Harsha Lokavarapu

5221 Ferrera Court
Pleasanton, CA 94588
Mobile: (925) 998-5486
Home: (925) 416-0490
lokavarapuh@gmail.com
https://github.com/hlokavarapu

Education

University of California, Davis	MS	Computational Geodynamics	2018
		Thesis Advisor: Professor Louise H. Kellogg	
University of California, Davis	BS	Computer Science	2015
	Minor	Applied Mathematics	2015

Appointments

2019-2020	Uber	Software Engineer
2014-2017	Computational Infrastructure for Geodynamics (CIG)	Junior Assistant Programmer
2012	Certify Data Systems (Humana)	Internship as Code Developer

Programming Languages, Computing Skills, and Work Experience

Fullstack Service Development

- Designed SQL schema to store hardware inventory, hardware reservation, and test case definition/result data
- Designed, architected and authored several golang services that:
  - stores an inventory of hardware assets and allows for employees to reserve assets
  - stores test case definitions such as regression, stress, and/or benchmarking for a set of hardware assets
  - stores test case results for a group of hardware assets
- Wrote a front end application to visualize test case defintions using React and Base Web

Data Analysis and Visualization

- I have worked with computing numerical models on several super computers housed at TACC. Several of these computational experiments generated datasets on the order of terabytes. Using numerical and visualization libraries in python, I have created several easy to share Jupyter notebooks to obtain numerical measures such as computational accuracy, stability, Kolmogorov Entropy, and others. Data formats that I am familiar with include HDF5, VTK, NETCDF and ASCII - (Git: Deep Carbon, Convergence Analysis, and more)
- In cases where the underlying dataset to analyze is less than several gigabytes, I have come to rely on using programming language R and its rich statistical libraries. The ease of use to write small scripts to quickly calculate statistical measures has moved me to prefer R over other remaining languages - (Git: Code)

- When dealing with extremely small datasets with the goal of quickly generating plots for discussion, I have come to rely on Gnuplot and its command line interface. Lately, I have been prefering python's matplotlib.pyplot module however - (Git: Code)
- Working with supercomputers with a predefined limit on core hour usage, I have developed experience making the most of the limited resources using various parallelization tricks. Many of these tricks require extensive experience with bash scripting in order to run numerous jobs, monitor the state of jobs, and data collection towards the end - (Git: Code)

Deep Learning

- Participant of Kaggle competitions:
  - House Prices: Advanced Regression Techniques
    * Use of Keras library to develop a regression model
    * Employed feature selection methods, feature crosses, normalization of data input, removal of outliers, and transformation of input data to correct for heavy left skew distributions
    * In combination with tensorboard, executed a hyperparameter search to discover optimal number of fully connected layers, number of neurons per layer, activation functions, and learning rates.
    * Currently placed in the 34th percentile within the competition.
  - Human Protein Atlas Image Classification
    * This competitions' challenges included multilabel classification, 4 channel input images, and imbalanced dataset. Less than 1% of the training image data consisted of 5 classes. Of the remaining classes, 3-4 classes were represented by 70% of the data set.
    * My submitted model took advantage of input images resized to 128 by 128, focal loss function, data augmentation, external data, and oversampling to counteract the imbalanced dataset.
    * Custom metrics were written to better measure the improvements of variuos experiments.
    * Experiments included transfer learning to VGG network, different network architectures such as Conv2D sequential architecture, and Netowk in Network architectures and more.
    * After the conclusion of the competition, the top 1% of the models consisted of ensembles of extremely large models handling input sizes of 512 by 512, 1024 by 1024 in compbination with transfer learning. Realizing the importance of computational power, I am currently experimenting with google cloud and their machine learning engine.
- Google cloud computing to train neural networks in a multi-GPU environment.

  Coursera Certifications:
  - Convolutional Neural Networks
  - Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
  - Neural Networks and Deep Learning

GPU Experience

- In order to resolve MHD dynamo of planets and stars, it is of utmost importance to resolve small scales. Due to the underlying mathematical principles, physical quantites of interest were transformed into spherical harmonic coefficients, whose computational complexity is O(N^3). Using CUDA in a multi-GPU environment, I pursed the optimization of spherical harmonic transform (SHT). Optimizations involved taking advantage of state of the art GPU libraries: CUDA Fast Fourier Transform (cuFFT) library, CUDA Basic Linear Algebra Subprograms (cuBLAS), and (CUB) library. This alogirthm was optimized to a time complexity of O(N^2logN) - Dev branch)
- Profile and test optimizations using strong and weak scaling tests
- Published results as poster [4], [6] (see below) at the 2014, 2015 Annual Fall AGU Meetings

Supercomputing Experience

- National Science Foundation (NSF) Texas Advanced Computing Center
  - Stampede and Stampede 2 - Ranked as 17th fastest supercomputer in top 500 list. Primarily utilized for running massive compuations using ASPECT to track interfaces in a multilayered buoyancy model over thousands of cores. Results and analysis were published in [1].
  - Maverick - A multi-gpu Nvidia K20 cluster. Both development and analysis of optimzation of SHT was primarily conducted on this cluster.

Open Source Code Development

- Advanced Solver for Problems in Earth's ConvecTion (ASPECT) written in C++. I am one of the active contributors of this project.
  - Design and implementation of parallel particle generation algorithms. (Git: PR)
  - Design and implementation of parallel particle interpolation algorithms including harmonic averaging and bilinear least squares. (Git PR: Harmonic Average, Bilinear least squares)
  - Design 2-D analytical solution to Stokes equations in order to benchmark the accuracy of particle algorithms in ASPECT. (Git PR: 1, 2)
  - Execute strong and weak scaling tests for original draft of publication [3] (see below), which was not included in the final publication
  - ASPECT contributions Git timeline
- State of-the-art model of the Earth's Geodynamo, Calypso written in FORTRAN 90. I am the 3/4 most active contributor.
  - Calypso contributions Git timeline
- Generalized reservoir modeling library (MS Thesis Project: Resecore) written in Python

Tools

- Structured Query Language
- React and Base Web for building front end applications
- GNS3 virtualization platform
- Zero-touch provisioning Arista and Juniper switches
- ISC DHCP server
- Paraview (Git: Code)
- Visit
- CMake, CTest - Build tool and Unit testing - (Git PR: 1, 2)
- CUDA - (Git: PR)(SHT = Spherical Harmonic Transform)
- Distributed memory parallelism - MPI for C++ and FORTRAN
- Shared memory parallelism - openMP
- SLURM HPC scheduler
- Profilers: gdb and cuda-gdb

Outside Interests

- Virtual Reality - (Git: VR-Experiments using A-Frame library) - JavaScript
- 3-D Design/Printing - Tinkercad

Publications

Refereed Journal Publications

Submitted

[1] L. H. Kellogg, D. L. Turcotte, M. Weisfeiler, H. Lokavarapu[*], S. Mukhopadhyay, (2018) "Implications of a Reservoir Model for the Evolution of Deep Carbon", *Earth and Planetary Science Letters*, Ms. Ref. No.: EPSL-D-17-01055

Accepted

[2] R. Gassmoeller, H. Lokavarapu[*], E. Heien, E. G. Puckett, and W. Bangerth, (2018) "Flexible and scalable particle-in-cell methods with adaptive mesh refinement for geodynamic computations", *Geochemistry, Geophysics, Geosystems* manuscript 2018GC007508R View Accepted Manuscript

Appeared

[3] E. G. Puckett, D. L. Turcotte, L. H. Kellogg, Y. He[†], J. M. Robey[*], and H. Lokavarapu[@] (2018) "New numerical approaches for modeling thermochemical convection in a compositionally stratified fluid", Special issue of . *Physics of the Earth and Planetary Interiors* associated with the 15th Studies of the Earth's Deep Interior (SEDI) Symposium (*Phys. Earth. Planet. In.)* **276**:10–35, 10.1016/j.pepi.2017.10.004 View Article

Poster Presentations

[1] L. H. Kellogg, H. Lokavarapu[*], D. L. Turcotte, and S. Mukhopadhyay (2017) "A reservoir model study of the flux of carbon from the atmosphere, to the continental crust, to the mantle", *Annual Geophysical Union Fall Meeting 2017* View Abstract

[2] J. Jiang, A. P. Kaloti, H. R. Levinson, N. Nguyen, E. G. Puckett, and H. Lokavarapu[@] (2016) "Benchmark Results Of Active Tracer Particles In The Open Souce Code ASPECT For Modelling Convection In The Earth's Mantle", *Annual Geophysical Union Fall Meeting 2016* View Abstract

[3] E. G. Puckett, D. L. Turcotte, L. H. Kellogg, H. Lokavarapu[@], Y. He[†], and J. M. Robey[*] (2016) "New Numerical Approaches To thermal Convection In A Compositionally Stratified Fluid", *Annual Geophysical Union Fall Meeting 2016* View Abstract

[4] H. Lokavarapu[@], and H. Matsui (2015) "Optimization of Parallel Legendre Transform using Graphics Processing Unit (GPU) for a Geodynamo Code", *Annual Geophysical Union Fall Meeting 2015* View Abstract

[5] J. A. Russo, E. H. Studley, H. Lokavarapu[@], I. Cherkashin, and E. G. Puckett (2014) "A New Monotonicity-Preserving Numerical Method for Approximating Solutions to the Rayleigh-Benard Equations", *Annual Geophysical Union Fall Meeting 2014* View Abstract

[6] H. Lokavarapu[@], H. Matsui, and E. M. Heien (2014) "Parallelization of the Legendre Transform for a Geodynamics Code", *Annual Geophysical Union Fall Meeting 2014* View Abstract

[@]Undergraduate Student
[*]Graduate Student
[†]Postdoctoral Scholar