

# Data Analysis with GPU-Accelerated Kernels

*Tuesday, July 28, 2020 5:30 PM (20 minutes)*

At HEP experiments, processing billions of records of structured numerical data can be a bottleneck in the analysis pipeline. This step is typically more complex than current query languages allow, such that numerical codes are used. As highly parallel computing architectures are increasingly important in the computing ecosystem, it may be useful to consider how accelerators such as GPUs can be used for data analysis. Using CMS and ATLAS Open Data, we implement a benchmark physics analysis with GPU acceleration directly in Python based on efficient computational kernels using Numba/LLVM, resulting in an order of magnitude throughput increase over a pure CPU-based approach. We discuss the implementation and performance benchmarks of the physics kernels on CPU and GPU targets. We demonstrate how these kernels are combined to a modern ML-intensive workflow to enable efficient data analysis on high-performance servers and remark on possible operational considerations.

## I read the instructions

## Secondary track (number)

**Author:** DUTTA, Irene (California Institute of Technology (US))

**Co-authors:** LU, Nan (California Institute of Technology (US)); Dr VLIMANT, Jean-Roch (California Institute of Technology (US)); NEWMAN, Harvey (California Institute of Technology (US)); SPIROPULU, Maria (California Institute of Technology (US)); REISSEL, Christina (ETH Zurich (CH)); RUINI, Daniele (ETH Zurich (CH)); PATA, Joosep (California Institute of Technology (US))

**Presenter:** DUTTA, Irene (California Institute of Technology (US))

**Session Classification:** Computing and Data Handling

**Track Classification:** 14. Computing and Data Handling