## SI2-SSI: Modules for Experiments in Stellar Astrophysics (MESA)

Frank Timmes (PI), Rich Townsend (Co-PI), Lars Bildsten (Co-PI), Bill Paxton (MESA's First Author), Josiah Schwab (Collaborator), Pablo Marchant (Collaborator), Rob Farmer (Collaborator)





- Performance
   State-of-the-art graph processing library
- Generality
   Covers a broad range of graph algorithms
- Programmability
   Makes it easy to implement and easy to impleme

Makes it easy to implement and extend graph algorithms from **1-GPU to multi-GPUs** 

• Scalability

Fits in (very) limited GPU memory space performance scales when using more GPUs

Continuous Integration
 Continuous delivery powered by jenkins.io









# SinCardio Open-Source, Multi-Physics, Cardiac Modeling and Simulation



Alexander D. Kaiser, PhD, Cardiothoracic Surgery, Stanford; Vijay Vedula, PhD, Cardiology, Stanford;

Nathan M. Wilson, PhD, MBA, Open-Source Medical Software Corporation; Shawn C. Shadden, PhD, Mechanical Engineering, UC Berkeley; Ellen Kuhl, PhD, Mechanical Engineering, Stanford; Alison L. Marsden, PhD, Pediatrics and Bioengineering, Stanford





Major support for this work was provided by the NSF SI2-SSI Collaborative Research Program (Awards #1339824 and #1663671).

## HydroShare: Cyberinfrastructure for Advancing Hydrologic Knowledge through Collaborative Integration of Data Science, Modeling and Analysis

David Tarboton, Ray Idaszak, Shaowen Wang, Jeffery Horsburgh, Dan Ames, Martyn Clark, Jon Goodall, Alva Couch, Hong Yi, Tony Castronova, Christina Bandaragoda, Rick Hooper

Advancing Hydrologic Understanding

- requires integration of information from multiple sources
- is data and computationally intensive
- requires collaboration and working as a team/community

HydroShare is a system to advance hydrologic science by enabling the community to more easily and freely share products resulting from their research, not just the scientific publication summarizing a study, but also the data and models used to create the scientific publication.

- **F**indable
- Accessible
- Interoperable
- **R**eusable
- Open data
- Transparency
- Research Reproducibility
- Enhanced trust in research



- Includes a repository for users to share and publish data and models in a variety of formats
- Includes tools (web apps) that can act on content in HydroShare

## Neutrino Radiation Hydrodynamics in GenASiS

Reuben D. Budiardja, Eirik Endeve, Christian Y. Cardall, R. Daniel Murphy The University of Tennessee -- Oak Ridge National Laboratory



SI2-SSE: Collaborative Research: Integrated Tools for DNA Nanostructure Design and Simulation

Pls: Shawn Douglas (UCSF) and Aleksei Aksimentiev (UIUC)



# LROSE:

# Lidar-Radar Open Software Environment



- LROSE is a joint 4-year project between Colorado State University and the National Center for Atmospheric Research funded by NSF SI2-SSI to develop common software for the LIDAR, RADAR and Profiler community
- The 1st LROSE community workshop was held in April 2017. Focus on key applications as 'building blocks', allowing users to assemble trusted, reproducible workflows to accomplish more complex scientific tasks
- LROSE "Blaze" released in 2018 in a "Virtual Toolbox" Docker container or as C++ compiled native apps focused on data Conversion, Display, and Gridding. Additional tools for QC, Echo, and Wind analysis are in development
  - nsf-lrose.github.io



HawkEye Display of Hurricane Irma (2017)



## **The Fruits of Provenance**

Emery R. Boose, Aaron M. Ellison, Elizabeth Fong, Matthew Lau, Barbara S. Lerner, Jackson Okuhn, Thomas Pasquier, Margo Seltzer





## Harmonically Mapped Averaging for Everyone

## **Ensemble Averages**

- Core of statistical mechanics
- Relate molecule coordinate averages to material properties
- Example: Pressure tensor

 $\mathbf{P} = \rho kT \mathbf{I} + \frac{1}{3V} \left\langle \sum_{i < j} \mathbf{f}_{ij} \mathbf{r}_{ij} \right\rangle$ 

## **Mapped Averages**

- Writes averages rigorously as deviation from an approximate starting point
- For crystals, a good starting point is a harmonic lattice
- Example: internal energy

$$U = \frac{3}{2}NkT + \left\langle U_{\text{config}} + \frac{1}{2}\mathbf{F} \cdot \Delta \mathbf{r} \right\rangle$$

## Performance: Heat capacity



## SSE project: Implement in these codes



## Fast Dynamic Load Balancing for Extreme Scale Systems

Cameron W. Smith, Gerrett Diamond, M.S. Shephard

Computation Research Center, Rensselaer Polytechnic Institute

discrete

event

simulation

Dynamic load balancing is a core tool needed to support automated simulations

Goal: Generalize a multicriteria procedure to:

- Applications past conforming unstructured meshes
- Execute on accelerator supported systems
- Developing the EnGPar multicriteria partition improvement procedures based on a distributed N-graph





## **BIG WEATHER WEB**

Carlos Maltzahn, Ivo Jimenez (UC Santa Cruz), Mohan Ramamurty (Unidata), Gretchen Mullendor (UND), Brian Ancell (Texas Tech), William Capehart (SD Mines), Clark Evans (UW Milwaukee), Robert Fovell, Kevin Tyle (UAlbany), Steven Greybush (PennState), Russ Schumacher (Colorado State), Kate Fossell (NCAR), Joshua Hacker, John Exby (Jupiter Intelligence)



Points:

easily shared artifacts,

community-improved

over time

#### Large ensemble shared among 7 universities:

Gretchen Mullendore (UND), Brian Ancell (Texas Tech), William Capehart (SDSM), Clark Evans (UW Milwaukee), Robert Fowell (U Albany), Steven Greybush (Penn State), Russ Schumacher (CSU).



SI2-SSE: Analyze Visual Data from Worldwide Network Cameras (Continuous Analysis of Many CAMeras, CAM<sup>2</sup>), NSFACI-1535108 PI: Yung-Hsiang Lu, yunglu@purdue.edu <u>https://cam2.ecn.purdue.edu</u>



#### SUMMARY OF 2017-2018



REST interface

#### NEW PUBLICATIONS

- IEEE Transactions on Cloud Computing
- IEEE International Conference on Multimedia Information Processing and Retrieval 2018.
- ACM Multimedia 2017
- IEEE International Conference on Information Reuse 2017
- "Parallel Video Processing using Embedded Computers", IEEE Global Conference on Signal and Information Processing 2017
- "Creating the World's Largest Real-Time Camera Network", Imaging and Multimedia Analytics in a Web and Mobile World 2017
- "Internet of Video Things in 2030: a World with Many Cameras", IEEE International Symposium of Circuits and Systems 2017.

#### MEDIA COVERAGE



Technology

#### CAM2 COMPONENTS



#### UNDERSTAND DATA BIAS





CAM2

#### USC INRIA ImageNet

#### GLOBAL SCALE







#### USER RECRUITMENT AND OUTREACH

Vertically Integrated Projects Midwest Undergraduate Research Workshop



#### PURDUE ■ More: ■ More: Purdue-based company developing software to improve customer service receives \$750,000 grant WEST LAFAVETTE. Ind -A grand aligned by the Exceedimentation to purd Purdue-based company developing software to improve customer service receives \$750,000 grant

Collaborative Research: Building Sustainable Tools and Collaboration for Volcanic and Related Hazards Program: Software Infrastructure for Sustained Innovation

Abani Patra

An open source multi-physics platform to advance fundamental understanding of plasma physics and enable impactful application of plasma systems

**Davide Curreli and Steve Shannon** 

## Parsl: A Python-based Parallel Scripting Library

http://parsl-project.org

Yadu Babuji\*, Kyle Chard\*, Ian Foster\*, Daniel S. Katz°, Mike Wilde\*, Anna Woodard\*, Justin M. Wozniak\* \*Computation Institute, University of Chicago and Argonne National Laboratory °National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

**Goals:** Easily write Python workflows that glue together external programs and Python functions; run them quickly and easily in parallel on diverse resources

#### Simple installation:

pip install parsl

Annotate functions to make Parsl apps

- Bash apps call external applications
- Python apps call Python functions

Apps run concurrently respecting data dependencies via futures. Natural parallel programming!

Parsl scripts are independent of where they run. Write once run anywhere!

Features: Multi-site work distribution, automatic elasticity, Gobus data staging, resilience, containers.



return "echo \$(( ( RANDOM )) &> {outputs[0]}"

```
@App('python', dfk)
def total(inputs=[]):
  total = 0
  for i in inputs:
    with open(i, 'r') as f:
        total += sum([int(line) for line in f])
  return total
```

# Create 5 files with random numbers
output\_files = []
for i in range (5):
 output files.append(generate(outputs=['random-%s.txt' % i]))

```
# Calculate the sum of the random numbers
t = total(inputs=[i.outputs[0] for i in output_files])
print (t.result())
```



## MetPy - A Python GEMPAK Replacement for Meteorological Data Analysis

**Ryan May** 

### **STORM**: a Scalable Toolkit for an Open Community Supporting Near Real-time High Resolution Coastal Storm Modeling

Hartmut Kaiser<sup>1</sup>, Joannes Westerink<sup>2</sup>, Rick Luettich<sup>3</sup>, Clint Dawson<sup>4</sup>

<sup>1</sup>Louisiana State University, <sup>2</sup>University of Notre Dame, <sup>3</sup>University of North Carolina, <sup>4</sup>University of Texas at Austin



## SI2-SSE: 3DSIM: A Unified Framework for 3D CPU Co-Simulation

**ANKUR SRIVASTAVA** 

## SI2-SSI: Sustainable Open-Source Quantum Dynamics and Spectroscopy Software

**Xiaosong Li** 

## SI2-SSI: EVOLVE Open MPI for Next Generation Architecture and Applications

Resilience

**File Operations** 

Runtime



Architecture

Aware Collective

**Operations** 

Tools

integration

Augment Open MPI to evolve and adjust to the challenges of new hardware environments, and to maintain the same level of efficiency for parallel scientific applications in a quickly transitioning hardware ecosystem.

- Efficient threading, including for MPI + X programming models, high injection rate
- Support hybrid architectures with direct data movement to/from accelerators
- Event-driven, architecture-aware collective communications
- Highly efficient MPI I/O support via
   OMPIO
- Fast and scalable runtime (PMIx) providing reduced memory consumption, fast startup and resilient capabilities

George Bosilca, Aurélien Bouteiller, Thomas Herault, and Edgard Gabriel

Hybrid MPI

## SI2-SSI: TESSE



## Task Based Environment for Scientific Simulation at Extreme Scale

Robert J. Harrison, Mohammad M. Javanmard, George Bosilca, Thomas Herault, Damien Genet, Edward F. Valeev

## • Challenge:

- Execute dynamic algorithms over irregular data on extreme scale hybrid machines using a task-based runtime
- Guarantee performance portability & productivity

Stony Brook University



### Recent Development

#### New Programming Model: Templated Task Graph (TTG) General purpose programming model implemented in C++ Applications composed as graphs of templated Ops encoding DAG of tasks instantiated at runtime Abstracts out the details of the execution runtime (PaRSEC and MADNESS already supported) void Plus::op(const kevT &kev. baseT::input values tuple type &&t. Tasks are parameterized baseT::output\_terminals\_type &out) { (loop index, label of node, auto x = baseT::qet<0>(t): couple of indices) auto y = baseT::get<1>(t); Minimize the known ::send<0>((int) (kev), x+v, out): task graph z Match tasks with keys Plus plus: Explicit send / implicit Times times: r receive connect<0, 0>(&plus, &times);

TensorFlow for general-purpose workloads

# ASTROLABE

## SI2-SSE: Visualizing Astronomy Repository Data using WorldWide Telescope Software Systems

P. Bryan Heidorn (PI) and Gretchen Stahlman, University of Arizona School of Information Julie Steffen (Co-PI) and Tom Hicks, American Astronomical Society

NSF SI2-SSE: 1642446

## **Project Objectives**

- WorldWide Telescope visual front-end for Astrolabe repository of legacy data.
- Tools for processing raw data into Astrolabe and visualization formats needed by WWT.
- Integration of the Unified Astronomy Thesaurus (UAT) into both Astrolabe and WWT.
- Data manipulation tools necessary to use WWT as a first look in archive browsing and retrieval.
- Workshops to identify community need for years 2 and 3.

### Year 1 Development

Port of key WWT functions to web interface.
 ~20k users/month since Oct

### Year 2 Development Status

- Integrating software with CyVerse
- Community outreach
- Ingesting data and metadata

#### Year 2 Astrolabe Tools

- JS9 WWT
- Glue
- Jupyter Notebook
- Topcat
- Customized metadata template
- UAT keywords
- DataCite DOIs

### Year 2 Outreach

- Splinter session workshop held at American Astronomical Society annual meeting in January 2018
- Workshop and hackathon held at Biosphere 2 in March 2018
- Paper to be published in ApJS: "Astrolabe: Curating, Linking and Computing Astronomy's Dark Data"







http://astrolabe.arizona.edu https://osf.io/sp349/





## SI2-SSE: GenApp - A Transformative Generalized Application Cyberinfrastructure

3d plots, atomic structures

audio, video https://genapp.rocks

#### <u>E. H. Brookes</u><sup>a</sup>, J. E. Curtis<sup>b</sup>, D. Fushman<sup>c</sup>, S. Krueger<sup>b</sup> and A. Savelyev<sup>a</sup>

<sup>a</sup>Department of Biochemistry & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, Texas <sup>b</sup>NIST Center for Neutron Research, Gaithersburg, Maryland <sup>c</sup>Department of Chem. & Biochem., University of Maryland, College Park, Maryland

> virtual cluster

node

virtual

cluster

node

virtual cluster node

virtual

cluster

node

#### Multiple Execution models e.g. elastic via Openstack

**CCP-SAS** 

**EPSRC** 







GenApp

## Application Characterization for Adaptive Computing Platform Determination for Computational and Data-Enabled Science and Engineering

Haiying Shen, Associate Professor University of Virginia, USA hs6ms@virginia.edu Walter B. Ligon, Associate Professor Clemson University, USA walt@clemson.edu

#### MOTIVATION



Hadoop Distributed File System (HDFS)

A Hadoop cluster



- Hadoop platforms with local storage and dedicated storage
- Different applications may benefit differently from the two platforms

### SOLUTIONS

- Determine thresholds to decide whether use:
  - local storage or remote storage
  - scale-up or scale-out nodes



#### PROJECT

- Application performance comparison (I/O, data, and CPU intensive)
- Best platform determination
- Adaptive hybrid Hadoop platform construction

1

0.8

0.6

0.4

0.2

CDF

#### **OBSERVATIONS**

- Remote file system Orange File System (OFS) outperforms local file system HDFS when the data size is large due to the fast I/O.
- HDFS outperforms OFS when the data size is small due to the network latency.
- Scale-up machines are better for small jobs but not large jobs.
- Scale-out machines bring more benefits to process a larger amount of data than scale-up machines.

#### RESULTS

-Hvbrid

----- THadoop

-----RHadoop

Hybrid: hybrid scaleup/out cluster with OFS THadoop: traditional scaleout Hadoop with HDFS RHadoop: traditional scaleout Hadoop with OFS





10 20 30 40 50 60 70 80 90



## **Software Elements to Enable Immersive Simulation**

Corey Nelson, Felix Newberry, John A. Evans, Kurt Maute, Alireza Doostan, <u>Kenneth E. Jansen</u> Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado, Boulder, CO 80309 **2018 NSF SI<sup>2</sup> PI Meeting** 



Acknowledgments: Funded by SI2 program of the NSF OAC-1740330. Collaboration of <a href="https://github.com/SCOREC/core">https://github.com/SCOREC/core</a> and <a href="https://github.com/PHASTA">https://github.com/SCOREC/core</a> and <a href="https://github.com/PHASTA">https://github.com/SCOREC/core</a> and <a href="https://github.com/PHASTA">https://github.com/SCOREC/core</a> and <a href="https://github.com/PHASTA">https://github.com/SCOREC/core</a> and <a href="https://github.com/PHASTA">https://github.com/PHASTA</a>.



## A Data-Centric Approach to Turbulence SImulation

Eric Peters, Riccardo Balin, Ryan Skinner, John A. Evans, Philippe R. Spalart, Alireza Doostan, <u>Kenneth E. Jansen</u> Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado, Boulder, CO 80309

2018 NSF SI<sup>2</sup> PI Meeting

- Separating turbulent boundary layers are poorly predicted by Low Fidelity Models (LFM)
- High fidelity models can predict them but at a cost far too high for design or uncertainty quantification (UQ)
- Goals of this project are:
- Leverage Multi-Fidelity Modeling (MFM) to accelerate confident design space exploration
- Contribute to Direct Numerical Simulation (DNS) data base of separating turbulent boundary layers
- Use Machine Learning (ML) to improve LFM turbulence modeling closures from DNS data
- <u>https://github.com/PHASTA</u>





NSCI SI2-S2I2 Conceptualization of *CFDSI*: Model, Data, and Analysis Integration for End-to-End Support of Fluid Dynamics Discovery and Innovation

<u>Kenneth E. Jansen</u> Jed Brown, Alireza Doostan, John A. Evans, John A. Farnsworth, Peter E. Hamlington, and Kurt K. Maute

Beverley J. McKeon

Caltech

Robert D. Moser

Rensselaer Mark S. Shephard, Onkar Sahni, and Cameron Smith



Acknowledgments: Funded under the SI2 program of the National Science Foundation OAC-1743178. https://www.colorado.edu/events/cfdsi/

## Extending the physics reach of LHCb in Run 3 using machine learning in the real-time data ingestion and reduction system

The LHCb detector is being upgraded for Run 3 (2021-2023), when the trigger system will need to process 25 exabytes per year.

- Currently, only 0.3 of the 10 exabytes per year processed by the trigger is analyzed using high-level computing algorithms; the rest is discarded prior to this stage using simple algorithms executed on FPGAs.
- To significantly extend its physics reach in Run 3, LHCb plans to process the entire 25 exabytes each year using software triggers running on a CPU farm. On average, this will require analyzing events 100 times faster than is possible today.

The primary objective of this project is to **expand the use of machine learning (ML) in the LHCb trigger**, to greatly improve its performance while satisfying its robustness and sustainability requirements. Specifically, ML algorithms will be developed to:

- replace the most computationally expensive parts the event pattern recognition;
- increase the performance of the event-classification algorithms; and
- reduce the # of bytes persisted per event without degrading physics performance.

#### Data-Driven Models for Predictive Molecular Simulations Atomic coordinates + quantum mechanical energies $+V^{NB}(1)$ PI: Paesani: Co-PIs: Götz & Zonca Permutationally Invariant Neural Networks $E = E[\rho]$ 1-bodv 2-hody Water Purification [Zn(l-L)(Br)]SCTT ZIDINA Catalysis density density density density isovalue: 0.01 A-3 isovalue: 0,001 A-3 isovalue: 0.0005 A-3 isovalue: 0.01 A<sup>-3</sup> functions of the types $\substack{j \neq i \\ k \neq i, j}$ $G_1^0$ $\mathbf{G}_1^{\mathbf{H}}$ $\mathbf{G}_m^{\mathbf{H}}$ $\xi = e^{-kR_{ij}}$ or $\xi = e^{-kR_{ij}}/R_{ij}$ Database 1B / 2B / 3B properties shown that our adQ tenas (Energies, charges, dipole morent $R_{ij}$ MB-nrg $V_{short}$ Optimizer ab initio **Database Objective function** 1B/2B/3Bpotential $Qutput|energy, \Delta\mu, \Delta V^{1B},$ tpu configurations generator $\Delta V^{2B}, \Delta V^{3B}$ $V^{2B/3B}$ XSEDE $\forall p d a te c_n P_r$ (c) (tar parameters **MB-nrg potential** V<sup>1B</sup>, V<sup>2B</sup>, V<sup>3B</sup>, V<sup>NB</sup> Initial configurations Initial **Biased MD** parameters configurations

SI2SSI: Collaborative Research: A Robust High-Throughput Ab Initio Computation and Analysis Software Framework for Interface Materials Science

Yifei Mo

## NSCI SI2-SSE: Multiscale Software for Quantum Simulations of Nanostructured Materials and Devices

- J. Bernholc, E. L. Briggs, W. Lu, C. T. Kelley, Z. Xiao, and J. Zhang North Carolina State University, Raleigh
- \* "Real" materials structures are often complex and cannot be reduced to a few hundreds of atoms
  - Process simulation requires large systems
- Materials Genome White House initiative to "deploy advanced materials twice as fast, at a fraction of the cost"
- National Strategic Computing Initiative (NSCI)
  - Pre-exascale, 150-300 pflops ~2018; Sustained exaflop ~2021
- Real-space multigrid (RMG) open-source software www.rmgdft.org
   High performance on supercomputers, clusters and desktops
- RMG-NEGF method for simulating nanoscale devices: I-V, gain, etc.



Novel negative differential resistance device

**NC STATE** UNIVERSITY



## The N-Jettiness Software Framework for Precision Perturbative QCD Calculations in Particle and Nuclear Physics

**Frank Petriello** 

- Python Analysis Infrastructure
  - Support static/dynamic/symbolic analysis, like LLVM for C/C++ static and KLEE for C
- Use in research
  - Enable researches like software engineering, programming languages, machine learning, AI and data science
  - Improve data processing effectiveness, reliability, stability and efficiency
  - Projects
    - data provenance tracking
    - white box tuning
    - data debugging
    - instability detection in floating point data processing
    - Synthesize workarounds fro cross-project bugs in data processing ecosystems
    - Intelligent debugging assistant by probabilistic inference
    - Automated programming assignment grading and synthesis

Andreas Stathopoulos, Computer Science, College of William & Mary

Applicability: Large matrices, a number of smallest / largest / interior e-values

- Theory: state-of-the-art, optimized, preconditioned methods
- Stability and robustness: solutions close to machine precision
- Tuned: full set of defaults and auto-tuning for end-users
- Flexible: full customizability for expert users
- Performance: HPC, parallel, block (tall skinny) methods, GPUs
- Interfaces: F77, MATLAB, Python, R, Julia, Nim

https://github.com/primme/primme



# AttackTagger: Early threat Detection for Scientific Cyberinfrastructure

bro IDS logs, etc.

Syslog, netflow, linux auditd,

- Employ factor graphs, a probabilistic graphical model, to capture attacker behavior and detect malicious activities
- Learning graph structure that represents dependencies and their strength among observed events and attack stages

Joint project: Cyber Security and Networking Division at the NCSA and the University of Illinois DEPEND Group <a href="http://depend.csl.illinois.edu/">http://depend.csl.illinois.edu/</a> <a href="http://depend.csl.illinois.edu/">http://security.ncsa.Illinois.edu/</a>



**Custom Data** 

 $e^2$ 

 $e^1$ 



# **Coastal Infrastructure Protection**



The entrance the 168th Street subway station in New York City. Taken by Seidenstud - 2006



Kyle T. Mandli

# **Computing Densities for Stochastic Differential Equations**

Harish S. Bhat, Applied Math, UC Merced (hbhat@ucmerced.edu)

Given an SDE  $dX_t = f(X_t; \theta)dt + g(X_t; \theta)dW_t$ would you like to compute its PDF? Estimate  $\theta$ ?

Density tracking by quadrature (DTQ) can help!

