# Track 6: Physics Analysis Tools Summary

<u>Stephan Hageboeck</u> David Heddle Alex Held Nicole Skidmore





Computing in High Energy & Nuclear Physics

## Statistical Inference and Fitting



## Statistical inference & fitting (Stephan, Alex)

## RooFit's new heterogeneous computing backend

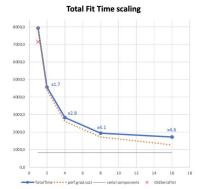
- Revised CUDA backend in ROOT 6.28.04!
  - Gives you great speedup for wide range of unbinned fits with many events
  - Average speedup of 25x (up to 40x!) in fits with 1M events on GeForce RTX 3070

## Automatic Differentiation Applied to RooFit

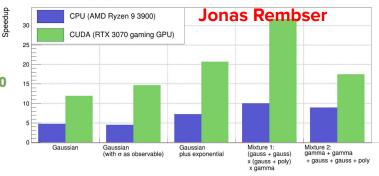
- Usage of clad for differentiation of RooFit models
- Advantages over numerical differentiation

## Build-a-Fit: RooFit Parallelisation and Benchmarking Tools

 Likelihood-gradient parallelisation to improve Higgs fits
 Zef Wolffs



#### RooFit: speedup in benchmark fits with BatchMode() relative to old RooFit (1 million events)





## Statistical inference & fitting (Stephan, Alex)

## New developments in Minuit2

- 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

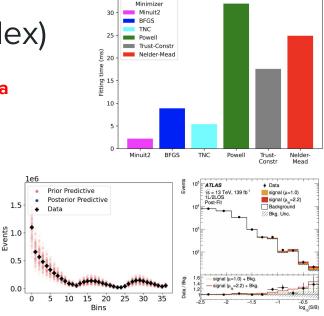
## Bayesian Methodology for Particle Physics with pyhf

Parallel Bayesan / Frequentist interface for HistFactory models in pyhf Malin Horstmann

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

Employing k-nearest neighbour method to assign event guality factors





Fitting time for scipy minimizers

Data GLUE Data  $\times O$ .  $M_{-}(GeV/c^2)$ 

#### **Zachary Baldwin**

0.5

0.0

## I/O and data formats



I/O and data formats (David, Nicole)

## **DUNE HDF5 Experience**

HDTree

#### **Barnali Chowdhury**

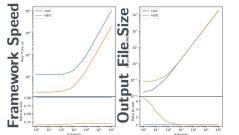
- HDF5 will be used as raw data file format in ProtoDUNE(s)-II.
  - Have successfully taken data in HDF5 from HD, VD Cold Box testing. ND-LAr Module0 data is in HDF5 format. 0

## Schema-Evolution and the TTree within HDF5

#### Tom Eichlersmith

#### Comparison to ROOT

Comparison Between HDF5 and ROOT : Production Mod



#### industry-standard file format. $\blacksquare$ C++ API currently available with performance on-par (if not exceeding) ROOT.

Utilize TTree-like data organization and access patterns while gaining the benefit of an



## 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

Florine de Geus

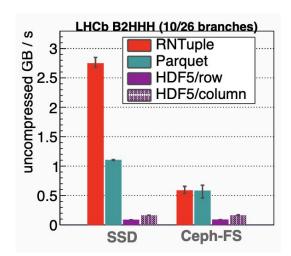
 Work in progress with first successful integration efforts: CMS & ATLAS frameworks, RDataFrame, RBrowser, XRootD, TTree data importer

## Integration of RNTuple in ATLAS Athena

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

#### 

Size on disk, zstd compressed "final-stage" ntuples



## 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

## Boosting RDataFrame performance with transparent bulk event

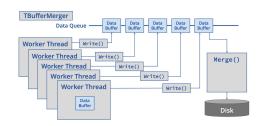
Enrico Guiraud

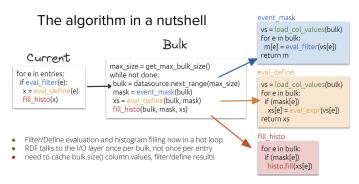
#### processing

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

#### 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;





## Machine learning





Computing in High Energy & Nuclear Physics

## Machine learning (David, Alex)

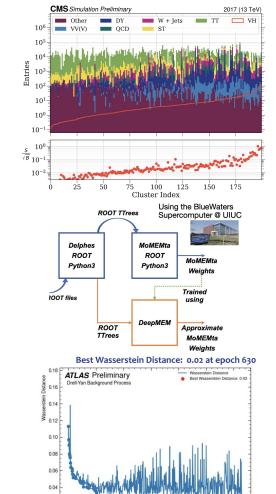
## Binning high-dimensional classifier output for HEP analyses through a clustering algorithm Niclas Eich

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
 Kiangyang Ju
 Generate background for the H→µµ decay in ATLAS



Training Epoch

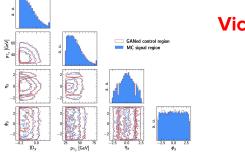
## Machine learning (David, Alex)

## Unbiased detection of data departures from expectations with machine learning Gaia Grosso

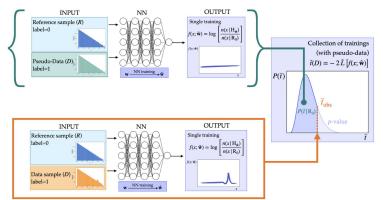
- Test goodness of fit using NPLM (New Physics Learning Machine)
- Approximate likelihoods using ML

## Data driven background estimation in HEP using Generative Adversarial Networks

GANs for generation of misidentified photon objects

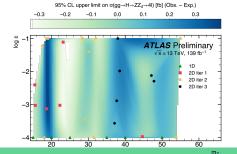


#### **Victor Lohezic**



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

active learning pipeline for efficient limit setting

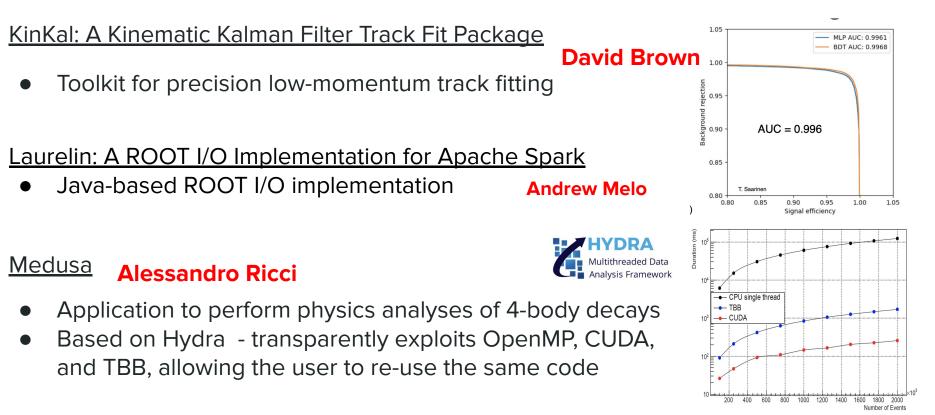


#### **Christian Weber**

## Reconstruction and amplitude fitting



## Reconstruction & amplitude fitting (David, Nicole)



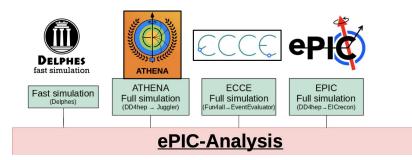
## Reconstruction & amplitude fitting (David, Nicole)

## Generator Tuning with MC uncertainties

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

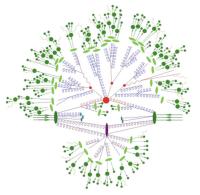
## ePIC: Common Physics Analysis Software for the EIC

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



# The ePIC Detector

Simulation of a proton-proton collision



**Christopher Dilks** 

**Xiangyang Ju** 

## Physics analysis workflows



## Physics analysis workflows (Nicole, Stephan)

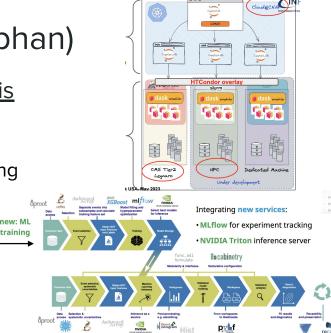
Benchmarking distributed-RDataFrame with CMS analysis workflows Daniele Spiga

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

Analysis Grand Challenge

**Alexander Held** 

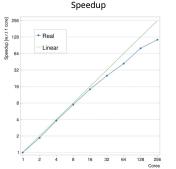
- Define a physics analysis task of realistic scope & scale
- Develop **analysis pipelines** that implements the task



## Vincenzo Padulano

Analysis Grand Challenge with a fully Pythonic RDataFrame

• "Multithreading or distributed execution just work"



16

## Physics analysis workflows (Nicole, Stephan)

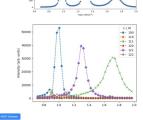
PyPWA: A Software Toolkit for Parameter Optimization and Amplitude Analysis Mark Jones

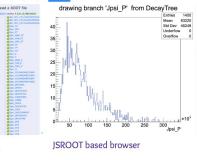
• Flexible toolkit for amplitude analysis in multi-particle final states

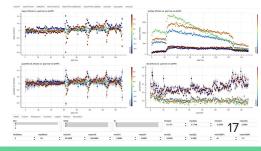
#### Chris Burr Analysis Productions: A declarative approach to ntupling

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

Lb	2LII								#546	9954
kel	NO / #5188856 / LE2LU									
28	iobs tested at commit -e-	cdf1f7cd bio	erecity & 12	tun						
	nge brænch 'mmulder-Lb2LT Li productions See menge n	L-NeVertex' inte	'nester' Fix #	er USUI se		reitrippi	gi witt	нреги	de most pres	ines.
	aks good! Liste conciened successfully.									
	Test summary statistics		Full production statistics							
	Events Processed	Mput Dota	Ourput Data	Memory	Foetprint	Input Das	• •	utput Da	a (astinate)	
Tete	485578	145.62 65	11.47 68	4.09 68		63,91 75		4,97 10		
	Salation are appropriated over a					ibline, t	.pia		12 25 pc	Della -
			elation				.çis iessages		i = 25 per	
Jot		20)		out Size . No	ertime Meen	Logo		ficial	-	
Job	as (showing 12 of 7	20) Test job s	cound dury		erdine Meen 13:29 2.2	Leg r		hai -	Estimated o	land
Job Job	os (showing 12 of 7	20) Test job o Events Pr	consoid dury 23-3	1. PS 84		Log a bry Wars	iessages Errer		Estimated o	kipul Kept
	os (showing 12 of 7 Name NAAJShajiy,1,521,m_1+p1e	20) Test jak a Eventa Po 697	consod dury 23.3 24.4	11.145 84 15.140 84	13:29 2.2	Loga ory Work 28 1981 50 109	iessages Errer	-	Estimated a focal Siza 28, 79–68	Kopt V
	as (showing 12 of 7 Name 1938,Nepp_Lb7Lem_topte 1938,Nepp_Lb7Lem_topte	20) Test job 6 Eventa Po 693 666	connol dury 23.2 24.4 23.4	11 MS 84 15 MB 86 10 MS 86	13:29 2.2 (82:02 2.16	Lege ary Wars 58 Mil 50 CN 58 Mil	trur –	-	Potimolog o Rotal Sizo 28, 79 68 29, 77 68	Kape V
Job HC HC HC	as (showing 12 of 7 Name NSS,Displa,Likiter_taple NSS,Displa_Likiter_taple NST,Displa_Likiter_taple	20) Test job e Eventa Po 666 656	cossed dury 23.3 24.4 23.4 23.4	11 PG 00 15 PG 00 17 PG 00	13129 2.2 (88)82 2.18 (39)45 2.27	Log / ory Wars 28 941 50 0% 50 108 60 135	forer Errer — —		Petimolog e Total Size 28, 79 - 68 29, 77 - 68 29, 59 - 68	Kopt Xopt V
	as (showing 12 of 7 None 1933, Pagle, J.M. m., top Ic 1932, Pagle, J.M. m., top Ic 1932, Pagle, J.M. m., top Ic 1935, Pagle, J.M. m., top Ic	20) Test jok e Evente Po 666 656 656 665 665 665 665 665 665 66	consol dury 25.3 24.4 23.4 25.4 26.4 26.4 26.4 26.4 26.4 26.4 26.4	11 P8 84 65 P0 86 69 P8 86 67 P0 86 64 P8 86 55 P0 86	13:29         2.2           (80:92         2.16           (30:45         2.27           (80:92         2.26           (12:45         2.28           (12:45         2.28           (12:45         2.18	Leg / Ery Num 28 94 50 09 50 09 50 108 50 108 5	- - -		Fetimated e Rotal Size 29, 77–68 29, 77–68 29, 77–68 20, 2–68 20, 2–68 30, 97–68	Algert Rage S S S S S S S
	tes (showing 12 of 7 Name BBS, Jegip, J.N. H., typic BBS, Jegip, J.N. H., typic BBS, Jegipe, J.N. H., typic BBS, Jegipe, J.N. H., typic BBS, Jegipe, J.N. H., typic BBS, Jegipe, J.N. H., typic	20) Test jok e Events Pe 605 605 789 605 789 605 605	connot dury 33.3 24.4 23.4 24.4 24.4 24.4 24.4 24.4	11 PS e4 5 P0 e5 10 PS e4 17 P0 e6 14 P8 e6 15 P5 e6 14 P8 e6	13129         2.2           180292         2.18           13045         2.27           180492         2.38           13045         2.27           13045         2.28           13045         2.28           13045         2.28           13045         2.38           13045         2.38           13045         2.38           13045         2.38           13045         2.38           13045         2.32	Lega 27 Norm 28 No. 29 Norm 29 Norm 20 No. 20 No	Errer - - -		Petimeted e Rotal Sile 20, 79 48 20, 79 48 21, 79 48 23, 79 48 23, 59 48 30, 27 48 3, 45 68	Kapt Xapt V V V V V V
	tes (showing 12 of 7 Name 1993;Unspipe, LixXim, Linfe 1993;Yappine, LixXim, Linfe 1993;Yappine, LixXim, Linfe 1993;Yappine, LixXim, Linfe 1993;Unspipe, LixXim, LixXim 1993;Unspipe, LixXim, LixXim	20) Test jok e Events Pe 605 605 789 605 789 605 605 605 605 605 605 605 605 605 605	connot dury 25.3 24.4 25.4 25.4 25.4 25.4 25.4 25.4 25.4	11 P8 04 5 P0 05 10 P8 04 17 P0 05 14 P8 04 15 P8 04 15 P8 04 14 P8 05	131/29         2.2           082/02         2.16           130/05         2.27           080/02         2.08           120/05         2.28           120/05         2.12           130/05         2.12           100/05         2.12           112/05         2.29           112/05         2.12           112/05         2.12           112/05         2.12           112/05         2.12           112/05         2.12           112/05         2.12	Leg / Norm 28 991 59 091 50 091 50 091 50 091 50 091 50 091	- - -	-	Patimated o Social Size 20, 77 68 20, 78 68 20, 27 68 20, 27 68 3, 48 68 3, 48 68 2, 77 68	Agent Kapt S S S S S S S S S S
	es (showing 12 of 7 https://www.es/status/ #83.5%spig.1.50.0%, confer #82.5%spig.1.50.0%, confer #82.5%spig.1.50.0%, confer #83.5%spig.1.50.0%, confer #83.5%spig.1.5%spig	20) Test job e Eventa Po 693 665 695 693 693 693 693 693 695 695 695	connet durg 25.3 24.4 25.4 25.4 25.4 25.4 25.4 25.4 24.4 22.4 26.4	11 PS 04 5 P0 05 10 PS 04 17 P0 05 17 P0 05 15 PS 05 14 P0 05 14 P0 05	(13)(29         2,2           (13)(29         2,14           (13)(15)         2,27           (13)(15)         2,28           (12)(15)         2,28           (13)(15)         2,28           (13)(15)         2,21           (13)(15)         2,22           (13)(15)         2,22           (13)(15)         2,23           (13)(15)         2,31           (13)(15)         2,36	Leg r avy Nexa 34 981 56 091 56 994 56 995 56 995 56 996 56 996			Patimated o Social Size 20, 79 65 20, 79 65 20, 79 65 20, 20 65 2, 85 68 2, 85 68 2, 77 68 20, 41 69	
	tes (showing 12 of 7 Name 1993;Unspipe, LixXim, Linfe 1993;Yappine, LixXim, Linfe 1993;Yappine, LixXim, Linfe 1993;Yappine, LixXim, Linfe 1993;Unspipe, LixXim, LixXim 1993;Unspipe, LixXim, LixXim	20) Test jok e Events Pe 605 605 789 605 789 605 605 605 605 605 605 605 605 605 605	connet durg 25.3 24.4 25.4 25.4 25.4 25.4 25.4 25.4 24.4 22.4 26.4	11 PS 04 5 P0 05 10 PS 04 17 P0 05 17 P0 05 15 PS 05 14 P0 05 14 P0 05	131/29         2.2           082/02         2.16           130/05         2.27           080/02         2.08           120/05         2.28           120/05         2.12           130/05         2.12           100/05         2.12           112/05         2.29           112/05         2.12           112/05         2.12           112/05         2.12           112/05         2.12           112/05         2.12           112/05         2.12	Leg r avy Nexa 34 981 56 091 56 994 56 995 56 995 56 996 56 996			Patimated o Social Size 20, 77 68 20, 78 68 20, 27 68 20, 27 68 3, 48 68 3, 48 68 2, 77 68	Agent Kapt S S S S S S S S S S







#### **ROOTInteractive**

### **Giulio Eulisse**

• Sophisticated visualisations for detector experts in ALICE

## New tools and experiences



## New tools and experiences in analysis (Stephan, David)

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

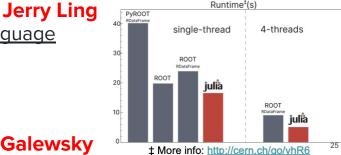
Managed to implement ATLAS analysis in Julia

## **ServiceX**

## Extracting Columnar Event Data From Experiment Specific Data Formats At Scale

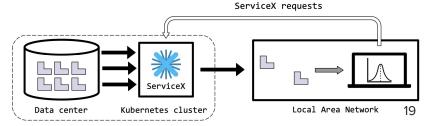
Data Management Package for the novel data delivery system, ServiceX, and Applications to various physics analysis workflows

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

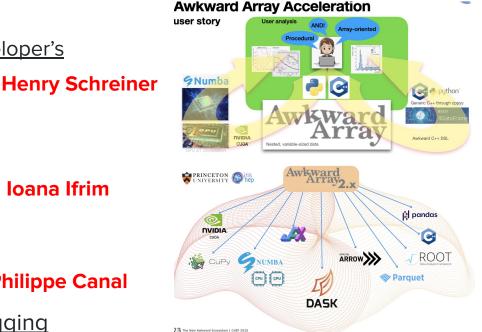


Ben Galewsky

KyungEon Choi



## New tools and experiences in analysis (Stephan, David)



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

- Conversion to and from RDataFrame
- Numba and cuda support

## <u>The New Awkward Ecosystem</u>

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

## Philippe Canal

## Interpreting C++20, with profiling and debugging

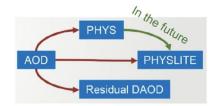
- cling (foundation of python bindings and I/O in HENP)  $\rightarrow$  clang
- Profile and debug interpreted code!

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

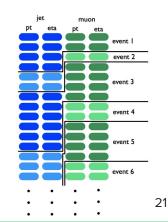
## **Nils Krumnack**

## CP Algorithms: A common corrections framework for ATLAS

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



Target size	MC	Data			
PHYS	50	30			
PHYSLITE	12	10			



#### 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

## **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"

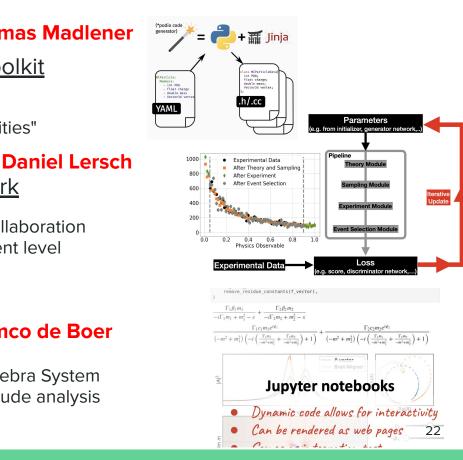
## The QuantOm Event-Level Inference Framework

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

## The ComPWA project

#### Remco de Boer

- 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
  - $\circ$  Schedule your jobs run in a distributed way
- Machine learning is everywhere; not only in Track 9
- Julia entered the stage
- CHENP is less catchy, but the "N" is certainly noticeable

# Thank you to all speakers of track 6!

