

APS Scientific Computation Seminar Series

Speaker: Yao-Lung Leo Fang
Assistant Computational Scientist
Computational Science Initiative
Brookhaven National Laboratory

Title: High-Performance Python GPU Programming with CuPy

Date: Monday, May 24, 2021

Time: 1:00 p.m. (Central Time)

Location: <https://bluejeans.com/5625919115?src=calendarLink>

Hosts: Mathew Cherukara and Nicholas Schwarz

Abstract:

Python is a powerful dynamic language enabling faster prototyping-to-production cycles. With its rich ecosystem based on NumPy and SciPy, many useful libraries for scientific computing, data science, and machine learning have been built and widely used. However, usually a penalty associated with such development efficiency is performance. It is known that Python CPU-based libraries often require some kinds of optimization strategies, such as parallelization, device offloading, just-in-time (JIT) compilation, etc, to match the performance to their pure C/C++ counterparts (if exist), and code porting and/or rewriting could be a tedious work, especially for domain scientists who might be less fluent or comfortable in low-level programming. In this talk, as a core contributor to CuPy — a NumPy-compliant GPU array library accelerated by CUDA — I will introduce how it can be used to significantly ease code porting and optimization. Followed by a brief introduction, I will explain how several useful, high-level features, including reduction operations and fast Fourier-Transforms, are optimized under the hood. In addition to its NumPy compliance that considerably lowers the bar to port NumPy code to GPU, CuPy also offers high extensibility with various forms of custom GPU kernels which I will also discuss. Finally, I will present the community effort to make CuPy interoperable with other Python libraries, such as Numba and mpi4py. If time permits, I will discuss CuPy's internal (including the on-the-fly kernel generation and compilation and the recent ROCm/HIP support for AMD GPUs) and the upcoming Python Array API standard which CuPy will comply with.

Yao-Lung “Leo” Fang is an Assistant Computational Scientist in Computational Science Initiative (CSI), Brookhaven National Laboratory (BNL). He received his B.S. in physics from National Taiwan University in 2009 and his Ph.D. in physics from Duke University in 2017. He is a theoretical physicist by training, specialized in quantum optics and open quantum systems. After joining BNL, his research interests gradually shifted towards high-performance computing (HPC) and quantum computing (QC). He is a current member of CSI's HPC and QC groups. His best known works include nonlinear quantum optics in the non-Markovian regime and GPU acceleration for X-ray ptychography. He is a regular contributor to several efforts within the Python scientific computing ecosystem and in particular the CuPy project. He has mentored several GPU Hackathon teams for Python GPU programming and is a member of the Consortium for Python Data API Standards.