Overview of GCL Projects

Michela Taufer
Global Computing Laboratory

Who we are

Dr. Michela Taufer
Jack Dongarra Professor

Dr. Silvina Caino-Lores
Postdoctoral Researcher

Dr. Ariel Rorabaugh
Postdoctoral Researcher

Dr. Leobardo Valera
Postdoctoral Researcher

Ian Lumsden
Doctoral Student

Paula Olaya
Doctoral Student

Nigel Tan
Doctoral Student

Kae Suarez
Doctoral Student

Clark Hathaway
Undergraduate Student

Sebastian Mobo
Undergraduate Student

Ria Patel
Undergraduate Student

Lauren Whitnah
Technical Writer

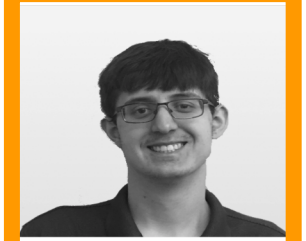
Who will present today

Dr. Michela Taufer
Jack Dongarra Professor

Dr. Silvina Caino-Lores
Postdoctoral Researcher

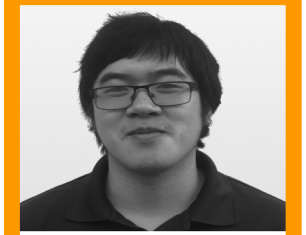
Dr. Ariel Rorabaugh
Postdoctoral Researcher

Dr. Leobardo Valera
Postdoctoral Researcher

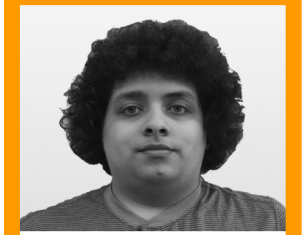


Ian Lumsden
Doctoral Student

aula Olaya
octoral Student



Nigel Tan
Doctoral Student



Kae Suarez
Doctoral Student

Clark Hathaway
Undergraduate Student

Sebastian Mobo
Undergraduate Student

Ria Patel
Undergraduate Student

Lauren Whitnah
Technical Writer

Our External Collaborators and Sponsors

LLNL: Stephen Herbein, Todd Galbin, Olga Pearce
LANL: Bob Bird
Sandia: Jay Lofsead
Oak Ridge National Laboratory: Travis Johnston
Cornell U.: Harrel Weinstein
Syracuse U.: Duncan Brown
Southern California University: Ewa Deelman, Rafael Ferreira da Silva
U. of North Texas: Sanjukta Bhowmick
U. Utah: Mary Hall
U. of New Mexico: Trilce Estrada
U. of Maryland: Abhinav Bhatele
U. Delaware: Rodrigo Vargas
UIUC: Victoria Stodden
RIKEN: Florence Tama and Osamu Miyashita

Workflows for in situ analytics

- Build workflows that combine machine learning and in situ data analytics approaches, workflow management methods, and high performance computing techniques
  - **A4MD: Build a framework to analyze molecular dynamics data as it is generated → Silvina**
  - **A4NN: Build a framework to design artificial intelligence (AI) models for specific novel scientific datasets → Ariel**
  - A4X: collect the lessons learned and integrate them into a general framework for in situ analytics, where X = {MD, NN, etc.}

Performance Analysis and Beyond

- Build a suite of ML methods to prevent and cure IO contentions in HPC systems → Mike
  - Prionn (to prevent contentions)
  - CanarIO (to cure contentions)
- **Augment Hatchet (a LLNL data analysis tool that can read HPC profiling data from different profilers) → Ian**
  - **Analyze performance of different MPI calls in benchmark applications**
- **Adopt and adapt HPC performance analytics method for scientific data modeling and predictions → Leobardo**
  - **Model soil moisture for precision agriculture and wildfire simulations**

Scientific Applications

- Definite and optimize models and algorithms, build workflows for and interfaces to CI, optimize performance, increase application scale, and tune accuracy
  - **Protein structure identification from protein diffraction patterns → Paula**
  - Metabolomics (i.e., high-throughput analytical chemistry and multivariate data analysis) in tripartite ecosystems (i.e., bacterium, roundworm, and insect)
  - **Large-scale particle-in-cell plasma simulations → Nigel**

Reproducibility and Traceability

- Build systems that support reproducibility and traceability
  - Link non-determinism with reproducibility issues (e.g. bugs, result differences) in scientific applications → Dylan and Kae
  - Build containerized environments for reproducibility and traceability of scientific workflows → Paula
- Define policies and practices for reproducible sciences
  - Reproduce results in high-profile scientific discoveries -- Laser Interferometer Gravitational-Wave Observatory (LIGO) and Event Horizon Telescope (EHT)
  - Gather the scientific community in a set of workshop to study robust science (i.e., reproducibility, permanence, and scalability) (recently awarded NSF PPoSS)