Workshop 3:
GPU compute job scheduler

Jeremy Jacobson
Lecturer

EMORY COLLEGE OF ARTS AND SCIENCES

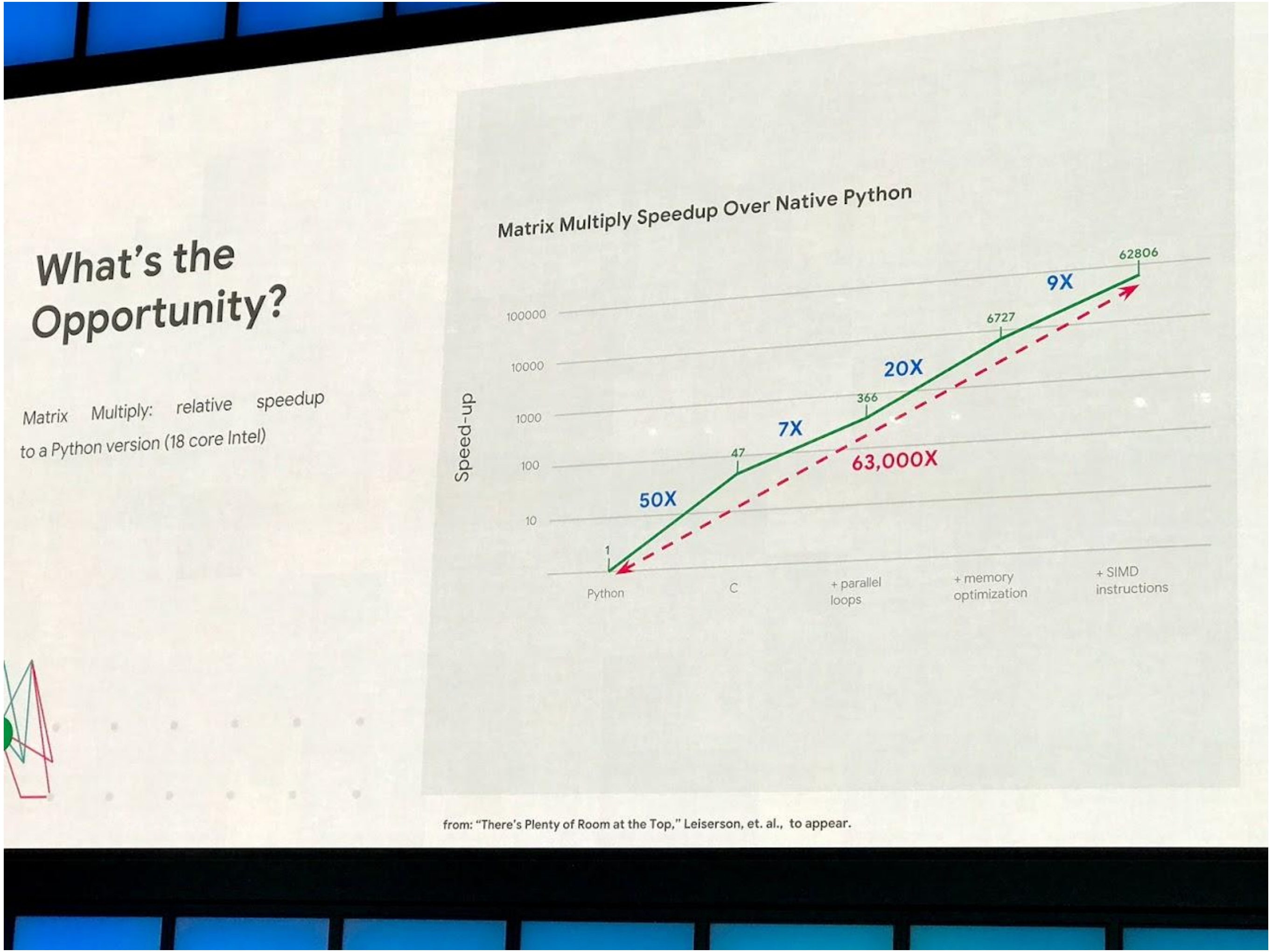
Department of Quantitative Theory and Methods

Overview

Main points

- (5-10 minutes) **Introductions**
- (5 minutes) **Workshop 2 review**
- (30 minutes) **Workshop 3 topics:**
  - **Installing software (review)**
  - **Slurm job scheduler**
  - **Jupyterlab on server**

Review: Workshop 2



What's the Opportunity?

Matrix Multiply: relative speedup to a Python version (18 core Intel)

Matrix Multiply Speedup Over Native Python

from: "There's Plenty of Room at the Top," Leiserson, et. al., to appear.

Review:
Matrix multiply on a GPU

- Pure Python
- Numpy
- PyTorch

Installing software

- Recall, anything you want to install on the server should be within a virtual environment of some kind
  1. Conda environments (python and data science tools)
  2. Venv (Python packages)
  3. Docker (only if software not use GPU)
  4. Enroot (if software uses GPU)

Slurm job scheduler

- There are 8 GPUs on the QTM server
- To share them among the growing number of QTM faculty, the server uses a job scheduler called Slurm
- Any software that uses the GPU or is memory/compute intensive should be run via Slurm

Jupyterlab/Rstudio on server

- Run code from last workshop: Start jupyterLab as batch job on GPU with time limit

References

- For another example of the workflow introduced today, with a few pro tips throw in, check out these blog posts (1, 2) from the lab of Harriet Alexander at the Woods Hole Oceanographic Institution.
- The sbatch file we wrote for our Slurm job to start Jupyter was based off the one found here at the Yale Center for Research Computing