

## Distributed statistical inference with pyhf enabled through funcX

*Thursday, May 20, 2021 3:13 PM (13 minutes)*

In High Energy Physics facilities that provide High Performance Computing environments provide an opportunity to efficiently perform the statistical inference required for analysis of data from the Large Hadron Collider, but can pose problems with orchestration and efficient scheduling. The compute architectures at these facilities do not easily support the Python compute model, and the configuration scheduling of batch jobs for physics often requires expertise in multiple job scheduling services. The combination of the pure-Python libraries pyhf and funcX reduces the common problem in HEP analyses of performing statistical inference with binned models, that would traditionally take multiple hours and bespoke scheduling, to an on-demand (fitting) “function as a service” that can scalably execute across workers in just a few minutes, offering reduced time to insight and inference. We demonstrate execution of a scalable workflow using funcX to simultaneously fit 125 signal hypotheses from a published ATLAS search for new physics using pyhf with a wall time of under 3 minutes. We additionally show performance comparisons for other physics analyses with openly published probability models and argue for a blueprint of fitting as a service systems at HPC centers.

**Primary authors:** FEICKERT, Matthew (Univ. Illinois at Urbana Champaign (US)); HEINRICH, Lukas Alexander (CERN); Dr STARK, Giordon Holtsberg (University of California,Santa Cruz (US)); GALEWSKY, Benjamin (Univ. Illinois at Urbana Champaign (US))

**Presenter:** FEICKERT, Matthew (Univ. Illinois at Urbana Champaign (US))

**Session Classification:** Virtualisation

**Track Classification:** Distributed Computing, Data Management and Facilities