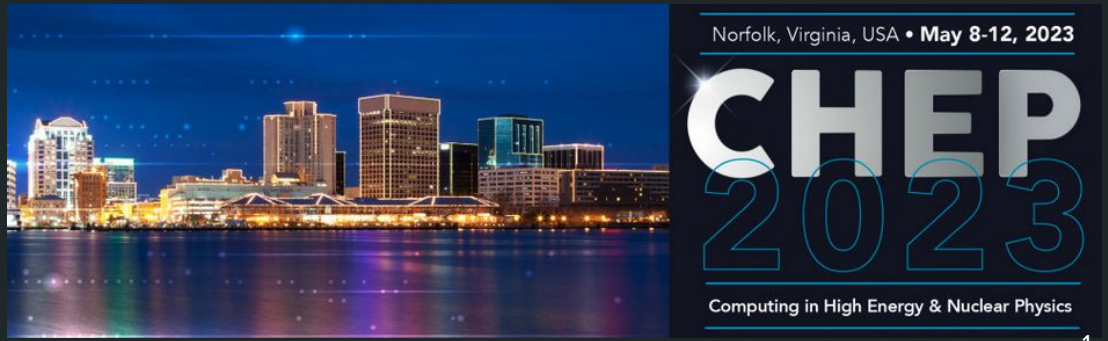
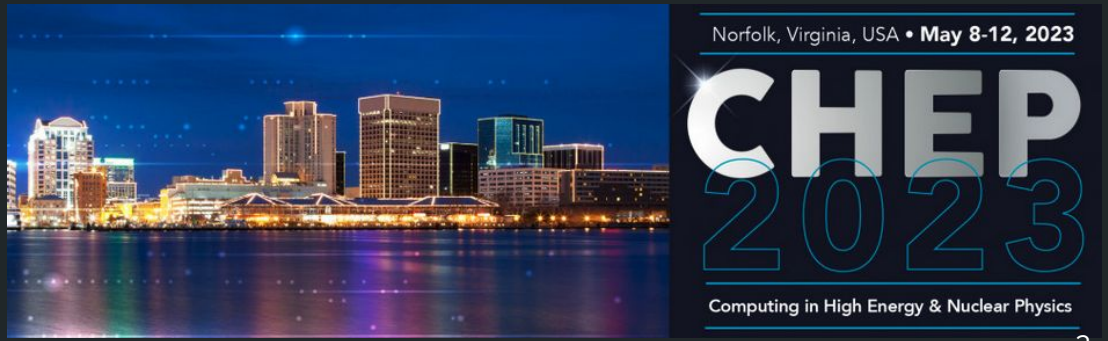
Track 6: Physics Analysis Tools Summary

Stephan Hageboeck
David Heddle
Alex Held
Nicole Skidmore



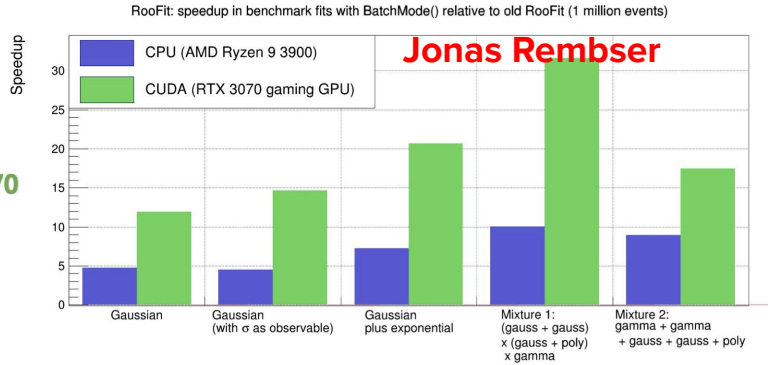
Statistical Inference and Fitting



Statistical inference & fitting (Stephan, Alex)

RooFit's new heterogeneous computing backend

- Revised CUDA backend in **ROOT 6.28.04**!
  - Gives you great speedup for wide range of unbinned fits with many events
  - Average speedup of **25x** (up to **40x!**) in fits with 1M events on **GeForce RTX 3070**

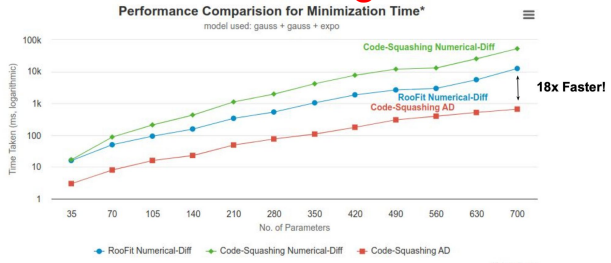


Automatic Differentiation Applied to RooFit

- Usage of clad for differentiation of RooFit models
- Advantages over numerical differentiation

Results

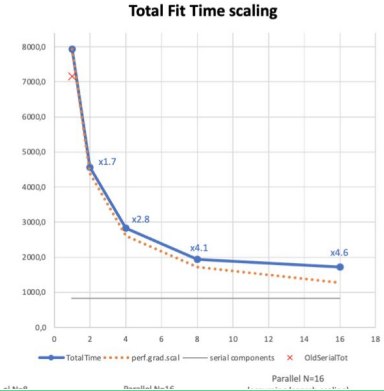
Garima Singh



Build-a-Fit: RooFit Parallelisation and Benchmarking Tools

- Likelihood-gradient parallelisation to improve Higgs fits

Zef Wolffs

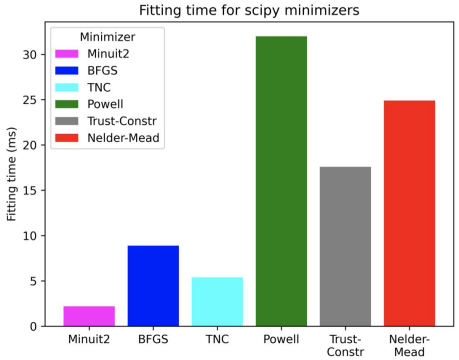


Statistical inference & fitting (Stephan, Alex)

New developments in Minuit2

Lorenzo Moneta

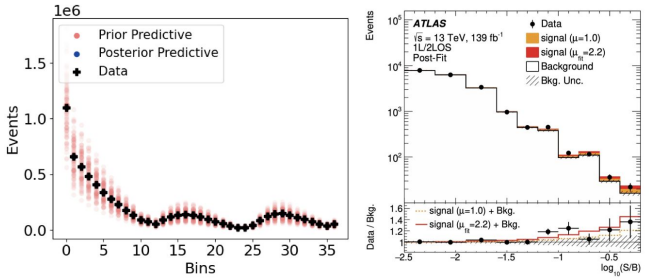
- Minuit is more than 50 years old but it seems to be still the best minimization algorithm for HEP fitting problems
- New algorithm (Fumili2) for least-square and binned likelihood fit



Bayesian Methodology for Particle Physics with pyhf

- Parallel Bayesan / Frequentist interface for HistFactory models in pyhf

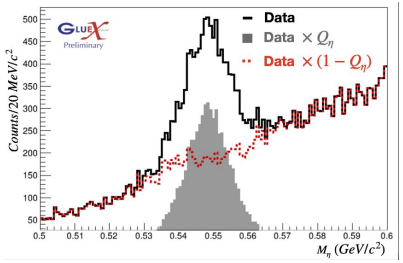
Malin Horstmann



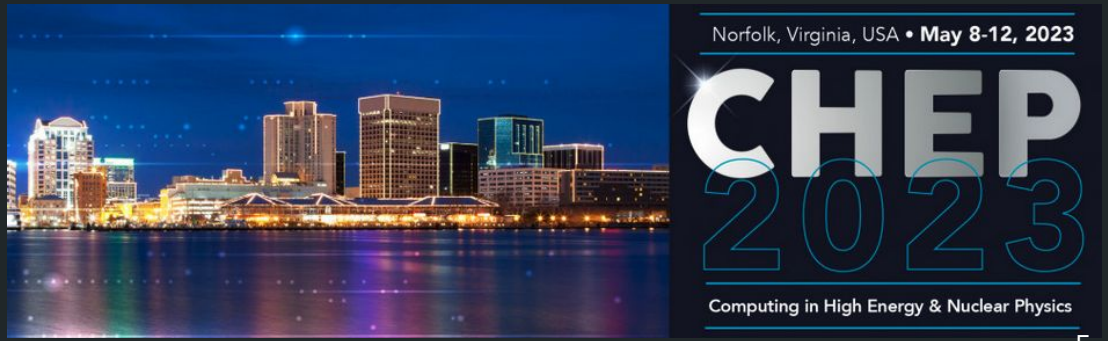
A multidimensional, event-by-event, statistical weighting procedure for signal to background separation

Zachary Baldwin

- Employing k-nearest neighbour method to assign event quality factors



I/O and data formats



I/O and data formats (David, Nicole)

DUNE HDF5 Experience

Barnali Chowdhury

- HDF5 will be used as raw data file format in ProtoDUNE(s)-II.
  - Have successfully taken data in HDF5 from HD, VD Cold Box testing. ND-LAr Module0 data is in HDF5 format.

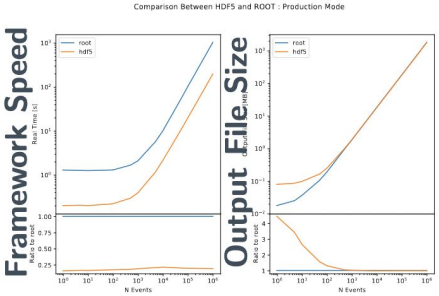
Schema-Evolution and the TTree within HDF5

Tom Eichlersmith

HDTree

- Utilize TTree-like data organization and access patterns while gaining the benefit of an industry-standard file format.
- C++ API currently available with performance on-par (if not exceeding) ROOT.

Comparison to ROOT



I/O and data formats (David, Nicole)

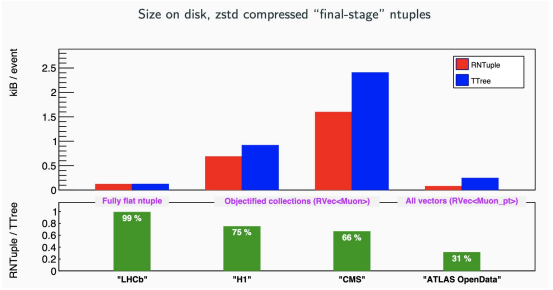
ROOT's RNTuple path to production

Jakob Blomer

- Smaller files, significantly faster reading – ideal format for HEP data

ROOT RNTuple is a leap in data throughput and storage efficiency

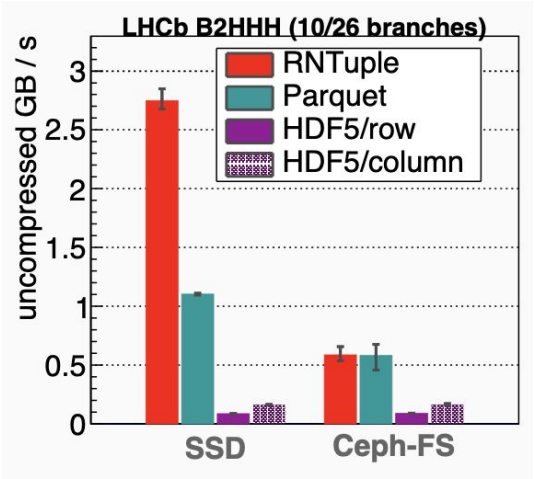
- Significantly smaller files and faster reads compared to TTree
- Efficient use of modern devices and storage systems such as SSDs, object stores, accelerators
- Work in progress with first successful integration efforts:
  CMS & ATLAS frameworks, RDataFrame, RBrowser, XRootD, TTree data importer



Integration of RNTuple in ATLAS Athena

Florine de Geus

- Support for RNTuple almost complete
- Improvements in file size and read throughput



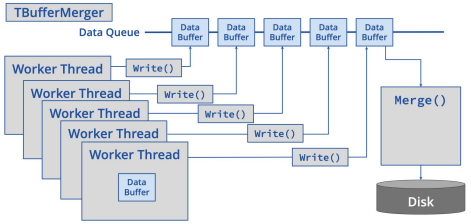
I/O and data formats (David, Nicole)

Improving ROOT I/O Performance for Analysis **Philippe Canal**

- Order of magnitude I/O improvements in highly parallel workflows

Thread safe statics

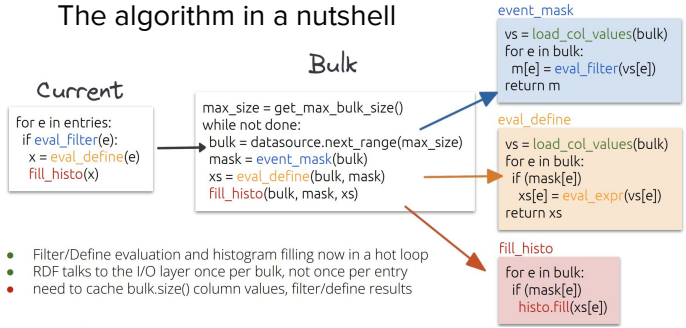
```
static std::vector<size_t> lengths{ []()
{
  std::vector<size_t> create_lengths;
  for (...) {
    create_lengths.push_back(...);
  }
  return create_lengths;
}() };
return lengths;
```



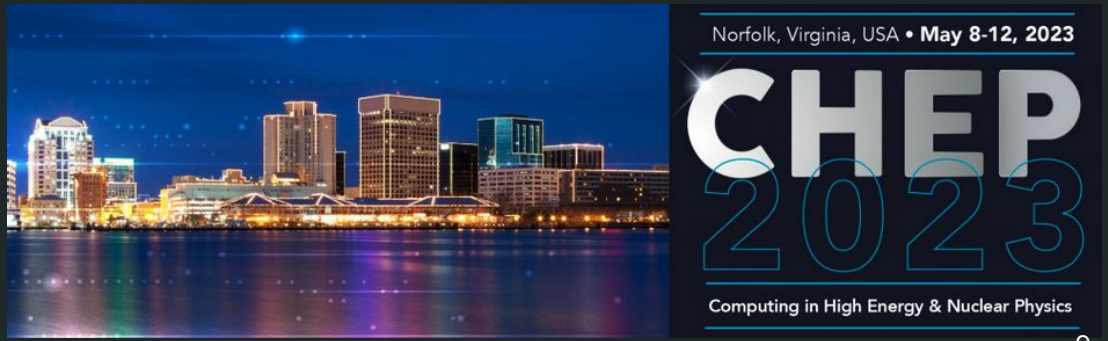
Boosting RDataFrame performance with transparent bulk event processing **Enrico Guiraud**

- Potential for a much faster RDF for common analysis use cases (for TTree and RNTuple)

The algorithm in a nutshell



Machine learning

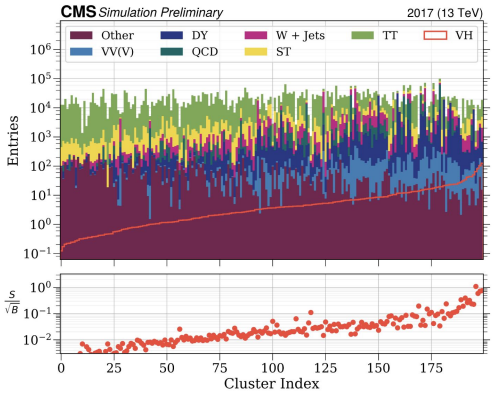


Machine learning (David, Alex)

Binning high-dimensional classifier output for HEP analyses through a clustering algorithm — Niclas Eich

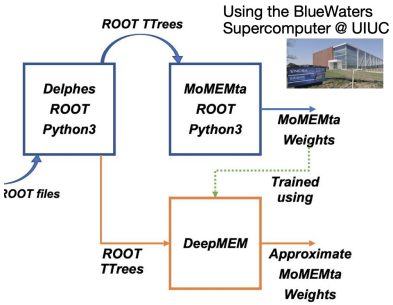
Binning High-Dimensional Classifier Output through Clustering:

- **Agnostic** binning algorithm in high-dimensional space
- No loss of information
- **Improvement of analysis sensitivity**



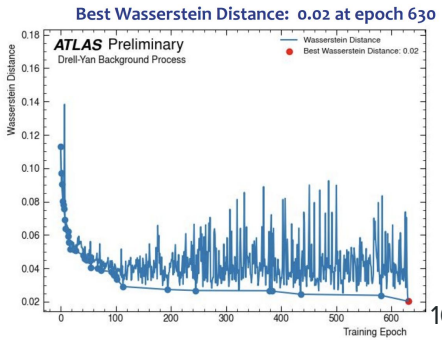
Deep Learning for Matrix Element Method — Mark Neubauer

- MEM computationally expensive ➜ approximate using deep learning



Deep generative models for generating Drell-Yan events in the ATLAS collaboration at the LHC — Xiangyang Ju

- Generate background for the H→μμ decay in ATLAS

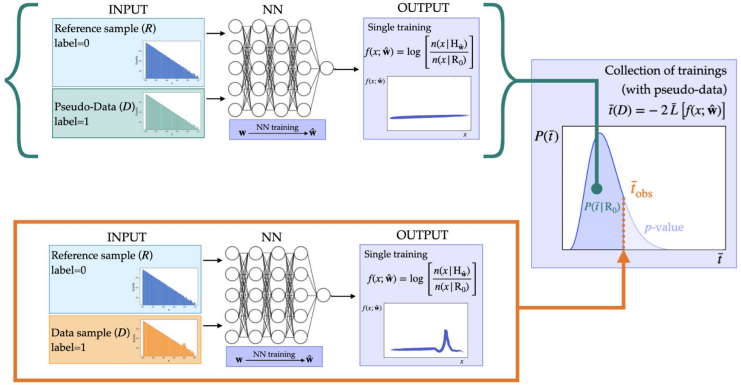


Machine learning (David, Alex)

Unbiased detection of data departures from expectations with machine learning

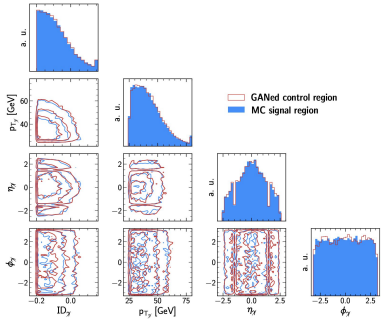
Gaia Grosso

- Test goodness of fit using NPLM (New Physics Learning Machine)
- Approximate likelihoods using ML



Data driven background estimation in HEP using Generative Adversarial Networks

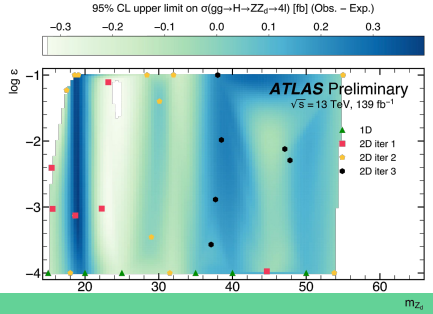
- GANs for generation of misidentified photon objects



Victor Lohezic

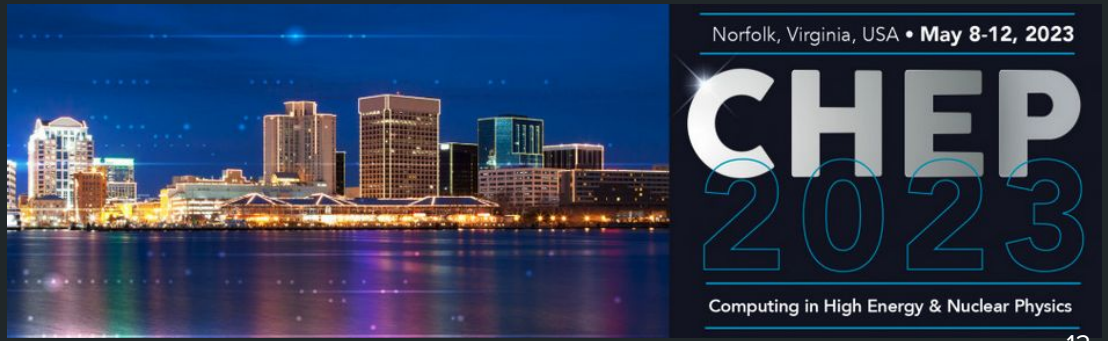
An Active Learning application in a dark matter search with ATLAS PanDA and iDDS

- active learning pipeline for efficient limit setting



Christian Weber

Reconstruction and amplitude fitting

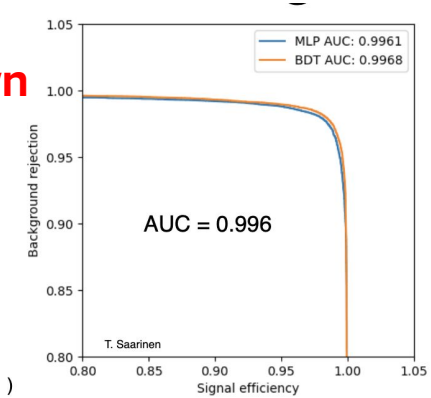


Reconstruction & amplitude fitting (David, Nicole)

KinKal: A Kinematic Kalman Filter Track Fit Package

David Brown

- Toolkit for precision low-momentum track fitting



Laurelin: A ROOT I/O Implementation for Apache Spark

- Java-based ROOT I/O implementation

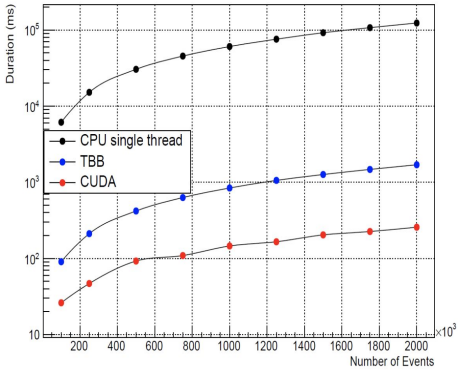
Andrew Melo

Medusa

Alessandro Ricci

HYDRA
Multithreaded Data Analysis Framework

- Application to perform physics analyses of 4-body decays
- Based on Hydra - transparently exploits OpenMP, CUDA, and TBB, allowing the user to re-use the same code



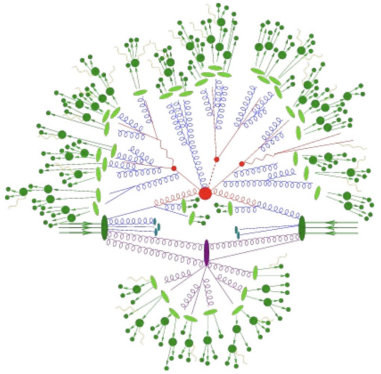
Reconstruction & amplitude fitting (David, Nicole)

Generator Tuning with MC uncertainties

Xiangyang Ju

- Examined the impact of adding the MC uncertainties into the Generator tuning

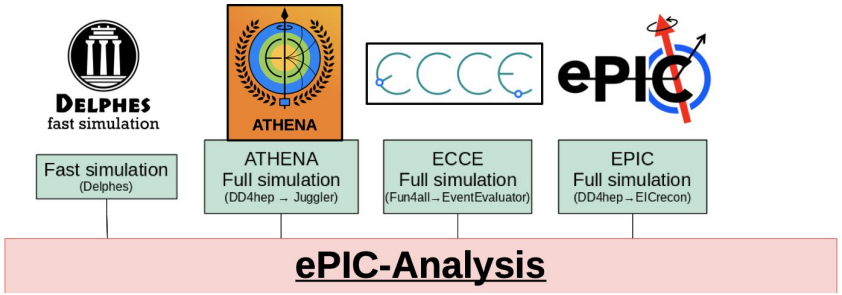
Simulation of a proton-proton collision

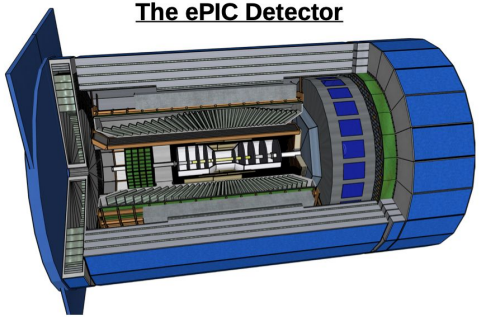


ePIC: Common Physics Analysis Software for the EIC

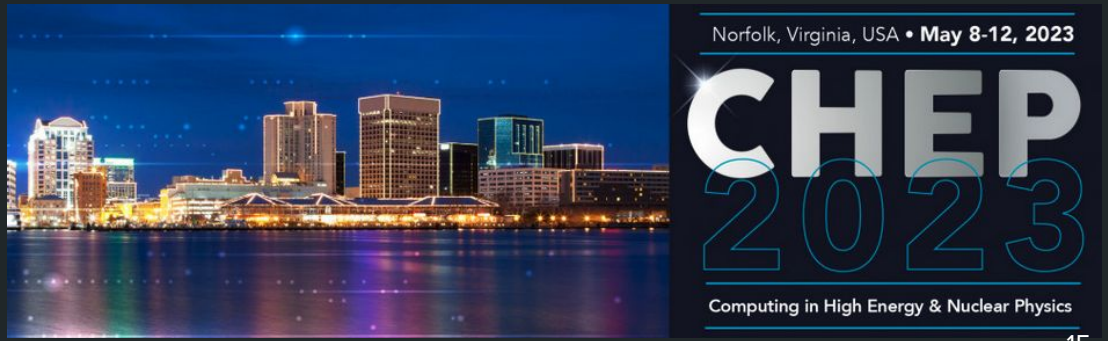
- Integrates various existing frameworks
- Directed Acyclic Graphs for N-D binning problem

Christopher Dilks

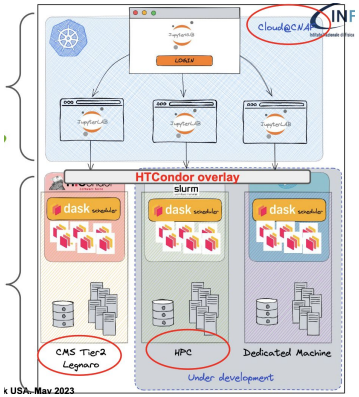




Physics analysis workflows



Physics analysis workflows (Nicole, Stephan)



Benchmarking distributed-RDataFrame with CMS analysis workflows

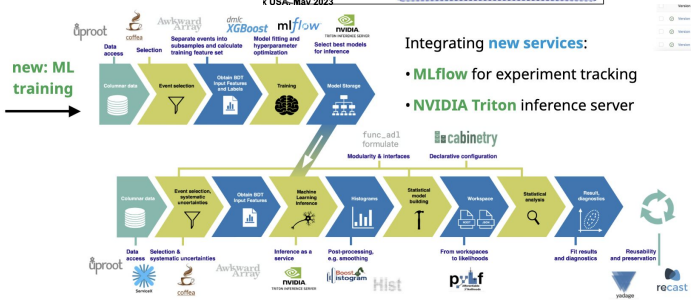
Daniele Spiga

- Factor 8 speedup (a lower limit) - interests from users is growing

Analysis Grand Challenge

Alexander Held

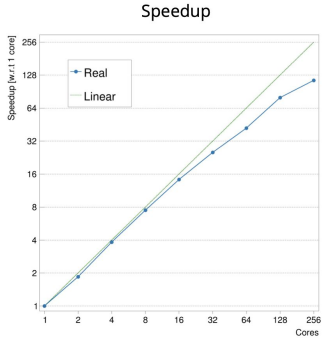
- Define a physics analysis task of **realistic scope & scale**
- Develop **analysis pipelines** that implements the task



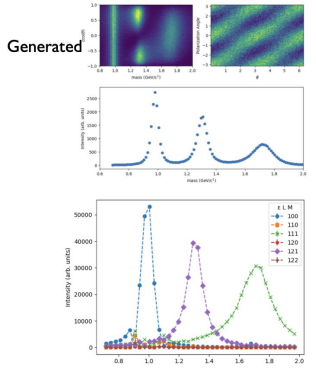
Vincenzo Padulano

Analysis Grand Challenge with a fully Pythonic RDataFrame

- "Multithreading or distributed execution **just work**"



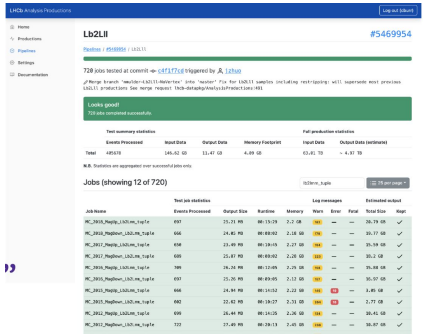
Physics analysis workflows (Nicole, Stephan)

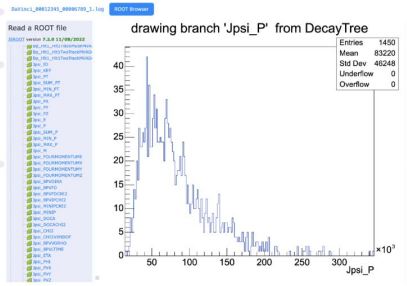


PyPWA: A Software Toolkit for Parameter Optimization and Amplitude Analysis

Mark Jones

- Flexible toolkit for amplitude analysis in multi-particle final states





Chris Burr

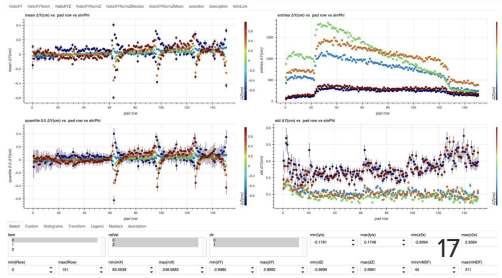
Analysis Productions: A declarative approach to ntupling

- Handles the processing of LHCb data
- "I often use the CI instead of local tests because it's so much easier to read"

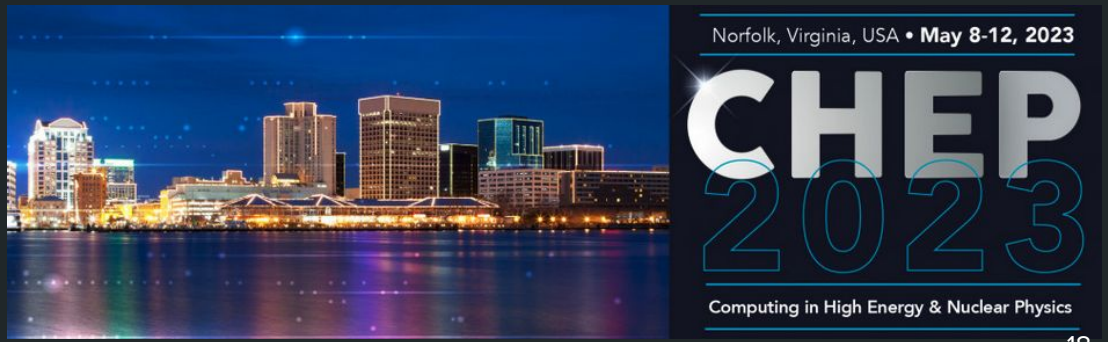
ROOTInteractive

Giulio Eulisse

- Sophisticated visualisations for detector experts in ALICE



New tools and experiences

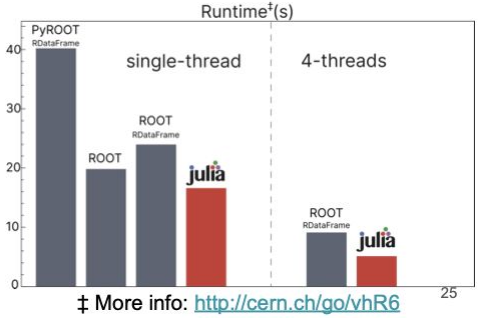


New tools and experiences in analysis (Stephan, David)

Jerry Ling

High-performance end-user analysis in pure Julia programming language

- Managed to implement ATLAS analysis in Julia



ServiceX

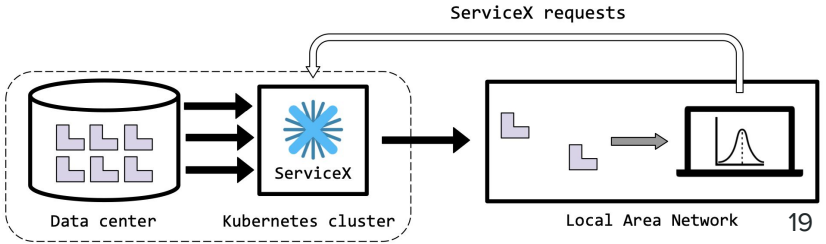
Ben Galewsky

Extracting Columnar Event Data From Experiment Specific Data Formats At Scale

KyungEon Choi

Data Management Package for the novel data delivery system, ServiceX, and Applications to various physics analysis workflows

- ServiceX runs remotely and delivers small subsets of experiment-specific data to "your laptop"
- Finds the files for you in rucio



New tools and experiences in analysis (Stephan, David)

<u>Awkward Just-In-Time (JIT) Compilation: A Developer's Experience</u>

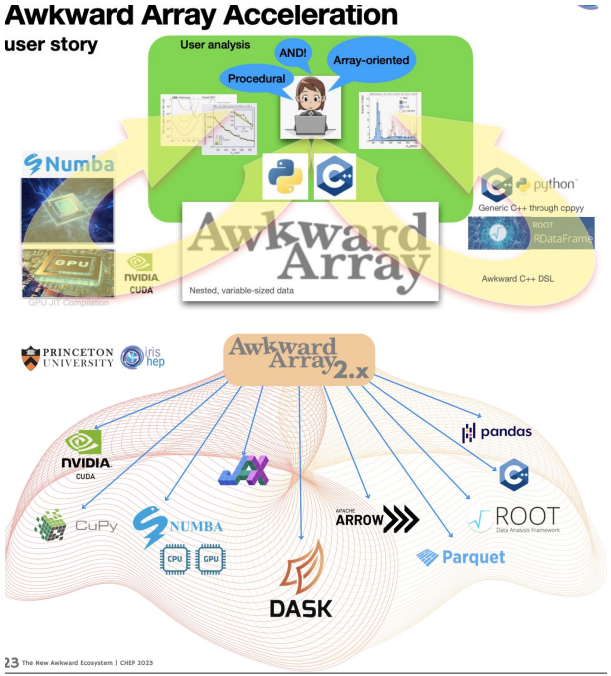
Henry Schreiner

- Conversion to and from RDataFrame
- Numba and cuda support

<u>The New Awkward Ecosystem</u>

Ioana Ifrim

- Multitude of conversion facilities
- Task graphs, CUDA, DASK, JAX



Philippe Canal

<u>Interpreting C++20, with profiling and debugging</u>

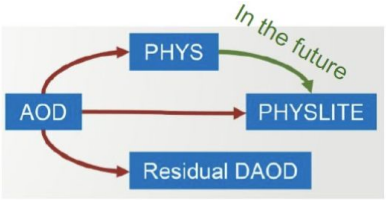
- cling (foundation of python bindings and I/O in HENP) ➔ clang
- Profile and debug interpreted code!

New tools and experiences in analysis II (Stephan, Alex)

Nils Krumnack

CP Algorithms: A common corrections framework for ATLAS

- "Specify what you want; not how you get it"
- Space-efficient tracking of systematic uncertainties



Jana Schaarschmidt

PHYSLITE - a new reduced common data format for ATLAS

- Smaller, lighter data format enabling physics analysis in Run 4
- To cover 80% of all analysis cases
- Saves CPU and storage

Target size	MC	Data
PHYS	50	30
PHYSLITE	12	10

Nils Krumnack

Columnar analysis and on-the-fly analysis corrections at ATLAS

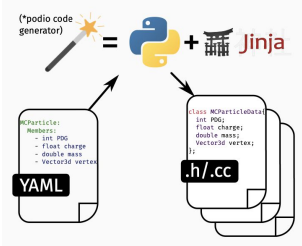
- Early stages. Goal: running on-the-fly directly from PHYSLITE in Run 4
- Correctionlib and triple-use CP tools studied to handle systematic uncertainties

New tools and experiences in analysis II (Stephan, Alex)

Thomas Madlener

podio v1.0 - A first stable release of the EDM toolkit

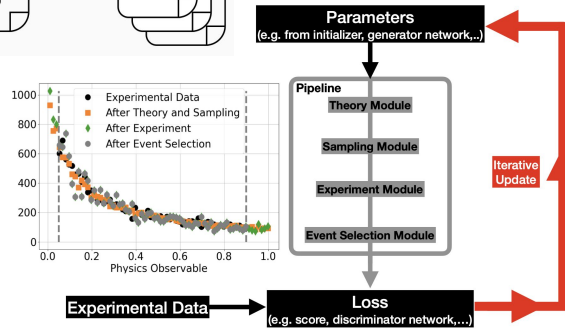
- Enabling the definition of EDMs
- "Heard at this conference: In use by several communities"



Daniel Lersch

The QuantOm Event-Level Inference Framework

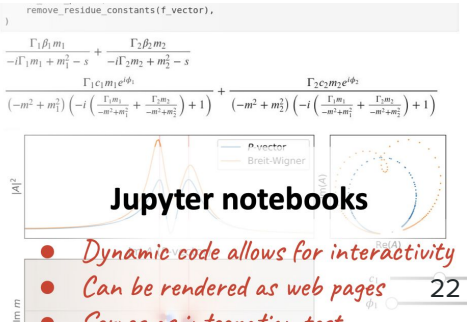
- QUAntum chromodynamics Nuclear TOMography Collaboration
- Goal to extract Quantum Correlation Functions at event level



Remco de Boer

The ComPWA project

- Speeding up amplitude analysis with a Computer Algebra System
- Write symbolic expressions; generate code for amplitude analysis in multiple programming languages



What I am taking away

- We will take a leap forward in terms of I/O with RNTuple
- There is a variety of analysis tools available to analyzers
  - Since a while!
  - But now not only the developers are talking to each other – the software does, too
  - And many are doing amazing things
- We are trying to take away the nitty-gritty details from the average analysis person, and sometimes we manage!
  - Calibrate your objects
  - Find your files
  - Derive your data
  - Schedule your jobs – run in a distributed way
- Machine learning is everywhere; not only in Track 9
- Julia entered the stage
- CHE*N*P is less catchy, but the "N" is certainly noticeable

Thank you to all speakers of track 6!