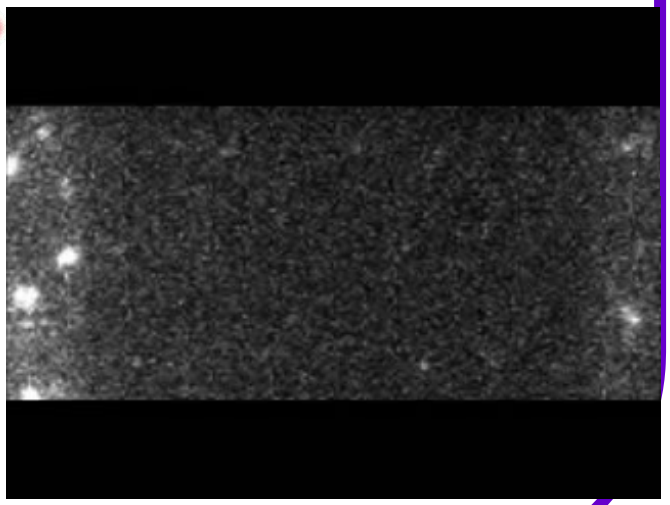


50 µm



Nanotechnology



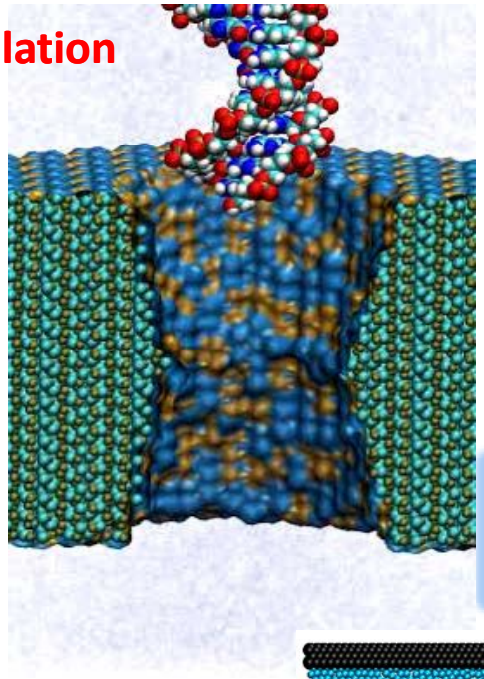


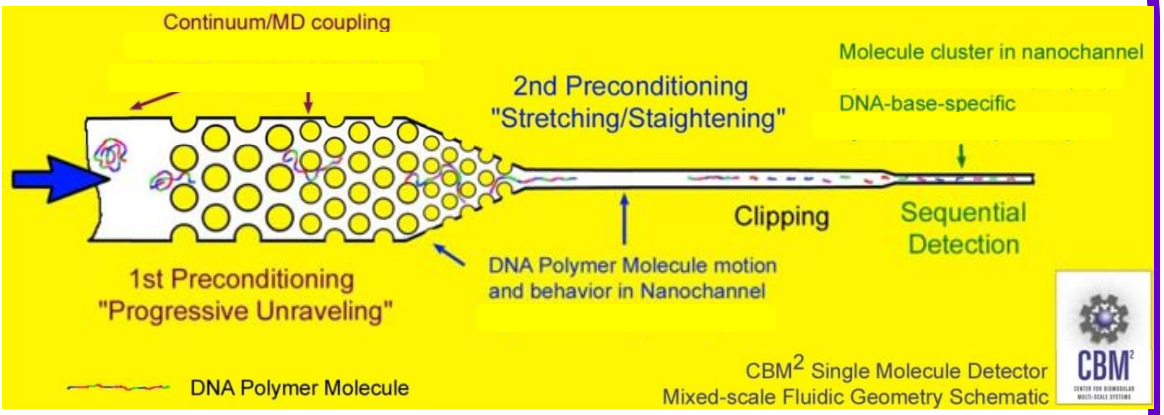
Menard and J. M. Ramsey, UNC

CyberTools Modeling of DNA Transport in Micro- / Nano-domains (Steve Soper, LSU)

MD Simulation

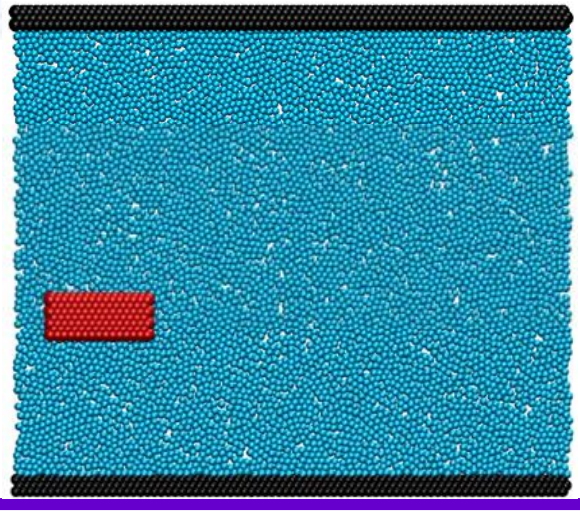
Aksimentiev, Heng, Timp, and Schulten, *BioPhy. J.* 87 (2004) 2086.

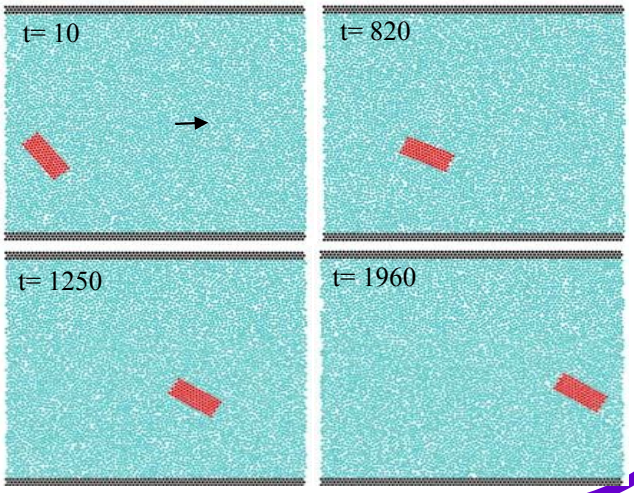




Problem – MD Simulations limited to ~100 ns; 'True' translocations are millisecond-scale events.

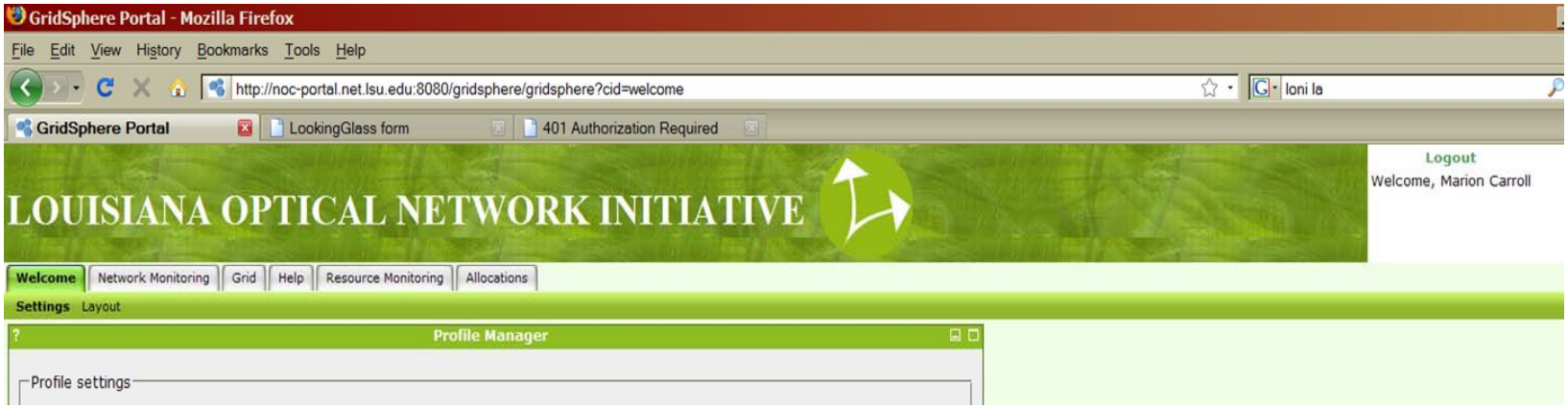
MD Simulation using Lennard-Jones Fluid (4M time steps, step ~2.16 ps; w ~ 25 nm; L ~ 27 nm).





Discovery of new *Alu* Subfamilies using HPC (Marion Carroll, XU)

TREE_PUZZLE and Maximum Likelihood Analysis



Queenbee HPC is employed to generate output files using TREE-PUZZLE that suggest divergence of uncharacterized Alu Y elements into subfamilies. Diagnostic mutations must then be described via sequence alignment in MEGA.

Tree-Puzzle Algorithm
(Marion Carroll, XU)

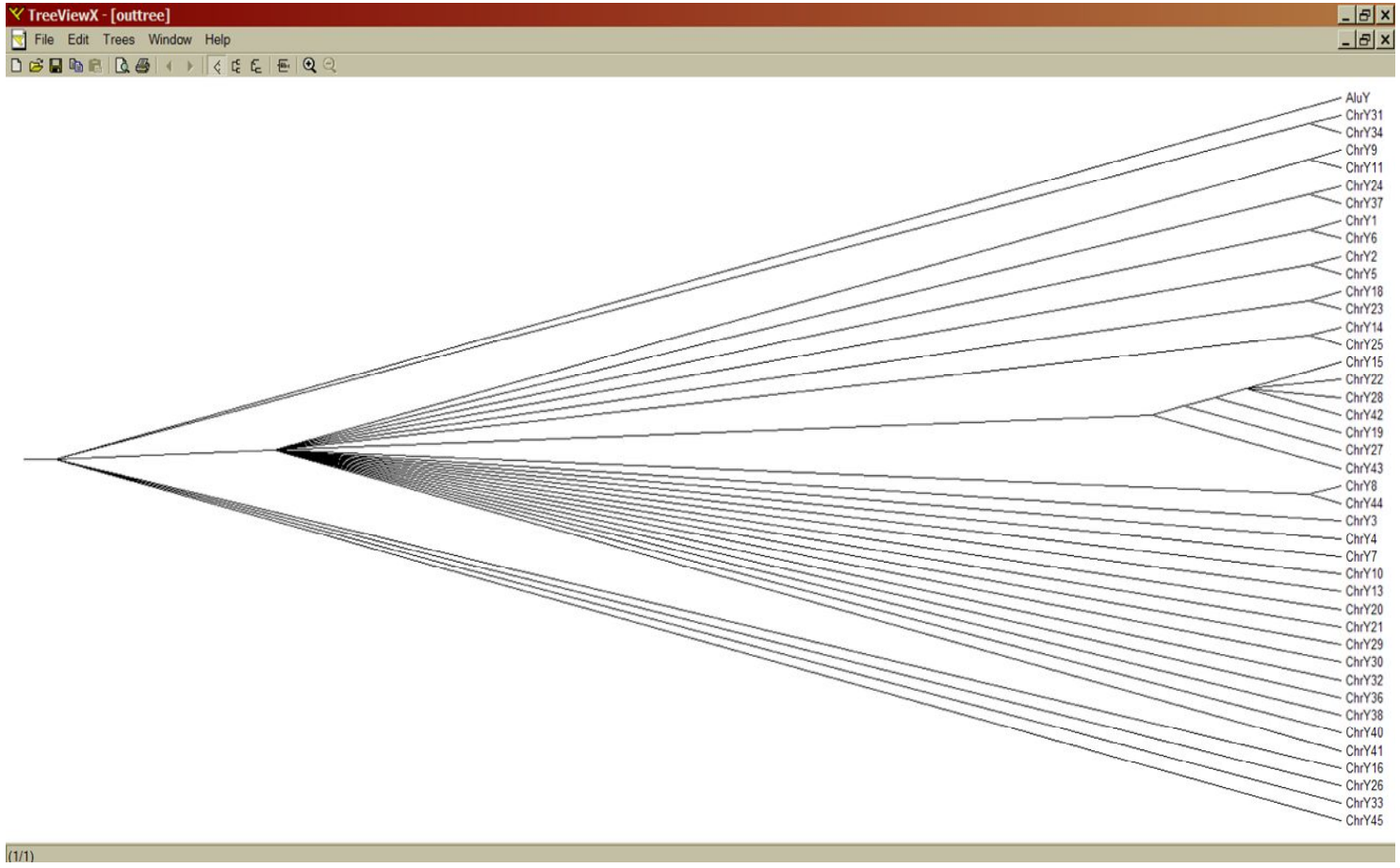
TREE-PUZZLE is an application run on Queenbee that reconstructs phylogenetic trees from nucleotide sequences by maximum likelihood.

TREE-PUZZLE conducts a number of statistical tests on the data set.

It does a tree search algorithm or quartet puzzling that allows analysis of large data sets using MPI.

Output of Tree-Puzzle Analysis (Marion Carroll, XU)

OUTTREE FILE



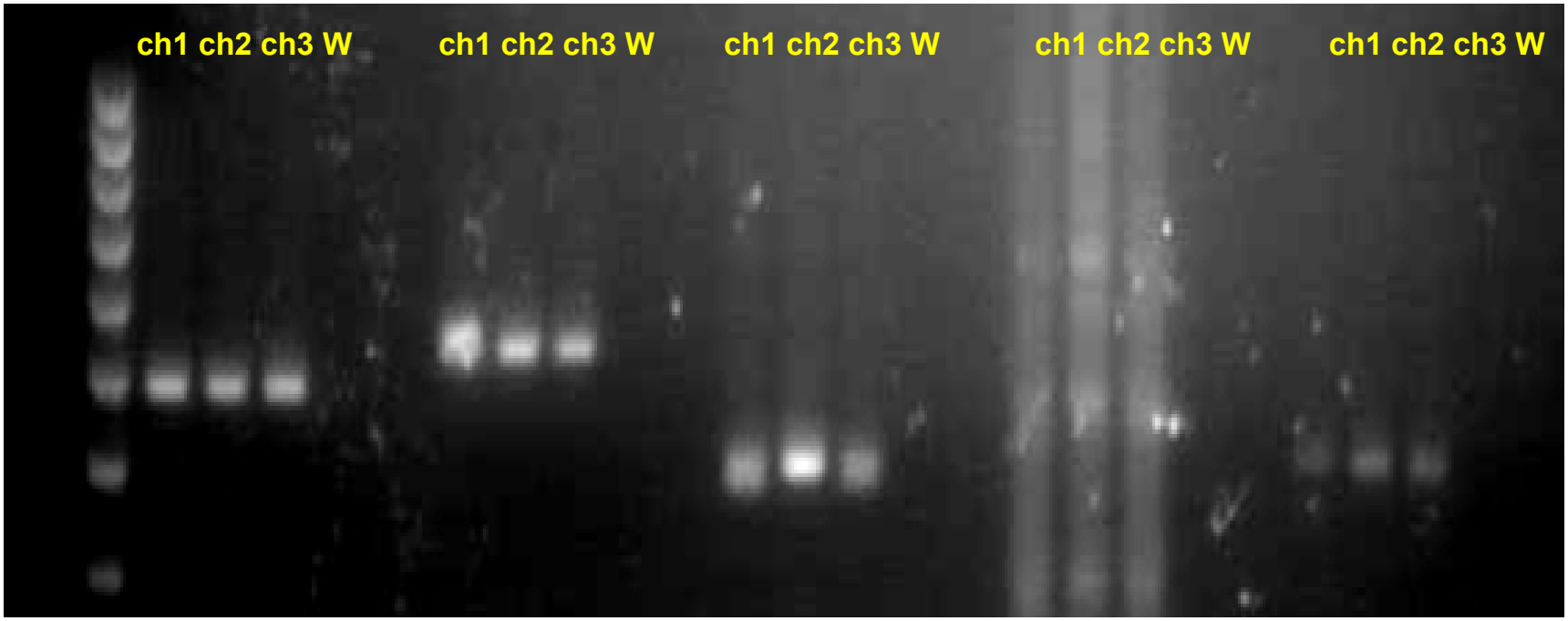
Tree-Puzzle Analysis Output (Marion Carroll, XU)

Subfamily Diagnostic Mutations

```
                                                
```

Experimental Verification
(Marion Carroll, XU)

Chimp Alu Polymorphic Display



Microfabrication Infrastructure
(Pin-Chuan Chen, LSU)

Forming Patterns
10^{-8}m => 10^{-1}m

X-ray lithography

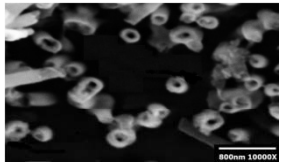
Micro-milling

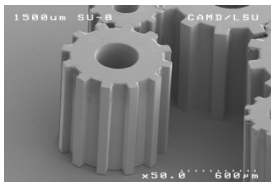
UV lithography

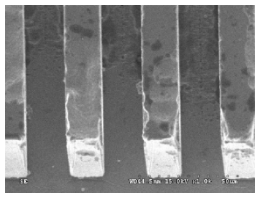
Excimer laser

Obducat
nano-imprinting

Filling Patterns (Metals)
10^{-8}m => 10^{-1}m

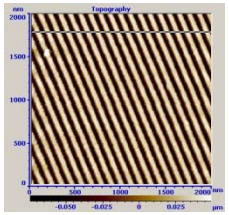






Replicating Patterns
10^{-8}m => 10^{-1}m

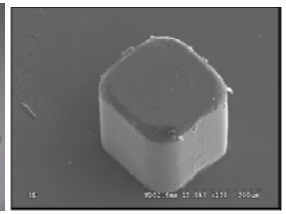
Battenfeld injection molding



50 nm grating

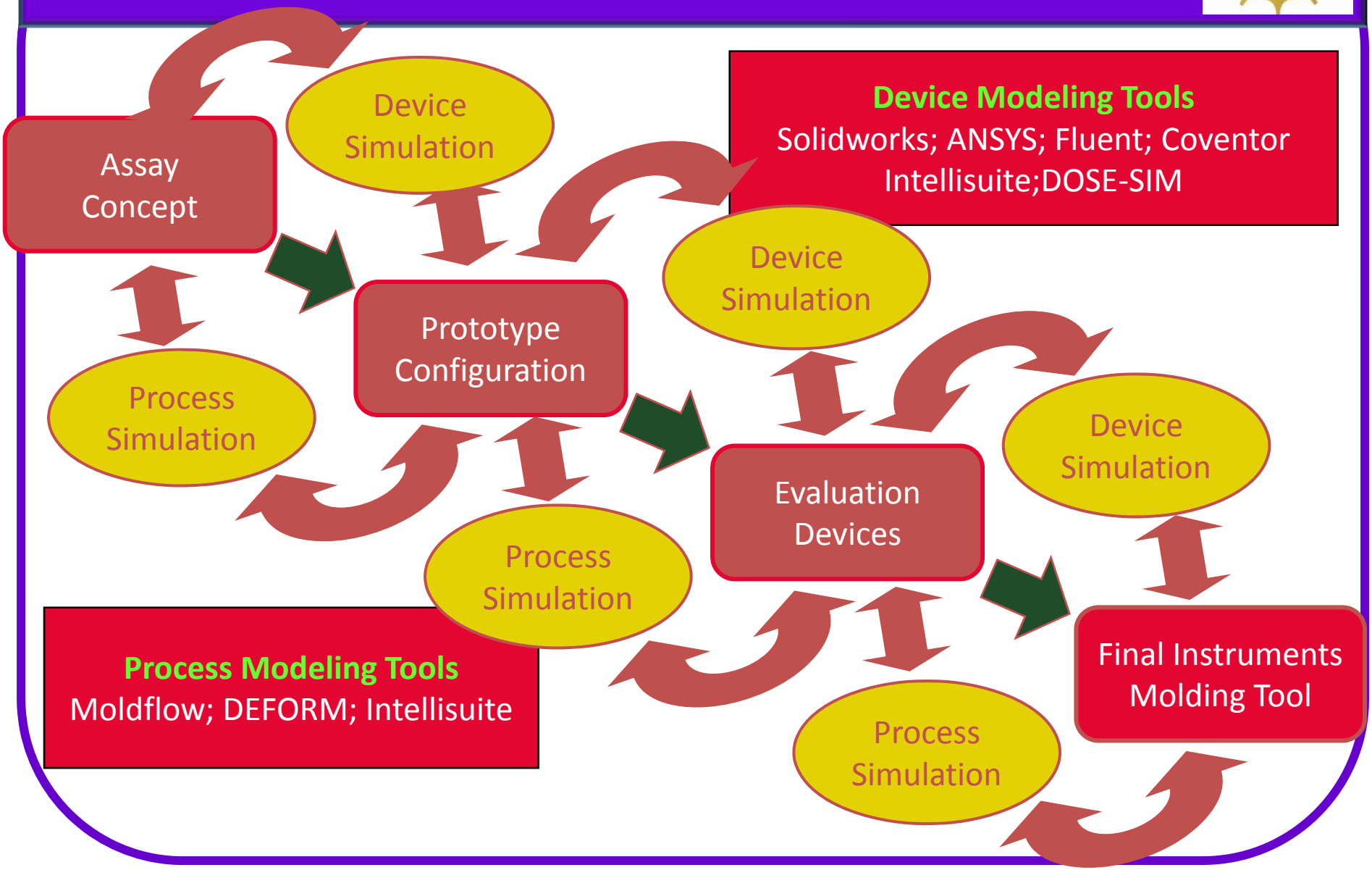
Jenoptik HEX 02

Double-sided
hot embossing



Injection molded
cube

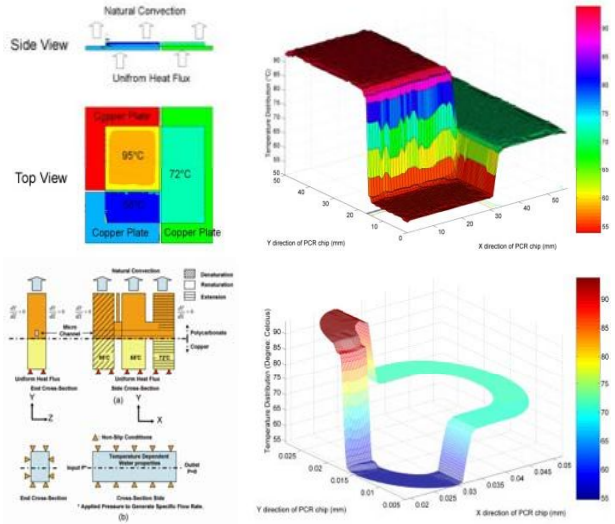
Design and Realization
(Pin-Chuan Chen, LSU)



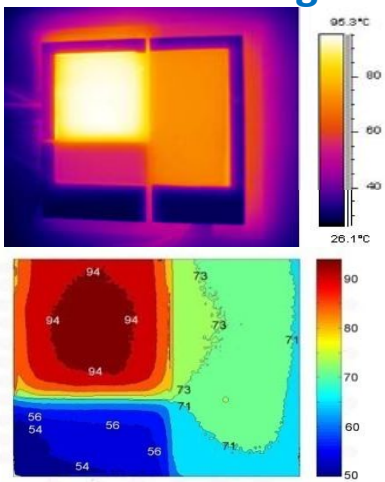
In-Plane Thermal Management (Pin-Chuan Chen, LSU)

Original CFPCR (3 cm X 4 cm)

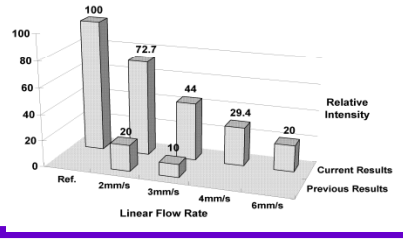
Finite Element Analysis



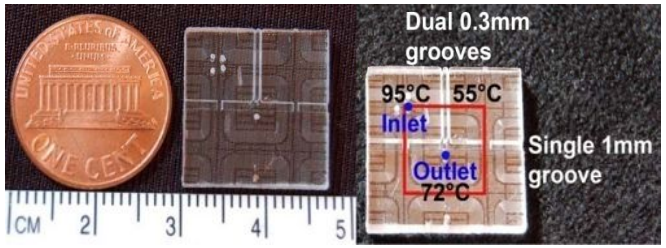
IR Camera Images



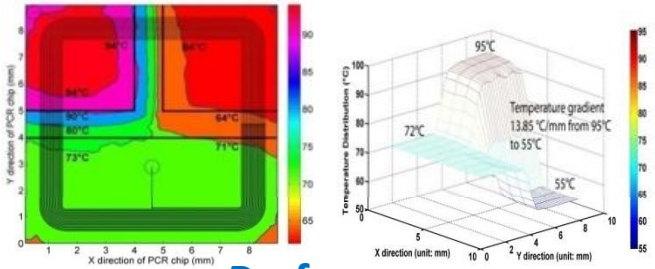
Improved Performance



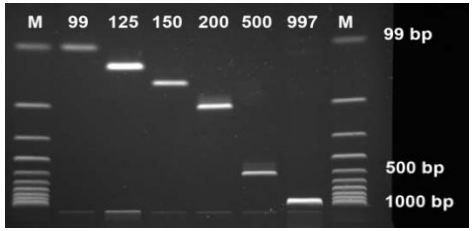
Small Area CFPCR (9 mm X 9 mm)

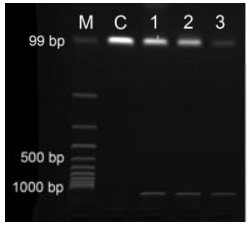


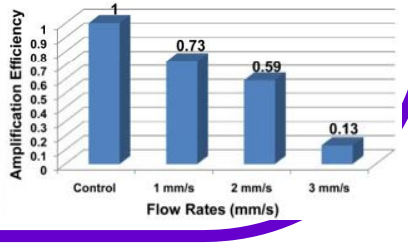
Finite Element Analysis



Performance

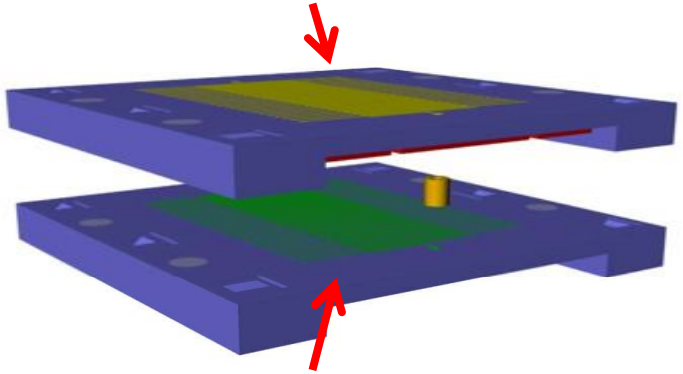




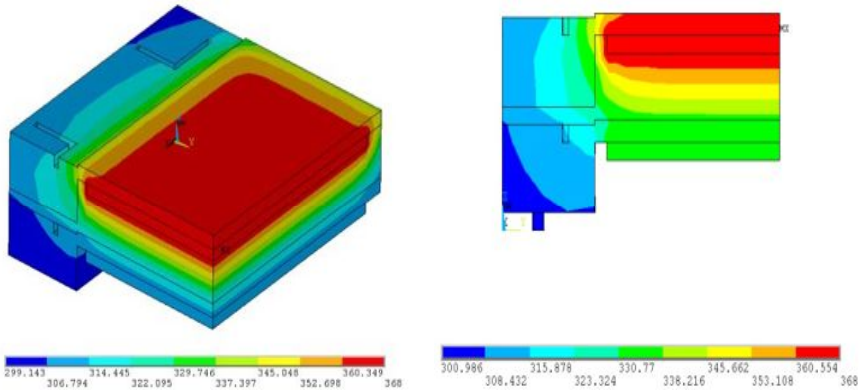


Stacked Thermal Management (Pin-Chuan Chen, LSU)

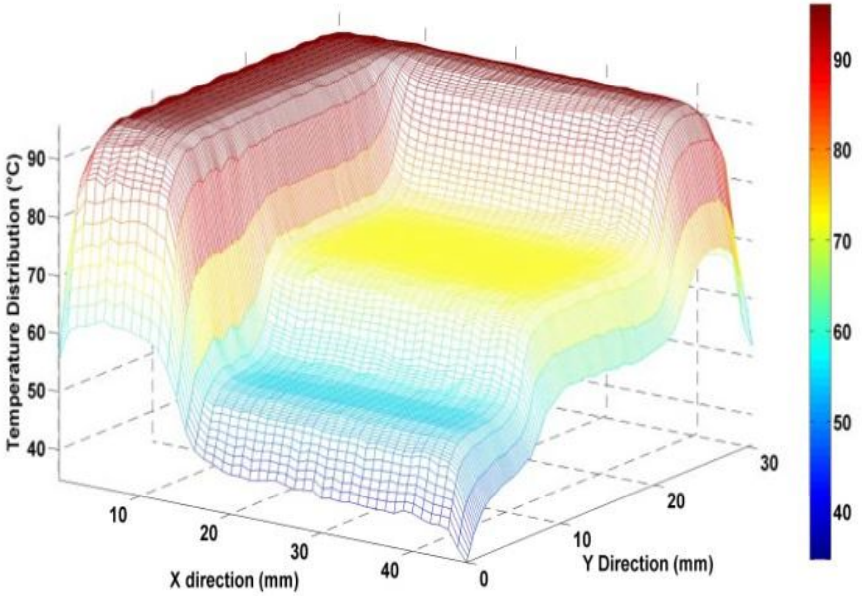
Microfluidic Module 1



Microfluidic Module 2

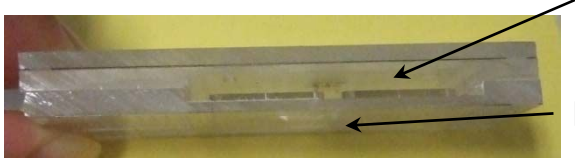


Finite Element Analysis to Understand Heat Transfer From Layer to Layer



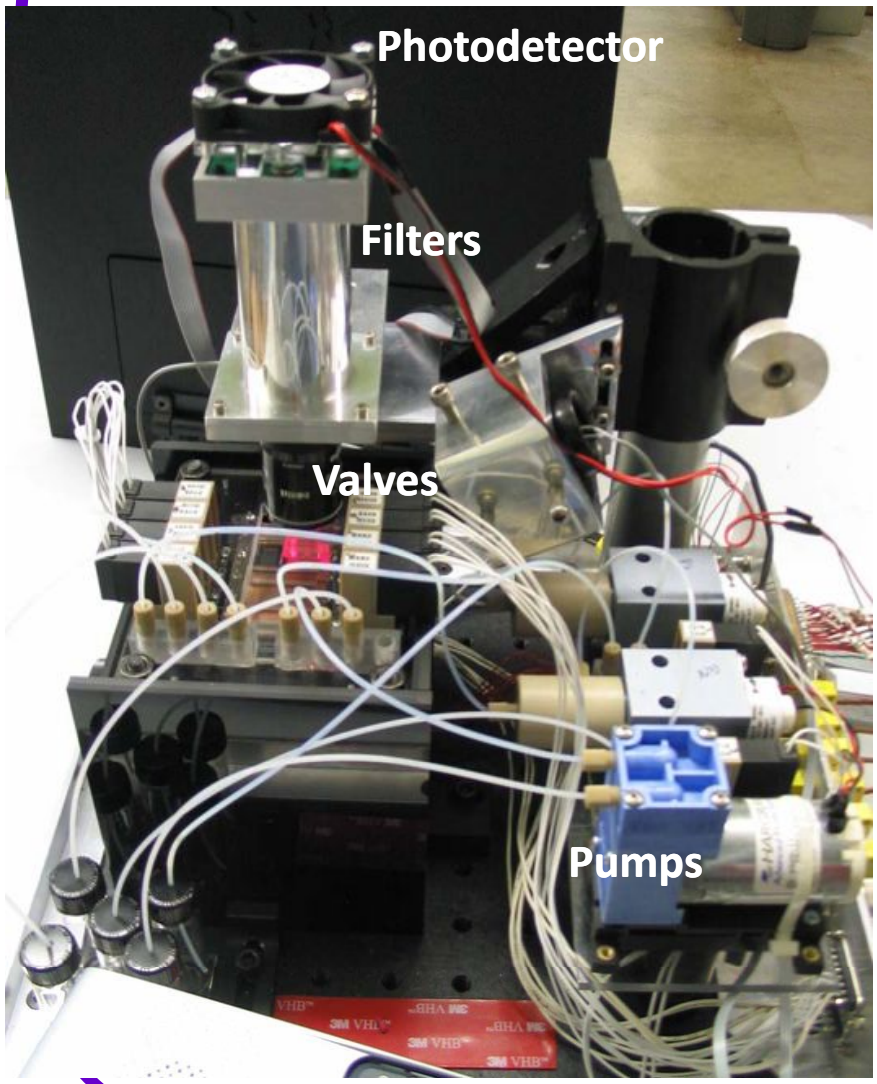
Thermal Reactor (CFPCR)

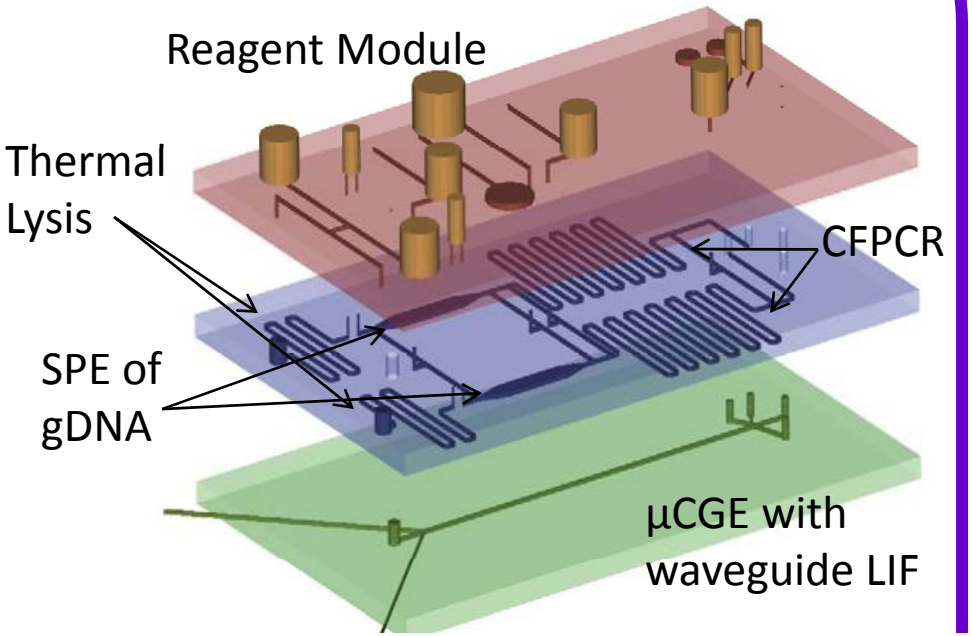
Air gap

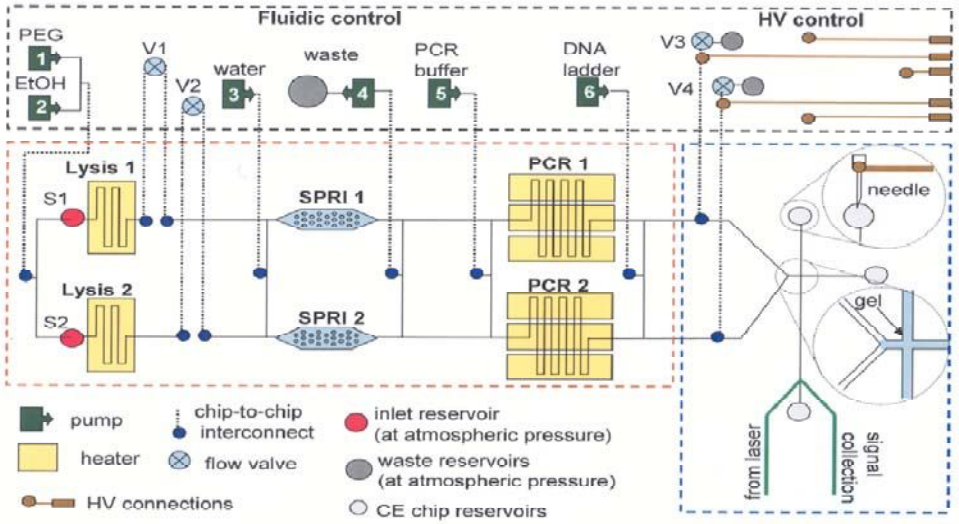


Electrophoresis

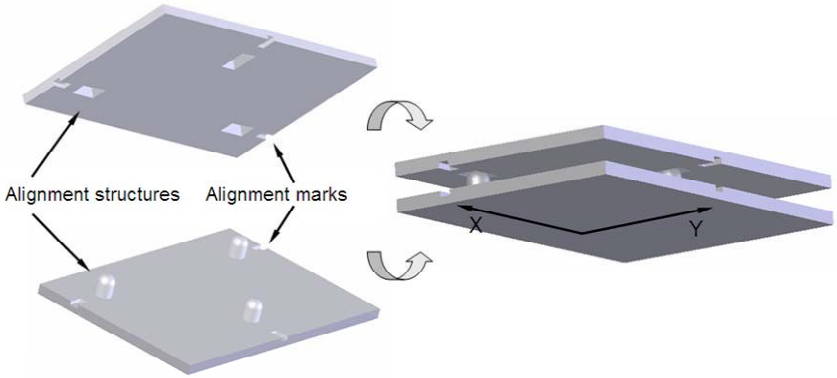
Genosensor for Human Indentification (Jason Emory, LSU)

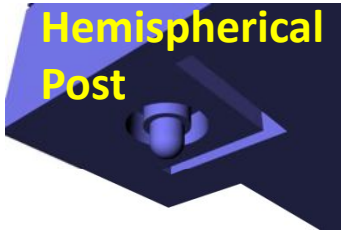


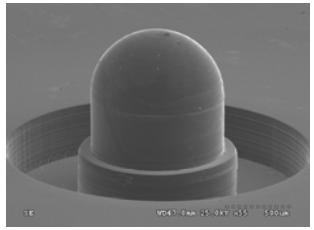


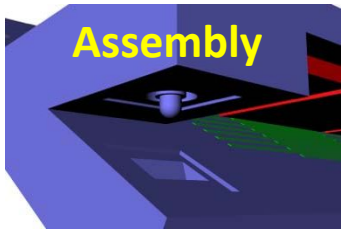


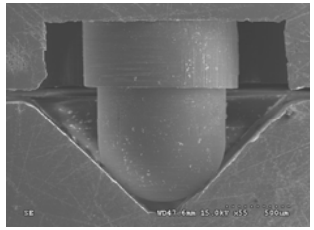
Passive Alignment Structures (Jason Emory, LSU)

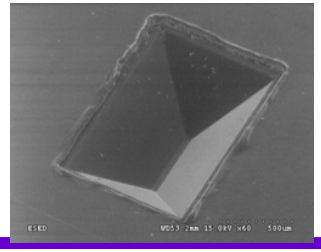


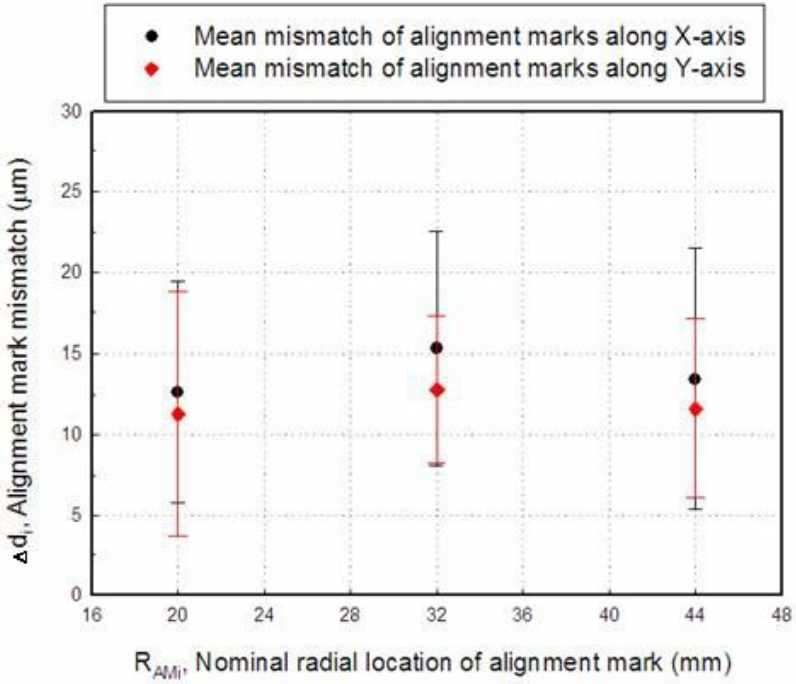






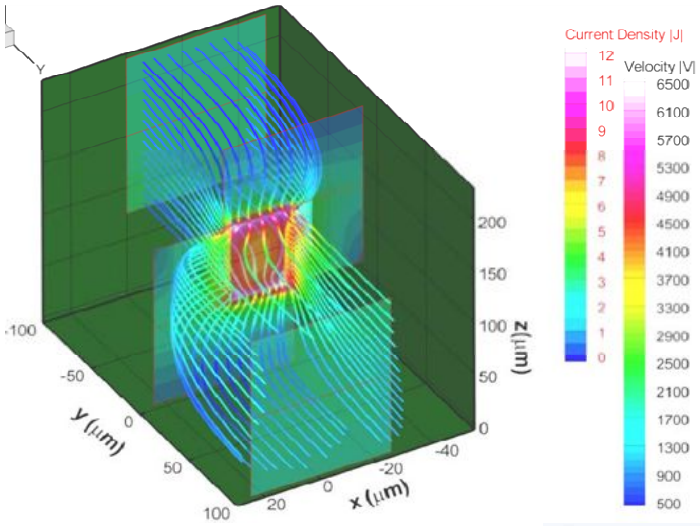


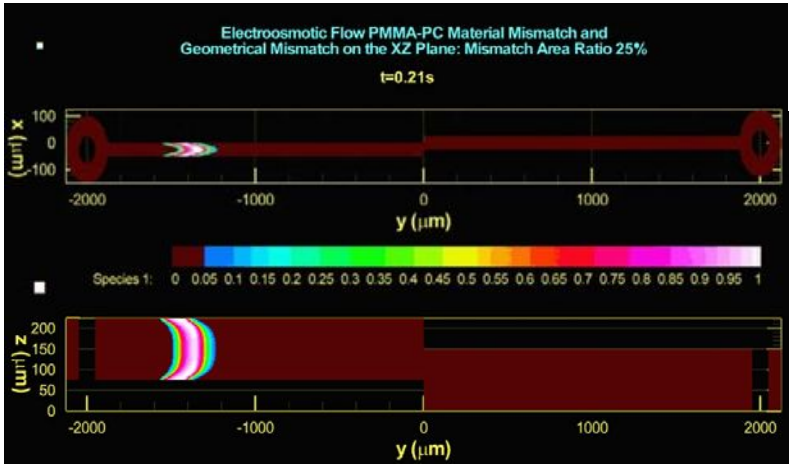




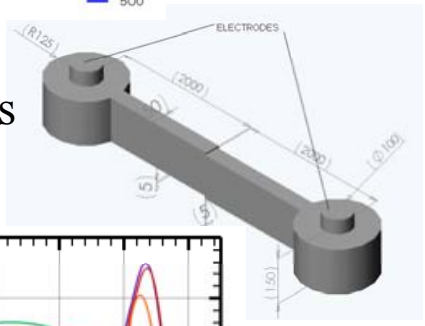
- **Mean lateral offset in X- and Y-axes 10-15 μms**
- **Not location dependent**
- **Nominal post height 925 μms**
- **Mean hot embossed post height 922 ± 2 μms**
- **Standard deviation < 6 μms**

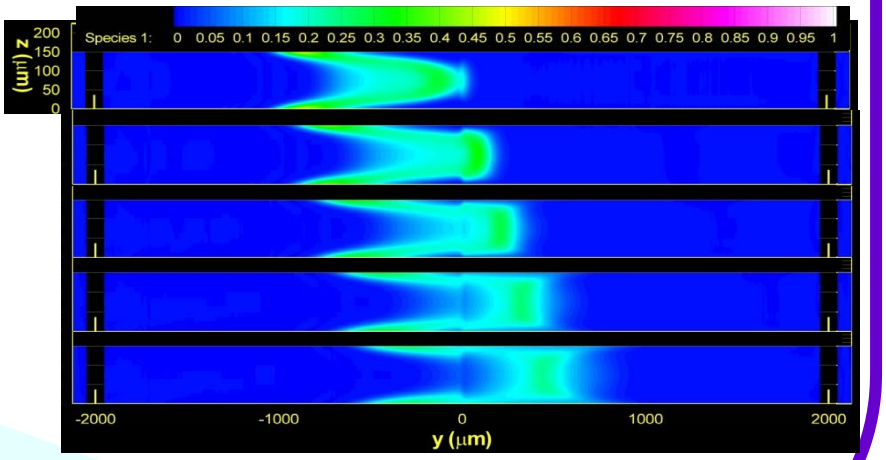
Interconnects for Fluid Transfer between Modules (Jason Emory, LSU)



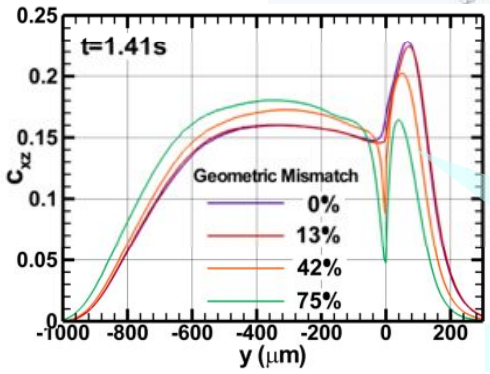


Q=7.03nL/s



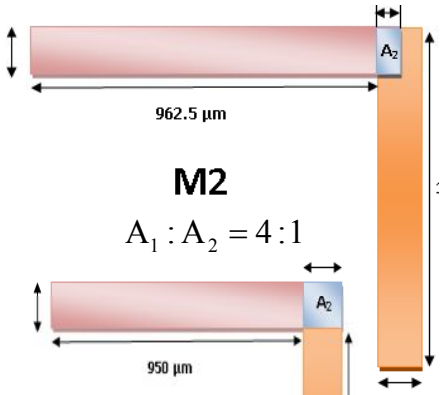


No effect for 13% geometrical mismatch



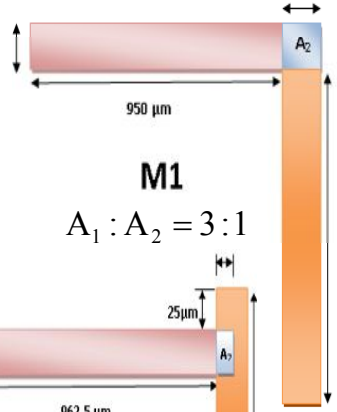
Plug re-concentration after PMMA-PC material mismatch

Interconnects: Channel Overlap Configuration (Jason Emory, LSU)



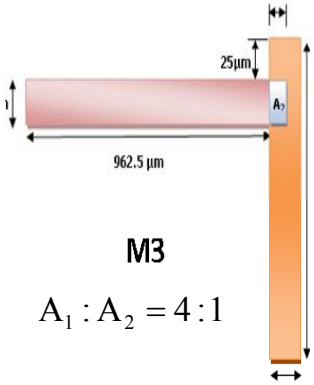
M2

$A_1 : A_2 = 4:1$



M1

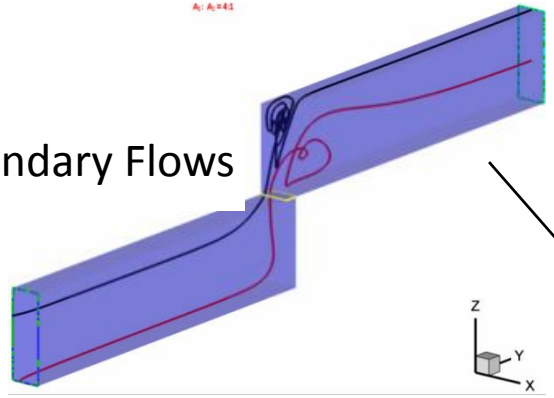
$A_1 : A_2 = 3:1$

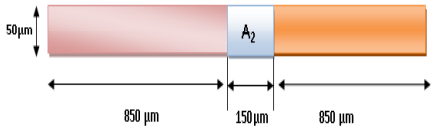


M3

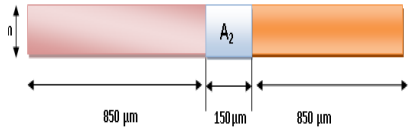
$A_1 : A_2 = 4:1$

- Secondary Flows

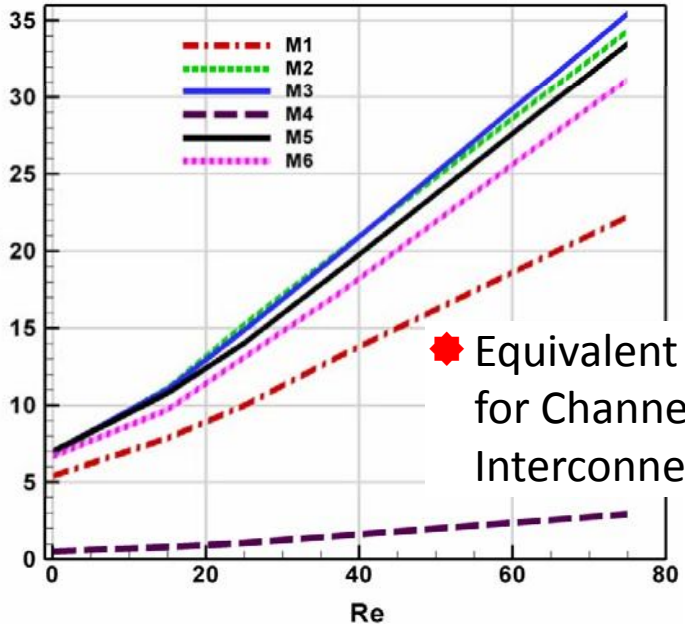




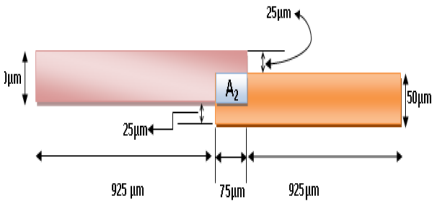
M4 $A_1 : A_2 = 1:1$



M5 $A_1 : A_2 = 4:1$



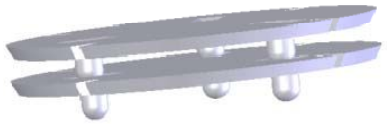
- Equivalent Lengths for Channel-Overlap Interconnects



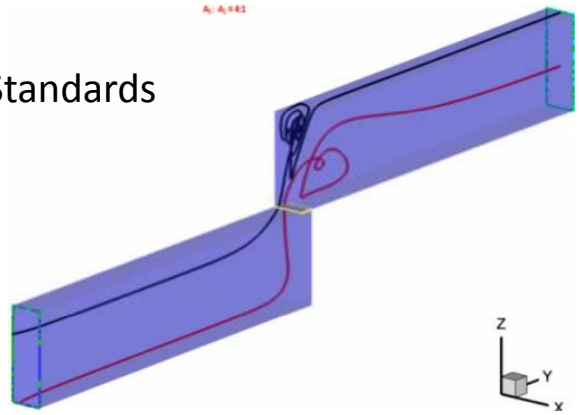
M6 $A_1 : A_2 = 4:1$

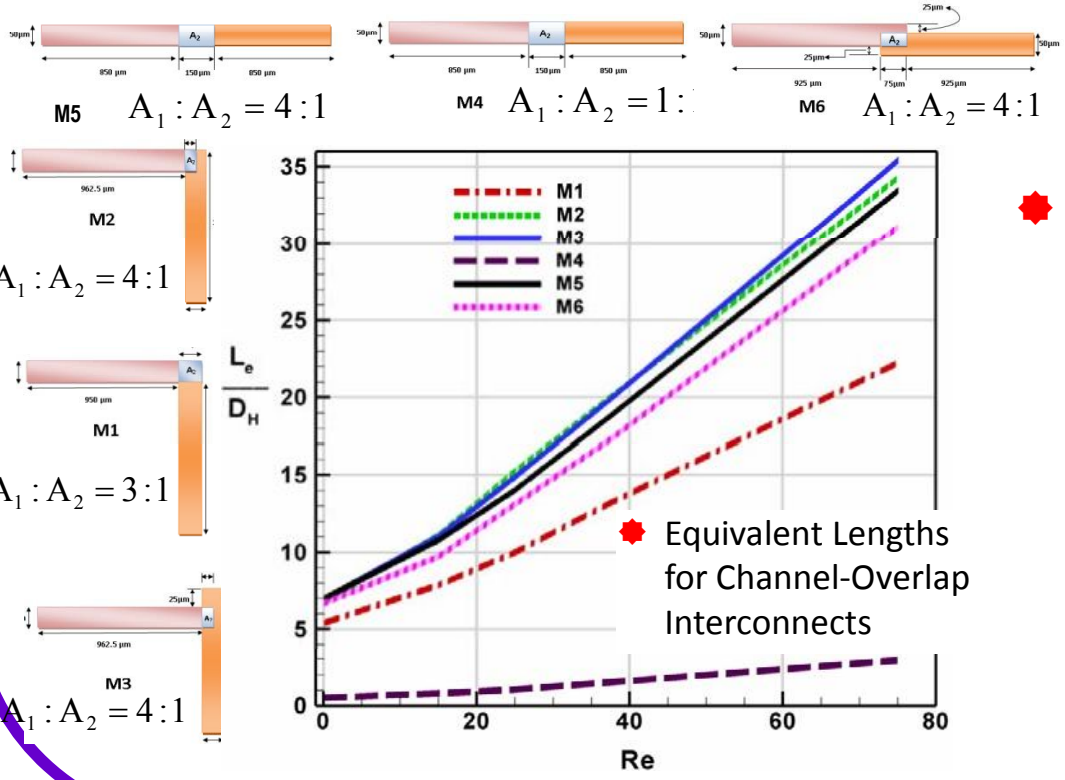
See Poster by S. Rani

Design Knowledge Base & Rules Through Simulation (Dimitris Nikitopoulos, LSU)



- Pressure Loss Through Interconnects
  - Determined according to ANSI/ASME Standards
  - Equivalent length dependence on
    - Reynolds number
    - Interconnect Configuration





- Equivalent Lengths for Channel-Overlap Interconnects

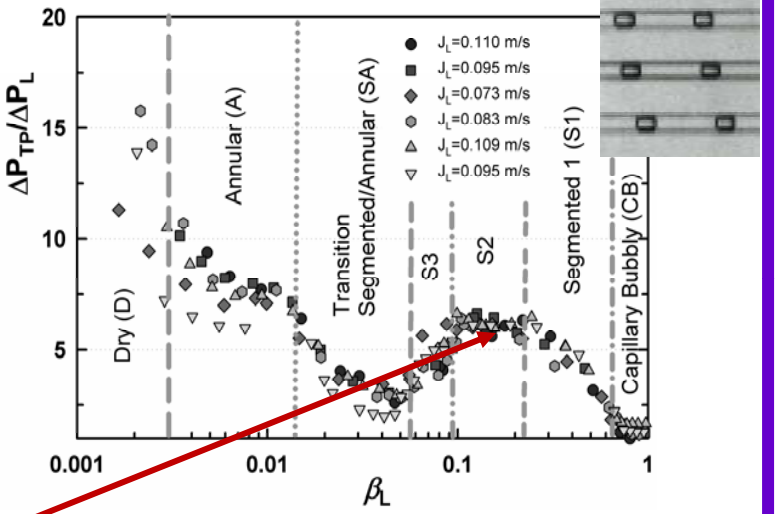
- HPC Utilization and Benefit
  - Migrate commercial codes on Queenbee (WP4)
  - Parametric study parallelization (WP1)
  - Interactive Post-processing and Data Management
  - *Full-system simulation* when component-by-component approach fails (e.g. processes involving heat and mass transfer)

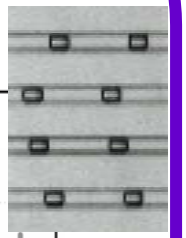
See Posters
by N. Kim (Exp.)
by E. D. Walker (Sim.)

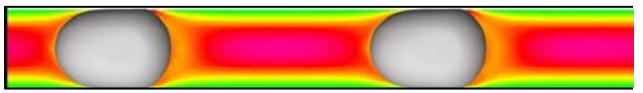
Understanding Multi-phase Micro-Fluidics (Dimitris Nikitopoulos)

- ☑ Code adapted to handle wall-interface interaction and break-up/coalescence
- ✸ Parallelization – Code (Walker; WP4 - Tyagi) – Run (WP1)
- ✸ Advanced interactive visualization tools (WP3 - Ullmer)
- ✸ Improvement of Accuracy/Performance
  - ✶ Multi-Grid algorithm for elliptic eqns. with discontinuous coefficients (Walker; WP4 – Aksyolu, Tyagi)
- ✸ Handle complex Cartesian geometries (Walker; WP4 - Tyagi)
- ✸ Computational Steering (WP1, WP3, WP4)

Segmented flows

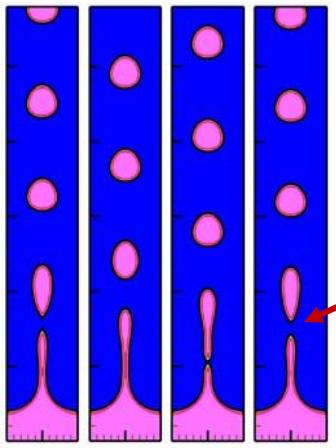






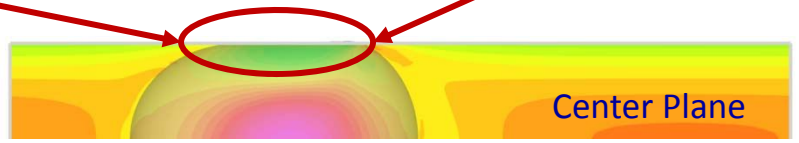
film break-down = contamination

Injection

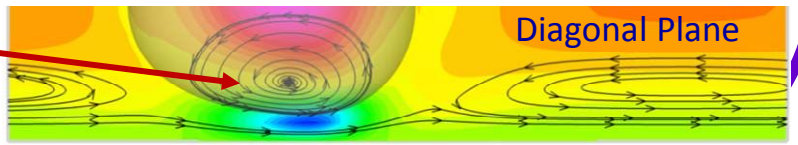


Scientific Challenges

- ✸ Pressure drop prediction
- ✸ Understand and model the physics
  - ✶ Interface/wall interaction
    thin films - wettability
  - ✶ Break-up (injection)
  - ✶ Coalescence

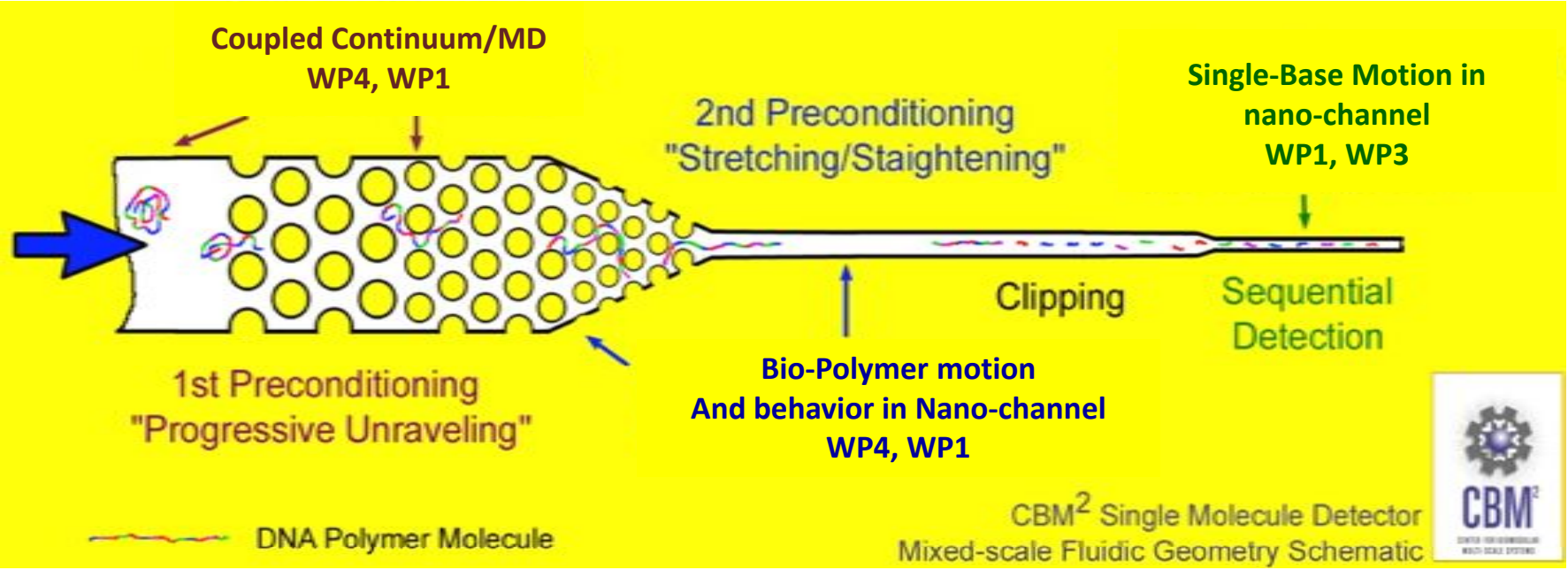


secondary flows
aid mixing



Multi-scale Application Test-Bed Example (Dimitris Nikitopoulos)

- Single-Molecule Multi-Scale Sensor
  - 1st Preconditioning: Milli- micro- to nano-scales
  - 2nd Preconditioning & Bio-polymer length meas.: micro- to nano-scales
  - Nano-channel Small Molecule Sensor: nano-/molecular scales

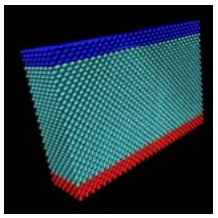


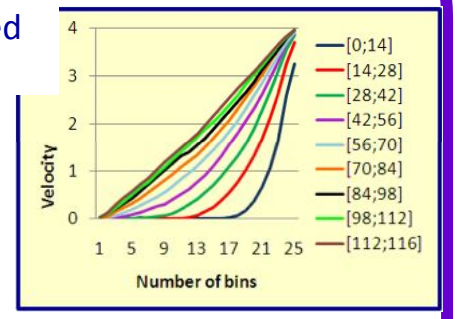
See Poster by T. Dufaud

Multi-scale Coupled MD-Continuum Simulation Tool (Dimitris Nikitopoulos)

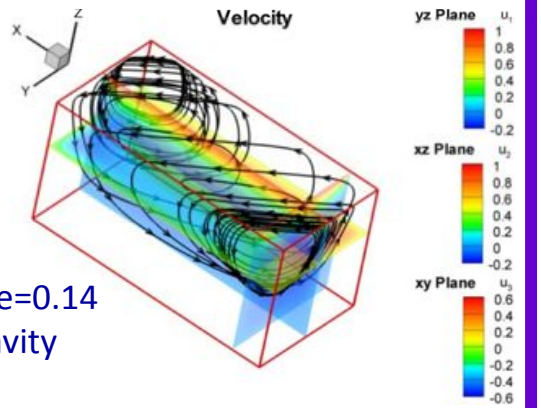
- Basic in-house MD code
  - ☑ Developed, parallelized, Tested (Couette, Poiseuille)
  - Documentation of the code for delivery to WP4
  - Migration to CACTUS (New Student, WP4-Tyagi, Schnetter, Kim)
- Continuum 3D N-S Parallel Code (Velocity/Vorticity)
  - ☑ Developed, parallelized (T.-Dervout*, Dufaud)
  - ☑ Tested on 3D driven cavity test problems (Dufaud)
  - Documentation of the code for delivery to WP4 (Dufaud)
  - Migration to CACTUS (New Student, WP4-Tyagi, Schnetter)
- Continuum-MD Coupling
  - Parallelization issues (Dufaud, New Student; WP4-Tyagi)
  - MD-Continuum code coupling using constrained dynamics under CACTUS (New Student; WP4-Tyagi, Schnetter, Kim; WP1)

Test Results: Impulsively Started Couette Flow

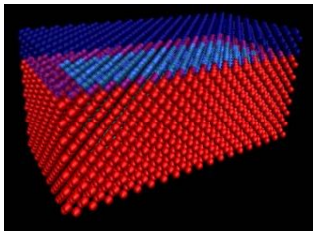




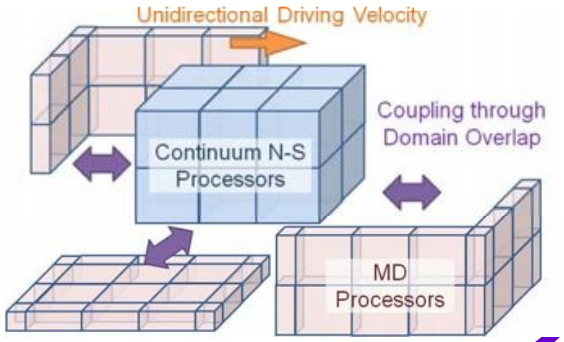
Test Results: Re=0.14 xy-driven cavity



MD Domain Layout for Driven- Cavity Test Problem

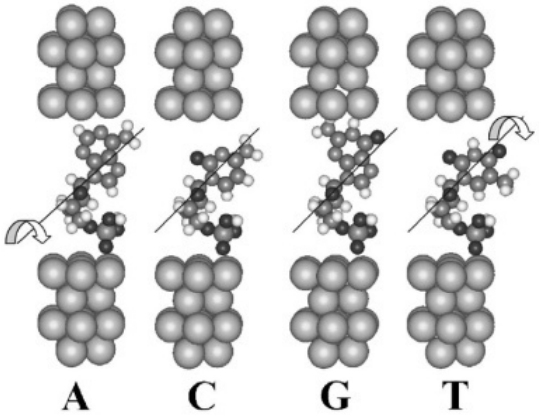


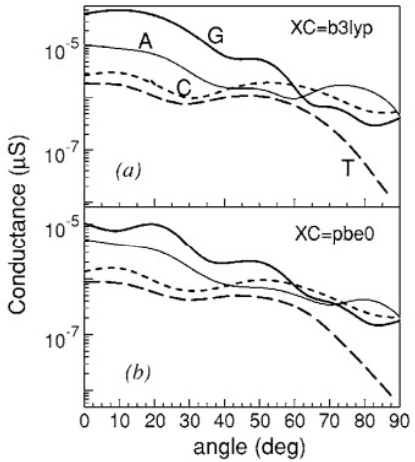
Problem Distribution Schematic for Coupled MD-Continuum Driven-Cavity Test Problem



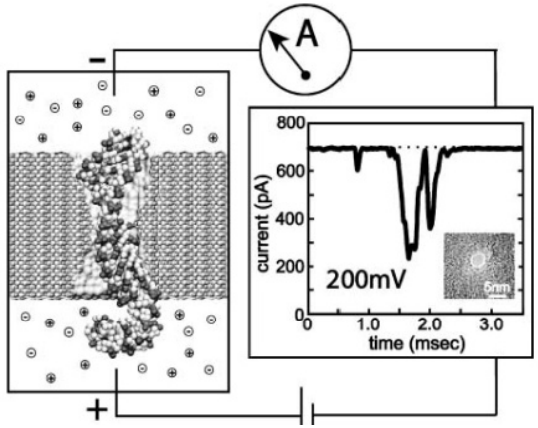
Atomistic Simulation of Biopolymer Transport through Nano-Domains (Dorel Moldovan, LSU)

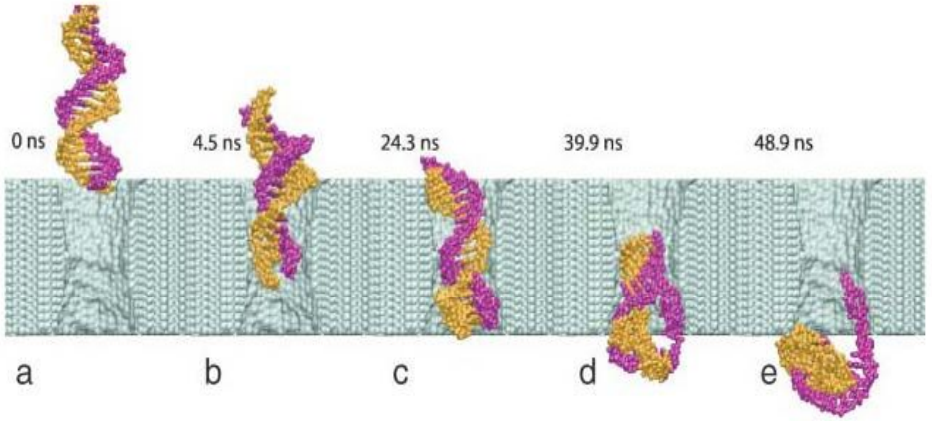
Motivation





R. Zikic et al. , Characterization of the tunneling conductance across DNA bases, Phys. Rev E 74, 011919 (2006)



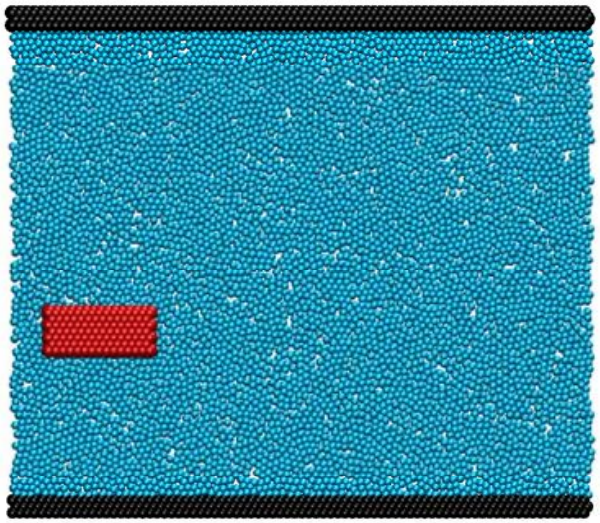


A. Aksimentiev et al., et al. , Microscopic kinetics of DNA translocation through synthetic nanopores, Biophys. J. 87 (2004) 2086

Atomistic Study of Biopolymer Transport through Nano-channels (Dorel Moldovan, LSU)

Methodology and Simulation System

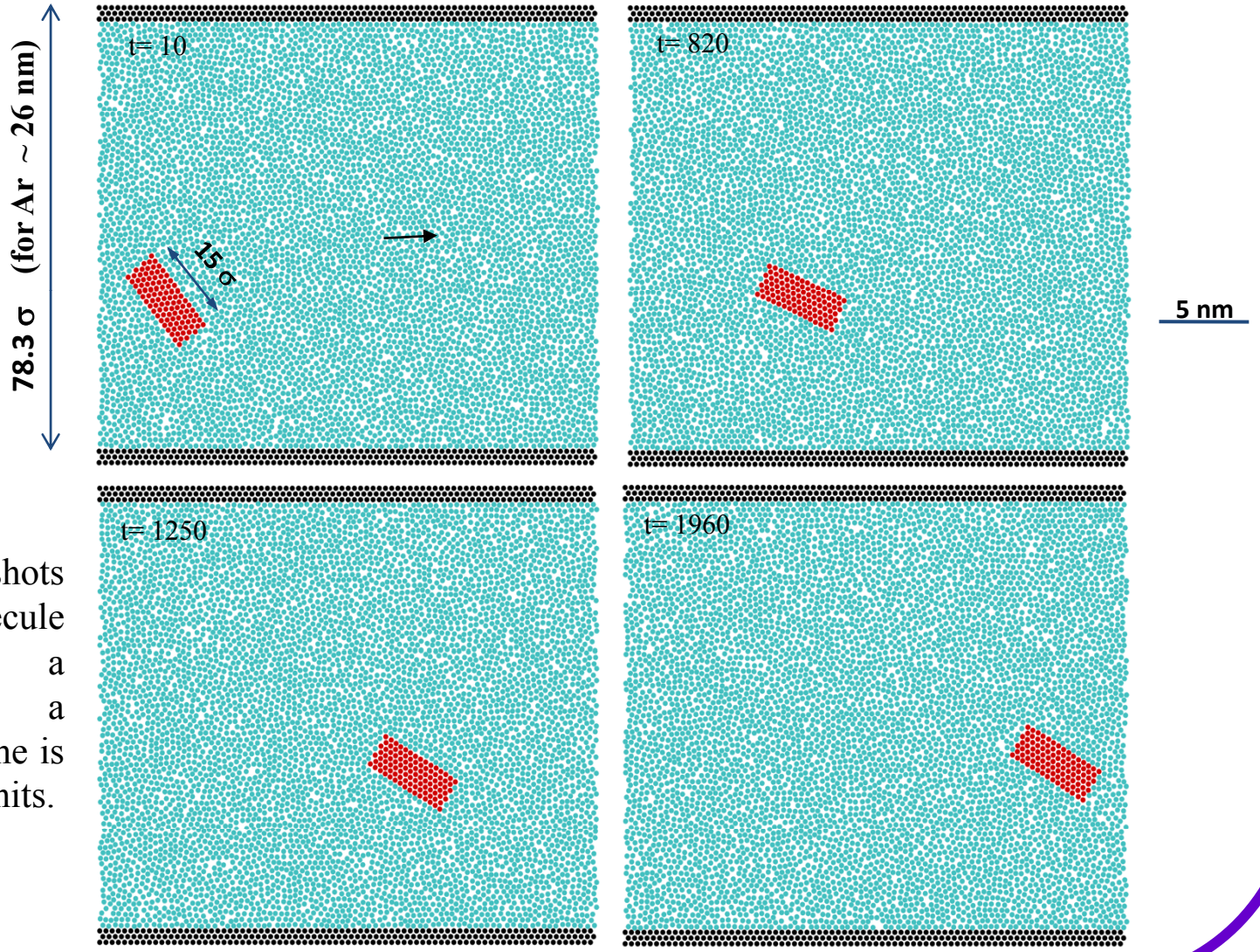
- MD simulations were performed with the software package LAMMPS
- The interactions between any pair of atoms are described by the Lennar-Jones potential.
- The two-dimensional system consists of ~6000 atoms and the molecule has an elongated shape of aspect ratio 2.6
- The simulations were conducted and analyzed in reduced units. The distances are expressed in units of σ, the energy in ε, the temperature in ε/k_B, the time in, $1/\sqrt{\varepsilon/m\sigma^2}$ the force in ε/σ, the density $1/\sigma^2$, etc.
- The simulations were carried out at temperature $k_BT/\varepsilon = 1.2$ and density $\rho/\sigma^2 = 0.81$.
- The Poiseuille flow was induced by introducing a "gravity" force that is applied parallel to the channel axis to each atom of the liquid and molecule.



$$V_{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

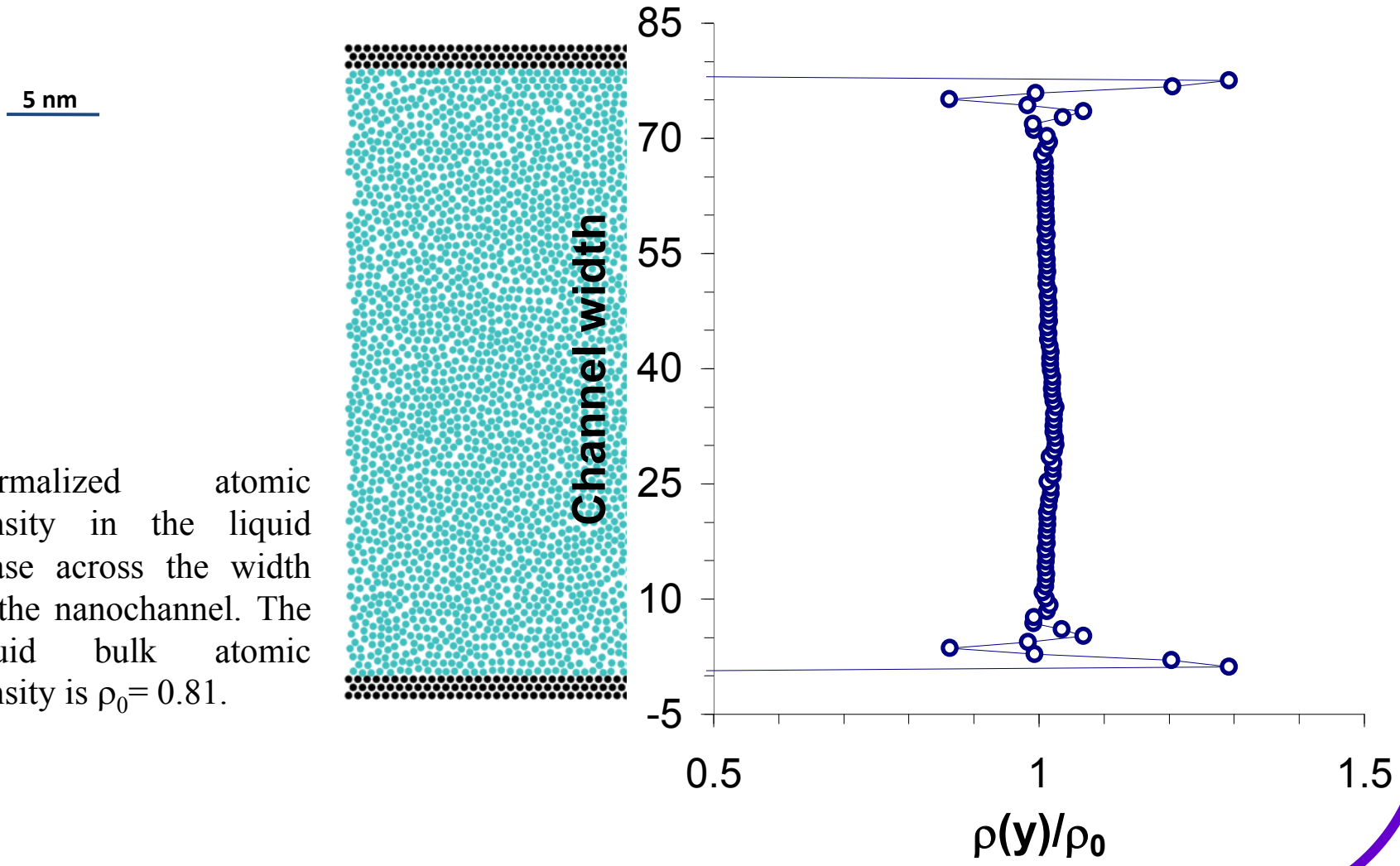
For Ar: $\sigma = 3.4$ Å, $\varepsilon/k_B = 120$K, m=40 a.u. accordingly the natural time unit is = 2.16 ps.

MD Simulation Results



Simulation snapshots of the molecule moving in a nanochannel in a Poiseulle flow. Time is given in reduced units.

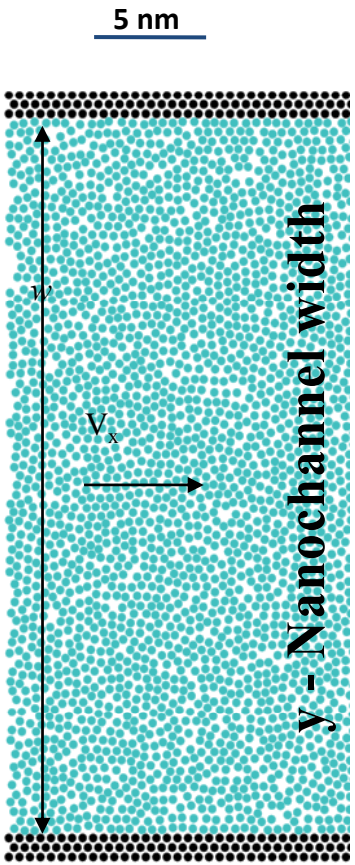
Atomic Layering Close to Walls (Dorel Moldovan, LSU)

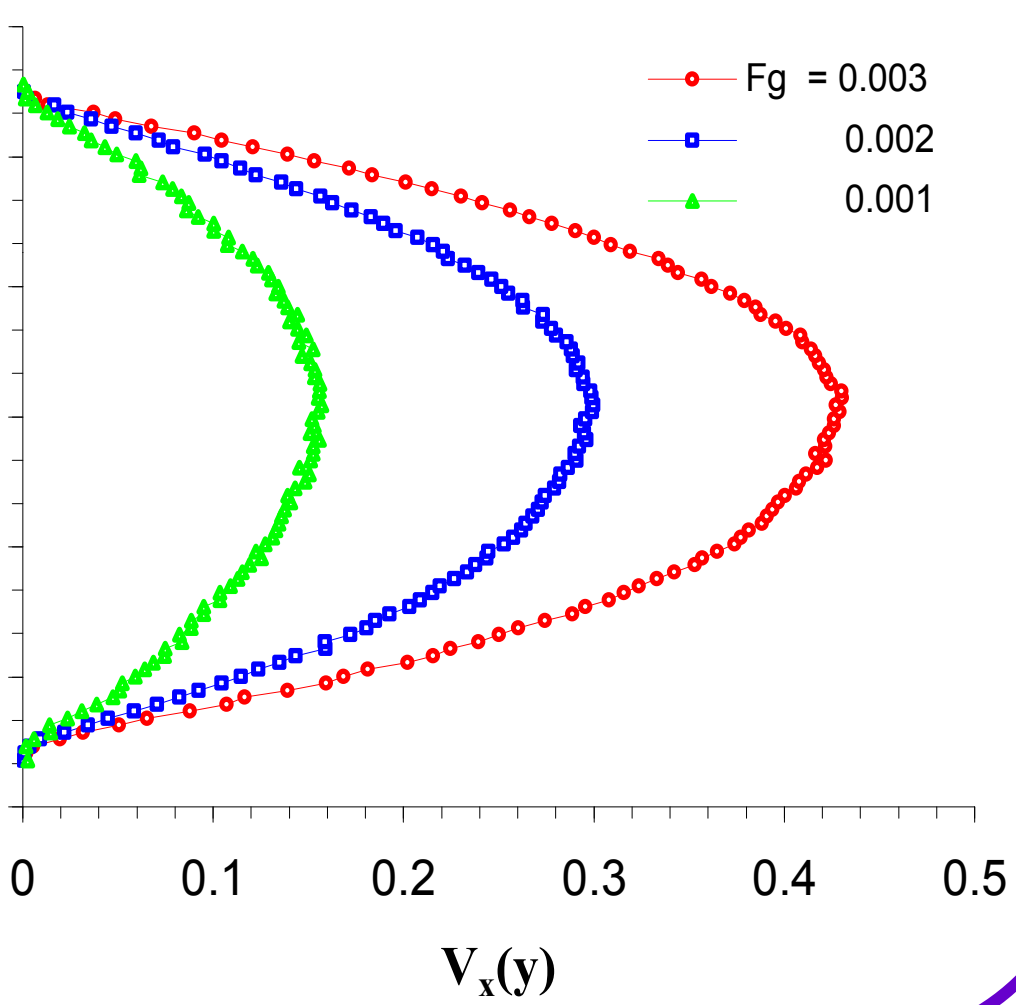


Normalized atomic density in the liquid phase across the width of the nanochannel. The liquid bulk atomic density is $\rho_0 = 0.81$.

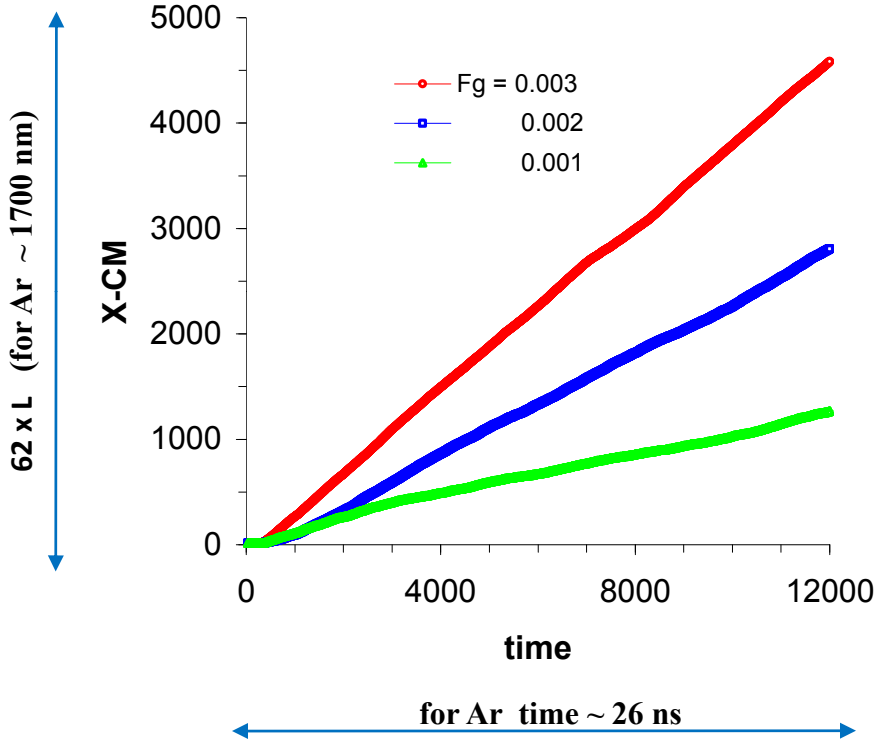
Transverse Velocity Profile (Dorel Moldovan, LSU)

Velocity profiles obtained from MD simulations of Poiseuille flow. The result are given for three values of the additional constant force, F_g=0.003, F_g=0.002, and F_g=0.001, applied to each "liquid" atom to generate the flow.

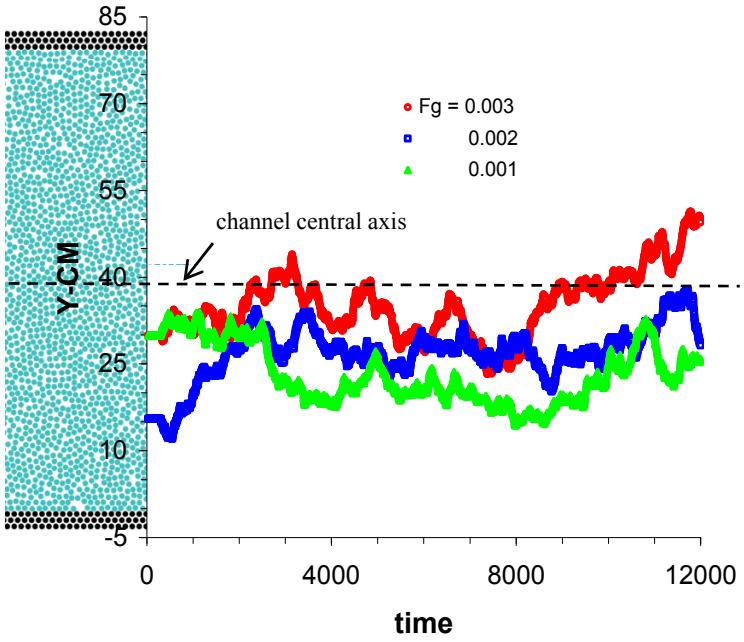




Time Evolution of Molecular CM during Translocation (Dorel Moldovan, LSU)



Variation of the x-component of the position of the molecule center of mass vs time for three flow regimes controlled by gravity forces: F_g=0.003, F_g=0.002 and F_g=0.001



Variation of the y-component of the position of the molecule center of mass vs time for three flow regimes controlled by gravity forces: F_g=0.003, F_g=0.002 and F_g=0.001

Education and Outreach: Professional Development Seminars for GS, PDF

CBM² Seminar Series

"Using LSU's High-Performance-Computers to Simulate Merging Stars"

by **Prof. Joel E. Tohline**

Department of Physics and Astronomy and Coast to Cosmos (C2C) Focus Area Lead
at Center for Computation and Technology (CCT)
Louisiana State University

Astronomers understand that the internal structure of individual stars, like our Sun, as well as the interactions between pairs of stars that orbit one another in so-called "binary star systems" is governed by essentially the same set of mathematical equations that govern fluid flows here on Earth. However, generally speaking, very large and very fast computers are required to solve this complex set of equations, especially in the case of strongly interacting binary systems. We are using high-performance-computers (HPCs) at LSU and across LONI (Louisiana Optical Network Initiative) to study the evolution of binary stars whose interactions are so strong that they eventually collide and merge. Such violent events in nature are thought to give rise to certain types of supernovae or even more energetic phenomena referred to as gamma-ray bursts (GRBs). Research by various groups within LSU's Center for Computation and Technology (CCT) has aided us in our pursuit of this challenging astrophysics goal, and is guiding our plans to effectively use future generations of HPC hardware.

Wednesday, June 25, 2008
Presentation at 4:00 pm
Life Sciences Annex A101
followed by refreshments at 5 pm

For more info contact: Dr. Maggie A. Witek mwitek@lsu.edu

CBM² Seminar Series

Scientific and Professional Writing

by **prof. Malcolm Richardson**

Dr. J.F. Taylor Professor of English
Department of English
Louisiana State University

Methods to create good scientific writing are not complex or mysterious but require certain kinds of preparation which are typically not taught during English writing courses either in the U.S. or abroad. These methods, which should be fully understood before the first word is written, include first an understanding of basic rhetorical principles of audience analysis and second an understanding of both the purpose of the entire scientific document and of its different parts (introductions, results, discussion, etc.). This presentation will focus on writing theses, dissertations, and academic articles, and will suggest practical ways to be a more efficient writer by planning ahead.

Monday, October 16 th 2006
Refreshments at 4:30 pm
Presentation at 5:00 pm
Life Sciences Annex A101

Contact: Dr. Maggie Witek mwitek@lsu.edu

Other E&O Activities

- Science Adventure Camps (Audubon Girl Scouts) –
  Goal: increase interest in science/engineering in females; one-week summer camps with experiments in chemistry, biology, environmental engineering, mechanical engineering, biological engineering.
- Project Science (Cain Center, LA Dept. Education) –
  Goal: provide linkages to university and community resources to build synergistic relationships among scientists and educators.
- *You Be the Chemist* Challenge (Exxon)
  Goal: Provide middle school students the opportunity to be exposed to rigorous chemistry concepts and gain experience in participating in academic exercises.

Other E&O Activities

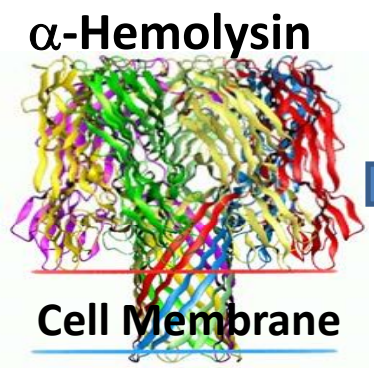
- **Science and Engineering Day @ LSU (08/01/08)**
  - Formal presentations and panel discussions on biological/medical technology needs; computational capabilities in microfluidics design; poster session
- **High School/Undergraduate Research Experiences**
  - *Ginger Granville* – Louisiana Arts and Science Academy High School, Microchip separation of *Alu* elements
  - *Jenny Hsu* – Princeton University, Novel Near-IR Fluorescent Dyes for Drug Discovery
- **Numerous Graduate Student Presentations at National/International Meetings**
  - Pin-Chuan Chen; Paul Okagbare; Jason Emory; Samuel Njorge; Matt Hupert (µTAS)
- **Publications** (10 faculty - 4 CHEM; 3 BS; 3 ME)
  - Team faculty members and their students published 68 papers in 07/08

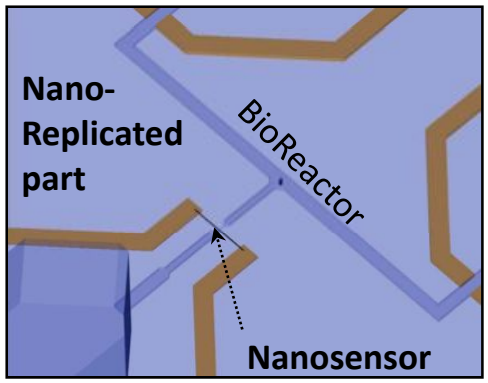
2007

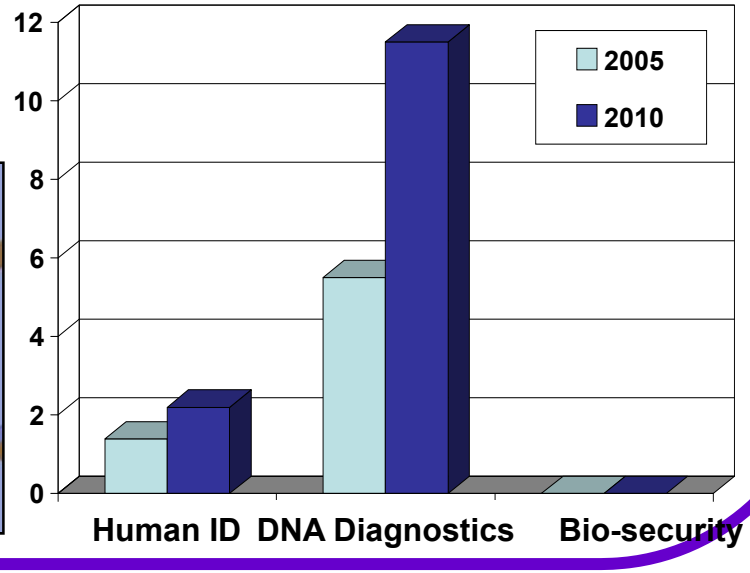
2008

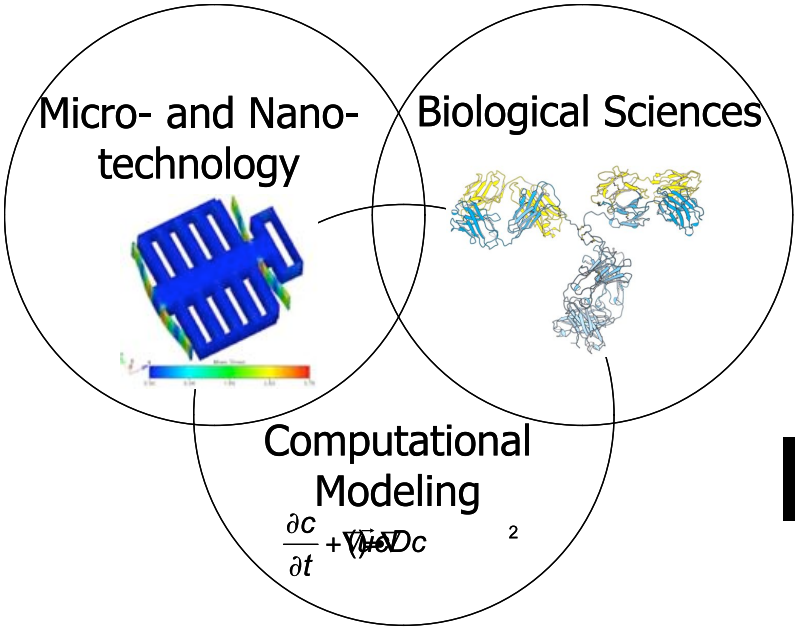
Other Activities (Patents and Entrepreneurship, Center Grants)

- **Statistics for Technology Transfer** – 10 disclosures and 4 Provisional Patents were filed in 2007/2008
- BioFluidica Microtechnologies, LLC – Commercial venue for new technologies emanating from CBM^2 (won two business plan competitions; CEO – Yohannes Desta, Ph.D. with Prof. Michael Murphy) – Development of point-of-use systems for human identification
- **CBM^2 submitted an ERC application in 2007** – of ~155 pre-proposals submitted, CBM^2 was selected at one of 34 for full proposal submission; was not selected for site-visit
- **CBM^2 submitting STC in 2008**







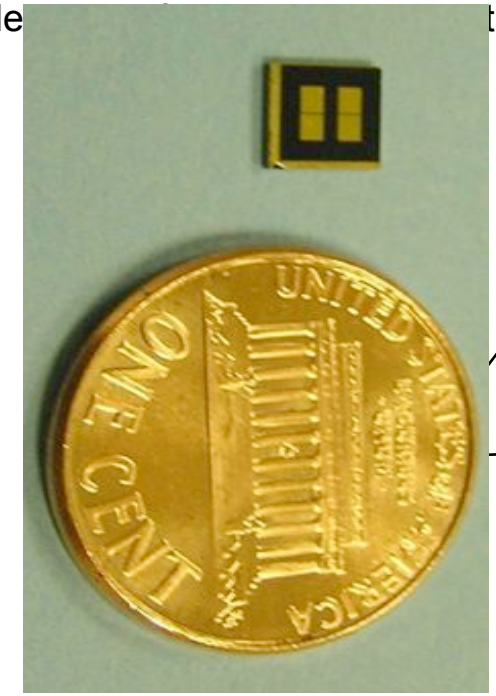


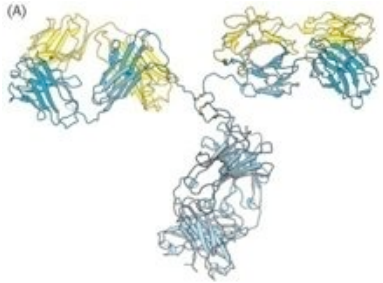
Immunosensors

Ricardo Cortez, Mehnaaz Ali,
Thomas Bishop, Kate Hamlington,
Jerina Pillert, and Mangilal Agarwal

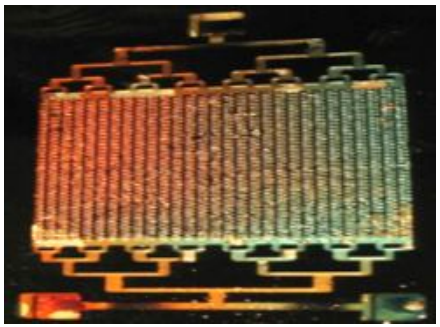
Antibody-based Biosensor

The system will be composed of microfluidic and immunosensing elements (antibodies) targeted for the analysis of biological or chemical agents. **Components**: microfluidic elements for sample pre-processing, nanoporous membranes for target pre-selection and carbon printed ele t.

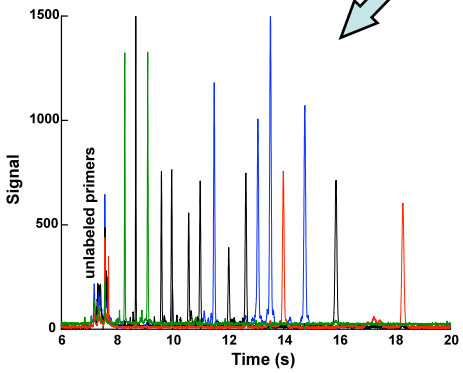




Biochemistry



Fluid flow / reactions in micromixer



Signal

Antibody-based Biosensor

Tulane	LaTech IfM	Xavier	UNO
Jerina Pillert Kate Hamlington Amit Jain Mehnaaz Ali Hank Ashbaugh Tom Bishop Diane Blake Ricardo Cortez Lisa Fauci Don Gaver	Senaka Kanakamedala Jie Liu Mangilal Agarwal Mark DeCoster Ji Fang Yuri Lvov	Robert Blake	Steven Rick

Experiments: characterization of antibodies, determination of assay parameters, preparation and reactivation of Apo-glucose oxidase, synthesis.

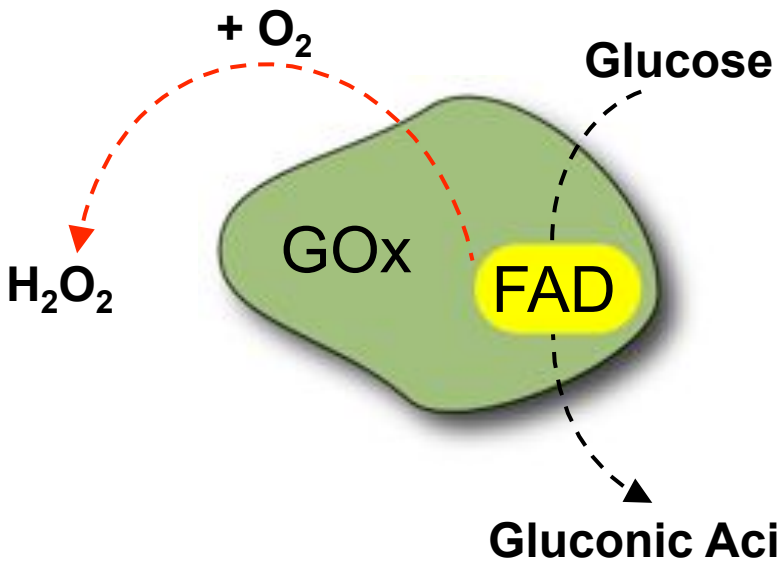
MD Simulations: antigens binding to antibody, energy minimization, loop structures, sequence alignment.

CFD Simulations: flows in microchannels, complex geometry, property optimization, reaction-diffusion-transport of concentrations, parallelization.

Manufacturing: microsensor layer fabrication, micromixer fabrication and evaluation, nanoporous membrane.

The immunosensor will use GOx mediated glucose oxidation for signal transduction

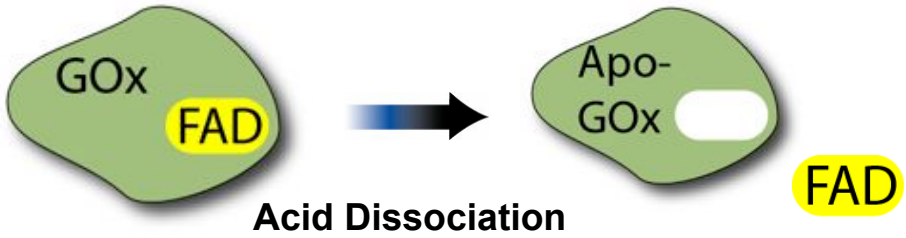
Glucose Oxidation



***Glucose oxidase** requires the cofactor **FAD** for the catalysis of glucose to gluconic acid. This process involves the initial reduction of FAD to $FADH_2$ and consequent oxidation by molecular O_2 generating H_2O_2*

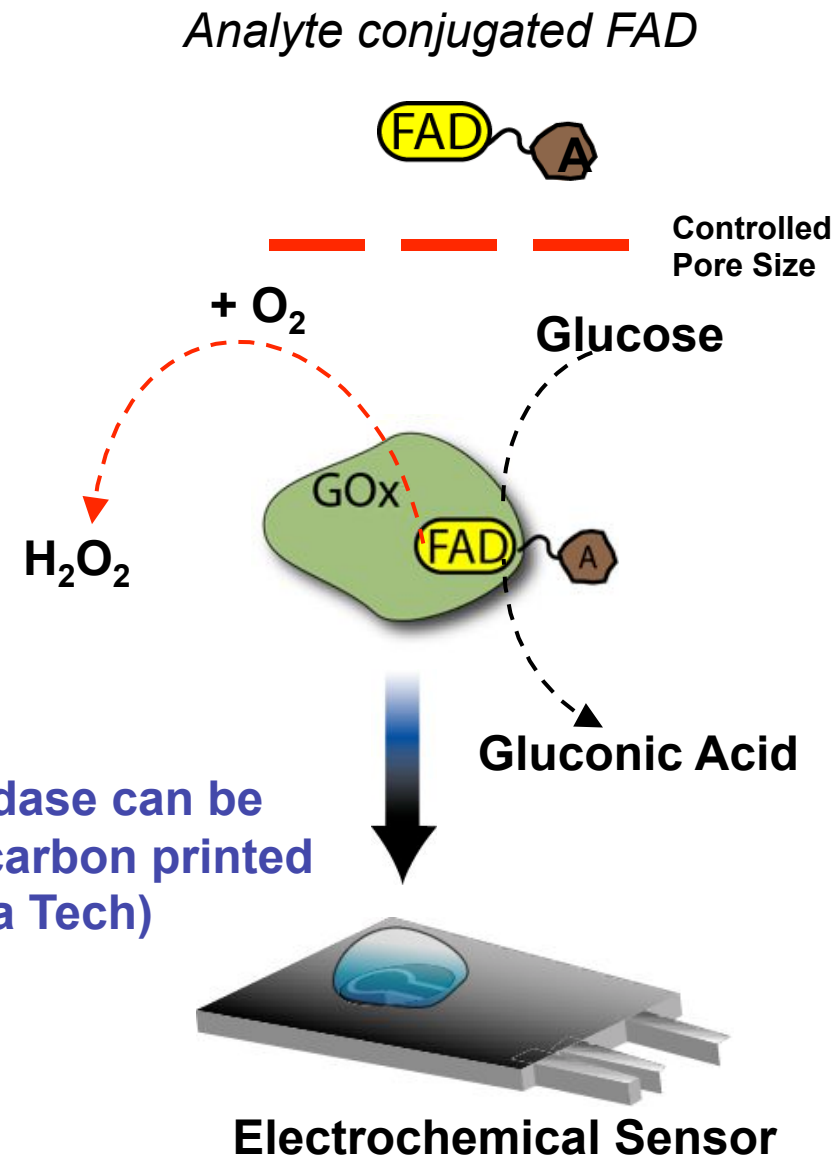
E-Chem Sensor

Enzyme activity** can be **modulated** by the removal and introduction of the cofactor **FAD**. The cofactor can be efficiently **dissociated** under acidic conditions to yield **apo glucose oxidase.



Thus the cofactor FAD can be conjugated to an analyte and utilized to modulate enzyme activity.

General Strategy for E-chem Immunoassay



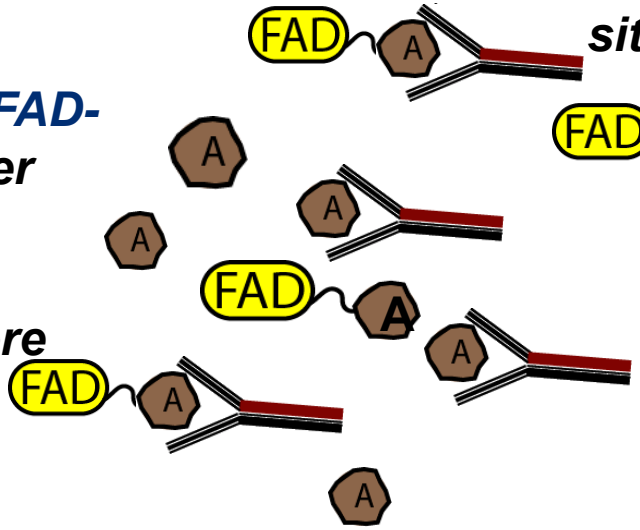
FAD-analyte mediated reactivation of the Apo-GOx in the presence of glucose.

Apo-Glucose oxidase can be coated onto the carbon printed electrode (IFM, La Tech)

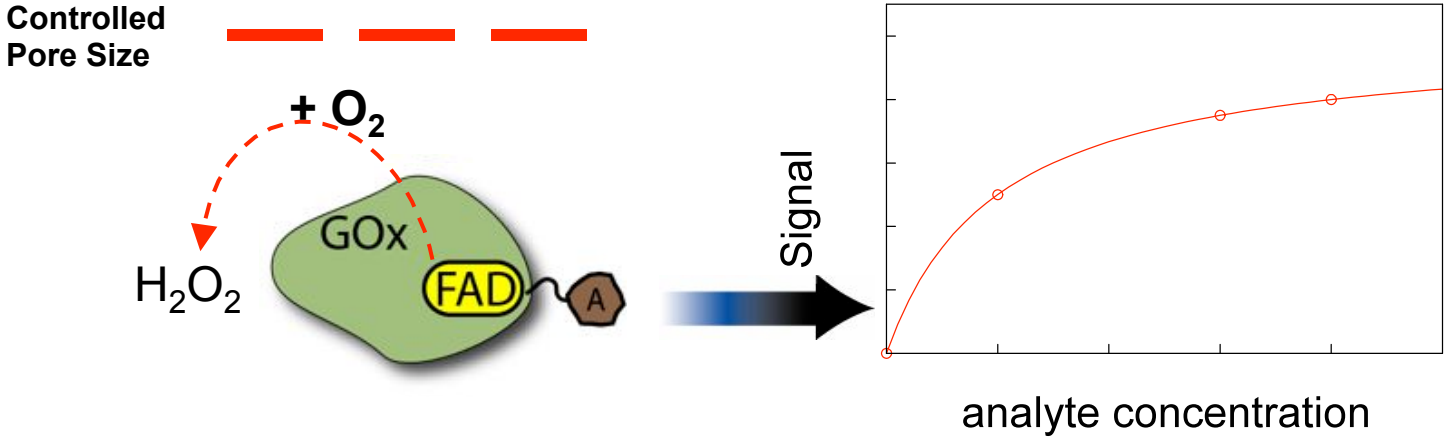
E-Chemical Immunoassay

Addition of analyte from serum or environmental sample competes with the FAD-analyte conjugate for antibody binding sites…

Antibody is mixed with FAD-analyte conjugate. Larger antibody-analyte-FAD conjugate is not able to penetrate the control pore size layer.

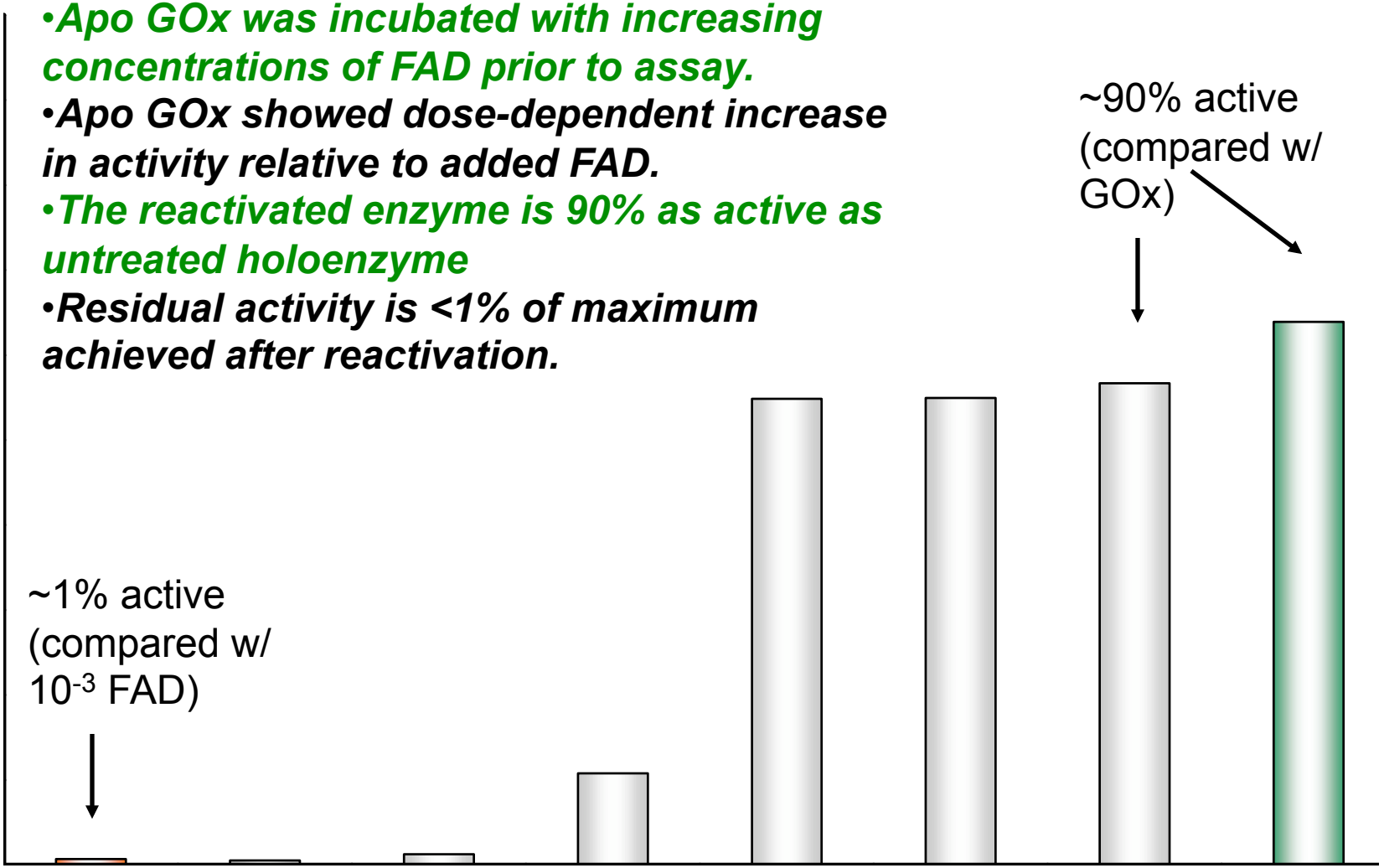


…and thus releases the conjugate. This FAD conjugate is small enough to enter pore and activate the enzyme



Reactivation of Apo-GOx

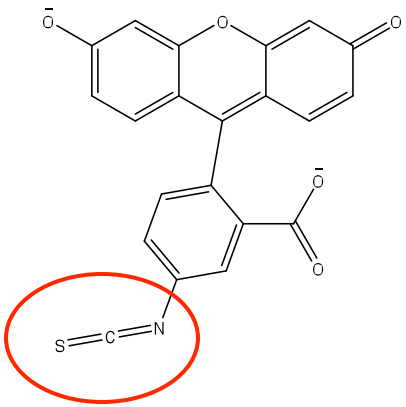
- ***Apo GOx was incubated with increasing concentrations of FAD prior to assay.***
- ***Apo GOx showed dose-dependent increase in activity relative to added FAD.***
- ***The reactivated enzyme is 90% as active as untreated holoenzyme***
- ***Residual activity is <1% of maximum achieved after reactivation.***



Antibody – Analyte Selection

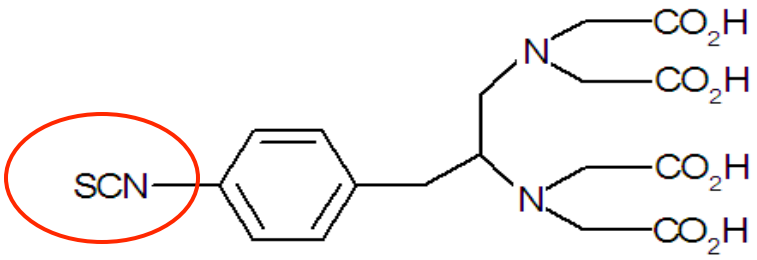
Clone Number	Ligand	K_d (M)	Availability
4-4-20	Fluorescein	1.5×10^{-9}	Invitrogen
M49209	Fluorescein	3.6×10^{-9}	Fitzgerald International
12F6	2,9-dicarboxyl-1,10 phenanthroline (DCP)	7.5×10^{-7}	Blake et al., (2004) *Bioconj. Chem.* **15**:1125.
12F6	UO_2^{2+}-DCP	9.1×10^{-10}	Ibid
4B33	EDTA	1.3×10^{-8}	Blake lab
4B33	Cu^{2+}-EDTA	2.2×10^{-9}	Blake lab

Analytes to be coupled to FAD



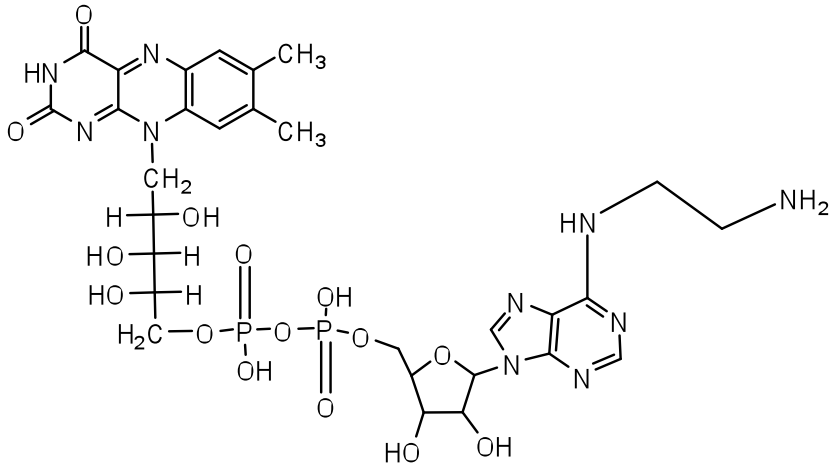
FITC

DCP



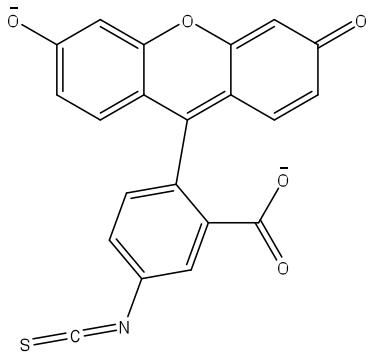
EDTA

Synthesis of FAD Conjugate

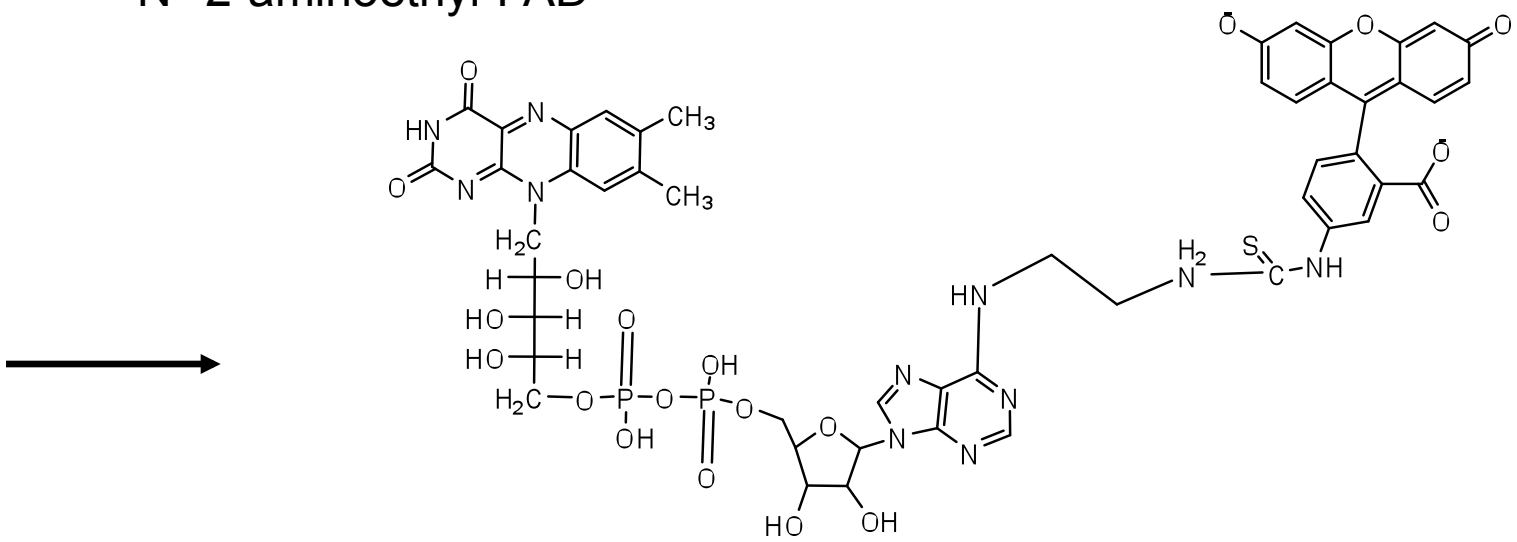


N^6-2-aminoethyl-FAD

+



FITC



FAD-FITC conjugate

Summary

Selection and Characterization of Antibodies

- *Commercial and in-house antibodies have been characterized.*

Apo-Glucose Oxidase has been prepared

- *Change in UV-VIS Spectra >300nm confirmed removal of FAD.*
- *Purification has been optimized to yield high quantity with low residual signal; storage conditions have been developed.*
- *Apo GOx has been transferred to LATech for sensor fabrication.*

FAD mediated Reactivation of Apo-Glucose Oxidase

- *Reactivation of Apo GOx was dependent on FAD concentration.*
- *Reactivated enzyme showed kinetics identical to native GO.*
- *Enzyme activity was not affected by components of the immunoassay.*

Synthesis of primary amine terminated FAD

- *N^6-2-aminoethyl FAD has been synthesized and characterized.*
- *This intermediate was used to synthesize FAD-analyte conjugates.*
- *The apo enzyme could be reactivated with the FAD-FITC conjugate.*
- *A newly synthesized bifunctional crosslinker is also being tested for the preparation of FAD conjugates.*

Antibody Bootcamp for Modelers

Experimental Rotation in Blake Lab

- Ashbaugh graduate student (Jain) spent one and half weeks in Blake lab learning experimental protocols for antibody sensing.

- Titer experiments performed to measure concentrations of antibody 5B2, Pb^{2+}-DTPA-benzyl-BSA conjugate, metal chelator (DTPA), and Pb^{2+}-DTPA.

- Enzyme-Linked ImmunoSorbant Assay (ELISA) used for titer of monoclonal antibody 5B2 and Pb^{2+} conjugate. Competitive inhibition ELISA used to infer the ability of DTPA and Pb^{2+} to bind to 5B2.

The Molecular Modeling Requires

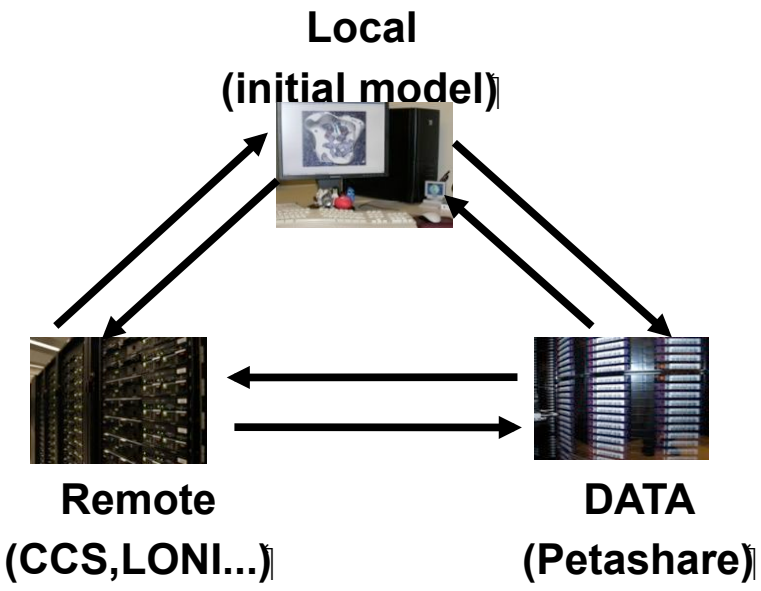
1)Creation of putative antibody models based on sequence; (Modeler)

2)Parameterization of the analytes that bind to the antibodies; (Gaussian)

3)Docking analytes in different potential antibody binding loops; (PackMol)

4)Optimization of the antibody-analyte interaction by *in silico point mutations (Methods under development,* ***REDS****)*

.



Biosensors: Computational Aspects MD

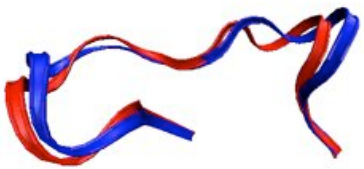
Simulations of 5B2 loop region (Test Cases)

- Binding of antigens to antibody occurs in loop domain. Aim to identify using simulations side chains in loop region that contribute to binding specificity to guide antibody engineering.

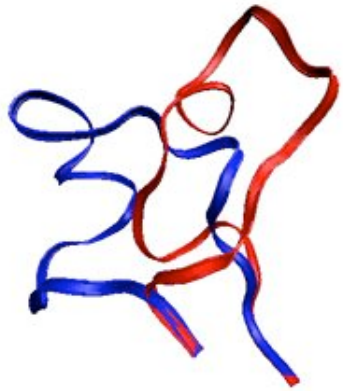
- In vacuo energy minimizations of 5B2 LC and HC loops confirm previous identification of metal binding residue Lys^{58}.

- Replica Exchange Molecular Dynamic performed of 5B2 in vacuo and implicit solvent to generate families of loop structures for minimization to determine robustness of predictions and identify spatial and dynamic correlations between key binding residues

- Initial findings: HC3 loop has more varied and flexible structure than the other five antibody loops



LC1



HC3

Biosensors: Computational Aspects MD

Simulations of 5B2 loop region (continued II)

- Replica Exchange (REMD):
  replica: several simultaneous simulations
  2 levels of parallelization
  exchange: simulations swap information

Simulation Characteristics

Loops

~1000 atoms => 2CPUS/sim
10ns run time => Gb's data;
Full REMD in 24hrs 64CPUS

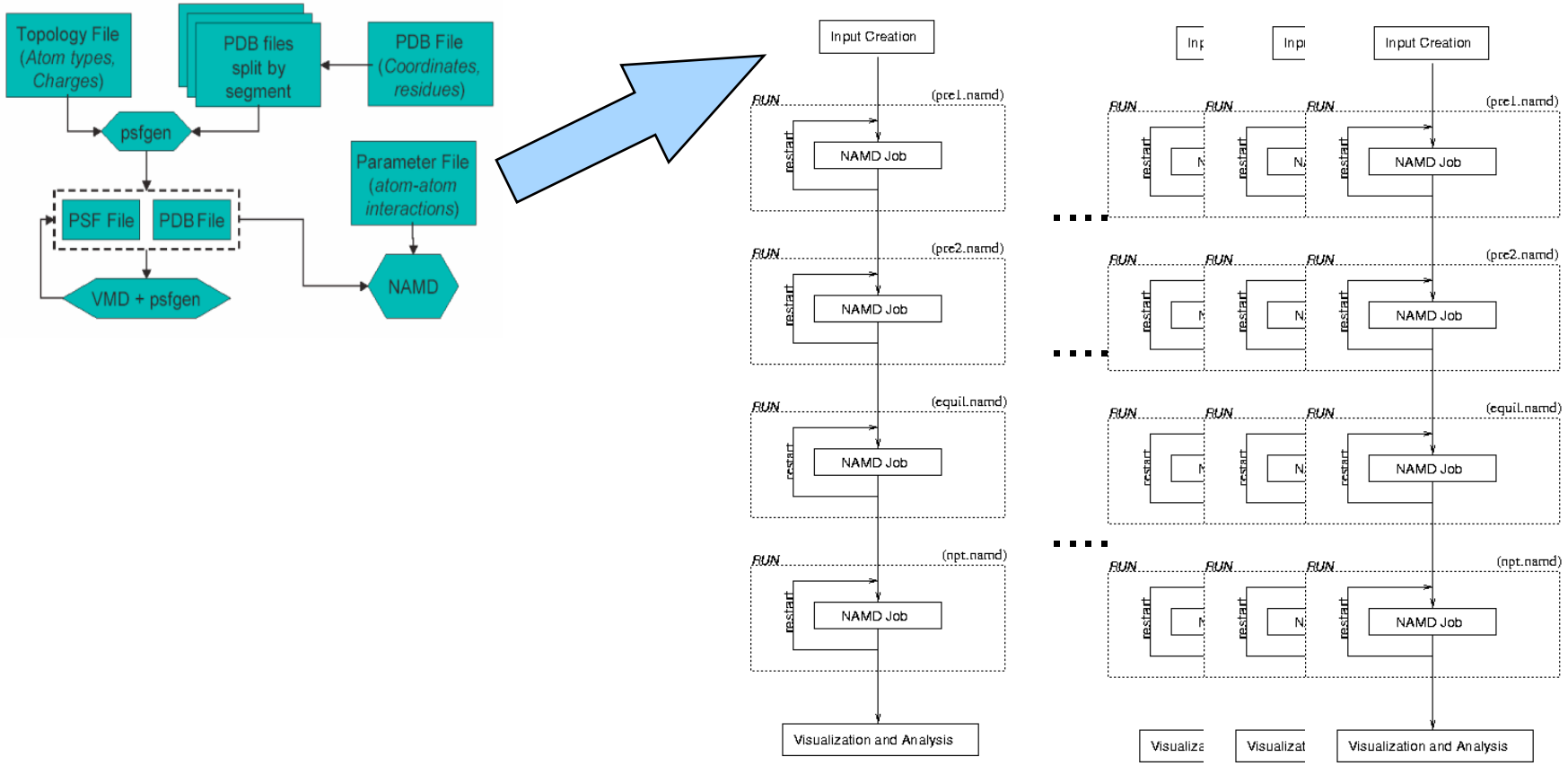
Full System

~10,000atoms => 4CPUS/sim
10ns run time => 10-100Gb data;
Full REMD in 2wks 64CPUS

Biosensors: MD Fast Track Study

A high throughput simulation workflow

- Bishop (CCS @ TU)
- Emir Embahsi & Tevik Kosar (CCT @ LSU)



CyberTools Connections

WP 1: Scheduling and Data Services.

The details of integrating our Molecular Modeling packages into WP 1 are being addressed by Drs. Thomas Bishop (Tulane) and Tevfik Kosar (LSU).

WP 2: Information Services and Portals.

Drs. Thomas Bishop and Tevfik Kosar are collaborating to bring Bishop's DNA folding simulations on-line. The Workflow resulting from this effort can be readily modified to investigate the antibody and analyte interactions.

WP 3: Visualization Services.

Work is in progress to create modules that will permit all scientists involved in the project to visualize molecular models and other results via a common user interface without the necessity of transferring data or installing software on local lab computers.

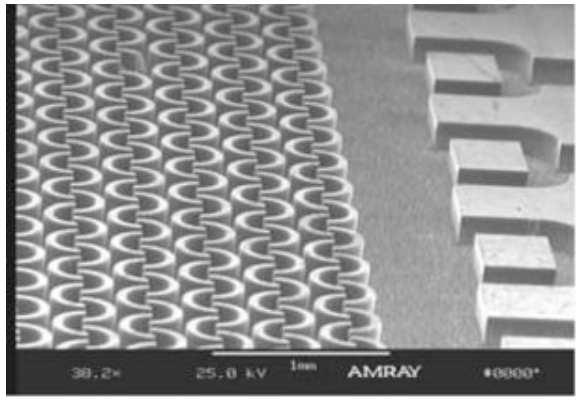
WP 4: Application Services and Toolkits.

Drs. Steven Rick (UNO) and Henry Ashbaugh are developing replica simulation techniques that will enable this group to efficiently identify antibody loop sequences that optimize the antibody-analyte interactions.

Fluid Mechanics and Transport

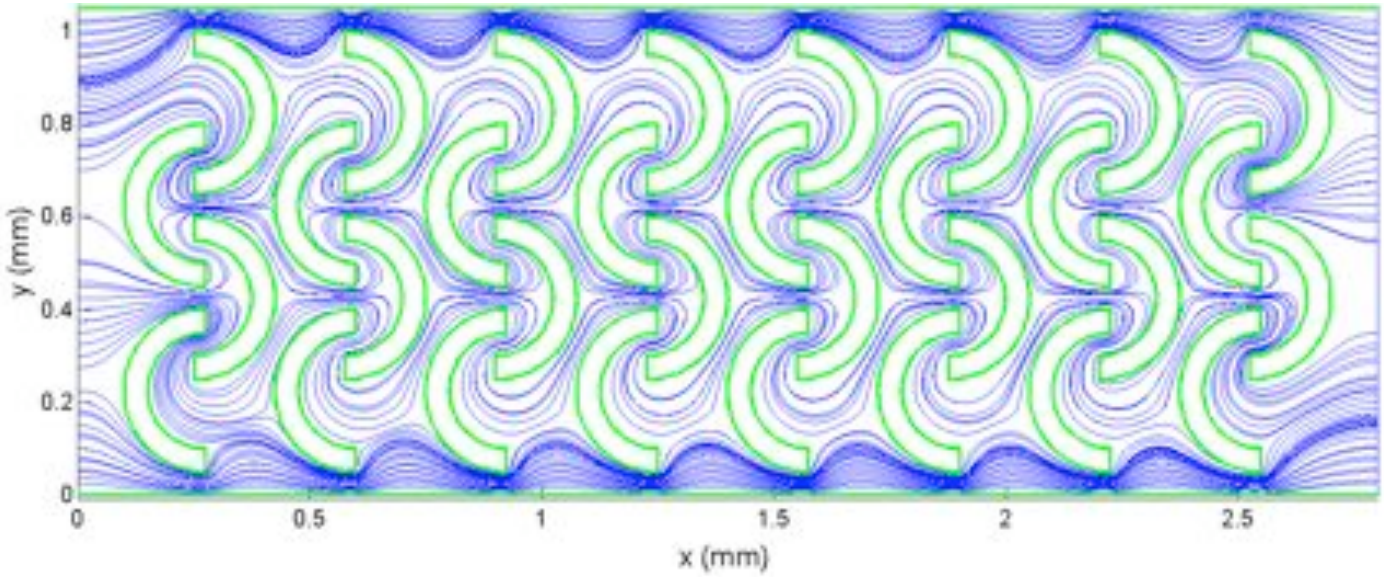
- **GOAL** ➔ *Computationally determine the optimal geometric configuration of the omega channel network to enhance mixing of two species.*
- Laminar flow field governed by continuity & Stokes equations: $\nabla P = \mu \nabla^2 \mathbf{u}$

  $\nabla \cdot \mathbf{u} = 0$
- Boundary Element Method determines velocities and surface stresses

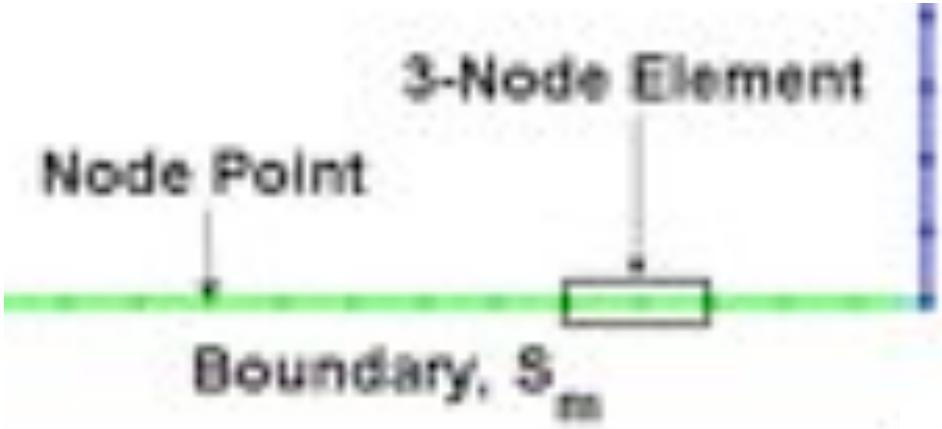


Omega channels developed by IfM

Streamlines resulting from constant pressure drop across model channel



Boundary Element Method



- Velocity **u** and stress τ are approximated as quadratic polynomials, and at each node point, satisfy

$$\mathbf{C}_{ki}u_i(\mathbf{x}) + \sum_{m=1}^{N}\int_{s_m}\mathbf{T}_{ik}(\mathbf{x},\mathbf{y})u_i(\mathbf{y})dS_m = \frac{1}{\mu}\sum_{m=1}^{N}\int_{s_m}\mathbf{U}_{ik}(\mathbf{x},\mathbf{y})\tau_i(\mathbf{y})dS_m$$

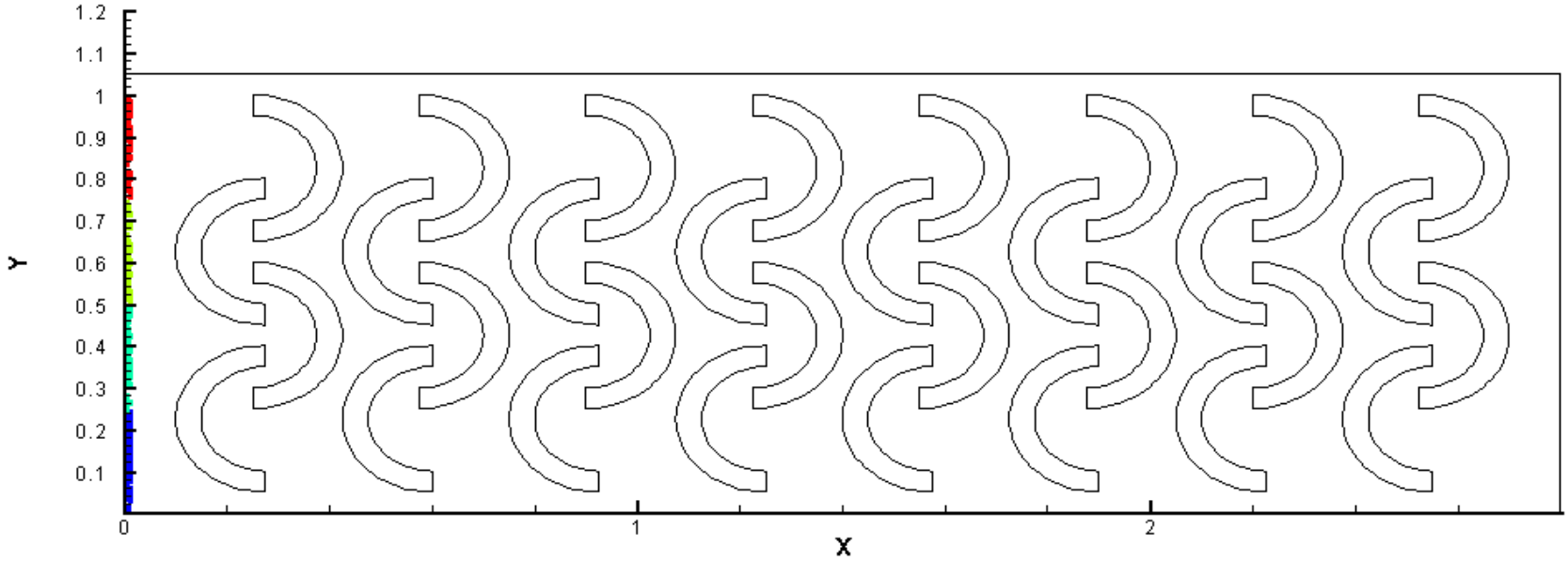
- Integral equation is expressed as system of linear equations:

$$H\mathbf{u} = G\tau$$

- Elements of H and G computed using Gaussian quadrature rules
- Optimization of simulation is being developed in conjunction with WP4 and will create a general purpose *CyberTool*.

Results: Particle Trajectory

For more info see our poster!



- Particles initially positioned along y-axis at x = 0
- Path of each particle traced as it flows through the domain
- Note inner particles travel more slowly than outer particles that migrate quickly across channel along outer walls
- Results suggest domain modification is important to improve mixing

Microsensor Mixing and Transport

- Analyte-FAD conjugate and analyte from serum compete to bind with antibody

- Binding and release occur spontaneously as analytes and antibody are transported by fluid motion

Each analyte/antibody satisfy a reaction-diffusion equation:

$$\frac{\partial C_i}{\partial t} + \nabla \cdot (\mathbf{u} C_i) = D_i \nabla^2 C_i + R_i$$

$\mathbf{u}$ – fluid velocity

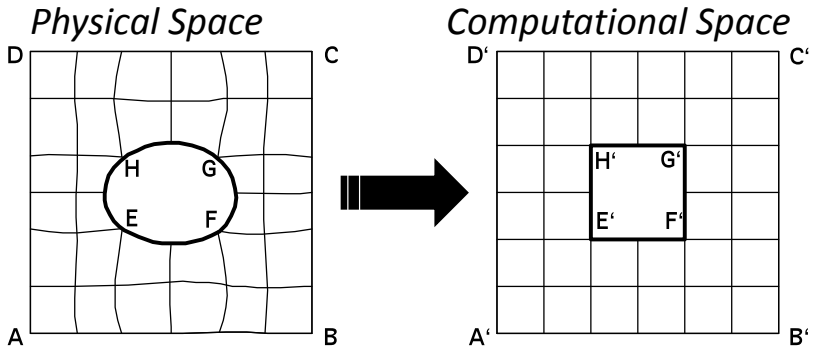
C_i – concentration of each species

D_i – diffusion coefficient

$R_i(C)$ – reaction term

Transport Methodology

- Transform equations into a boundary-fitted coordinate system

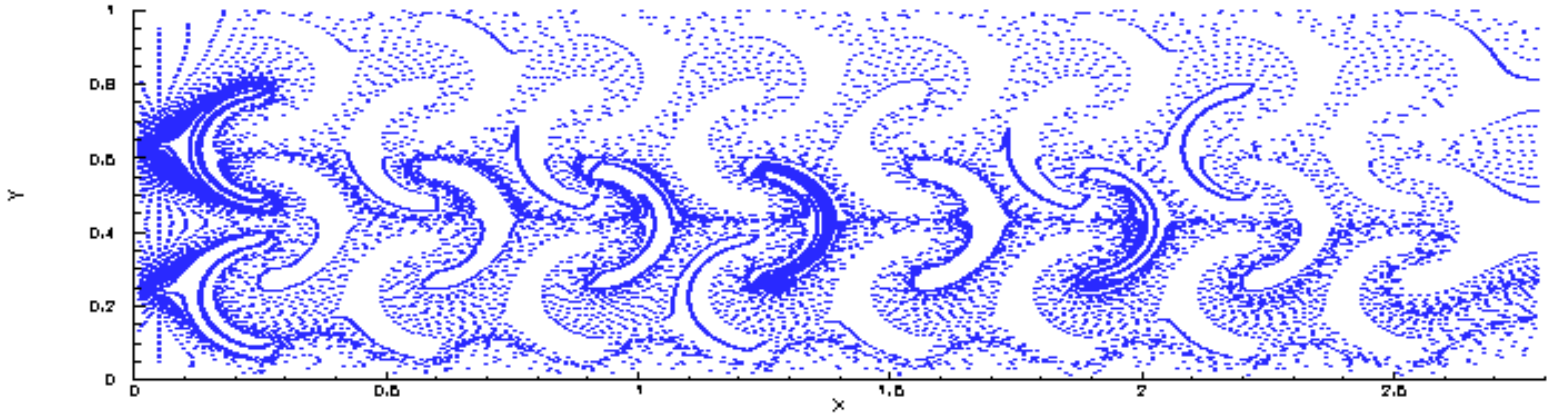


- Use the Finite Volume Method to solve for *concentration*
- Note: velocity field obtained from BEM code
- Multiblock approach for omega channels
- Working with Dr. Blake for reaction/diffusion rates

CyberTools Connections

Current Work:

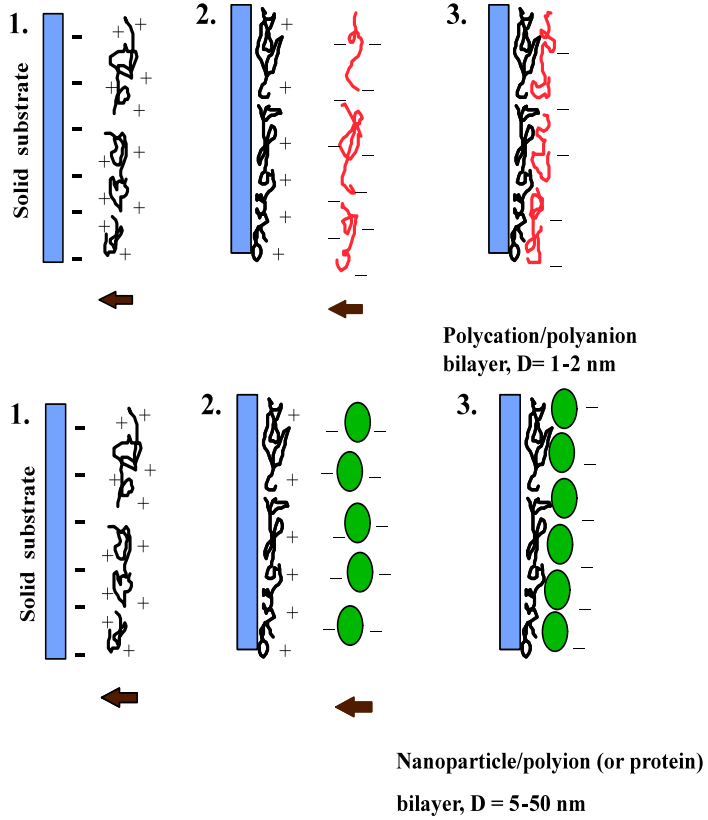
- Parallelization of Stokes flow problem for use in the HPC environment (WP4: Mayank Tyagi, Shantenu Jha, Sanjay Kodiyalam)
  - *OpenMP*
- Visualization of model problem using TecPlot (with WP3)
- Generalization of code to develop a *CyberTool* package that solves Stokes flow equations



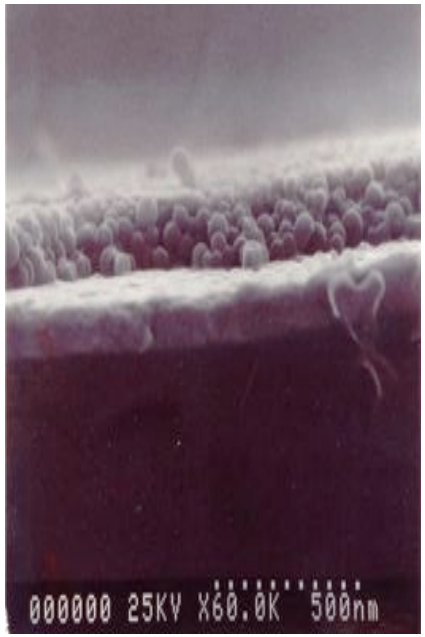
Future Work:

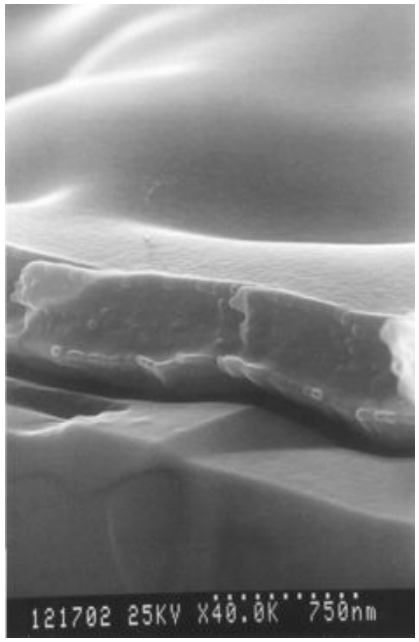
- Parallelization of source code including transport

Layer-by-Layer (LbL) Nanoporous Membrane for Immunoassay (sensor): technology for enzyme deposition



Scheme of the layer-by-layer nanoassembly by alternate adsorption of polycations and polyanions or nanoparticles





SEM cross-section images of (glucose oxidase/PAH)$_{22}$ multilayer on quartz (left), and (40 nm silica/PAH)$_6$ film on silver electrode (right).

Polymer-based Electronic Microsensor Fabrication

Silicon wafer with oxide layer

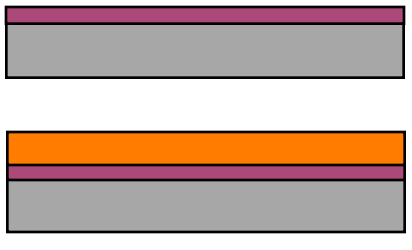
Spin coat PR 1813 resist layer

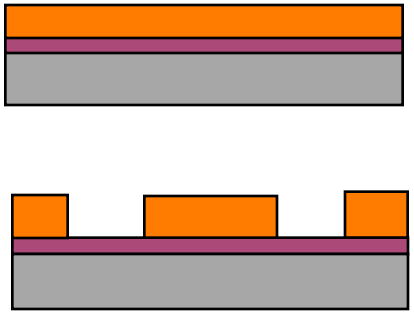
Pattern PR 1813 resist

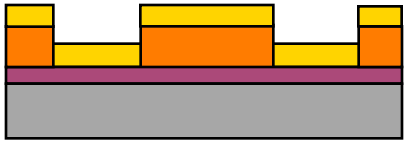
Sputter gold electrodes

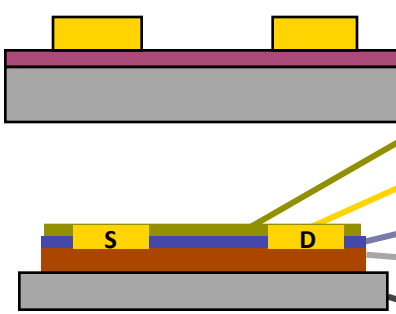
Lift-off PR 1813 resist

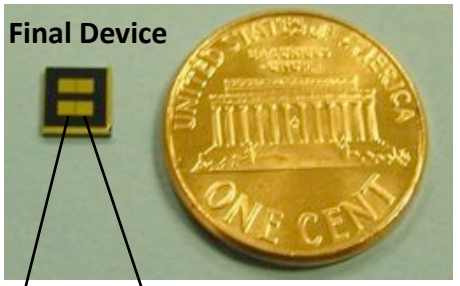
Final Device

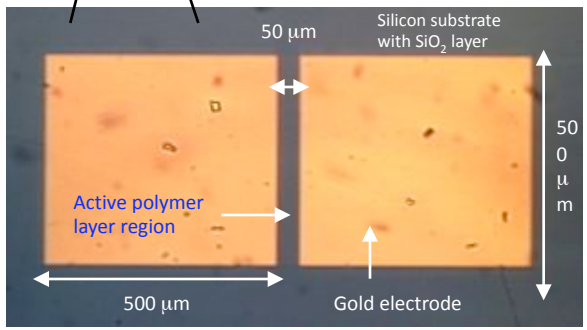


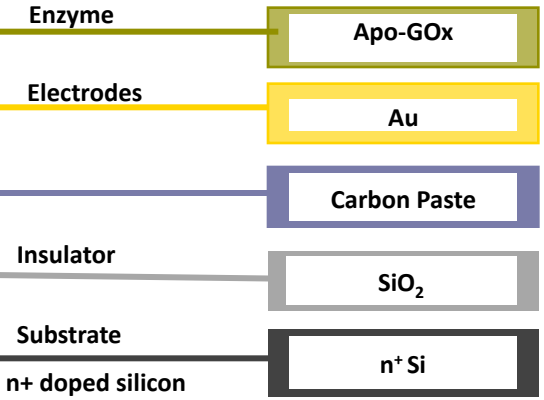












Micromixer Fabrication

Silicon wafer
with oxide layer

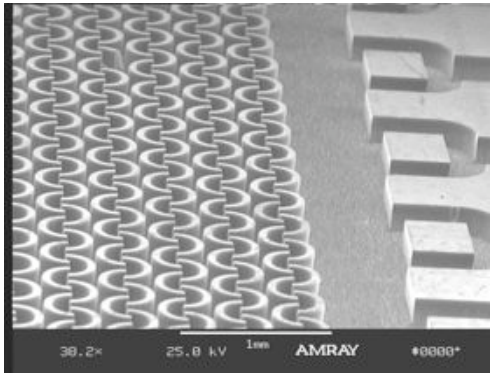
Spin coat
PR 1813 resist layer

Pattern
PR 1813 resist

Etch SiO_2

Etch Silion

Etch SiO2 & Bond
Top Glass Substrate



SEM → Omega Channel Micromixer

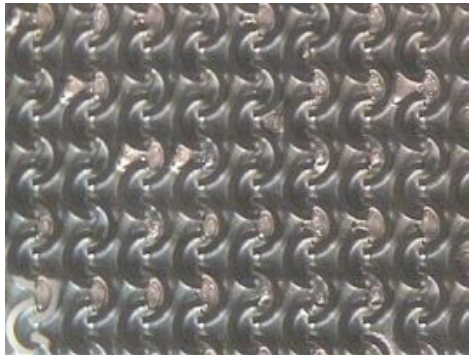
Fabrication
- Lithography
- ICP
- Bonding

Challenges
- Connectors

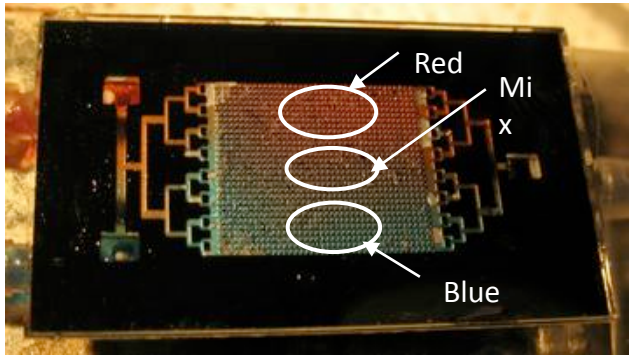
Modifications
- New set of connectors from Upchurch Scientific are being tested and evaluated

Micromixer Evaluation

Straight Channel



Omega Channel

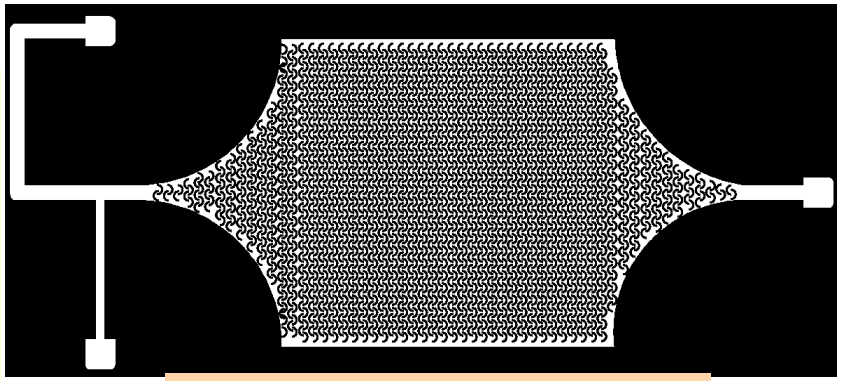


Micromixer

Challenges
- Laminar Flow
- Mixing only at the center of the device

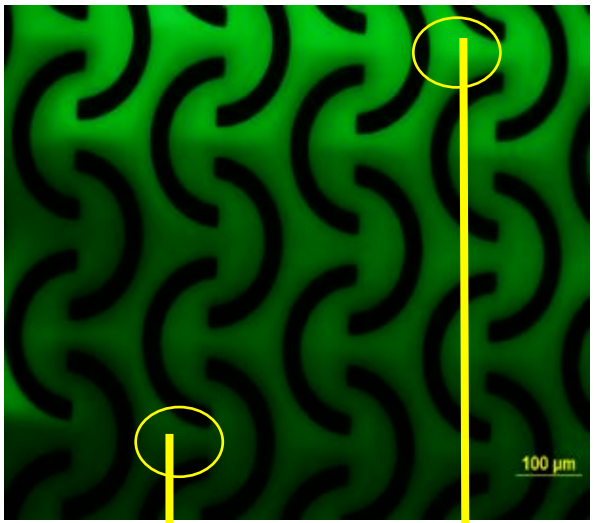
Modifications
- Designed ‘T’ shape inlet and outlet for initial mixing

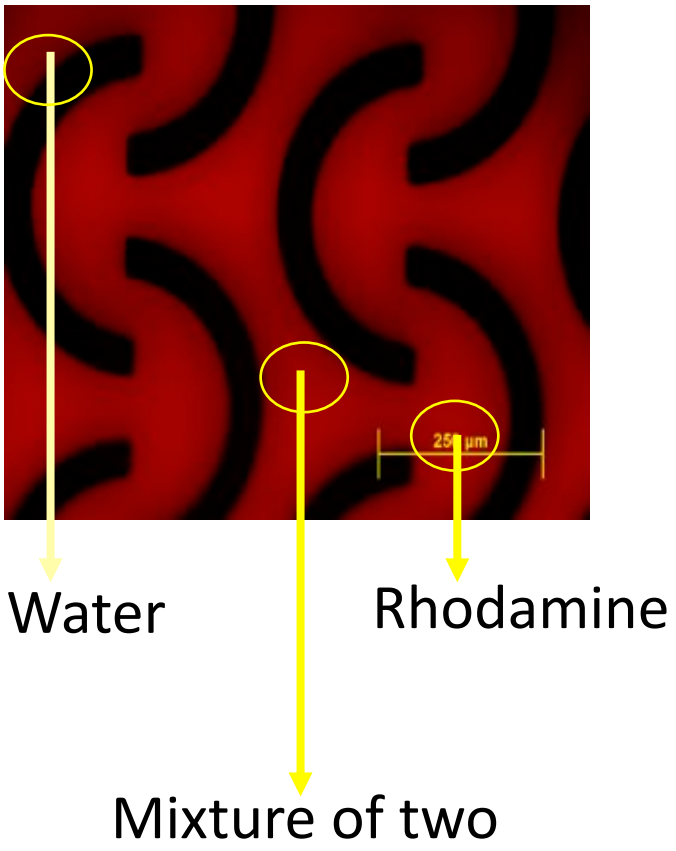
Quantification
- Image analysis software
- Using fluorescent dyes for better signal/noise



New Micromixer Design

Fluorescent dyes in the omega channels: VISUALIZED WITH MICROSCOPY AND DIGITAL CAMERA





Summary and Conclusions

- Microfluidic Component
  - Fabricated and evaluated two sets of micromixers
  - Designed new micromixer based on the results obtained (fabricated)
- Nanoporous Membrane
  - LbL nanoassembly is being evaluated for fabricating nanoporous membrane; ***New:*** in collaboration with Dr. Scott Gold (Louisiana Tech.) for modeling of flow through porous membrane.
- Reproducibility
  - Evaluating PEDOT and carbon nanotube based microsensor for reproducibility, Selectivity, and Life Time
- Testing Procedures
  - Currently testing florescent dyes and particles (proposed) for evaluating micromixers.
  - Currently evaluating microscale sensor system based on carbon nanotubes
  - Carbon-based Electrodes
  - Under testing and fabrication
  - *CyberTools* Connection
  - Access Grid (AG) video conference with Tulane (23 July 2008): simulations and experimental data. Evaluation of Cybertools link to visualization software: VisIt 1.9.1 (Windows version, DeCoster) thru the **Cactus Code** Link.
  - *VISIT OUR GRADUATE STUDENT POSTER! -- SENAKA KANAKAMEDALA-*

Final Remarks

Interactions between the laboratory and the molecular dynamics groups

•Provided new models of protein structure that allow testing of hypotheses in silico prior to time-consuming laboratory experiments.

•Validated *in silico* predictions based on laboratory experiments.

•Identified methodological refinements based on laboratory results.

Interactions between the laboratory and the micromanufacturing groups

•Broadened the kinds of hardware and electronics that can used to construct the sensors.

•Prompted micromanufacturers to examine paradigms used to validate their devices (e.g. they have modified the molecular identity and concentration ranges of reagents used to test their microscale mixers).

Interactions between the fluid mechanics and the micromanufacturing groups

•Provided data to determine boundary conditions used in the simulations.

•Established realistic geometries for fluid flow domains.

•Delineated a plan to reduce the number of fabrication trials needed to optimize the final device.

Science Driver: Bio-Transport Computations

Computing of Transport Processes in Biological Systems

S. Acharya[1,2] (Lead), D.Moldovan[1], R. Devireddy[1],

D. Nikitopoulos[1], A. Gilmanov[1,2]

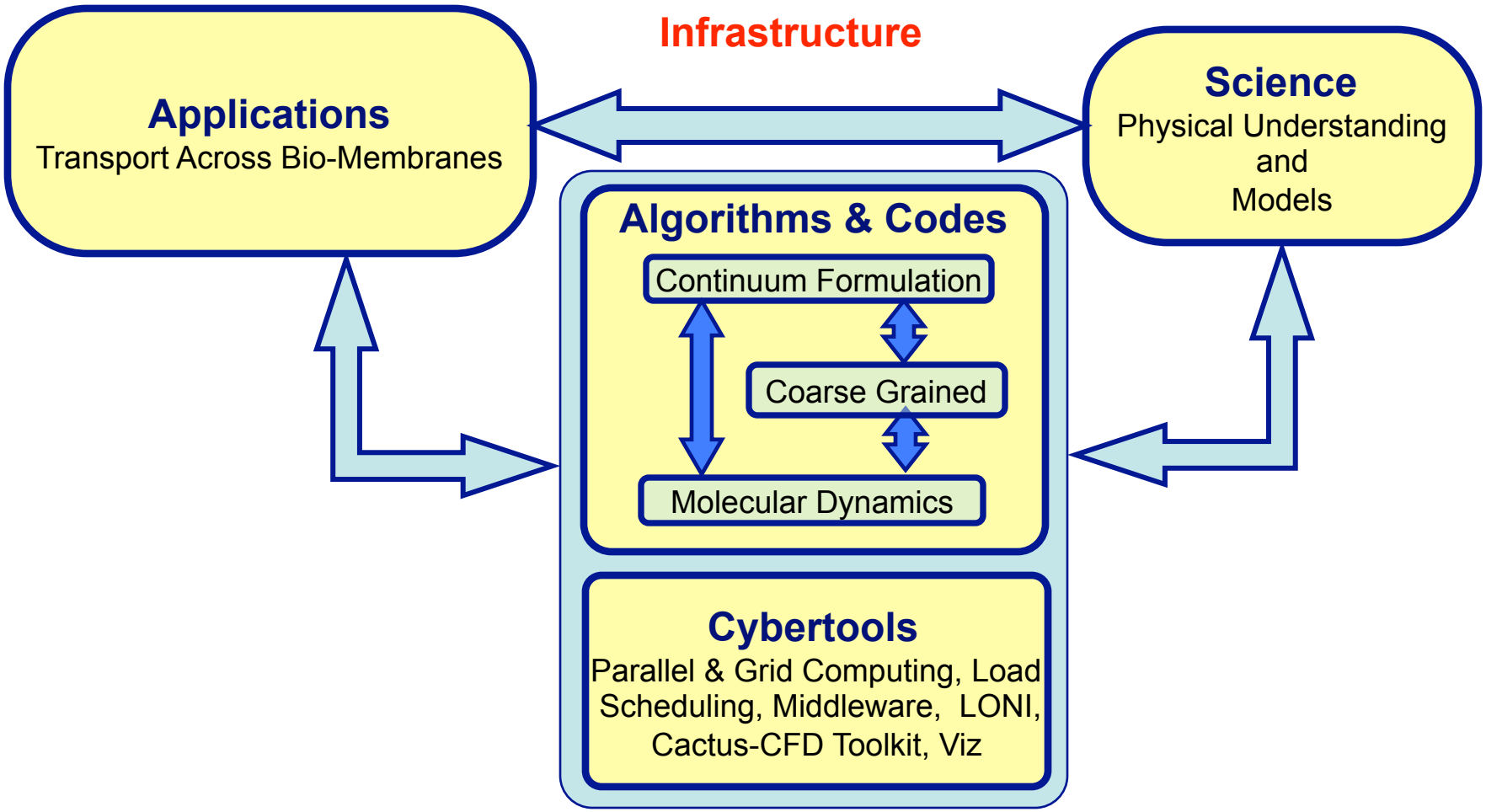
Louisiana State University

[1]Mechanical Engineering Department

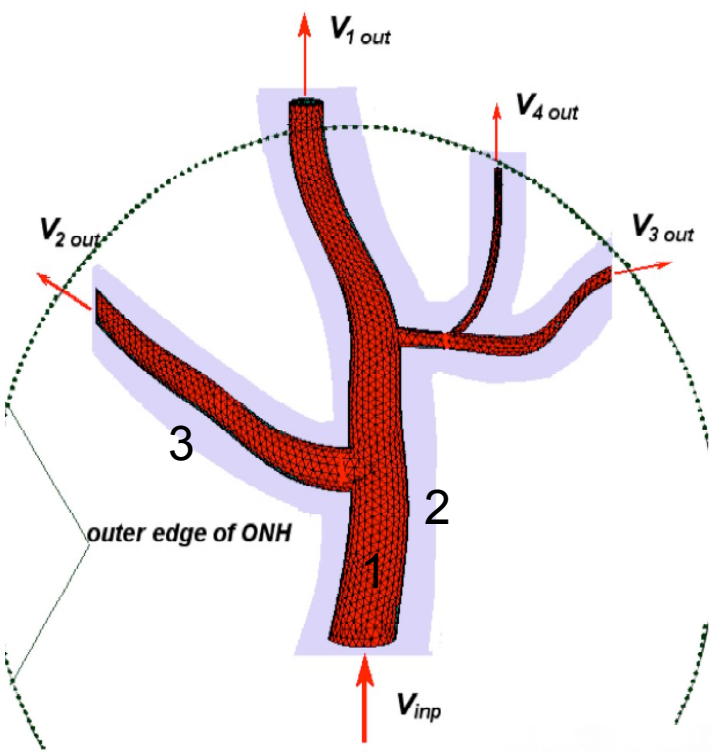
[2]Center for Computation and Technology

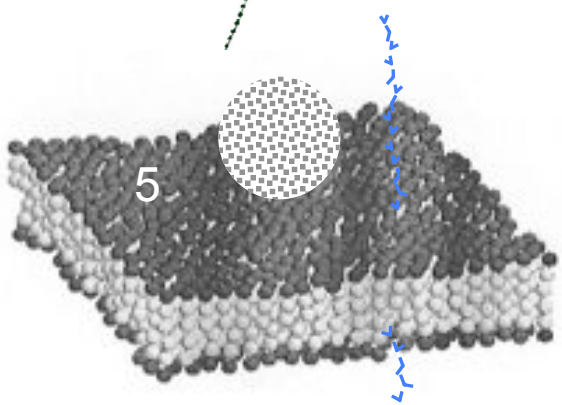
Graduate Students: R. Alapati, P. Kalghatgi, T. Gilmanov

Support from the NSF EPSCoR Program & the LA-BOR Is gratefully acknowledged


Infrastructure
Applications
Transport Across Bio-Membranes
Science
Physical Understanding
and
Models
Algorithms & Codes
Continuum Formulation
Coarse Grained
Molecular Dynamics
Cybertools
Parallel & Grid Computing, Load
Scheduling, Middleware, LONI,
Cactus-CFD Toolkit, Viz

- Prediction and understanding of oxygen transport in biological systems
  1. Continuum flow in larger vessels-Navier Stokes
  2. Porous media transport across vessel walls & tissues-Brinkmann
  3. Structural deformation of vessels/tissues-
  4. Particle flow in capillaries-Lattice-Boltzmann
  5. Atomistic transport across cellular interfaces-Molecular Dynamics
  6. Upscaling from atomistic to continuum





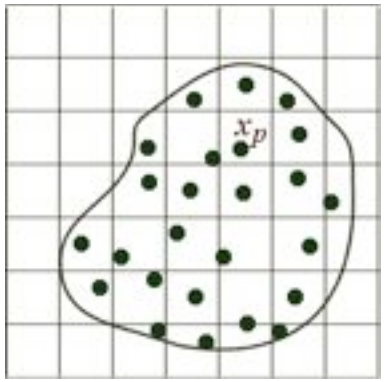
- Development of computationally efficient numerical methods or algorithms needed for biological transport calculations
  - ✓ Structural calculation using a meshless particle method
  - ✓ Flow-Structure Interaction (FSI) methodology using Immersed Boundary Method (IBM)

  → Year 1
- Contributing to improved science-understanding of small molecule flow/transport physics under asymmetric concentrations and applied stresses
  - ✓ Asymmetric calculations of molecule/particle transport across lipid bilayers

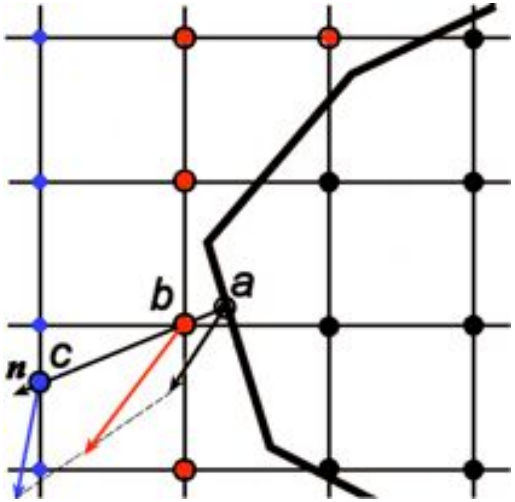
  → Year 1
- Contributing to improved computational infrastructure- collaborating with the cybertools group responsible for developing the CFD toolkit
  - ✓ Development of cactus-compatible routines for transport and flow calculations
  - ✓ Validation Studies

  → Year 1
- Contributing to improved science-understanding of oxygen flow/ transport physics under elevated pressures

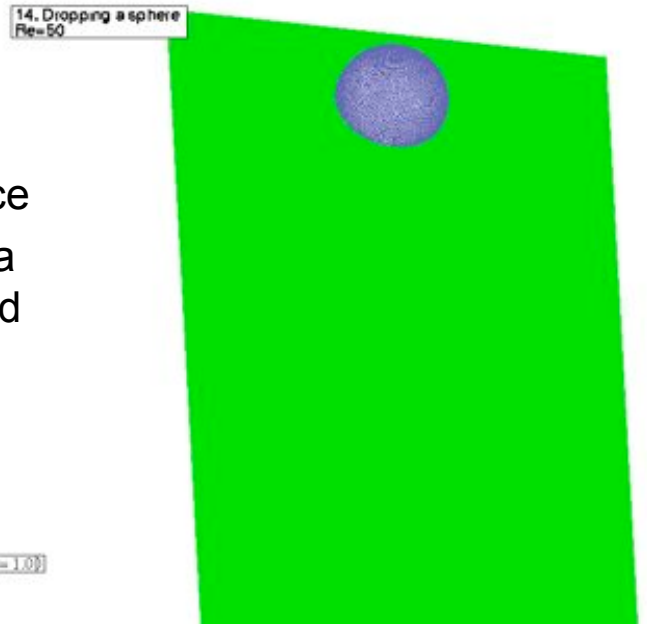
- Continuum flow and transport calculations
  - Multiblock structured grid with continuous grid lines across block interfaces
  - Fractional step algorithm with staggered grid locations for the velocity (stored at cell faces)
  - Pressure-poisson equation for pressure
  - Consistent second order differencing for diffusion and pressure terms and upwind biased differencing for the convective terms
  - Explicit and implicit second order temporal differencing
  - Flow-structure interaction
    - Particle-based meshless calculations for structural deformations (called material point method-MPM)
  - Immersed Boundary Methodology (IBM) for resolving boundary conditions along moving interfacial surfaces
  - Flow-Structure Interaction for Biosystems

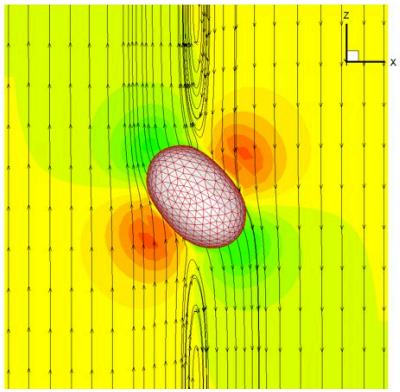


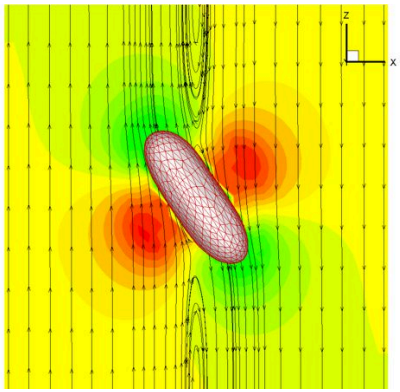
Background grid for solution of momentum equations

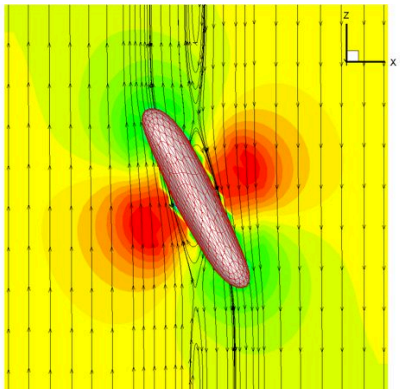


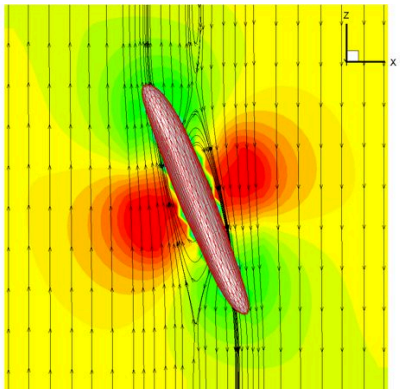
- Material-Point Method (MPM) for structural deformations
  - Arbitrary distribution of points on the solid body/surface
  - Material points are solved (deformation & stress) on a background grid that is independent from the fluid grid
  - Flow-structure coupling through boundary/interface conditions
  - Flow around deforming surface handled through IBM









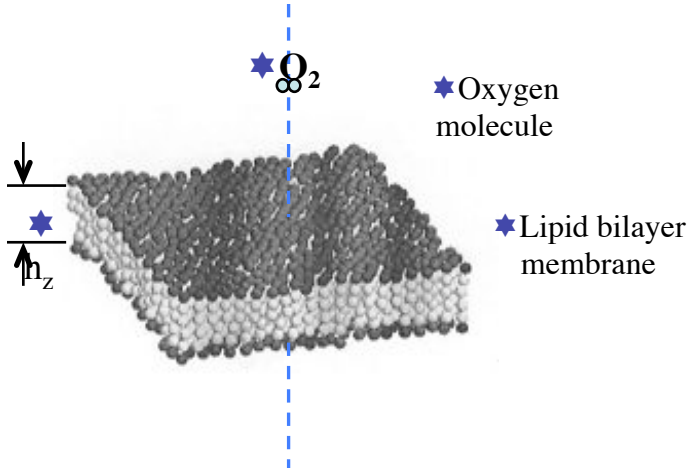


- Collaborating with the WP4 group for the development of a CFD Toolkit;
- Finite volume, multi block;
- Data array structure consistent with current structure in Cactus;
- Multi-block grid from commercial grid generators;
- Baseline code developed for laminar flow; several benchmarks being run to provide WP4 input-output files for Toolkit verification and validation;
- Long term plans are to transition to the Toolkit for the biosystems transport simulation;

- Implemented suggestions for improved performance of parallel code—seen improvements
- Discussions ongoing with Viz groups to get better access to better visualization codes (WP3)
- Discussions ongoing on use of a Lattice Boltzmann code for particle simulations
- Discussions ongoing on most effective ways of doing CFD-MD coupling

- Diffusion rate and permeability coefficients across vessel walls and tissues for different conditions are generally not known reliably (difficulty in in situ measurements)

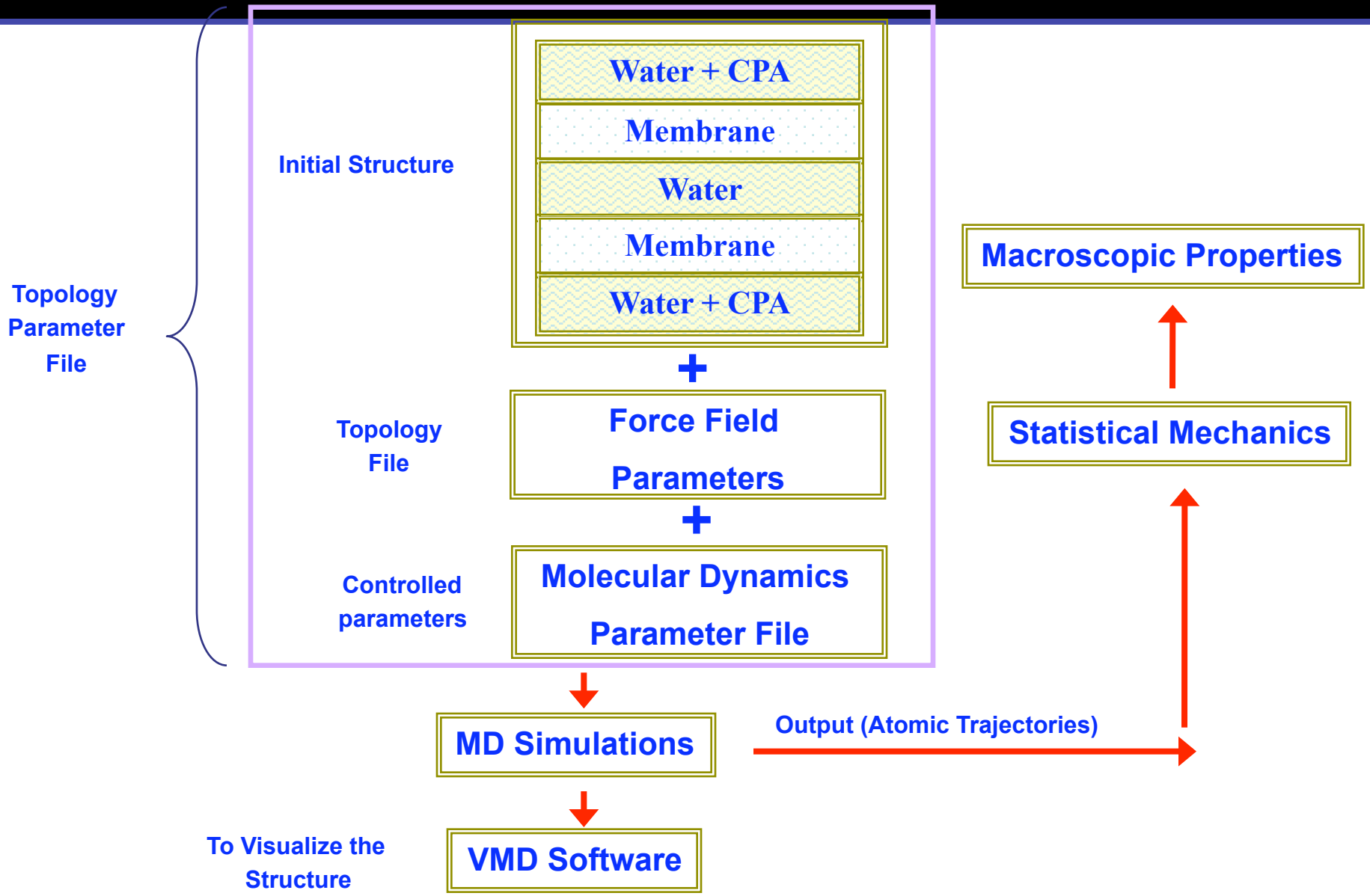
- Specifically designed MD simulations under different conditions can provide:
  - atomistic insight and molecular mechanism underlying the transport of O_2 across a lipid bilayer membrane in order to determine which details are important for the permeation process.
  - Derive the oxygen diffusivities, D_{O2}, inside the inhomogeneous region of a lipid bilayer.
  - Derive permeation rates, P_{O2}, indirectly via computation of the free energy and diffusion rate profiles of a O_2 molecule across the lipid bilayer.



$$P = \frac{1}{\int_{z_1}^{z_2} \frac{\exp(\Delta G(z)/RT)}{D(z)} dz}$$

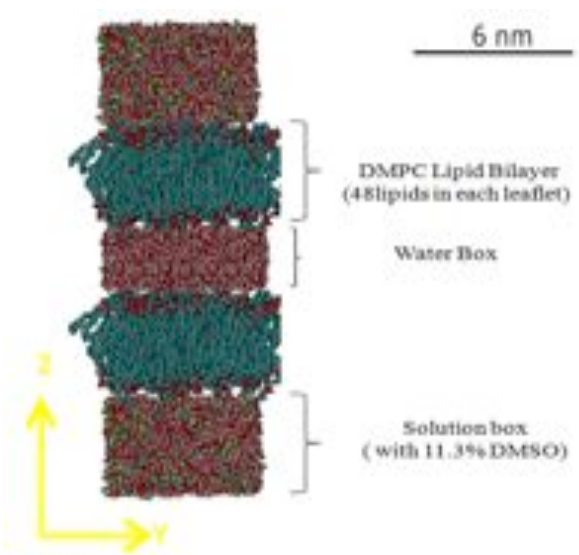
$$D = \frac{1}{3}\int_0^{\infty} \langle v(0) \cdot v(t) \rangle dt$$

MD Simulation Using "GROMACS"

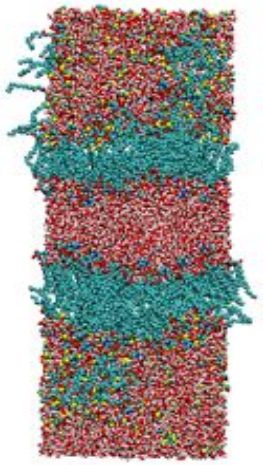


Structural changes in Lipid Bilayers

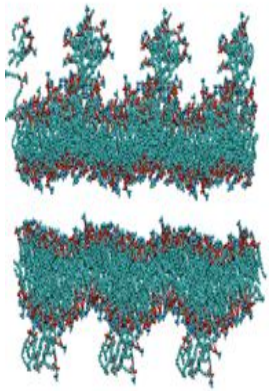
Initial system



(a) Whole system

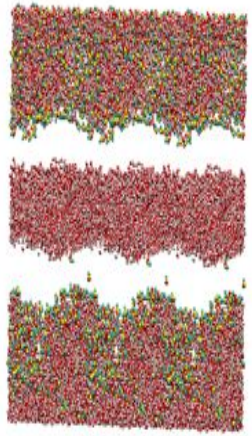


(b) Lipids only



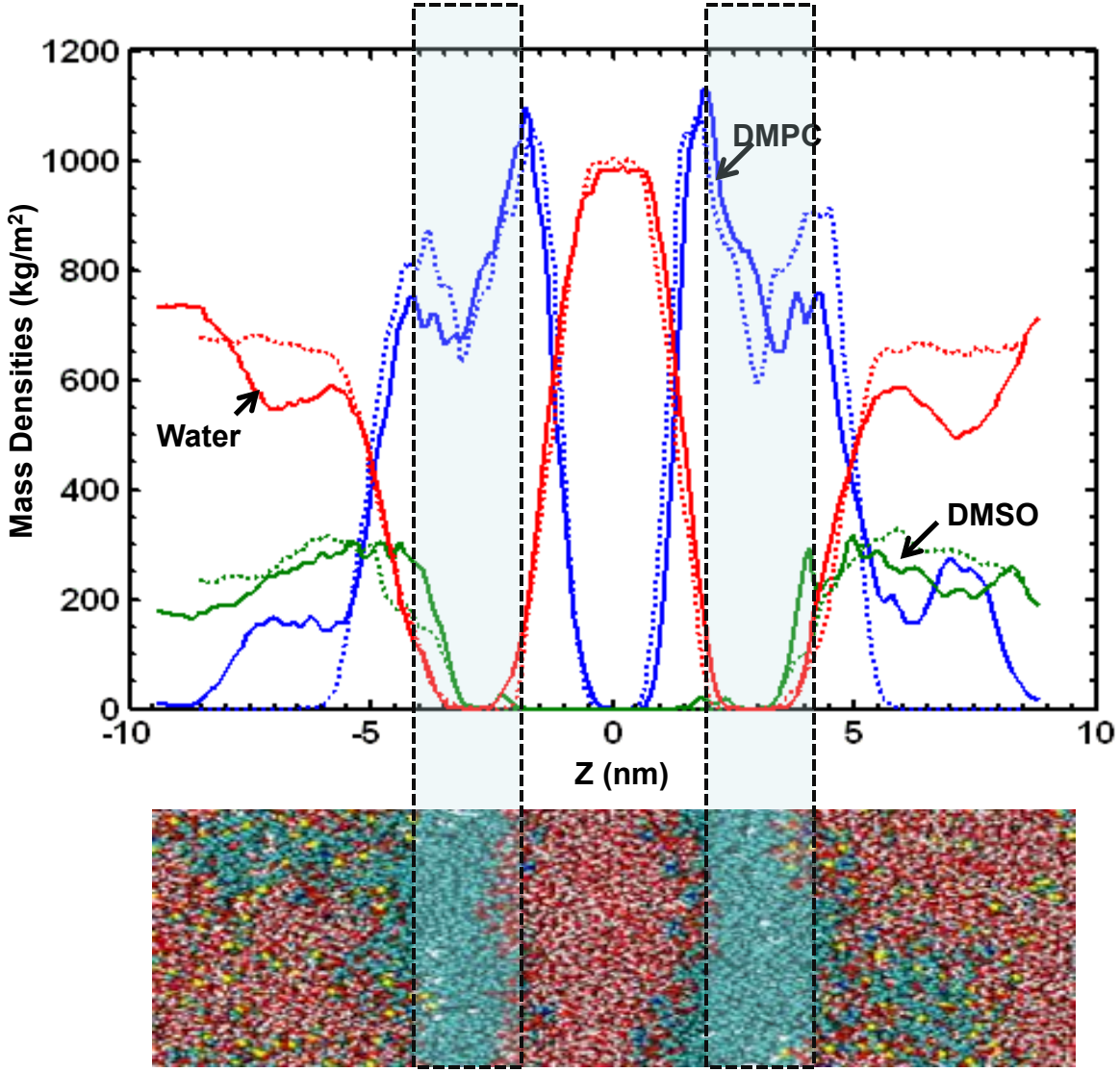
t=50ns

(c) Solution only



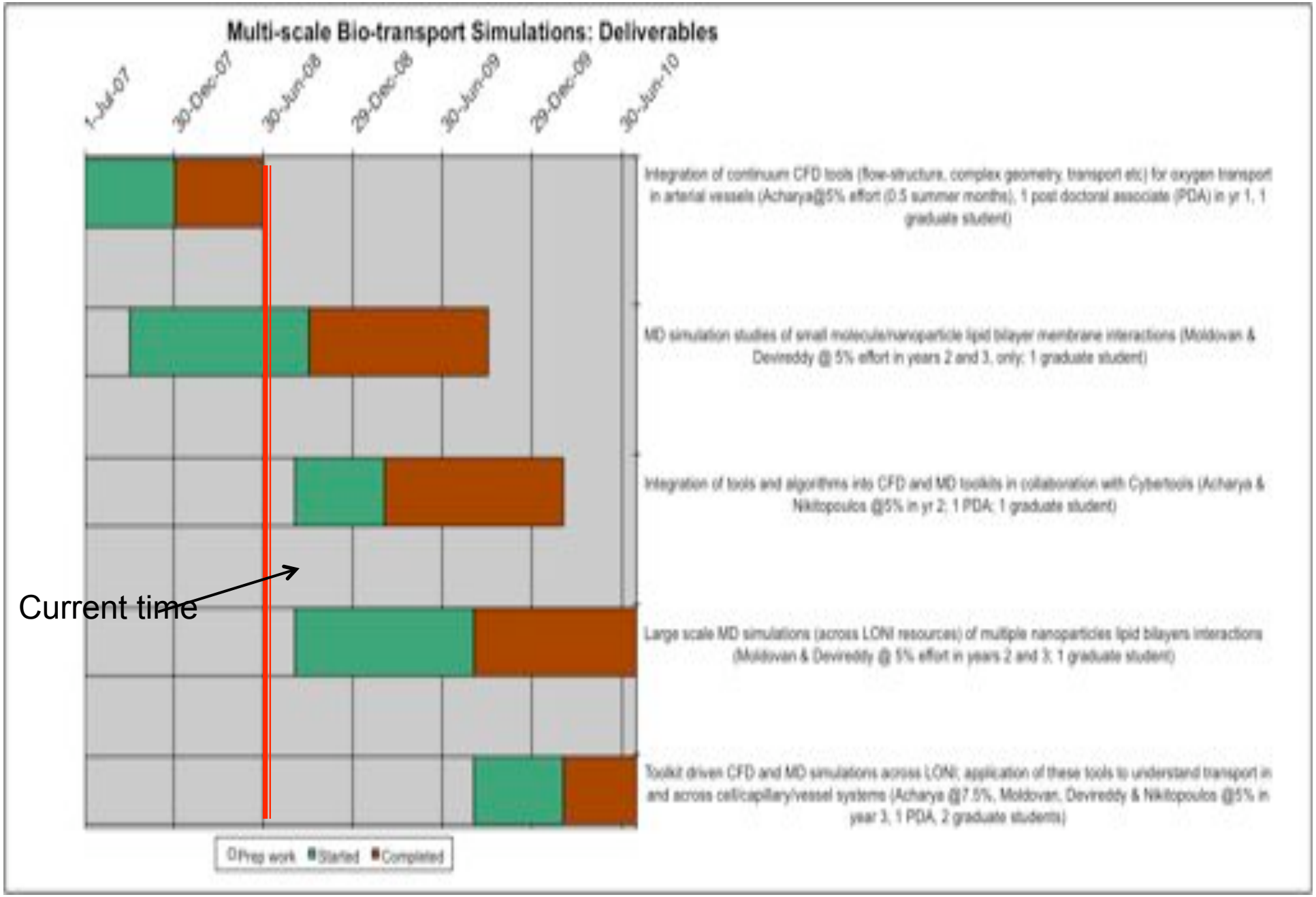
- **No penetration of water molecules**
- **Data analyzed for mass density profiles, radial distribution functions, tail order parameters, and water orientation profile**

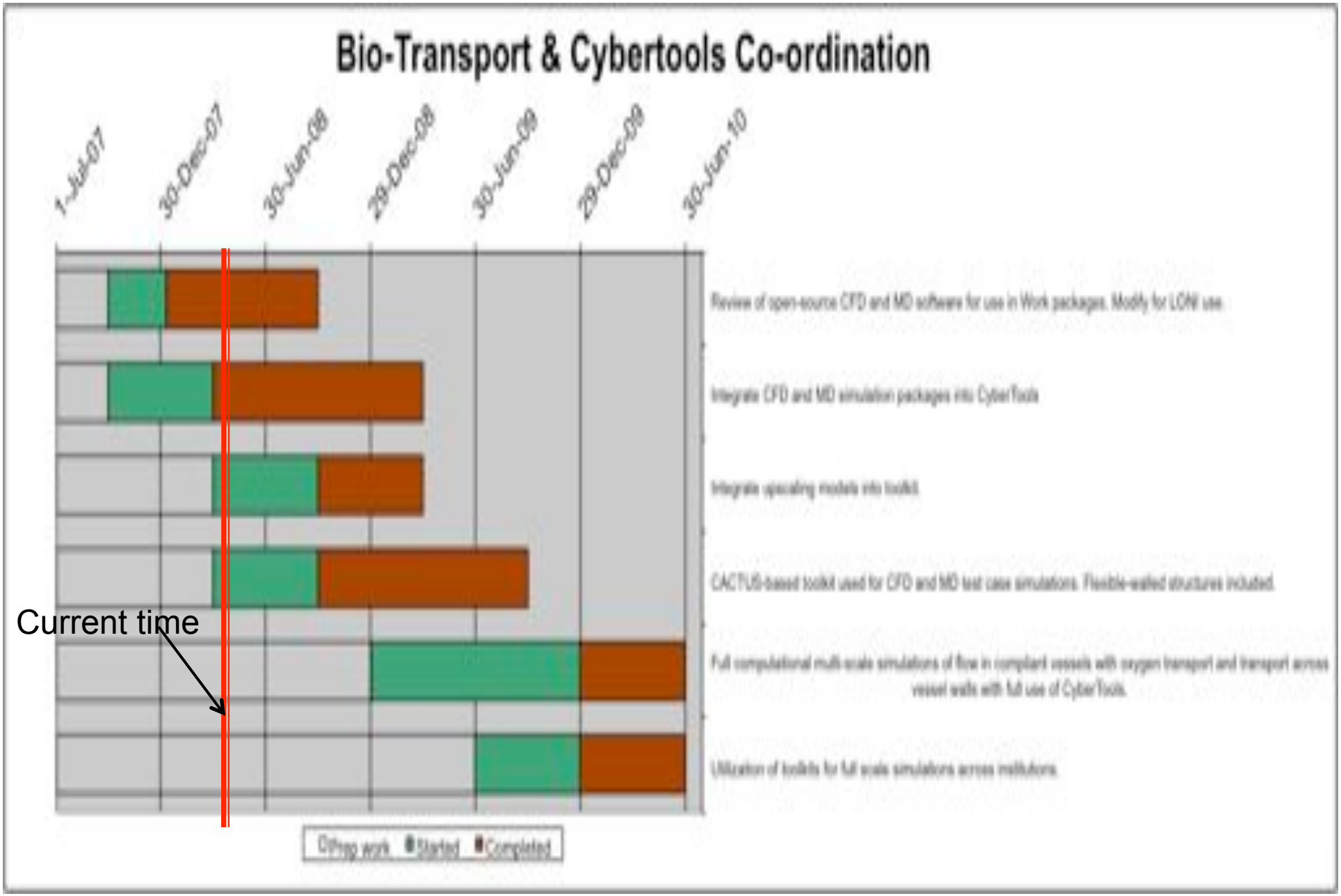
Mass density profiles of : DMPC, DMSO, and water



10ns profiles: dotted line ,50ns profiles: solid line

- CFD
  - ✓Improvements to the IBM (pressure interpolation)
  - ✓Working on the MPM for greater robustness (implicit, parallel)
  - ✓Simulation of transport in flexible tubes
- MD
  - ✓ Simulation of small molecules across lipid bi-layers
- Collaboration with WP4
  - ✓ Regular meetings with the WP4 team
  - ✓Development of a simplified CFD code with data array structure consistent with Cactus for implementation as part of the CFD Toolkit
- CFD-MD Coupling
  - ✓Discussion on coupling strategy and approaches


Multi-scale Bio-transport Simulations: Deliverables
1-Jul-07
30-Dec-07
30-Jun-08
29-Dec-08
30-Jun-09
29-Dec-09
30-Jun-10
Integration of continuum CFD tools (flow-structure, complex geometry, transport etc) for oxygen transport in arterial vessels (Acharya@5% effort (0.5 summer months), 1 post doctoral associate (PDA) in yr 1, 1 graduate student)
MD simulation studies of small molecule/nanoparticle lipid bilayer membrane interactions (Moldovan & Devireddy @ 5% effort in years 2 and 3, only; 1 graduate student)
Integration of tools and algorithms into CFD and MD toolkits in collaboration with Cybertools (Acharya & Nikitopoulos @5% in yr 2; 1 PDA; 1 graduate student)
Current time
Large scale MD simulations (across LONI resources) of multiple nanoparticles lipid bilayers interactions (Moldovan & Devireddy @ 5% effort in years 2 and 3; 1 graduate student)
Toolkit driven CFD and MD simulations across LONI; application of these tools to understand transport in and across cell/capillary/vessel systems (Acharya @7.5%, Moldovan, Devireddy & Nikitopoulos @5% in year 3, 1 PDA, 2 graduate students)
Prep work
Started
Completed


Bio-Transport & Cybertools Co-ordination
1-Jul-07
30-Dec-07
30-Jun-08
29-Dec-08
30-Jun-09
29-Dec-09
30-Jun-10
Review of open-source CFD and MD software for use in Work packages. Modify for LDNI use
Integrate CFD and MD simulation packages into CyberTools
Integrate upscaling models into toolkit
CACTUS-based toolkit used for CFD and MD test case simulations. Flexible-walled structures included
Full computational multi-scale simulations of flow in compliant vessels with oxygen transport and transport across vessel walls with full use of CyberTools
Utilization of toolkits for full scale simulations across institutions
Current time
Other work
Started
Completed

- Development of improved CFD methodologies for biological systems (complex geometries, moving boundaries, multi-scale phenomena)

- Utilization of CFD and MD methodologies for improved understanding of transport processes in biological systems

- Supporting the development of Toolkit infrastructure for open source, scalable code for community usage

- CFD-MD integration for resolving/integrating atomistic effects

- Future interactions will also include the visualization groups and the portals group

Cybertools WP4
Application Toolkits

S. Jha, J. Kim, E. Schnetter, M. Tyagi

CCENTER FOR COMPUTATION & TECHNOLOGY

WP4: The Mission

- Capture and analyze the application characteristics and requirements of the science drivers
- Facilitate the use of computational infrastructrure, including but not limited to LONI, for advancing science
  - Short-term (6-12 months): help deploy applications and the design of tools to facilitate utilisation of infrastructure
  - Longer-term (1-3 years): design of application managers and toolkits – that abstract the common requirements and usage modes of applications
- Work not only with Science Drivers to provide direct support, but also interface with other Cybertool WPs

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

WP4: Personnel

- Science Drivers:
  - Sumanta Acharya, Prasad Kalghatgi
  - Don Gaver, Jerina Pillert, Kate Hamlington, Dave Halperin
  - Steve Soper, Dimitris Nikitopoulos, Eamonn Walker
  - Tom Bishop
- HPC/LONI/CyD:
  - Honggao Liu (LONI), Dan Katz and Joohyun Kim (CyD)
  - Hartmut Kaiser, Sanjay Kodiyalam
- WP4 funded personnel:
  - Joao Abecasis (GA)
  - *Nayong Kim* (USC) and *Jeff Ko* (KISTI, Korea)

CCENTER FOR COMPUTATION & TECHNOLOGY

WP4-SD Interaction

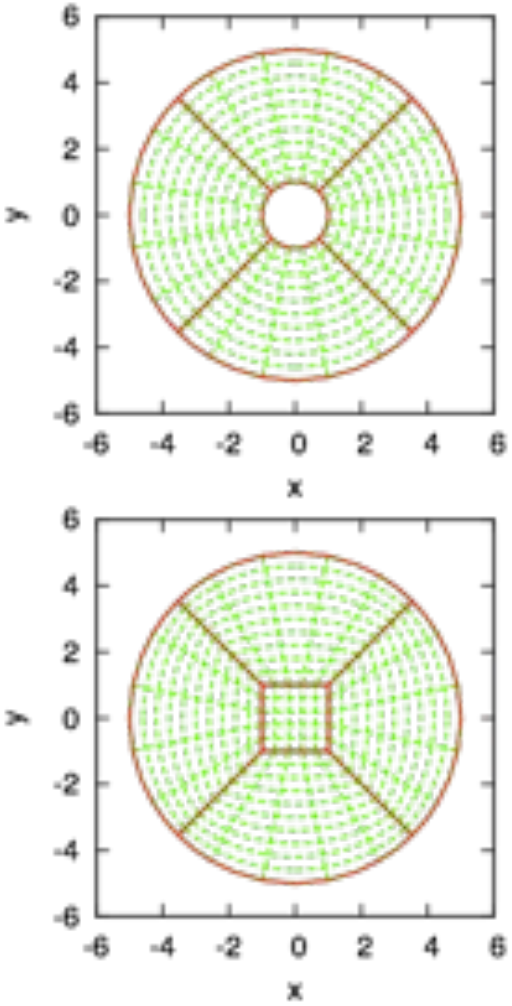
- Analyse the requirements SDs, into existing (fast track) or need-to-be-developed (deep track) capabilities
  - Regular bi-weekly meetings
- SD1 (BioTransport):
  - Multi-block support for implicit solvers [Prasad]
  - Immersed boundary support for moving geometry
- SD2 (Fluid Structure Interaction):
  - OpenMP version for BEM code [Jerina, Kate]
- SD3 (BioSensor):
  - Fast Track: vorticity formulation + driven cavity
  - Deep Track: coupling CFD + MD appropriate interface
- Infrastructure development for all SDs (with WP1,2)
  - Initial sketch of general purpose application manager

	Biotransport	Fluid-Structure Interaction	BioSensor	Capabilites that Exist
Numerical Schemes				
BE Method		Y		
Finite Difference	Y		Y	Y
Finite Volume	Y			Y
Numerical Solvers				
Lapack		Y		SCALAPACK
Hyper	Y			UNIGRID
PetSc				UNIGRID
MultiGrid			Y	MUDPACK
Explicit				
Domain Representation				
Uniform Grid		Y	Y	Y
Single Block			Y	Y
Multiblock	Y			Y
AMR	Y			Y
Unstructured (Meshless)			Y	

Computational Infrastructure	Biotransport	Fluid-Structure Interaction	BioSensor	Work Package
Parallelization Scheme				
OpenMP		Y	Y	WP4
MPI	Y		Y	WP4
Cactus Features				
Checkpointing	Y	Y	Y	WP4
Error Handling	Y	Y	Y	WP4
Visualization (post-processing)	Y	Y	Y	WP3
Visualization (Steering)			Y	WP3
Distributed Data Mgmt, Handling and Archiving	Y	Y	Y	WP1
Efficient I/O			Y	WP1, WP4
Distributed Job Launch/Mgmt	Y		Y	WP1, WP2, WP4

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

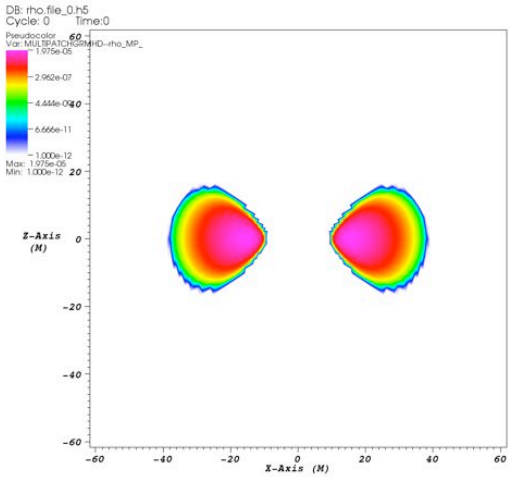
Multi-Patch Systems in the Cactus Framework



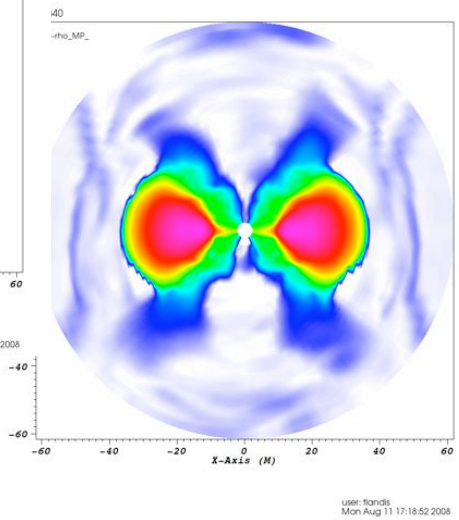
- Spherical (smooth) outer and/or inner boundaries
- No coordinate singularities (*z* axis, origin)
- Adapted to interesting features (neutron stars, boundaries, objects and their trajectories)
- Can choose angular and radial resolution independently

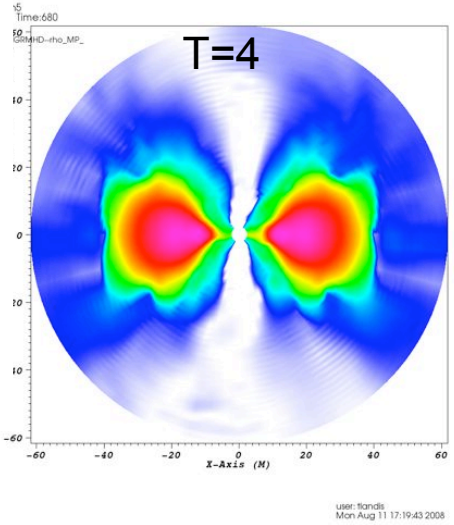
CENTER FOR COMPUTATION & TECHNOLOGY

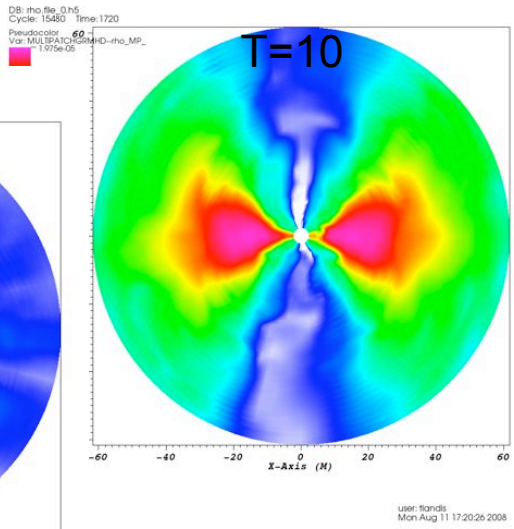
Astrophysical Application: Magnetised Torus



Schwarzschild (black hole) background







T=0

T=2

T=4

T=10

Colour map:

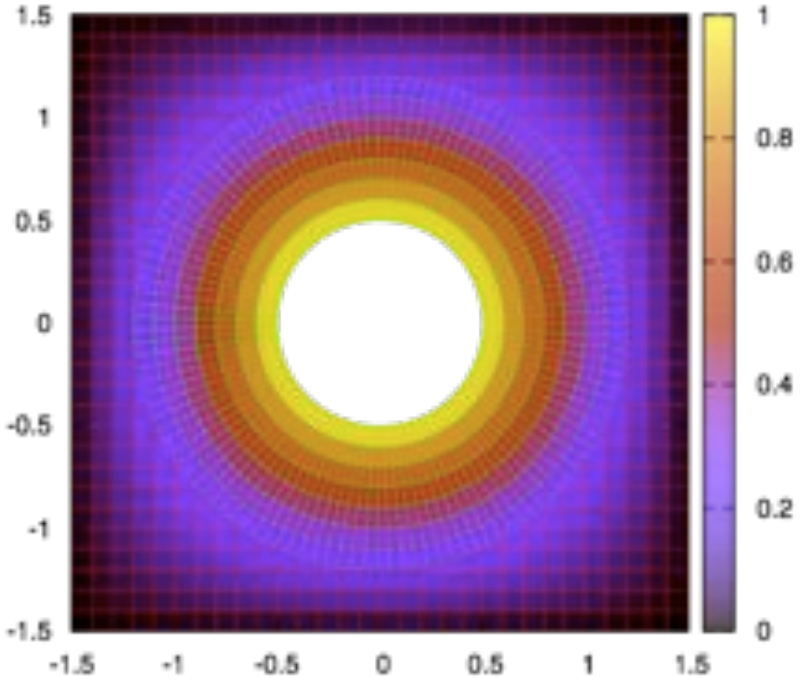
fluid density

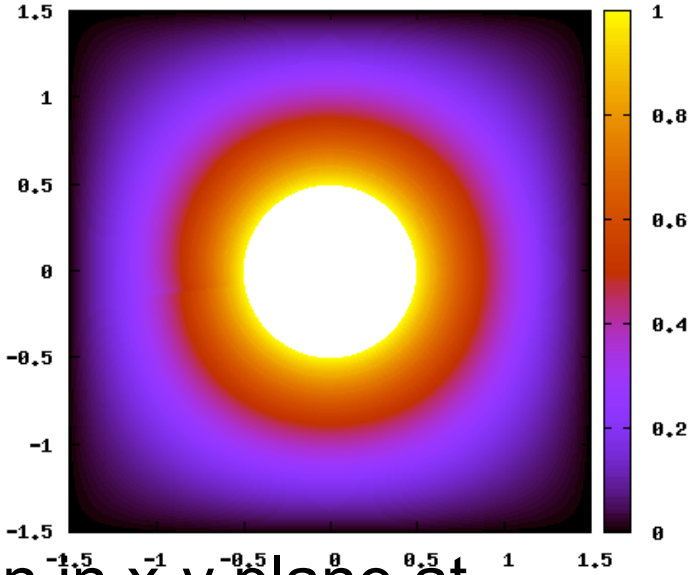
Initially weak poloidal magnetic field loops grow and make torus unstable

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

Test Problem: Diffusion Equation

Domain:
Cube minus Cylinder





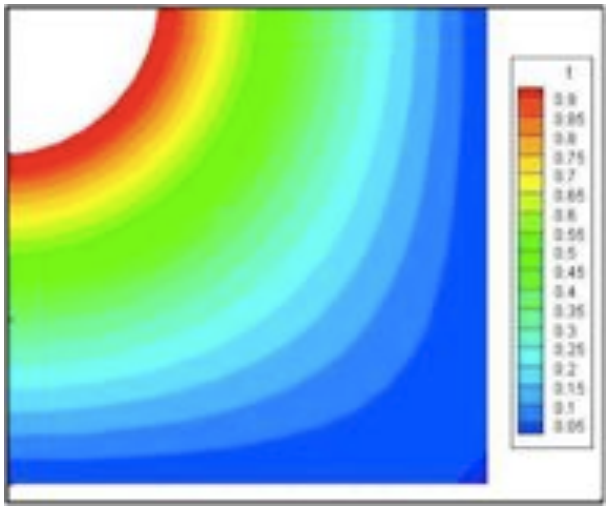
Solution in x-y plane at T=0.5

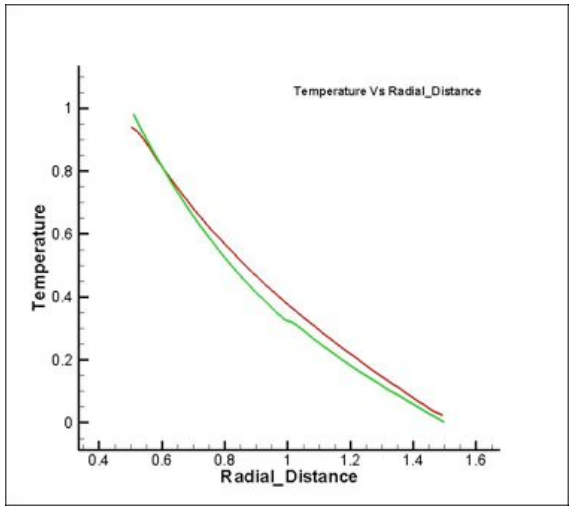
Grid Structure in x-y plane

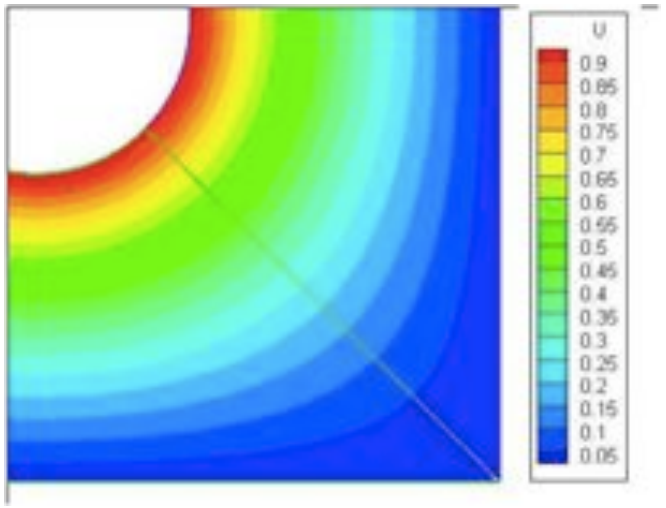
See P. Kalghatgi's poster for a comparison to S. Acharya's Science Driver

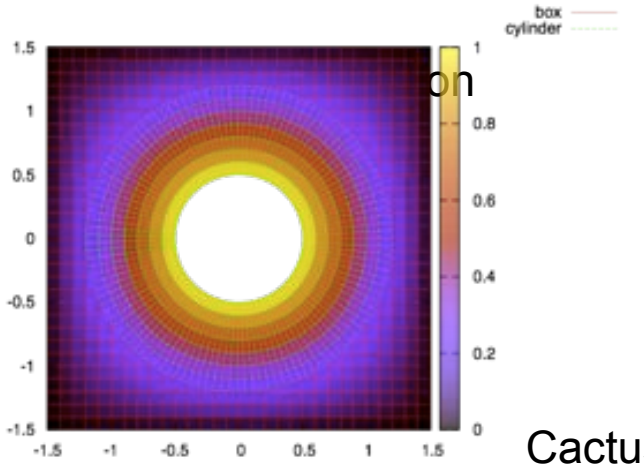
LSU
CENTER FOR COMPUTATION & TECHNOLOGY

Test Problem: Diffusion Equation









on

Comparison

CFD Module simulation

Courtesy P. Kalghatgi (see poster)

Domain (3D):
Cube minus Cylinder,
x-y central plane shown

Cactus multi-patch

grid structure

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

Cactus: Overview

- Cactus (www.cactuscode.org) is a software framework for collaborative development, primarily developed at LSU
- Very successful in astrophysics (used by >200 publications, >30 student theses)
- Provides computational infrastructure and supports application toolkits (e.g. CCTK, Einstein Toolkit)

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

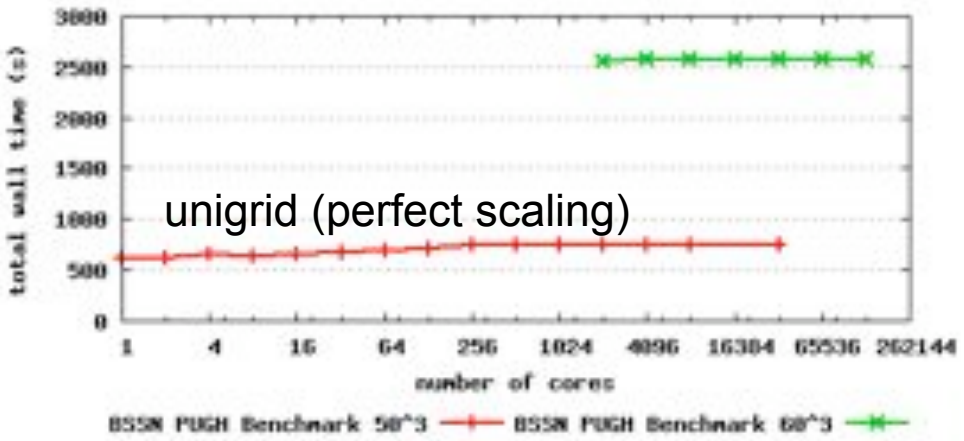
Cactus: Separation of Concerns

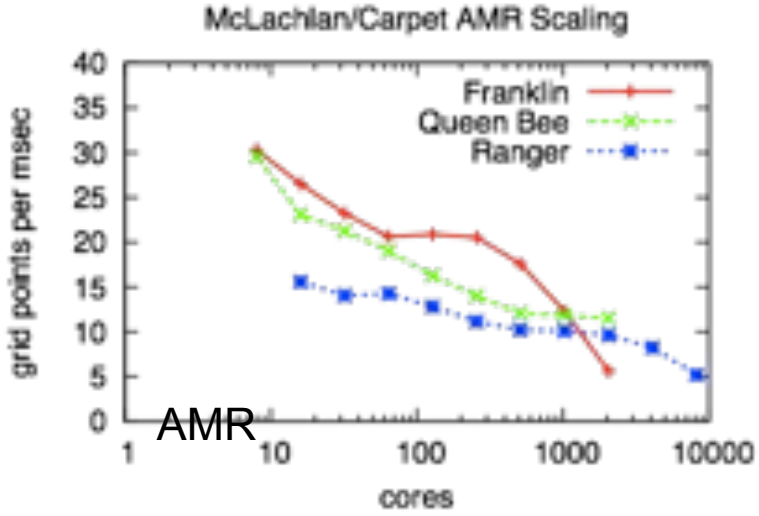
- Physics: equations, stability, modelling
- Discretisation: differencing, numerical analysis, conservation, constraints
- Domain: mesh, parallelisation, load balancing, cache efficiency
- Computer science: module interfaces, scheduling, efficient I/O, visualisation

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

Cactus: Parallelisation

- Supports both OpenMP (simpler) and MPI (more efficient) parallelisation strategies
- Provides Adaptive Mesh Refinement (AMR) and multi-patch domains with Carpet driver (www.carpetcode.org)
- Can e.g. perform automatic loop optimisations (cache blocking) at run time





CENTER FOR COMPUTATION & TECHNOLOGY

Cactus: Job Handling

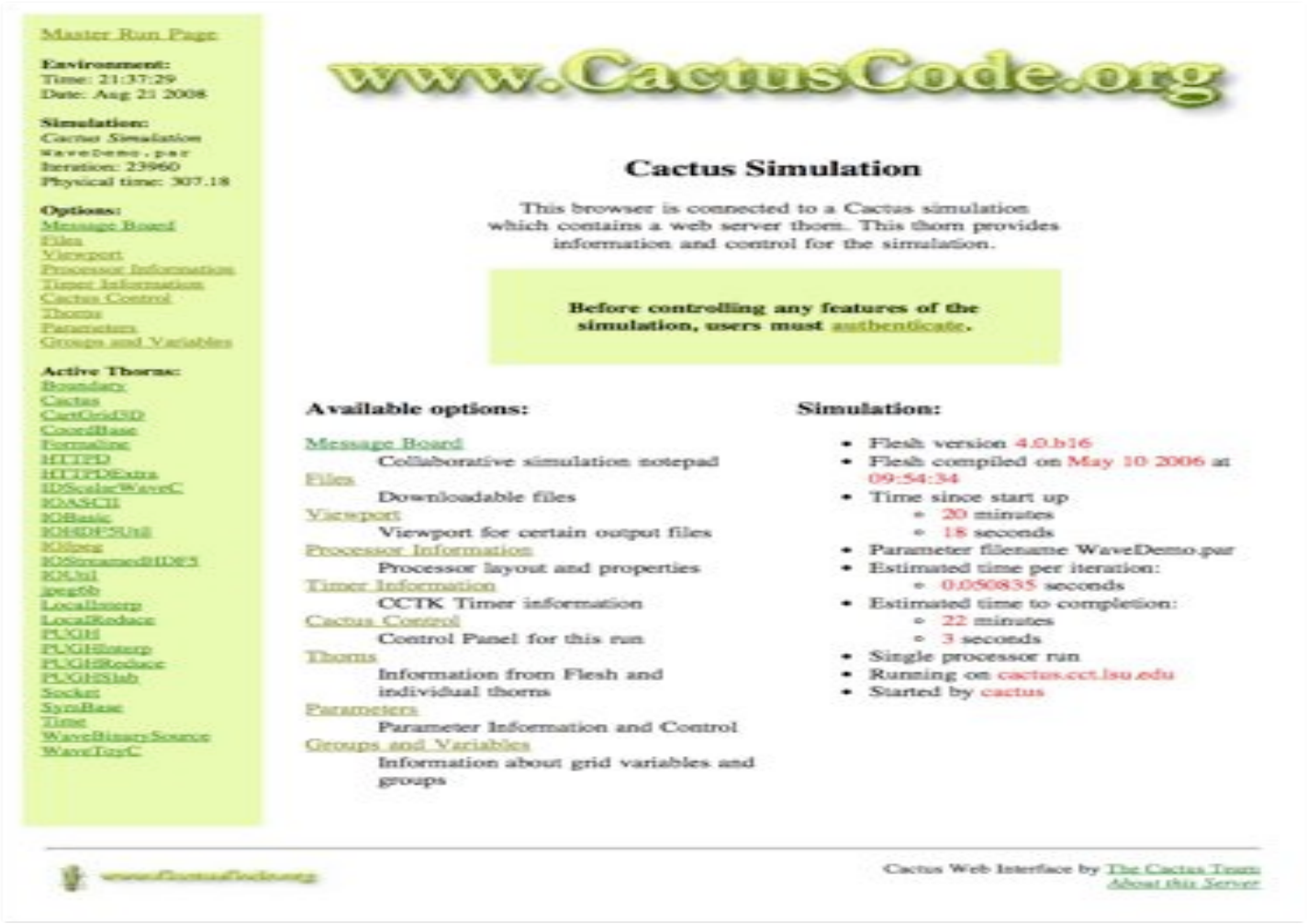
- [Show portal listing Cactus jobs] (http://devportal.cct.lsu.edu:8081/gridsphere/gridsphere)
- [Interact with perpetual Cactus simulation] (http://cactus.cct.lsu.edu:5555/)

CENTER FOR COMPUTATION
& TECHNOLOGY

Cactus Simulation Portal

	simulation	by user	parameter file path	running on URL	started TZ America/Chicago	last updated TZ America/Chicago
	QC-0	eschnett	qc0-reference.par /work/eschnett/simulations/qc0-0001/output-0000-active	qb007	0:00 hours ago Aug 21, 2008 8:50:53 PM	0:00 hours ago Aug 21, 2008 8:50:53 PM
	QC-0	eschnett	qc0-reference.par /work/eschnett/simulations/qc0-0000/output-0000-active	qb003	0:00 hours ago Aug 21, 2008 8:50:49 PM	0:00 hours ago Aug 21, 2008 8:50:49 PM
	Announce to an RDF metadata server	eschnett	sendrdf.par /work/eschnett/simulations/sendrdf-0001/output-0000-active	qb047	0:22 hours ago Aug 21, 2008 8:29:04 PM	0:22 hours ago Aug 21, 2008 8:29:04 PM
	Cactus Simulation	defseg	test-formaline-10000.par /home/defseg/timings/one-host	louie126	1001:24 hours ago Jul 11, 2008 3:26:42 AM	1001:24 hours ago Jul 11, 2008 3:26:42 AM
	Cactus Simulation	defseg	test-formaline-1000.par /home/defseg/timings/one-host	louie126	1001:46 hours ago Jul 11, 2008 3:04:24 AM	1001:46 hours ago Jul 11, 2008 3:04:24 AM
	Cactus Simulation	defseg	test-formaline-1000.par /home/defseg/timings/one-host	louie126	1001:54 hours ago Jul 11, 2008 2:56:26 AM	1001:54 hours ago Jul 11, 2008 2:56:26 AM
	Cactus Simulation	defseg	test-formaline-100.par /home/defseg/timings/one-host	louie126	1001:56 hours ago Jul 11, 2008 2:54:49 AM	1001:56 hours ago Jul 11, 2008 2:54:49 AM
	Cactus Simulation	defseg	test-formaline-10000.par /home/defseg/timings/one-host	louie126	1002:08 hours ago Jul 11, 2008 2:42:33 AM	1002:08 hours ago Jul 11, 2008 2:42:33 AM
	Cactus Simulation	defseg	test-formaline-1000.par /home/defseg/timings/one-host	louie127	1002:34 hours ago Jul 11, 2008 2:16:38 AM	1002:34 hours ago Jul 11, 2008 2:16:38 AM
	Cactus Simulation	defseg	test-formaline-100.par /home/defseg/timings/one-host	louie127	1002:35 hours ago Jul 11, 2008 2:16:00 AM	1002:35 hours ago Jul 11, 2008 2:16:00 AM
	Cactus Simulation	defseg	test-formaline-10000.par /home/defseg/timings/one-host	louie127	1002:48 hours ago Jul 11, 2008 2:02:48 AM	1002:48 hours ago Jul 11, 2008 2:02:48 AM
	Cactus Simulation	defseg	test-formaline-1000.par /home/defseg/timings/one-host	louie126	1003:19 hours ago Jul 11, 2008 1:31:43 AM	1003:19 hours ago Jul 11, 2008 1:31:43 AM

CENTER FOR COMPUTATION
& TECHNOLOGY



Cactus: On-Line Visualisation

Master Run Page

Environment:
Time: 21:38:40
Date: Aug 21 2008

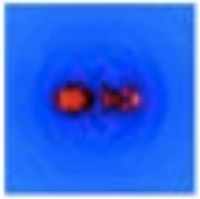
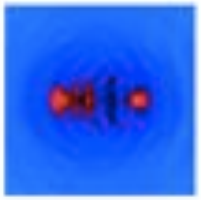
Simulation:
Cactus Simulation
waveDemo.par
Iteration: 25360
Physical time: 325.13

Options:
Message Board
Files
Viewport
Processor Information
Timer Information
Cactus Control
Thorns
Parameters
Groups and Variables

Viewport

This page displays certain types of the output files from the download page as images (currently only JPEGs [mime type image/jpeg]).

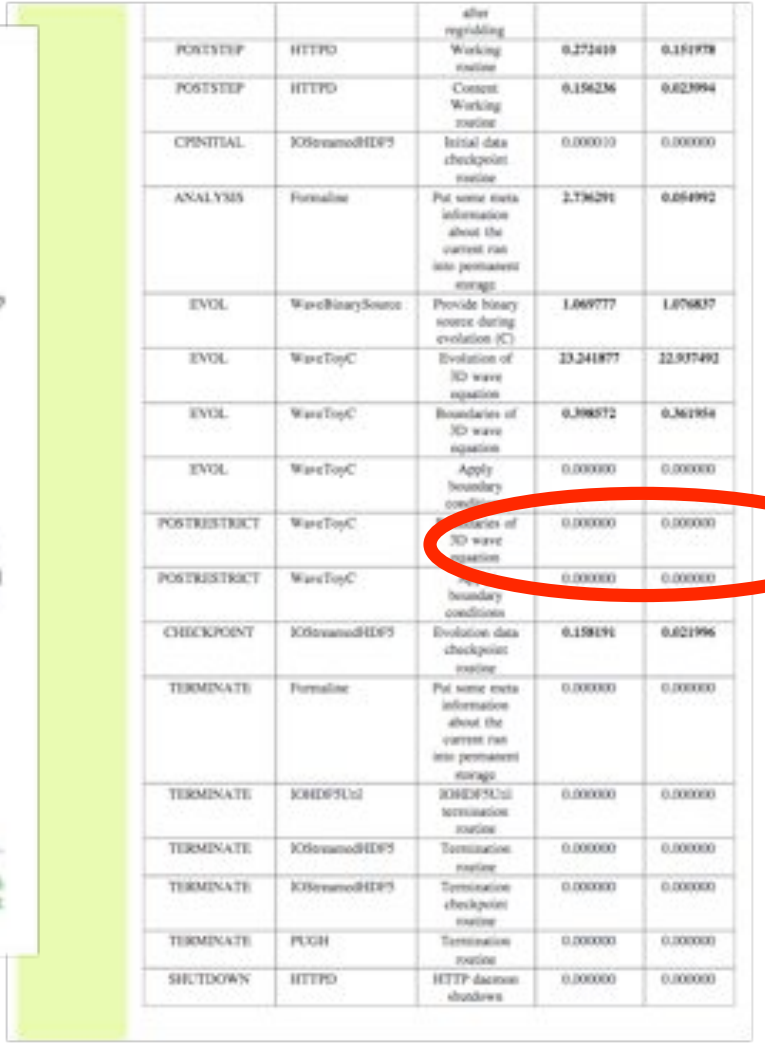
Many IO methods have *steerable* parameters which allow you to e.g. add fields and customise behaviour. Depending on your authorisation, you can access the parameter steering page

Variable Slice	Description	Image
WAVETOY::phi xz_[20]	*Jpegs of slices*	
WAVETOY::phi yz_[20]	*Jpegs of slices*	

Cactus Web Interface by The Cactus Team
About this Server

LSU
CENTER FOR COMPUTATION
& TECHNOLOGY

Cactus: Steering, Profiling



		after regridding		
POSTSTEP	HTTPD	Working routine	0.272410	0.151978
POSTSTEP	HTTPD	Control Working routine	0.156236	0.023994
CPINITIAL	IOStreamedHDF5	Initial data checkpoint routine	0.000010	0.000000
ANALYSIS	Formaline	Put some extra information about the current run into permanent storage	2.736291	0.004992
EVOL	WaveBinarySource	Provide binary source during evolution (C)	1.069777	1.076837
EVOL	WaveToyC	Evolution of 3D wave equation	23.241877	22.937492
EVOL	WaveToyC	Boundaries of 3D wave equation	0.398572	0.361954
EVOL	WaveToyC	Apply Boundary condition	0.000000	0.000000
POSTRESTRICT	WaveToyC	Boundaries of 3D wave equation	0.000000	0.000000
POSTRESTRICT	WaveToyC	Apply boundary conditions	0.000000	0.000000
CHECKPOINT	IOStreamedHDF5	Evolution data checkpoint routine	0.158191	0.021996
TERMINATE	Formaline	Put some extra information about the current run into permanent storage	0.000000	0.000000
TERMINATE	IOHDF5Util	IOHDF5Util termination routine	0.000000	0.000000
TERMINATE	IOStreamedHDF5	Termination routine	0.000000	0.000000
TERMINATE	IOStreamedHDF5	Termination checkpoint routine	0.000000	0.000000
TERMINATE	PUGH	Termination routine	0.000000	0.000000
SHUTDOWN	HTTPD	HTTP daemon shutdown	0.000000	0.000000

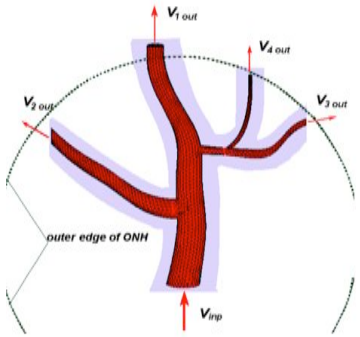
LSU
CENTER FOR COMPUTATION
& TECHNOLOGY

Next Steps

- Benefit from other ongoing Cactus projects:
  – XiRel (improve AMR; data handling)
  – Alpaca (performance/correctness tools)
  – ParCa (connect to PARAMESH solver)
- Generalise elliptic solver interfaces for AMR /multi-patch

CENTER FOR COMPUTATION & TECHNOLOGY

WP4: Connection to SD1 (Biotransport)



Main driver for “multiblock finite volume method”

Continuum flow and FSI calculations

- Multiblock structured grid (Biosensors need this capability)
- Flow-Structure interaction (Science need for BEM also)
- Particle-based meshless calculations for structural deformations (Material point method, MPM)
- Immersed Boundary Methodology (IBM) for resolving boundary conditions along moving interfacial surfaces

Non-continuum Effects

- Atomistic (Molecular-Dynamics) simulations of particle/molecule transport across cellular interfaces
- Upscaling or coarse-graining calculations for averaged property information needed for continuum calculations

CENTER FOR COMPUTATION & TECHNOLOGY

SD2: Flow around Ω-obstructions

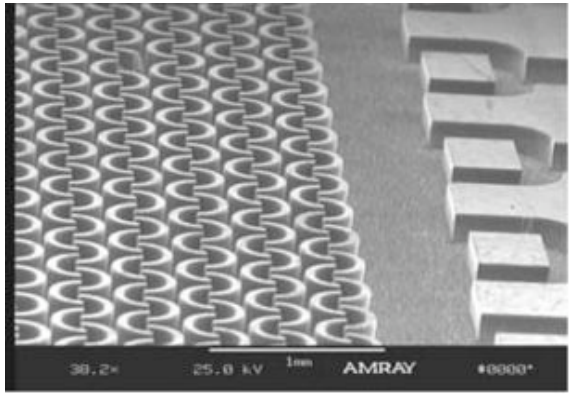
(slide courtesy: Gaver Group)

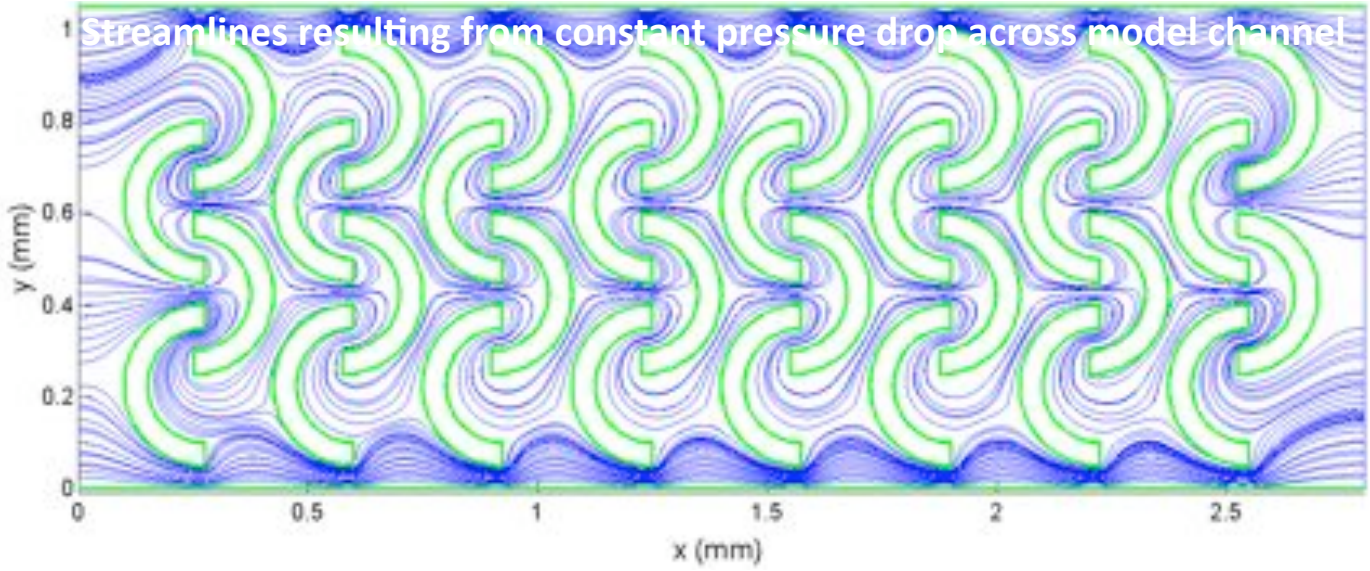
- **GOAL** ➔ *Computationally determine the optimal geometric configuration of the omega channel network to enhance mixing of two species.*
- Laminar flow field governed by continuity & Stokes equations:

$$\nabla \cdot \mathbf{u} = 0$$

$$\nabla P = \mu \nabla^2 \mathbf{u}$$

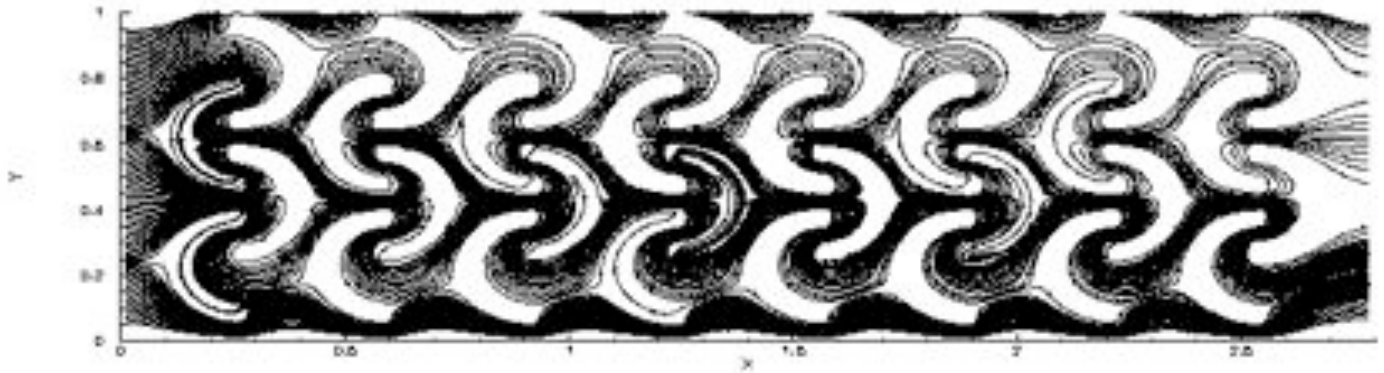
- Boundary Element Method determines velocities and surface stresses



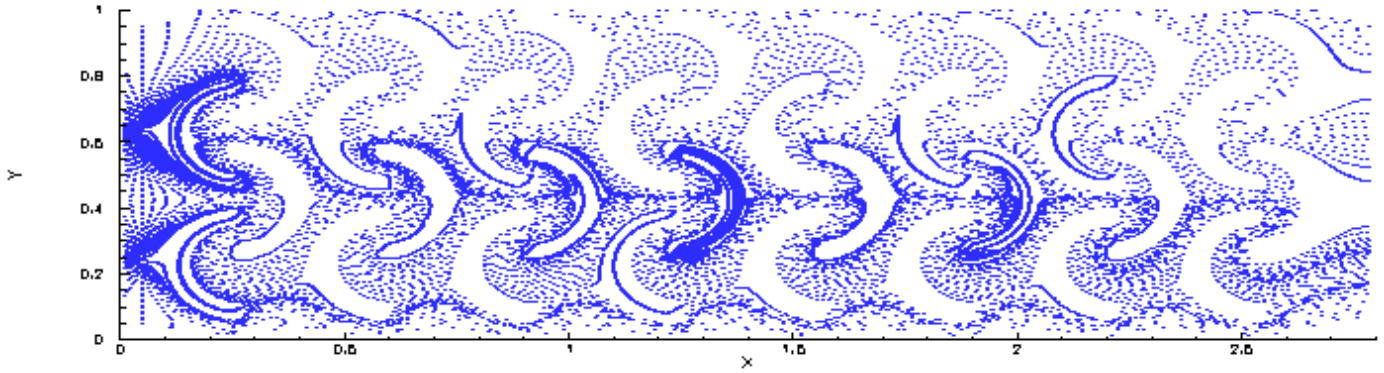


CENTER FOR COMPUTATION
& TECHNOLOGY

WP4: Link to SD2
Model Microchannel Problem



Streamlines



Particle Traces

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

WP4: Link to SD2

(slide courtesy: Gaver Group)

- Current Work:
  - Parallelization of Stokes flow problem for use in the HPC environment (WP4: Mayank Tyagi, Shantenu Jha, Sanjay Kodiyalam, Yaakoub El-Khamra)
    - *OpenMP*
  - Visualization of model problem using TecPlot
  - Generalization of code to develop a *CyberTool* package that solves Stokes flow equations

- Future Work:
  - Parallelization of source code including transport

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

WP4: Links to SD3

(Slide Courtesy: Dimitris)

- Multi-Phase flow Simulation Tool (WP4, WP3, WP1)
  - Parallelization
    - Implementation of parallelized Multi-Grid solver (WP4)
    - Distribution of different multi-processor simulations to groups of processors for efficient parametric studies (WP1)
  - Advanced interactive visualization tools (WP3)
  - Improvement of Accuracy/Performance
    - Implement Multi-Grid algorithm designed to handle elliptic equations with discontinuous coefficients (WP4)
    - "Poisson" solver for the pressure
  - Extend code capabilities to handle complex Cartesian geometries
    - Domain Decomposition (WP4)
    - Multi-blocking (WP4)
  - Computational Steering (WP1, WP3, WP4)

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

WP4: Link with SD3 (Coupling CFD-MD)

- Basic MD code
  - Developed
  - Parallelized in one dimension
  - Tested on simple 2D flows
    - Couette
    - Poiseulle
  - Modification of MD code to accommodate more diverse BC and parallelization in two dimensions (in progress)
  - **Documentation of the code for delivery to WP4 (in progress)**
- Continuum 3D N-S Parallel Code (Velocity/Vorticity Formulation)
  - Developed (international collaboration)
  - Tested on 3D driven cavity test problem - Re[0.1,5000] (in progress)
  - **Documentation of the code for delivery to WP4 (in progress)**
- Continuum-MD Coupling (In progress)
  - **Will work with WP4 to develop tools to**
    - Build a Modular Continuum-MD Parallel Simulation Environment under CACTUS

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

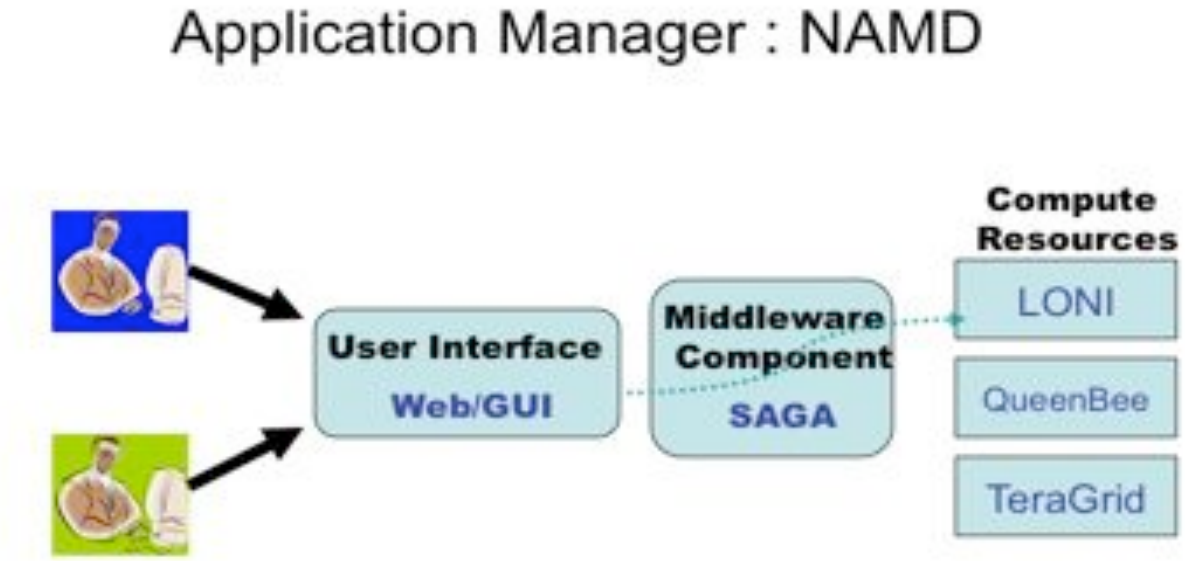
WP4: Connection to WP1-3

- WP1 (Scheduling and Data Services):
  - Work with WP1, CyD, LONI/HPC to define infrastructure and deployment requirements (eg PetaShare, SAGA)
  - Facilitating high-throughput MD and other simulations with data-intensive complex data-management needs
- WP2 (Info Services and Portals):
  - Applications Managers being developed using SAGA, which will integrate with portals and gateways
- WP3 (Visualization Sevices):
  - Exploring with applications use of Vish, VISIT
  - Common interface for accessing visualization (SAGA)

CENTER FOR COMPUTATION & TECHNOLOGY

Application Manager

Provides support for uniform usage patterns and interface to heterogeneous resources



LSU
CENTER FOR COMPUTATION & TECHNOLOGY

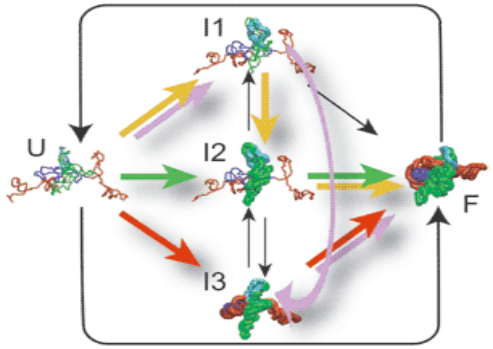
Application Manager: Sailent Points

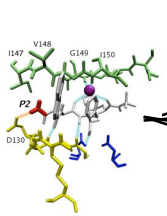
- Uniform: Provides single interface to heterogeneous and distributed resources
- Generic: Infrastructure can be embedded into either a portal or into a GUI
  - Also lightweight, flexible, modular
  - Easy to deploy
- Can support:
  - Other MD packages (e.g., LAMMPS)
  - Other Usage Modes (e.g., High-throughput) (WP1)
  - Complicated workflow driven computation (WP1)

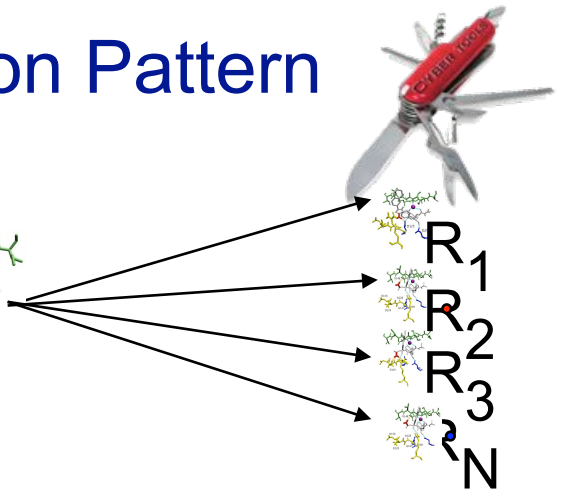
CENTER FOR COMPUTATION & TECHNOLOGY

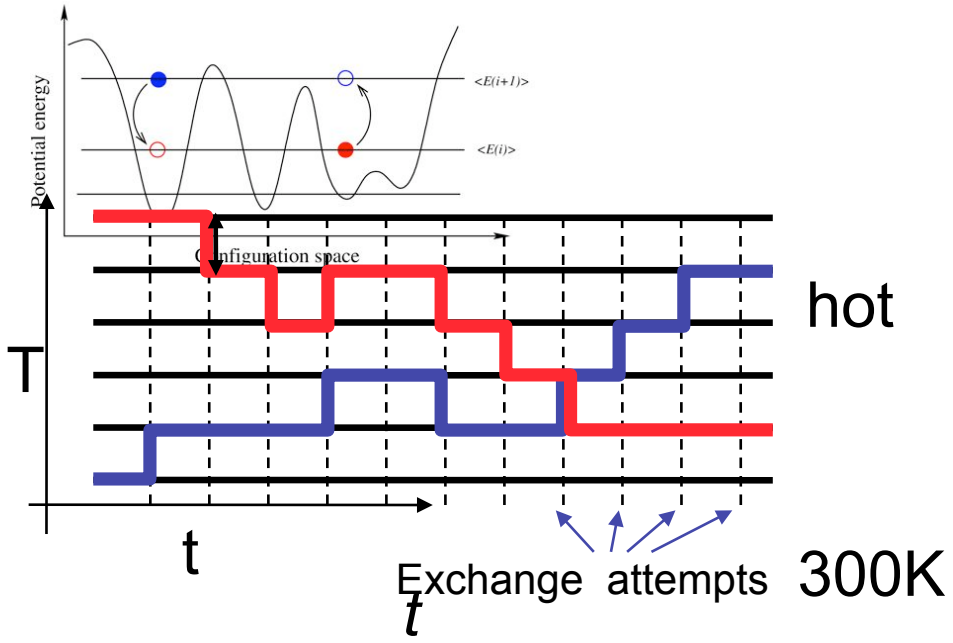
Replica-Exchange Application Pattern

- Task Level Parallelism
  - Embarrassingly distributable!
  - Loosely coupled
- Create replicas of initial configuration
- Spawn 'N' replicas over different machine
- Run for time *t ;* Attempt configuration swap
- Run for further time t; Repeat till finish



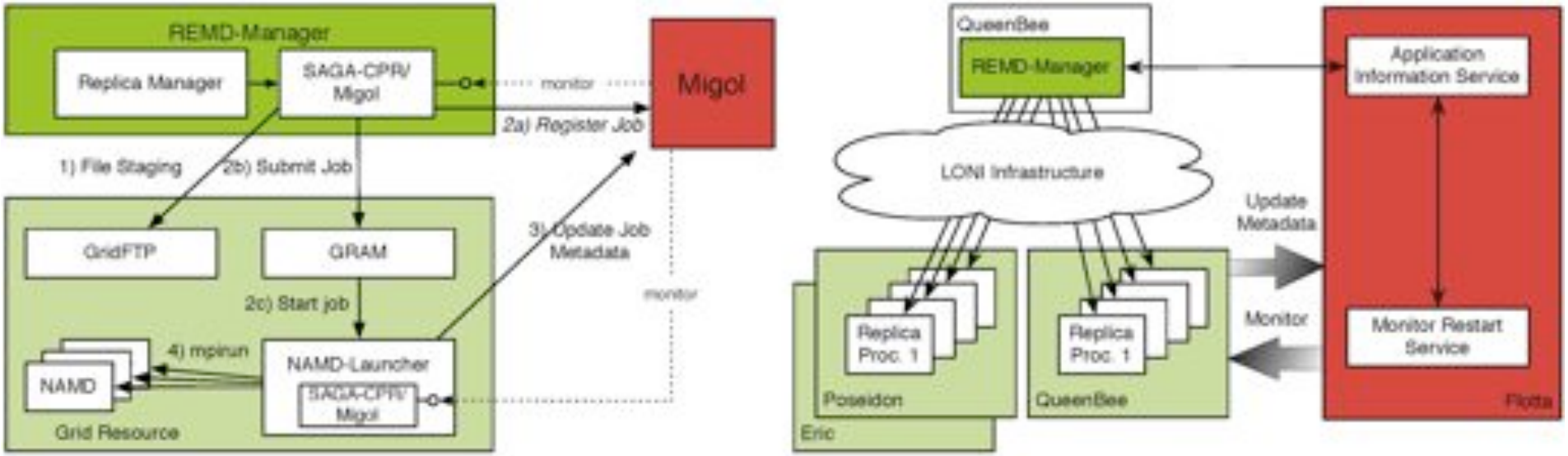


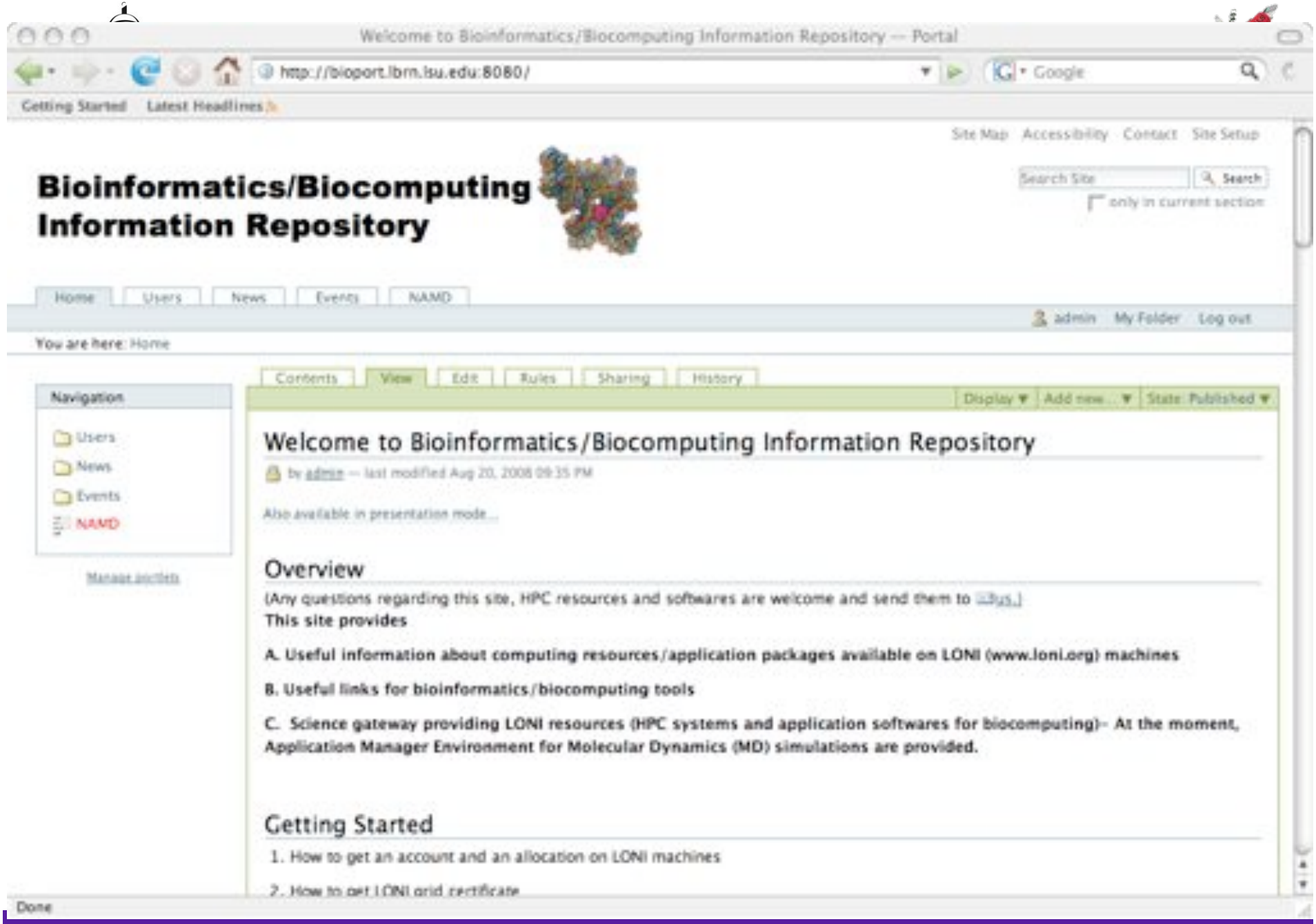


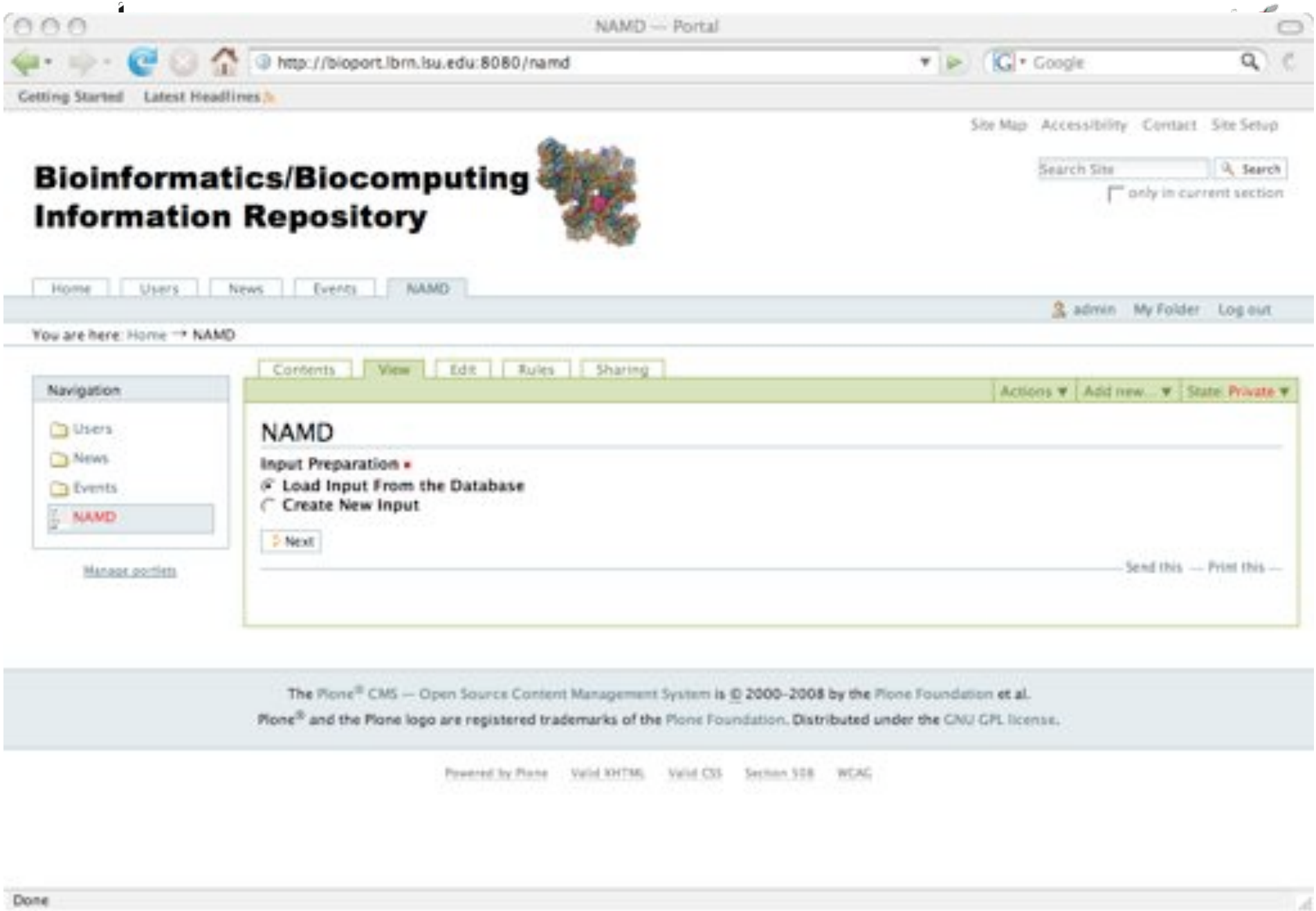


CENTER FOR COMPUTATION
& TECHNOLOGY

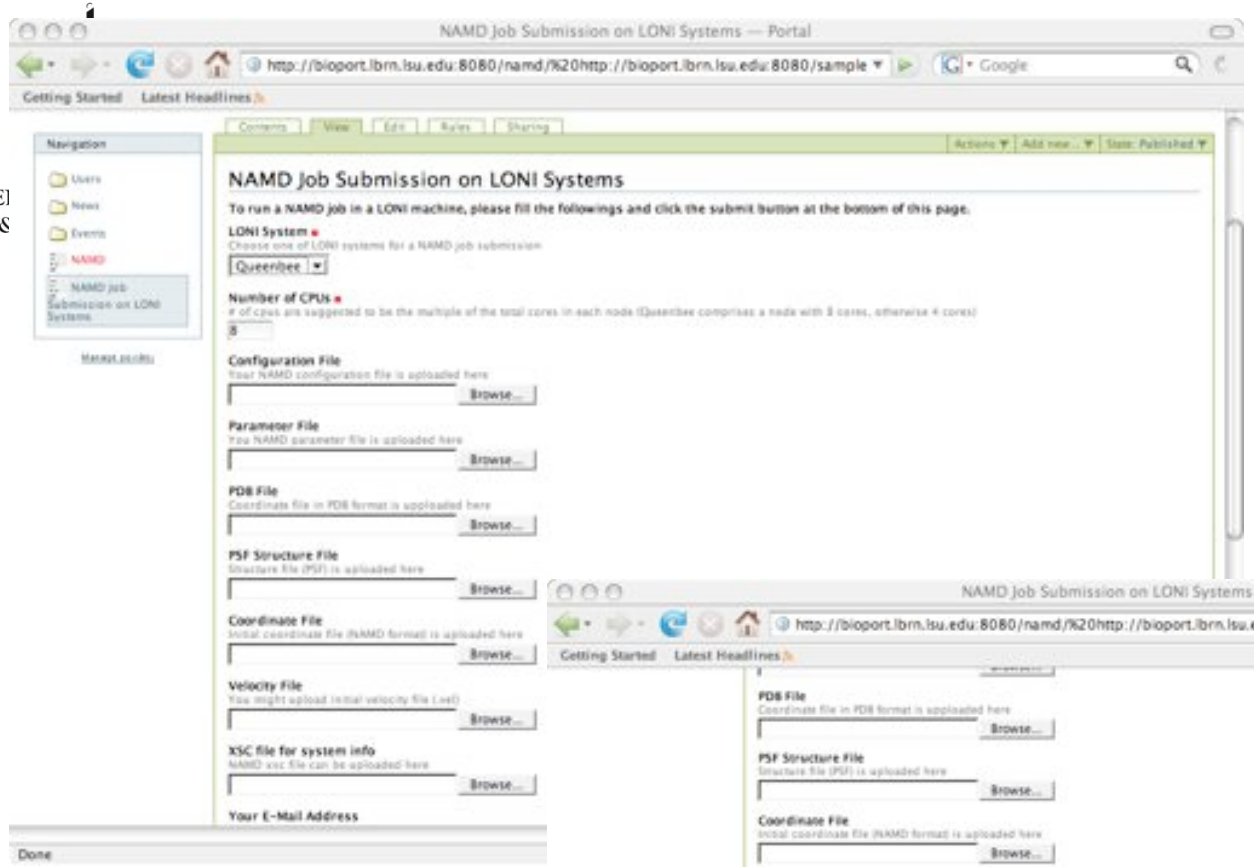
Replica-Exchange Manager

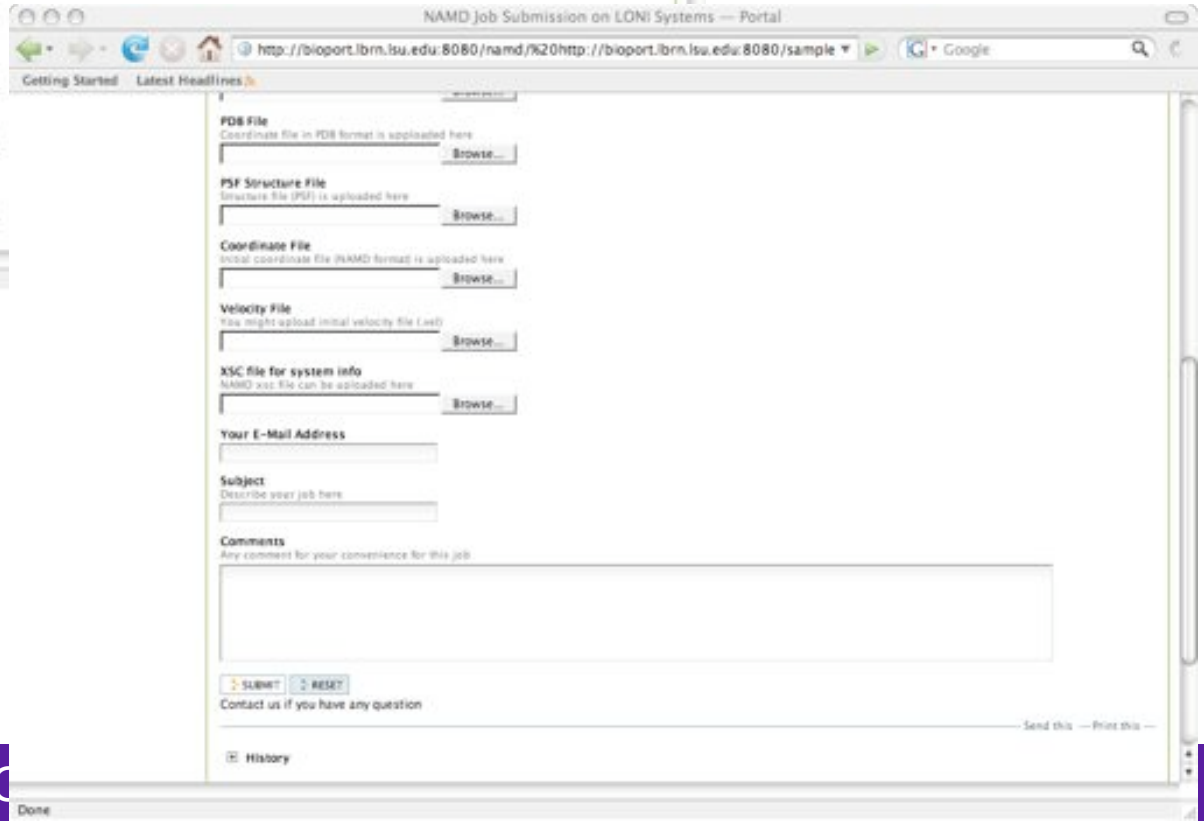



Welcome to Bioinformatics/Biocomputing Information Repository -- Portal
http://bioport.lbrn.lsu.edu:8080/
Google
Getting Started
Latest Headlines
Site Map
Accessibility
Contact
Site Setup
Bioinformatics/Biocomputing
Information Repository
Search Site
Search
only in current section
Home
Users
News
Events
NAMD
admin
My Folder
Log out
You are here: Home
Contents
View
Edit
Rules
Sharing
History
Navigation
Users
News
Events
NAMD
Display
Add new...
State: Published
Welcome to Bioinformatics/Biocomputing Information Repository
by admin -- last modified Aug 20, 2008 09:35 PM
Also available in presentation mode...
Overview
(Any questions regarding this site, HPC resources and softwares are welcome and send them to [bys.])
This site provides
A. Useful information about computing resources/application packages available on LONI (www.loni.org) machines
B. Useful links for bioinformatics/biocomputing tools
C. Science gateway providing LONI resources (HPC systems and application softwares for biocomputing)- At the moment, Application Manager Environment for Molecular Dynamics (MD) simulations are provided.
Getting Started
1. How to get an account and an allocation on LONI machines
2. How to get LONI grid certificate
Done

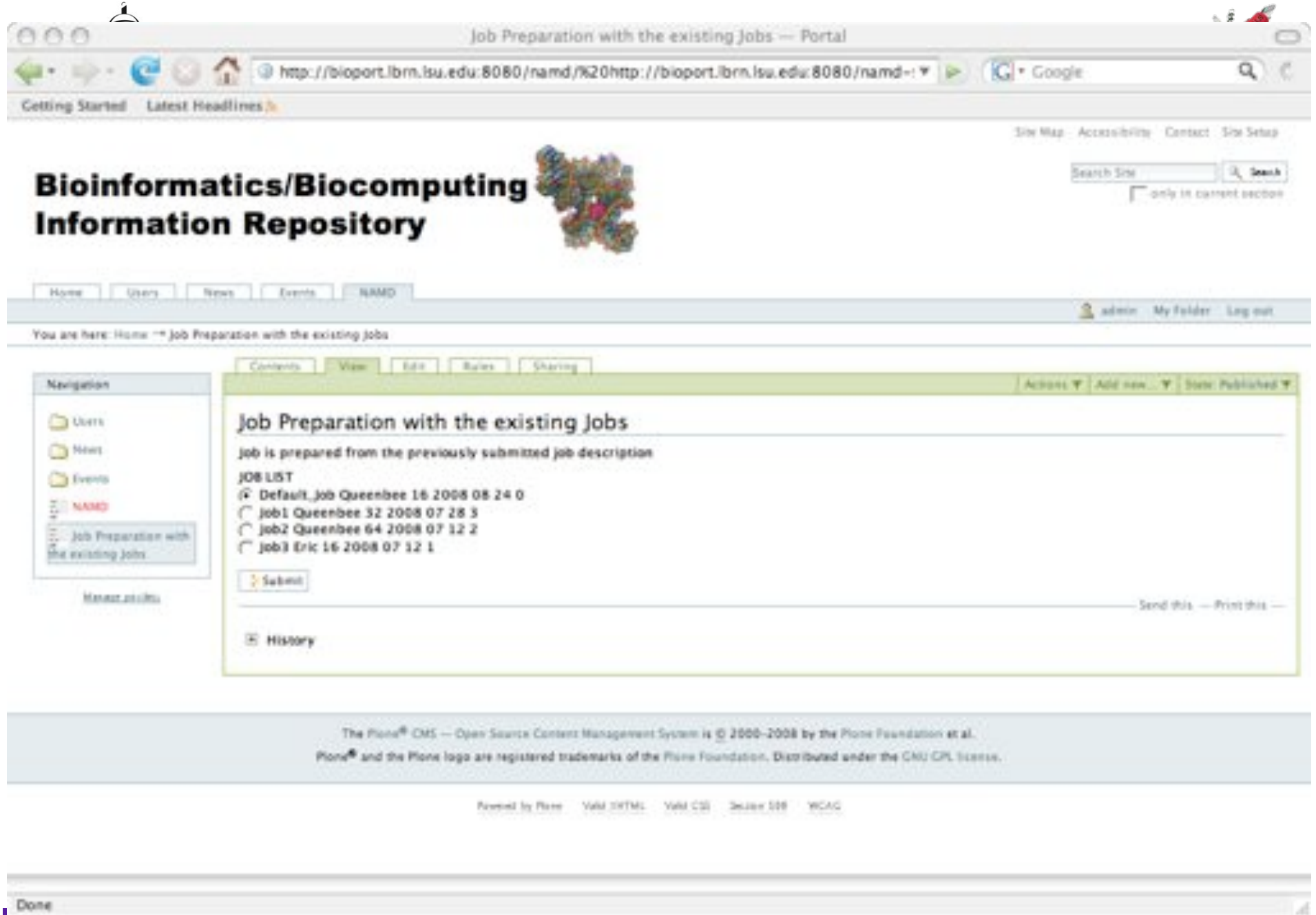

NAMD — Portal
http://bioport.lbm.lsu.edu:8080/namd
Google
Getting Started
Latest Headlines
Site Map
Accessibility
Contact
Site Setup
Bioinformatics/Biocomputing
Information Repository
Search Site
Search
only in current section
Home
Users
News
Events
NAMD
admin
My Folder
Log out
You are here: Home → NAMD
Contents
View
Edit
Rules
Sharing
Actions
Add new...
State: Private
Navigation
Users
News
Events
NAMD
Manage portlets
NAMD
Input Preparation *
Load Input From the Database
Create New Input
Next
Send this
Print this
The Plone® CMS — Open Source Content Management System is © 2000-2008 by the Plone Foundation et al.
Plone® and the Plone logo are registered trademarks of the Plone Foundation. Distributed under the GNU GPL license.
Powered by Plone
Valid XHTML
Valid CSS
Section 508
WCAG
Done

CENTE
&


NAMD Job Submission on LONI Systems — Portal
http://bioport.lbrn.lsu.edu:8080/namd/%20http://bioport.lbrn.lsu.edu:8080/sample
Getting Started
Latest Headlines
Navigation
NAMD Job Submission on LONI Systems
To run a NAMD job in a LONI machine, please fill the followings and click the submit button at the bottom of this page.
LONI System
Queenbee
Number of CPUs
Configuration File
Parameter File
PDB File
PSF Structure File
Coordinate File
Velocity File
XSC file for system info
Your E-Mail Address
Done


NAMD Job Submission on LONI Systems — Portal
http://bioport.lbrn.lsu.edu:8080/namd/%20http://bioport.lbrn.lsu.edu:8080/sample
Getting Started
Latest Headlines
PDB File
PSF Structure File
Coordinate File
Velocity File
XSC file for system info
Your E-Mail Address
Subject
Comments
SUBMIT
RESET
Contact us if you have any question
History
Done

CCT: C


Job Preparation with the existing Jobs — Portal
Bioinformatics/Biocomputing
Information Repository
Job Preparation with the existing Jobs
Job is prepared from the previously submitted job description
JOB LIST
Default_Job Queenbee 16 2008 08 24 0
Job1 Queenbee 32 2008 07 28 5
Job2 Queenbee 64 2008 07 12 2
Job3 Eric 16 2008 07 12 1
Submit
History
Done

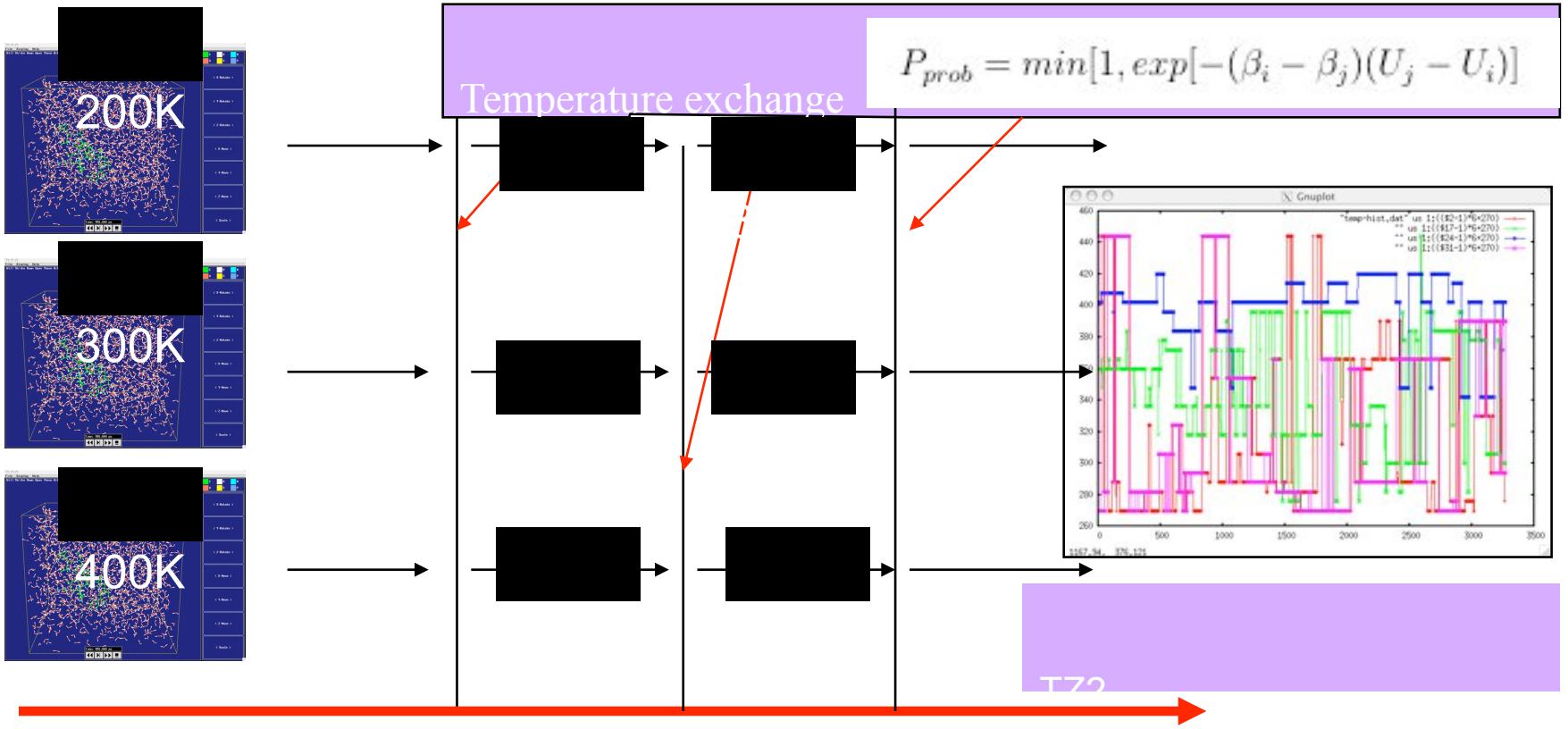
LSU
CENTER FOR COMPUTATION & TECHNOLOGY

Application Manager: Sailent Points

- Uniform: Provides single interface to heterogeneous and distributed resources
- Generic: Infrastructure can be embedded into either a portal or into a GUI
  - Also lightweight, flexible, modular
  - Easy to deploy
- Can support:
  - Other MD packages (e.g., LAMMPS)
  - Other Usage Modes (e.g., High-throughput) (WP1)
  - Complicated workflow driven computation (WP1)

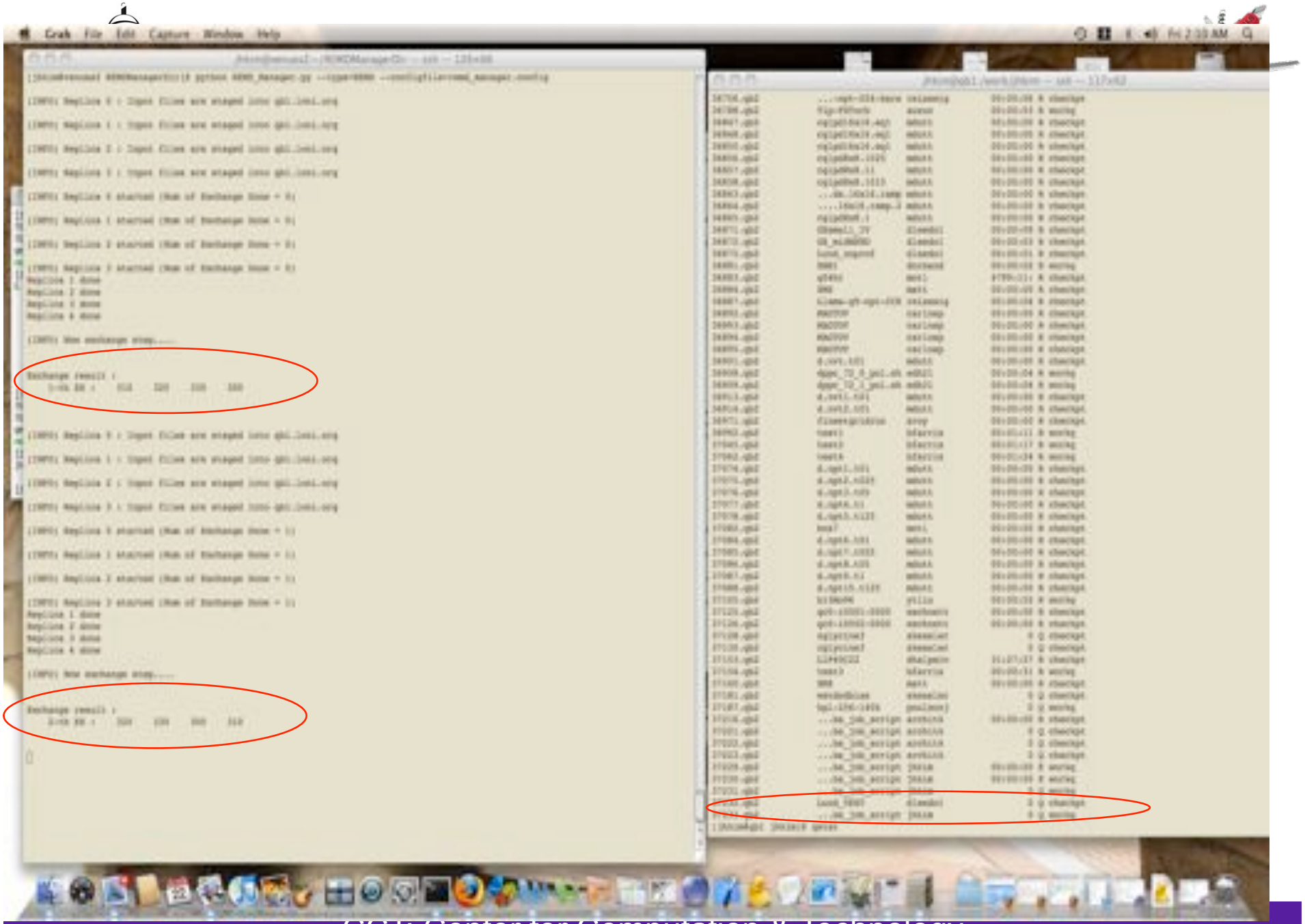
CENTER FOR COMPUTATION & TECHNOLOGY

REMD for β-hairpin folding



- 16-64 replicas
- 250K-500K
- More than 10ns
- PMF via. WHAM or Probability
- Free Energy Surface along a reaction coordinate

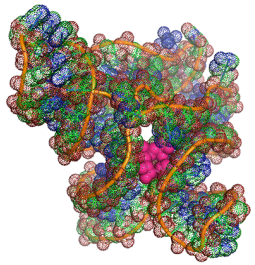
3ns, 30 replicas



```
Replica 1 done
Replica 2 done
Replica 3 done
Replica 4 done

(INFO) Now exchange step....


Exchange result :
    2-th EX :    320    330    300    310



(INFO) Replica 0 : Input files are staged into qbl.loni.org

(INFO) Replica 1 : Input files are staged into qbl.loni.org

(INFO) Replica 2 : Input files are staged into qbl.loni.org

(INFO) Replica 3 : Input files are staged into qbl.loni.org

(INFO) Replica 0 started (Num of Exchange Done = 2)

(INFO) Replica 1 started (Num of Exchange Done = 2)

(INFO) Replica 2 started (Num of Exchange Done = 2)

(INFO) Replica 3 started (Num of Exchange Done = 2)
Replica 3 done
Replica 1 done
Replica 2 done
Replica 4 done

(INFO) Now exchange step....


Exchange result :
    3-th EX :    330    320    310    300



(INFO) Replica 0 : Input files are staged into qbl.loni.org
```

LSU
CENTER FOR COMPUTATION & TECHNOLOGY

REMD Simulation



RNA Riboswitch (SAM-I)

- 50,000 atoms (explicit water)
- 16-32 replicas (2-3 LONI/TeraGrid)
- Each replica : 48-64 cpu mpi job (total : more than 1000 cpus)
- 2-3 days : 10-20 ns for a replica (total : 160 ns-600 ns)

➔ Provides information corresponding to multi-ms time scale dynamics

WP1
Scheduling and Data Services
Demonstrations

Tevfik Kosar, Sumeet Dua, Nate Brenner et al.

CENTER FOR COMPUTATION
& TECHNOLOGY

WP1 in a Nutshell

- Motivation: Enable domain scientists to focus on their primary research problem, assured that the underlying infrastructure will manage the low-level cpu scheduling and data handling issues.
- Use Case: A domain scientist should be able do:
  - Submit a simulation with a single click
    - Which may run on hundreds of processors across the state & access distributed data
  - Get informed when results are ready
- All low level details should be transparent to the domain scientist
  - site selection, scheduling, data movement, fault tolerance, automation ..etc

WP1 Team

- **Senior Personnel:** Allen, Brenner, Katz, Kosar (LSU), Box, Dua (Tech)

- **WP-1 Funded Personnel:**

  Gaduate Students: Esma, Jagadish, Mehmet, Zhiefeng (LSU), Thanadech (Tech)

  Postdocs: TBD

- **WP-1 Supporting Personnel:**

  Staff: Prats, Honggao (LONI), Archit, Andrei (LSU)

  Students: Vinay, Ibrahim, Jack, Ismail, Emir, Sirish (LSU), Pradeep, Harpreep (Tech)

WP1 Progress

- Basic Grid services deployed across LONI
  - Lustre, Globus, Condor, GridFTP
- Distributed storage (PetaShare) deployed across six LONI sites
  - 170 TB usable (220 TB raw), unified name space
- User friendly PetaShare client tools developed
  - petashell, petafs, pcommands, petasearch
- Stork data scheduler enhanced
  - Whole datasets, parallel streams, checksums
- End-to-end workflow management of several science driver applications enabled
- New site selection algorithms developed
- New data mining algorithms developed

WP1 Demonstrations

1. End-to-end workflow management
2. Dynamic site selection
3. Distributed data access & retrieval
4. Protein structure classification tools
5. Medical Image classification tool
6. Discovery of DNA folding units

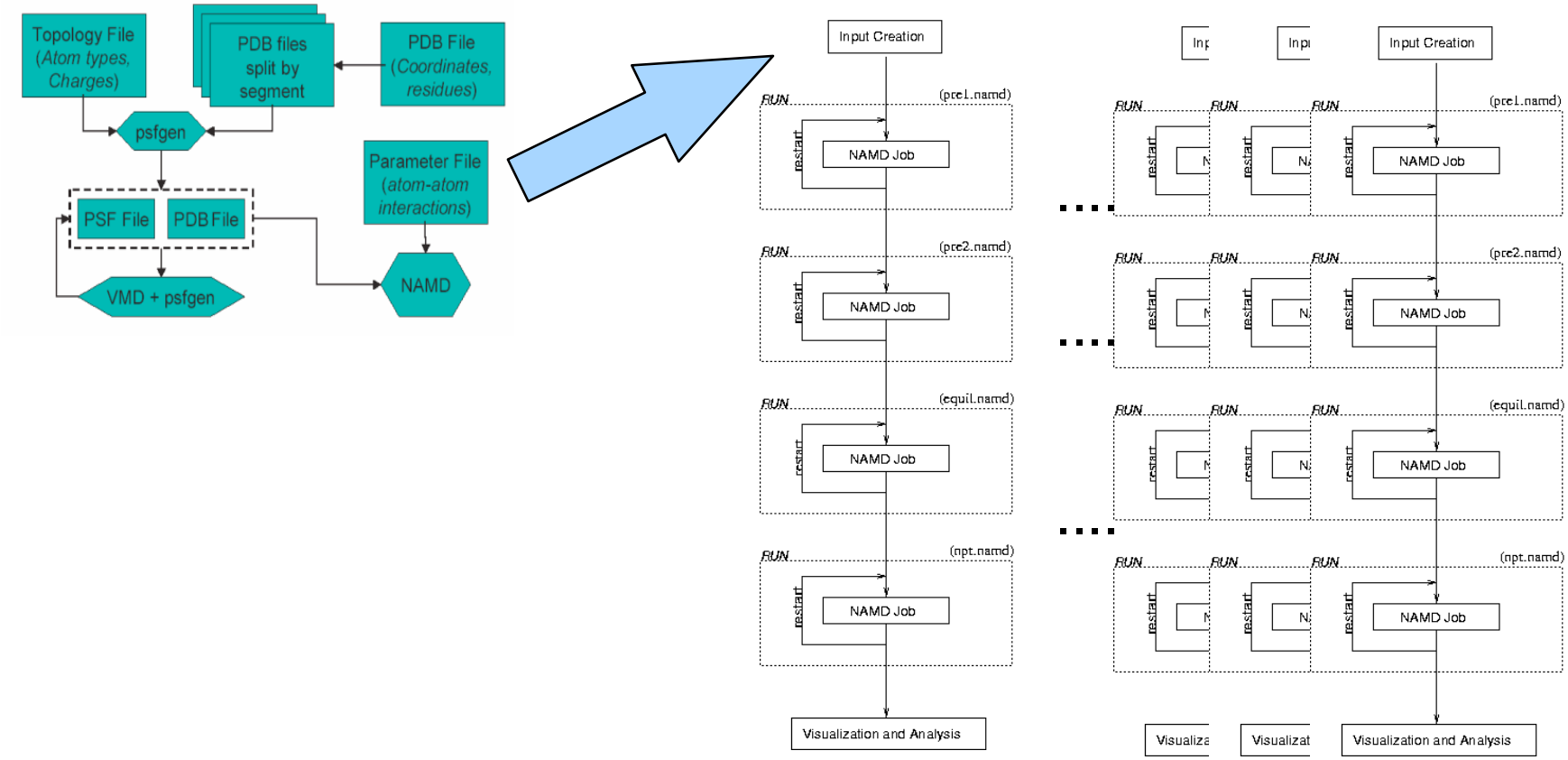
DEMO - 1:

End-to-end Workflow Management for DNA folding

E. Bahsi, T. Kosar (LSU), T. Bishop (Tulane)

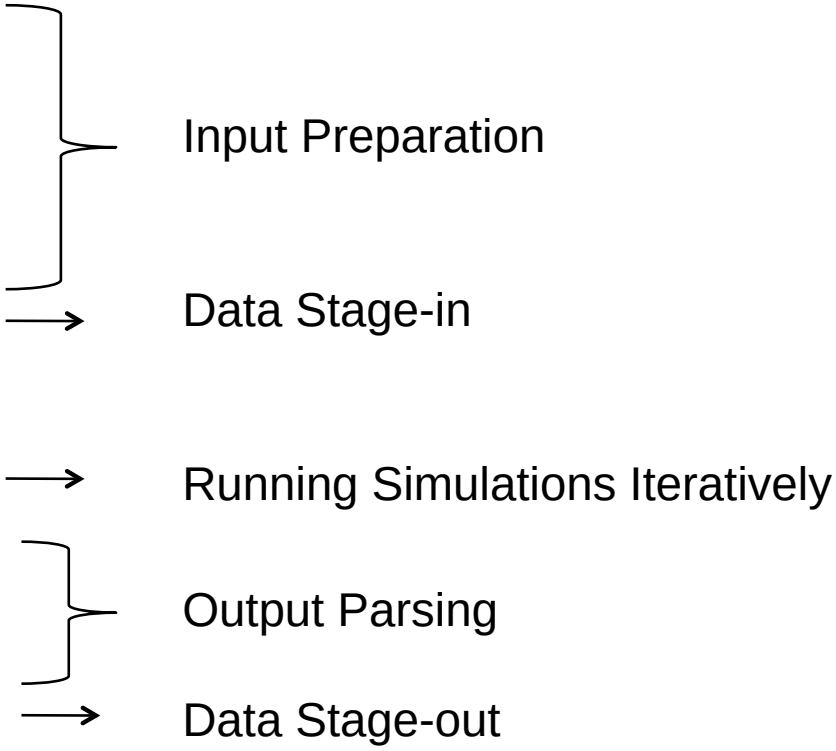
Biosensors: MD Fast Track Study

A high throughput simulation workflow:



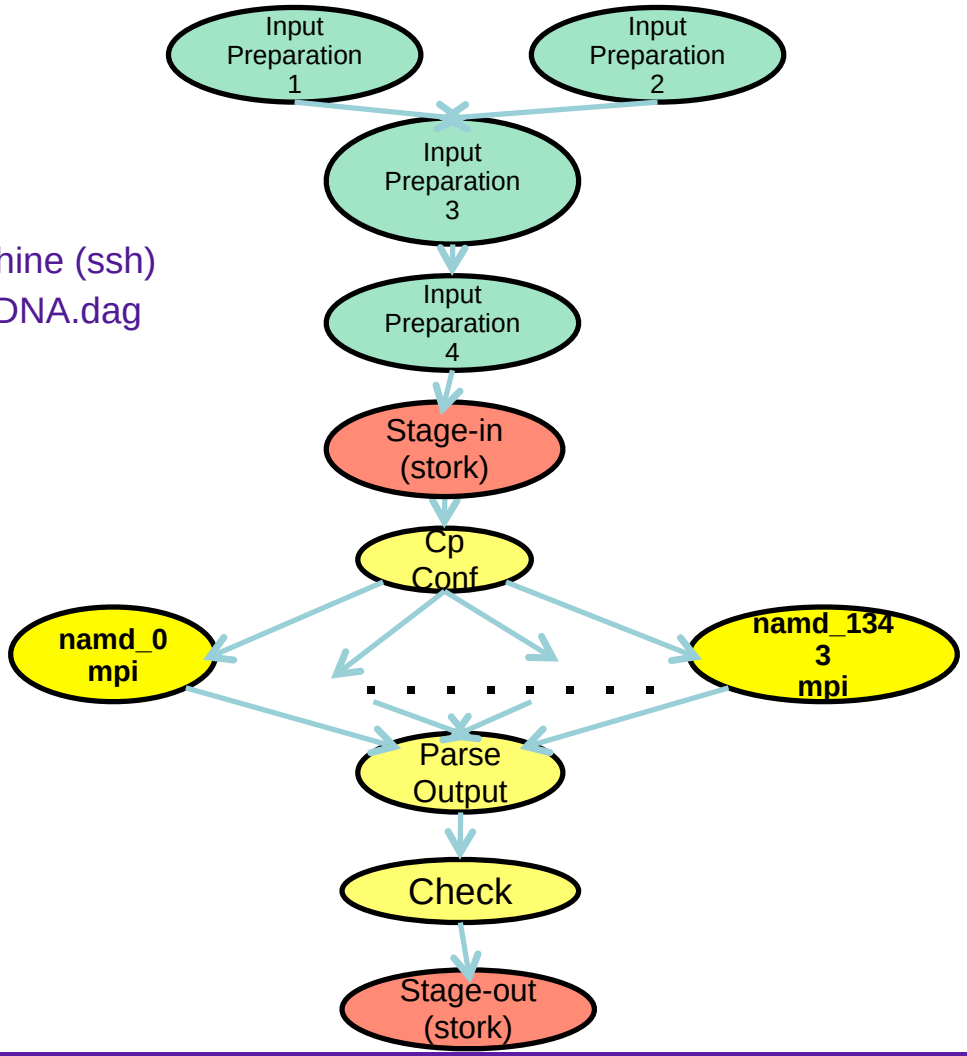
Running DNA Folding Application step-by-step (Before)

1. Connect to local machine (ssh)
2. Run 01-setup.tcsh
3. Run 02-mk-dna.awk
4. Run 03-setup-amber.tcsh
5. Run 04-setup-sims.tcsh
6. Run 05-rsync
7. Connect to cluster (ssh)
8. Run 06-namd
9. Run 07-min1.analysis
10. Run 08-check.sims
11. Run 09-rsync
12. Connect to local machine (ssh)



Workflow-enabled Application (After)

1. Connect to local machine (ssh)
2. Condor_submit_dag DNA.dag



Advantages

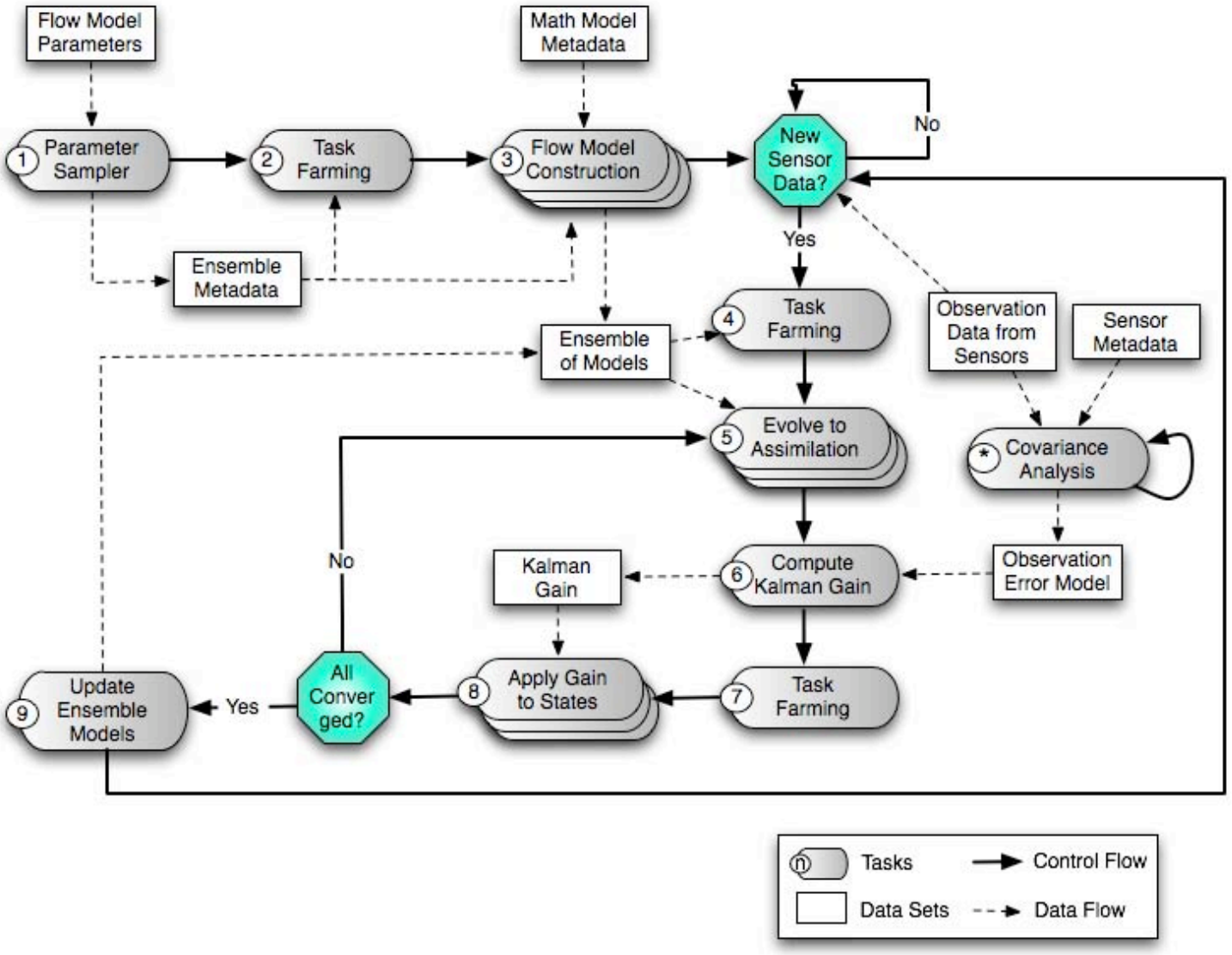
- Babysitting for workflow
- Stork for Data Transfer
- Parallelization of mpi jobs
- Submit file Generation

DEMO - 2:

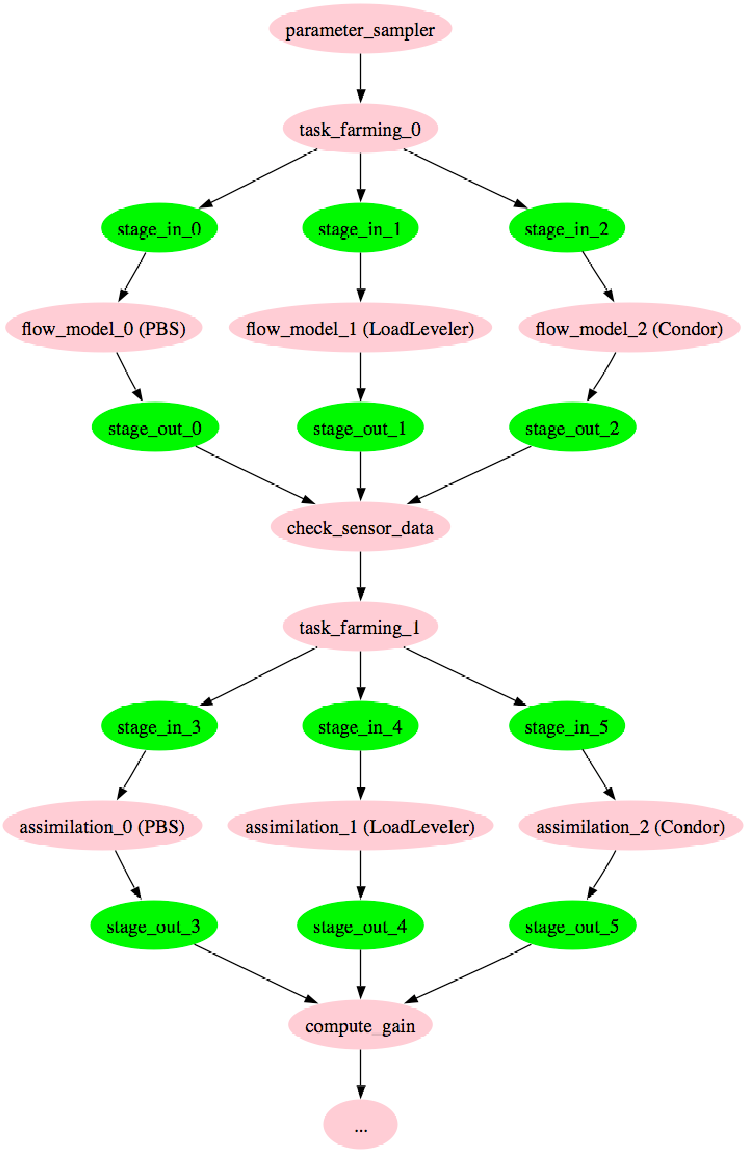
Dynamic Site Selection for Reservoir Modeling

E. Bahsi, T. Kosar, G. Allen, M. Tyagi, C. White (LSU)

Reservoir Modeling Workflow



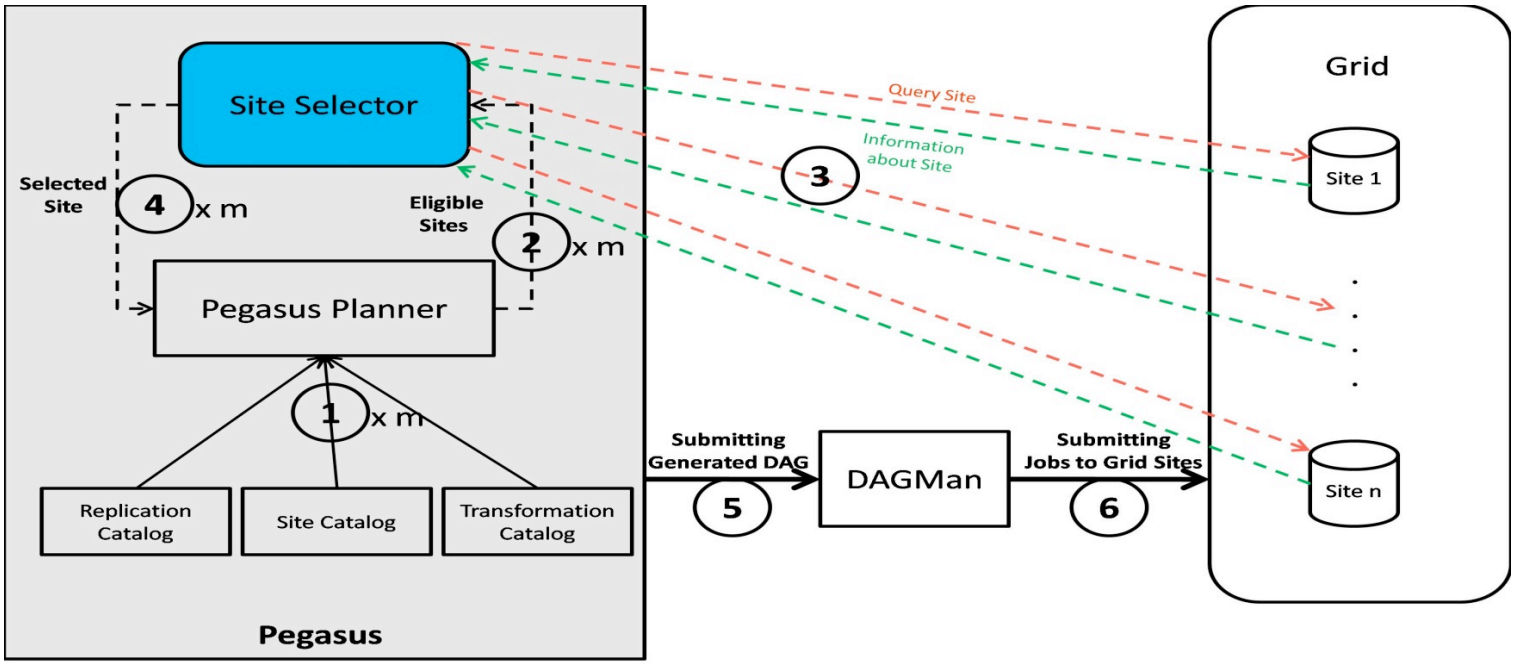
Concrete Workflow Mapping



CENTER FOR COMPUTATION & TECHNOLOGY

Site Selection Mechanism

- Two Site Selectors are implemented
- Querying Sites for information about jobs and queue (# of free nodes, total # of nodes, # of jobs in the queue)



DEMO - 3:

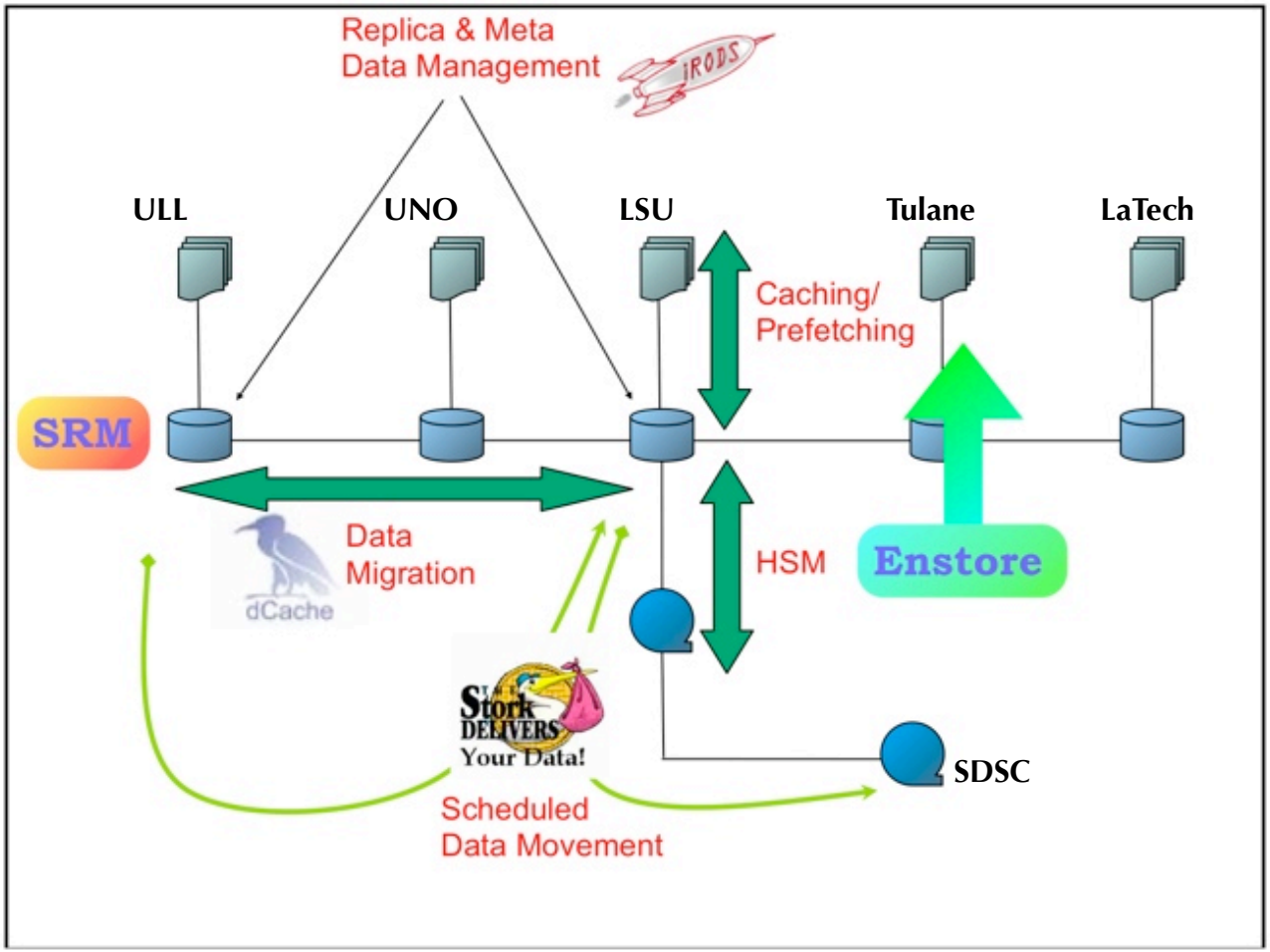
Distributed Data Access & Retrieval

I. Akturk, T. Kosar, X. Wang (LSU) et al.

PetaShare Core

petafs

User-level Virtual File System
- NO need to change OS/kernel
- NO need to change code
- NO relinking
- NO recompiling



petashell

POSIX Shell Interface
- All of the above
- Without privileged access

Web interface:

PetaSearch

DEMO - 4:

Protein Structure Classification Tool

P. Chowriappa, S. Dua (LaTech), H. Thompson (LSUHSC)

Synopsis of Cybertools Efforts

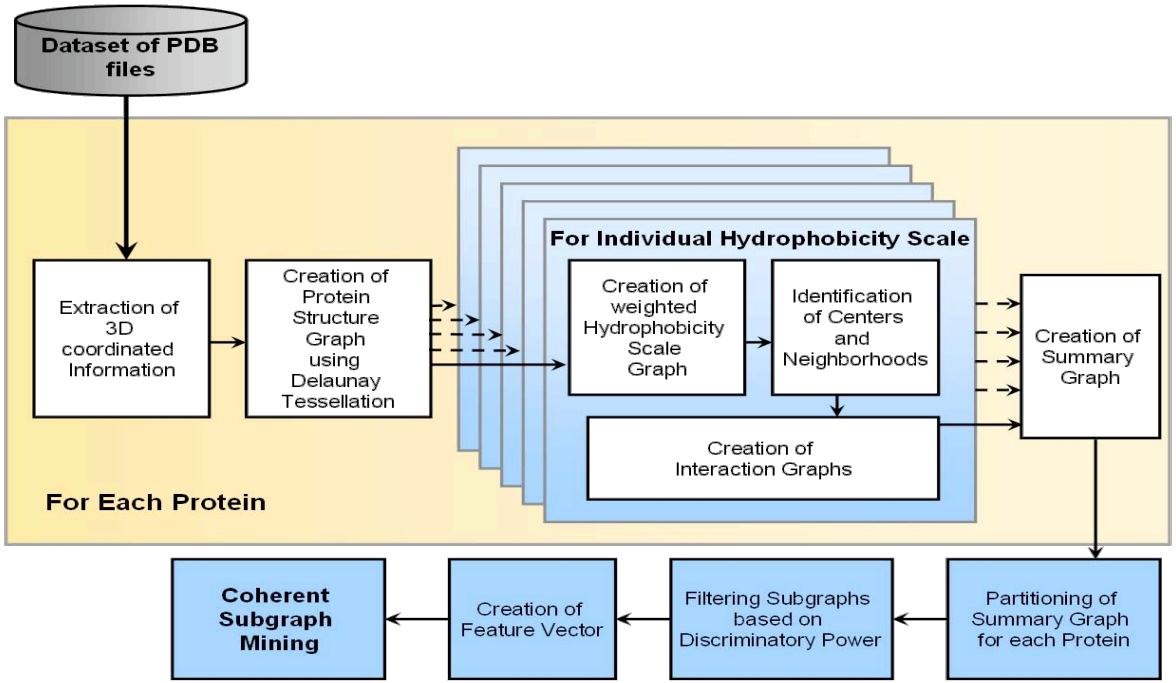
(S. Dua et al. @ LA Tech, H. Thompson et al. @

1. Information fusion algorithms (automated metadata extraction and information retrieval for data mining)
   - Fusion of stereochemical properties for automated protein core discovery and classification
   - Fusion of synchronization experiments in gene expression analysis (and gene ranking)
- Medical Image Classifier systems
   - Patient classification for Diabetic Retinopathy images

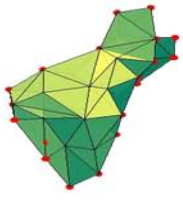
Information fusion: Integration of protein stereochemical properties for

- Protein sequence based tools are not sensitive enough to discover similarity between proteins because of the exponential growth in diversity of sequences.
- We have developed a Graph Theory based Data Mining Framework to extract and isolate protein structural features that sustain invariance in evolutionary proteins.

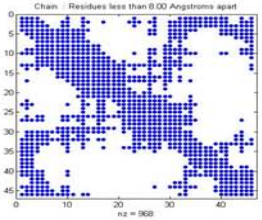
We have hypothesized that proteins of the same homology contain conserved hydrophobic residues that exhibit analogous residue interaction patterns in the folded state.



Methodology

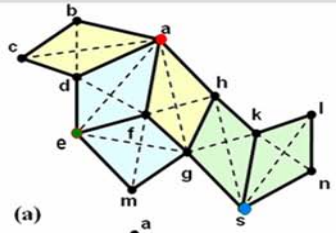


(a)

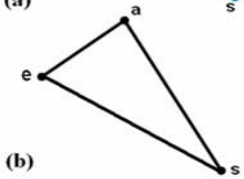


(b)

Creation of Protein Structure Graphs $G_{Hydn}(P)$
Example: SCOP Class Small Protein: 1CRN(A)
(a) Protein Structure Graph using Delaunay Tessellations with distance set at 8 Angstroms
(b) The Adjacency Matrix of the Protein Structure Graph



(a)



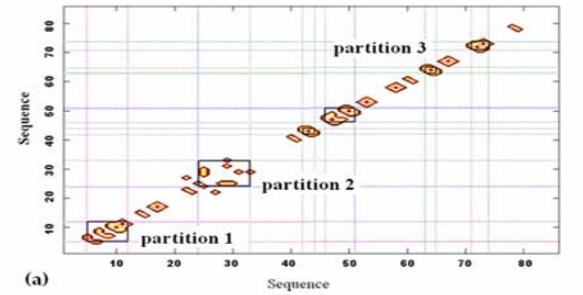
(b)

t_1	a=10	b=7	c=8	d=5
t_2	a=10	e=12	f=9	d=5
t_3	a=10	f=9	g=3	h=6
t_4	e=12	f=9	g=3	m=7
t_5	s=9	g=3	h=6	k=4
t_6	s=9	k=4	l=3	n=5

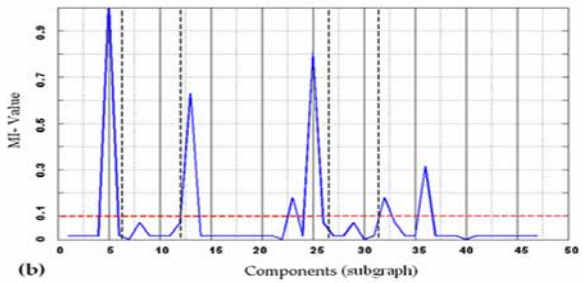
$C(t_1)$=a $C(t_2)$=e $C(t_3)$=a $C(t_4)$=e
$C(t_5)$=s $C(t_6)$=s

$H(a)=\{t_1,t_3\}$ $H(e)=\{t_4,t_2\}$
$H(s)=\{t_5,t_6\}$

Identification of centers and neighborhoods.
a) Representation of the central residue "C" uniquely colored as "e" (green), "a" (red), "s" (blue), and their respective neighborhoods "H" (identified here by the shades of tetrahedra). The table shows each tetrahedron (t), its respective residues, and its respective weights.
b) The resultant Interaction Graph (IG), where a proximity edge is drawn between centers if they share vertices in common.

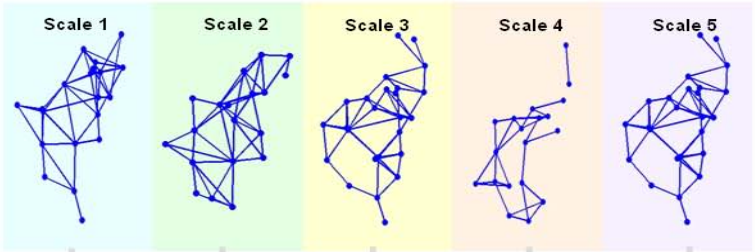


(a)



(b)

Summary Graph of protein 1CRN(A)
a) Identification of connected components
b) Filtering of connected components using Mutual Information



With five different hydrophobicity scales, we obtain a set of five Interaction Graphs *(IG)* representing protein 1CRN(A). The vertices for each *IG* have a common vertex set *V* ($G_{Hydn}(P)$), but possess different edge sets.

Protein Mining (snapshot of results)

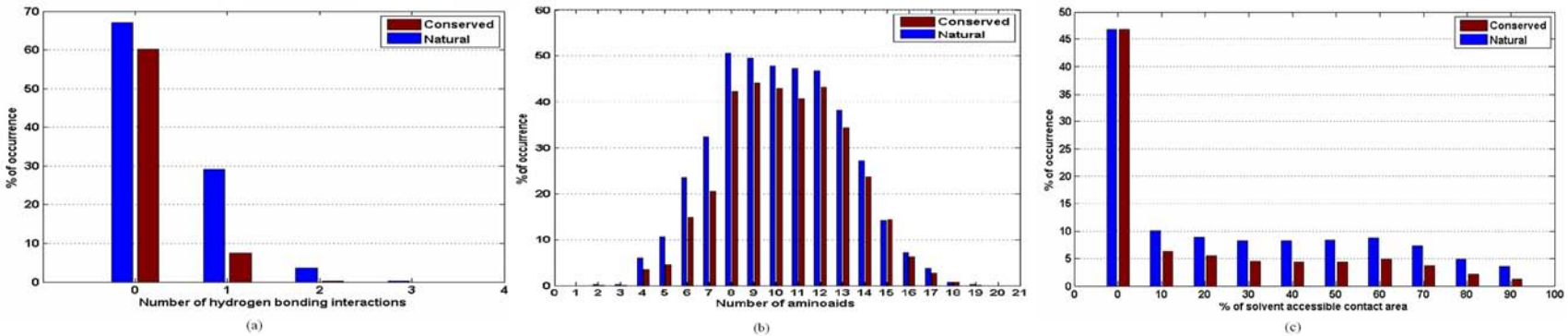
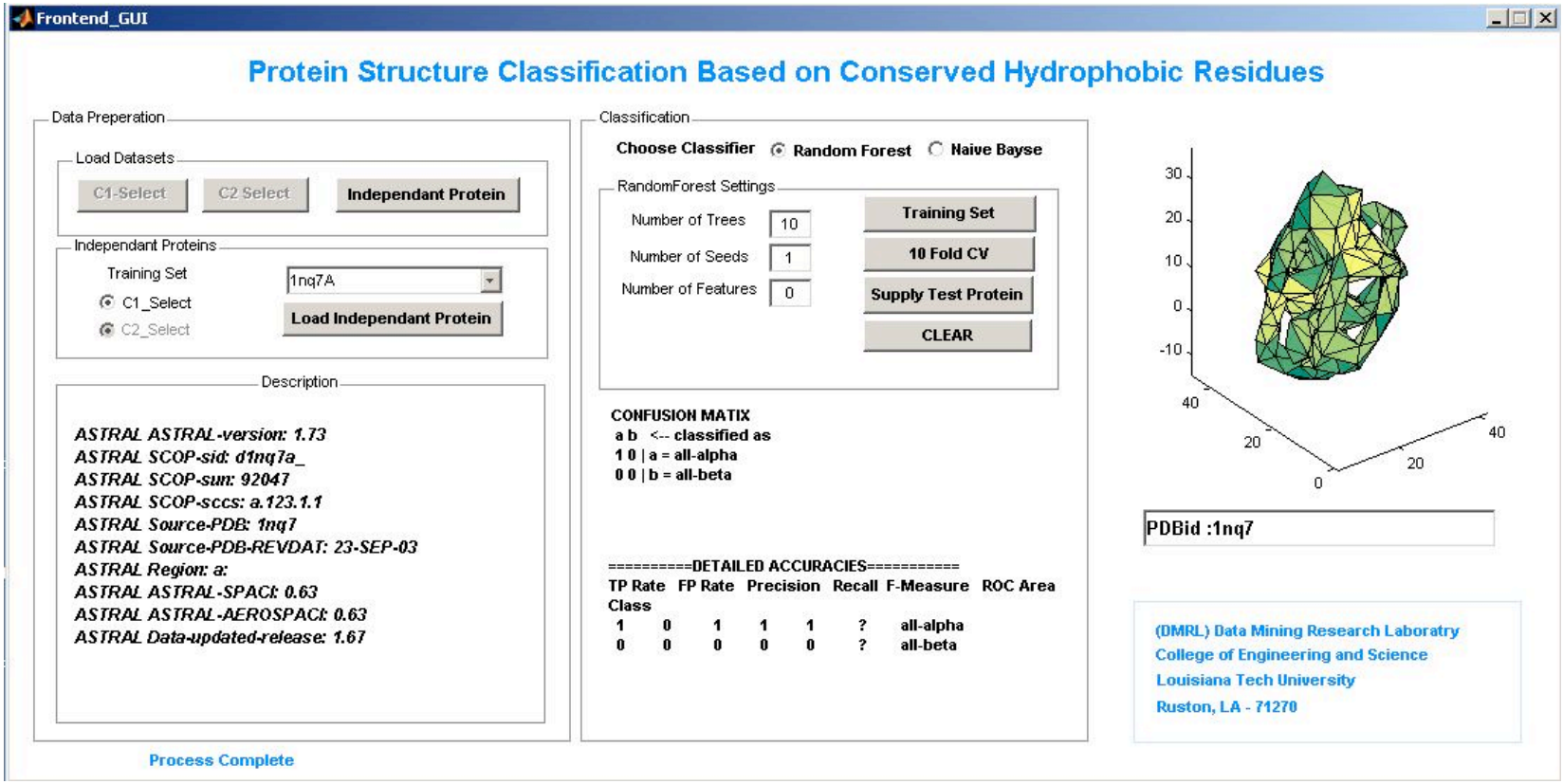


Fig. Composition of amino acids in conserved residues of the summary graphs compared with the entire protein representative set. On the Y-axis is the percentage of amino acids and on the X-axis: a. hydrogen bonding interactions, b. Ooi number in an 8 Å radius around the amino acid and c. solvent accessible contact area as a percentage of residue accessibility.

•Ref.: P. Chowriappa, S. Dua, J. Kanno and H. Thompson, *"Protein Structure Classification Based on Conserved Hydrophobic Residues"*, to appear in the *IEEE/ACM Transactions on Computational Biology and Bioinformatics.*

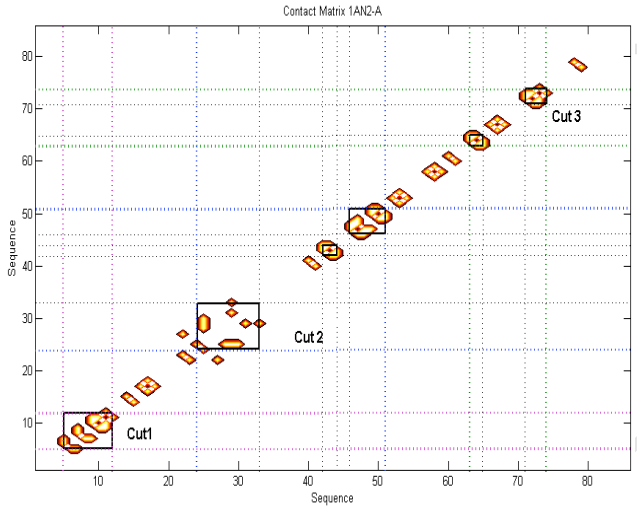
•Ref.: S. Dua, P. Chowriappa and R. Rajagopalan, "*Spectral Coherence Feature Extraction from Stereochemical Scales for Protein Classification", under review for IEEE/ACM Transactions on Computational Biology and Bioinformatics.*

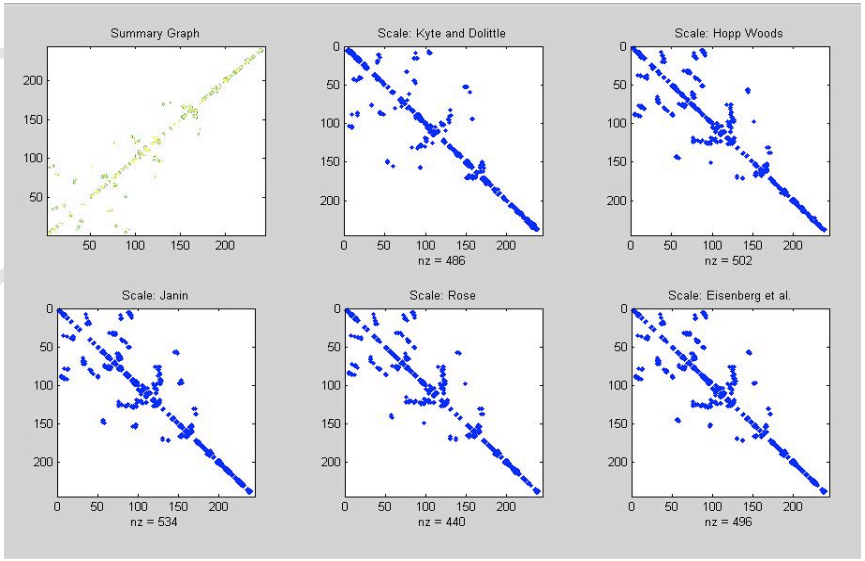
Tool Features



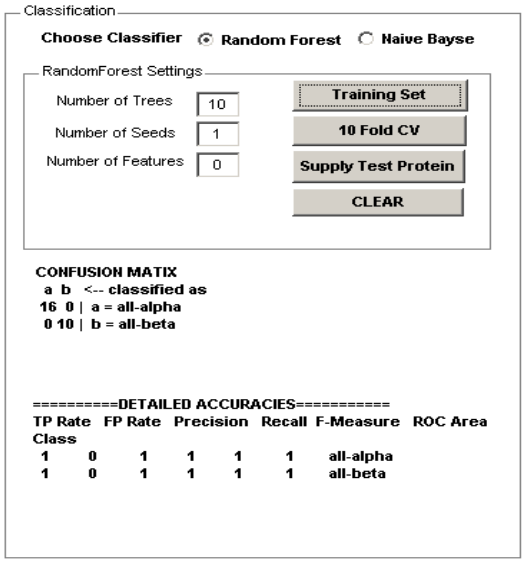
- Provides for the identification of conserved regions within proteins of the same family
- Integration of five physico-chemical properties
- Classification using Random Forest and Naïve Bayes classifier
- Provides for classification of independent proteins into specific classes

In Depth Analysis



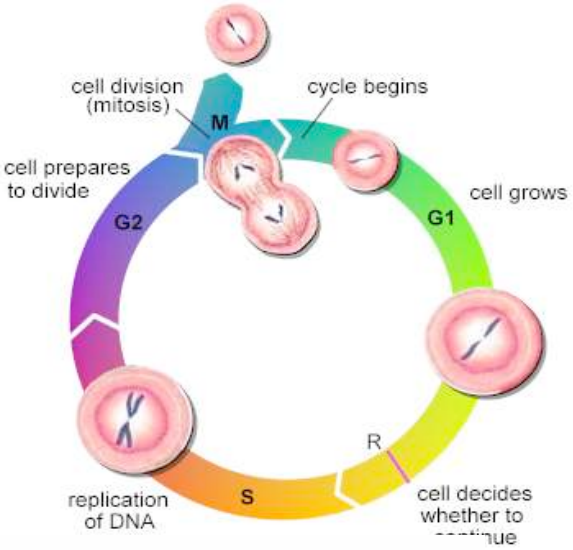


- Provides a graphical representation of the Summary Graph for better viewing of conserved hydrophobic residues
- Gauge the classification performance using standard measures of calibration



Information fusion: Gene Ranking through fusion of Synchronization

- The *cell cycle,* or *cell-division cycle*, is the series of events that take place in a cell leading to its replication.
- The cell-division cycle is one of the most fundamental processes of life, allowing cells to multiply and faithfully pass on their genetic information to future generations.
- The first critical task in understanding such cyclic systems is to identify the genes that are periodically expressed during the cell cycle – focus of our work.



Our Approach

A. DATA PREPROCESSING
A.1 Interpolation
A.2 Denoising
A.3 Intersecting Gene Sets
A.4 Normalization

→ **B. DATA ALIGNMENT** →

C. INFORMATION FUSION
cdc15 ⊕ cdc28

↓

D. FEATURE EXTRACTION
D.1 Principal Component Analysis
D.2 Skewness Moment
D.3 Kurtosis Moment

←

E. CLASSIFICATION USING GENETIC ALGORITHMS
E.1 Training: Top-ranked (300, 500, 800) genes
E.2 Testing: Published benchmark (B1, B2 and B3: known cyclic genes)

Gene ranking (snapshot of

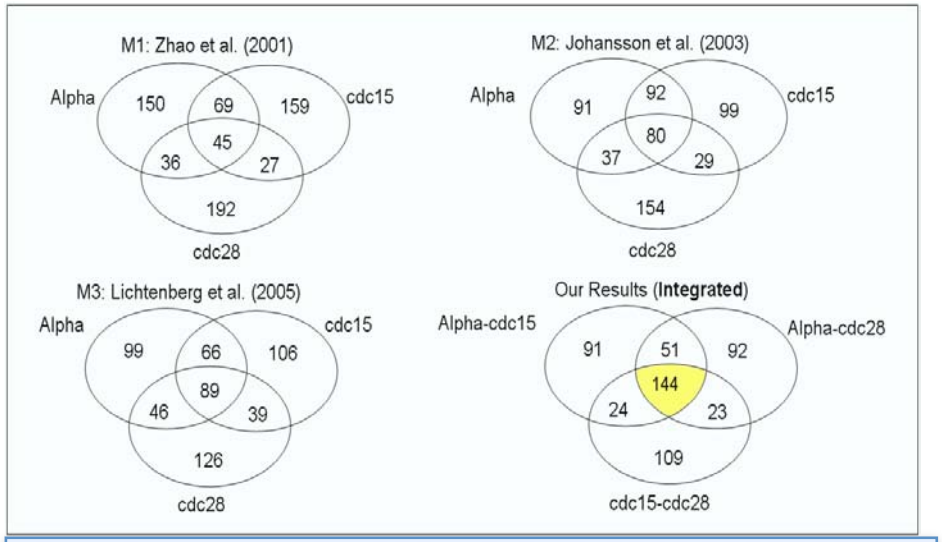


Fig. Agreement across experiments. Venn diagram based on the top 300 genes from each experiment are shown for the methods that provide ranked lists for the individual and integrated experiments.

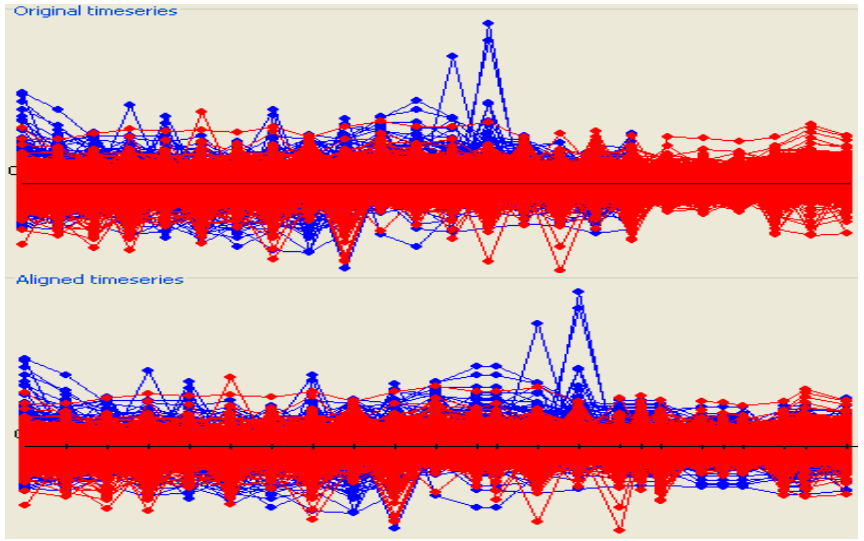


Fig. Data alignment for alpha and cdc15 datasets.

References: A. Alex, S. Dua, P. Chowriappa, "Gene Ranking through the Integration of Synchronization Experiments", to appear in the Proceedings of 2008 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (IEEE-CIBCB08).
S. Dua, P. Chowriappa and A. E. Alex; "Ranking through Integration of Protein-similarity for Identification of Cell-cyclic Genes", to appear in the Proceedings of the Biotechnology and Bioinformatics Symposium (BIOT-2008).

Conclusion and Directions Information Fusion and Data Mining

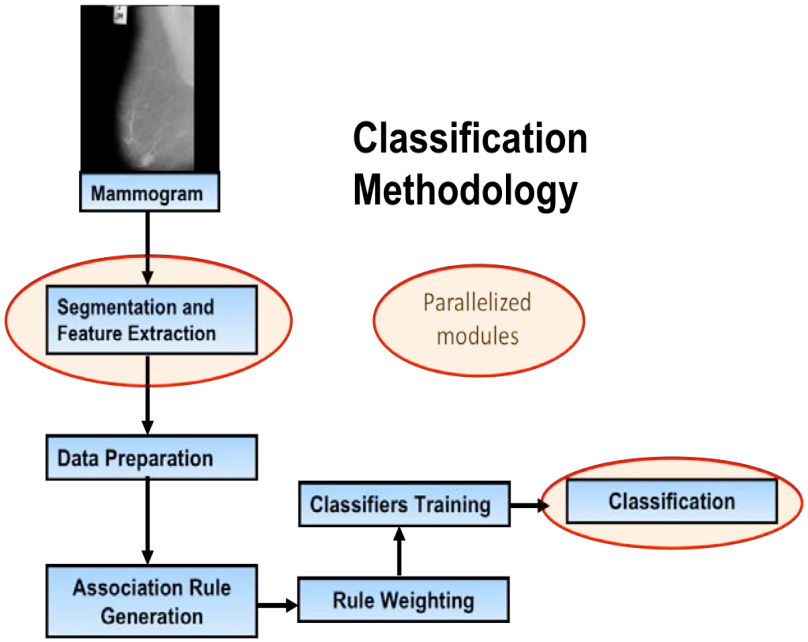
- In conclusion, the work has demonstrated evaluation studies on independent sets of protein classes for performance benchmarking purposes.
  - Other uses: hypothesis generation, protein model verification, and classification.
  - 1 IEEE-TCBB, 1- IEEE-CIBCB and 1-BioT publication.
- The work is a result of collaboration between investigators from:
  - Louisiana Tech University
  - Louisiana State University Health Sciences Center at New Orleans.
- Have an independent tool to share with biologists (available through our website).
  - Port tool for specific protein biotechnologist from LSUHSC (April-09, thanks to H. Thompson)
- Current effort: We are developing an efficient parallelized version of the algorithm for analyzing entire PDB (Oct. 2008).

DEMO - 5:

Medical Image Classification Tool

S. Dua, H. Singh (LaTech), H. Thompson (LSUHSC)

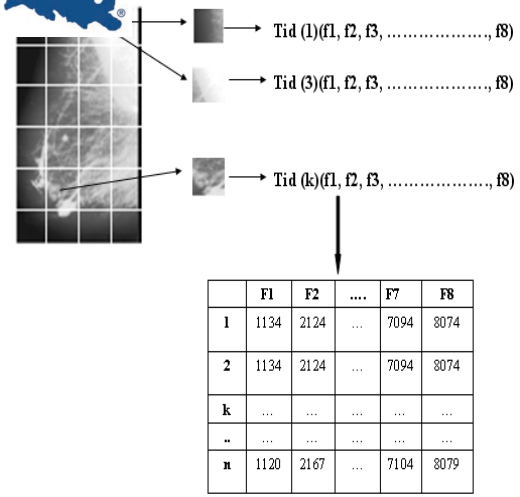
Mammogram Classification using Weighted Rules based Classification



- We have developed a novel method for the classification of medical images (mammograms) using a unique weighted association rule based classifier.
- Isomorphic association rules are derived between various texture components extracted from segments of images,
- These discriminatory rules are then used for the classification through exploitation of their intra- and inter-class

• Rigorous experimentation has been performed to evaluate the rules' efficacy under different classification scenarios.
• The algorithm delivers accuracies as high as 89%, which far surpasses the accuracy rates of other rule based classification techniques.

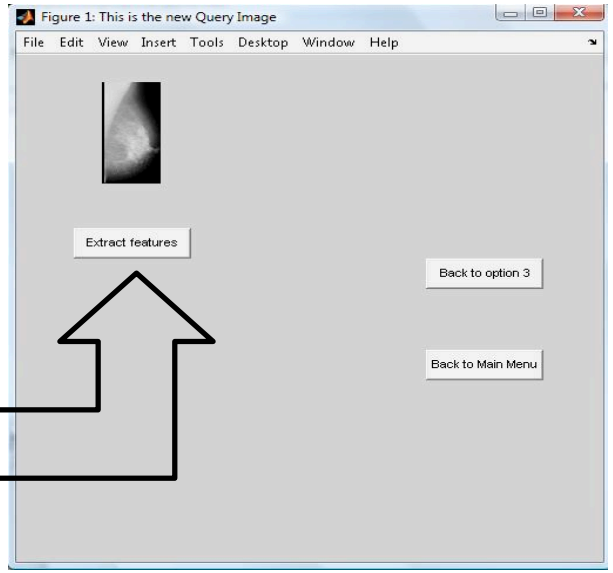
Segmentation and Feature Extraction



	F1	F2	....	F7	F8
1	1134	2124	...	7094	8074
2	1134	2124	...	7094	8074
k	...	...	...	...	...
..	...	...	...	...	...
n	1120	2167	...	7104	8079

- **Each image is divided into NxN segments**
- **Eight texture features extracted from each segment**

Figure 2. Segmentation and feature extraction



Click feature extraction

Feature Label.	Feature	Calculation
F1	Energy	$\sum_{i=0}^{n}\sum_{j=0}^{n}\{p(i,j)\}^2$
F2	Contrast	$\sum_{i=0}^{n}\sum_{j=0}^{n}(i-j)^2 p(i,j)$
F3	Local Homogeneity	$\sum_{i=0}^{n}\sum_{j=0}^{n}\frac{p(i,j)}{1+(i-j)^2}$
F4	Correlation	$\sum_{i=0}^{n}\sum_{j=0}^{n}((ij)p(i,j)-\mu_x\mu_y)/\sigma_x\sigma_y$
F5	Entropy	$-\sum_{i=0}^{n}\sum_{j=0}^{n}p(i,j)\log p(i,j)$
F6	Cluster Shade	$\sum_{i=0}^{n}\sum_{j=0}^{n}(i-M_x+j-M_y)^3 p(i,j)$
F7	Information measure of correlation	$Hxy-Hxy1/\max\{Hx,Hy\}$
F8	Maximum Probability	$\max_{i,j} P(i,j)$

where,

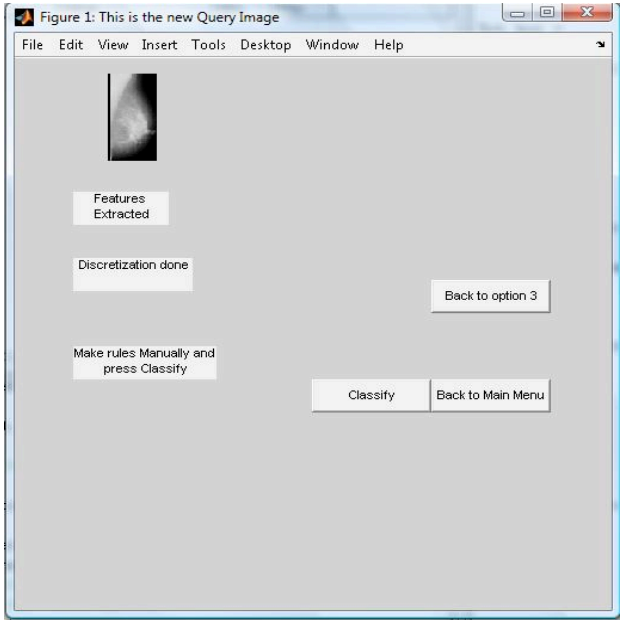
$M_x=\sum_{i=0}^{n}\sum_{j=0}^{n}ip(i,j)$ $M_y=\sum_{i=0}^{n}\sum_{j=0}^{n}jp(i,j)$

$P_x=\sum_{j=0}^{n}p(i,j)$ $P_y=\sum_{i=0}^{n}p(i,j)$

$Hx=-\sum_{i=0}^{n}P_x(i)\log P_x(i)$, $Hy=-\sum_{j=0}^{n}P_y(j)\log P_y(j)$

$Hxy1=-\sum_{i=0}^{n}\sum_{j=0}^{n}P(i,j)\log\{P_x(i)P_y(j)\}$

Table 1. Texture features



Classification

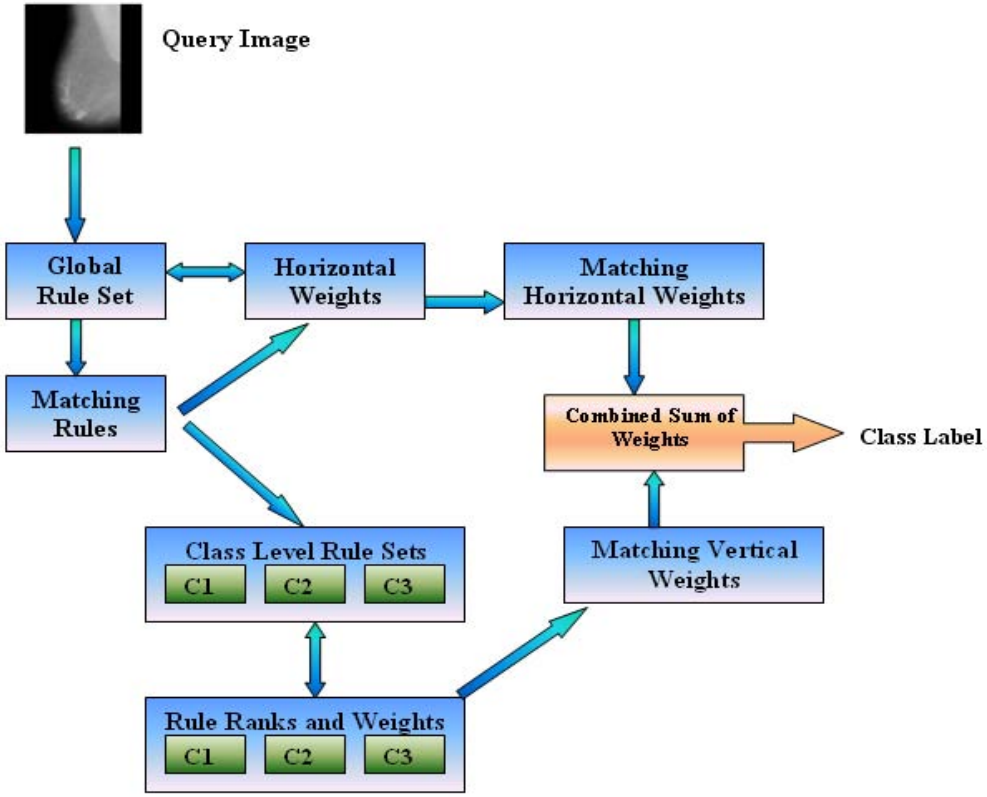
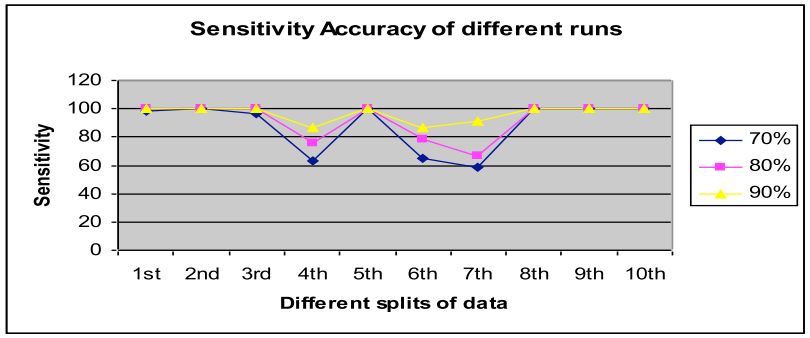


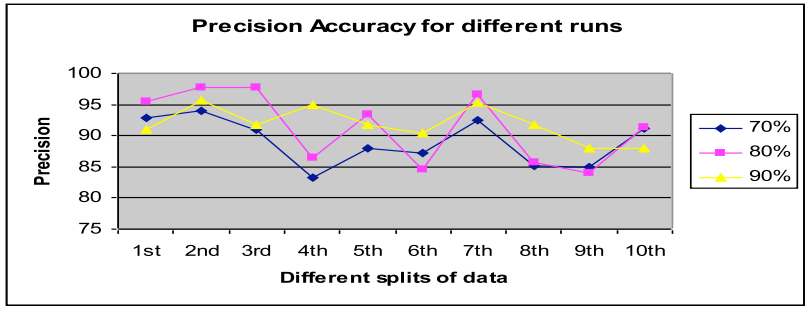
Figure. Classification Mechanism

- Form horizontal weights of rules
- Form vertical weights for rules
- Take query image and find matching rules
- Find corresponding horizontal and vertical weights
- Add these weights to form cumulative sum
- Classify to the class with highest weight
- Display images from same class

Mammogram classification
(snapshot of results)



(a)



(b)

The change of Precision (a) and Recall (b) with different percentages of training versus testing data.

True Classes / Reported Classes

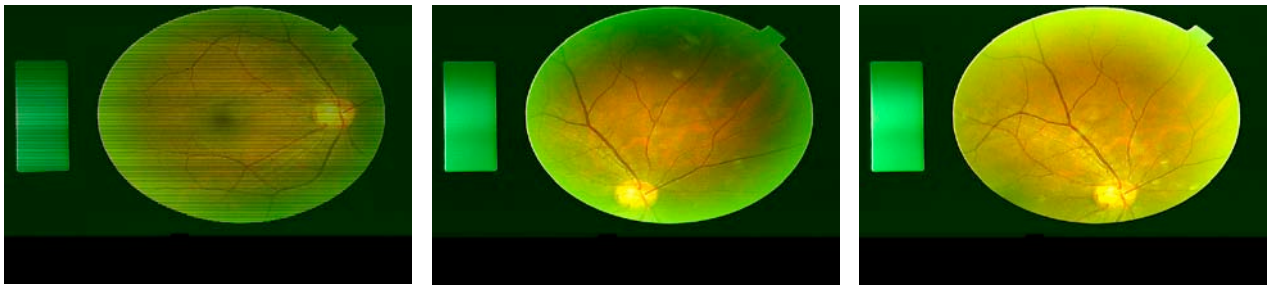
	Normal	Benign	Malign
Normal	22	0	0
Benign	1	5	0
Malign	1	0	3

The confusion matrix for three classes considered for classification. The number indicates the number of cases reported.

Reference: S. Dua, H. Singh, H.W. Thompson, "Associative Classification of Mammograms using weighted Rules based Classification", under review for Expert Systems and Applications Journal (Elsevier).

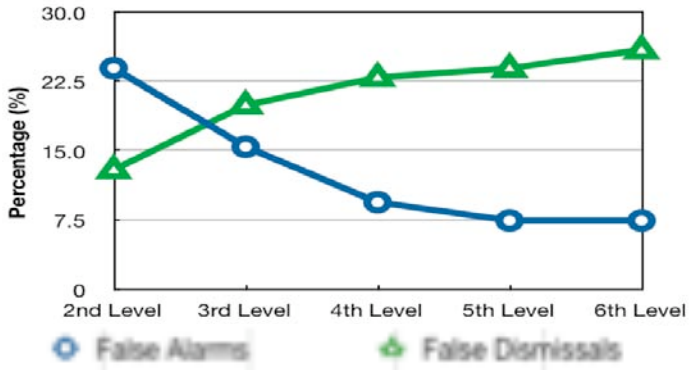
Diabetic Retinopathy Patient Classification

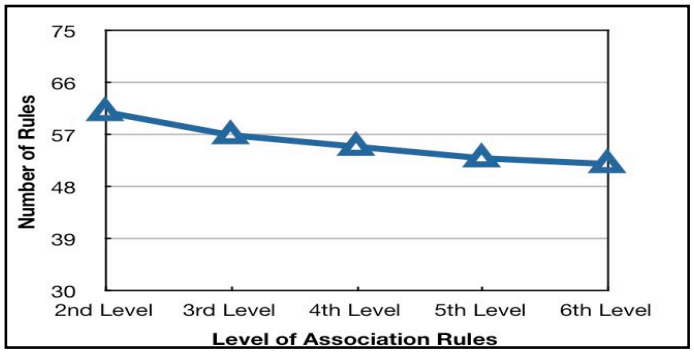
- Patient classification in medical imaging has a range of applications spanning both the biomedical and healthcare delivery domains.
- We have developed a unique classifier for automated integration and classification of images of patients .
- Patients were suffering from either Non-proliferative Diabetic Retinopathy (NPDR) or Proliferative Diabetic Retinopathy (PDR).



Diabetic Retinopathy Patient

Patient set id	Common rules	FA (avg.)	FD (%)	FA (%)
p01	42	455	0	30
p02	309	409	0.48	24
p03	4	420	0	33
p04	15	351	3.6	30
p05	15	465	0	36
p06	40	505	15	32
p07	728	114	0.14	9
p08	27	457	0.92	29
p09	671	101	0.4	8





Reference: S. Dua, V. Jain, H.W. Thompson, "Patient Classification using Association Mining of Clinical Images", appeared in the Proceedings of The Fifth IEEE International Symposium on Biomedical Imaging (ISBI '08).

Conclusion and Directions Image Classification

- We can autonomously classify images based on discovered content, rather than user-supplied metadata.
  - 1 IEEE-ISBI publication, 1 under review.
- The work is a result of collaboration between investigators from:
  - Louisiana Tech University
  - Louisiana State University Health Sciences Center at New Orleans.
- The tool is not specific to mammograms or DR images.
  - Can we easily extended (without recoding) to other image domains.

DEMO - 6:

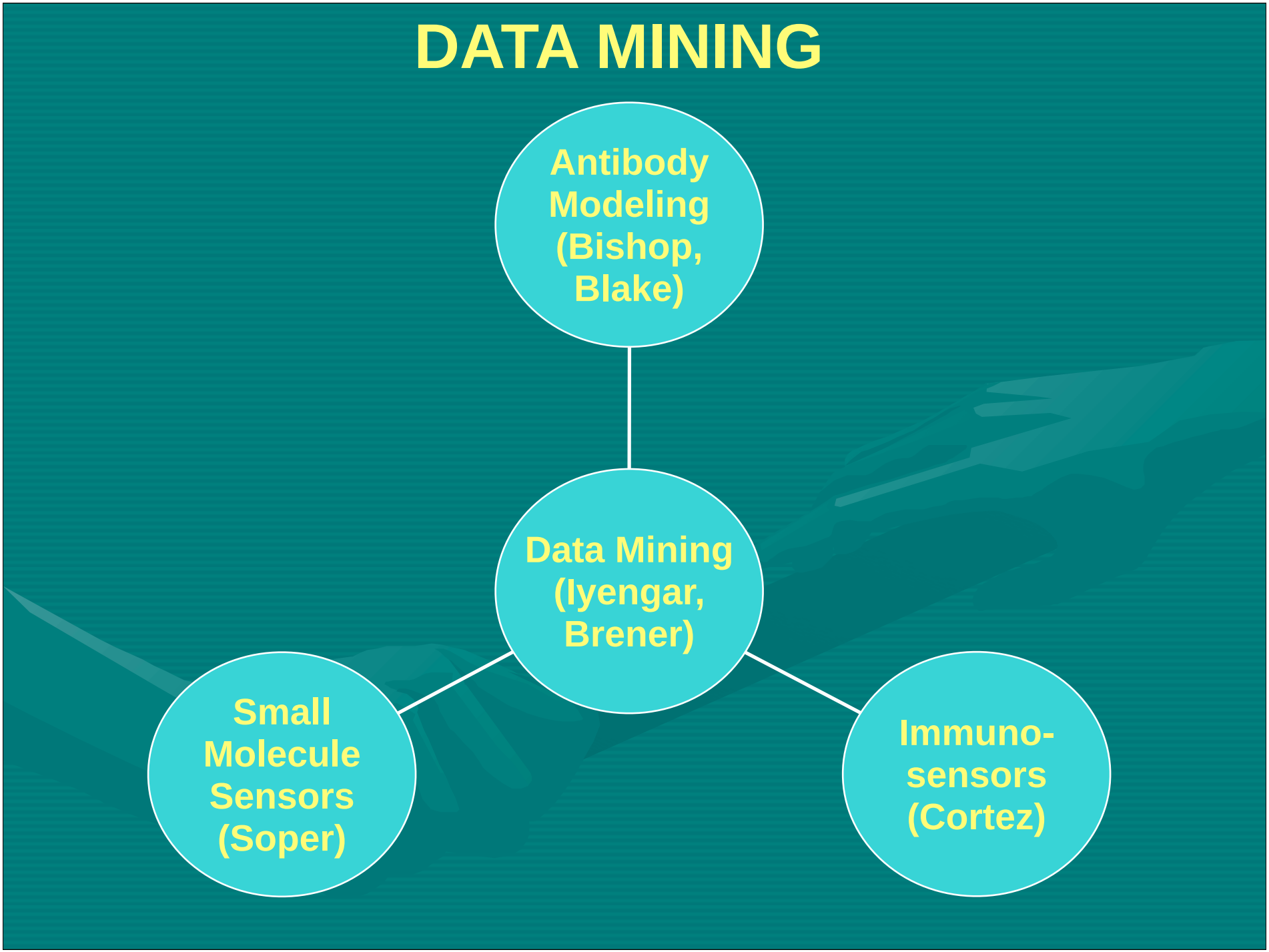
DNA Folding Units Discovered by Data Mining

N. Brenner et al (LSU)

IMAGE FUSION AND DATA MINING

Faculty:	**Dr. S. Sitharama Iyengar (LSU)** **Dr. Nathan E. Brener (LSU)** **Dr. Bijaya B. Karki (LSU)** **Dr. Hilary Thompson (LSUHSC)**
Project Coordinator:	**Dr. Dimple Juneja**
Graduate Students:	**Dr. Hua Cao** **Rathika Natarajan** **Archit Kulshrestha** **Harsha Bhagawaty** **Asim Shrestha** **Jagadish Kumar** **Gaurav Khanduja** **Dipesh Bhattarai**
Integration with other investigators:	**Dr. Allen, Dr. Acharya, Dr. Bishop, Dr. Blake, Dr. Soper**
Collaborators:	**LSU Health Sciences Center (LSUHSC)** **LATech** **Air Force Institute of Technology**

DATA MINING


Antibody Modeling (Bishop, Blake)
Data Mining (Iyengar, Brener)
Small Molecule Sensors (Soper)
Immuno-sensors (Cortez)

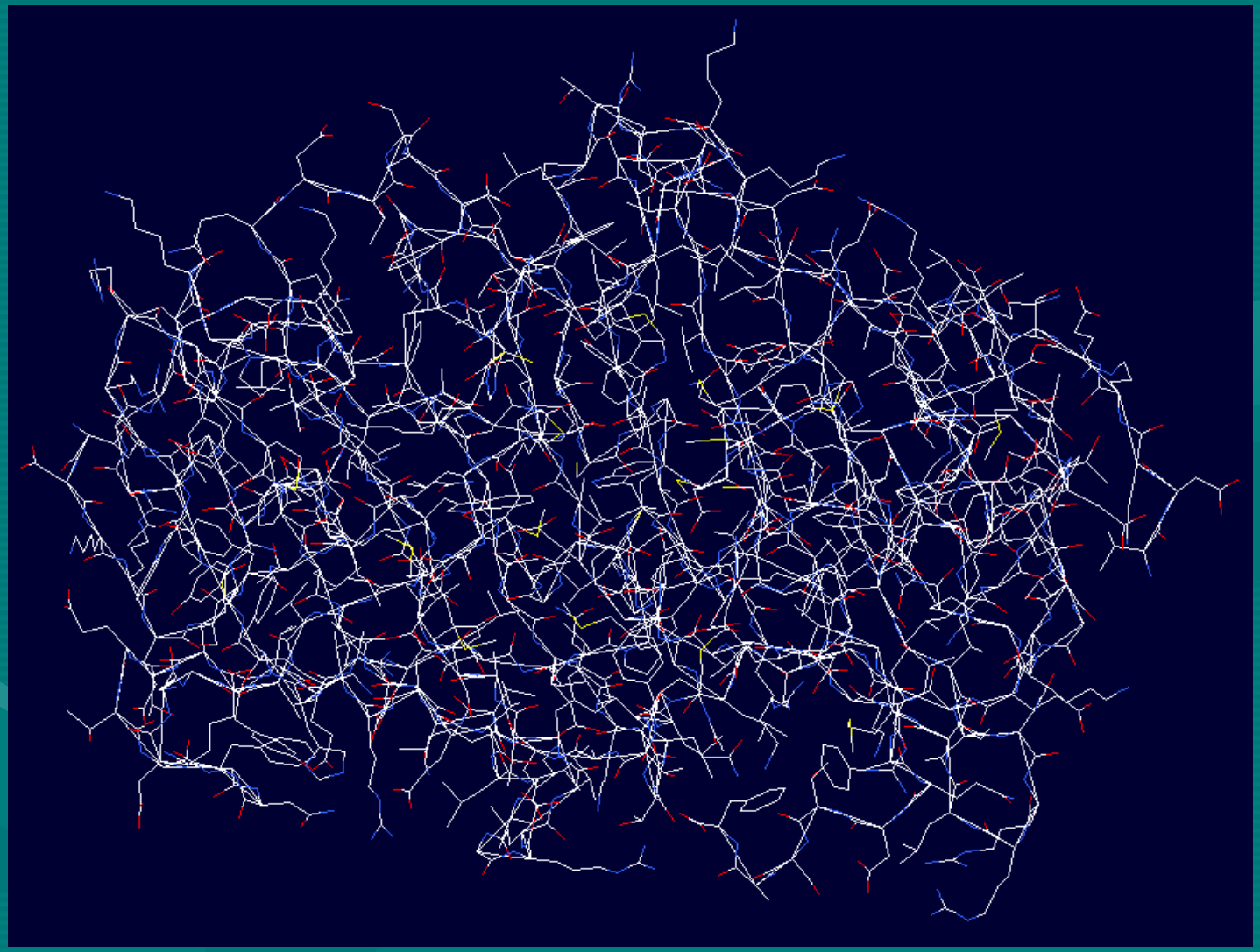
Data Mining Algorithms

- Searching for features of interest in large data sets
- Potential CyberTools applications:
  - Antibody modeling (Bishop, Blake)
  - Small molecule sensors (Soper)
  - Immunosensors (Cortez)
- Test problem
  - Protein Databank (PDB). Look for common protein folding units (can be of variable length)

New Data Mining Algorithm

- New efficient clustering algorithm to classify proteins according to common folding units. Based on conformational angle representation to reduce parameters.
  - Represent the protein structure as a series of conformational angles
  - Partition the proteins into fragments (folding units) of a specified size
  - Cluster the fragments into groups

Example of Randomly Selected Protein



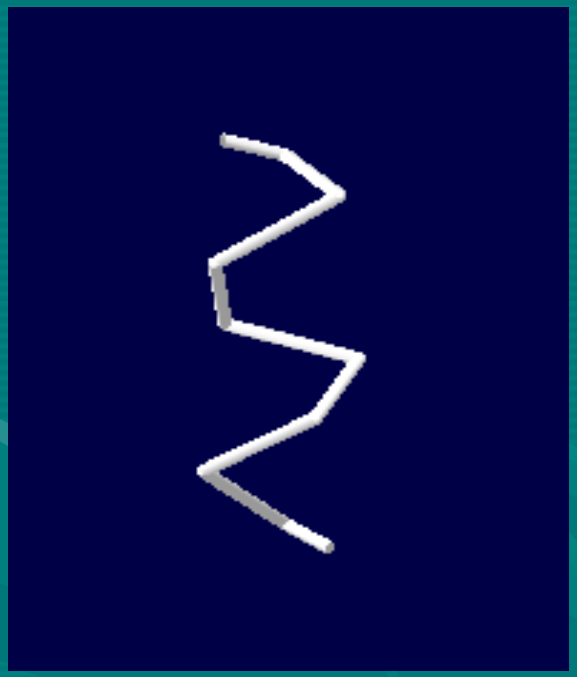
1mka

Common Folding Units Discovered by Data Mining

Randomly Selected Proteins

1ash, 1bsr, 1cca, 1cew, 1clm, 1crn, 1cct, 1erb, 1fut, 1hng, 1hoe, 1lbu, 1mka, 1mng, 1pkp, 1udi, 1utg, 1yal, 2vab, 5pti

3698 fragments



From 1mka
α helix

Group 1 514 fragments

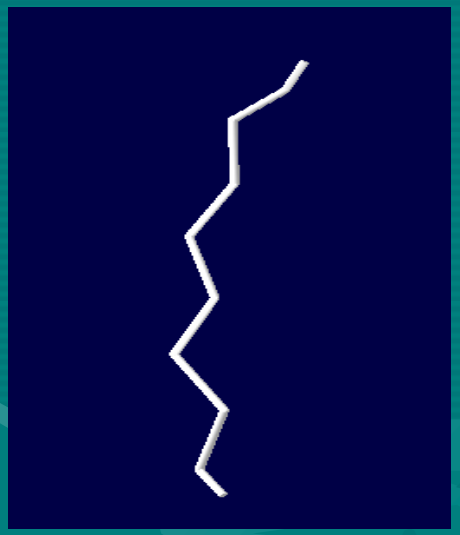
Amino Acid	phi	psi
GLN	-60.078	-41.741
LEU	-69.310	-35.875
VAL	-65.116	-46.320
GLY	-67.025	-36.399
PHE	-62.244	-39.936
TYR	-66.128	-38.417
LEU	-64.114	-37.476
GLY	-70.167	-32.912

Common Folding Units Discovered by Data Mining

Randomly Selected Proteins

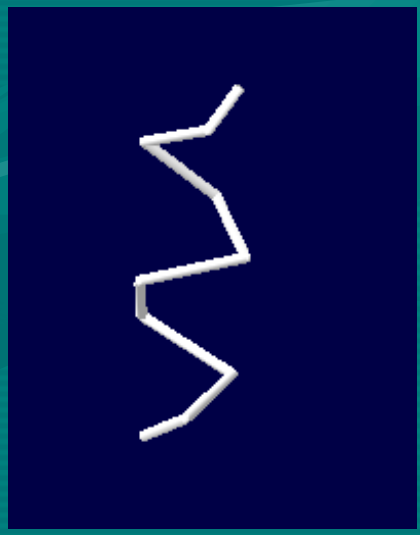
1ash, 1bsr, 1cca, 1cew, 1clm, 1crn, 1cct, 1erb, 1fut, 1hng, 1hoe, 1lbu, 1mka, 1mng, 1pkp, 1udi, 1utg, 1yal, 2vab, 5pti

3698 fragments



Group 2
188 fragments
From 1erb
β pleated sheet

Group 3
79 fragments
From 1bsr

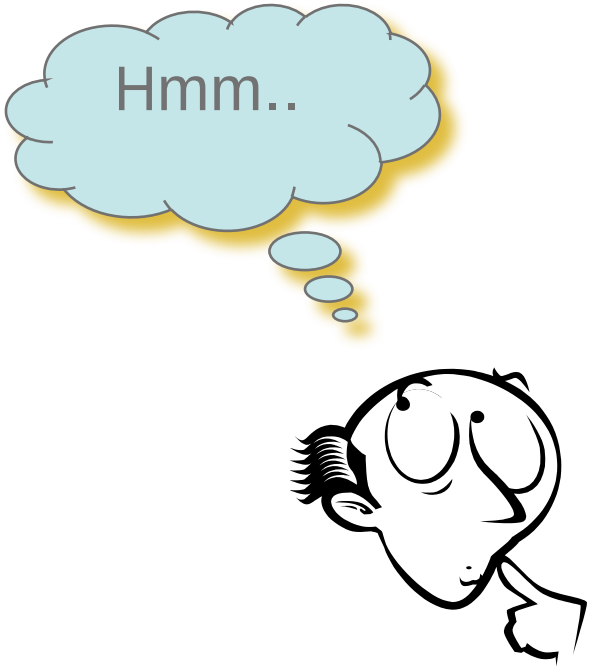


Group 4
61 fragments
From 1lbu

Milestones and Future Work

- Oct 2007- Jan 2008
  - Designed new data mining algorithm
- Jan 2008- Aug 2008
  - Implemented new algorithm for large data sets
  - Tested algorithm on Protein Data Bank
  - Verified that algorithm finds features of interest (common protein folding units)
  - This data mining tool runs fast and handles large data sets
- Future Work
  - Apply this software tool to the data used by the science drivers (Bishop, Blake, Soper, Cortez)

Thank You!



WP3:
Visualization

Faculty and staff:

LSU CCT+CS: B. Ullmer, W. Benger, A. Hutanu, J. Ge

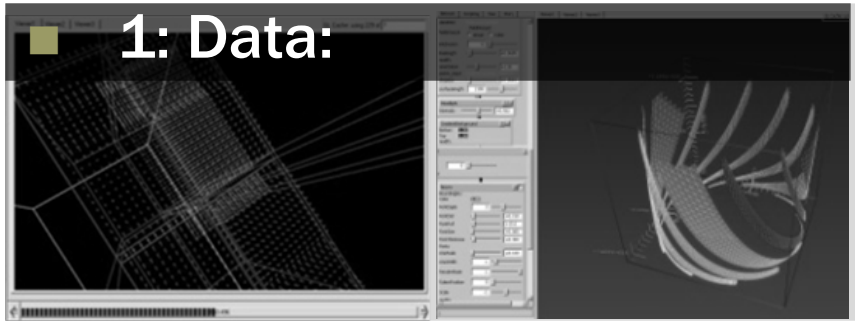
ULL/LITE: **C. Cruz-Neira,** R. Jindal, M. Miller

LSU CS: S.S. Iyengar, N. Brener, B. Karki

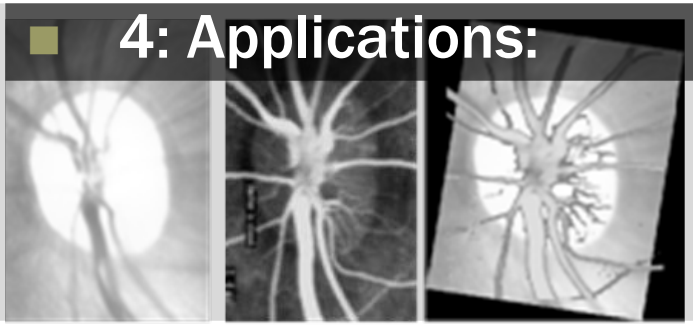
Southern: A. Jana

Review of project components

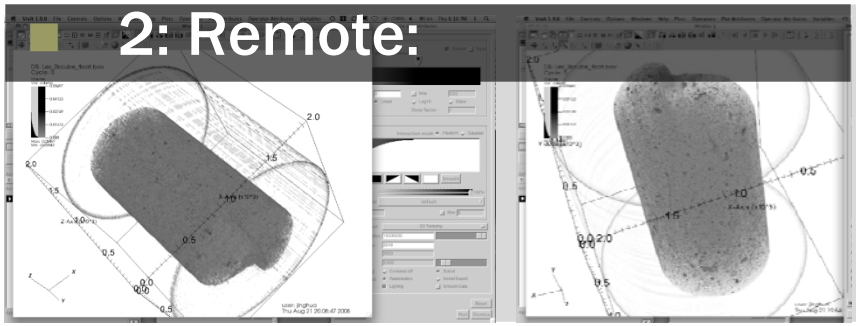
- 1: Data:



- 4: Applications:

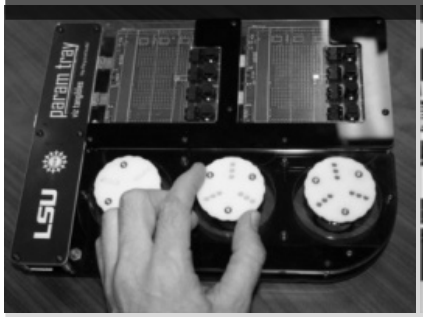


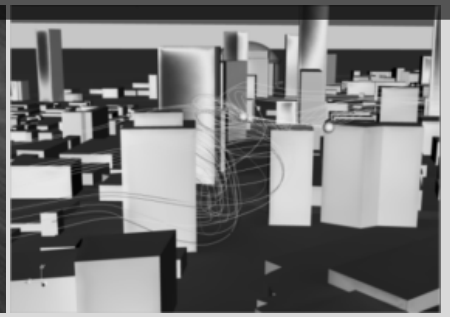
- 2: Remote:

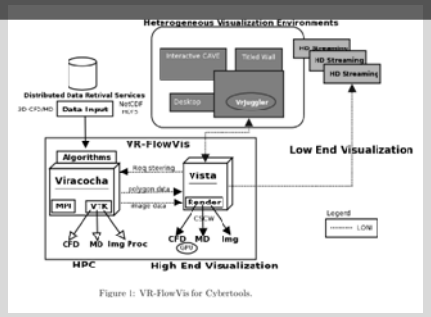


- Outreach:

- 3: Advanced:

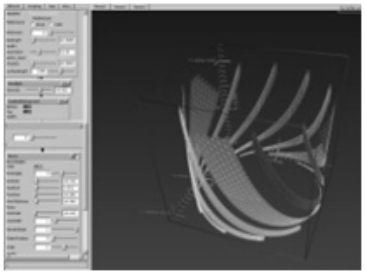






Review of project components

- **3.1: Viz/data integration:** *Benger, Ritter, Jiao,* Shetty
- **3.2: Remote streaming:** Hutanu, Ge, Amatya
- **3.3: Advanced viz environ.:** Cruz-Neira, Ullmer, Shetty, Natesan
- **3.4: Applications:** Iyengar, Brener, Karki, *Benger*
- **3.5: Outreach (LIGO, Southern):** Ullmer, Jana, Toole



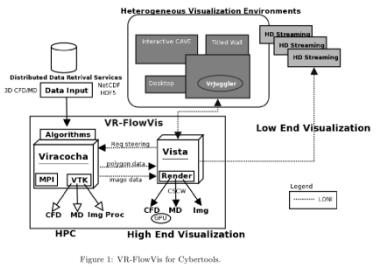
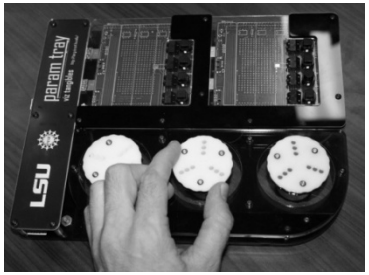
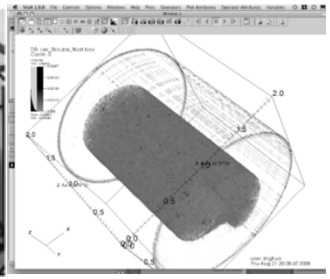
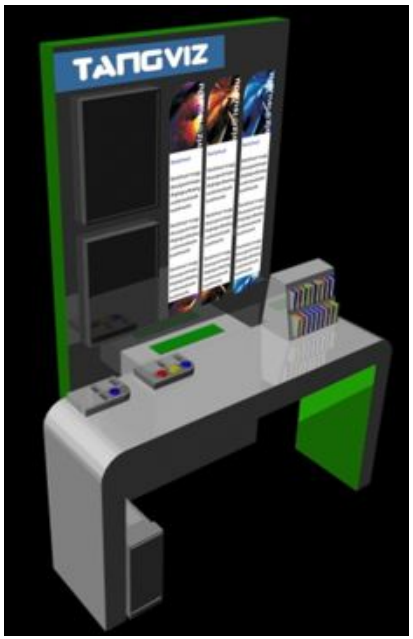


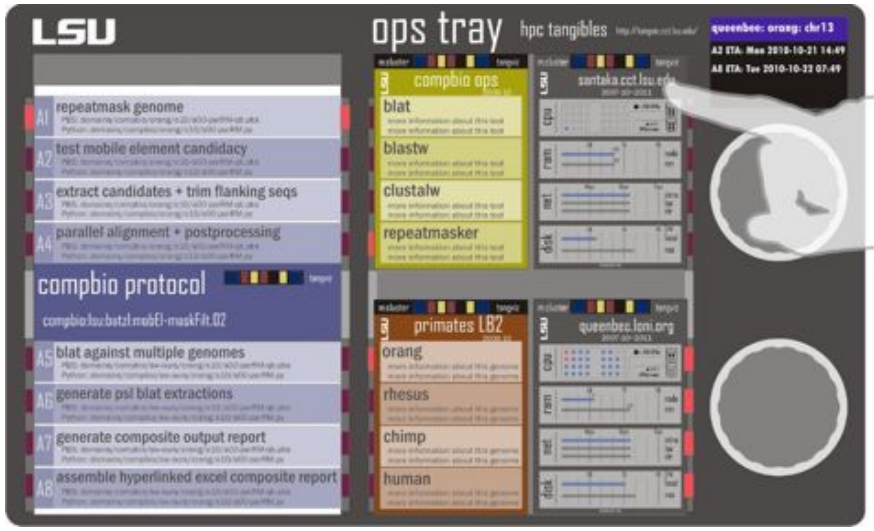
Figure 1: VR-FlowVis for Cybertools.



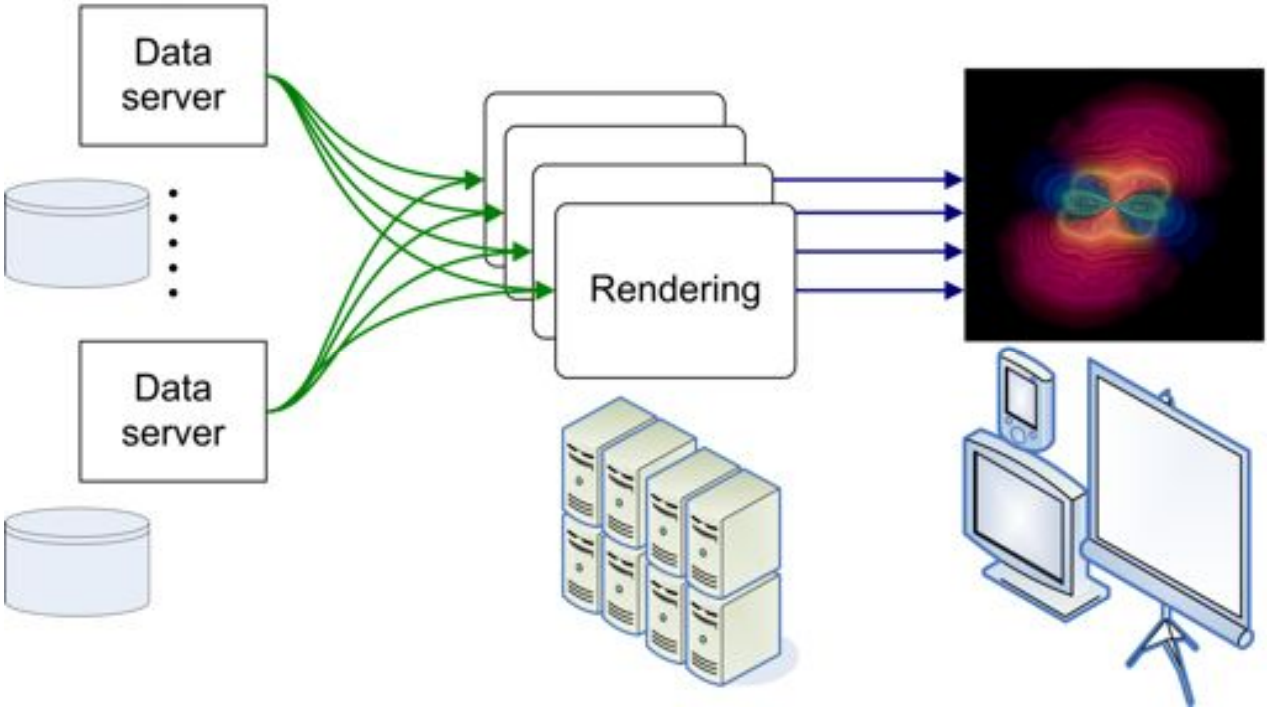


Viz tangibles + LIGO outreach update





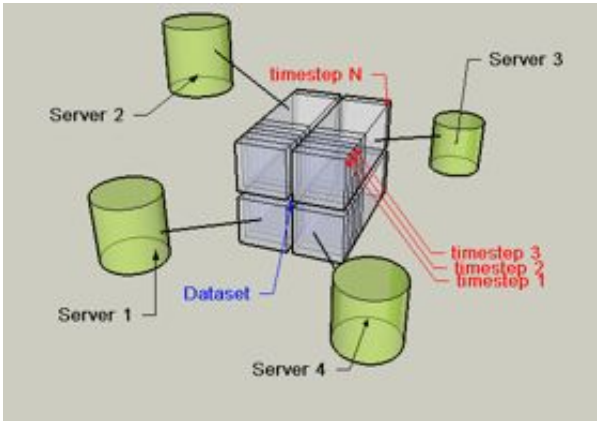
Example: remote visualization

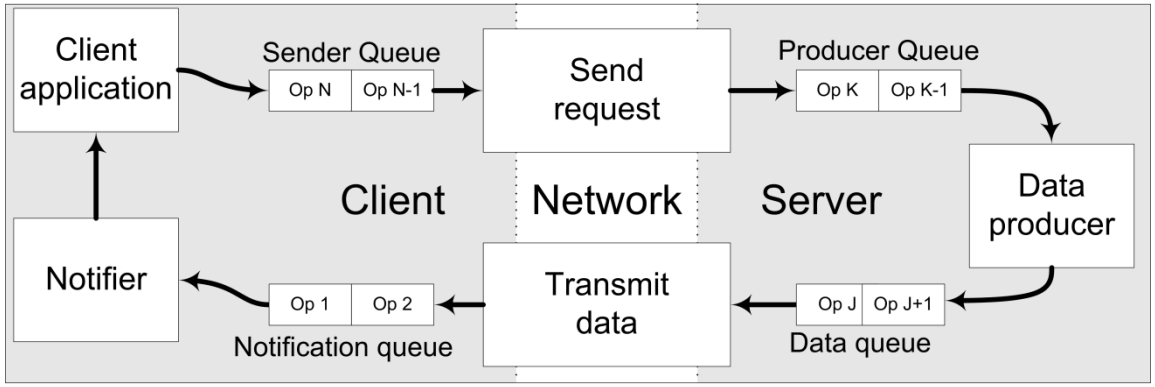


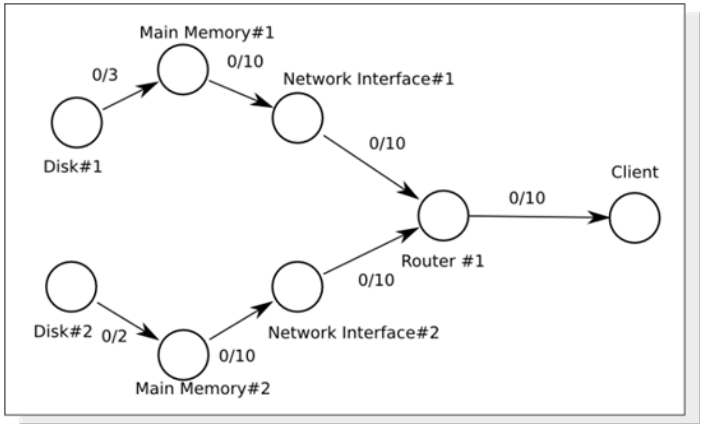
- Goal: Optimization of visualization of large data through parallelism and intelligent resource selection
- CyberTools Components: PetaShare, Visit/Equalizer/Vish, VRFlowViz, SAGE/UltraGrid, Science Drivers data

Data

- Use distributed data servers
- Designed an algorithm to use information about network topology and link capacity to optimize throughput
- Flexible, pipelined high-performance data transfer system





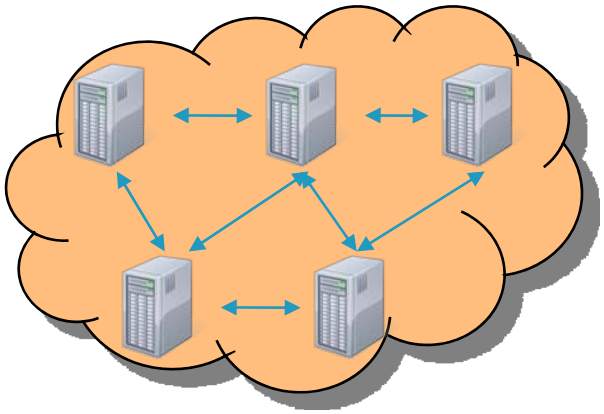


Data

- Status:
  - Data transfer system implemented
  - Optimization algorithm for prefetching and deterministic network links designed
- Next steps:
  - Tuning the data transfer system and integrate in visualization application (currently using Petashare)
  - Implement optimization algorithm and integrate in data transfer system
  - Benchmark suitable transport protocols

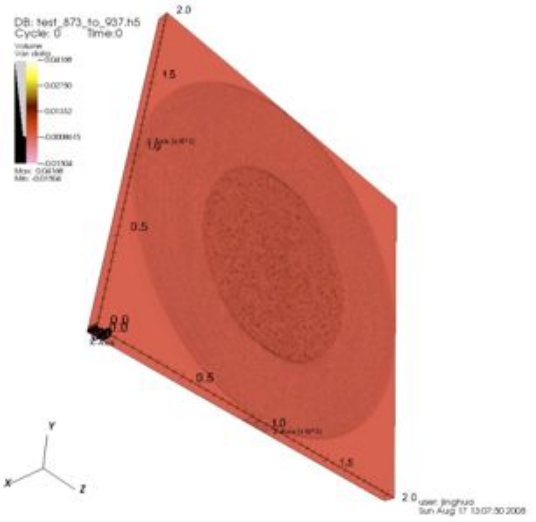
Rendering

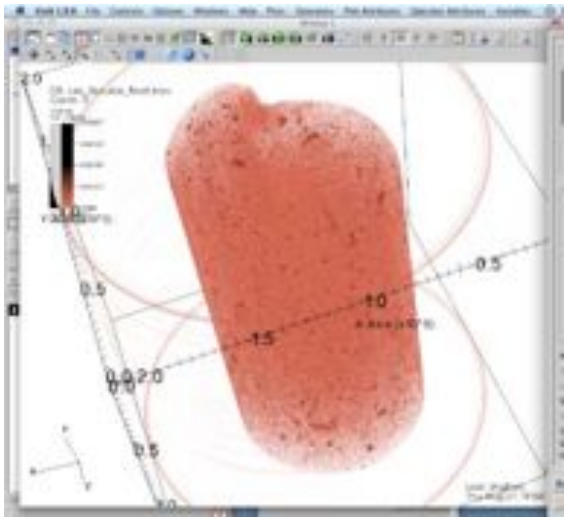
- Use HPC and visualization clusters to render large datasets
- Choose rendering options (data distribution, image distribution or hybrid) and configuration



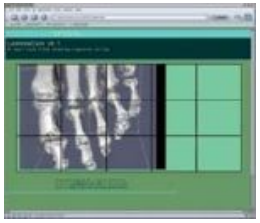
Rendering

- Currently investigating two visualization systems: visit and Equalizer, testing with an example dataset
- Visit: complete visualization system, does not yet support hardware – accelerated parallel rendering
- Equalizer: flexible system designed for parallel rendering
- Future steps: benchmark and compare various rendering options, long-term: automatic tuning of configuration parameters





Streaming



- Transport resulting images to user
- Options: hardware-assisted, software (integrated in the application or external)
- Current status:
  - successfully used hardware-assisted system based on HD videoconferencing set-up used for HPC classes. Advantage: can be used with any visualization application, Disadvantage: poor scalability
  - Evaluating software-only rendering options (SAGE, visit & VISH built-in streaming)
- Future work: automatic tuning of video streaming parameters

ULTRAGRID

SAGE

Demos

- Data: synchrotron x-ray tomography of flame retardant in polystyrene solution. 32Gb (2048x2048x2048) for single dataset
  - Simple image data set for development, will move to using more viz from CyberTools
- Demo 1: parallel software rendering (raycasting) with ViSit on HPC cluster
- Demo 2: parallel hardware-accelerated rendering (texture mapping) with equalizer on two high-end visualization workstations

IMAGE FUSION

Faculty: Dr. S. Sitharama Iyengar (LSU)
Dr. Nathan E. Brener (LSU)
Dr. Bijaya B. Karki (LSU)
Dr. Hilary Thompson (LSUHSC)

Project Coordinator: Dr. Dimple Juneja

Graduate Students: Dr. Hua Cao
Rathika Natarajan
Harsha Bhagawaty
Asim Shrestha
Jagadish Kumar
Gaurav Khanduja
Dipesh Bhattarai

Integration with Collaborators: Dr. Acharya
LSU Health Sciences Center (LSUHSC)

Visualization: Image Fusion

- Combining relevant information from two or more images with different modalities into a single image. Current applications in biomedical computing, remote sensing.
- Important tool for dynamic data driven computing scenarios for automated data extraction.
- Test problem: Branching arterial images (Thompson)
  - Content change and non-uniform distributed intensities of the involved images
  - Automate to support end-to-end workflows

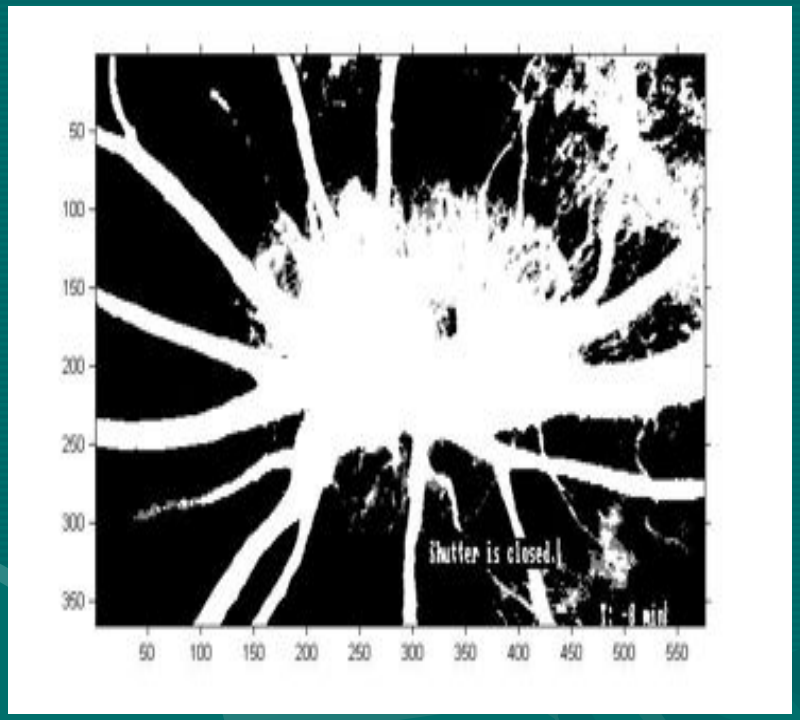
Image Registration: Adaptive Exploratory Algorithm

New algorithm to identify control points for image registration

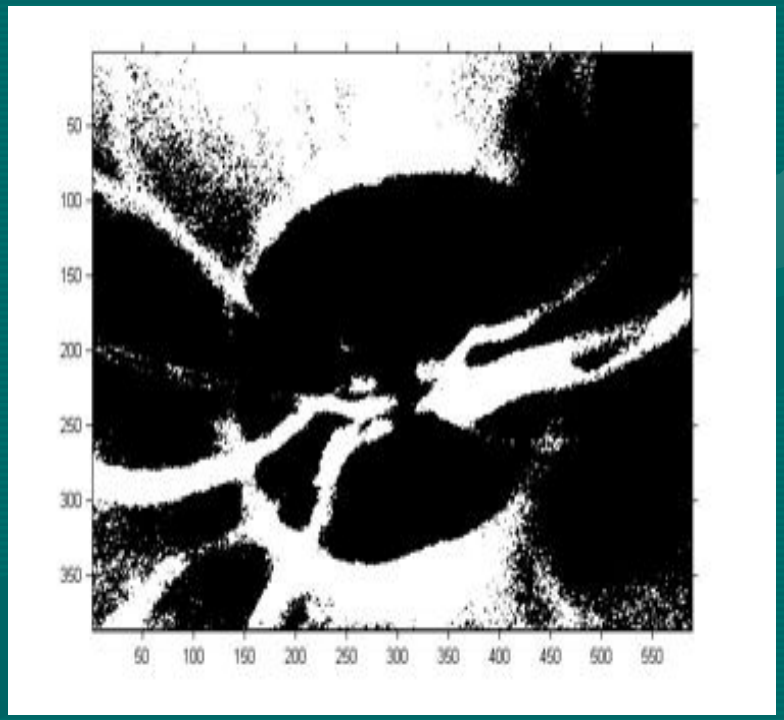
Image 1

Image 2

Image Registration: Mutual Pixel Count Algorithm



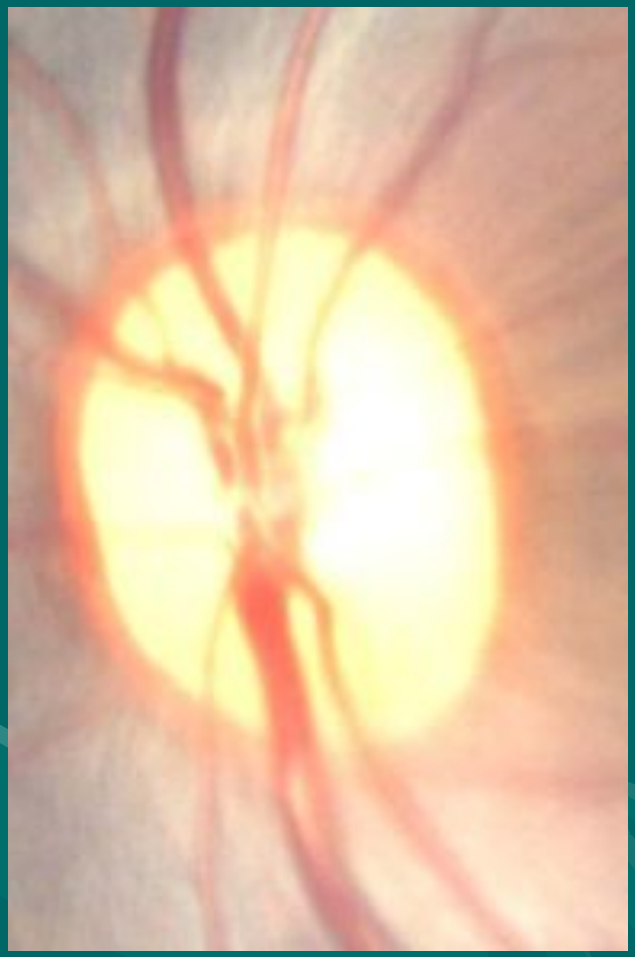
BW image of Image 1



BW image of Image 2

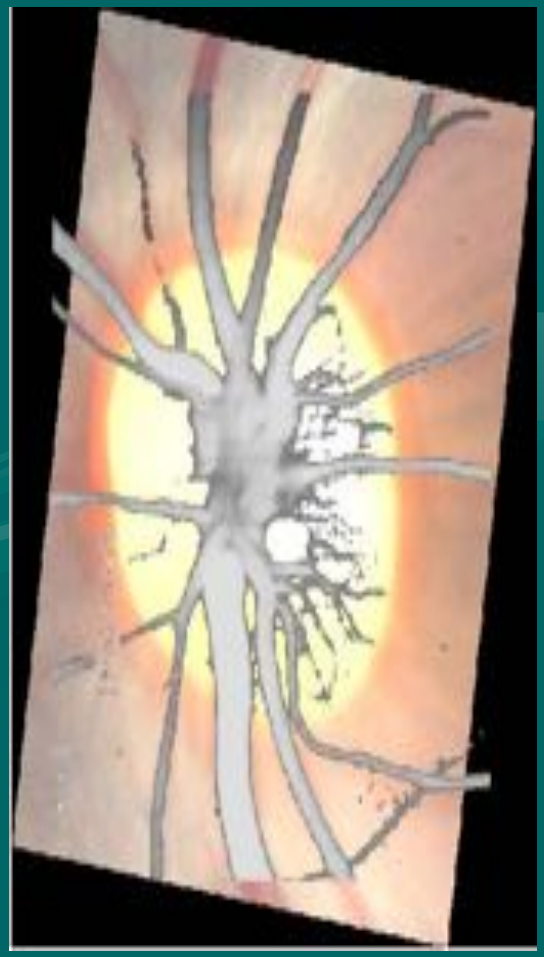
New algorithm iteratively varies control points to improve accuracy of registration.

Fusion of Images



A

B



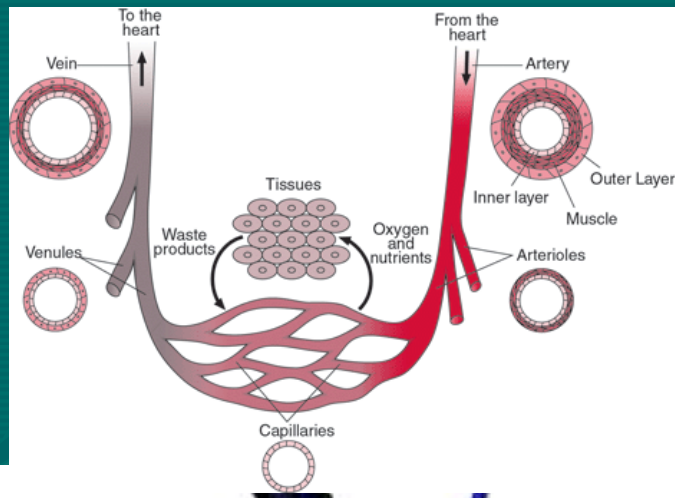
C

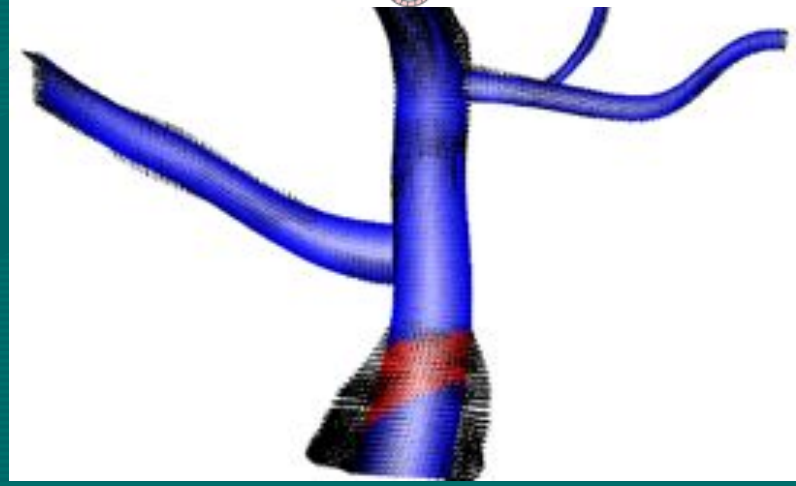
C is the composite (fused) image of A and B.

Application of Image Fusion Technique to Biotransport

- Acharya and colleagues are researching transport processes to model blood flow through arteries.
- Starting point is accurate mesh for artery structure.
- Our data fusion algorithms will generate fused images with more detailed geometric information than individual images leading to more accurate meshes.

Arterial Vessels





Publications

1. K. Manikandan, Debnath Pal, S. Ramakumar, Nathan E. Brener, S. Sitharama Iyengar and Guna Seetharaman, “Functionally Important Segments in Proteins Dissected Using Gene Ontology and Geometric Clustering of Peptide Fragments”, Genome Biology, Vol. 9, Issue 3, article R52 (2008).

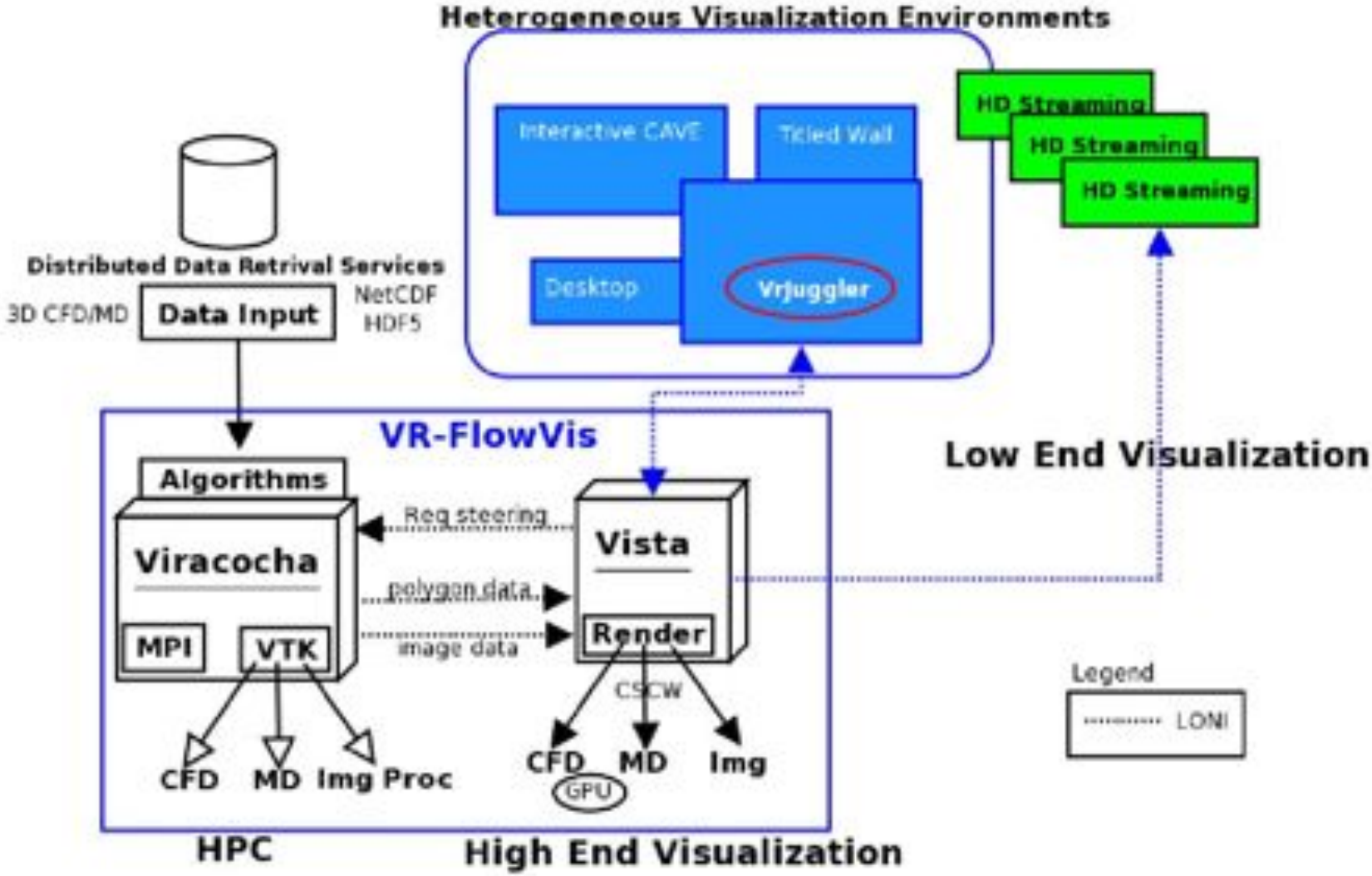
2. Hua Cao, Nathan Brener, S.S. Iyengar, "High performance Adaptive Fidelity Algorithms for Multi-Modality Image Fusion" , Submitted to IEEE Transactions on Computers.

Milestones

- Oct 2007- Jan 2008
  - Designed new data fusion and data mining algorithms
- Jan 2008- Aug 2008
  - Implemented algorithms on large scale data sets
  - Developed cyber tools for data fusion and data mining applications
  - Published papers and one PhD graduated and two MS students (Oak Ridge National Lab)
- Future Work
  - We will be providing software tools to be used by the science drivers

VRFlowVis

- Nikhil Shetty & Vignesh Nateshan



Education: Year 1

Deliverables in K-12 & Undergraduate training:

- **High School Apprenticeships:**

LaTech → Science project on glucose sensor
Tulane HSC → Preparation of apo glucose oxidase

- **Design academic year projects on topics of the grant:**

We will begin implementing it in 2008-09. A meeting is scheduled to discuss possible projects, venues for students to carry them out, supervisors, etc. One Tulane student is already on board.

- **Create summer research opportunities targeting primarily minorities:**

A 5-week program at Tulane in summer 2008 took place in June-July. There were 7 students from Tulane, Dillard, Xavier. The projects included microorganism swimming and disease transmission modeling.

2008 Summer research program in Computational Science at Tulane University

- **Modeling epidemics and disease transmission of the West Nile Virus using continuous and discrete models with space**

  Justin Walbeck & Timothy Clinton (Tulane), Caira Dyer (Dillard)

- **Numerical models of jellyfish motion**

  Namdi Brandon (Tulane) and Barry Jackson (Xavier)

- **Modeling Microorganism Locomotion with Stokes Flow**

  Austin Griffith & Maren Leopold (Tulane)

- **Mathematical models of disease transmission**

  Cavin Ward-Caviness (Tulane)
  senior project

Education: Year 1

Deliverables in Graduate training:

•**Summer Internships or extended visits to other institutions:**

This included sending Tulane students to IfM or CCT for extended visits. It also included sending students to other institutions for the summer.

•Emir Bahsi (LSU Graduate Student) visited Tulane (May 2008)

•Jerina Pillert & Kate Hamlington (Tulane Students) had virtual meetings with LSU CCT & IfM (July 2008)

•Senaka Kanakamedala (fM) visit to Tulane planned (September 2008)

•**Multi-institutional dissertation committees:**

Cortez is on Hamlington's committee (Gaver, BME, Tulane)
Bishop is on Henry's committee (D. Blake, Biochem, Tulane)
DeCoster is on Kanakamedala's committee (Lvov, Chem, LaTech)

Education: Year 1

Deliverables in Postdoctoral training:

- **Cross Institutional mentoring and training:**

Mehnaaz Ali (Tulane), Mangilal Agarwal (LaTech IfM), Yuen Yick Kwan from Purdue will join Tulane in August 2008

Mehnaaz Ali: A Biochemistry postdoc visited IfM to learn about the facility and microfabrication techniques.

Mangilal Agarwal: An Electrical Engineering postdoc visited Tulane to learn about biochemistry and molecular biology.

Outreach: Year 1

•**Publications:** Including joint authorship across institutions and participants, general audience articles

•**Scientific and nonscientific conference presentations:**

- BMES (Oct 2008) – Kate Hamlington
- ACS (August 2008) – Diane Blake (as we speak!)
- ACS (August 2008) – Mangilal Agarwal (last Tuesday)

EPSCoR RII Education & Outreach: Evaluation of Implementation

Mary Jo McGee-Brown, External Evaluator
Linda Ramsey, Internal Evaluator

EPSCoR RII Evaluation Data Collection Methods

- Participant Observation at Meetings
- Program Documents analysis
- Informal Interviews
- Data & data sets received through e-mail
- Program Products analysis (newsletters, brochure, web sites, etc.)
- End-of-year Surveys (37.7% response rate from research scientists from 7 institutions; 26.5% response rate from grad students & post docs from 5 institutions)

EPSCoR RII Collaboration

- **Interdisciplinary collaboration** – characterized by the majority (85%) of Research Scientist survey respondents and Graduate student and Postdoctoral student survey respondents (100%) as effective, and an essential part of their ongoing work
- **Interdisciplinary Mentoring** – Half (50%) of Research Scientist survey respondents characterized themselves as interdisciplinary mentors for undergraduates, graduate students and/or postdoctoral fellows

EPSCoR RII Collaboration

- **Inter-institution collaboration** – the majority (70%) of Research Scientist survey respondents indicated inter-institution collaboration has been implemented, and remaining respondents indicate that inter-institution collaboration has not begun in their group or partner institution

Expectations for Year Two EPSCoR RII Collaboration

- Most expect increased collaboration during Year 2
- Some groups are examining effective models of face-to-face inter-institutional collaboration they have used in order to expand implementation
- The IOCOM communication network with desktop capabilities is expected to enhance inter-institutional and WP/Science Driver communication
- Some WPs have indicated the need for increased communication and collaboration with Science Drivers as products become ready for testing

Conclusions: Reaching Project Objectives (K-12, higher education and the general public)

- Project leaders and Research Scientists have implemented or are developing multiple effective programs and have distributed multiple informative products to educate and engage all targeted groups (K-12 students and faculty, undergraduate and graduate students, general public) through outreach activities that reflect EPSCoR RII participant expertise and scientific and engineering achievements.

Conclusions: Support for Tenure-Track Junior Faculty

- Multiple grant programs have been developed and effectively implemented to support a large number (77) of non-tenured and tenure-track junior faculty in seeking advice from federal granting agencies for writing successful research grant proposals; presenting invited talks at national or international conferences; collaborating with scientists at national labs and industries; and exploration of novel research.

Conclusions: Engagement of Minorities/Underrepresented Groups

- Some individual Research Scientists (45% of survey respondents) and groups, a few outreach programs, and a few grant programs have addressed the objective of engaging minority and underrepresented groups in STEM activities, research and graduate programs, and others are in the planning stages for Year Two.

Conclusions: Collaboration

- Interdisciplinary collaboration and mentoring has been effectively implemented in most groups and partner institutions, and program participants characterize it as an essential component of their ongoing work
- Inter-institution collaboration during Year One has been effectively implemented through some programs across some partner institutions and not implemented in others

Overall Project Effectiveness in Education and Outreach

- Data from multiple sources indicate that leaders and participants in the systemic EPSCoR RII Project have effectively planned, developed and/or implemented most proposed programs for all Education and Outreach objectives during Year One.

Emerging EPSCoR RII STEM Education & Outreach Models

- Regular Education & Outreach meetings
- Tri-state (AL-MS-LA) EPSCoR RII Conference
- Multi-focused grants for Junior Faculty
- LSU graduate student Professional Development Seminar Series
- Science & Engineering Research Day for Graduate & Undergraduate Students

Emerging EPSCoR RII STEM Education & Outreach Models

- Tulane Undergraduate Research Program - recruitment from minority/underrepresented groups
- Collaboration models involving graduate students, post docs, &/or Research Associates
- Expertise & achievement dissemination & networking (cross-institution and cross-state collaboration, internal grants, SoS, publications, project products, etc.)

Model Development for Internal Dissemination & Exportability

- Generate descriptive program information and participant awareness strategies
- Identify proposed impacts across participant levels
- Identify appropriate interdisciplinary and inter-institution participation
- Generate & analyze program impact data
- Identify what was effective & why; what was not effective & why not
- Generate a clear and complete model description
- Plan dissemination and exportability strategies

Enabling Distributed Applications with SAGA

João Abecasis, Shantenu Jha, Hartmut Kaiser, Joohyun Kim, André Merzky, and Ole Weidner

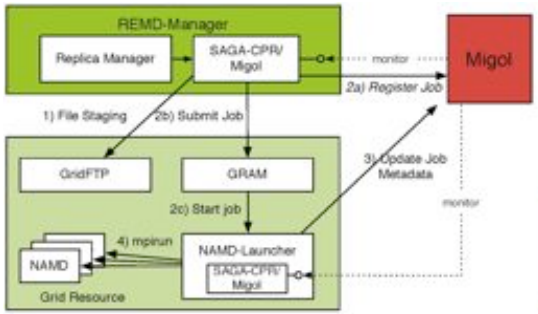
Center for Computation & Technology, Louisiana State University, Baton Rouge, U.S.A.

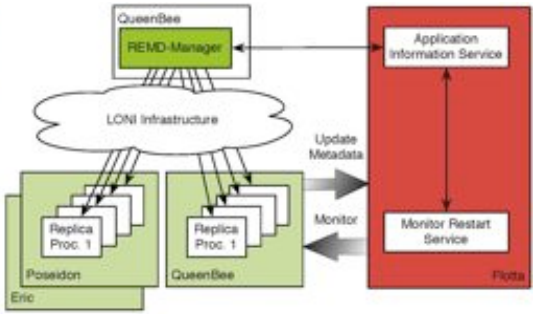
Abstract

The Simple API for Grid Applications (SAGA), a proposed recommendation of the Open Grid Forum (OGF), defines a high-level programmatic interface for developers of Distributed Applications [1]. The fundamental idea of SAGA is to lower the barrier for applications and application scientists to utilize distributed infrastructure. SAGA provides a simple, uniform, stable interface to the most often required functionality in order to construct general purpose, extensible and scalable applications.

Our group has lead the SAGA effort, starting from the specification effort at the OGF to providing the first C++ implementation [2]. We are also developing several different novel applications, using SAGA to harness the power of distributed infrastructure.

SAGA has already been used to develop different types of distributed applications. Namely, (i) converting legacy applications to utilize distributed resources; (ii) development of applications based upon abstractions and frameworks that are themselves developed using SAGA; (iii) first principles applications, explicitly cognizant of the fact that they will operate in a distributed environment, where the application logic is coupled with the distributed logic. SAGA supports the development of these applications and many others, thus providing a tool to develop a broad and general class of applications.

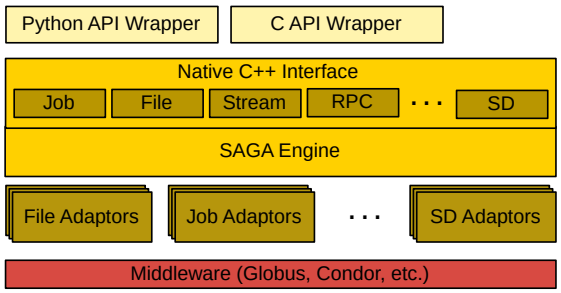


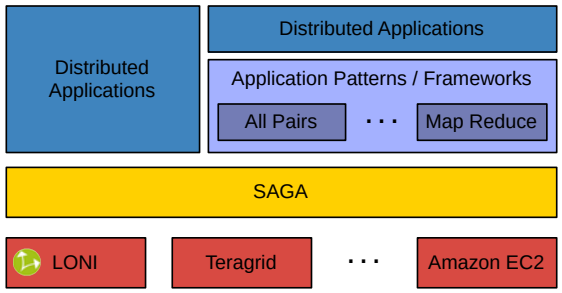


Simple, Powerful Abstraction Layer

SAGA facilitates the use of distributed infrastructure by providing a simple interface across different middleware distributions and environments. Therefore once an application has been written using SAGA it can be deployed and run on any environment in which SAGA is supported.

We are developing adaptors for the most commonly occurring distributed environments. Additionally SAGA provides the abstractions from which commonly occurring execution patterns and usage modes can be supported. For example for data-intensive applications, we create a framework that supports the common MapReduce pattern. Applications involving basic functionality such as searching, can then be deployed over distributed environments





Connections with CyberTools

SAGA is being used within the Cybertools project in several critical ways:

- It is being used to create a general purpose "Application Manager", that will enable many science drivers to utilize remote LONI machines without any changes to the execution environment. In particular it can be used to support specific application usage patterns, for example, it has been used for distributed replica-exchange (RE) simulations using NAMD. The same infrastructure can be used for use with other codes such as LAMMPS, etc. The figure above provides details on how SAGA is used to implement RE.
- SAGA will be the interfaced with Cactus applications to use Information Services and other advanced CyberInfrastructure features.
- SAGA will also provide the basic capability for interfacing multi-physics applications (via extension to the API to support messaging)

References

1. Goodale, T, Jha, S, Kaiser, H, Kielmann, T, Kleijer, P, Merzky, A, Shalf, J, Smith, C, (2007) GFD-R-P.90 A Simple API for Grid Applications (SAGA), Open Grid Forum
2. SAGA C++ Project [Online]. http://saga.cct.lsu.edu

Acknowledgements: This work was supported by NSF, the Louisiana Board-of-Regents and CCT funds.

Distributed Visualization for Cybertools Project Applications - EAVIV

Vinay C. Amatya[1,2,3], Andrei Hutanu[1,2,3], Jinghua Ge[1,3], Shalini Venketaraman[1,3], Cornelius Toole[1,2,3], Gabrielle Allen[1,2,3]

1Louisiana State University
2Department of Computer Science , LSU
3Center For Computation Technology, LSU

Abstract

Visual verification of theories and data from experiments and simulations follow a chain of processes. From formatting data in suitable format to streaming data to the points of interpretation like rendering or analysis; to rendering the data in the rendering farms or in a local cluster; retrieving the rendered data and convert them to the image pixel and streaming those pixels live to the desired destinations local or remote. Our aim is to investigate each of these steps and offer a better tool, algorithm or mechanism to smoothen and optimize each of the aforementioned steps in the visualization pipeline. We have discretely looked into each of these steps with a certain degree of success and further discuss the idea to realize the project and share our experience.

Goal: *Optimization of visualization of large data through parallelism and intelligent resource selection*

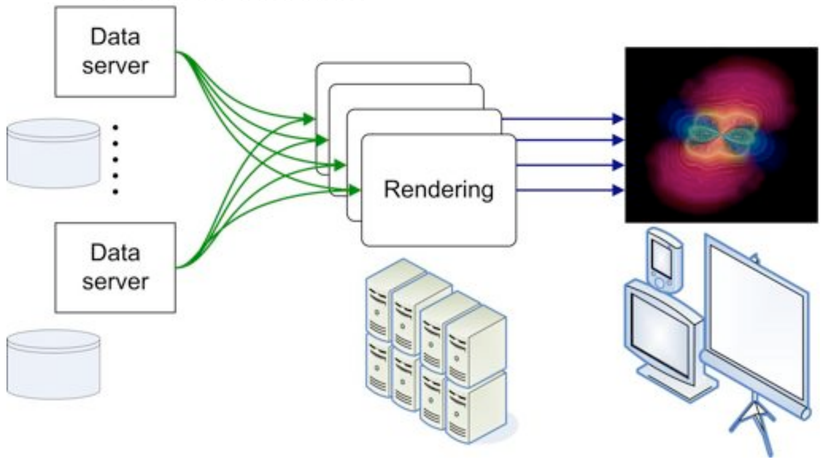
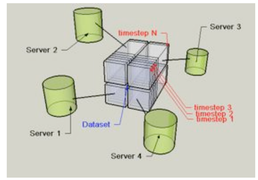


Fig. 1 Distributed Visualization Pipeline

Data

Use distributed data servers
Designed an algorithm to use information about network topology and link capacity to optimize throughput
Flexible, pipelined high-performance data transfer system



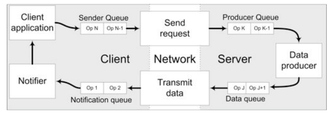


Fig.2 & 3 Overview of distributed data server implementation and actual data transfer process

Status:
Data transfer system implemented
Optimization algorithm for prefetching and deterministic network links designed
Next steps:
Tuning the data transfer system and integrate in visualization application (currently using Petashare)
Implement optimization algorithm and integrate in data transfer system

Rendering

Use HPC and visualization clusters to render large datasets
Choose rendering options (data distribution, image distribution or hybrid) and configuration

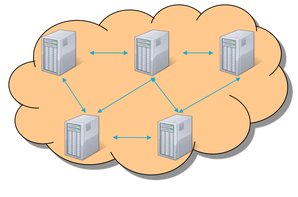


Fig.4 Distributed Rendering

Currently investigating two visualization systems: Visit and Equalizer, testing with an example dataset

Equalizer

Flexible system designed for parallel rendering
Supports distributed rendering and frame compositing, Multiple Decomposition – Recomposition Modes

Status:
Equalizer has been tested on scientific experimental data

Future steps: benchmark and compare various rendering options, long-term: automatic tuning of configuration parameters; test on display walls, cave displays, scalable rendering

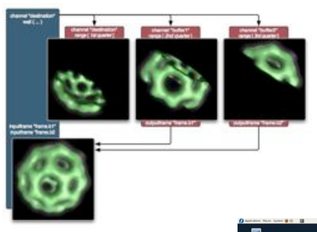


Fig. 7 An example of compound specification for sort-last rendering using 3 channels

Fig.9 Distributed Visualization

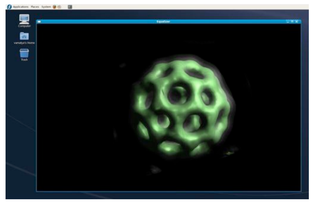


Fig.8 Snapshot of display for Equalizer running on two separate machines graphics cards

Visit

Complete visualization system, does not support hardware – accelerated parallel rendering
Free, Open Source, Platform Independent, distributed, parallel
Distributed architecture allows to take advantage of both compute power of large parallel computer and local graphics hardware
Rendering on remote parallel machine, while display on local machine

Status:
Visit has been tested on scientific experimental data

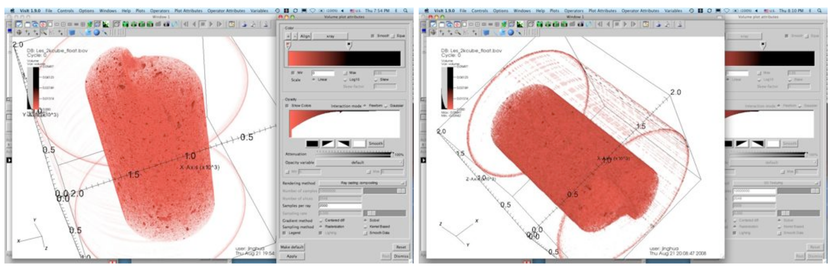


Fig.5 &6 Visit Snapshotson of Scientific Experimental Data

Streaming

Transport resulting images to user
Options: hardware-assisted, software (integrated in the application or external)
Current status:
Successfully used hardware-assisted system based on HD videoconferencing set-up used for HPC classes.
Advantage: can be used with any visualization application,
Disadvantage: poor scalability,
Evaluating software-only rendering options (SAGE, visit built-in streaming)
Advantage: scalable, but needs high bandwidth
Future work: automatic tuning of video streaming parameters

Connections with CyberTools

Workpackage-3: Visualization Services

Acknowledgements

We would, sincerely, extend our gratitude to all the IT persons in LSU for their active assistance. We would like to heartily thank the facility and manpower lent to our cause by LSU and CCT. Last but not the least we offer deepest thanks to the NSF for funding the project.

Stress and Deformation Behavior during Demolidng in Nanoimprint Lithography

Zhichao Song[1], Lance Brumfield[1], Junseo Choi[1], Alborz Amirsadeghi[1], Jaejong Lee[2], and Sunggook Park[1]

[1]Mechanical Engineering Department and Center for BioModular Multiscale System, Louisiana State University, Baton Rouge, LA 70803, USA

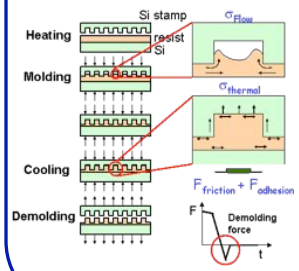
[2]Korea Institute of Machinery & Materials, Yuseong, Daejeon, 305-600 Korea

Abstract

Most of structural failures in nanoimprint lithography occur during demolding, a process to separate the stamp from the molded substrate. In this work, we studied stress and deformation behavior for the molded polymer layer using numerical simulation. Via parametric studies, a general rule to improve the demolding process has been proposed.

Introduction

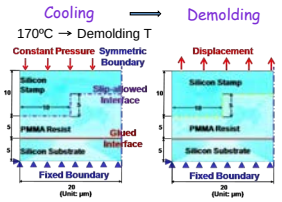
Demolding in nanoimprint lithography



- NIL has potential as a production-type tool to fabricate micro- and nanostructures in polymer via molding.
- Demolding is a process to overcome all the chemical and mechanical interactions between stamp and substrate.
  → Most of imprint failures occur at this process step.
- A systematic study on demolding is needed to develop processes leading to low stress and deformation in the molded substrate.

Simulation Method

2-D model for demolding simulation



- Assumptions
  - Governed by continuum mechanics.
  - PMMA is initially filled into stamp.
  - No initial stress.
  - Sliding is allowed, but not separation between stamp and PMMA.
  - All materials are isotropic
- Simulation was performed using ANSYS 10.0.

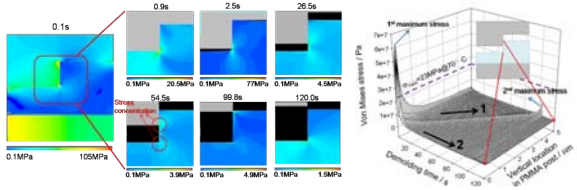
Materials

- PMMA: viscoelastic → 10 element Maxwell model was used.
- Si: linear elastic → E = 128GPa, ν=0.28, α=2.5×10^{-6} /℃

Simulation Results

1. Single symmetric structure

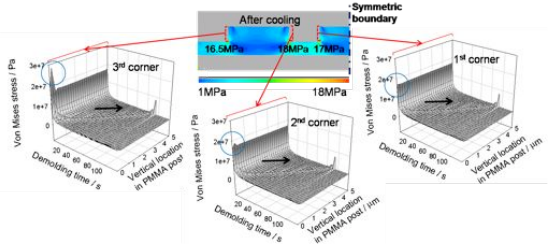
Stress evolution during demolding



- Local stress evolution during demolding shows two maximums: at the beginning and end of demolding.
  → Demolding failure can also occur at the end of demolding.

2. Multiple symmetric structure

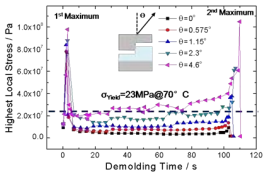
Stress evolution at different side walls



- Higher stress is shown at the outmost structures.
  → An auxiliary outer structure could protect the inner structures.

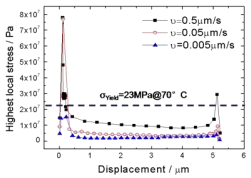
3. Parametric studies

σ_{max} vs. demolding angle



- To have accurate alignment in demolding direction is critical.!

σ_{max} vs. demolding rate

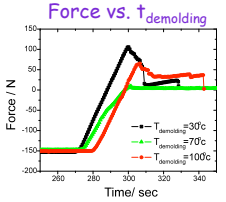


- High demolding rate leads to high local stress.

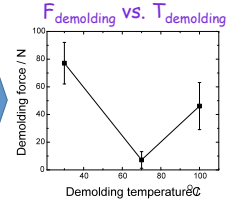
Experimental Verification (Effect on demolding temperature)

1. Demolding force measurements

Force vs. $t_{demolding}$



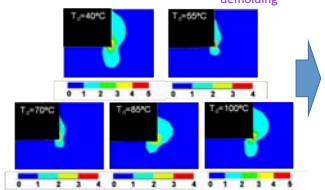
$F_{demolding}$ vs. $T_{demolding}$



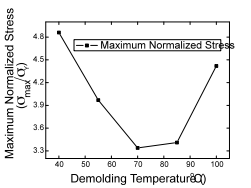
- A minimum in $F_{demolding}$ vs. $T_{demolding}$ curve → 70℃

2. Demolding simulation

Normalized σ vs. $T_{demolding}$



Normalized σ_{max} vs. $T_{demolding}$



- Stress is normalized by σ_{Yield} at each $T_{demolding}$.
- A minimum in normalized σ_{maxd} vs. $T_{demolding}$ curve → 70℃

Connections with CyberTools

- The ability of the FEM simulation for complicated yet actual structures will enable prediction of the demolding process as well as determination of a range of process parameters for successful demolding even at the stage of a process design in an economical and reliable way.
- For more accurate simulation, it is also necessary to incorporate nanoscale phenomena such as non-Furrier type heat conduction and nanoscale friction.

→ This requires more powerful computational tools, for which supports from CyberTools are critical.

Acknowledgements

This research was supported by the Center for Nanoscale Mechatronics & Manufacturing (CNMM), one of the 21st Century Frontier Research Programs from the Ministry of Science and Technology, KOREA (Grant No. M102KN0-1000706K1401-00710), and by the Louisiana Board of Regents – RCS (Contract No.LEQSF (2006-09) – RD – A – 09) and NSF-EPSCoR RII.

Workflow Enabling Large Scale Scientific Applications via Pegasus

Emir M Bahsi, Tevfik Kosar

Center for Computation and Technology, Louisiana State University

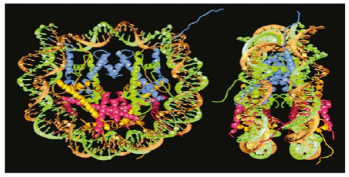
Abstract

Our first goal is end-to-end automation of two large scale applications: DNA folding and reservoir uncertainty analysis. Our implementation is based on Pegasus workflow tool that uses Condor, Condor-G, DAGMan, and Stork.

Our second goal is to implement a site selector that aims to achieve intelligent resource selection and load balancing among different grid Resources.

DNA Folding

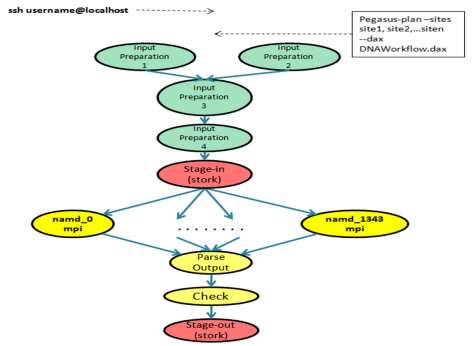
In order to identify how DNA sequence proteins rotate the global structure and dynamics of chromatin, Dr. Bishop and his research team developed suite of scripts and following is the way how scripts run by a scientist manually:



Folded DNA Structure

- **Input preparation 1:** Downloading pdb files.
- **Input preparation 2:** Processing pdb files via 3DNA software.
- **Input preparation 3:** Creating additional files.
- **Input preparation 4:** Creation of proper directories and files for each sims.
- **Stage in:** Transferring input data to clusters via *rsync* command.
- **Connect:** ssh to cluster
- **PBS Submission:** All simulations are executed via submitting a pbs submit file that submits each simulation sequentially.
- **Parse Output:** Energy value 2000 is parsed for each simulation
- **Check:** Checked each output and simulations are decided as passed or failure
- **Stage-out:** Outputs are transferred back to local machine via *rsync* command

We have designed a workflow in Pegasus using those scripts:



Pegasus workflow of MD Simulations

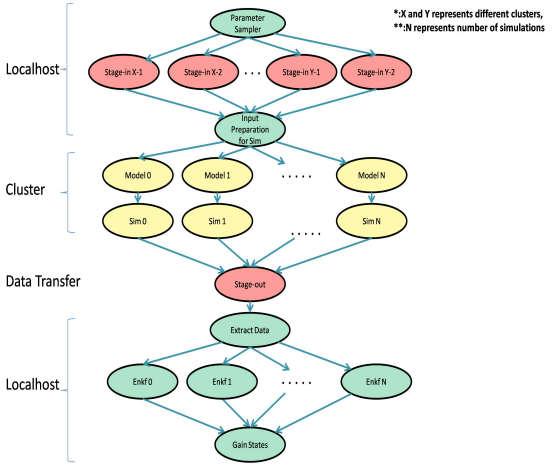
UCoMS

UCoMS (Ubiquitous Computing and Monitoring System), which is an ongoing project with the collaboration of Louisiana universities, aims to discover and manage energy sources.

There are four steps that are commonly performed for uncertainty analysis:

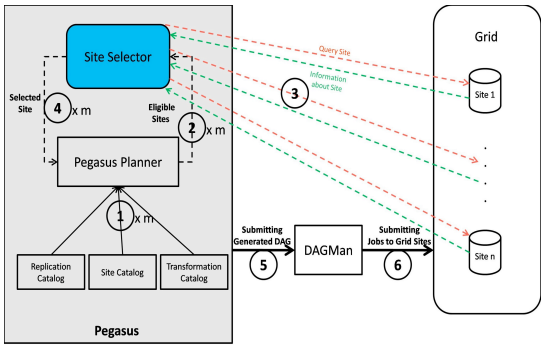
- Seismic inversion
- Reservoir modeling
- Flow simulation
- Post processing

We have designed workflow in Pegasus by using those steps to have end-to-end processing:

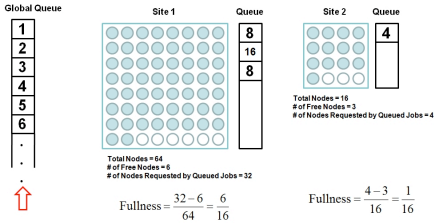


Pegasus workflow of UCoMS Project

Site Selection



$$\text{Fullness} = \frac{\#\text{of Nodes Requested by Queued Jobs} - \#\text{of Free Nodes}}{\text{Total \# of Nodes}}$$



$$\text{Fullness} = \frac{32-6}{64} = \frac{6}{16} \qquad \text{Fullness} = \frac{4-3}{16} = \frac{1}{16}$$

Improvements Achieved by Pegasus

- **Automation of Tasks:** saves time to scientists instead of running programs manually in a sequence.
- **Reliability:** Most of the failures can be corrected via retry mechanism.
- **Separation of Computing and Data Tasks:** Different type of tasks are handled differently. Stork is used for data transfers.
- **Running Jobs on Heterogeneous Batch Systems:** Pegasus uses Condor-G at the bottom level to submit jobs to different batch systems(PBS, Loadleveler, Condor, etc.)
- **Resource Independency:** Since Pegasus generates proper files for each site, scientists do not have to write different scripts for each site.
- **Utilization of Resources to Gain Extra Performance:** Our site selection mechanisms aim to get high throughput by mapping large number of jobs to least loaded sites therefore they give better performance comparing to simple site selection algorithms.

Connections with CyberTools

In our studies we attempted to introduce workflow concepts to science drivers in two large-scale scientific applications: DNA Folding and UCoMS. Both applications are workflow-enabled and have gained reliability and performance improvements.

We are currently in the process of workflow-enabling Coastal Ocean Observing and Prediction projects.

Acknowledgements

This project is in part sponsored by the National Science Foundation under award numbers CNS-0619843 (PetaShare) and EPS-0701491 (CyberTools), and by the Board of Regents, State of Louisiana, under Contract Number NSF/LEQSF (2007-10)-CyberRII-01.

Enhancements in Stork Data Placement Scheduler

Mehmet Balman, Tevfik Kosar

Center for Computation and Technology, Louisiana State University

STORK: A Scheduler for Data Placement Activities

Data management has been a crucial problem in every stage of computer engineering, from micro to macro level systems. We focus on data access and data placement problems in distributed systems for large scale applications. We study aggregation of requests in order to increase the throughput especially for transfers of small data files. We also explore the possibility of an efficient error detection and reporting system for distributed environments. In addition, we are investigating techniques to make use of replicas for multi-source downloads. We also work on several enhancements like file similarity analysis and semantic compression methods to reduce total transfer size.

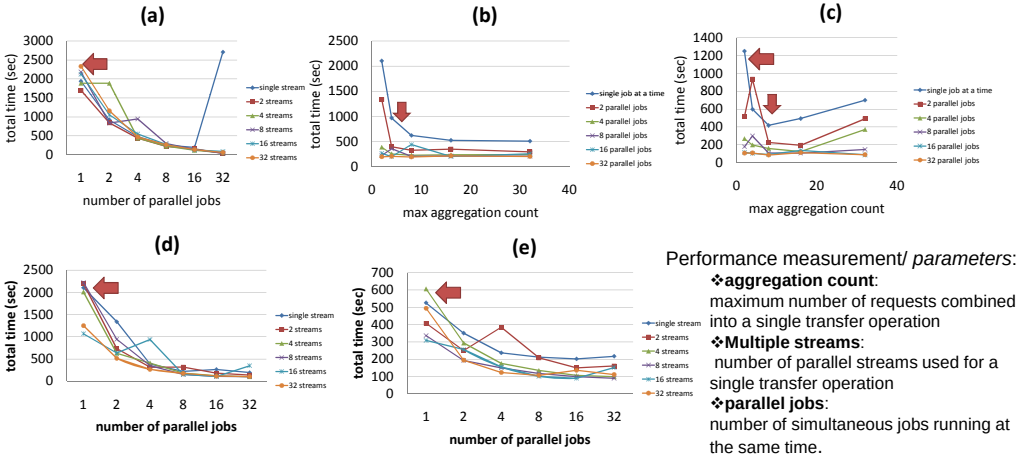
Aggregation of Data Placement Jobs

According to the file size and source/destination pairs, data placement jobs are combined and processed as a single transfer job. Information about the aggregated job is stored in the job queue and it is tied to a main job which is actually performing the transfer operation such that it can be queried and reported separately.

We have seen vast performance improvement, especially with small data files, simply by combining data placement jobs based on their *source* or *destination* addresses. The main performance gain comes from decreasing the amount of protocol usage and reducing the number of independent network connections. Thus, Stork makes better use of underlying infrastructure by coordinating and arranging data placement jobs.

Experiments on LONI (Louisiana Optical Network Initiative)

✓ **Test-set:** 1024 transfer jobs from Ducky to Queenbee (rtt avg 5.129 ms) - 5MB data file per job



Performance measurement/ *parameters*:

- **aggregation count**: maximum number of requests combined into a single transfer operation
- **Multiple streams**: number of parallel streams used for a single transfer operation
- **parallel jobs**: number of simultaneous jobs running at the same time.

Fig: *Effects of parameters over total transfer time of the test-set*
(a) without job aggregation – number of parallel jobs vs number of multiple streams
(b) transfer over single data stream – aggregation count vs number of parallel jobs
(c) transfer over 32 streams – aggregation count vs number of parallel jobs
(d) at most 2 requests are aggregated – number of parallel jobs vs multiple streams
(e) at most 16 requests are aggregated – number of parallel jobs vs multiple streams

Connection with CyberTools

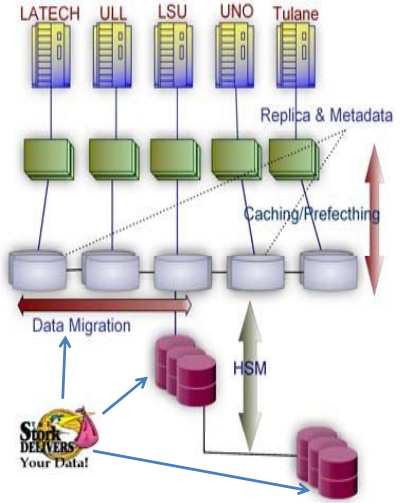
PetaShare Core Architecture

Two types of data movement:

- First, data needs to be perfected from low level storage layers to the higher levels such that management of data access has to be handled in an efficient manner.
- Second, data should be migrated between those five contributing institutions; moreover, data should be scheduled and moved between distributed sites and the clients.

Protocols:

file:/	->	local file
ftp://	->	FTP
http://	->	HTTP
gsiftp://	->	GridFTP
srb://	->	SRB (Storage Resource Broker)
irods://	->	iRODS

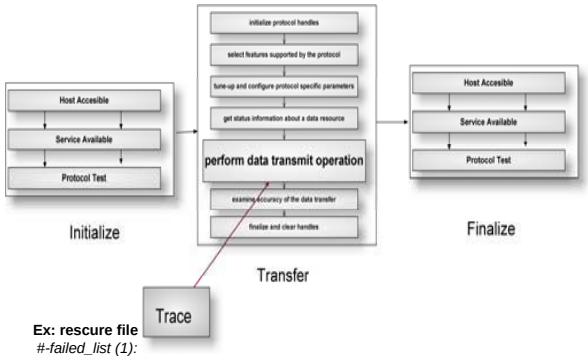


Error Detection and Recovery

Stork, data placement scheduler, checks network connection and availability of data transfer protocol beforehand with the help of new network exploration module. We have implemented error detection and classification as new features inside Stork. Our experiments, in which we are generating artificial errors for testing purpose, shows that current data transfer protocol are not always able to generate adequate log information; therefore we also focus on tracing the transfer job and preparing the infrastructure to explore dynamic instrumentation while transfer is in progress.

stork.globus-url-copy:
(supports wildcards and recursive copy)

Stork GridFtp data transfer module is able to verify the successful completion of the operation by controlling checksum and file size. Besides, it can recover from a failed operation. In case of a retry from a failure, scheduler informs the transfer module to recover and restart the transfer using the information from a rescue file created by the transfer module.



Stork.globus-url-copy features

`-ckp | -checkpoint -`
use a rescue file for checkpointing
`-ckpdebug | -checkpoint-debug`
`-ckpfile <filename> | -checkpoint-file <filename>`
checkpoint filename. Default is "<pid>.rescue"
`-cksm | -checksum >`
checksum control after each transfer
`-pchck | -port-check`
check network connectivity and availability of the protocol

Ex: rescure file
#-failed_list (1):
gsiftp://dsl-turtle06.csc.lsu.edu/tmp/test/x_ttest2.tar file:///tmp/x_ttest2.tar
#-expanded_url_list (7):
gsiftp://dsl-turtle06.csc.lsu.edu/tmp/test/out file:///tmp/out
gsiftp://dsl-turtle06.csc.lsu.edu/tmp/test/test2/test22/ file:///tmp/test2/test22/
#-transferred_list (2):
gsiftp://dsl-turtle06.csc.lsu.edu/tmp/test/test1/ file:///tmp/test1/
gsiftp://dsl-turtle06.csc.lsu.edu/tmp/test/test2/ file:///tmp/test2/

Acknowledgement

DSL Distributed System Laboratory
www.dsl.csc.lsu.edu

www.storkproject.org

This project is in part sponsored by the National Science Foundation under award numbers CNS-0619843 (PetaShare) and EPS-0701491 (CyberTools), and by the Board of Regents, State of Louisiana, under Contract Number NSF/LEQSF (2007-10)-CyberRII-01.

Low Cost Fabrication of Micro- and Nanopores in Free-Standing Polymer Membranes for Study of Lipid Adsorption

Junseo Choi, Anish Roychowdhury and Sunggook Park

Department of Mechanical Engineering and Center for Bio-Modular Multiscale Systems, Louisiana State University

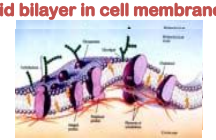
Abstract

We present low cost fabrication of a large area, free-standing SU-8 membranes with perforated micropores up to 4 inch diameter. For the fabrication, a combination of imprint lithography and a sacrificial layer technique was employed in order to obtain a clean, fully released, and mechanically stable membrane. The fabricated membrane was used to selectively immobilize lipid vesicles at the micropores in the membrane. This result indicates that the perforated polymer membranes with micro- and nanoscale pores have potential as a platform for fundamental study of biological systems. We also show integration of the polymer membrane into microfluidic devices made of polydimethysiloxan (PDMS), which allows for in-situ study of lipid adsorption.

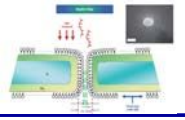
Objective

- To study an adsorption behavior of lipid at the micro pores in the polymer membrane made by a combination of imprint lithography and a sacrificial layer technique.
- Immense research potential in biological systems [1], protein lipid studies, DNA sequencing [2], polymer photonics and component for BioMEMS.
- Need to develop a low cost, parallel process allowing for controlled pores size, location and mechanical stability for fabrication of the perforated membranes.

Lipid bilayer in cell membrane [1]

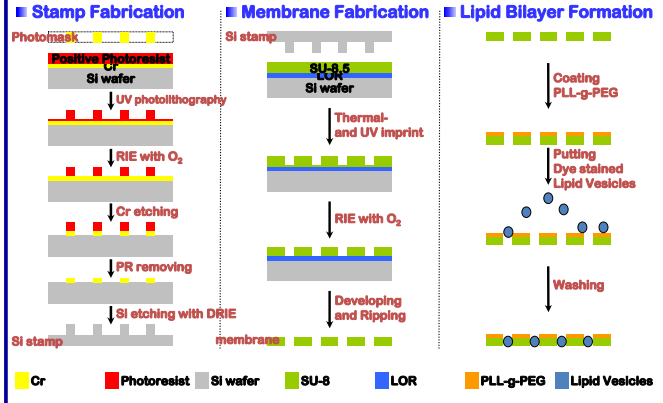


Nanopores for DNA sequencing [2]



Experiments

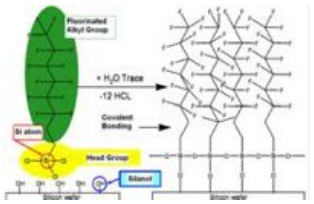
1. Fabrication Process



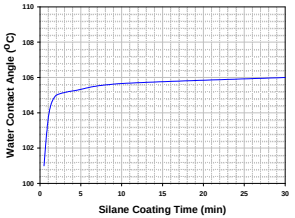
2. Surface Treatment on Si Stamp

A vapor phase deposition process was employed in a home-made chemical vapor deposition chamber to coat stamp surface with fluorinated silane to reduce adhesion of stamp surface. [3]

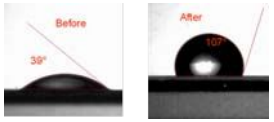
Chemistry of Silane Adsorption



Optimization of Deposition Time



Water Contact Angle Measurement

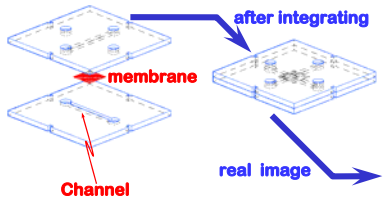


- The silane coating increased hydrophobicity of Si stamp surfaces.
- The optimum value of silane deposition time was 10minutes.

3. Integration of the Membrane into Microchips

The membrane with perforated micropores was sandwiched by two PDMS microfluidic devices. The microchannels in the two PDMS devices were so aligned to be perpendicular to each other.

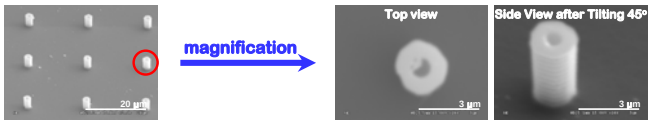
Schematic Images



Photograph

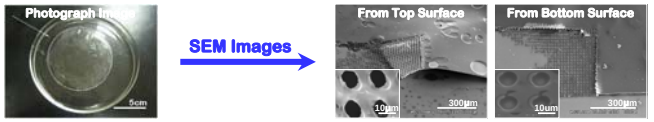
Results

1. Stamp Fabrication



- DRIE results in almost vertical sidewalls and a scallop-like features on the sidewalls.
- The smallest feature obtained by photolithography and Si DRIE is pillars of 2 μm diameter.

2. Free Standing SU-8 Membrane

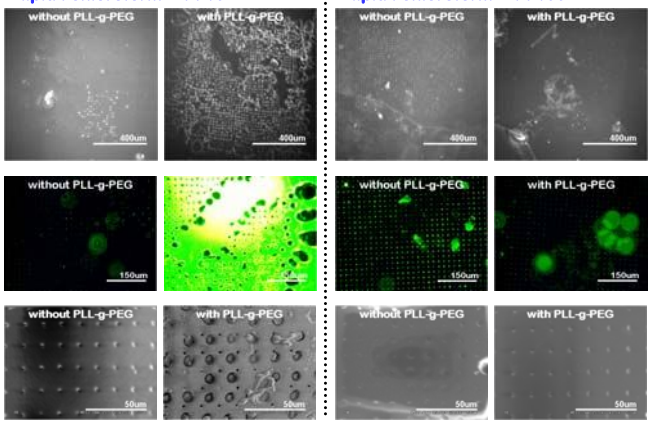


- Clean and fully released polymer membrane, 4 inch size, was achieved.

3. Microscope, Fluorescence and SEM Images after Staining with Dye

lipid : chloroform = 1 : 10

lipid : chloroform = 1 : 100



- Lipid vesicles adsorb at the micropore sites in the SU-8 membrane.
- Lipid adsorption behavior was affected by quality of imprinted patterns and dilution of lipid solution
- When the membrane surface was treated with PLL-g-PEG prior to the lipid adsorption, the fluorescence signal becomes weaker. This indicates a possible formation of lipid bilayers at the pore sites.

Conclusions

- Technology to produce free-standing perforated membranes in SU-8 layers using all parallel processes were successfully developed using a novel combination of imprinting lithography and sacrificial layer techniques.
- Lipid vesicles were selectively immobilized at the fabricated pores in the membranes. Lipid adsorption behavior depend on the concentration of the lipid solution, surface treatment as well as the fouling of the membrane surface due to imprint failure.
- The number of active pores can be controlled simply by using microchannels of different widths. This will also alleviate the requirement of high accuracy in aligning two PDMS devices for bonding.

References

- http://med.tn.tudelft.nl/~hadley/nanoscience/week4/4.html
- S. M. Lqbal et al, *Nature Nanotechnology* 2007, 2, 243-248
- H. Schift et al, *PSI scientific reports* 2004

Development and Application of Material Point Method for Structure Calculations in Biological Systems

Timur Gilmanov[1], Anvar Gilmanov[2], Sumanta Acharya[1, 2]

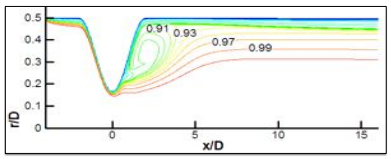
[1]LSU, Mechanical Engineering
[2]LSU, Center for Computation and Technology

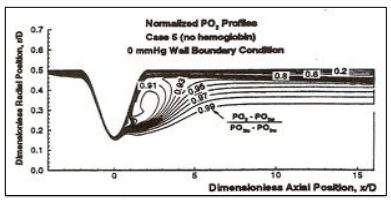
Abstract

For many problems in biological systems it is essential to take into account the deformation of tissue under the action of fluid/blood flow and conversely, the influence of the tissue deformation on the fluid flow.

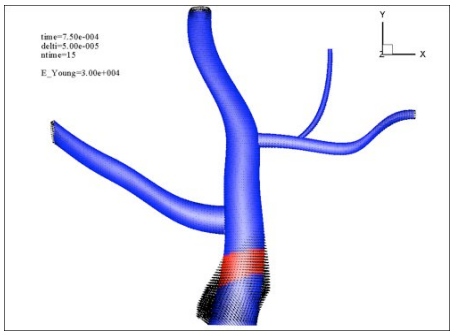
Fluid-Structure interaction (FSI) problems are considered as a solution of a coupled system of equations describing the behavior of fluid and structure which act on each other across the common boundaries.

An improved FSI technique is being developed that will efficiently handle problems with large structural deflections.

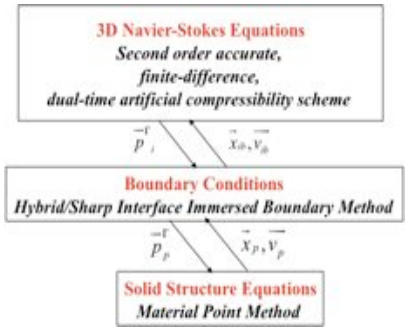


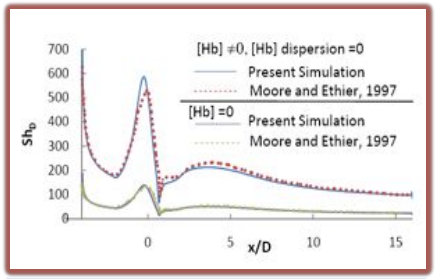


Simulation of Blood Flow-Oxygen Concentrations in Arterial Vessels with Stenosis
Top: Our Computations Bottom: Results from Moore & Ethier

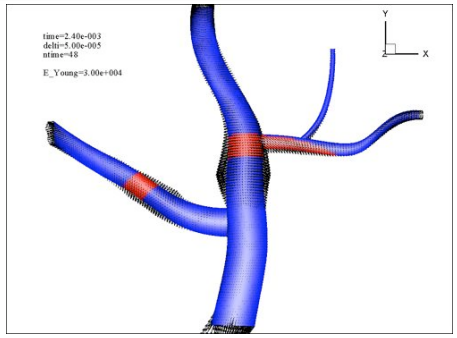


FSI Algorithm





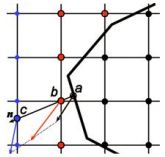
Surface Mass Transfer Rate in Arterial Vessels with Stenosis:
Comparison of our Computations (solid) with Moore & Ethier



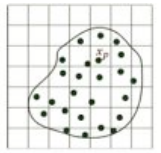
Simulation of Solid Structure: Deformation of arterial structure under the pulsation of pressure in the form of progressive wave. The pressure is given as

$$\Delta P(s,t) = f(s - c_0\, t)$$

IBM



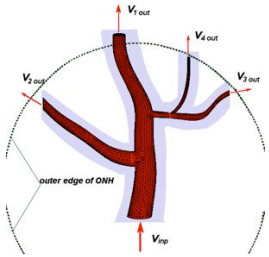
MPM



Background grid for solution of momentum equations

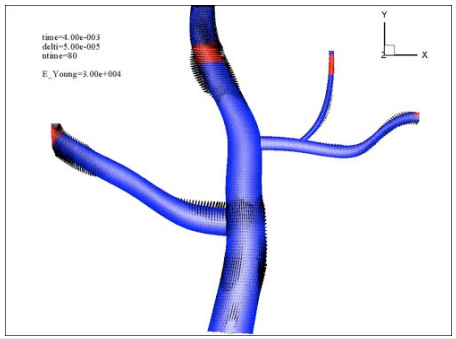
Connections with CyberTools

A key goal of the proposed work is to understand the oxygen transport behavior in arterial vessels.



An example of a branched arterial structure is shown in the figure. We will utilize the FSI algorithm developed to solve the problem & understand the essential flow physics.
We will use Cybertools and the CFD toolkit to solve highly resolved calculations that include:
- FSI involving tissue deformation;
- diffusional transport and consumption in tissue/walls;
- upscaling of atomistic simulations for transport properties.

WP4 Collaborators: S. Jha, M. Tyagi, E. Schnetter



Acknowledgements

This work is currently supported by the NSF EPSCoR.
The simulations were run on LSU's and LONI's HPC resources.

DEVELOPMENT OF CFD MODULES FOR CFD-TOOLKIT

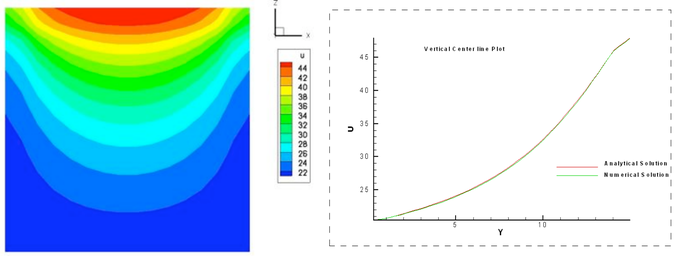
Prasad Kalghatgi , Sumanta Acharya
LSU, Mechanical Engineering
WP4 Co-ordinators: Dr. Mayank Tyagi Dr. Shantenu Jha Dr. Erik Schnetter

ABSTRACT

Bio-fluid systems of interest are invariably characterized by low Re flows in deformable complex domains such as blood flow in arteries. 3D-simulation of such a flow scenario requires an adept flow solver handling complex domain and low Re flows.

A block structured finite volume code to solve unsteady incompressible Navier-Stokes Equations is being developed with CGNS grid interface. A hybrid Staggered/Non-Staggered formulation is being used and is specifically suitable for implementing the immersed boundary method on curvilinear meshes. This feature is useful for biological systems.

Validation of Diffusion on Multi-block

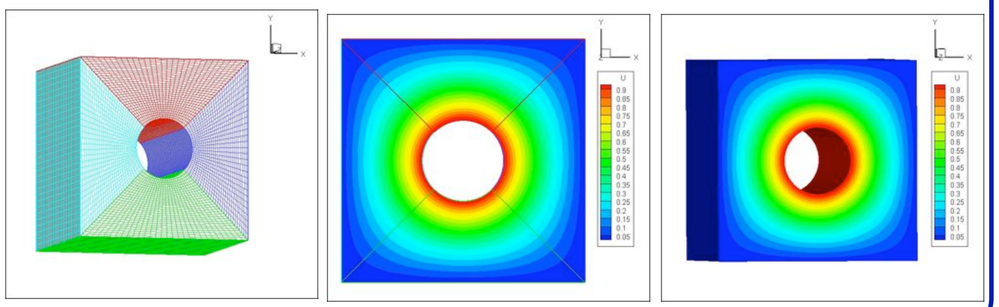


Governing Equations

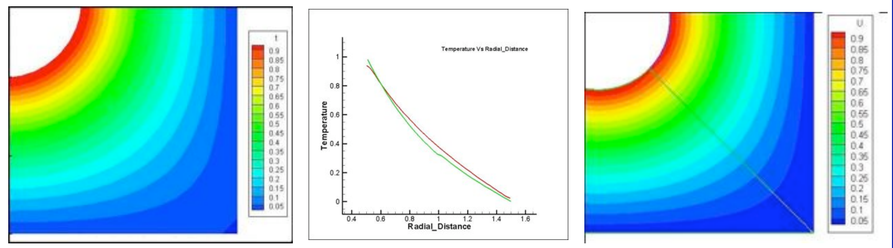
$$\frac{\partial v_i}{\partial x_i} = 0$$

$$\frac{\partial v_i}{\partial t} + \frac{\partial v_i v_j}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \frac{\mu}{\rho}\frac{\partial^2 v_i}{\partial x_j \partial x_j}$$

Diffusion in Curvilinear Multi-block Mesh



Comparison With Multi-Patch Simulation in Cactus



Multi Patch Simulation

*Courtesy Dr. Schnetter

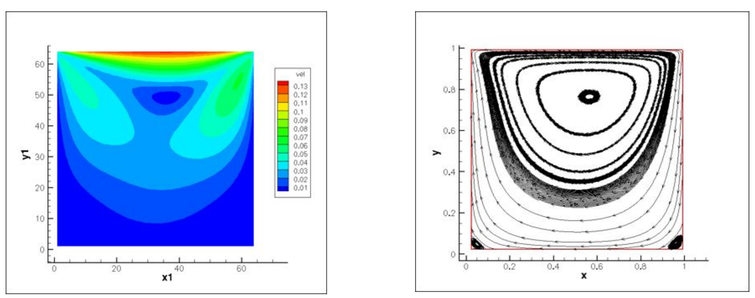
Comparison of CFD Modules with Multi-Patch

CFD Module Simulation

Data Structure

4D array Data Structure
U[m][j][k][l]
M → Block ID (grid Hierarchy)
J, k, l-→ local grid index

Flow Simulation Results
Lid Driven Cavity, Re = 100



Velocity Contours

Stream Lines

Features of Code

- CGNS interface, imports MB grids & BC's from commercial grid generator.
- Staggered/Non-staggered approach on MB curvilinear grid.
- 2nd order accurate FV discretization
  (CD for diffusion & QUICK for convection, second order time integration)
- Fractional Step Method for pressure momentum coupling.
- BC's tagged to each boundary cell face to support partial block connectivity.
- Hypre solver for efficient parallel calculations of algebraic system of equations.

Connections with CyberTools

Subroutines To be Ported in Cactus

ReadCg_Grid.c
ReadCg_Bc.c
Allocate_Mem.c
Geom.c
Metric.c
Set_Ghost_zone.c
Exchange_Ghost_Zone.c
Diffusion_Flux.c
Convective_Flux.c
PressurePoission.c
FractionalStep.c

These Modules will be ported as Cactus thorns

Acknowledgements

This work is currently supported by the NSF EPSCoR.
The simulations were run on LSU's and LONI's HPC resources.

Towards Cyber Infrastructure for Dynamic Storm Surge Predictions

Archit Kulshrestha[1], Harsha Bhagawaty[2], Gabrielle Allen[3], Nathan Brener[4], S.S.Iyengar[5]

[1,2,3,5]Center for Computation & Technology, Louisiana State University

[1,2,3,4,5]Department of Computer Science, Louisiana State University

Abstract

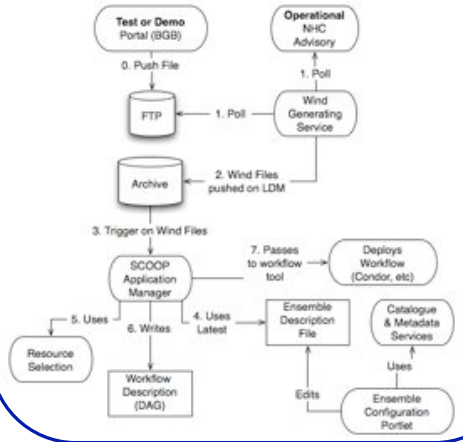
The Louisiana Coastal Area presents an array of rich and urgent scientific problems, such as hurricane forecasting or wetland erosion, that require new computational approaches. Dynamic and adaptive capabilities are crucially important for many of these problems, providing the ability to integrate coupled models with real-time sensor information, or to enable deadline based scenarios and emergency decision control systems.

This poster describes a scenario where new real time data driven algorithms could improve decision support systems for responding to the effects of hurricane and severe storm events. Motivated by the SURA Coastal Ocean Observing and Prediction workflow, we illustrate how dynamic selection of runtime parameters for storm surge models can effect both the accuracy and total runtime of the system. Research on algorithms for dynamic data driven application systems (DDDAS) is important for many science drivers in CyberTools which involve real time data or control systems.

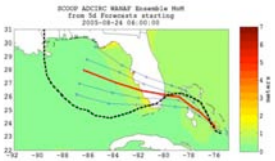
Introduction

Hurricane are a major threat to life and property in Louisiana and other coastal areas. Hurricane Katrina and Hurricane Rita demonstrated that the key to saving lives is timely prediction and ample warning. In order to predict the effect of hurricanes and disseminate the results to the proper authoritoes it is important that the whole process be automated and .the system be capable of making dynamic decisions based on available information. As part of the SCOOP program an end to end system was developed that reacts to coastal events and triggers various wind, surge and wave models. The SCOOP system relies on the SCOOP archive to receive and archive various data products and trigger coastal models upon their arrival. The output from these models is then ingested back into the archive and visualized. In this work we study the use of different grids for storm surge predictions and the effect on the turn around time and accuracy of the models. We propose that a optimal schedule can be estabilished that provides fast turn around times and accurate results dynamically based on the threat level.

SCOOP Workflow

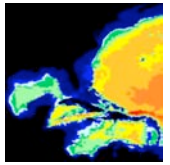


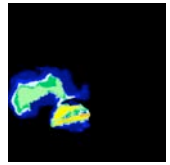
Results



Results were obtained by running the storm surge simulation for Katrina storm using ADCIRC on the two grids described in the poster. The results showed that a 7 day forecast ran for 1hr 39mins on 64 processors of the LONI queenbee machine while the smaller grid took less than 10 minutes to run.Larger grids with over 2 million nodes will take more than 6 hours to complete which is too long for emergency response purposes. Future work will focus on maing dynamic decissions on which gris to use.

ADCIRC Grids





Two different grids were used one with 31435 nodes covering a large area and other with 598240 nodes which is a high resolution grid of the New Orleans Lake Pontchartrain area.
Finer grids improve the accuracy of the results but also take longer to run. Scalng issues cause the time taken to generate the results to increase. Depending on the urgency of the impending storm/hurricane, an optimal grid can b e chosen so as to predict a path of reasonable accuracy in a short span of time and provide accurate results for emergency planning.

Connections with CyberTools

This work is connected to CyberTools WP1 due to its use of the LONI resources. WP3 will provide the visualization services and WP4 will provide the dynamic decission algorithms.

Acknowledgements

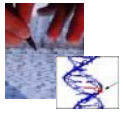
We would like to thank the entire SCOOP team, Brett Estrade, LONI support staff, Werner Benger, Hartmut Kaiser, RakeshYadav.

An Automated Genosensor System using Modular Microfluidics

M. L. Hupert,[1] H. Wang,[1] H.-W. Chen,[1] M. A. Witek, [1] S. K. Njoroge, [1] W. Stryjewski, [1] D. Patterson, [1] P. Datta, [2] P. Chen, [3] B. H. You, [3] J. W. Guy, [1] J. Goettert, [2] D. E. Nikitopoulos, [4] M. C. Murphy, [3] M. A. Batzer, [4] and S. A. Soper [1,3]

[1]- Department of Chemistry, [2]- Center for Advanced Structures and Devices (CAMD), [3]- Mechanical Engineering, [4]- Department of Biological Sciences
Louisiana State University, Baton Rouge, LA 70803

Introduction



www.edeincorp.com

Genotyping

Pathogen Detection

Molecular testing and genotyping is of significant importance in diagnostics/ prognostics of disease states or detection and identification of biopathogens. These tests typically require multiple laboratory operations performed by highly trained personnel using a collection of task-specific instruments. For example, genetic testing involves the following complex set of steps; (i) cell lysis; (ii) nucleic acid extraction; (iii) amplification; (iv) sequence variation discrimination; (v) detection of reaction products.

We are developing a universal, portable instrument for automated sample preparation and genetic testing. All of the bioanalytical processing, from sample reception to readout, is done on a disposable, plastic microfluidic chip. The operation of the chip is provided by electronic, optical, and hydraulic controls located off-chip. The sequence of sample processing steps performed on the flow-through plastic biochip includes cell lysis, DNA extraction, polymerase chain reaction (PCR) with or without ligase detection reaction (LDR), and DNA universal array read-out or micro capillary electrophoresis.

The unique feature of the system is that it can be easily reconfigured and used with other test specific chips for a much broader range of applications based on nucleic acid testing such as human identification and pathogen detection. The CyberTools are being used to optimize the design and performance of this system.

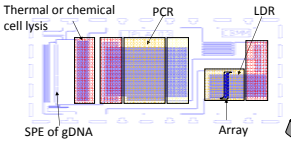
System Overview and Operation

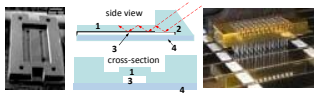
Load sample and reagents
⇩
Insert chip into instrument
⇩
Press "RUN" and wait for results

- **Instrument measures 12" x 12" x 8" (electronics, optics)**
- **Fully integrated (load sample and reagents only operator requirement)**
- **Fluidic chips are hot-embossed or injection molded (no active components)**
- **Low cost per integrated chip**
- **Off-chip active components (reusable)**
- **Reconfigurable – performs different molecular assays**
- **Fast assays (< 30 min)**
- **Computational simulations used for component optimization**

System Configurations

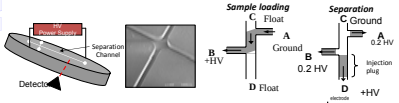
Cell Lysis, DNA Immobilization (SPE), PCR, LDR





microarray readout

SEM of microarray chip; Schematic of microarray unit; 1- integrated waveguide; 2- coupling prism; 3 – microfluidic channel; 4 – coverplate



Cell Lysis, DNA Immobilization, PCR

micro capillary electrophoresis

Due to material requirements, the SPE and PCR functions must be carried out using a polycarbonate (PC) and poly(methyl methacrylate), PMMA, is employed for the electrophoresis and microarray readout.

System integration using Passive Alignment Structures

The final system has fluidic modules that must be interconnected with no leaks to provide the necessary processing steps. Our approach to module integration is based on the implementation of passive assembly technology in molded polymers. Screw theory can be applied to the design of appropriate combinations of kinematic pairs that do not over-constrain the assembly

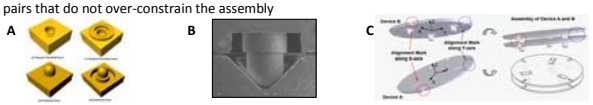
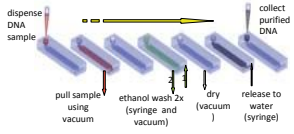


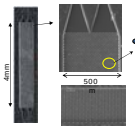
Figure A presents the regular design of a hemisphere tipped recess and the resulting hemispherical pin (a, b) and the modified design of the hemisphere tipped recess with the annular structure (c, d). (B) is a cutaway image of an assembled hot embossed stepped alignment structure, showing the stepped hemisphere-tipped pin in the v-groove. (C) is test bed for evaluating the use of passive alignment structures for the assembly of polymer microfluidic devices.

Developed Methodology and Technologies

Purification of DNA using Solid Phase Extraction (SPE)

Molecular tests require obtaining pure extracts of nucleic acids (i.e., removal of proteases, enzyme inhibitors, salts, dyes and other contaminants). Generally, solid phase extraction includes three steps; (i) selective immobilization of nucleic acids from a crude sample on an activated surface; (ii) removal of contaminants through washing; (iii) release of purified nucleic acids. The extraction bed is created following embossing of the fluidic network using a simple UV activation step of PC.



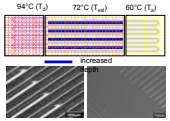


Microposts configured within the capture bed used to increase the load level of gDNA.

Binding capacity of capture bed: **190 ±40 ng of the gDNA**

Surface loading density: **790 ng/cm²** (0.26 fmol/cm²)

Continuous Flow - Polymerase Chain Reaction (CF-PCR)



Polycarbonate (PC)

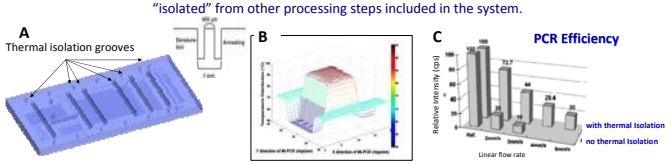
- Thermal zones ratio: D:A:E = 1:1:4

Channel width: 80 μm
Channel length: 1450 mm

Reactions such as the polymerase chain reaction (PCR) or Ligase Detection Reaction (LDR) rely on subjecting the reaction mixture to predefined thermal cycles. Conventional methods are chamber-type processes in which a stationary reaction mixture is alternately heated and cooled. Continuous flow (CF) thermal cycling is based on flowing the reaction mixture in a microchannel repetitively through different isothermal zones – primary advantage of CF format is FAST cycling .

The cocktail with DNA template is transported to the **CFPCR**, hosting 30 cycles. The reaction is performed with 3 temperature zones; 10 s for denaturation, 10 s for annealing and 30 s for extension with sample flowing at 0.6 μL/min. Longer extension times are achieved using a deeper (240 μm) channel in that zone.

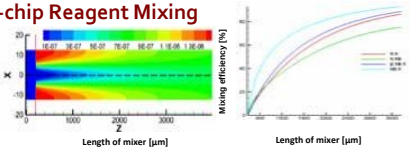
Biochemical reactions that are thermally controlled and using active heating/cooling elements must be "isolated" from other processing steps included in the system.



(A-B) A cross-section view of grooves and fins and (B) the temperature distribution of CFPCR reactor obtained via FE simulations with ANSYS. (C) The relative intensity of amplification efficiency at each flow rate compared to the reference –commercial thermal cycler.

On-chip Reagent Mixing

The Reynolds number in microfluidic channels is low, resulting in laminar flow. Therefore, the mixing of binary or multicomponent fluid streams in a microchannel relies mainly on diffusion. Diffusional mixing is fast in the nanoscale, but slow in the micro- and macroscale, owing to the nonlinear dependence of diffusion time with distance ($x^2 = 2Dt$).



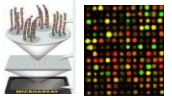
Numerical simulations aid in the design of the most effective mixer for reaction buffers.

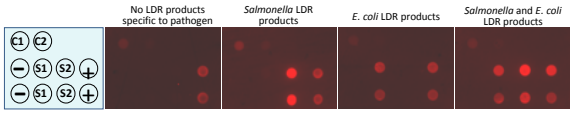
Microarray Readout - Pathogen Detection

rapid, selective, specific, and simultaneous detection

- Foodborne pathogens (*Escherichia coli* O157:H7 and *Salmonella)*
  - important targets for the control of food safety and public health
  - approximately 70 million illnesses and 5,000 deaths each year in the US
  - food safety testing and healthcare costs total nearly 10 billion dollars

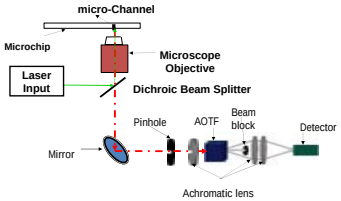
Sharma, V.K.; Carlson, S.A. *Appl. Environ. Microbiol.* **2000**, *66*, 5472-5476
Ahn, S.; Wolf, D. R. *Anal. Chem.* **2005**, *77*, 5041-5047





Results of pathogen detection; C1, C2: 20 μM and 10 μM Cy5-$(T)_{10}$-NH_2 spotting and immobilization control; - -negative control; S1- probe targeting *E. coli* O157:H7 *eaeA* gene; S2 - probe targeting *Salmonella sipB/C* gene; + - hybridization positive control.

Laser-Induced Fluorescence Detector for Multi-Color Analysis

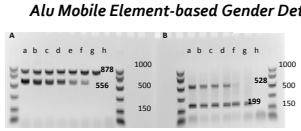


Acousto-optic tunable filter (AOTF) based Fluorescence Reader

- Solid state electron-optic device which is compact, with no moving parts and relaxes stringent alignment requirements
- Acoustic (vibrational) waves at RF are used to separate a single wavelength from multicolor source
- The wavelength of light selected is a function of the frequency of RF applied to the crystal, user selectable λ
- High efficiency device with transmissions at selected wavelength as high as 98%
- AOTF coupled to a detector offer a more rugged, flexible, field deployable and adaptable system

Microchip Capillary Gel Electrophoresis

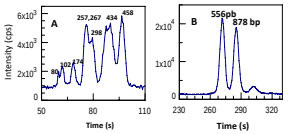
***Alu* Mobile Element-based Gender Determination**



Gel electrophoresis of amplicons of (A) *AluSTXa:* Male (M): 556 and 878 bp, Female (F): 878 bp only and (B) *AluSTYa*: M: 199 and 528 bp, F: 199 bp. Lanes (a-h) are different F:M DNA ratios; (a) 0:10, (b) 1:9, (c) 3:7, (d) 5:5, (e) 7:3, (f) 9:1, (g) 10:0, (h) 0:0.

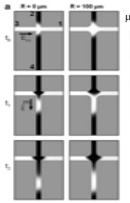
The PCR cocktail contained both *Alu STXa* and *STYa* primer sets, targeting different loci on the X and Y chromosomes. For the *STXa* a filled site in X chromosome would give an 878 bp product and for an empty site in Y chromosome, a 556 bp product. For *STYa*, a 528 bp and 199 bp products were expected for filled sites in Y chromosome, and an empty site in X chromosome, respectively. For both loci, males are distinguished as having 2 amplicons, while PCR with female DNA gives one.

Experimental results



Electropherograms of (A) pUC18 (stained with POPO-3, E = 200 V/cm) and (B) amplicons from typing Alu STXa for male DNA (E = 125V/cm).

Numerical Simulations



Numerical simulations of loading and dispensing of sample into cross injectors with different geometries. a The pictures present only the central part of the simulated microchannels. 1 waste, 2-buffer, 3-sample, and 4-separation microchannel. Loading: 1 at 33.4 V, 3 at GND, 2 and 4 float. Dispensing: 1 and 3 float, 2 at GND, 4 at 66.8 V. $E_{load} = E_{disp} = 167$ V/cm; $t_{load} = 7.5$ s, $t_{disp} = 11$ s. t_0 corresponds to the start of dispensing; t_1 and t_2 were taken at 1 s increments.

Connections with CyberTools

- Numerical simulations of fluid flow, heat transfer as well as electrical transfer, are currently used as an aid to develop design rules for multi-module biosensors required for constructing systems. Currently, only component level simulations can be carried out, once Fluent and ANSYS are migrated to HPC, system-level optimization simulations will be carried out.
- The use of simulations will minimize the number of test devices/systems that need to be fabricated using lithographic or HPMM processing, significantly reducing cost and time associated with the development of prototype devices.
- Current Experimental results will be used for algorithm and simulation optimization and verification. Once CyberTools are verified, these tools can be used for future system design.
- System modeling and numerical calculations include system assembly, material mismatch and selection, integration, geometrical architectures, thermal management, mixing etc.

Acknowledgments

This work is generously supported by the National Science Foundation (EPS-0346411) and the State of Louisiana Board of Regents Support Fund.

Small Molecule Sensor: HTS for Drug Discovery

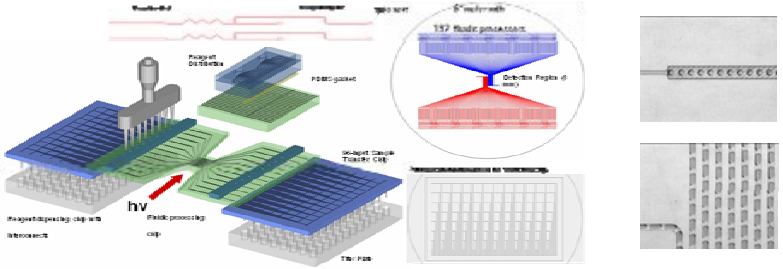
Paul Okagbare, Jason Emory, Wonbae Lee, Subramanian Balamurugan, Namwon Kim, Dimitris Nikitopoulos, Michael C. Murphy and Steven A. Soper

Department of Chemistry and Mechanical Engineering
Louisiana State University

INTRODUCTION

We are developing a high throughput screening (HTS) modular microfluidic system that will provide high levels of automation and the ability to carry out HTS campaigns in a variety of small laboratory/company settings embarking upon drug discovery projects. The fluidic elements will be fabricated into polymers using replication technologies from masters developed through a variety of processing techniques. The basic fluidic element will consist of: (i) Arrays of capillaries oriented in a footprint to match a conventional 96-well titer plate to feed inhibitor (small molecules) candidates from the titer plate into the fluidic system; (ii) Passive micro-mixer consisting of high aspect ratio microstructures for minimizing mixing time; (iii) Nanoreactors consisting of aqueous fluidic droplets suspended in an immiscible fluid and; (iv) Highly sensitive fluorescence reader to monitor enzymatic activity.

To test and evaluate the performance of the proposed system, we will screen inhibitors, from small molecule combinatorial libraries, of L1-EN. L1 genomic elements are active autonomous human mobile elements making up approximately 17% of the genome. At this point there are approximately 45 diseases caused by L1-driven events. We will use the normal oligonucleotide motif recognized by L1-EN and insert fluorescent labels onto this oligonucleotide. In order to perform a large number of assays, we will investigate and implement a parallelized detection scheme based on a CCD detector. The parallelized measurements will include fluorescence cross-correlation spectroscopy (FCCS). A rendition of this system is shown schematically below.



Integrated fluidic system for performing HTS assays. The system is configured on a 6" wafer and the wafer shown has 192 processors. The wafer consists of a stack of fluidic chips, with one chip used for containing the substrates and buffer reagents required for the HTS, a 96-element transfer chip to move drug candidates to the processor wafer and the HTS processor chip containing passive mixers, 2-phase flow generator and detector elements. The system is operated in a 2-phase flow format with an inert separator liquid to significantly increase processing throughput. The figures to the right show aqueous droplets or plugs suspended in a fluorinated hydrocarbon carrier fluid.

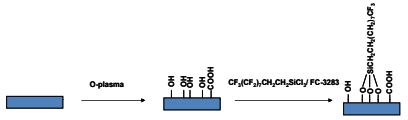
Connections with CyberTools

CyberTools will enable rapid progress and understanding of the scientific and engineering processes used to optimize and rationally guide design, construction and operation of components and eventually, systems comprised of thermal and fluidic components using the HPC platform. In addition, hybrid-codes (continuum/MD codes) will be transitioned to the HPC platform as well for evaluating multi-phase flows in mixed-scale structures . Specifically:

- **Source Code Review:** Review of existing open-source CFD and MD software, which will be included in WP4 and modified for LONI use. This will provide us the ability to perform system-level optimizations.
- **Microchannel/nanochannel Geometry**: Mixed-scale CFD simulations are being developed to monitor fluid transport in microchannels (continuum formulations) and nanochannel domains (MD formulations).
- **Test Case Simulations:** CFD and MD simulation methods to predict experimental observables in sensor test cases operated in mixed-scale domains.
- **CFD/MD Integration with CACTUS:** CACTUS-based toolkit used for CFD and MD test case simulations.
- **Non-Newtonian Fluidic simulations:** For molding high aspect ratio structures over mixed scales, new codes will be developed to model molding and demolding of fluidic components using flowing polymer melts.
- **Toolkit-based Simulations:** Utilization of toolkits for full scale simulations across institutions.

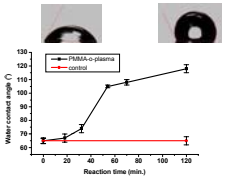
Surface Engineering for Droplet Microfluidics

- In droplet microfluidics, each droplet of approximately 1 nL volume is used as a microreactor. The controlled and rapid mixing of fluids in the droplet reactors results in decreased reaction times. In this format a large number of reactions can be carried out in parallel using small volumes of the reagents.
- In order to carry out the kinetics of the reaction of L1-EN with the library of compounds, droplets of the enzyme, the substrate and the compound libraries should be formed in the immiscible carrier perfluoro liquid (FC-3283) and mixed precisely.
- In PMMA and PC microfluidic devices, the aqueous droplets are stable only for a few minutes. In order to form the stable two-phase flow for longer time, the surface of the microfluidic device should be render hydrophobic.
- We modified the surface of the PMMA microfluidic device, by reacting with the perfluoro compounds as shown in Scheme-1.
- The water contact angle measurement showed that the modified PMMA surface is hydrophobic.

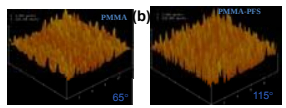


Scheme 1. Reaction scheme used for the fluorination of the PMMA surface

- Water contact angle (WCA) of PMMA (65± 2°) was increased (118± 3°) upon modification.
- The AFM images show that the reaction does not increase the surface roughness. The RMS surface roughness of these films were for PMMA 14. 87 nm and after the reaction 13.17 nm.
- This method is applicable to modify surfaces of other polymer such as polycarbonate (PC).

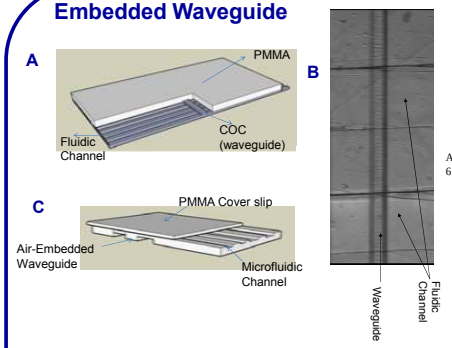


Extent of the reaction of perfluoro silane as a function of time as measured by water contact angle measurement



AFM images of (a) cleaned PMMA and (b) perfluoro silane modified PMMA surfaces

Embedded Waveguide



(A). Schematic of COC core embedded waveguide showing the integrated fluidic geometry. (B) Optical image of the fabricated fluidic device with orthogonal air-embedded COC core waveguide. (C) Schematic of air embedded waveguide fabricated on COC wafer.

Single Molecule tracking

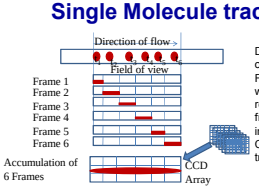


Diagram showing the migration of single molecule along the Field of View of the CCD which generated streak image resulting from accumulation of frames with charges collected in adjacent pixels when the CCD is operated in frame transfer mode

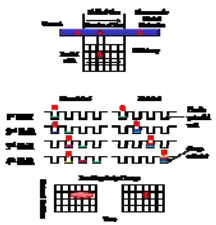
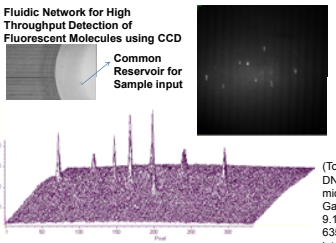


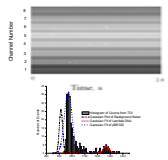
Diagram showing the Time-Delayed Integration (TDI) operational mode of a CCD camera, which occurs when there is a match between the parallel shift rate of the CCD and the rate of travel of emitting objects through the excitation volume. All photons are accumulated into the same potential well of the CCD when these rates are matched. If the timing is mismatched, the fluorescence photons are spread across multiple pixels, resulting in a smeared image.

High Throughput SMD with Frame Transfer CCD



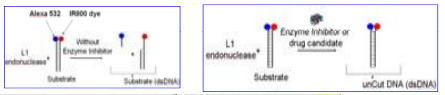
(Upper Right Pane;) Single DNA molecules migrating through a PMMA multi-channel chip (sample conc. = 100 pM). Image acquired with 10X/0.5 objective and 3 x 3 binning of CCD pixels, exposure time = 200 ms with CCD operated in Frame transfer mode. (Lower Panel) 3-D image showing intensity distribution of single DNA molecules in one frame of the CCD image

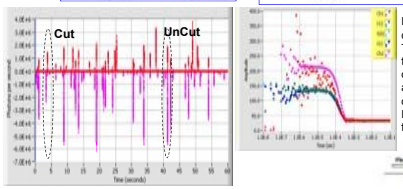
High Throughput SMD with CCD Operated in TDI Mode



(Top Panel) TDI image of λ-DNA and pBR322 DNAs traveling electrokinetically through 8 microfluidic channels with an orthogonally situated Gaussian laser beam (25 mM borate buffer pH 9.1; shift rate 8 ms; E = 125 V/cm; 10 mW; λex = 635 nm). (Bottom Panel) Histograms of the peak intensities versus number of events from TDI images shown in Figure 4a. The histograms were fit to Gaussian functions from which the mean burst amplitude and standard deviations were derived (mean = 1268, standard deviation = 23 for λ-DNA and mean = 974, standard deviation = 17 for pBR322). A Gaussian curve of the noise was also plotted to determine the detection threshold level of 952.

Fluorescence Cross-Correlation Spectroscopy

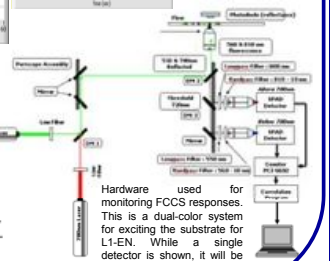




Fluorescence assay for determining L1-EN activity. The FCCS response is used to determine the population of cut vs. uncut DNAs that are labeled with two dyes, one on each strand. The FCCS response shown is from DNA cleaved by L1-EN.

Crosscorrelation function G(τ) : Mathematically, the normalized cross-correlation function G(τ) is calculated as time average (< >) of the product of the fluorescence fluctuations of two fluorescent species i at times t and j at t + τ, normalized by the product of the time-averaged fluorescent signals of two species i and j. z

$$G(\tau) = \frac{\langle \delta F_i(t)\delta F_j(t+\tau)\rangle}{\langle F_i(t)\rangle\langle F_j(t)\rangle}$$



Hardware used for monitoring FCCS responses. This is a dual-color system for exciting the substrate for L1-EN. While a single detector is shown, it will be combined with a CCD.

Acknowledgements We thank the National Science Foundation (EPS-0346411), National Institute of Health (1R01EB006639-01A1) and Louisiana Board of Reagents for financial support of this work.

Design and Fabrication of Small Footprint Continuous Flow PCR Devices for a Multi-Well CFPCR Platform

D. S. Park[1], P.-C. Chen[1,2], B. H. You[1,2], N. Kim[1,2], T. Park[1,2], T. Y. Lee[1,2], P. Datta[3], Y. Desta[4], S. A. Soper[1,5], D. E. Nikitopoulos[1,2], M. C. Murphy[1,2]

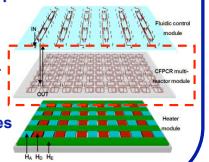
[1]Center for Bio-Modular Multi-Scale Systems, [2]Department of Mechanical Engineering, [3]Center for Advanced Microstructures and Devices (CAMD), [4]BioFluidica Microtechnologies, [5]Department of Chemistry, Louisiana State University, Baton Rouge

Abstract

Small footprint (8 mm x 8 mm) continuous flow (CF) PCR devices were designed, fabricated, and used to amplify DNA fragments as part of a multi-well CFPCR platform for high throughput (HT) PCR applications. A variety of spiral CFPCR devices were designed and fabricated by UV-LIGA technique for a nickel large area mold insert (LAMI) and grooves and fins by micro-milling for a brass LAMI. Double-sided micro molding in polycarbonate (PC) with two LAMIs was done using hot embossing. The molded PC chips were sealed in a custom-designed thermal fusion bonding apparatus. Small footprint, 20- and 25-cycle CFPCR devices for a CFPCR multi-reactor chip successfully amplified 99-bp target DNA fragments from a 48k bp λ-DNA template.

Motivation and Objective

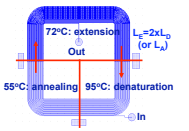
- High demand for a highly parallel, polymerase chain reaction (PCR) multi-reactor platform: exploration of the accumulated genetic information from the Human Genome Project
- Incorporation of CFPCR devices into a polymer, 96-well titer plate format (120 mm x 96 mm) for a HT CFPCR multi-reactor platform
  - CFPCR multi-reactor module
  - Heater module
  - Fluidic control module
- Small footprint, CFCPR devices for a 1st generation, double-sided CFPCR multi-reactor module chip
  - Optimization of the geometry for CFPCR devices
  - Verification of DNA amplification capability



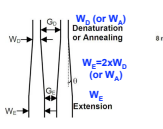
CFPCR multi-reactor platform

Design

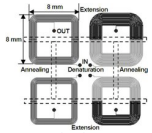
- Effective footprnit of each spiral CFPCR device: 8 mm by 8 mm
- Residence time ratio of 1:1:4 for denaturation : annealing : extension
  - Length and width of the microchannels in extension zones doubled compared to those in denaturation or annealing zone
- Various dimensions: microchannel widths of 10~40 μm, wall widths of 10~55 μm, microchannel depth of 40 μm (six types of devices) → a group of twelve CFPCR devices for 20- and 25-cycles
- Sharing of temperature zones for adjacent CFPCR devices



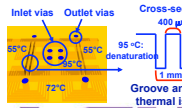
Spiral CFPCR device with three temperature zones

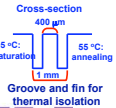


Converging/diverging section

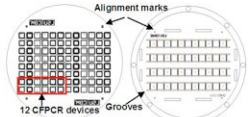


Four CFPCR devices sharing temperature zones (dotted lines: grooves)





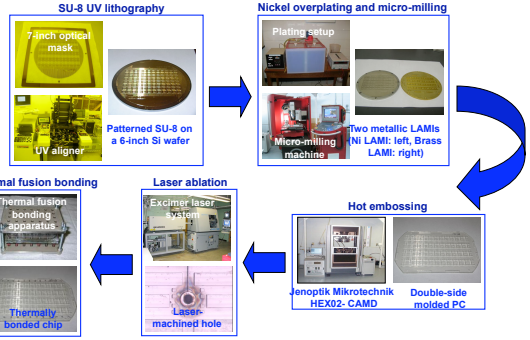
Groove and fin for thermal isolation



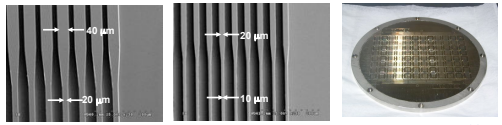
Nickel LAMI | Brass LAMI

Fabrication

- Fabrication sequence



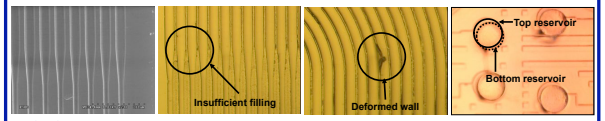
- Nickel large area mold insert (LAMI)



SEM images of microchannels of PCR in 40 μm thick Ni LAMI

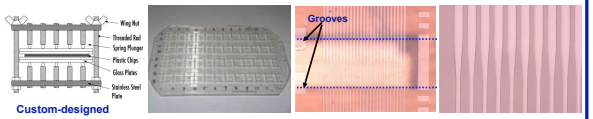
Ni LAMI mounted in SS hot embossing fixture

- Double-side hot embossing in polycarbonate (PC)



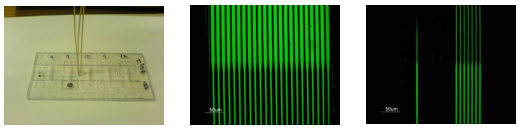
CFPCR device with 20 μm/ 40 μm wide microchannels | Insufficiently filled 10 μm wide channel walls | Deformed 10 μm wide channel walls | Four fluidic interconnects in a double-side molded chip

- Thermal fusion bonding



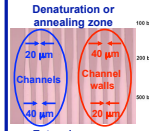
Custom-designed thermal bonding apparatus | Sealed CFPCR multi-reactor chip in PC | Sealed microchannels with grooves underneath | Close-up view of sealed microchannels

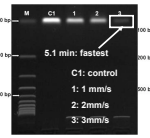
- Leakage testing with fluorescent dye


10-well CFPCR chip with capillaries | 20/40 μm wide channels with 40/20 μm wide walls | Severe air blockage in 10 μm wide channels

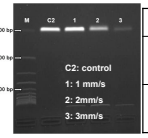
DNA Amplification

- Small footprint, 20- and 25-cycle CFPCR devices with 20 μm/40 μm wide microchannel (40 μm/20 μm wide channel walls) used
- DNA template: 48k bp λ-DNA c1857Sam7
- Primers to generate 99 bp target DNA fragments
- Thermal cycling with three copper plates, strip heaters, and TCs (94°C for denaturation, 63°C for annealing, and 72°C for extension)
- Different flow velocities: 1 mm/s, 2 mm/s and 3 mm/s corresponding to 0.048 μl/min, 0.096 μl/min, and 0.144 μl/min





Gel image for 20-cycle CFPCR devices



Gel image for 25-cycle CFPCR devices

# of cycles	Flow velocity	Reaction times
20-cycles	1 mm/s	15.4 min
	2 mm/s	7.7 min
	3 mm/s	5.1 min
25-cycles	1 mm/s	20.6 min
	2 mm/s	10.3 min
	3 mm/s	6.9 min

Reaction times for each flow velocity

Conclusions

- Optimization of the geometry for the 1st generation 96-well CFPCR multi-reactor chip throughout manufacturing processes
  - 20 μm/40 μm wide microchannel walls for structural rigidity
  - 40 μm/20 μm wide microchannels for smooth fluid control
- Successful demonstration of DNA amplification capability in small footprint CFPCR devices
  - Reaction times as fast as 5.1 min for 20-cycle CFPCR devices at 3 mm/s
- Development of a heater module and a fluidic control module for complete realization of the high throughput CFPCR platform under way

Connections with CyberTools

- Simulation challenges for micro-scale devices in large area format (120 mm x 96 mm)
- Device simulation
  - Temperature distribution over the whole CFPCR multi-reactor module
  - Fluid flow and heat transfer for individual CFPCR devices
  - Tracking plugs through multi-well devices
- Process simulation
  - Filling behavior analysis in molding process
  - Failure analysis in de-molding: thermal stress and local deformation

References

[1] M. Hashimoto, P.-C. Chen, M. W. Mitchell, D. E. Nikitopoulos, S. A. Soper, and M. C. Murphy, 2004, Lab Chip, pp. 638-645

[2] D. S. Park, P.-C. Chen, B. H. You, N. Kim, T. Park, T. Y. Lee, P. Datta, Y. Desta, S. A. Soper, D. E. Nikitopoulos, and M. C. Murphy, 2008, Hilton Head Workshop 2008, pp. 114-117

Acknowledgements

This work was supported by the National Science Foundation under Grant Number EPS-0346411 and the State of Louisiana Board of Regents Support Fund, Industrial Ties Program through Grant Number LEQSF (2005-08)-RD-B-04. The authors thank the Center for Advanced Microstructures and Devices (CAMD) at Louisiana State University for the microfabrication support.

Design Optimization and Realization of an Electrophoretron Cycler

Céline Ramet[1], Dimitris Nikitopoulos[1], Michael Murphy[1], Steve Soper[2]

[1]Department of Mechanical Engineering, Louisiana State University, Baton Rouge

[2]Department of Chemistry, Louisiana State University, Baton Rouge

Abstract

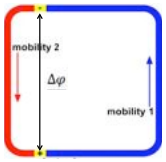
The aim of this work is to design and realize a continuous flow micro-fluidic device for PCR (Polymerase Chain Reaction) or LDR (Ligase Detection Reaction), two cyclic reactions needed in DNA analysis. The device, called an Electrophoretron, combines electroosmotic and electrophoretic effects to induce cyclic motion of buffer, DNA and other reactants in a single-loop micro-channel, with only one constant difference of potential applied.

It should be noted that the following study is conducted assuming PCR conditions (buffer: PCR buffer at pH 8.3, species: DNA) in a Polycarbonate device; however, other applications are possible (e.g. mixing).

Principle[1]

Channel 2:

- Chemical reversal of the EOF
- EOF and Electrophoresis complementary



Channel 1:

- No EOF treatment
- EOF and Electrophoresis adverse

Figure 1: Scheme of an Electrophoretron

Theoretical Analysis

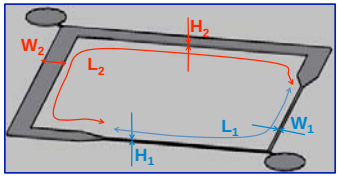


Figure 2: Example of a 3D Electrophoretron Design

Assumptions

- Electrical Debye Layer infinitely thin
- Electrodes effects negligible
- Bends effects negligible
- DNA Diffusion negligible
- Geometry transition effects negligible

Solving Electrical potential and Flow gives velocity profiles in channel i:

$$w_i(x,y) = \frac{\Delta\varphi}{L_i}\left(\mu_{ef,i} - \frac{\pi^2}{4}F_i(x,y)\frac{\frac{\mu_{eo1}H_1W_1}{L_1} + \frac{\mu_{eo2}H_2W_2}{L_2}}{\frac{H_1W_1}{L_1}g_1 + \frac{H_2W_2}{L_2}g_2}\right)$$

$F_i(x,y)$ and g_i are quickly converging series depending on channel i geometry
$\mu_{eo,i}$: electroosmotic mobility of channel i
$\mu_{ef,i}$: effective mobility of channel i (sum of electroosmotic and electrophoretic)

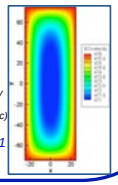
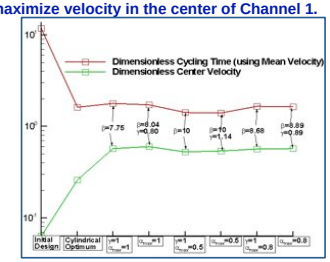


Figure 3: Example of Electroosmotic Velocity Profile in Channel 1 (Configuration from[2])

Optimization

Using theoretical analysis results and Matlab Optimization toolbox, we investigate design variations (α: L_1/L_2, β: W_2/W_1, γ: H_2/H_1) to maximize velocity in the center of Channel 1.



Note:
Function used detects only local maxima => Initial conditions and limits have a huge impact on the optimized solution.

Figure 4: Comparison between different Optimum Configurations

Simulations

Simulations allow to take into account: bends effects and electrodes influence, as well as DNA diffusion. Designs were realized with AutoCAD 2008, and mesh and running were done using Coventorware 2006 (Solver: Fluent).
These confirmed cycling (Figure 6) as well as validated theoretical velocity profiles shape (Figures 7 & 8)

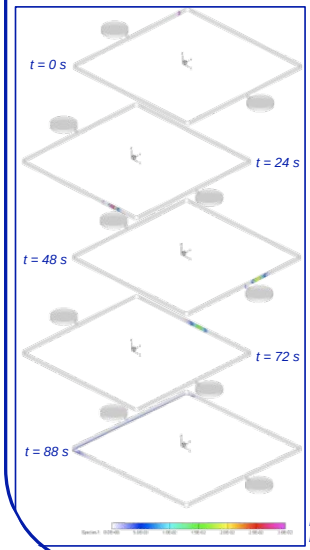


Figure 6: Simulated Motion of DNA for 24 mm long COC Electrophoretron α=2.79, β=γ=1

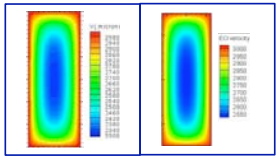


Figure 7: Comparison between Velocity Profiles found via Simulation[2] and Computation of Theoretical Calculations- Velocities are in μm/s- α=1.97, β=γ=1

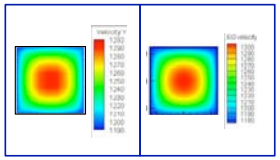


Figure 8: Comparison between Velocity Profiles found via Simulation & Computation of Theoretical Calculations -Velocities are in μm/s- α=0.5, β=10, γ=1

Main Practical Limitations

- 96 well format: footprint 8x8 mm

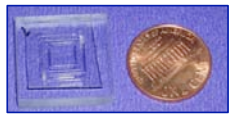
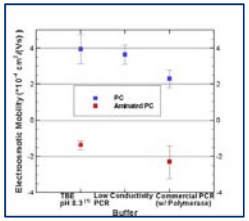


Figure 9: Initial design chip hot embossed in Polycarbonate (PC)

- EOF: large uncertainty & protein (e.g. Polymerase for PCR) influence

Figure 10: Comparison of Different Measurements of Electroosmotic Mobilities inside PC microchannels



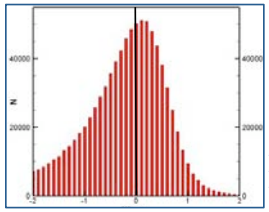


Figure 11: Monte-Carlo Simulations: Mean Velocity Repartition in Channel 1 due to Electroosmotic Mobility Variations and Geometry Variations-α=2.88 β=γ=1

- Hydrolysis: out electrodes
- Visualization of EOF only challenging (absolutely neutral dye and/or particles needed)

Connections with CyberTools

Current simulations have been done with commercial software on individual workstations. Obtaining simulation results to support an optimization study of a complex, multi-physics device/process which involves a large number of parameters is untenable through present means. CyberTools (e.g. WP1, WP3, WP4) enables the use of High Performance Computing for the solution of the governing equations and their coupling with optimization algorithms on a user-friendly platform (e.g. CACTUS). This results in faster and more efficient design of devices/systems for the science-driver application (geno-sensor), as well as better understanding of the related physics through visualization.

References

[1] Choi et al., Journal of Chromatography A, 924, 53-58 (2001)
[2] Elmajdoub, LSU thesis (2006)

Acknowledgements

All CBM² members, especially Murphy's and Soper's groups, Jason Guy and Proyag Datta (Microfab.).
Fundings: NIH-BRP, NSF-EPSCoR

Real-time Information Services for Scientific Applications

Katerina Stamou[1,2], Gabrielle Allen[1,2], Erik Schnetter[1,3]

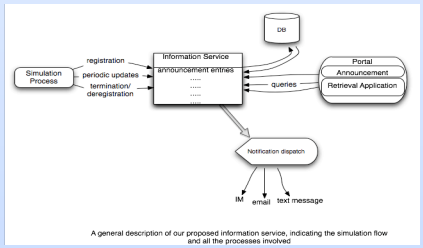
[1]Center for Computation and Technology LSU
[2]Department of Computer Science LSU
[3]Department of Physics and Astronomy LSU

Abstract

Distributed scientific communities can immensely benefit from having a central infrastructure for keeping important information, given the overall collaborative nature of their running simulations, that are usually conducted in different local sites. Such data needs to be structured in a way that can be described and queried with precision and speed, for later retrievals.

Core Scenario/Motivation

A scientist who would like to run her simulation, submits her task to run to a cluster of machines. When the requested resources become available, the task is selected and executed. During initialization phase, the simulation registers with an application information service, and dispatches basic execution details. As information arrives a notification service informs collaborating scientists. While the simulation runs, it periodically updates the current application status and information.



A general description of our proposed information service, indicating the simulation flow and all the processes involved

Drawing from this use case, we were motivated into investigating the following:

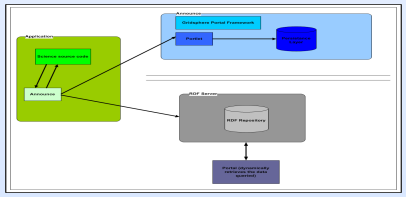
•Measure the performance, functionality and usability of both Announce and Formaline Cactus thorns by running actual tests on several LONI clusters

•Integrate both technologies within LONI portal through a smooth migration and interoperation process

•Extend the SAGA Advert Service, by incorporating it into the suggested information system

Announce & Formaline in Cactus

Cactus provides a vast array of application-specific scientific functions, through extension libraries, called 'thorns', while taking advantage of modern large scale, parallel computing resources.



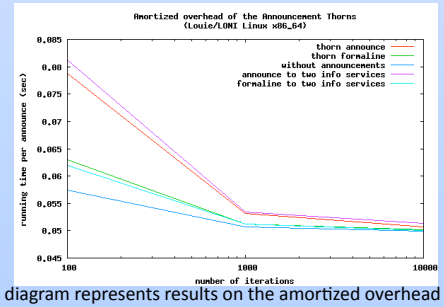
The ``Announce'' thorn was developed to automatically communicate general information about Cactus simulations to a service viewable from a portal.

A different Cactus thorn, ``Formaline'' similarly preserves important results and metadata about simulations by announcing them to an information service where they can be kept for prolonged periods of time, for later analysis.

We are conducting tests, in order to measure and compare the efficiency of both the "Announce" and "Formaline" systems:

- net impact of each thorn on the overall simulation time
- amortized overhead of the systems
- user interface responsiveness

* backend store database scaling, under increasing amounts of announcement entries

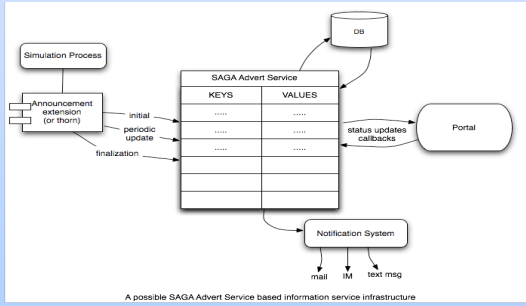


The above diagram represents results on the amortized overhead time of the announcement thorns, in relation to the number of iterations and the simulation time

SAGA-Advert Service

A full-fledged information service, built on top of the SAGA-Advert service, would require:

- From a client-side perspective, an extension library for the simulation application (or a thorn, in the case of Cactus), that would communicate the simulation metadata to the advert service, using the SAGA library interface.
- An information service, build around the Advert Service, which would be able provide means for publishing and retrieving metadata, as well as providing notification services.
- An application for visualizing the stored metadata, giving convenient access through an interface for querying and viewing real-time as well as historic archived simulation information. Preferably, this could be implemented as an web portlet on top of a portal platform, like gridsphere.



A possible SAGA Advert Service based information service infrastructure

Connections with CyberTools

Through this effort, we build on, and extend existing well-established architectures and packages, i.e. the Cactus toolkit, and the SAGA library. This work will lead to a general information service and announcement mechanism which can be easily incorporated into any CyberTools simulation code.

Acknowledgements

We thank Ian Kelley and Thomas Radke, co-authors of Announce and Formaline for their cooperation and assistance in this work. Also, Hartmut Kaiser, SAGA lead architect, for providing us with useful advice on SAGA advert service.

Cybertools is supported by NSF/ EPSCoR Award No. EPS-0701491

Distributed Data Management in CyberTools

Ibrahim H Suslu[1], Xinqi Wang[1], Ismail Akturk[1], Tevfik Kosar[1]

[1]Louisiana State University, Center for Computation and Technology

Abstract

CyberTools will provide services such as information processing, data management, storage, and co-scheduling for the science projects in LONI environment. Data management services help manage large amount of experimental and observational data. The larger data require the better data management tools need to be developed. User friendly and intelligent data management tools will decide what type of remote data access technique to use either remote I/O or staging, client tools can access remote data using three different interfaces: petashell, petafs, and pcommands, and the ontology based metadata search gives logical filenames that match the semantic search criteria, and then, each logical file name has corresponding set of physical file names, for the searched data entity.

Remote I/O and data staging are the most widely used data access methods for large scale distributed applications with non-local data sources. We are developing such a model for the CyberTools data management clients which will choose the most appropriate data access method for applications. We define the parameters that potentially affect the end-to-end performance of the distributed applications which need to use remote and distributed data.

Extendable metadata management is essential in large-scale distributed data management, traditional metadata management is not sufficient to provide interoperability for large-scale, physically and semantically heterogeneous dataset. We present a semantically enabled metadata management framework based on ontology. We seek to address two main issues: data integration for semantically and physically heterogeneous distributed knowledge stores, and semantic reasoning for data verification and inference in such a setting. Our metadata management aims to enable data interoperability among otherwise semantically incompatible data sources, cross-domain query capabilities and multi-source knowledge extraction.

Modeling to Access Remote Data

Preliminary Model

Parameters:

Staging

$$T_s = T_{in} + E + T_{out}$$

Where

$$T_{in} = R_{r_{in}} + N_{s_{in}} + W_{l_{in}} + R_{l_{in}}$$

$$T_{out} = W_{l_{out}} + R_{l_{out}} + N_{s_{out}} + W_{r_{out}}$$

Remote I/O

$$T_r = R_{r_{in}} + N_{r_{in}} + E + N_{r_{out}} + W_{r_{out}}$$

For Remote I/O to be more efficient than staging:

$$R_{r_{in}} + N_{r_{in}} < R_{r_{in}} + N_{s_{in}} + W_{l_{in}} + R_{l_{in}}$$

$$\Rightarrow N_{r_{in}} - N_{s_{in}} < W_{l_{in}} + R_{l_{in}}$$

and

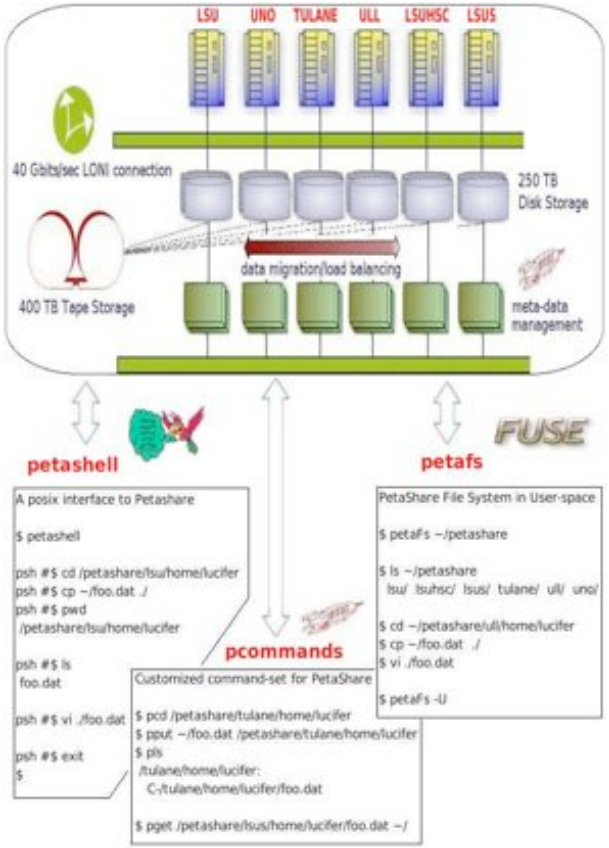
$$N_{r_{out}} + W_{r_{out}} < W_{l_{out}} + R_{l_{out}} + N_{s_{out}} + W_{r_{out}}$$

$$\Rightarrow N_{r_{out}} - N_{s_{out}} < W_{l_{out}} + R_{l_{out}}$$

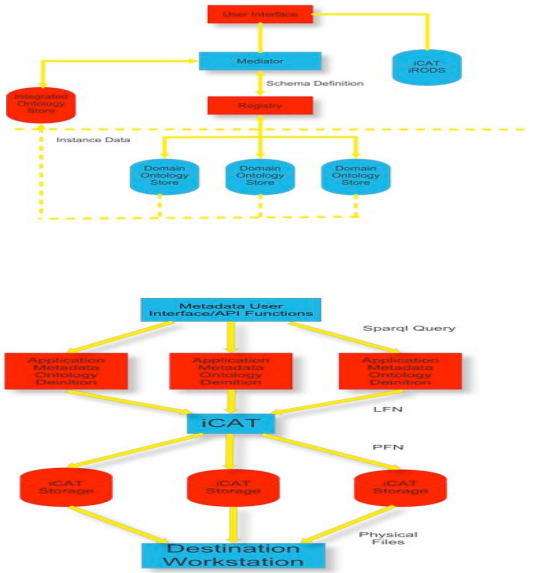
If remote I/O library performs good in data transfer over network, or local disk performance is slow, remote I/O might be advantageous over staging, otherwise, staging method would perform better.

Petashare

CyerTools's distributed data management infrastructure PetaShare provides scientists with access to data widely distributed across multiple geographically far-apart institutions. So far, three PetaShare clients have been developed to provide three distinct access modes to underlying PetaShare infrastructure: PetaFs allows PetaShare infrastructure to be mounted as a folder on any Linux machines; PetaShell provides a regular shell environment for access to PetaShare, user can use regular Unix commands to access PetaShare under both PetaFs and PetaShell; Pcommand, on the other hand, provides an set of PetaShare enabled commands user can use to directly access PetaShare.



Semanticly Expanded Data Access



Connections with CyberTools

- ✓**Among CyberTools related projects, Petashare will seek to provide an iRODS based distributed data management infrastructure in which data can be more easily located, understood and retrieved.**
- ✓**We seek to address the problem of lack of integration of data produced by different scientific domains through semantic enabled metadata management.**
- ✓**Petshare user client interfaces (petaFs, petaShell, pCommands) allow transparent data management, so that the CyberTools scientists can focus on their own research and the content of the data rather than how to manage it.**
- ✓**Staging and Remote I/O model can be applied to most data intensive distributed cyberTools applications to decide the best data access model for those applications.**

Acknowledgements

This project is in part sponsored by the National Science Foundation under award numbers CNS-0619843 (PetaShare) and EPS-0701491 (CyberTools), and by the Board of Regents, State of Louisiana, under Contract Number NSF/LEQSF (2007-10)-CyberRII-01.

LIGO Outreach Tangibles: Integration of Tangible Interaction and Visual Computing as Gateways to Science

Cornelius Toole, Jr. , Zachary Dever, Alvin Wallace, Jr., and Brygg Ullmer
Louisiana State University
Department of Computer Science and
Center for Computation and Technology

LIGO Outreach Tangibles

The LIGO Outreach Tangible Kiosk is a platform that combines visual computing, tangible interaction, and visual & physical design efforts to deliver engaging educational content on science topics related to the Laser Inferometer Gravitational Observatory project. Longer term, we also seek to provide a path for accessing and engaging with high-end computational resources used for scientific inquiry or their byproducts. Here we describe work in progress in the development of this kiosk platform and a game activity based upon a key LIGO outreach activity. We also describe future iterations for this platform.

Motivation

- To stimulate interests and reinforce science instruction in formal and informal contexts
- To help fill gap, in costs and flexibility, between two types of successful LIGO outreach activities: Exploratorium-developed exhibits($10K-50K in costs) and "science snacks"($0-50)
- To address tech literacy/usability gaps by using tangible interaction to wield chains of complex, digital actions through simple physical interactions

Connections to Cybertools

- When deployed in places with network connectivity, can be used to provide access to key cyber-infrastructure
- As we extend this tangible interaction kiosk platform to other applications, users will be able to access scientific data repositories, visualization services

Acknowledgments

- This project was supported by the Louisiana Board of Regents, contract #33027 (2008-2009) and the LSU Huel D. Perkins Doctoral Fellowship.
- We would like to thank John Douthat, Ian Wesley-Smith, Srikanth Jandhyala, Rajesh Sakaran and Kexi Liu

Children driving hurricane visualization(left) and handheld microscope(right) with tangibles

Images from the 1977 Ray and Charles Eames film, “Powers of Ten”

Touchscreen tablet + Powermate® knob enclosure(right) and three-wheeled parameter tray(left)

Design sketches of LIGO Outreach Kiosk for museum settings(left) and for classroom settings(right)

LIGO Kiosk Content and Activities

- A wide range of applications can be delivered through the tangible kiosk platform
- For initial phase of project we chose to develop a game centered around a film, which is pre-screened by most visitors to the LIGO Science Education Center
- We plan to develop information visualization content based upon galaxy catalog data

Powers of Ten Game

- Based upon a Ray and Charles Eames 1977 film that depicts the relative scale of the universe
- Our game pits player against each other as they try to correctly match orders of magnitude with images

Galaxy Catalog Visualization

- Corso et al provide a tool that analyzes LIGO run data along with Compact Binary Coalescence Galaxy catalog data to visualize the sensitivity of LIGO thus showing which galaxies can be observed

Future Work

- Limited deployment at science education centers and middle schools
- More content development with aid of scientist consultant
- Finalize evaluation plan with educational consultant
- Integration of novel tangible interaction devices
- Design a high level communication framework capable of supporting both local and remote interaction for easier integration of other applications
- Longer term development of tangibles kiosks for access to high performance computing applications such as large data visualization, remote collaborative visualization and computational steering.

References

1. Corso, B., Benger, W., and Gonzalez, G. Visual Representation of Inspiral Group Galaxy List, Technical Report, 2008.
2. Sankaran, R., Ullmer, B., Jandhyala, S., Kallakuri, K., Sun, S. and Laan, C. Blades and Tiles: An Extensible Hardware Architectural Approach for Ubiquitous Interaction Devices. In Proc. of Ubicomp'07, 2007.
3. Ullmer B., Sakaran, R., Jandhyala, S., Tregre, B., Toole, C., Kallakuri, K., Laan, C., Hess, M. Harhad, F., Wiggins, U., and Sun, S. Tangible Menus and Interaction Trays: Core Tangibles for Common Physical/Digital Activities, In Proc. of TEI'08., 2008

Predicting Optimal Level of Parallelism in Wide Area Data Transfers

Esma Yildirim[1], Dengpan Yin[2], Tevfik Kosar[3]

[1,2,3]Center for Computation and Technology,
[1,2,3]Louisiana State University

Abstract

Using multiple parallel streams for wide area data transfers may yield much better performance than using a single stream, but overwhelming the network by opening too many streams may have an inverse effect. The congestion created by excess number of streams may cause a drop down in the throughput achieved . Hence, it is important to decide on the optimal number of streams without congesting the network. Predicting this 'magic' number is not straightforward, since it depends on many parameters specific to each individual transfer. Generic models that try to predict this number either rely too much on historical information or fail to achieve accurate predictions.

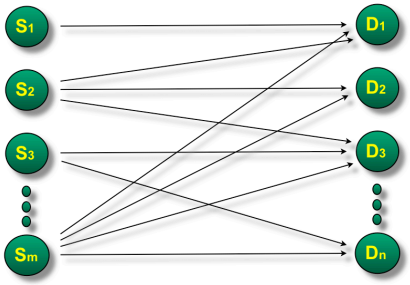
We present a set of new mathematical models which aim to approximate the optimal number of streams to open with least historical information and lowest prediction overhead. An algorithm is also introduced to select the best combination of historical information to do the prediction. We measure the feasibility and accuracy of the proposed prediction models by comparing to actual GridFTP parallel data transfers.

We are able to predict the throughput of any parallelism level accurately by using only little historical information. The decision of the correct parallel stream number will provide us with the optimal data transfer rate without congesting the network. Current data schedulers (e.g. Stork) could use these insights to optimize multiple data transfers and do intelligent decisions without requiring a large amount of historical information about bulk data transfer characteristics.

Data Scheduling Problem

Transfer *k* files between *m* sources and *n* destinations

- Ordering requests (considering priority, file size, etc.)
- Throttling - deciding number of concurrent transfers (considering available target storage space, network capacity, etc.)
- Tuning protocol transfer parameters (considering current network condition)



Models

Mathis Throughput Equation

$$Th \le \frac{MSS}{RTT}\frac{c}{\sqrt{p}} \quad \text{for } n \text{ streams} \quad Th_n \le n\frac{MSS}{RTT_n}\frac{c}{\sqrt{p_n}}$$

Th = Throughput
MSS = Maximum Segment Size
c = Constant
p = Packet Loss Rate

Model Th with a partial order-c equations → **Approach 1**

Model Th with a full second order equation → **Approach 2**

Approach 1

$$p'_n = p_n\frac{RTT_n^2}{c^2 MSS^2} = a'n^{c'} + b'$$

$$Th_n = \frac{n}{\sqrt{p'_n}} = \frac{n}{\sqrt{a'n^{c'} + b'}}$$

Approach 2

$$p'_n = p_n\frac{RTT_n^2}{c^2 MSS^2} = a'n^2 + b'n + c'$$

$$Th_n = \frac{n}{\sqrt{p'_n}} = \frac{n}{\sqrt{a'n^2 + b'n + c'}}$$

We only need 3 throughput measurements of different parallelism levels to solve the above equations: Th_{n1}, Th_{n2} and Th_{n3}

$$\frac{n_3^{c'} - n_1^{c'}}{n_2^{c'} - n_1^{c'}} = \frac{\frac{n_3^2}{Th_{n_3}^2} - \frac{n_1^2}{Th_{n_1}^2}}{\frac{n_2^2}{Th_{n_2}^2} - \frac{n_1^2}{Th_{n_1}^2}}$$

$$a' = \frac{\frac{n_2^2}{Th_{n_2}^2} - \frac{n_1^2}{Th_{n_1}^2}}{n_2^{c'} - n_1^{c'}}$$

$$b' = \frac{n_1^2}{Th_{n_1}^2} - a'n_1^{c'}$$

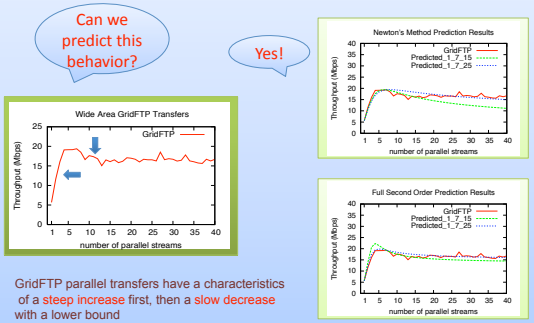
To be able to solve ***c'*** we apply an approximation method called Newton's Method

$$a' = \frac{\frac{n_3^2}{Th_{n_3}^2} - \frac{n_1^2}{Th_{n_1}^2} - \frac{n_2^2}{Th_{n_2}^2} - \frac{n_1^2}{Th_{n_1}^2}}{n_3 - n_1 \quad n_2 - n_1}$$

$$b' = \frac{\frac{n_2^2}{Th_{n_2}^2} - \frac{n_1^2}{Th_{n_1}^2}}{n_2 - n_1} - (n_1 + n_2)a'$$

$$c' = \frac{n_1^2}{Th_{n_1}^2} - a'n_1^2 - b'n_1$$

Application



GridFTP parallel transfers have a characteristics of a steep increase first, then a slow decrease with a lower bound

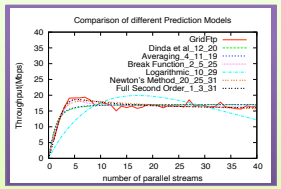
Best Parallelism Data

Algorithm
Input: *m* throughput values of different parallelism levels

For i=1 to *m*-2
 For j=i to *m*-1
  For k=j to *m*
  - Calculate *a', b'* and *c'*
  - if the coefficients are within certain boundaries
  - Calculate *Error Rate*

$$Err_{i,j,k} = \sqrt{\sum (Th_{actual} - Th_{predicted})^2}$$

Find *i,j,k* that gives minimum *Err*



Connections with CyberTools

The Scheduling and Data Services (Work Package 1) work package of CyberTools aims to support a distributed data archival for all LONI projects and management of reliable and efficient data retrieval for Science Drivers who would like to store and access their data. PetaShare will handle all the low-level data handling issues such as data-aware storage systems and schedulers which support application areas that include coastal and environmental modeling, geospatial analysis, bioinformatics, medical imaging, fluid dynamics, petroleum engineering , numerical relativity and high-energy physics. The Stork data scheduler will further be developed to allow on-demand queuing, scheduling and optimization of data placement jobs. With the optimization of parallel stream number, the maximum throughput can be gained for data placement jobs without congesting the network and in this project we find a methodology to decide optimal stream number via mathematical models.

Acknowledgements

This project is in part sponsored by the National Science Foundation under award numbers CNS-0619843 (PetaShare) and EPS-0701491 (CyberTools), and by the Board of Regents, State of Louisiana, under Contract Number NSF/LEQSF (2007-10)-CyberRII-01. Thank You!

ASSEMBLY TOLERANCE ANALYSIS FOR INJECTION MOLDED MODULAR, POLYMER MICROFLUIDIC DEVICES

Byoung Hee You[1, 3], Daniel S. Park[3], P.-C. Chen[1, 3], Sudheer D. Rani[1, 3], Dimitris E. Nikitopoulos[1, 3], Steven A. Soper[2, 3], and Michael C. Murphy[1, 3]

[1]Deparement of Mechanical Engineering, [2]Department of Chemistry, [3]Center for Bio-Modular Multi-Scale Systems
Louisiana State University, Baton Rouge, LA 70803, U.S.

Validation of an assembly tolerance analysis for the assembly of modular, polymer microfluidic devices was performed using simulation and experiments. A set of three v-groove and hemisphere-tipped post joints was adopted as a model assembly. Monte Carlo methods were applied to the assembly function to simulate the assembly. The estimated mismatches were 109±13 μm and 20±14 μm along X- and Y-axes, respectively. The estimated vertical gap between the modules at the alignment standards along the X- and Y-axes 291±33 μm and 291±34 μm, compared to the designed value of 300μm. The measured lateral mismatches were 103±6 μm and 16±4 μm along X- and Y-axes, respectively. The vertical gaps measured for the assemblies were 316±4 μm and 296±9 μm at the X- and Y-axes. The models have significant potential for enabling the realization of cost-effective mass production of modular instruments.

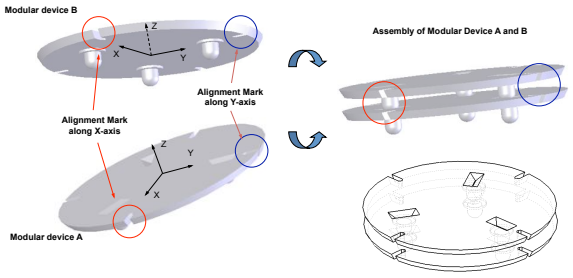
Motivation

- Modular, polymer microdevices for biochemical analysis
- Assembly technologies enable integration of modules without optical alignment

Objective

- Development of micro-assembly technology for cost- effective mass production of modular, polymer microdevices

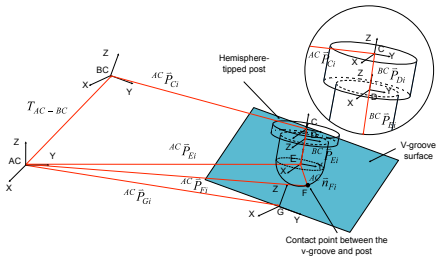
Design of Assembly Scheme



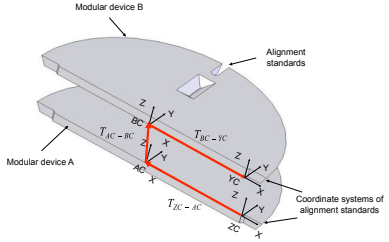
A set of three hemisphere-tipped posts and v-grooves was developed using kinmatic design[1]. The assembly features can prevent under-constraint and over-constraint in assembly so that precise, inexpensive assembly, enabling reliable microfluidic interconnections, can be achieved.

[1] You, Byoung Hee, Chen, Pin-Chen, Guy, Jason, Datta, Proyag, Nikitopoulos, Dimitris E., Soper, Steven A., and Murphy, Michael C., 2006, "Passive alignment structures in modular, polymer microfluidic devices," *Proceedings of ASME International Mechanical Engineering Congress and Exposition, Chicago, 5-10, November, IMECE2006-16100.*

Modeling of Assembly Function

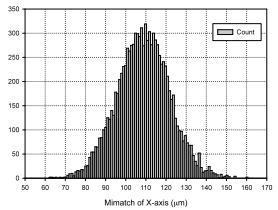


A kinematic chain between a hemisphere-tipped post and v-groove in the assembled system

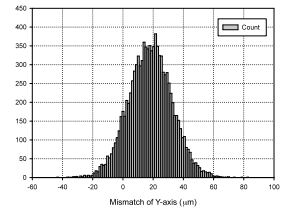


A kinematic chain between the alignment standards of the modules to estimate the mismatch of assembly.

Monte Carlo Simulations

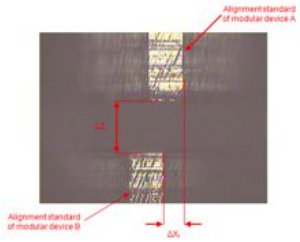


Mismatch of the assembly along the X-axis (mean = 109 μm and standard deviation = 13 μm)



Mismatch of the assembly along the Y-axis (mean = 20 μm and standard deviation = 14 μm)

Experiments



A micro photograph of an alignment standard on the X-axis

Measured mismatches along the X- and Y-axes

X-axis	Gap at X-axis	Y-axis	Gap at Y-axis
103±6 μm	316±4 μm	16±4 μm	296±9 μm

Conclusions

The modular devices were assembled. The simulation and experimental results showed accordance with each other. The developed assembly and tolerance analysis is applicable to the design of cost-effective mass production of modular, polymer microfluidic devices.

Connections with CyberTools

Assembly tolerance analysis using Monte Carlo methods can predict the accuracy of assemblies in virtual space using computation. It requires ten thousand or more assemblies virtually generated for the simulation of mass production of modular, polymer microfluidic devices. Efficient computation is necessary for accurate prediction. More complex models are needed for larger assemblies

Acknowledgments

National Science Foundation (EPS-0346411)
National Institutes of Health (NIH R24-EB-002115-03)
State of Louisiana Board of Regents

Evaluation of Microsensor and Micro-Mixer for Biosensor Applications

Senaka Kanakamedala[1], Mangilal Agarwal[1], Ji Fang[1], Yuri Lvov[1], and Mark A. DeCoster[1,2]
[1]Institute for Micromanufacturing, [2]Biomedical Engineering, Louisiana Tech University, Ruston LA, 71270

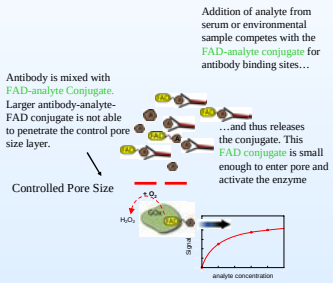
Abstract

Most currently available immunosensors are designed to detect high molecular weight molecules (primarily proteins and infectious agents); where as low molecular weight chemical agents still remain a challenge to detect. In this project, we proposed to develop a miniaturized immunosensor platform with the versatility to simultaneously detect a large number of low molecular weight agents, including environmental contaminants, serum constituents, and chemical warfare agents. We proposed the development of a micro-mixer for handling fluids and electrochemical detection system for the new sensor format. Initial experiments were performed using electrode made of poly(3,4-ethylenedioxythiophene) poly(styrenesulfonate) (PEDOT-PSS) conducting polymer and carbon nanotubes (CNTs). Different techniques such as electrochemical polymerization, chemical and electrochemical deposition and spin coating were applied for the fabrication of such carbon nanotubes or polymer-based electrodes. A novel omega shaped micro-mixer was fabricated to enhance the mixing of antibody with analyte. The fluid flow and mixing action in microchannels were observed by injecting two test fluids of different colors. From both simulation and experimental results, a laminar flow of specified fluid was observed in the devices with straight channels, whereas a turbulent type of flow was observed in devices with omega channels.

Introduction and Background Work

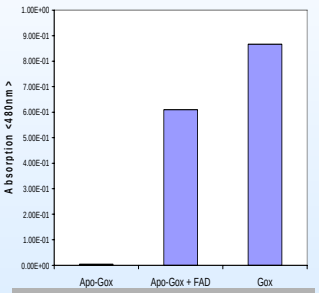
The enzyme glucose oxidase (GOx) is the key component of many commercially successful biosensors. In order to work as a catalyst, glucose requires a cofactor, Flavin Adenine Dinucleotide (FAD). Figure 1(b) shows the absorption characteristics of apo-Gox, apo-Gox + FAD and GOx in the presence of glucose. Apo-GOx did not give much response indicating the complete removal of FAD from GOx.

Figure 1(a)



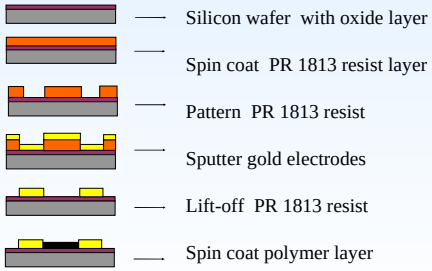
E-Chemical Immunoassay

Figure 1(b)



Absorption Characteristics of Apo-GOx, Apo-GOx + FAD and GOx in the presence of glucose in UV-vis Spectroscopy.

Immunosensor Fabrication

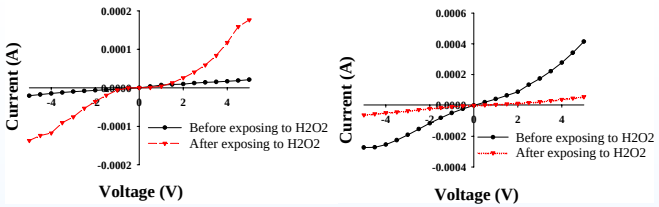


Sensor chip with gold electrode pads and active polymer region for sensing application.

The sensor chips fabricated from the process described above has two sensors and is 4 x 4 mm (length x width) in dimensions.

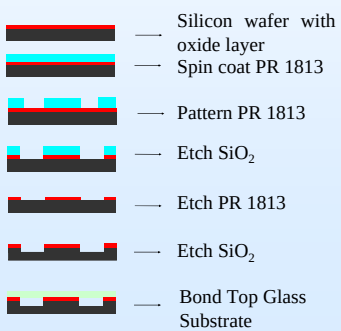
Immunosensor Results

The electrical characteristics of the PEDOT-PSS and carbon nanotube sensor devices were investigated as a function of time and found that the devices are tending to stabilize after 3-5 days.

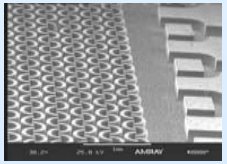


Effect of H_2O_2 on the electrical characteristics of (a) PEDOT-PSS (b) Carbon nanotube film .

Micro-Mixer Fabrication



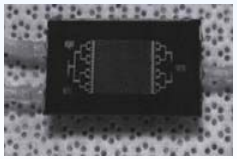
Fabrication
- Lithography
- ICP Etching
- Bonding



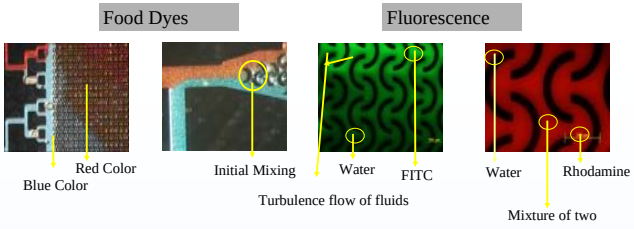
SEM → Omega Channel Micromixer

Micro-Mixer Testing Results

The fluid flow and mixing in microchannels were observed by injecting two sets of different fluorescent dyes along with water respectively (FITC (green) and water, RITC (red) and water).



In the first design the flow is laminar until the fluids reached omega channel and mixing was observed only in middle region of the channel. But in the current omega channel design mixing was observed even in the region that is near to the inlet section.



Summary and Conclusion

- Omega channel design has benefits of better mixing over straight channel.
- PEDOT and Carbon nanotubes may be combined with micro-mixer and immunoassay for the integrated sensor.
- Simulation and modeling may lead to better design of sensor system.

Integration With Cyber Tools

Dec 13, 2007 → Immunosensor Kickoff Meeting.
April 28, 2008 → Visit Tulane Group (Dr. Blake).
May 12, 2008 → Team and Project Leader Meeting.
May 20, 2008 → Coordinating with CFD/MD Team (Dr. Gaver) (Tulane).
May 30, 2008 → All Hand Meeting (BoR, EPSCoR)- Modeling Discussions.
July 23, 2008 → Video Conference with Tulane Group- Simulation Discussions.

Acknowledgments

Acknowledgments are due to the Institute for Micromanufacturing for providing the research facility and to NSF EPSCoR research Infrastructure Improvement (RII) Award.

Coupling Antibody Binding to Enzyme Activation in a Miniaturized Immunosensor

Mehnaaz F. Ali[1], Robert C. Blake II[2], Thomas C. Bishop[3], Amit S. Jain[3,4], Henry S. Ashbaugh[4], Steven W. Rick[5] and Diane A. Blake[1]

[1]Department of Biochemistry, Tulane Univ. Hlth. Sci. Ctr. New Orleans LA, 70112

[2]Division of Basic Pharmaceutical Sciences , Xavier University of Louisiana, New Orleans, LA 70125

[3]Department of Mathematics and [4]Chemical/ Biomolecular Engineering, Tulane University, New Orleans LA, 70118

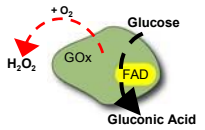
[5]Department of Chemistry, University of New Orleans, New Orleans LA, 70148

Abstract

The scope of this work is to develop antibody-based sensors for the detection of low molecular weight elements associated with environmental pollution, serum constituents and chemical warfare agents. Specifically, antibody binding will be coupled to the activation of the enzyme glucose oxidase upon the positive detection of the relevant model analyte. An important aspect of the immunosensor is the ability to modulate the activity of the glucose oxidase with the removal and addition of its co-factor flavin adenine dinucleotide (FAD). This aspect of the enzyme can be efficiently harnessed to provide an electrochemical signal transduction system. In order to facilitate the miniaturization and thus efficacy of the proposed 'hand-held' device, it is advantageous to utilize this electrochemical signaling modality in combination with antibody binding events. Another important component of the overall goals of this project involves the close collaborations with theoretical and computational methods groups. This work will facilitate appropriate sensor design and determine physical and chemical limitations to the final project.

Immunosensor will use GOx Catalyzed glucose oxidation for Signal Transduction

Glucose Oxidation



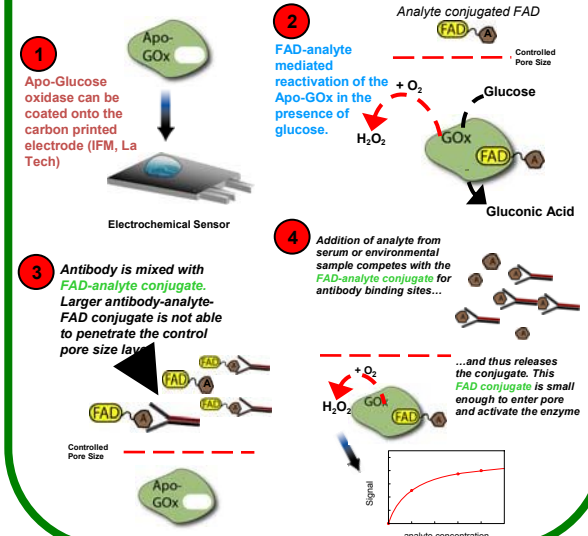
***Glucose oxidase** requires the cofactor **FAD** for the catalysis of glucose to gluconic acid. This process involves the initial reduction of FAD to $FADH_2$ and consequent oxidation by molecular O_2 generating H_2O_2*

E-Chem Sensor

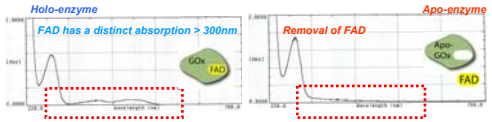
Enzyme activity** can be **modulated** by the **removal** and **introduction** of the cofactor **FAD**. The cofactor can be efficiently **dissociated** under acidic conditions to yield **Apo-glucose oxidase.

Thus the cofactor FAD can be conjugated to a model analyte and utilized to modulate enzyme activity.

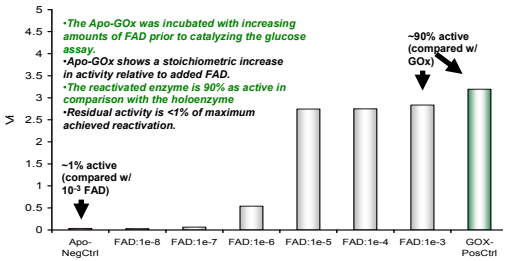
General Strategy for E-chem Immunoassay



Absorption Spectra of FAD Region

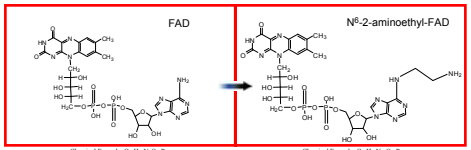


Reactivation of Apo-GOx



Synthesis of N^6-2-aminoethyl-FAD

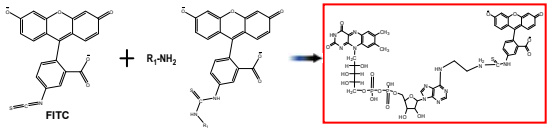
via Dimroth Arrangement



Externally oriented adenine moiety of the FAD, allows introduction of a reactive primary amine group.

Alkylation of N-1 position of the adenine ring with ethyleneimine, followed by Dimroth rearrangement to N-6 under mild aqueous conditions.

FITC-FAD conjugate via Thiourea Linkage



Antibody – Analyte Selection

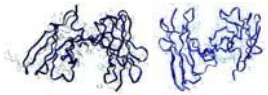
Clone Number	Ligand	K_d (M)	Reference
12F6	2,9-dicarboxyl-1,10-phenanthroline (DCP)	7.5×10^{-7}	Blake et al. (2004) *Bioconj. Chem.* **15**:1125.
12F6	UO_2^{2+}-DCP	9.1×10^{-10}	Blake et al. (2004) *Bioconj. Chem.* **15**:1125.
4B33	EDTA	1.25×10^{-9}	Unpublished data
4B33	Cu^{2+}-EDTA	2.2×10^{-9}	Unpublished data

Molecular Dynamics of Antibody Binding Regions

will provide insight into antibody performance and aid in optimization of immunosensor

Simulations of antibody complementarity determining regions

1)Homology modelling based on antibody sequence (Blake, Bishop, Jain)

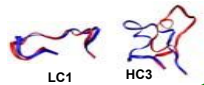


Different orientations of the antibody 5B2 (both light and heavy chains) modeled on antibody 1NGP

2) *In vacuo*, implicit and explicit solvent simulations of antibody 5B2 LC and HC loops confirm previous identification of metal binding residue Lys[58] (Blake, Jain, Rick and Ashbaugh)

- Replica Exchange Molecular Dynamic performed of antibody 5B2 in vacuo and implicit solvent to generate families of loop structures for minimization to determine robustness of predictions and identify spatial and dynamic correlations between key binding residues (Blake, Jain, Rick and Ashbaugh)

- Initial findings: HC3 loop has more varied and flexible structure than the other five antibody loops



Connections with Cyber Tools

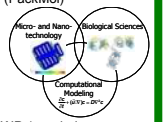
The Molecular Modeling component of this project requires:

1) Creation of putative antibody models based on sequence; (Modeler)

2) Parameterization of the analytes that bind to the antibodies; (Gaussian)

3) Docking analytes in different potential antibody binding loops; (PackMol)

4) Optimization of the antibody-analyte interaction by *in silico* point mutations. (Methods under development)



These tasks will require the following Work Packages:

WP 1: Scheduling and Data Services.
The details of integrating our Molecular Modeling packages into WP 1 are being addressed by Drs. Thomas Bishop (Tulane) and Tevfik Kosar (LSU).

WP 2: Information Services and Portals.
Drs. Thomas Bishop and Tevfik Kosar are collaborating to bring Bishop's DNA folding simulations on-line. The Workflow resulting from this effort can be readily modified to investigate the antibody and analyte interactions.

WP 3: Visualization Services.
Work is in progress to create modules that will permit all scientists involved in the project to visualize molecular models and other results via a common user interface without the necessity of transferring data or installing software on local lab computers.

WP 4: Application Services and Toolkits.
Drs. Steven Rick (UNO) and Henry Ashbaugh are developing replica exchange simulation techniques that will enable this group to efficiently identify antibody loop sequences that optimize the antibody-analyte interactions.

Acknowledgement Statement

The authors gratefully acknowledge the National Science Foundation (NSF) for their financial support of this research. This material is based upon work Supported by the NSF/EPSCoR under Award No. (EPS-0701491). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do notnecessarily reflect the views of the NSF.

ASYMMETRY OF STRUCTURAL CHARACTERISTICS OF LIPID BILAYERS INDUCED BY DIMETHYLSULFOXIDE

Raghava Alapati, Dorel Moldovan, and Ram Devireddy
Department of Mechanical Engineering, Louisiana State University, Baton Rouge ,LA -70803

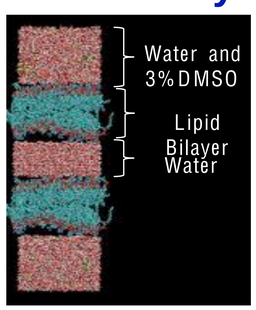
Abstract

Dimethylsulfoxide (DMSO) is one of the most widely used solvents in cell biology and cryopreservation. During a typical cryopreservation protocol the DMSO composition of aqueous buffers inside and outside of the cell is known to differ considerably. To model and understand the structural changes in cell membranes in such a situation we performed molecular dynamics (MD) simulations of an idealized lipid bilayer membrane which separates two aqueous reservoirs with and without DMSO. Zwitterionic dimyritoylphosphatidylcholine (DMPC) lipid bilayers was chosen as model membrane. Various structural and ordering parameters characterizing the DMPC lipid bilayers asymmetrically exposed to water and 3 mol% DMSO solution were evaluated.

Simulation Methodology

- MD simulations were performed using GROMACS software[1].
- The system consists of two DMPC lipid bilayers (consisting of 96 lipid molecules or 48 DMPC lipids in each leaflet) water placed in between the two bilayers and 3mol% DMSO-water solutions on either side of the lipid bilayers.
- The simulation are performed at const pressure (1atm) using semi-isentropic pressure coupling and at constant temperature characterizing liquid crystalline phase of lipid bilayers[2].
- Force field parameters for bonded and non-bonded are taken from Berger et al[3].
- An energy minimization based on steepest descent algorithm was initially applied to the structure prior to actual run.

Initial System



Lipids :: 192

Water :: 8133

DMSO:: 165

Figure 1. Simulation system showing the initial arrangement of the two lipid bilayers(48 lipids in each leaflet). Each bilayer has one side in contact with water and the other one in contact with 3% DMSO solution.

Simulation Results

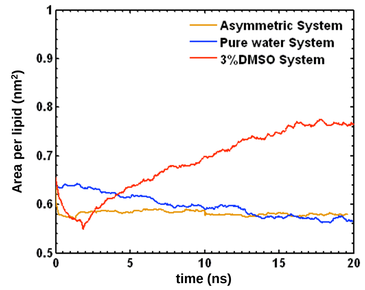


Figure 2. The area per lipid versus simulation for time three membrane systems: i) the DMPC 3 mol% DMSO asymmetric system, ii) the symmetric DMPC membrane embedded in a 3 mol %DMSO solution, and iii) the DMPC membrane in pure water

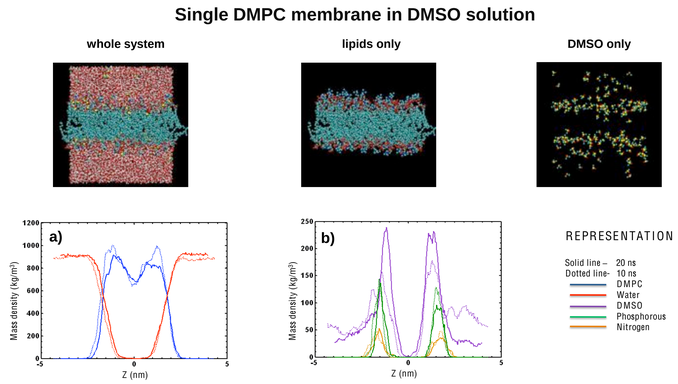


Figure 3: A single DMPC lipid bilayer exposed symmetrically to 3 mol % DMSO solution. Mass density profiles at 10 and 20 ns. a) Mass density profiles of lipids and water b) Mass density profiles of DMSO, phosphorous, and nitrogen.

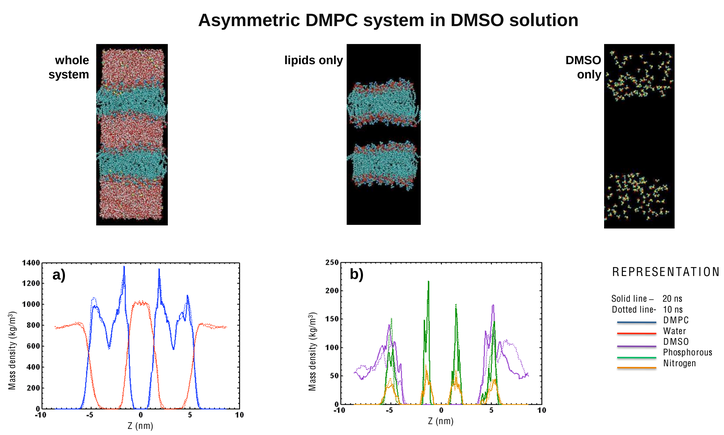


Figure 4: Two DMPC lipid bilayers system exposed asymmetrically to water and to 3 mol % DMSO solution. Mass density profiles at 10 and 20 ns. a) Mass density profiles of lipids and water b) Mass density profiles of DMSO, phosphorous, and nitrogen.

Conclusions

- In the asymmetric DMPC bilayer system the average area per lipid remains constant even after 20 ns; similar to the bilayer membrane immersed in pure water.
- The DMSO molecules cause large structural rearrangements within the outer lipid leaflets exposed to DMSO-water solution and has a reduced effect on the inner leaflets exposed to pure water.
- There is no evidence for DMSO penetration through the lipid bilayer[4].

References

1. E.,Lindahl, et al., J. Molec. Mod., 7,306 (2001).
2. H.J.C., Berendsen, et al., Biophys. J., 81, 3684 (1984).
3. O., Berger, et al., Biophys. J., 72, 2002 (1997).
4. D., Moldovan, et al., App. Phys. Lett., 91, 204104 (2007).

Connections with CyberTools

We are working with CyberTools team to develop a toolkit for job management, data analysis, and visualization. CyberTools (e.g. WP1, WP3, WP4) enables the use of High Performance Computing for the large scale atomistic simulations of lipid membrane systems by enabling the use of a user-friendly interface for submitting and monitoring multiple MD jobs.

Acknowledgements

The authors gratefully acknowledge the National Science Foundation for their financial support through grant NSF-EPSCoR RII Award No. EPS-0701491.

Automated Multi-Modality Image Fusion

Hua Cao[1], Nathan Brener[2], S. S. Iyengar[3], Bahram Khoobehi[4] Hilary Thompson[5]

[1,2,3]Computer Science Department, LSU

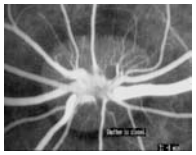
[4,5]LSU Student Health Center

Abstract

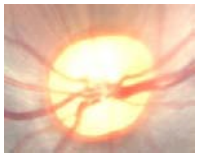
This study has made two new contributions to the image fusion area. The new contributions consist of the Adaptive Fidelity Exploratory Algorithm (AFEA) and the Heuristic Optimization Algorithm (HOA). The AFEA and HOA algorithms have been applied on two modalities of images of branching arterial structures. An optimal fusion result has been achieved by giving the visualization of a color image with a complete grayscale image overlay. Control points are detected at the vessel bifurcations using the AFEA algorithm. Shape similarity criteria are used to match the control points that represent same salient features of different images. The HOA algorithm adjusts the initial good-guess of control points at the sub-pixel level in order to maximize the objective function Mutual-Pixel-Count (MPC).

I. Edge Detection

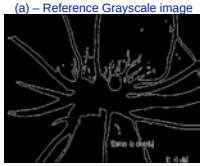
The Canny operator finds edges by looking for local maxima of the gradient of the input image (see Figure 1). It uses two thresholds for detecting strong and weak edges. Canny operator is less likely than the others to be "fooled" by noise, and more likely to detect true weak edges. Therefore, this approach chose Canny Edge Detector to extract the branching arterial structures from the binary images.



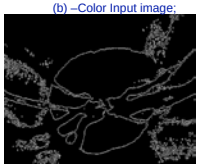
(a) – Reference Grayscale image



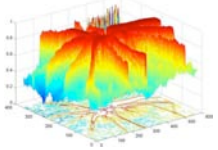
(b) –Color Input image;



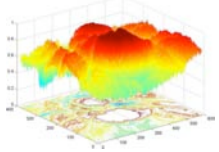
(c) - Canny edges of the reference image;



(d) – Canny edges of the input image;



(e) – 3D shaded surface plot of the reference image;



(f) – 3D shaded surface plot of the input image.

Figure 1 – Reference and input images, Canny edges and 3D surface plots.

II. Control Point Detection

Good-guess of the initial control point selection ensures fused image generated at an efficient computational time. Bad control point selection will significantly increase the computation cost, or even cause the image fusion to fail. Vessels or other factors may cause images to not necessarily match the arterial structures. Even when structure and function correspond, the mismatch still happens sometimes if inconsistence exists between structural and functional changes. Furthermore, grayscale images usually have higher resolution and are rich in information, whereas color images have lower resolution and are indeed abstract with some details or even missing some small vessels. Practically, those situations are unavoidable and will create difficulties in extracting the control points because the delineation of the vessel boundaries may not be precise. In this study, control points are detected using the AFEA algorithm (see Figure 2 and 3).

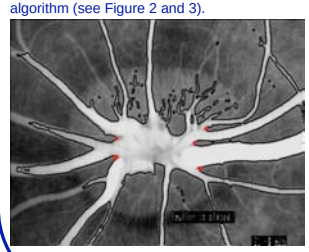


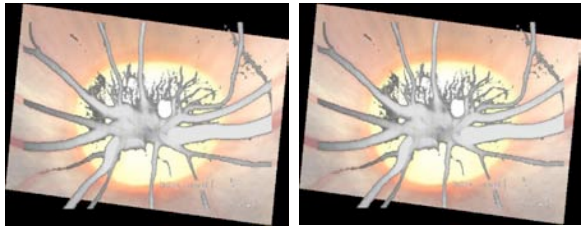
Figure 2 - Grayscale reference image's control points

Figure 3 –Color image's control points

III. Heuristic Optimization

An optimization procedure is required to adjust the initial good-guess control points in order to achieve the optimal result. The process can be formulated as a heuristic problem of optimizing an objective function that maximizes the Mutual-Pixel-Count between the reference and input images. The algorithm finds the optimal solution by refining the transformation parameters in an ordered way. By maximizing the objective function, one image's vessels are supposed to be well overlaid onto those of the other image (see Figure 4).

(a) Objective Function MPC = 5144 (b) Objective Function MPC = 7396 (c) Objective Function MPC = 7484



(e) Objective Function MPC = 7681 (f) Objective Function MPC = 7732

Figure 4. Fused image improvement during the iteration.

IV. Objective Function

Mutual-Pixel-Count measures the arterial structure overlap for corresponding pixels in both images. It is assumed that the vessels are represented by 0 (black pixel) and background is represented by 1 (white pixels) in the binary 2D map. When the artery pixel's transformed (u, v) coordinates on the input image correspond to the artery pixel's coordinates on the reference image, the MPC is incremented by 1. MPC is assumed be maximized when the image pair is perfectly geometrically aligned by the transformation (see Figure 5). After pre-processing, the binary images of the reference and input images are obtained, i.e. I_{ref} and I_{input}. Only black pixels from both images contribute to MPC. The ideal case is that all zero pixels of the input image are mapped onto zero pixels of the reference image. The problem can be mathematically formulated as the maximization of the following objective function:

$$f_{mpc}(x,y,u,v) = \sum_{I_{ref}(x,y)=1}^{u,v \in ROI} I_{input}(T_x(u,v), T_y(u,v)) \quad \text{and} \quad I_{input}(u,v)=1$$

where f_{mpc} denotes the value of the Mutual-Pixel-Count. T_x and T_y are the transformations for u and v coordinates of the input image. The ROI (Region-of-Interest) is the vessel region where the MPC is calculated on.

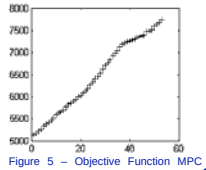


Figure 5 – Objective Function MPC improvement during the iteration.

V. Transformation Model

The 2D affine transformation model is applied to register the input image pixels into those of the reference image. The affine model has the capability to measure the lost information such as skew, translation, rotation, shearing and scaling that maps finite points to finite points and parallel lines to parallel lines.

$$\begin{pmatrix} U \\ V \\ 1 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & b_1 \\ a_3 & a_4 & b_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

VI. Connections with CyberTools

State-of-the-art imaging devices can quickly acquire multi-sensor 2D or 3D images. These images can further be transformed and merged into a single volume and thus combine the information of different modalities. Fusing images captured by different sensors (multimodal analysis) is able to provide the related staff with the complementary information, and thus help them more thoroughly understand both of the functional and structural information.

The application of this novel approach to branching arterial images demonstrates our new data fusion technique. This new algorithm can easily be extended to a number of other types of images.

Acknowledgements

Authors are grateful to Dr. Khooobehi, Dr. Thompson, and Dr. Ning for their support and help during this project. This work is funded in part by the NSF/EPSCOR RII grant.

Data Mining

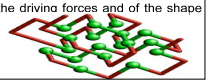
Asim Shrestha, Dimple Juneja, Nathan E. Brener, S. Sitharama Iyengar.
Louisiana State University

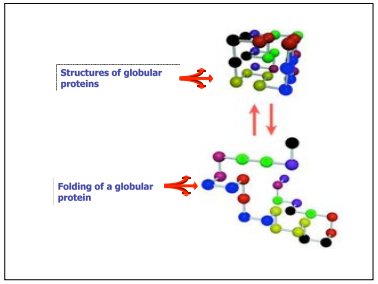
A Robust Data Mining Algorithm for Clustering of Similar Protein Folding Units

The properties of a protein depend on its sequence of amino acids and its 3D structure which consists of multiple folds of the peptide chain. If some of the properties depend primarily on the folding structure, then proteins with certain folding units may exhibit properties specific to those units. In that case, a classification of proteins based on folding units would facilitate the selection of proteins with certain desired properties.

THE PROTEIN FOLDING PROBLEM

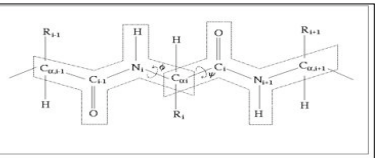
- Understanding and predicting the three-dimensional structures of proteins from their sequences of amino acids requires both basic knowledge of molecular forces and sophisticated computer programs that search for the correct configurations
- **The Objective:** The aim of the efforts in conformational searching is to use only knowledge of Amino acid sequence to predict protein structure. The points of conformance include:
  - Ability to distinguish a successful prediction from a failure
  - Enhanced knowledge of the driving forces and of the shape of the energy landscape
  - Faster Search Strategies





Data Mining Algorithm

- Application of Clustering data mining algorithm to a large medical data sets.
- An example: Protein Folding
- The properties of a protein depend on its sequence of amino acids and its 3D structure which consists of multiple folds of the peptide chain.
- If some of the properties depend primarily on the folding structure, then proteins with certain folding units may exhibit properties specific to those units.
- In that case, a classification of proteins based on folding units would facilitate the selection of proteins with certain desired properties.



n is computed from $\overrightarrow{C_{\alpha i}C_i} \times \overrightarrow{C_{\alpha i}N_i}$

p is computed from $\overrightarrow{N_iC_{\alpha i}} \times \overrightarrow{N_iC_{i-1}}$

q is computed from $\overrightarrow{C_iN_{i+1}} \times \overrightarrow{C_iC_{\alpha i}}$

ϕ is computed from n × p

ψ is computed from q × n

GROUPING ALGORITHM

- The peptide chain is decomposed into a series of overlapping fragments of length 8:

Fragment 1: [$(\phi,\psi)_1$ $(\phi,\psi)_2$ $(\phi,\psi)_3$ $(\phi,\psi)_4$ $(\phi,\psi)_5$ $(\phi,\psi)_6$ $(\phi,\psi)_7$ $(\phi,\psi)_8$]

Fragment 2: [$(\phi,\psi)_2$ $(\phi,\psi)_3$ $(\phi,\psi)_4$ $(\phi,\psi)_5$ $(\phi,\psi)_6$ $(\phi,\psi)_7$ $(\phi,\psi)_8$ $(\phi,\psi)_9$]

Fragment 3: [$(\phi,\psi)_3$ $(\phi,\psi)_4$ $(\phi,\psi)_5$ $(\phi,\psi)_6$ $(\phi,\psi)_7$ $(\phi,\psi)_8$ $(\phi,\psi)_9$ $(\phi,\psi)_{10}$]

.....

- We define the distance between two points A_i and A_j, DIST(A_i , A_j) , as

$$DIST(A_i, A_j) = ((\phi_{i1}-\phi_{j1})^2 + (\psi_{i1}-\psi_{j1})^2 + (\phi_{i2}-\phi_{j2})^2 + (\psi_{i2}-\psi_{j2})^2 + \ldots + (\phi_{i8}-\phi_{j8})^2 + (\psi_{i8}-\psi_{j8})^2)^{1/2}$$

where

$A_i = [(\phi_{i1}, \psi_{i1}) , (\phi_{i2}, \psi_{i2}) , \ldots (\phi_{i8}, \psi_{i8})]$

$A_j = [(\phi_{j1}, \psi_{j1}) , (\phi_{j2}, \psi_{j2}) , \ldots (\phi_{j8}, \psi_{j8})]$

- For every $(\psi_{im}-\psi_{jm})$, if $|\psi_{im}-\psi_{jm}|>180$, then we will use $360 - |\psi_{im}-\psi_{jm}|$, and similarly for $(\phi_{im}-\phi_{jm})$

- Let j be the index that labels the groups. We define the center of group j , C_j , as

$$C_j = [(\phi_{j1}, \psi_{j1}) , (\phi_{j2}, \psi_{j2}) , \ldots (\phi_{j8}, \psi_{j8})]$$

where

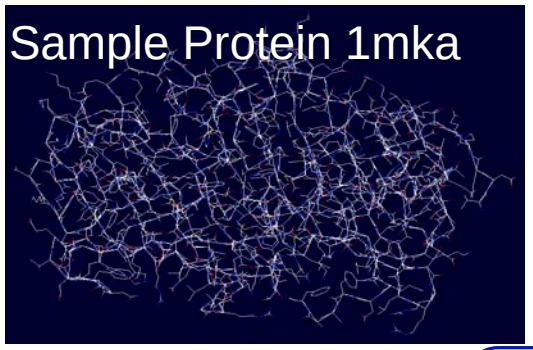
$\phi_{jm} = \Sigma\phi_{im} / N_j$

$\psi_{jm} = \Sigma\psi_{im} / N_j$

(i = 1, 2, .. N_j ; m = 1, 2, ... 8),

N_j is the number of points in the group, and the sum is over i. Such groups are regarded as folding units in our current work.

Sample Protein 1mka



CONCLUSIONS & FUTURE WORK

- This describes a data mining algorithm that can be used to classify proteins according to similar folding units.
- This classification has the potential to facilitate the selection of proteins with specific desired properties.
- The preliminary implementation of the algorithm indicates that it has the capability to discover common folding units in proteins and can be generalized to large sets of proteins.
- This technique will be explored in the context of geno/small molecule sensors (Soper)
- Identification of similar features would enhance the design features of genomic or immuno Sensors.

Common Folding Units Discovered by Data Mining

Proteins

1ash, 1bar, 1cca, 1cew, 1clm, 1crn, 1cct, 1erb, 1fut, 1hng, 1hoe, 1lbu, 1mka, 1mng, 1pkp, 1udi, 1utg, 1yal, 2vab, 5pti

3698 fragments

Group 1 514 fragments

Amino Acid	phi	psi
GLN	-60.078	-41.741
LEU	-69.310	-35.875
VAL	-65.116	-46.320
GLY	-67.025	-36.399
PHE	-62.244	-39.936
TYR	-66.128	-38.417
LEU	-64.114	-37.476
GLY	-70.167	-32.912

From 1mka α-helix

Group 2
188 fragments
From 1erb
β sheet

Group 3
79 fragments
From 1bsr

Group 4
61 fragments
from 1lbu

Table 1: A short list of proteins that were randomly selected

PDB Entry	Name of the Protein	Amino Acids Selected	Points Derived
1ash	HEMOGLOBIN (DOMAIN ONE)	1 – 146	138
1bsr	RIBONUCLEASE(BOVINE, SEMINAL) [CHAIN A]	1 – 124	115
1cca	CYTOCHROME C PEROXIDASE	4 – 294	282
1cew	CYSTATIN	9- 116	99
1clm	CALMODULIN (PARAMECIUM TETRAURELIA)	4 – 147	135
1crn	CRAMBIN	1 – 46	37
1cti	CYTIDINE DEAMINASE	4 – 294	285
1erb	RETINOL BINDING PROTEIN COMPLEX WITH N-ETHYL RETINAMIDE 2	2 – 174	164
1fut	RIBONUCLEASE F1	1 – 107	98
1hng	CO2 (RAT) (CHAIN B)	2 –176	166
1hoe	ALPHA-AMYLASE INHIBITOR HOE-467A	1-74	65
1lbu	HYDROLASE METALLO (ZN) DD-PEPTIDASE	1 – 213	204
1mka	BETA-HYDROXYDECANOYL THIOL ESTER DEHYDRASE (CHAIN A)	1- 171	162
1mng	MANGANESE SUPEROXIDE DISMUTASE (CHAIN A)	1 – 203	194
1pkp	RIBOSOMAL PROTEIN S5	4 – 148	137
1udi	URACIL-DNA GLYCOSYLASE	18 – 244	218
1utg	UTEROGLOBIN (OXIDIZED)	1 – 70	61
1yal	CARICA PAPAYA CHYMOPAPAIN	1 – 218	209
2vab	MHC CLASS I H-2KD HEAVY CHAIN	1 – 274	265
5pti	TRYPSIN INHIBITOR	1 – 58	49

Acknowledgements

This project is funded in part by NSF/EPSCoR RII

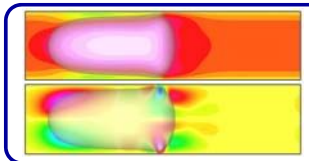
Connections with CyberTools

- Connected to WP1
- This technique will be explored in the context of geno/small molecule sensors (Soper)
- Identification of similar features would enhance the design features of genomic or immuno Sensors.

Numerical Simulations of Micro-Scale Segmented Two-Phase Flows for Bio-Analytical Chip Applications

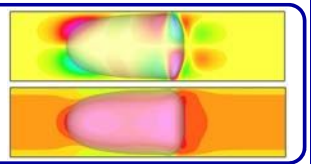
Eamonn D. Walker[1,2], Dimitris E. Nikitopoulos[1,2], Dorel Moldovan[1,2], Mayank Tyagi[3,4], Michael C. Murphy[1,2], Steven A. Soper[1,2,5], Gretar Tryggvason[6]

[1]Mechanical Engineering, [2]Center for Bio-Modular Multi-scale Systems, [3]CCT, [4]Petroleum Engineering , [5]Chemistry, LSU, Baton Rouge, LA; [6]Mechanical Engineering, WPI, Worcester, MA



Abstract

Segmented flows in micro-channels are of great interest to bio-analytical applications. They can reduce reagent volumes and enable high throughput without cross-contamination. A research code is used and being improved to predict such flows accurately & efficiently. This is done in close collaboration with the CyberTools group at CCT. On the scientific level, hierarchical disjoining pressure models and/or local molecular dynamics-based simulations need be implemented to accurately represent surface property effects, nano-scale effects in the thin films, breakup and coalescence.



About the Code

Formulation

- Incompressible, Isothermal Navier-Stokes Equations (each fluid)
- Jump conditions across interfaces (interfacial force balance)
- Boundary conditions specific to problem
  - Segmented Flow in Micro-channel
    - No-slip on channel walls
    - Periodic boundary conditions in stream-wise direction

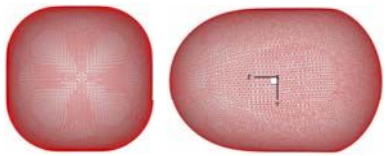
Solution Methods

Eulerian Governing Equations

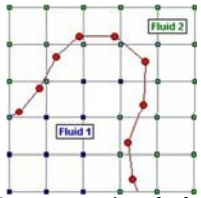
- Solved with a standard two-step projection method
  - Calculate pseudo-velocities (ignoring pressure effects)
  - Solve "Poisson" equation for pressure (satisfying continuity)
  - Correct velocities from pressure with an Euler step
  - Velocity used to advect the front and update velocity and pressure
- Elliptic "Poisson" solver for the pressure
  - MUDPACK open source code libraries
  - Use of Multi-grid method for increased efficiency
  - Red/black Gauss-Seidel successive over-relaxation (SOR)
  - Includes OpenMP parallelization capabilities
- Eulerian grid is fixed, regular, Cartesian and staggered (vel. & pres.)

Handling Interfaces – Front Tracking

- Fluid interface approximated as a front
  - Adaptive, unstructured, triangulated grid on front
  - Connected marker points (front nodes) advected with front
  - Marker points added/deleted as needed to maintain grid quality



3-D front obtained from simulation



2-D representation of a front

- Front-to-fixed-grid communication
  - Front used to assign fluid properties to fixed grid nodes
  - Surface tension
    - calculated from front curvature
    - distributed as a weighted source term to fixed grid nodes
- Fixed-grid-to-front communication
  - Calculated fixed-grid velocities applied to advect the front points
  - Grid-front interactions use smoothing to avoid excessive gradients

Objectives

- Accurately predict segmented multi-phase flows
- Validate predictions by comparing to experiment (poster by N. Kim et al.)
- Study segmented flow conditions difficult to realize in the laboratory
- Explain pressure-drop variation trends measured in the experiments
- Predict thin film disintegration for wall contamination assessment
- Examine sensitivity to perturbations in droplet size and pitch

Relevant Physical Parameters

Capillary Number

$$Ca = \frac{\mu_C J}{\sigma} = \frac{\text{Viscous Forces}}{\text{Capillary Forces}}$$

Density Ratio

$$\gamma = \frac{\rho_d}{\rho_C}$$

Reynolds Number

$$Re = \frac{\rho_C J w}{\mu_C} = \frac{\text{Inertial Forces}}{\text{Viscous Forces}}$$

Viscosity Ratio

$$\kappa = \frac{\mu_d}{\mu_C}$$

Weber Number

$$We = Ca \cdot Re$$

Channel Wall Surface Energy (Wettability)

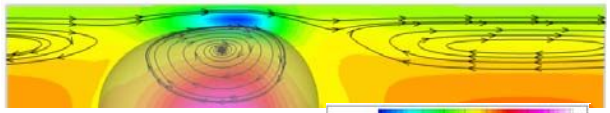
Results at Low Capillary Numbers

Periodic Segmented Liquid-Liquid Micro-channel Flow

Ca=0.002, We=0.008, Re=4, $\kappa=\mu_d/\mu_C=1.4$, $\gamma=\rho_d/\rho_C=0.55$

Contours of stream-wise velocity relative to that of the droplet and streamlines inside and around one droplet of the segmented flow obtained from 3-D, two-phase (liquid-liquid) simulations in square micro-channel (w=200μm).

Diagonal Plane



Center Plane

Thin-film physics and wall wettability must be properly introduced

Resolving thin film physics and properly introducing wall wettability into the simulation are critical in accurately predicting pressure drop and wall contamination probability.

Challenges

Scientific

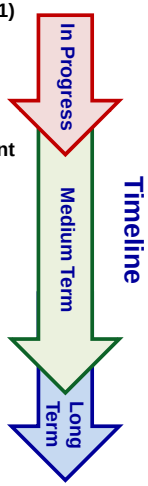
- Introduce thin-film and wall wettability physics
  - Hierarchical disjoining pressure models (Short-term)
  - Local Molecular-Dynamics coupled with Continuum (Long-term)
- Physical coalescence/break-up criteria (as above)
- Investigate a large multi-parameter space

Computational

- Improve code efficiency/accuracy/performance
  - Parallelization
  - New, elliptic solver algorithm suited to property discontinuities
- Post-processing & visualization of large number of large datasets
  - Parametric studies
  - Development of practical correlations from numerical data

Connections with CyberTools

- Multi-Phase flow Simulation Tool (WP4, WP3, WP1)
  - Parallelization
    - Implementation of parallelized Multi-Grid solver (WP4)
    - Distribution of different multi-processor simulations to groups of processors for efficient parametric studies (WP1)
  - Advanced interactive visualization tools (WP3)
  - Improvement of Accuracy/Performance
    - Implement Multi-Grid algorithm designed to handle elliptic "Poisson" equations with discontinuous coefficients (WP4)
  - Extend code capabilities to handle complex Cartesian geometries
    - Domain Decomposition (WP4)
    - Multi-blocking (WP4)
  - Computational Steering (WP1, WP3, WP4)



Acknowledgements

Special thanks to the NSF EPSCoR RII grant to the State of Louisiana, which has made this research and the collaboration with the CyberTools group possible, and the LA Board of Regents Dean's Fellowship program which is funding Mr. E. D. Walker.

Medical Image Classification Using Weighted Association Rules Based Classifier

Harpreet Singh[1], Sumeet Dua[1], Hilary W. Thompson[2]

[1]Data Mining Research Laboratory, Computer Science Program, Louisiana Tech University, Ruston, LA – 71272

[2]Department of Biostatistics, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA 70112

Contact e-mail: *sdua@coes.latech.edu*; Phone: 318-257-2830; *http://dmrl.latech.edu*

Abstract

Medical images are widely used by physicians to diagnose various diseases. Advanced in automated image collection routines coupled by our reliance on medical imaging for diagnostic discovery, treatment planning and decision support research has fuelled the demand for automated image mining and classification routines. The high volume of medical images coupled with the difficulty to discover features of interest in them poses an interesting algorithmic development challenge to autonomously interpret them for clinical decision support and early diagnosis. We present a new image representation scheme, a preprocessing method, and a computational framework for the classification of mammograms using Weighted Association Rules (WAR-BC). The framework is demonstrated here for mammogram classification but the classification theory allows extendibility to other domains. In mammogram classification an accuracy of 89% is achieved over ten repetitions, far surpassing the accuracy of other techniques. High Precision (96%) and Recall (91%) values show the strength of the proposed technique. We conclude that Association Rules can be effectively used to uncover the isomorphisms present in images which can be used for the classification of images for content-based image retrieval applications.

METHODOLOGY

The methodology is divided into four parts: Preprocessing, Segmentation and Feature Extraction, Association Rule Generation and Classifier Training, and Classification. During Preprocessing, a mammogram is converted to its binary image, and connected components are found from this image. Then, the image is segmented using these components. The segment boundary is smoothed. Finally, the black background is deleted from the mammogram, and histogram equalization is performed to remove noise.

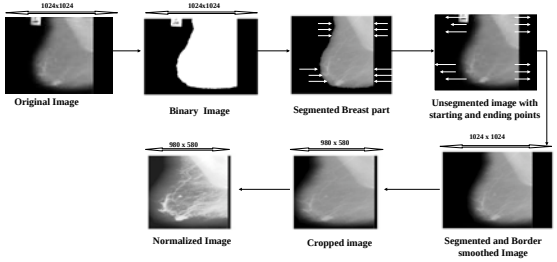


Figure 1. Mammogram Preprocessing for Label and Noise Removal

The preprocessed image is divided into non-overlapping segments of size 20x20 to capture the local relationships present in the image. Once the image has been segmented into blocks, eight texture features are extracted from each segment. Each vector is given a unique Segment ID, which, in our case, is the number of the segment from which the features were extracted, e.g. *TID 1 (f1, f2, f3……..f8) and TID 2(f1, f2, f3……..f8).* We use eight of the fourteen Haralick[1] coefficients. Since the classification mechanism is a stand alone tool, any set of nonharalick features can be used.

Once the features have been extracted from each segment, the image can be considered a transaction database where one transaction is one row of the database or the features extracted from one segment. The next step is to uncover the isomorphisms present by using association rules. An association rule is of the form *f1 (1134),f2(2124) → f8(8074)*with Support = 40% and Confidence = 80%, given by following formulas.

$$\text{Support}(f1(1134), f2(2124) \rightarrow f8(8074)) = \frac{\text{Number of Transactions having } (f1(1134), f2(2124), f8(8074)}{\text{Total Number of Transactions}}$$

$$\text{Confidence}(f1(1134), f2(2124) \rightarrow f8(8074)) = \frac{\text{Number of Transactions having } (f1(1134), f2(2124), f8(8074)}{\text{Number of Transactions having only } (f1(1134), f2(2124))}$$

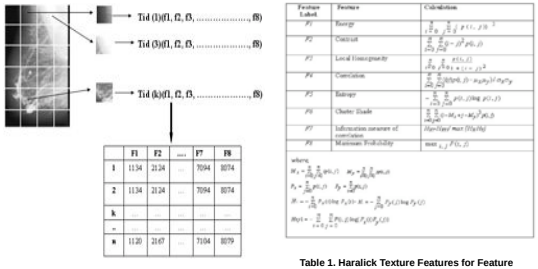


Figure 2. Segmentation and Feature Extraction

Feature Label	Feature	Calculation
f_1	Energy	$\sum_i \sum_j \frac{p(i,j)^2}{R^2}$
f_2	Contrast	$\sum_i \sum_j (i-j)^2 \frac{p(i,j)}{R^2}$
f_3	Local Homogeneity	$\sum_i \sum_j \frac{1}{1+(i-j)^2} \frac{p(i,j)}{R^2}$
f_4	Correlation	$\sum_i \sum_j \frac{(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x \sigma_y}$
f_5	Entropy	$-\sum_i \sum_j p(i,j) \log p(i,j)$
f_6	Cluster Shade	$\sum_i \sum_j (i+j-\mu_x-\mu_y)^3 p(i,j)$
f_7	Information measure of correlation	$\frac{Hxy - Hxy1}{\max\{Hx, Hy\}}$
f_8	Maximum Probability	$\max_{i,j} P(i,j)$

Table 1. Haralick Texture Features for Feature Representation

In our case, we only consider rules which have a minimum support of at least 4% and a confidence of at least 90%. Rules from images in each class are combined to form a separate class-level rule set for each class. Further, the class level rule sets are combined to form a global rule set. Weights are given to global rules according to their presence across images of the same class and across different classes (Horizontal). Class-level rule sets are ranked according to decreasing confidence/support pairs, and the highest ranked rule gets the highest weight (Vertical). For classifying a new image, both Horizontal and Vertical weights are added to find the weight of a matching query rule.

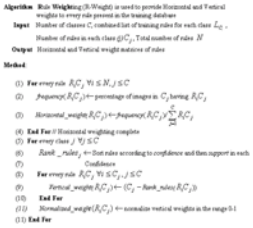


Figure 3. Pseudo Code for Rule Weighting

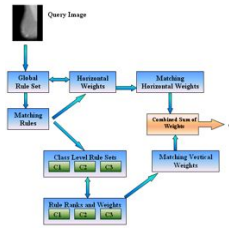


Figure 4. Classification Mechanism

For a query image the whole training procedure, except for rule weighing, is performed. Then, each rule from the query image is matched with the global rule set to find its *horizontal weight* and with class-level rule sets to find its *vertical weight* in each class. The horizontal and vertical weight of the rule is added and then multiplied by the number of items present to get a score for the rule. The procedure is repeated for each rule in the query image. Finally, the scores of the matching rules are added on a class-by-class basis and a cumulative sum is calculated for each class. The image is classified to the class with the highest cumulative sum.

RESULTS

A well known Mammography dataset called MIAS[2] is used for experiments. It consists of a total of 322 mammograms of which 208 are Normal, 63 are Benign, and 51 are Malignant. To make an accurate comparison with other existing techniques, we use the same data for training/testing (90/10).

Our technique (WAR-BC) outperforms others in the 10-fold technique. Class level accuracies are then found to see which class performs worst. We also run experiments with less training/testing data (70/30, 80/20) to check the accuracy.

To check the efficacy of association rules generated by our technique, we provide association rules as input to classifier F-KNN (called in F-KNN2) instead of raw haralick features. The increase in accuracy of F-KNN shows the strength of rules generated. Another set of experiments are carried out using only the Region of Interest (ROI) information for abnormal mammograms. Finally the classification is performed with respect to mammogram density of Fatty, Glandular, and Dense.

	BP NN	JAC	ARC-BC	F-KNN	PNC2	WAR-BC
1st	96.87	69.342	80	59.37	53.13	93.75
2nd	90.62	86.373	93.33	46.87	56.25	90.62
3rd	90.62	77.586	86.67	56.25	62.5	93.75
4th	78.125	72.912	76.67	71.87	62.5	84.37
5th	81.25	78.224	70	53.12	59.37	93.75
6th	84.375	77.055	76.67	75	20	81.25
7th	65.625	77.691	83.33	65.25	68.75	90.62
8th	75	73.752	76.67	56.25	20	90.62
9th	56.25	82.123	76.67	53.12	68.75	87.5
10th	93.75	79.819	83.33	59.37	12.5	90.62
Avg. Acc	81.2485	77.4877	80.334	59.647	48.375	89.685

Figure 5. Comparison of Different Techniques Using WAR-BC

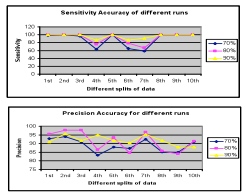


Figure 6. Sensitivity and Precision of WAR-BC over 10 Different Runs

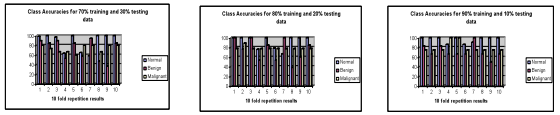


Figure 7. Class Level Accuracies for Different Training/Testing Pair of Data

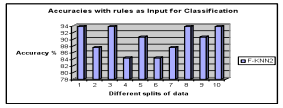


Figure 8. F-KNN2 Showing the Efficacy of Association Rules

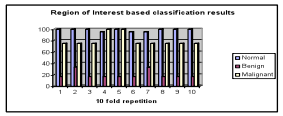


Figure 9. ROI Based Classification Results

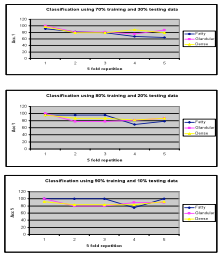


Figure 10. Density Based Classification

CONCLUSION

We have presented a novel framework for the improvement of mammogram classification which includes an improved preprocessing method, a new image representation scheme, and a new rule weighting strategy. Results demonstrate that our technique is superior to existing techniques. For further reference to this research please consult [3].

REFERENCES

[1] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. on SMC,* IEEE SMC Society, Piscataway, NJ, Nov. 1973, pp. 610-621

[2] MIAS Database, *The PCCV Project*: Benchmarking Vision Systems, http://peipa.essex.ac.uk/info/mias.htm.

[3] S. Dua, H. Singh and H. W. Thompson, "Associative Classification of Mammograms using Weighted Rules based Classification", Elsevier Expert Systems with Applications (submitted).

Connections with Cyber Tools

This work relates to the WP1 (Data Services) aims for the design and development of tools for metadata extraction and data mining services. The work is also connected with the WP2 (Information Services) of Cyber Tools development with regards to information discovery algorithms. The work is a result of collaboration between investigators from Louisiana Tech University and Louisiana State University Health Sciences Center at New Orleans.

Acknowledgements

-NSF/LA-BoREPSCoR program
-NIH/INBRE program

Computational Model of a Microfluidic Mixing Chamber for Miniaturized Immunosensor Devices

Katharine Hamlington[1], Jerina Pillert[1], David Halpern[2], Donald P. Gaver[1]

[1]Tulane University
[2]University of Alabama

Abstract

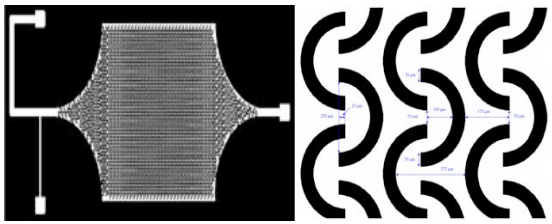
We are using numerical simulations to determine the appropriate geometric configuration of a microfluidic channel network to enhance the mixing and subsequent reactions of biological or chemical species. Convective mixing on the microscale can be difficult with low Reynolds number flows; typically, this requires long length- and time-scales to allow molecular diffusion between laminar streams. Decreasing these mixing scales in a microfluidic network allows for the creation of a portable sensing device that can readily detect harmful biological or chemical agents. The Institute for Micromanufacturing at Louisiana Technological University has developed a microfluidic system consisting of a network of novel omega-shaped channels designed to enhance mixing by introducing circulatory flows. Our goal is to investigate and optimize the design of the "omega channels" for mixing in an immunosensor device. Solvent dynamics in complex geometric domains involving omega-shaped obstructions were computationally determined by solving the equations for Stokes flow using the boundary element method. We analyzed improvement of mixing in the omega channels by examination of convective flow fields.

Miniaturized Immunosensor Devices

- Detect biological or chemical agents by antigen-antibody binding and subsequent signal detection
- Useful in disease diagnosis, detection of food toxins, and environmental monitoring for harmful chemical agents
- Compact size, portability, low cost, and ease of operation would enable widespread use by general public

The Mixing Dilemma

- Two species must mix and bind to produce signal
- Fluid flow is purely laminar at microscale (non-turbulent) and mixing occurs only due to diffusion, requiring long length- and time- scales
- Omega channels developed by LaTech may induce transverse and circulatory flows to promote mixing between two species



GOAL: ***Computationally determine the optimal geometric configuration of the omega channel network to enhance mixing of two species.***

Governing Equations

- Inertial effects negligible at microscale; Reynolds number (Re) << 1
- Incompressible flow governed by continuity and Stokes equations

$$\nabla \cdot \mathbf{u} = 0$$

$$\nabla P = \mu \nabla^2 \mathbf{u}$$

Boundary Element Method

- Used to solve for velocity field and surface stresses
- Green's theorem is applied to Stokes equations to obtain an integral equation linking the velocity and stress on the boundary surface *S*

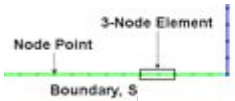
$$\bigstar \; C_{kl} u_l(\mathbf{x}) + \int_S \mathbf{T}_{ik}(\mathbf{x},\mathbf{y}) u_i dS_y = \frac{1}{\mu} \int_S \mathbf{U}_{ik}(\mathbf{x},\mathbf{y}) \tau_i dS_y$$

- U and T are kernels based upon the free-space Green's function

$$\mathbf{U}_{ik} = -\frac{1}{4\pi}\left(\delta_{ik} \log|\mathbf{x}-\mathbf{y}| - \frac{(x_i - y_i)(x_k - y_k)}{|\mathbf{x}-\mathbf{y}|^2}\right)$$

$$\mathbf{T}_{ik} = -\frac{1}{\pi}\left(\frac{(x_i - y_i)(x_j - y_j)(x_k - y_k) n_j(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^4}\right)$$

- Only boundary is discretized into *N* 3-node quadratic elements



- Discretized integral equation (★) expressed as system of linear equations

$$H\mathbf{u} = G\tau$$

Computational Domain

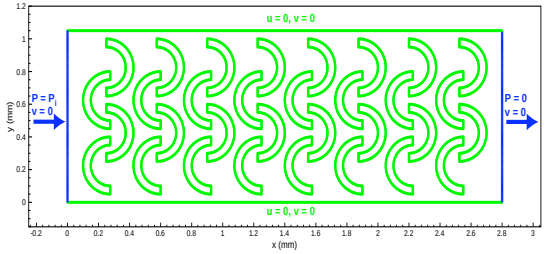


Figure 1: Schematic of computational domain.

- Omegas modeled with manufactured dimensions
- Pressure drop imposed over channel length; P_i = 3.5 MPa
- No slip boundary condition on upper and lower walls and obstructions
- Fluid viscosity μ = 1 Pa•s

Results

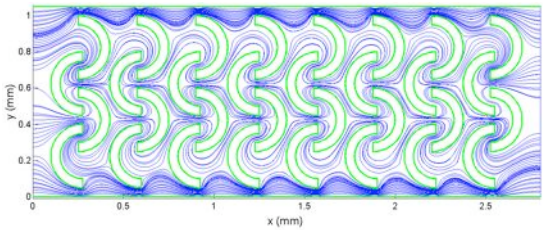


Figure 2: Streamlines of fluid flow in omega channel domain.
Flow rate = 1 mL/min, Average velocity = 105 mm/s, Re = 0.03

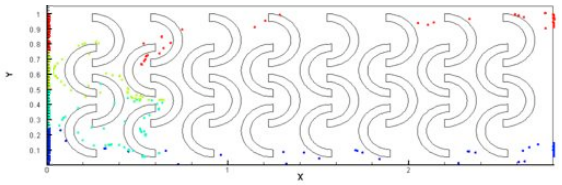


Figure 3: Snapshot of a particle trajectory field.
Particles initially positioned along y-axis at x = 0 as shown.
Path of each particle is traced as it flows through the domain.

Conclusions

- Closeness of streamlines and particle trajectories indicates that omega channels may enhance mixing by decreasing the diffusion length scale.
- The lack of obvious vortices suggests that domain modification may be important to improve mixing.
- We will continue to optimize the geometry for mixing in terms of initial concentrations and overall chamber size by incorporating the convection-diffusion-reaction equations of the species into the calculations.

Connections with CyberTools

- We have worked closely with Dr. Mayank Tyagi and the WP4 team to parallelize our boundary element code using OpenMP for use in the HPC environment.
- WP3 is helping us visualize our velocity field using Tecplot.
- As a long term goal, we are developing a *CyberTool* package to solve Stokes flow equations for use by the scientific community.

Acknowledgements

We thank Mr. John Sullivan who aided in the development of this poster and Dr. Hideki Fujioka who provided invaluable computational assistance. ***This work was funded by NSF EPSCoR.***

Multi-Phase Flow in Polymer Microfluidic Systems

Namwon Kim, Estelle T. Evans, Daniel S. Park, Subramanian Balamurugan, Steven A. Soper, Dimitris E. Nikitopoulos and Michael C. Murphy

Department of Mechanical Engineering and Chemistry
Louisiana State University, Baton Rouge, LA 70803

Abstract

Multiphase flow is realizing its potential in various lab-on-a-chip devices as a promising candidate for the flow control methods. For the better understanding and application of multiphase flow in microfluidic systems, fundamental physical studies are required. An experimental investigation of multi-phase flows in polymer microfluidic channels replicated using hot embossing of poly-methyl-methacrylate (PMMA) and polycarbonate (PC) with micro-milled brass mold inserts was performed. Deionized water and dry air were used for gas-liquid two-phase flow and deionized water and fluorinated hydrocarbon fluid were used for liquid-liquid immiscible flow.

Introduction

Motivation

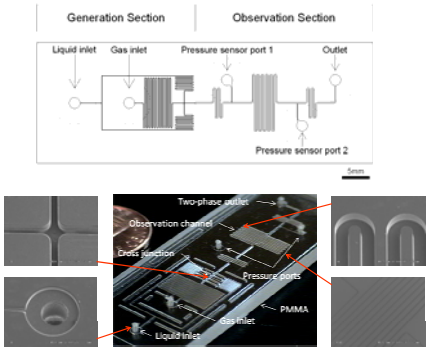
- Reduced use of reagents in a microfluidic network
  - Substitution of inert fluid (gas, immiscible liquid) in aqueous reagent instead of filling the entire channel
- Fast mixing and minimized dispersion of reagents
  - Two-phase flow: mixing is Intensified by internal convective motion in a liquid plug

Objectives

- Experimental Investigation of multi-phase flows
  - Two-phase flow regimes
  - Stability of gas bubbles and liquid plugs
  - Two-phase pressure drops
  - Effect of surface properties (Surface energy, roughness)

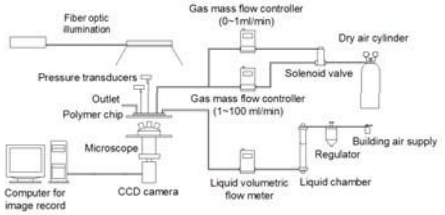
Gas-liquid two-phase flow

Design for generation and test of two-phase flow



MECHANICAL ENGINEERING DEPARTMENT
TO PREDICT ◆ TO DESIGN ◆ TO PERFORM

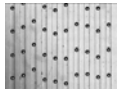
Experimental Apparatus



Two-phase flow patterns

$$\beta_L = \frac{Q_L}{Q_L + Q_G} = 1 - \beta_G$$

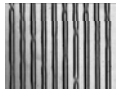
Where β_L = liquid volumetric flow ratio
β_G = gas volumetric flow ratio
Q_L = liquid volumetric flow rate
Q_G = gas volumetric flow rate



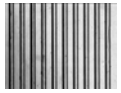
Capillary bubbly flow, $0.66 \le \beta_L \le 1$
$L_b / w < 1$
- Regular distribution of bubbles
- No randomly dispersed bubbles
where L_b = gas bubble length, L_p = liquid plug length and w = observation channel width

Segmented-1, -2, -3, $0.06 \le \beta_L \le 0.66$
Segmented-1: $L_b / L_p < 1$, $L_b / w < 5$
Segmented-2: $L_b / L_p > 1$, $L_b / w < 5$
Segmented-3: $L_b / L_p > 1$, $L_b / w > 5$

Segmented-annular flow
$0.018 \le \beta_L \le 0.06$
- Beginning of coalescence of neighboring bubbles with thinning of liquid plug

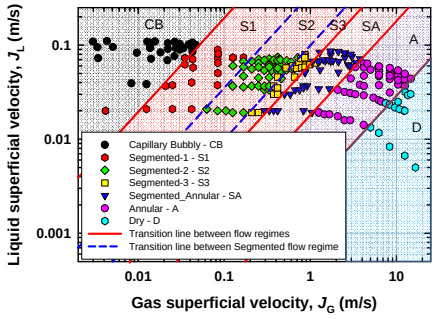


Annular flow
$0.0029 \le \beta_L \le 0.018$
- Ring shaped liquid film flow along the channel wall in an irregular pattern

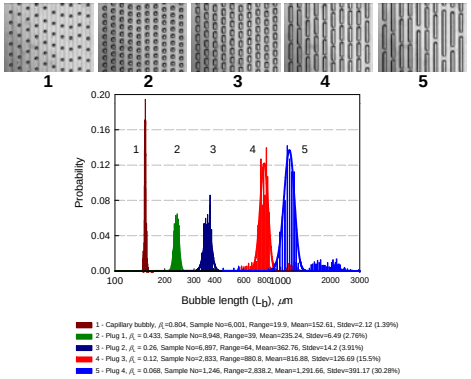


Dry flow
$0 \le \beta_L \le 0.0029$
- Liquid confined to the corners of the rectangular channel

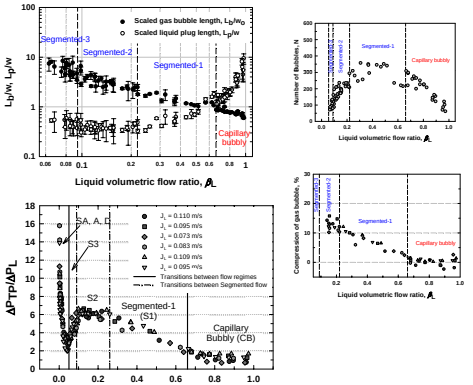
Flow map



Regularity of flow



Length of gas bubble and liquid plug & pressure drop



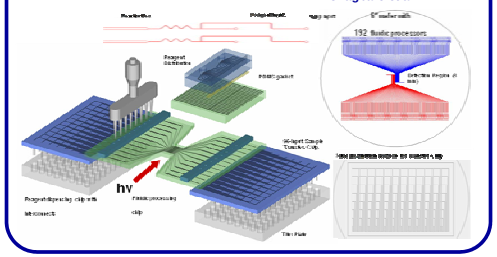
$$\Delta P_{TP} = \Delta P_{Friction} + \Delta P_{Gauge} + \Delta P_{Contraction}$$

$$\Delta P_{TP} = 0.89 \left[\frac{32}{D_h^2} \left(L_G + \frac{\mu_L}{\mu_G} L_L \right) + \frac{4kV}{3D_h} \left(\frac{\sin\theta_{apex}}{\theta_{apex}^2} + \frac{\sin\theta_{apex}}{\theta_{apex}^2} \right) + K_{ent} \frac{\eta}{\eta} \right]$$

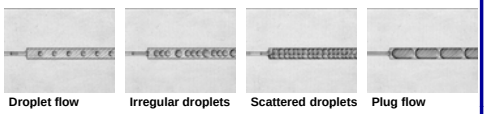
Liquid-liquid immiscible flow

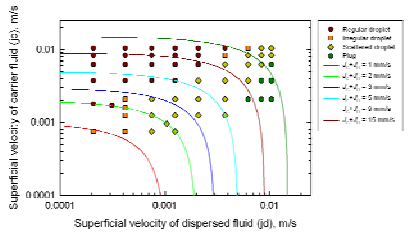
Small Molecule Sensor: HTS for Drug Discovery

- Okagbare et al.

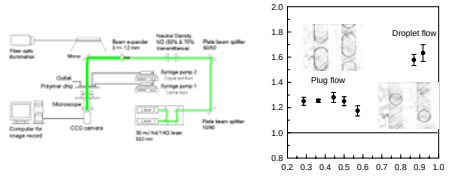


Liquid-liquid Immiscible Flow pattern and map





Droplet and plug velocity measurement



Conclusions

Multi-phase flow patterns, maps, transition between flow regimes, regularity of flows and length of gas bubble and liquid plug were determined for each case. Gas-liquid two-phase flow pressure drops were measured and each flow regime identified on the basis of topological observations is associated with different trends of the pressure drop variation with respect to volumetric flow ratio. While the Lockhart-Martinelli correlation showed good agreement with results at the Capillary bubbly, Segmented-annular, Annular and Dry flow regimes, a new linear model was developed for the segmented regime, which was divided into three more specific flow regimes, segmented 1, 2 and 3.

Acknowledgements

This work is supported by the National Science Foundation under grant EPS-034641 and MRI grant NSF-9977576(CTS) as well as the State of Louisiana Board of Regents under grant LEQSF(2005-06)-ENH-TR-20.

Transport of Molecular Clusters Through Nano-Scale Channels

Nancy Lekpeli [1], Dorel Moldovan[2], and Dimitris Nikitopoulos[2]
[1]Univerite Claude Bernard, Lyon, France
[2]Department of Mechanical Engineering, Louisiana State University, Baton Rouge, Louisiana 70803

Abstract

We report molecular dynamics (MD) simulation results depicting the behavior of a single molecule in a nanochannel under "gravity" driven Poiseuille flow of a Lennard-Jones liquid. The goal of this research is to further the understanding of the mechanism(s) underlying single molecule translocation through nanochannels. Recent experimental and simulations studies suggest that a similar system involving translocation of DNA molecules through nano-pores could be developed into an ultrafast method of DNA sequencing.

Simulation Methodology

➢The simulation system consists of two walls, a slab of liquid, and a solid molecule (i.e. atomic cluster) embedded in the liquid (Fig. 1). Periodic boundary conditions were applied in the x-direction (the channel axis)

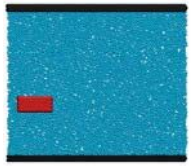


Fig. 1 Schematic representation of the simulation system

➢ MD simulations were performed with the software package LAMMPS

➢The interactions between any pair of atoms are described by the Lennar-Jones potential.

➢The two-dimensional system consists of about 6000 atoms and the molecule has an elongated shape of aspect ratio 2.6

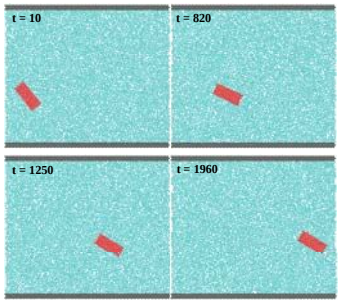
➢The simulatioons were conducted and analyzed in reduced units

➢The simulations were carried out at temperature $k_BT/\varepsilon=1.2$ and density $\rho/\sigma^2=0.81$.

➢The Poiseuille flow was induced by introducing a "gravity" force that is applied parallel to the channel axis to each atom of the liquid and molecule.

Simulation Results

Fig. 2 Simulation snapshots of the molecule moving in a nanochannel in a Poiseulle flow. Time is given in reduced units.



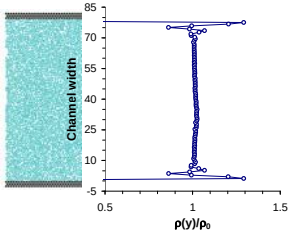


Fig. 3 Normalized atomic density in the liquid phase across the width of the nanochannel. The liquid bulk atomic density is ρ_0= 0.81.

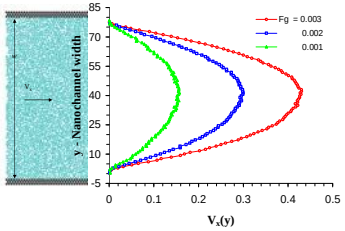


Fig. 4 Velocity profiles obtained from MD simulations of Poiseuille flow. The result are given for three values of the additional constant force, F_g=0.003, F_g=0.002, and F_g=0.001, applied to each "liquid" atom to generate the flow.

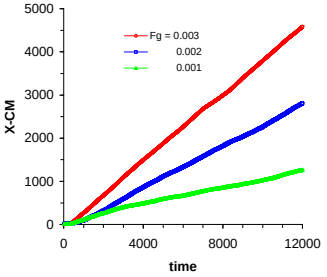


Fig. 5 Variation of the x-component of the position of the molecule center of mass versus time for three flow regimes controlled by gravity forces: F_g=0.003, F_g=0.002 and F_g=0.001

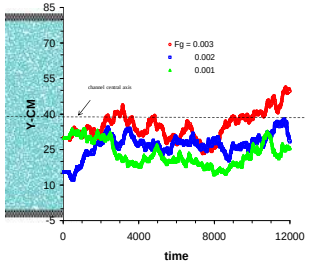


Fig. 6 Variation of the y-component of the position of the molecule center of mass versus time for three regimes controlled by gravity forces: F_g=0.003, F_g=0.002 and F_g=0.001
.

Conclusions

➢The simulations show that close to the channel walls the liquid is indeed more structured and therefore has significantly different rheological properties compared to the bulk liquid.

➢The velocity profile the of the liquid in Poiseuille flow, indicate that the deviation between continuum and MD prediction is indeed very small.

➢The MD simulations indicate that the presence of af a large shear rate in the flowing liquid promotes the motion of the molecule towards the center of the nanochannel.

➢MD simulations can provide detailed atomistic understanding of the mechanism of by which a single biomolecule translocates through long nanometer-narrow channels. In addition by providing dynamical snapshots of the translocation process can they act as computational microscopes therefore having great potential for assisting the design and development of nanochannel-based biosensor systems.

Connections with CyberTools

We are working with CyberTools team to develop a toolkit for job management, data analysis, and visualization. CyberTools (e.g. WP1, WP3, WP4) enables the use of High Performance Computing for the large scale atomistic simulations of single molecule translocation through nano-channels by enabling the use of a user-friendly interface for submitting and monitoring multiple MD jobs.

Acknowledgements

The authors gratefully acknowledge the National Science Foundation for their financial support through grant NSF-EPSCoR RII Award No. EPS-0346411.

Coupling an Einstein and an Euler code via the Cactus framework

Oleg Korobkin[1,2], Erik Schnetter[1,2]

[1] Center for Computation and Technology, Louisiana State University

[2] Department of Physics and Astronomy, Louisiana State University

Introduction

Cactus is a state-of-the-art high-performance software framework for 3D numerical simulations. The Cactus framework allows to create multiphysics applications, which combine independently developed highly sophisticated codes with various numerical discretization schemes. Here, we demonstrate the coupling between the Einstein evolution code, implemented with high-order finite differences on a non-overlapping multiblock domain (QUILT), and a relativistic hydrodynamical Euler code, implementing a finite volume scheme on overlapping patches (DiFranco). The coupling mechanism is provided by the Einstein toolkit (thorns ADMBase and TmunuBase). Both interacting codes were adjusted to satisfy a few coupling requirements, which now allows us to interface them with other codes used in the numerical relativity community.

Einstein's equations:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = T_{\mu\nu}$$

Constraint equations

$${}^3R - K^{ij}K_{ij} + K^2 = 16\pi\rho$$

$${}^3\nabla_i (K^{ij} - g^{ij}K) = -8\pi J^j$$

Evolution equations

$$\partial_t g_{ab} - (1+\gamma_1)\beta^k \partial_k g_{ab} = F^{(g)}_{ab}$$

$$\partial_t \Pi_{ab} - \beta^k \partial_k \Pi_{ab} + \alpha g^{ki} \partial_k \Phi_{iab} - \gamma_1\gamma_2 \beta^k \partial_k g_{ab} = F^{(\Pi)}_{ab}$$

$$\partial_t \Phi_{iab} - \beta^k \partial_k \Phi_{iab} + \alpha \partial_i \Pi_{ab} - \gamma_2 \alpha \partial_i g_{ab} = F^{(\Phi)}_{iab}$$

Matter fields equations:

$$T_{\mu\nu}{}^{;\nu} = 0$$

Fluid evolution equations in conservation form

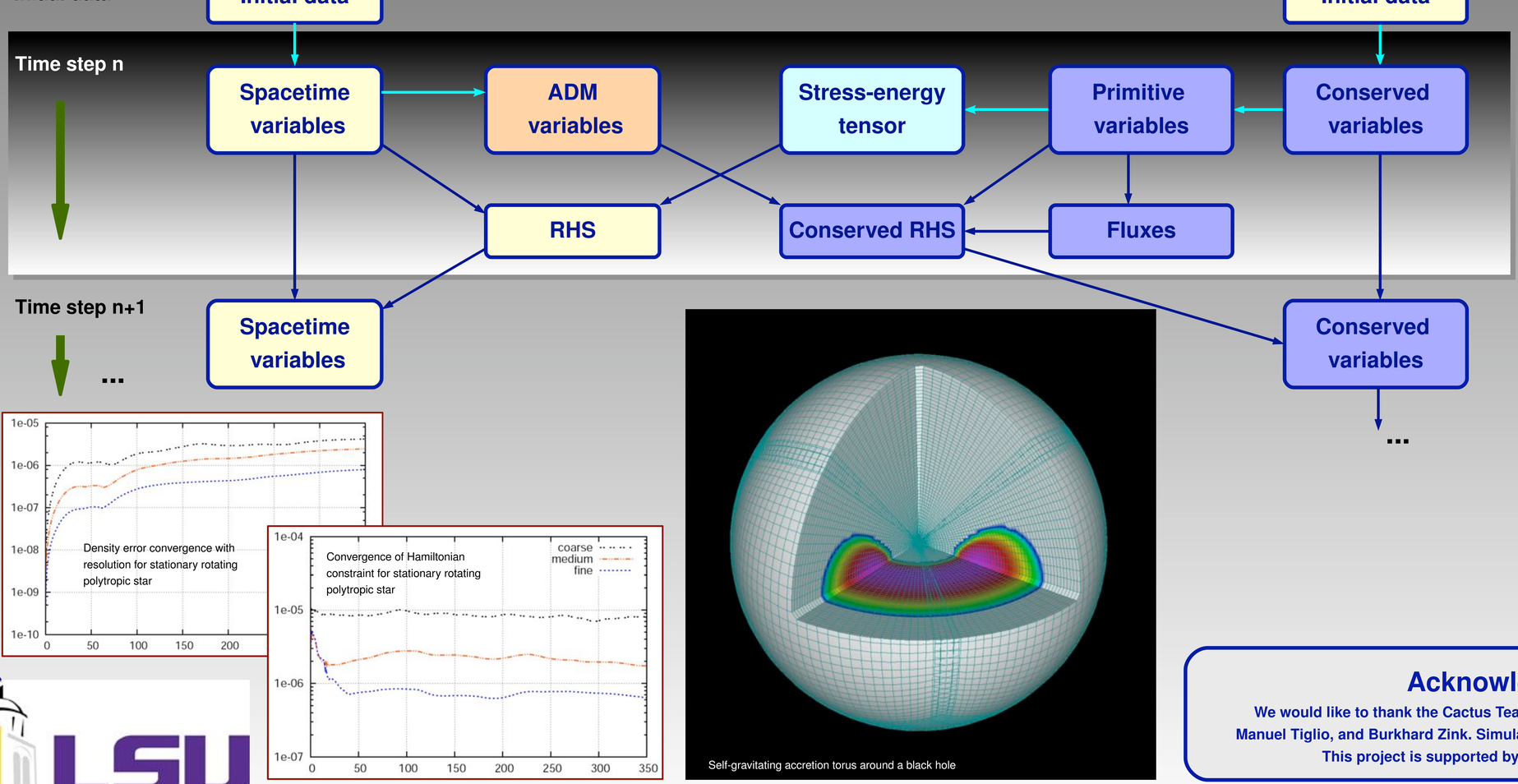
$$\partial_t Q + \partial_i F^i(P) = S(P)$$

Variables

- Formulation-specific: $g_{ab}, \Pi_{ab}, \Phi_{iab}$
- ADM: g_{ij}, K_{ij}, α, β^i

Variables

- Stress-energy tensor
- Primitive: ρ, p, u, u_x, u_y, u_z
- Conservative: dD, dQ_t, dQ_x, dQ_y, dQ_z



Connections with CyberTools

- **WP 1: Scheduling and data services**
  Cactus has been used for parallel computation on different architectures and provides tools for parallel data access
- **WP 2: Information services and portals**
  Cactus connects to portals via thorns Announce and Formaline
- **WP 3: Visualization services**
  Cactus IO formats are supported by remote visualization tools (such as VisIt)
  Cactus has a built-in web-server that can display results interactively.
- **WP 4: Application services and toolkits**
- **Science Driver: e.g., Biotransport**
  Multiblock infrastructure from Cactus can be used to solve problems in biotransport

Acknowledgments

We would like to thank the Cactus Team, Christian D. Ott, Christian Reisswig, Manuel Tiglio, and Burkhard Zink. Simulations were performed on LONI machines. This project is supported by NSF grant 0721915 (“Alpaca”)

An Algorithmic Tool for Protein Structure Classification based on Conserved Hydrophobic Residues

Pradeep Chowriappa[1], Sumeet Dua[1], Jinko Kanno[2], and Hilary Thompson[3]
[1] Data Mining Research Laboratory, Computer Science Program, Louisiana Tech University, Ruston, LA – 71272.
[2]Mathematics and Statistics Program, Louisiana Tech University, Ruston, LA 71270.
[3]Department of Biostatistics, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA 70112.

Abstract

Motivation: Protein folding is frequently guided by local residue interactions that form clusters in the protein core. The interactions between residue clusters serve as potential nucleation sites in the folding process. Evidence postulates that the residue interactions are governed by the hydrophobic propensities that the residues possess. To this end, an array of hydrophobicity scales have been developed to determine the hydrophobic propensities of residues under different environmental conditions. We hypothesize that proteins of the same homology contain conserved hydrophobic residues that exhibit analogous residue interaction patterns in the folded state.

Product: We developed a graph theory based data mining tool to extract and isolate protein structural features that sustain invariance in evolutionary related proteins, through the integrated analysis of five well-known hydrophobicity scales over the 3-D structure of proteins. The results demonstrate that discriminatory residue interaction patterns shared among proteins of the same family can be employed for both the structural and the functional annotation of a variety of proteins. We obtained an average of 90% accuracy in protein classification with a significantly small feature vector compared to previous results in the area. The tool is usable for a variety of proteins from distinctively different families and classes.

Background

Researchers have investigated the correlation of hydrophobic interactions to similarities in 3-D structural elements, and have exhibited and exploited property conservation at these sites. Although using different approaches, each model suggests a common and unexpected feature of protein packing that proteins significantly rely on based on few members of the set of conserved residues.

- Paiardini et al [1] and Reddy et al, (CKAAPS)[2], using Multiple Sequence Alignment (MSA) techniques show that a significant correlation exists between sequence, structure, and Conserved Hydrophobic Contacts (CHC) that remain invariant during long evolutionary periods.
- Tsai et al [3] propose a method using a scoring function based on the physicochemical properties of hydrophobicity, compactness, solvent accessibility of surface area (ASA), and segmentation to test the validity of fold unit definition based on eigenvector analysis.
- The models proposed by Muppirala et al [4] and Huang et al [5], quantitatively measure the individual contributions of amino acid residues by enumerating contacts between a hydrophobic residue and its surrounding area within a protein structure.

Methodology

We have developed a data mining tool for the integrated analysis of five popular hydrophobicity scales to enhance the detection of structurally conserved regions among homologous proteins, which we believe will also be useful for classification purposes. Incorporating the metric of mutual information to identify compact structural units, we extract frequently occurring patterns using a discriminative weighing function. By doing so, we reduce our feature space and show that the reported conserved hydrophobic residues are sufficient to differentiate between proteins at both the class and fold levels of the Structural Classification of Proteins (SCOP) hierarchy.

Rank	1	2	3	4	5	6	7	8	9	10
Kyte and Doolittle	ARG	LYS	ASP	GLU	ASN	GLN	HIS	PRO	TYR	TRP
Hopp Woods	TRP	PHE	TYR	ILE	LEU	VAL	MET	CYS	ALA	HIS
Janin et al	LYS	ARG	GLU	GLN	ASP	ASN	TYR	PRO	THR	HIS
Rose et al	LYS	ASP	GLU	GLN	ASN	PRO	ARG	SER	THR	GLY
Eisenberg et al	ARG	LYS	ASP	GLN	ASN	GLU	HIS	SER	THR	PRO
Rank	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Kyte and Doolittle	SER	THR	GLY	ALA	MET	CYS	PHE	LEU	VAL	ILE
Hopp Woods	THR	GLY	PRO	ASN	GLN	SER	ASP	GLU	LYS	ARG
Janin et al	SER	ALA	GLY	TRP	MET	PHE	LEU	VAL	ILE	CYS
Rose et al	ALA	TYR	HIS	LEU	MET	TRP	VAL	PHE	ILE	CYS
Eisenberg et al	TYR	CYS	GLY	ALA	MET	TRP	LEU	VAL	PHE	ILE

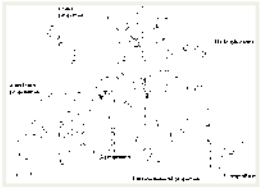


Table 1. Ranks of amino acid based on propensities assigned by the five hydrophobic scales [6], and all the known Amino Acid Indices listed by **(http://www.genome.ad.jp/aaindex/)**.

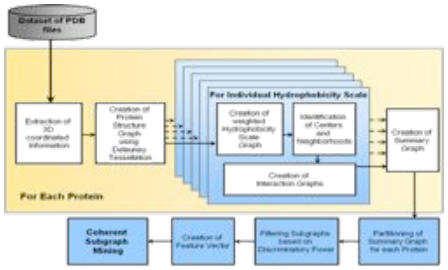
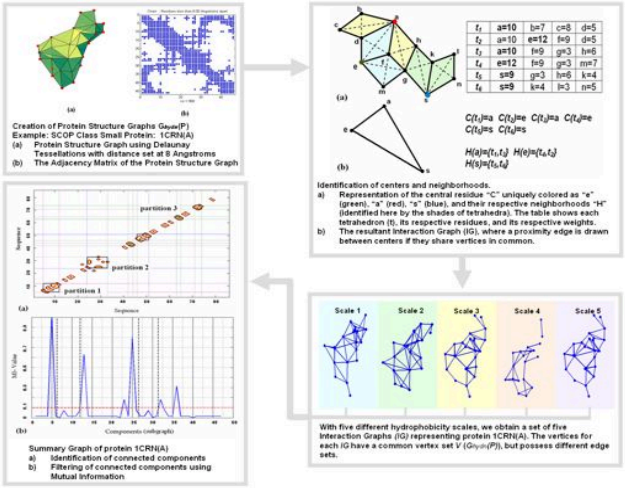


Figure 1. Overview of Methodology

The Process of Feature Extraction



Validation

Binary Class Classification:

Constraints: Proteins filtered under < 35% pair wise sequence similarity to remove highly homologous proteins, with resolution<=3 and R factor <= 1.0.

(Class Level) Dataset C1:

Consists of proteins from Classes:

All Alpha- Nuclear Receptor Ligand-binding domain proteins (NB 16 proteins), against All Beta-Prokaryotic serine proteases family (PSP, 10 proteins).

(Fold Level) Dataset C2:

Consists of proteins from Folds of Class All-Beta:

Eukaryotic serine proteases family (ESP, 19 proteins) and Prokaryotic serine protease family (PSP, 19 proteins).

Results reported in Table 2.

Dataset	Method	Features	Accuracy (%)
C1	DT	20646	100
	AD	23130-37394	96-100
	LFM-Pro	5282	100
	Ours	38	100
C2	DT	15895	95
	AD	18491- 32569	93-95
	LFM-Pro	2180	100
	Ours	29	96.55

Table 2. Comparison of results obtained from Binary Class classification.

Structural Class	Precision (%)
Coiled Coil Proteins (A)	100
All Beta Proteins (B)	90.9
Alpha/Beta Proteins (C)	81.8
Overall Accuracy	90

Table 3. Results obtained from Multiclass Classification using the proteins of CKAAPS database.

Multi Class Classification:

Constraints: We have selected ten proteins from each class, resulting in a dataset consisting of 30 proteins which satisfy a RMSD of <=3.0 and a Z-score of >=4.5.

The proteins used in our study are located in the fssp-ckaaps-1.2 database **(ftp://ftp.sdsc.edu/pub/sdsc/biology/ckaap)** provided in [2] and belong to three structural protein classes: Coiled Coil, all-Beta, and alpha + beta.

Results reported in Table 3.

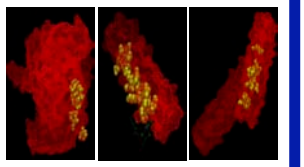


Figure 2. Reported Conserved Residues for Proteins of CKAAPS

The conserved residues reported in sample proteins 1BJ4(A), 1CZQ(A), and 1FE6(A) as shown in Figure 2. Outputs were generated using PDBSUM Astex Viewer.

Conclusion

In conclusion, the proposed tool provides an efficient means to integrate different scales for protein analysis. This study further reinforces, with newer evidence, that the identification of conserved hydrophobic residues is vital to the exposition of the folding of proteins and further aids in the functional annotation of proteins and possible mutational studies.

For further details of the proposed methodology and validations refer [7].

Reference

[1] A. Paiardini, F. Bossa and S. Pascarella, "Evolutionarily conserved regions and hydrophobic contacts at the super family level: The case of the fold-type I, pyridoxal-5'-phosphatedependent enzymes," *Protein Sciences*, vol. 13, pp. 2992-3005,2004.

[2] B.V.B. Reddy, W.W. Li, I.N. Shindyalov, and P.E. Bourne, "Conserved Key Amino Acid Positions (CKAAPs) Derived From the Analysis of Common Substructures in Proteins," *PROTEINS: Structure, Function and Genetics*, vol. 42, pp. 148-163, 2001.

[3] C.-J. Tsai and R. Nussinov, "Hydrophobic folding units derived from dissimilar monomer structures and their interactions," *Protein Science*, vol. 6, no. 1 pp. 24-42, 1997.

[4] U.K. Muppirala and Z. Li, "A simple approach for protein structure discrimination based on the network pattern of conserved hydrophobic residues," *Protein Engineering, Design & Selection*, vol. 19, no. 6, pp. 265-275, 2006.

[5] E.S. Huang, S. Subbiah and M. Levitt, "Recognizing Native Folds by the Arrangement of Hydrophobic and Polar Residues," *Journal of Molecular Biology,* vol. 252, pp. 709-720, 1995.

[6] K.M. Biswas, D.R. DeVido, and J.G. Dorsey, "Evaluation of methods for measuring amino acid Hydrophobicities and interactions," *Journal of Chromatography A*, vol. 1000, no. 1, pp. 637-655, 2003.

[7] Pradeep Chowriappa, Sumeet Dua, Jinko Kanno, Hilary W. Thompson, "Protein Structure Classification Based on Conserved Hydrophobic Residues," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (to appear)*.

Connections with CyberTools

This work relates to the WP1 (Data Services), especially towards meeting its aims for design and development of tools for metadata extraction and data mining services. The work is also connected to WP2 (Information Services) of CyberTools development with regard to information discovery algorithms. The work is a result of collaboration between investigators from Louisiana Tech University and Louisiana State University Health Sciences Center at New Orleans. While the work has demonstrated evaluation studies on restricted sets of proteins for performance comparison purposes, it is easily extendible to a variety of proteins available from the Protein Data Bank and other databases as they are discovered and made available.

Acknowledgements

-NSF/LA-BoREPSCoR program
-NIH/INBRE program

Data Mining Research Laboratory

Design and Performance Analysis of a Distributed HPC Molecular Dynamics Code on Distributed Resources

Promita Chakraborty, Shantenu Jha

Center for Computation & Technology, and Department. of Computer Science

Louisiana State University

Baton Rouge, Louisiana 70803, USA. {promita, sjha}@cct.lsu.edu

Abstract: Molecular Dynamics (MD) is an important step towards understanding the behavior of biological systems. Even though, the state-of-the-art computational resource and techniques have made it feasible to do large-scale MD simulations, MD remains a very compute-intensive job, especially for large realistic systems. Thus there exists interest in developing and understanding the performance of MD code on multiple distributed HPC resources. To realize this goal, we have developed a preliminary parallel and distributed MD code and benchmarked it on the LONI (Louisiana Optical Network Interface) grid, which uses dedicated light-path networks to connect supercomputers across Louisiana. This work is motivated by an attempt to utilize new infrastructure and to devise new programming strategies for MD simulations, with a focus on distributing the workload across various machines while retaining the advantages of parallelization within a single machine. Current practice is to distribute the job amongst various machines only when the resource requirement is more than the capacity of a single machine. In our work, we divided the job into several workloads, even if a single machine was capable of handling it. We tested our developed code on upto three distributed resources of LONI, namely Bluedawg, Zeke and Ducky. These are IBM P5 clusters with 114 nodes per cluster. Based on performance data, we show that without any serious optimization, the performance degradation as defined by total CPU-hrs on multiple machines is about 10-20% of the performance over a single machine, which has useful consequences when time to finish is critical. Based on this analysis, the users of grid resources can pick their choice between two different strategies: (1) optimize the CPU-hours used, and live with the huge wait time involved, (2) or use an extra 10-20% CPU-hours and optimize the overall job throughput. Our analysis has the potential to be extended to parallel codes other than MD. We believe that this study can lead to a better job distribution and resource allocation to optimize throughput.

Molecular Dynamics

- Atoms stretch, vibrate, and rotate about the bonds in response to intermolecular and intra-molecular forces.
- Involves both bonded and non-bonded forces
- Equations to determine velocity and displacement:
  - $(V_i)_{t+1} = (V_i)_t + (F_i / M_i)_t$
  - $(X_i)_{t+1} = (V_i)_t + (X_i)_t$

The parallel MD code

- Extended on the existing Mindy code for MD
- Computation of non-bonded forces is the most compute-intensive part, as seen from profiling of the code
- Non-bonded part takes 72-80% of computation time
- Bonded parts (to compute bond-angles, improper and dihedral angles) takes 3-5% of computation time

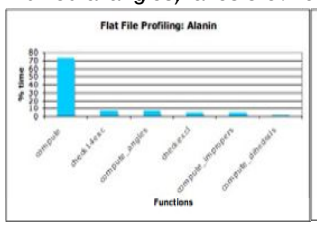


Fig: Flat-file profiling (Alanin)

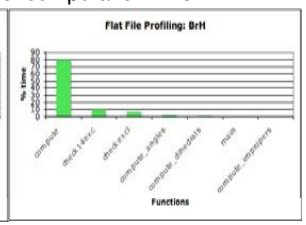
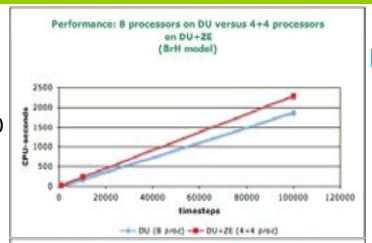


Fig: Flat-file profiling (BrH)

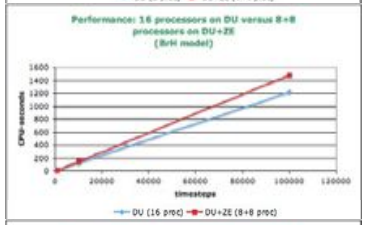
Performance Analysis

(a)



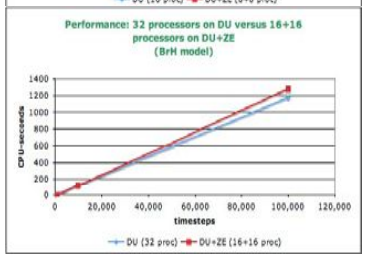
timesteps	Total CPU-hrs (sec) DU (8 proc)	DU+ZE (4+4 proc)
100	1.5	1.7
1,000	19.5	23.8
10,000	191.5	234.3
100,000	1882.7	2308.1
1,000,000		28132.5

(b)



timesteps	Total CPU-hrs (sec) DU (16 proc)	DU+ZE (8+8 proc)
100	0.0	0.0
1,000	11.1	13.4
10,000	123.7	149.6
100,000	1219.0	1481.9
1,000,000		57494.3

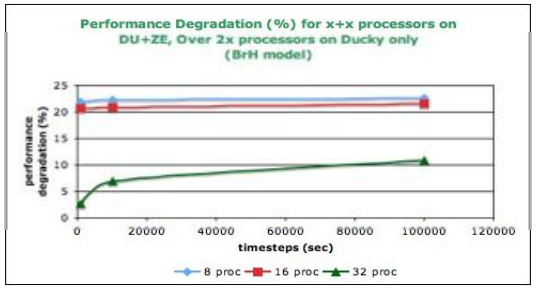
(c)



timesteps	Total CPU-hrs (sec) DU (32 proc)	DU+ZE (16+16 proc)
100	0.0	0.0
1,000	3.7	3.8
10,000	116.0	124.0
100,000	1175.5	1282.7
1,000,000	11977.0	13276.9

Fig: Performance degradation when distributed across 2 LONI machines is about 10-20%; (a), (b), (c) shows actual values; (d) shows percentage-comparison; BL, DU, ZE are LONI machines, representing Bluedawg, Ducky, Zeke.

(d)



The graphs and tables shown depicts performance for Bacteriorhodopsin protein dataset. Similar performance is seen on Alanine protein dataset - a much smaller size dataset.

Conclusion

- Performance degradation of compute-intensive code like MD is 10-20% of the performance over single machine
- Hence, users can choose any one:
  - optimize CPU-hours used, and try to get maximun resource over a single machine
  - optimize job throughput, because of lesser wait time, by using extra 10-20-% CPU-hours

Test-sets: Proteins

- **Alanin** (Alanine)
  - size 66 atoms
- **BrH** (Bacteriorhodopsin)
  - size 3762 atoms

Alanine

Source: www.wiki.org

BrH

Source: www.pdb.org

Test-bed: LONI (Louisiana Optical Network Initiative)

LONI is nation's one of the first network to connect Louisiana's supercomputers with fast light-paths so as to minimize communication delay.

We used LONI because, we wanted to know:

- The behavior of a distributed code, assuming minimum communication delay possible
- Exactly performance degradation a code can have, when run in distributed mode in supercomputers
- We used 3 LONI machines: Bluedawg, Ducky, Zeke
- LONI uses Loadleveler for job scheduling, which implements the backfill algorithm on FIFO scheme.

Acknowledgements / Sponsorships

This research has been supported by:
- Center for Computation and Technology (CCT)
- Cybertools Project