Julia for AGC
Atell-Yehor Krasnopol'ski
IRIS-HEP Fellow
Taras Shevchenko
National University of
Kyiv

Jerry Ling
Harvard University

Alexander Held
UWM
Analysis Grand Challenge

- columnar data extraction from large datasets
- processing of that data (event filtering, construction of observables, evaluation of systematic uncertainties) into histograms
- statistical model construction and statistical inference
- relevant visualisations for these steps
Analysis Grand Challenge

Find more here:
agc.readthedocs.io
Why Julia

- Perfect for Physics and Mathematics
- Fast by design, not because of packages & JIT-compiled
- Can interact with C, FORTRAN & Python
- Proven to be efficient for HEP: github.com/JuliaHEP
- www.juliahep.org
- arxiv.org/abs/2306.03675

(towardsdatascience.com/r-vs-python-vs-julia-90456a2bcbab)
For this task

- Less than 100 lines of code for the main loop
- Plotting, distributed computing, and working with complex data structures could not be easier
- Syntax & general experience
- Some tools were a bit raw when we started
Results

- The whole pipeline (except ML-related parts)
- Generating correct histograms with native Julia up to bin migrations
Results

Distributed version

Scaling of LHC_AGC.jl

![Graph showing scaling of LHC_AGC.jl with number of cores on the x-axis and 1/process time (1/s) on the y-axis. The graph compares real scaling (solid line and markers) and ideal scaling (dashed line).]
Results

- Convenient visualisation tools
- The workspace can be fully exported to a JSON file compatible with Cabinetry/Pyhf
- Found some issues in the reference implementation
What Have I Learned?

- First time working with distributed computing
- Some statistical insights about the inner workings of AGC
- Got more understanding of HEP