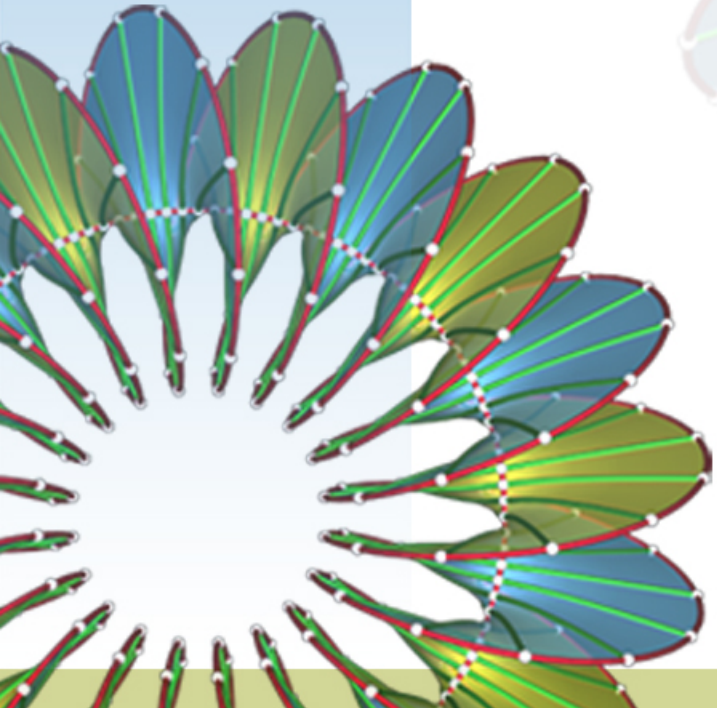


Software at NSF/OCI

Daniel S. Katz

Program Director, Office of Cyberinfrastructure

Create and maintain a software ecosystem providing new **capabilities** that advance and accelerate scientific inquiry at unprecedented complexity and scale

Support the foundational **research** necessary to continue to efficiently advance scientific software

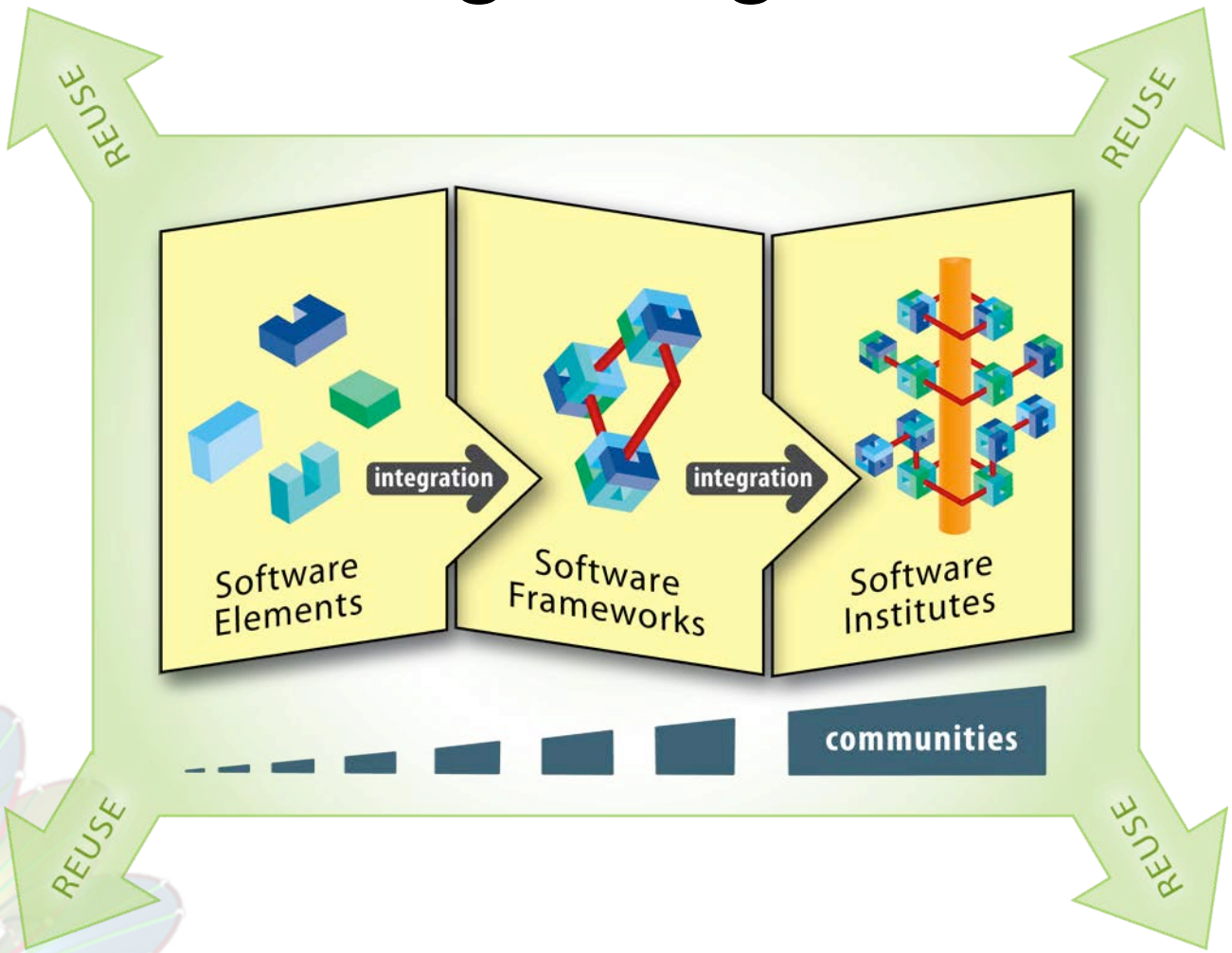
Enable transformative, interdisciplinary, collaborative, **science and engineering** research and education through the use of advanced software and services

Transform practice through new **policies** for software addressing challenges of academic culture, open dissemination and use, reproducibility and trust, curation, sustainability, governance, citation, stewardship, and attribution of software authorship

Develop a next generation diverse workforce of scientists and engineers equipped with essential skills to use and develop software, with software and services used in both the research and **education** process

NSF Software Vision: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf12113

Putting it together



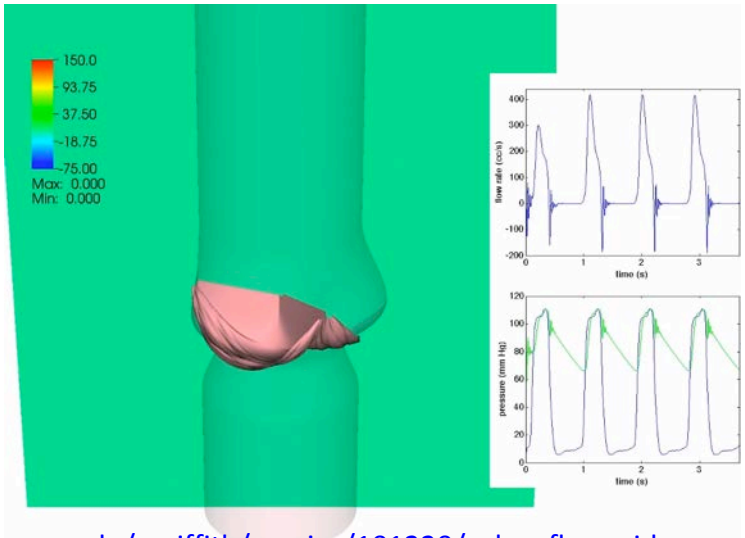
Look at NSF 11-539 and 11-589 for past SI^2 solicitations; current NSF 12-576; new solicitations coming; consider supplements, EAGERs, etc.

SI^2 Solicitation and Decision Process

- Cross-NSF software working group with members from all directorates
- Discusses solicitations, determines who will participate in each
- Discusses and participates in review process
- Work together to fund worthy proposals

- How does SI^2 fit with other NSF programs that support software?
  - See: Implementation of NSF Software Vision - http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504817

SSE: Parallel and Adaptive Simulation Infrastructure for Biological Fluid-Structure Interaction – PI: B.E. Griffith, NYU School of Medicine



http://cims.nyu.edu/~griffith/movies/101220/valve_flow_side.mov

Goals

Many problems in **biological fluid dynamics** involve **fluid-structure interaction (FSI)** and require localized regions of **high spatial resolution**.

Project aims to enhance and extend the IBAMR software framework for biological fluid dynamics.

IBAMR provides parallel implementations of the **immersed boundary (IB)** method for FSI and extensions of the IB method that support an **adaptive mesh refinement (AMR)** approach to providing high spatial resolution.

IBAMR leverages high-quality open-source libraries (e.g. SAMRAI, PETSc, `libMesh`).

Impact

Primary product is the open-source IBAMR software.

Project aims to extend types of problems to which IBAMR may be applied (e.g. to nanoscale systems that require accounting for thermal fluctuations).

Project also aims to improve performance (e.g. via improved solver methods and implementations).

Modular design facilitates third-party extensions.

Hosted at http://ibamr.googlecode.com.

Status

Present distribution is via open `svn` repository; plan formal release later in Fall 2012.

Actively used for a variety of projects at NYU and at Northwestern, Tulane, UCLA, U. Cincinnati, U. Glasgow (UK), U. North Carolina-Chapel Hill, U. Utah, and others.

Beginning work with Kitware, Inc.

Project Team

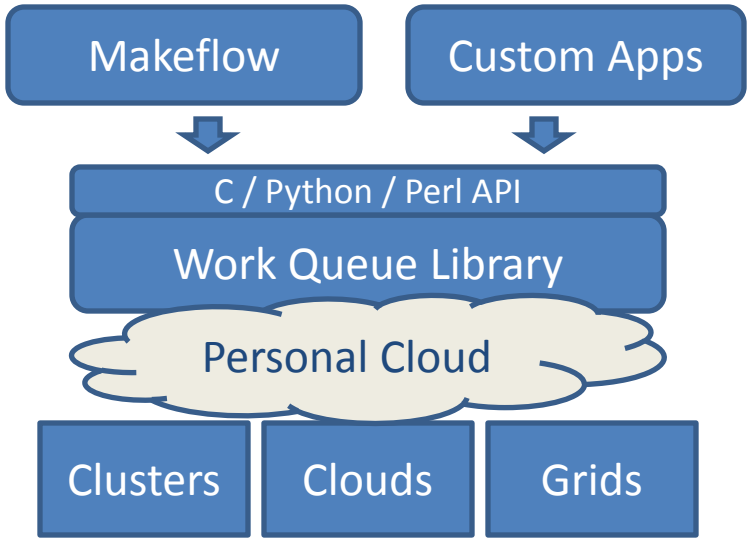
A. Donev, B.E. Griffith, D.M. McQueen, C.S. Peskin (NYU); A.P.S. Bhalla, N.A. Patankar (Northwestern); R. Ortiz (Kitware, Inc.)

NYU School of Medicine

Prepared: 9/20/2012

SI2-SSE: Bridging Cyberinfrastructure with the Cooperative Computing Tools

PI: Douglas Thain, University of Notre Dame



Goal

Enable researchers to quickly create scalable applications that harness all available clusters, clouds, and grids.

Users are primarily researchers in bioinformatics, molecular dynamics, and high energy physics.

Open source software (GPL) released 3 times/year, runs on Linux/Mac/Windows and tested on 40+ platforms using the NSF/UW Build and test facility.

Community engagement through annual user group meeting at ND, remote video presentations, tutorials at conferences, and active mailing list.

Measure of success: Publications describing fundamentally new results made possible

Components and Impact

- **Makeflow** - Make-based workflow system use for portability by bioinformatics community.
- **Work Queue** – API and runtime for creating dynamics scalable applications used in the molecular dynamics community.
- **Parrot** – Virtual filesystem for accessing remote data from HDFS, iRODS, FTP, used by CMS to access software and data over the wide area.
- **Chirp** – Personal file server for exporting and sharing data over WAN, used by the ATLAS expt.

Status

Version 3.6.0 available at www.nd.edu/~ccl

New features selected with community input and released 3 times/year.

Five outreach events so far in 2012, we would be happy to schedule one upon request.

Transitioning to public repository, bug tracking, etc, to better involve community.

Project Team

PI: Douglas Thain

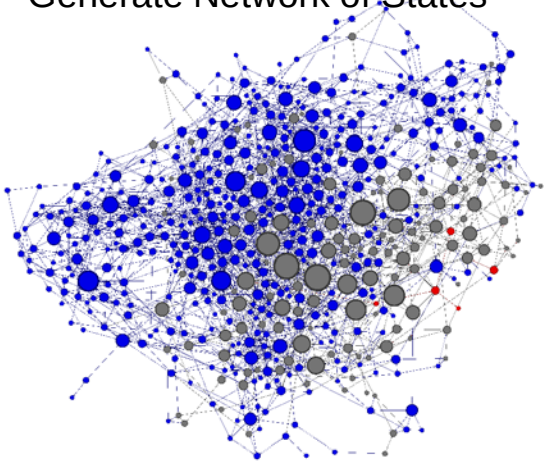
Graduate Students: Michael Albrecht, Patrick Donnelly, Dinesh Rajan, Li Yu

Prepared: 25 Sep 2012

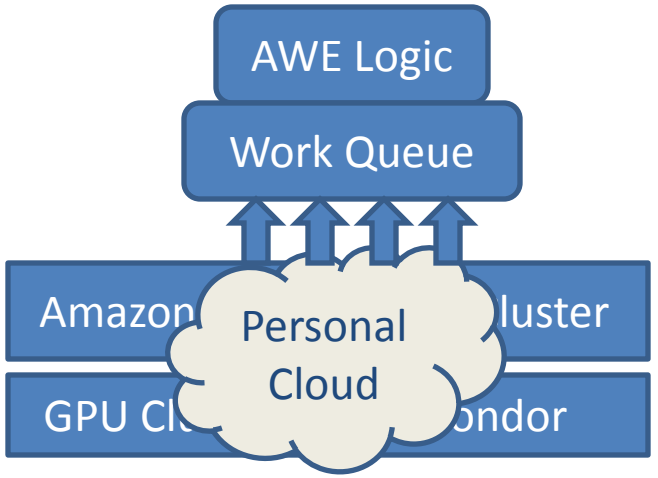
Example: Adaptive Weighted Ensemble

1: Start with Standard MD Simulation

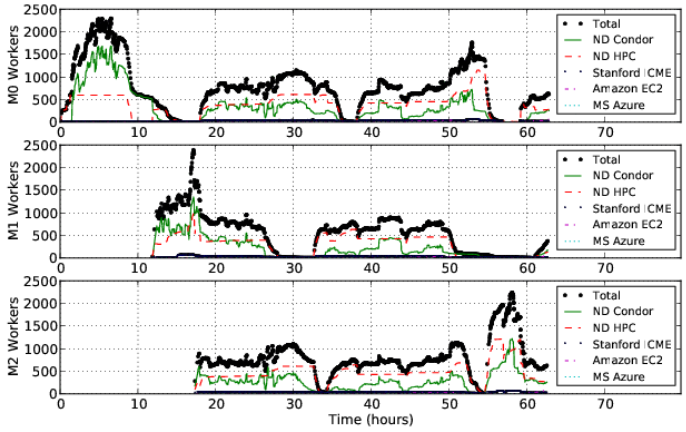
2: Research Goal: Generate Network of States



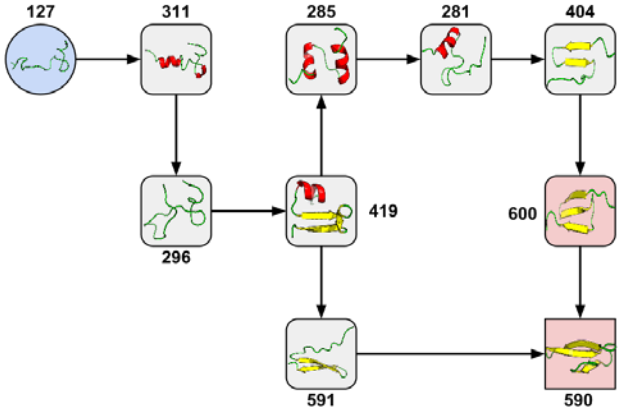
3: Build AWE Logic Using Work Queue Library



4: Run on 2000+ CPUs/GPUs at Notre Dame, Stanford, and Amazon



5: New State Transition Discovered!

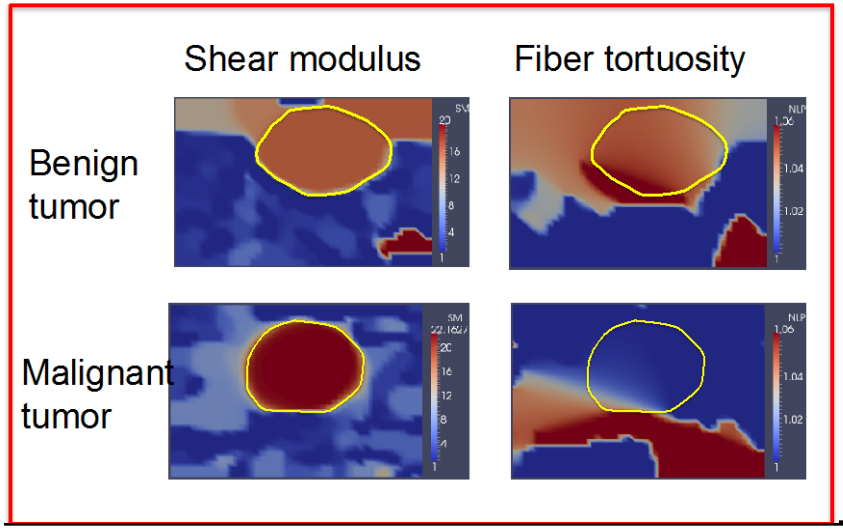


Badi Abdul-Wahid, Li Yu, Dinesh Rajan, Haoyun Feng, Eric Darve, Douglas Thain, Jesus A. Izaguirre, **Folding Proteins at 500 ns/hour with Work Queue**, IEEE e-Science Conference, 2012

SSE Advanced Software Infrastructure for Biomechanical Inverse Problems

Assad A Oberai (RPI), Chris Carothers (RP) and Paul E Barbone (BU)

In-vivo Biomechanical Images



Goal

In-vivo assessment of tissue microstructure **via multiscale inverse** analysis.

Applications: diagnosis of cancerous tumors, wound healing, vulnerable plaque, etc.

Key features: Modular software for different material types, inverse analysis, homogenization and robust optimization techniques.

Measurement of success: number of biomedical labs using NLACE and the breadth of applications.

Impact

NLACE: A code for creating biomechanical images through inverse analysis.

To be used for disease identification and biomechanics/mechanobiology.

NLACE can be extended to solve other material parameter identification problems: optical, electrical and thermal.

Status

Software releases: planned in summer 2013

Number of labs using NLACE: Seven

Application areas: cancer diagnosis, vulnerable plaque, gel dosimetry, wound healing, bone health and cellular imaging.

Project Team

Graduate students: Tom Seidl, Constance Liu, and Mohit Tyagi

International partners: 1

09/25/2012

Imaging Tissue Microstructure

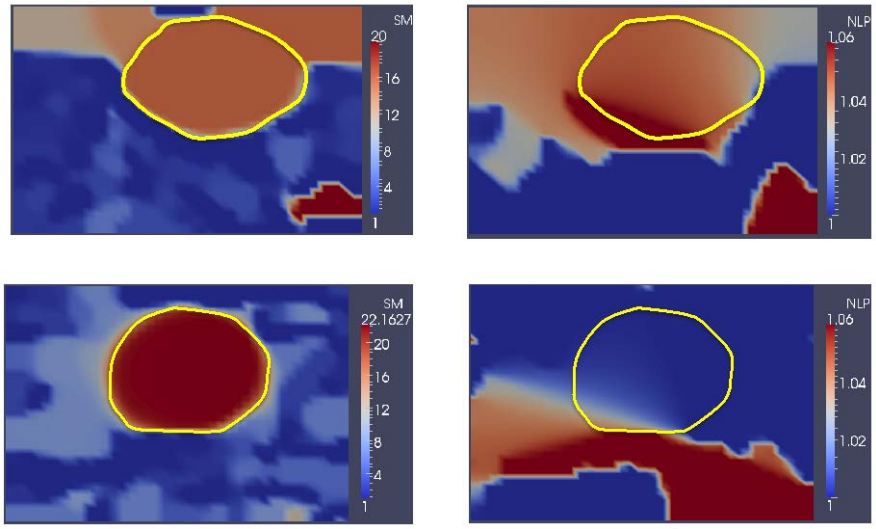
In-vivo Biomechanical Images

Shear modulus

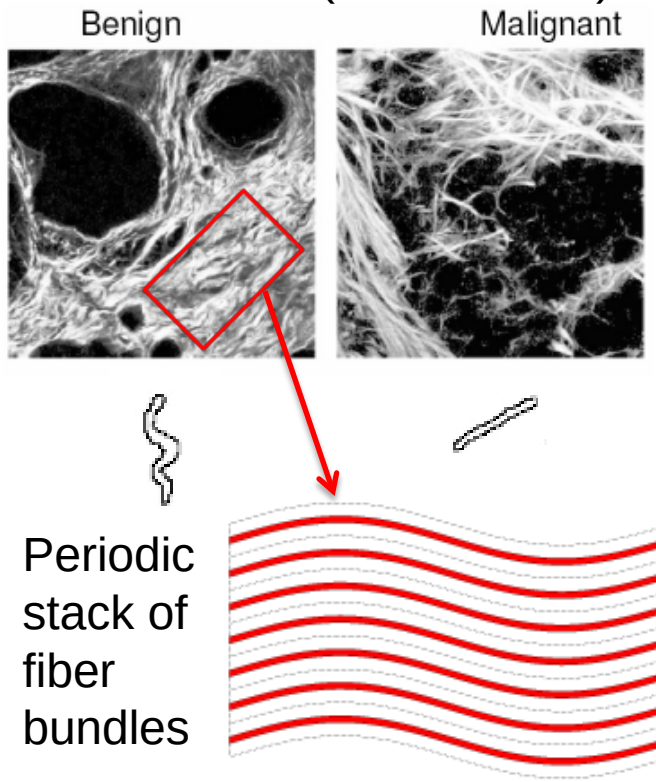
Fiber tortuosity

Benign tumor

Malignant tumor



Collagenous breast tumor microstructure (Falzon et al.)



Displacement measurements on patients

NLACE

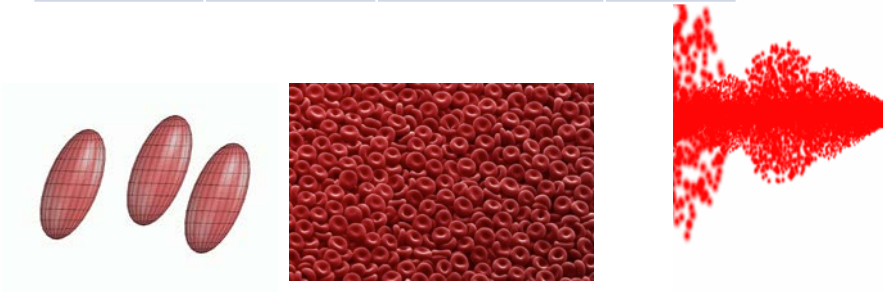
Homogenized Constitutive Model

SSE: Software for integral equation solvers on heterogeneous architectures

George Biros *(UT Austin)*, Denis Zorin *(NYU)*

Target App: Microcirculation of blood flow

RBCs	points/cell	unknowns/core	cores
50M	84	~450K	196K



simulate accurately plasma and cells

Goal

Primary applications: blood flow simulation in small capillaries; complex fluids; microfluidics.

Such problems require fast multipole methods (FMM) and boundary integral equation solvers (BIE). We are developing new parallel algorithms for these computational kernels.

We have designed extensive verification tests for each of the main computational kernels of the software. Algorithmic robustness to near collisions and long time integration.

Impact

Massively parallel FMM code (pKIFMM) @ padas.ices.utexas.edu; BIE code (MoBo) in development.

FMM is applicable to many areas of science: plasma physics, complex fluids, scattering, protein electrostatics, elastodynamics, and electromagnetics.Several groups in national labs and universities are currently using pKIFMM.

Status

pKIFMM: released;
MoBo: (BIE) is under development

Project Team

UT Austin: Bryan Quaife(postdoc), Logan Moon (research scientist), Dhairya Malhotra (PhD student)

NYU: Eduardo Corona (PhD student)

Sep/28/2012

Algorithmic challenges

- Complex geometry
- Nonlinearity
- Algorithmic parallel scalability

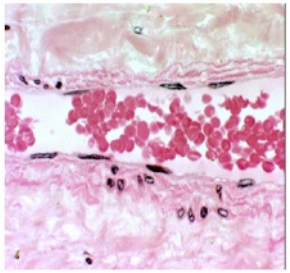
Implementation challenges

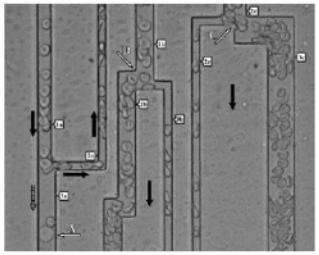
- Multiple phase algorithms
- Manycore architectures

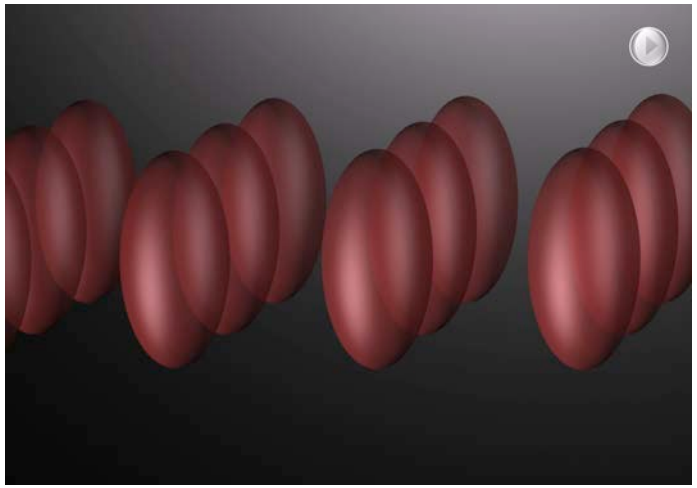
Software challenges

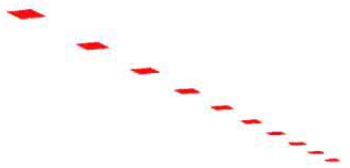
- Verification and validation
- Code complexity
- Sensitivity to problem setup

  limits generality



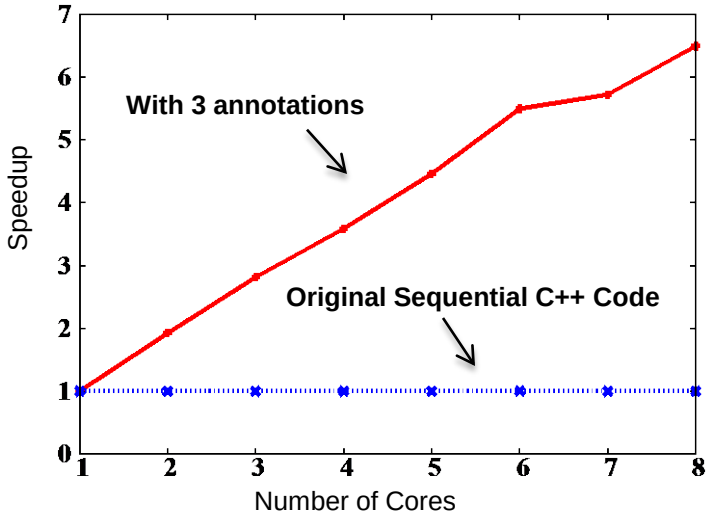






Accelerating Research with Implicitly Parallel Programming

PIs: David August and David Walker, Princeton University



Speedup of Parasite Simulation (2402 lines of code)

- Goal: Accelerate the pace of scientific research
- How: A high-level programming platform to achieve scalable performance with minimal effort
- Target: Existing long running sequential scientific applications
- Metrics: Performance and productivity

Impact

- Increased performance
- More time doing science, less time programming
- Improve how scientists use parallel machines

Status

- Problem identified: 114 scientists surveyed
- New implicitly parallel system designed
- System tested on few applications from the field
- Delivered improved performance by an order of magnitude

Project Team

- D. August and D. Walker (PIs)
- N. Johnson, P. Prabhu, J. Huang, H. Kim, S. Beard, T. Oh, F. Liu (graduate students)

09/27/2012

The Old Approach

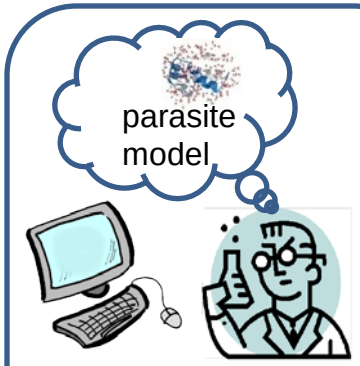
pthreads

MPI

"Explicit parallel programming too hard"

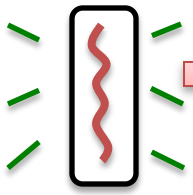
Unhappy Scientist

Implicit Parallel Programming Workflow



"If my simulations ran twice as fast, I will graduate a year earlier"

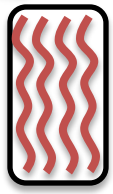
1 Scientific insights expressed



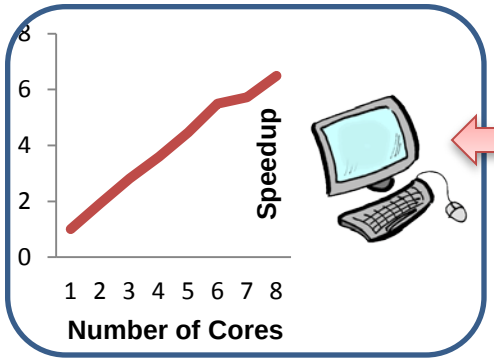
Sequential code

Parallelizing compiler

2 Automatic parallelization



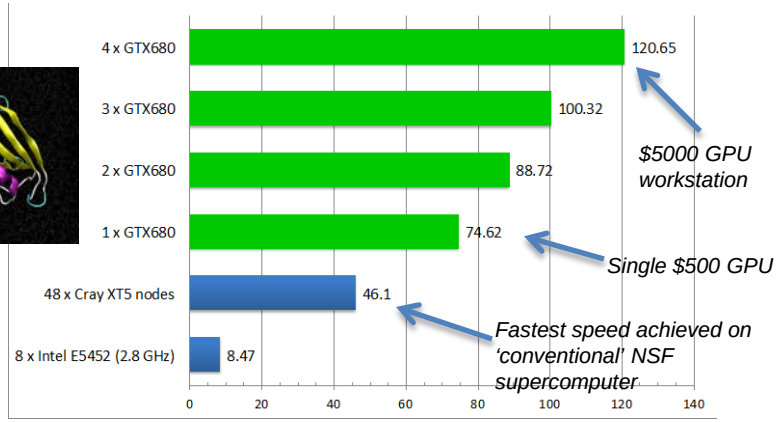
Parallelized Code



3 Scalable speedup achieved on real hardware

SSE: Sustained Innovation in Acceleration of Molecular Dynamics on Future Computational Environments: Power to the People in the Cloud and on Accelerators

Ross C. Walker (SDSC/UCSD) & Adrian E. Roitberg (UFL)



Throughput in ns/day for molecular dynamics simulation of DHFR (23K atoms) on CPUs (blue) and GPUs (Green)

Goals

Bring previously unattainable performance in Molecular Dynamics simulations to the graduate student's desktop.

Supercomputers are oversubscribed and do not lend themselves to 'experimentation' – This work brings supercomputer performance to the workstation transforming the way in which Molecular Dynamics simulation can be used in Molecular Biology research?

The key aspects of the engineering process is the development of general purpose GPU/MIC kernels that are high performing. Testing and validation is critical. Any approximations made must be extensively shown to have no impact on the results. New hybrid precision models allow us to accelerate MD simulations without sacrificing accuracy.

Success is measured by performance achieved, usage of the software and citation of the software in the literature.

Impact

Acceleration of all major aspects of condensed phase molecular dynamics simulation. http://ambermd.org/gpus/

Ubiquitous support for real time MD simulation on elastically scalable cloud resources. Interactive drug discovery for all. http://aws.amazon.com/gpu/

The impact on the Molecular Dynamics community is huge. Previously unprecedented simulation timescales are now possible on desktop resources.

Interactive simulation allows for experimentation on the fly, public interaction with all atom simulation of enzymes allows explorative education of how enzymes work.

Status

AMBER / AmberTools v11 with GPU support released Apr 2010. Major GPU update released May 2011.

AMBER / AmberTools v12 released Apr 2012. Major Update for NVIDIA Kepler released Aug 2012.

Citation count in excess of 1,500 per year / Over 12,000 downloads of AmberTools V12 from unique IPs. Over 600 sites using AMBER.

Project Team

Dr. Romelia Salomon (postdoc, UCSD), Mr. Jason Swails (Grad Student, UFL), Scott Le Grand (Senior Scientist, Amazon Web Services)

Industry Partners: NVIDIA / Microsoft / Intel / Amazon

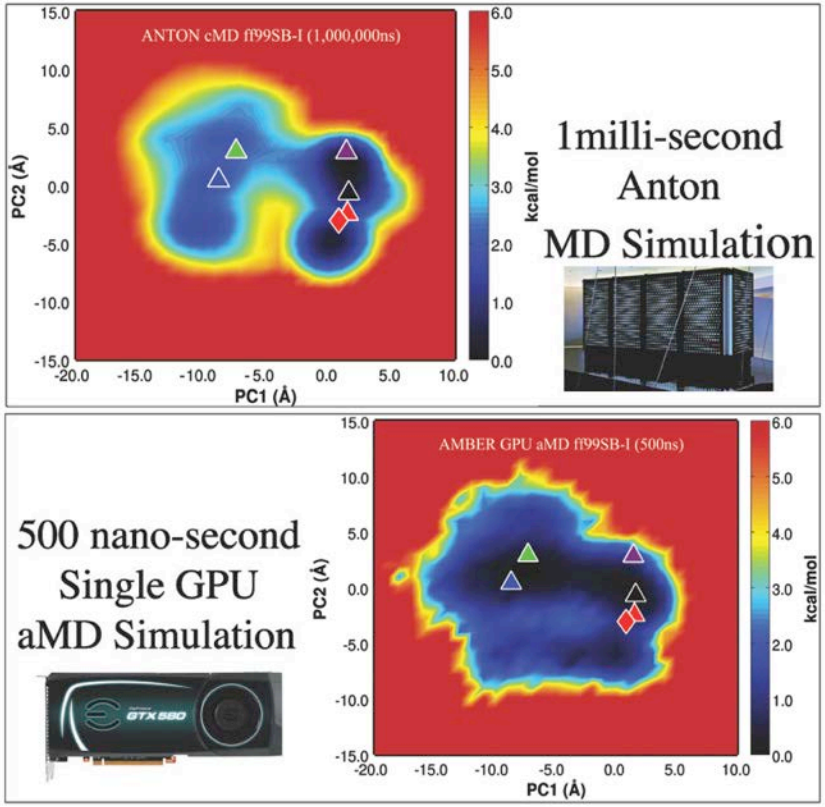
NVIDIA Microsoft amazon web services intel

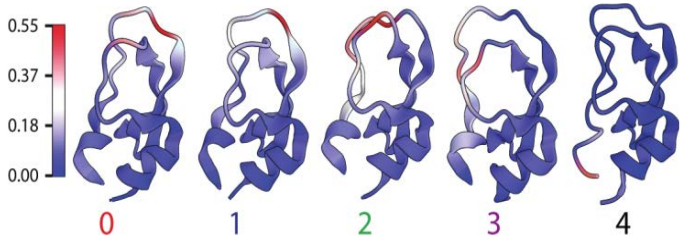
NSF

10/1/2012

Extreme Acceleration of Molecular Dynamics Simulations: Sampling for the 99%

- Pierce, L.C.T., Salomon-Ferrer, R. de Oliveira, C.A.F. McCammon, J.A. Walker, R.C., "*Routine access to millisecond timescale events with accelerated molecular dynamics.*", Journal of Chemical Theory and Computation, **2012**, in press, DOI: 10.1021/ct300284c





SI2-SSE funded accelerated MD simulation running for 1 week on $500 GPU samples the same phase space as 1 month long 1ms conventional MD simulation on multi-million dollar custom built ASIC.

SI2-supported Software

- IBAMR: An adaptive and distributed-memory parallel version of the immersed boundary (IB) method
- AmberTools
- Amber
- VOStat Web computing
- Eclipse Parallel Tools Platform
- SciDB
- CyberGIS Gateway
- Performance Analyzer for R programs
- R Virtual machine
- METIS - Serial Graph Partitioning and Fill-reducing Matrix Ordering
- ParMETIS - Parallel Graph Partitioning and Fill-reducing Matrix Ordering
- hMETIS - Hypergraph & Circuit Partitioning
- Scientific Cloud Computing Virtual Platform
- Scientific Cloud Computing Interface Tools
- Java GUI for the FEFF9 Spectroscopy Code
- Silver Extensible Meta-Programming Language
- Copper Parser and Context-Aware Scanner Generator
- Parallel Kernel Independent Fast Multipole Method
- Massively Parallel Quantum Chemistry (MPQC)
- Integral Library
- General Atomic and Molecular Electronic Structure System (GAMESS)
- ORCA
- MADNESS
- Automated web-based carbon-data ingestion tool
- Parallel implementation of linear geostatistical inverse problem algorithm
- Massive matrix multiplication algorithm
- Spatiotemporally aggregated uncertainty quantification for linear Bayesian inverse problems algorithm
- Lattice QCD library in CUDA for GPUs
- GenoSets: Visualizing Comparative Genomics
- Elemental distributed-memory dense linear algebra
- FLAME dense linear algebra library
- Berkeley Open Infrastructure for Network Computing (BOINC)
- Sage
- Sage-combinat
- Sigiri Task Management for Clouds
- Karma Provenance Capture
- In-Situ Partial Order Analyzer
- Distributed Analyzer of MPI Programs
- Symbolic GPU Program Analyzer
- Prover of User GPU Programs
- Dynamic Verifier of C Pthread Programs
- Graphical Explorer of MPI Programs
- Eddy Murphi - Parallel and Distributed Murphi
- Extensible Utah Multicore
- Differential Geometry Software Project
- OpenUH Compiler
- Abstract Data and Communication Library
- TAU Performance System
- Program Database Toolkit
- Ocelot Infrastructure (for architecture emulation and dynamic translation)
- EVPath event transport middleware
- ADIOS adaptable IO System (maintained and distributed by ORNL)
- Path integral Quantum Monte Carlo
- CUAHSI Hydrologic Information System
- HydroDesktop: CUAHSI Hydrologic Information System Desktop Application
- HydroServer: CUAHSI Hydrologic Information System Server
- Block: DMRG for quantum chemistry
- Cooperative Computing Tools
- User Level Cache Control in Multicores
- Balanced Work Stealing for Time-Sharing Multicores
- High Performance MPI for InfiniBand, IWARP and RoCE
- Pegasus Workflow Management System
- Swift Parallel Scripting Language
- Rocks Cluster Toolkit
- Rocks Cluster Toolkit Source Code Repository
- Application Fingerprint Source Code Repository
- GridFTP data transfer server and client
- Grid Resource Allocation and Management (GRAM) server and clients
- Grid Security Infrastructure (GSI) security libraries
- Globus Online hosted research data management services
- Graph Annotation Format
- Differential Equations Analysis Library
- Advanced Solver for Problems in Earth ConvecTion (ASPECT)
- Trilinos