Sustainable Research Pathways (SRP) Workshop 2021

STAFF PROJECTS

SUSTAINABLE HORIZONS INSTITUTE

BERKELEY LAB

TABLE OF CONTENTS

INTRODUCTION

The goal of the Sustainable Research Pathway (SRP) program is to foster collaborations between aspiring scientific professionals and Berkeley Lab staff by identifying shared interests for pursuing summer research experiences. In this booklet, you will find staff projects submitted by group leaders or staff scientists across the Computing Sciences area at Berkeley Lab. Keep in mind that the staff projects listed are not the only ongoing projects. Please visit the website links embedded in the group's name to find more information about additional areas. We welcome your collaboration ideas related to staff interests. Your interests could be directly related to projects described below and, on the website, or tangential, but there must be commonalities between your ideas and staff interests to match and move forward with the summer research experience. The collaboration should be equally beneficial to the program participant and Berkeley Lab staff.

Please read through the projects to find information such as the field or research area, abstract, and relevant skills, background, or interests. Projects have been listed in order of their respective departments. Feel free to email info@shinstitute.org if you have any questions or concerns.

APPLIED MATHEMATICS

1. Center for Computational Sciences and Engineering

Andy Nonaka, Staff Scientist

Multiphysics Modeling and Simulation

Field or research area: Applied Mathematics; Computational Science

Abstract: We have a variety of different multiphysics modeling projects in CCSE every summer, spanning a range of length and time scales. We model these phenomena using structured grid approaches and/or particles, and often run our simulations on supercomputers. Past projects have included the modeling of biological cells, nanoscale membranes, composite materials, hurricanes and more.

Relevant skills, background, or interests: Applied Math, Computational Science, Fluid Dynamics. Programming experience desirable but not necessary.

2. Scalable Solvers Group

Sherry Li, Group Lead and Senior Scientist

Intelligent Eigensolver

Field or research area: Linear Algebra; Randomization

Abstract: Combine machine learning and randomized sketching to choose an efficient algorithm for computing eigenpairs of a large sparse matrix.

Relevant skills, background, or interests: Numerical Linear Algebra, Randomized Algorithms, High Performance Computing

Autotuning HPC applications on exascale machines

Field or research area: High Performance Computing; Machine Learning

Abstract: Develop new autotuning algorithms in the GPTune Gaussian Process Bayesian optimization framework (https://github.com/gptune/GPTune) and apply them to tune the large scale DOE application codes.

Relevant skills, background, or interests: Statistics, Optimization, High Performance Computing, Python Programming, C Programming, Message Passing Interface Programming.

COMPUTER SCIENCE

3. Computer Architecture

John Shalf, Group Lead and Department Head for Computer Science

Data Analytics for Next Generation HPC Systems

Field or research area: Computer Science

Abstract: As part of designing the next HPC system, we are currently analyzing usage data of current systems to determine how systems are used and what future choices that usage motivates. In this project, we are looking for a student intern that will process and understand this usage data for the purpose of illustrating as well as creating new metrics to discover new insights. Most of analysis will be done in python but alternatives are possible.

Relevant skills, background, or interests: Good knowledge of programming and data representation are good, as well as curiosity to discover new metrics and insights, in order to generate future research directions.

Quantum Networking and Entangled States

Field or research area: Computer Science; Quantum Computing

Abstract: Looking for students interested in learning about quantum networking. This includes modeling of devices that can transcode quantum entangled states from the electrical domain to the optical domain.

Relevant skills, background, or interests: Quantum Networking, Device Modeling, Electrical Engineering, Optics

Hardware Architectures and Accelerators

Field or research area: Graph Analytic Algorithms

Abstract: Looking for students interested in developing specialized hardware architectures and accelerators for scalable graph analytics algorithms.

Relevant skills, background, or interests: Hardware Architectures, Accelerators, Graph Algorithms

4. Computer Languages and Systems Software

Damian Rouson, Group Lead

Full-Stack Acceleration of Predictions for Climate Change's Regional Impacts

Field or research area: High Performance Computing

Abstract: Representing the complex topography and fine-scale physics necessary to predict regional impacts of global climate change will require computing resources that outpace even the world's fastest supercomputers for the foreseeable future. The Intermediate Complexity Atmospheric Research (ICAR) model addresses this problem by downscaling: simulating high-resolution regional climate using input from lower-resolution, global climate models. Although downscaling reduces the problem size, the increased grid resolution imposes greater demands on accurately representing fine-scale phenomena such as atmospheric turbulence. This project will introduce a turbulence model into ICAR and explore ways to accelerate ICAR using partitioned-global address space (PGAS) parallel programming models to exploit artificial intelligence (AI) accelerator hardware. The project will leverage implementations of PGAS programming models developed by CLaSS researchers: the UPC++ template library and the OpenCoarrays Fortran 2018 parallel runtime library. Initial work on this new project focuses on using parallel performance analysis to identify execution bottlenecks. Future work will involve developing a UPC++ communication kernel for ICAR to leverage UPC++'s support for applications with unpredictable communication patterns and dynamic load-balancing requirements. Possible applications for AI will include using machine learning to tune the turbulence model or to produce initial estimates for solution increments for implicit time advancement schemes.

Relevant skills, background, or interests: Ideal candidates would have one or more of the following attributes:

- Faculty or students in computer science, the physical sciences, engineering, or mathematics
- Experience developing applications using parallel programming models such as MPI, OpenMP, or a PGAS model such as UPC++ or the parallel features of modern Fortran
- Experience in the application of AI to computational science problems, and/or a background in atmospheric sciences or fluid dynamics, especially turbulent flow.

Additional Information: Graduate-level or exceptional undergraduate students preferred.

GASNet

Field or research area: High Performance Computing Middleware

Abstract: As we approach the exascale era, trends in hardware and algorithms are pushing a demand for new programming models, as alternatives to the conventional "MPI+X." Many such alternatives are based on the PGAS (Partitioned Global Address Space) programming model. The PGAS model provides the abstraction of a global memory address space that is logically partitioned with a portion local to each process, thread, or rank. GASNet is a portable, high-performance PGAS communication library implemented directly over the native communication APIs of modern high-performance computing (HPC) networks. GASNet serves as the communication backend for several parallel programming languages and libraries, including UPC++, Legion, Chapel among others. GASNet provides a tighter semantic match between the needs of parallel programming models and modern network hardware capabilities than would be possible using MPI. Microbenchmarks demonstrate that GASNet performance is competitive with and often outperforms comparable communication operations provided by various MPI implementations. We seek collaborators interested in high-performance computing networks and parallel computing. Collaborators who are designing programming models might benefit from learning about and using GASNet as a back-end, and obtain experience running on world-class supercomputing hardware.

Relevant skills, background, or interests: Faculty or students in Computer Science or a related field.

- Experience with the design and implementation of applications using, or middleware implementing, the Partitioned Global Address Space (PGAS) model.
- Parallel programming in High Performance Computing applications and/or middleware using MPI, CUDA or a PGAS model.
- Experience in parallel/distributed algorithm design and implementation, especially with respect to the use of non-blocking communications and asynchronous execution.

Additional Information: Graduate-level or exceptional undergraduate students in Computer Science or a related field.

For more information, please visit https://gasnet.lbl.gov.

UPC++

Field or research area: Parallel Programming Models

Abstract: As we approach the exascale era, trends in hardware and algorithms are pushing a demand for new programming models, as alternatives to the conventional "MPI+X." UPC++ is a C++-based alternative which is based on the PGAS (Partitioned Global Address Space) programming model. The PGAS model provides the abstraction of a global memory address space that is logically partitioned with a portion local to each process. In addition to the Remote Memory Access (RMA) capabilities common to PGAS models, UPC++ provides Remote Procedure Call (RPC) as a mechanism to move computation to the data (where RMA moves data to the computation). UPC++ provides high-level productivity abstractions appropriate for PGAS programming such as: RPC, RMA, remote atomic memory operations, locality-aware APIs for user-defined distributed objects, and robust support for asynchronous execution to hide communication costs. Together, these features enable agile, lightweight communication such as arises in irregular applications, libraries and frameworks running on exascale systems. We seek collaborators interested in high-performance parallel computing. Collaborators who are designing parallel applications might benefit from learning about and using UPC++, and obtain experience running on world-class supercomputing hardware.

Relevant skills, background, or interests:

Have one or more of the following attributes:

- Faculty or students in computer science, science or engineering.
- Faculty or students in a science or math field with experience in parallel simulation.
- Experience with the design and implementation of applications using parallel programming models such as MPI, OpenMP, CUDA, OpenACC, or a PGAS model.
- Experience in parallel/distributed algorithm design and implementation, especially with respect to the use of non-blocking communications and asynchronous execution.

Additional Information: For more information, please visit https://upcxx.lbl.gov.

5. Performance and Algorithms

Khaled Ibrahim, Computer Scientist

Distributed DNN Performance Analysis and Optimization For HPC Systems

Field or research area: High Performance Distributed Computing

Abstract: Optimizing the performance of distributed DNN is critical for achieving a reasonable training time. DNN performance depends on a wide set of parameters involving the DNN model, the framework in use, the underlying system architectures, etc. This project aims at exploring various strategies performance analysis of distributed training of DNN and explore different techniques for optimizing performance. Our research efforts at the performance and algorithms group involve developing performance models and analysis techniques, performance tuning and optimization, and the development of application-specific runtime systems.

Relevant skills, background, or interests: Python Programming, Performance Modeling and Analysis, C/C++ programming, Familiarity with Distributed Computing

COMPUTATIONAL SCIENCE

6. Applied Numerical Algorithms Group

Daniel Martin, Group Lead and Staff Scientist

Ice sheet modeling

Field or research area: Scientific Computing; Climate

Abstract: The response of the Greenland and Antarctic ice sheets to the changing climate remains the single greatest uncertainty in projections of sea level rise. The BISICLES ice sheet model, developed at LBL, is an adaptive mesh refinement model designed to efficiently and accurately understand and project ice sheet dynamics. Project ideas include improving numerics of the model, adding additional physics to the model, and simulation-based projects to understand specific climate scenarios.

Relevant skills, background, or interests: Specific background depends on the specific project focus, but generally: applied mathematics, software development (C++, python), glaciology, climate science, etc.

Additional Information: This is a placeholder for a range of project ideas under the ice sheet modeling effort. We can come up with specific projects based on your skills and interests.

7. Computational Chemistry, Materials & Climate Group

Bert de Jong, Group Lead & Senior Scientist

Science, Computer Science, and Math For Quantum Computing

Field or research area: Quantum Computing

Abstract: Our research in quantum computing includes 1) the development of a software stack, consisting of programming models, compilers, validation and debugging and optimization capabilities, 2) the development of algorithms for scientific applications, 3) performing science experiments on quantum computers.

Relevant skills, background, or interests: There are opportunities for computer scientists, mathematicians focused on linear algebra, quantum information scientists, and domain scientists with some knowledge of quantum computing.

AI for chemistry and materials

Field or research area: Machine Learning

Abstract: Machine learning (ML) has the potential to accelerate scientific discovery in chemistry and materials. Our group is developing new ML approaches to increase the accuracy of ML models, and is developing techniques that would enable inverse or rational design of molecules and materials for various scientific application domains.

Relevant skills, background, or interests: Background in chemistry or materials with a strong affinity to computational methods and some experience with Machine Learning.

Scalable Computational Chemistry

Field or research area: High Performance Computing

Abstract: CCMC is involved in the development of scalable computational chemistry and materials codes, focused on next-generation hardware (GPU, TPU, etc.) and the exascale. There are opportunities to integrate new methods or ideas developed by SRP participants.

These codes have always room for improvement and could benefit from insights from computer scientists and mathematicians.

Relevant skills, background, or interests: Background in scalable computational chemistry and materials or computer science or applied math.

Scientific Insights Through HPC Computational Chemistry and Materials Simulations

Field or research area: Quantum Chemistry; Materials

Abstract: Our group does not only develop scalable tools for solving computational chemistry and materials problems, we're also computational scientists in pursuit of solving complex scientific problems relevant to the Department of Energy. Interests lie in Carbon capture, solar harvesting, energy storage, thermoelectric and thermocaloric materials, and drug design.

Relevant skills, background, or interests: Chemistry or materials background. Some affinity with computational chemistry or materials codes.

8. Computational Biosciences

Kristofer Bouchard, Group Lead and Staff Scientist

Deep Phylogeny

Field or research area: Genetics; Deep Learning

Abstract: The tree of life is a fundamental concept in biology, describing the relationship of different species to each other and their origins. Current approaches for determining (classifying) where an observed genetic sequence should be placed are time-consuming. We have recently shown that deep learning is a promising approach, and there is much work to do in this space.

Relevant skills, background, or interests: Experience in pytorch and High Performance Computing. Experience with hyper parameter optimization (e.g., RayTune) would be good.

‘Governing equations’ of neural dynamics from noisy data

Field or research area: Statistical Machine Learning

Abstract: Brain function is an emergent property of the coordinated activity of multiple neuronal types that are widely distributed across brain regions. That is, the interaction of many components over time produce something that is fundamentally different than can be

predicted from the parts. Current statistical-machine learning tools provide insight into the properties of individual neurons and can characterize the properties of entire populations. However, understanding how neurons with different morphological, electrophysiological, and transcriptomic properties differentially contribute to population dynamics is a major hurdle. Such an understanding is required to give biological interpretation to dynamical systems models of populations and to treat neurological disorders through targeted pharmacological interventions and optogenetic control. Addressing this challenge will require novel methods for extracting nonlinear 'governing equations' from high-dimensional, noisy time-series data with unobserved influences. A central area of research and development in our group is focused on developing methods to address this long-term, grand challenge problem.
Relevant skills, background, or interests: Time-Series Analysis, Statistical Mechanics, Machine Learning, Information Theory, Statistical Models, Dimensionality Reduction

9. Computational Cosmology Center

Peter Nugent, Senior Scientist and Division Deputy for Scientific Engagement, Department Head for Computational Science

Gravitationally Lensed Supernovae

Field or research area: Astrophysics
Abstract: In 1929 Edwin Hubble rocked the physics community with his announcement that the universe was expanding - characterized shortly thereafter by his eponymous constant. Nearly 80 years later the field of cosmology has advanced tremendously with the discoveries of the accelerated expansion of the universe and precision measurements of both the universe's mass and energy densities. Amazingly, the most contested value in cosmology today is the first one that was measured, Hubble's Constant. There is currently a 3+ sigma disagreement between local values, determined by supernovae and a maser, and distant ones from Baryon Acoustic Oscillations and the Cosmic Microwave Background. This implies that there are unknown systematics in either (or both) measurements or new physics. The key to unraveling this mystery lies in a method first proposed in 1964 by Sjur Refsdal - the measurement of time delays from gravitationally lensed supernovae. In the past years we have made great strides towards this goal with the discovery of the first multiply lensed Type Ia supernova and a new method which will increase the discovery rate of these objects by over an order of magnitude in the next decade.
Relevant skills, background, or interests: Python, Basic Astronomy, Physics, Calculus

10. Machine Learning and Analytics

Talita Perciano, Research Scientist

Transformers for science

Field or research area: Computer Science; Data Analysis

Abstract: This project aims to explore Transformer-based architectures for the analysis of scientific data. There are several different types of Transformers proposed in the literature, however, little research has been done towards using and developing similar architectures for the analysis of large scientific datasets.

Relevant skills, background, or interests: Python (or other) Programming language, Deep Learning, Scientific Data Analysis, Image Processing.

Image reconstruction algorithms for scientific data

Field or research area: Computer Science, Applied Mathematics

Abstract: Image reconstruction is essential in several data acquisition pipelines happening at DOE data facilities such as the ALS. Image modalities such as microCT for example, rely on the acquisition of a projection image for each angle while a sample is being rotated. In order to obtain a final 3D image representation of the targeted sample, these projections need to be processed by an image reconstruction algorithm. This project aims to develop efficient reconstruction algorithms using mathematically grounded approaches and deep learning methods.

Relevant skills, background, or interests: Python, other Programming Language, Applied Mathematics, Deep Learning, Image Processing.

Probabilistic Graphical Deep Learning

Field or research area: Computer Science; Applied Mathematics; Applied Statistics

Abstract: A graphical model or probabilistic graphical model (PGM) or structured probabilistic model is a probabilistic model for which a graph expresses the conditional dependence structure between random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning. One of the main advantages about this techniques is its ability to use prior information (physical constraints) related to the data. This becomes very important when analyzing scientific data. This project aims to develop efficient PGM-based algorithms to tackle problems such as image segmentation, image denoising, feature tracking, data reduction, data fusion, etc. We use mainly Markov Random Fields and Conditional Random Fields, and some of our approaches combine these methods with deep learning algorithms.

Relevant skills, background, or interests: Python and C++ Programming Languages, Data Analysis, Statistics, Applied Mathematics, Image Processing, Deep Learning, Machine Learning.

Quantum Image Processing for Scientific Data

Field or research area: Quantum Computing; Computer Science; Applied Mathematics

Abstract: Quantum computing is a research area that has received a great amount of attention in the last few years. In this project, we aim to take advantage of quantum computing theory to develop quantum image processing tools suitable to the analysis of scientific data. This includes the development of new quantum circuits for quantum image representation and for analysis algorithms (feature extraction, template matching). We aim to develop concrete proof-of-concept tools that run on NISQ devices.

Relevant skills, background, or interests: Python, Other Programming Language, Deep Learning, Scientific Data Analysis, Image Processing.

11. Scientific Data Management

Kesheng (John) Wu, Group Lead and Senior Computer Scientist

Tools to Enhance Non-volatile Memory for Scientific Applications

Field or research area: Data Management; Memory Management

Abstract: Non-volatile memory (NVRAM) technology supports faster read and write operations than conventional disk-based storage systems, however, it also has its own limitations such as limited number of write cycles. In the Scientific Data Management research group, we are interested in developing new tools that could effectively utilize their faster read write capability while minimize the impact of their limitations. For example, we are interested in techniques to unify the Application Programming Interface (API) for DRAM and NVRAM, indexing techniques for multi-dimensional scientific data, and so on.

Relevant skills, background, or interests: Data Management, Memory Systems, Indexing Techniques

12. Integrated Data Frameworks

Sean Peisert, Staff Scientist

Secure High-Performance Computing

Field or research area: Secure High Performance Computing

Abstract: The Computational Research Division has an immediate opening for a graduate student to perform research and development in the area of the use of trusted execution environments (TEEs) in high-performance computing (HPC) domains. The goal of this work is to enable high-performance scientific computing of sensitive data without significantly compromising usability or performance. The work includes software development of and experimentation with security and privacy technologies critical to facilitate data collection and sharing. The goal of this work is to contribute to developing and implementing new trusted execution environment architectures and related low-level system software appropriate to the threat model and performance requirements of high-performance scientific computing. Current commercial TEEs are inadequate for HPC for a variety of reasons. Our solution involves a RISC-V based architectural development, development of and modifications to low-level operating system elements, and implementation and experimentation.

Relevant skills, background, or interests: Work will focus on operating system kernel elements. Research problems include security and performance elements, and tradeoffs between the two.

Required Skills: Experience in software engineering methodologies and in writing low-level system software, such as operating system kernels and drivers. Proficiency and experience in programming languages including low-level programming languages used in systems software, such as C/C++ and Rust, and higher-level languages used in data analysis, such as Python. Proficiency with UNIX tools and computer systems. Demonstrated ability to work independently and collaboratively in a diverse interdisciplinary team and contribute to an active intellectual environment. Excellent written and oral communication.

Additional Information: Interested students should provide a description of your relevant experience and a description of how you feel the project fits in with your future thesis/dissertation goals.

13. Scientific Networking Division

Mariam Kiran, Research Scientist

Networks for Machine Learning

Field or research area: Machine Learning; Networks

Abstract: Applications such as edge computing, federated learning, or general collecting data from, nodes need reliable networks to collect and send data for training a global Machine Learning model. How can we build reliable networks for Machine Learning applications?

Relevant skills, background, or interests: Transmission Control Protocol, Networking Knowledge, Machine Learning Knowledge, Deep Learning Knowledge

FIELD OR RESEARCH AREA INDEX

RELEVANT SKILLS, BACKGROUND, OR INTERESTS INDEX