UNIVERSITAS STUDII PADUANI MCCXXII

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Current research & interests

Mikele Milia

ESC22 @Bertinoro

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Present
March 2026

Ph. D. Student @ Department of Information Engineering, University of Padova

Research - Analysis of genetic data and integration with clinical and other omic data in patients with Amyotrophic Lateral Sclerosis

Supervisor
Barbara Di Camillo

February 22
September 22

Research Grant (Type A) @ Department of Information Engineering, University of Padova

Research - Analysis of transcriptomics and genetics data, with development of Python & R software

Supervisor
Barbara Di Camillo

March 19
October 19

M. Sc. in Computer Engineering @ Department of Information Engineering, University of Padova

Thesis - Data-driven approach to inform a multi-agent spatio-temporal simulator of tumor micro-environment

Supervisor
Barbara Di Camillo

September 14
July 19

B. Sc. in Computer Engineering @ Department of Information Engineering, University of Padova

Thesis - Qualitative analysis of quality value smoothing and its impact on the compression of High Throughput Sequencing data

---

September 09
June 14

High School Diploma in Computer Science @ ITIS Severi

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

WALKTHROUGH

Focused on?

Analysis of genetic data and integration with clinical and other omic data in patients with Amyotrophic Lateral Sclerosis

Why it is important?

ALS is a progressive neurodegenerative disease with unknown cause

At most **10%** cases have known **genetic** etiology or family history

Other omics could explain yet unknown aspects of ALS

In which data am I interested?

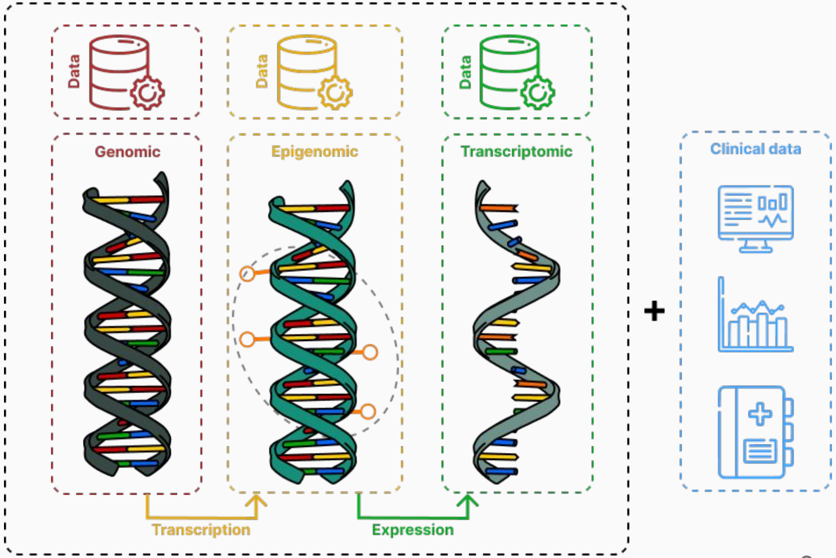
Genomics data can reveal and explain disease predisposition

Epigenomics data can reveal how environment affects DNA folding

Transcriptomics data can quantify gene expression levels (i.e., response to environment)

Clinical data can provide informative covariates

Omics-data integration and analysis



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

OMICS DATA

Genomics data (static information)

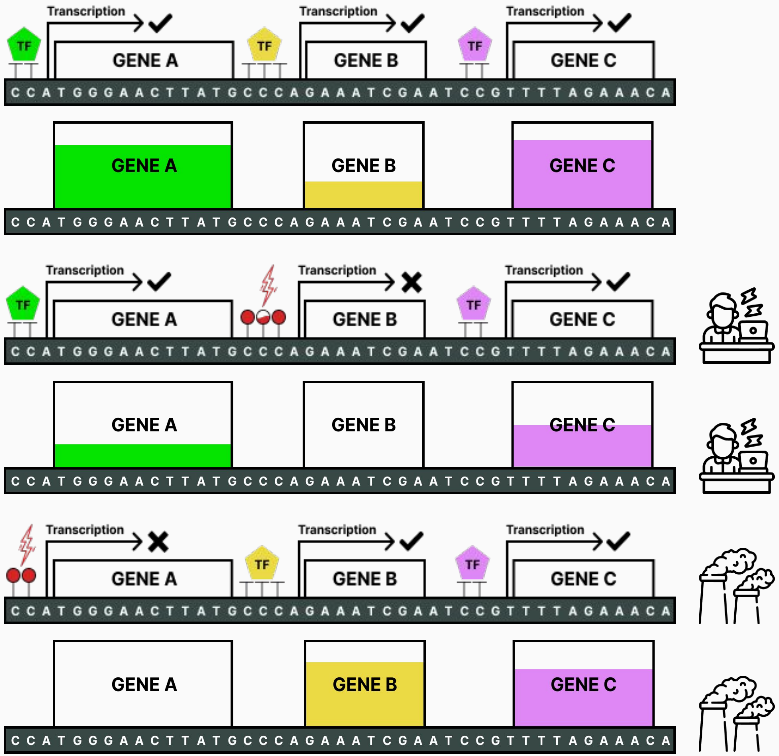
- contains information about **genotype**
- gene variations can be discretized as **0 / 1 / 2**

Epigenomics data (quasi-static information)

- contains information about **environmental effects on DNA**
- epigenetic modification can be:
  - discretized as **0 / 1 / 2**
  - expressed as **epigenetic marker level**

Transcriptomics data (dynamic information)

- contains information about transcribed genes w.r.t **environmental effects on DNA**
- provides **genes expression level**



01 MAIN RESEARCH

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

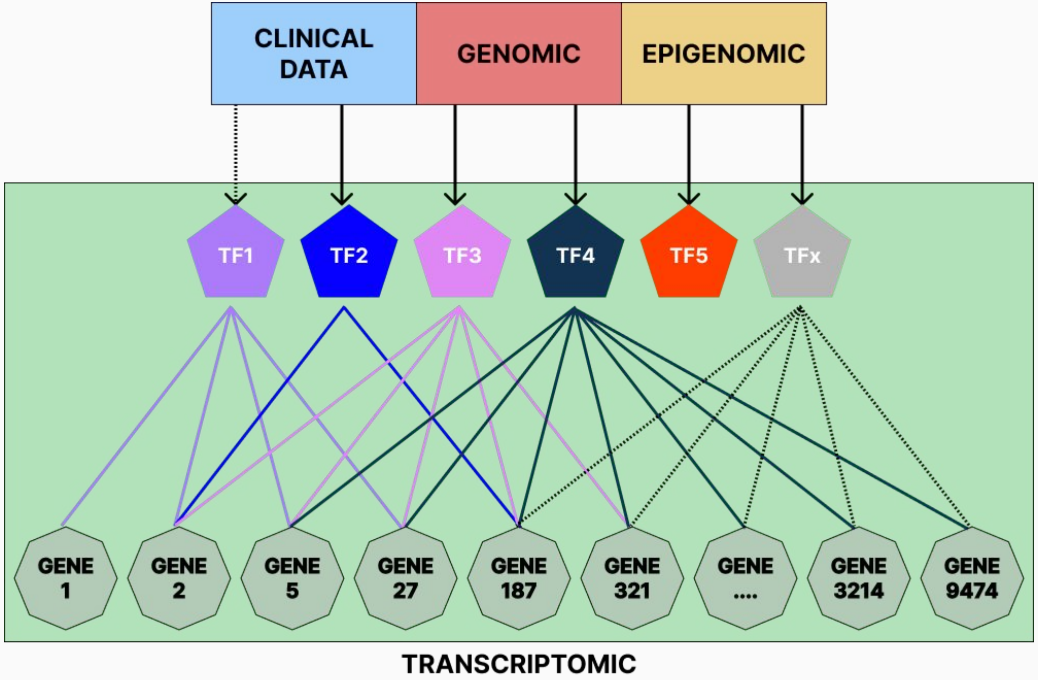
Integration and analysis of clinical and multi-omics ALS data through the development of computational methods and pipeline

HOW

Analysis of (epi)genomic data

Analysis and integration of epigenomic, transcriptomic and genomic data

Gene Regulatory Network reconstruction



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

My main interests are …

- Efficient large-scale data (e.g., genomics, …) analysis and integration
- Optimization algorithms (e.g., Genetic, Ant Colony , Tabu Search, …)
- Cluster computing
- (Un)supervised Machine Learning Algorithms
- Deep Learning Neural Networks

Green Computing for particle physics

Francesco Minarini
francesco.minarini@studio.unibo.it

$whoami

2019
M.Sc in Applied physics

2020
IT consultant for corporate

2021
Research grant @ INFN-CNAF

2022
PhD student @ Unibo with support from E4 computer engineering

$head -15 PhD_project

The overall request for high performance computing resources is ever-growing, for instance:

- *Research* (Big Data + advanced AI becoming a paradigm in many disciplines).
- *Industry* (AI is spreading in every tech business).

Two main issues emerge [1]:

1. *(technical)* If left uncontrolled, the carbon footprint of computing might just be unsustainable and keeping up with computing requirements might be impossible.

2. *(ethical)* Whoever owns the best hardware cuts out competitors and "softlocks" users.

What I'm trying to explore, in co-op with E4 computer engineering, is:

> What hardware architecture can be implemented to curb energy consumption without putting a penalty on the end user performance [2]? Will this be "domain agnostic"?

> What enhancements can be achieved through software? Is there a way to use AI for AI's sake? [3]

$echo Thank You!

[1] Schwartz et al. "*Green AI*" arXiv:1907.10597v3

[2] Bartolini et al. "*Monte Cimone: Paving the Road for the First Generation of RISC-V High-Performance Computers*" arXiv:2205.03725v1

[3] Bartolini et al. "COUNTDOWN: a Run-time library for Performance-Neutral Energy Saving in MPI Applications" arXiv:1806.07258v2

Lightning Talk ⚡

Andy Morris – University of Warwick

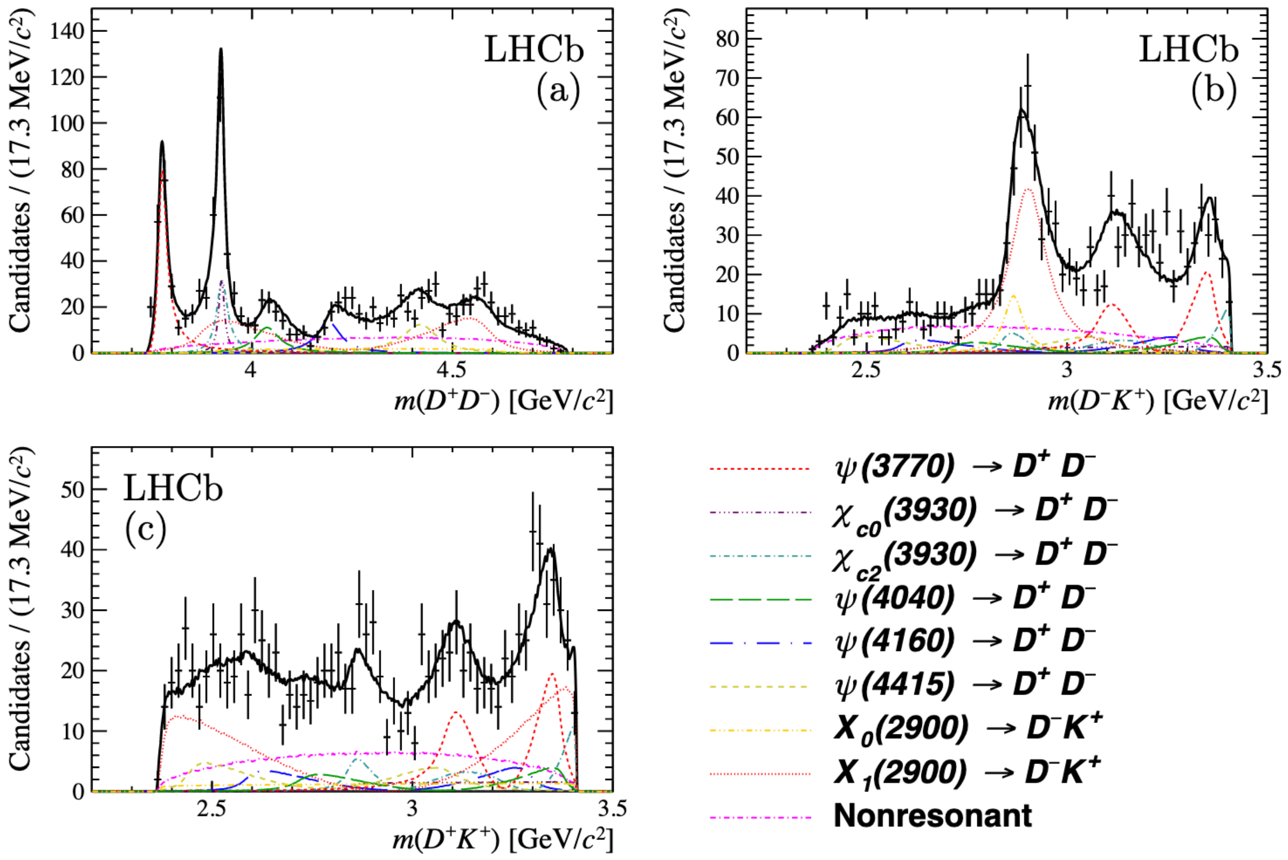
Andy.Morris@warwick.ac.uk – 03/Oct/2022

Intro: Non-TD Laura^{++} Briefly

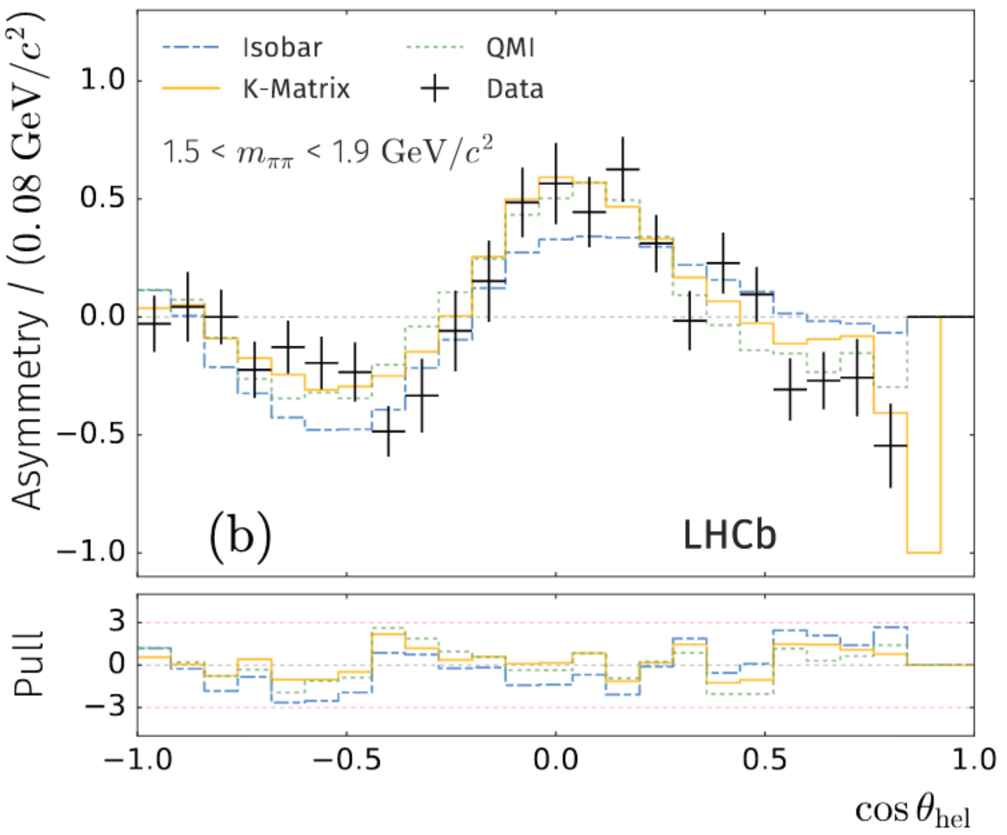
- Laura^{++} is a Dalitz plot fitter maintained mostly by people at or formerly at Warwick [1]
- It's based on Laura which is an older DP fitter written in Fortran
  - Laura^{++} is written in C^{++} 17 and built using CMake
- The current stable build of Laura (v3r5) can model the time-integrated Dalitz distribution of (pseudo)scalar meson three-body decays into three (pseudo)scalars
  - Notable publications using Laura include the recent $B^0 \to DDK$[2] paper as well as $B^+ \to \pi\pi\pi$[3]
- Once a model is fit, Laura can plot the results by generating MC using that model. This is essentially a Monte-Carlo integration of the amplitude model

Plots of pre-existing Laura results

$B \rightarrow DDK$



$B^+ \rightarrow \pi\pi\pi$

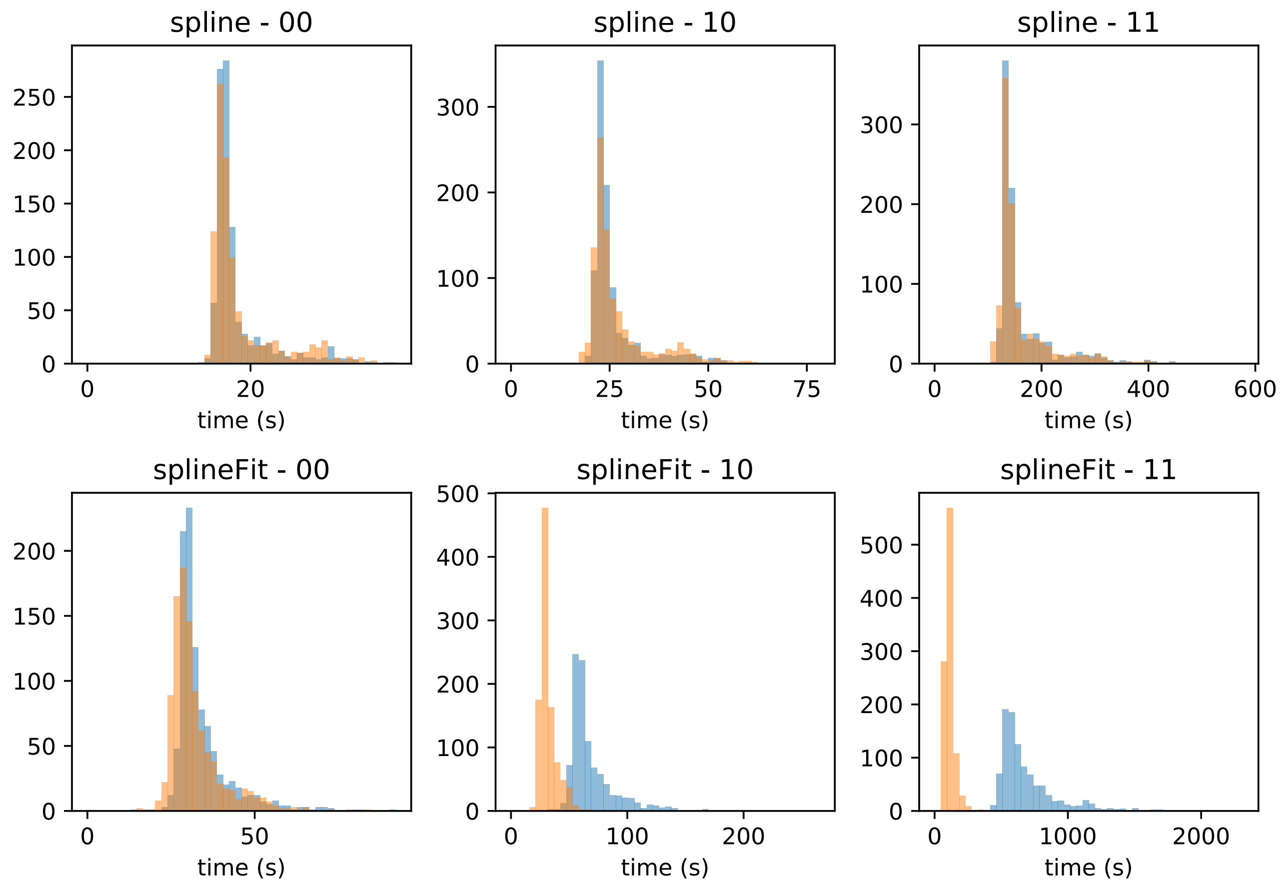


Time dependency in Laura

- The first thing to be considered for the time-dependent model is the theoretical description before any experimental effects are considered
- The theoretical model for time-dependent decays is given:
  - $$\Gamma\left(B_q^0/\overline{B}_q^0 \to f(t)\right) \propto \Big[\left(\left|A_f\right|^2 + \left|\overline{A}_f\right|^2\right)\cosh\left(\frac{\Delta\Gamma_q}{2}t\right) + 2\Re\left(\frac{q}{p}A_f^*\overline{A}_f\right)\sinh\left(\frac{\Delta\Gamma_q}{2}t\right) \pm \left(\left|A_f\right|^2 + \left|\overline{A}_f\right|^2\right)\cos(\Delta m_q t) \mp 2\Im\left(\frac{q}{p}A_f^*\overline{A}_f\right)\sin(\Delta m_q t)\Big]$$
- Here $B_q^0/\overline{B}_q^0$ represents the flavour at time t=0 and A_f represents the amplitude model
- This is then further complicated by the effects of Decay Time Acceptance and Decay Time Resolution

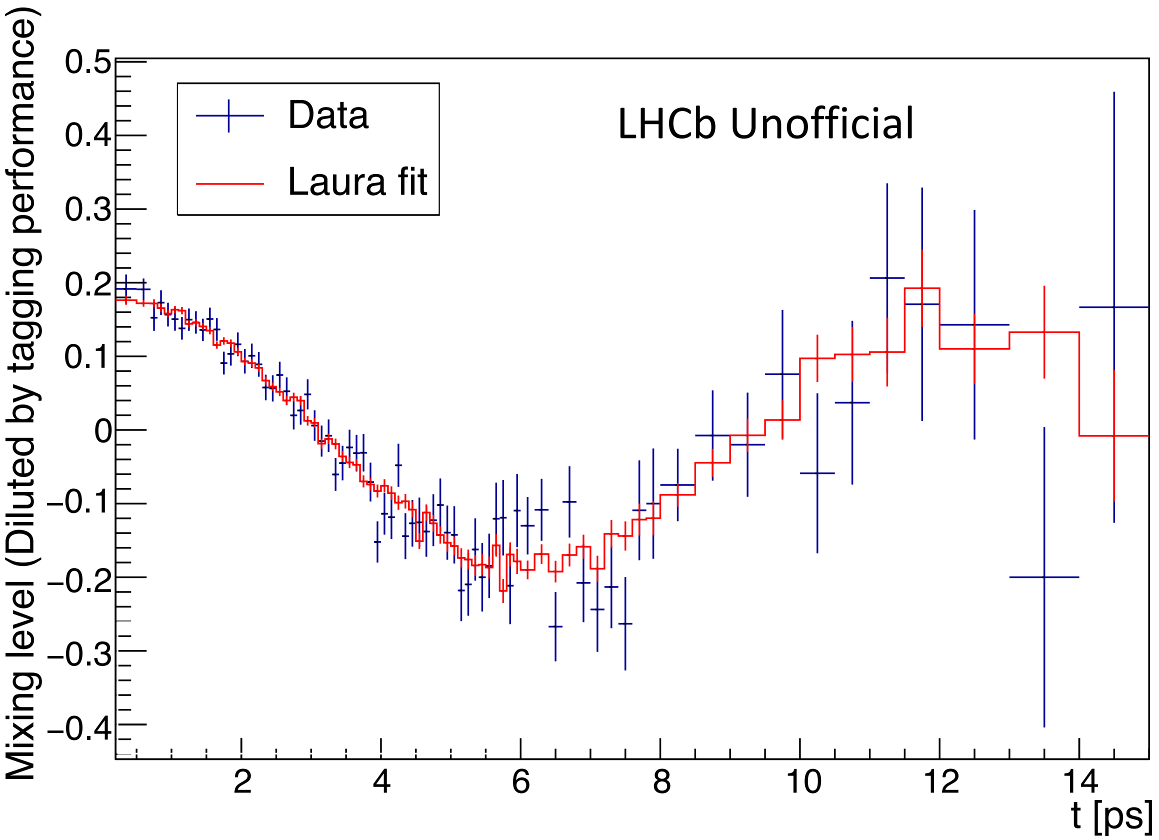
Further speed improvements

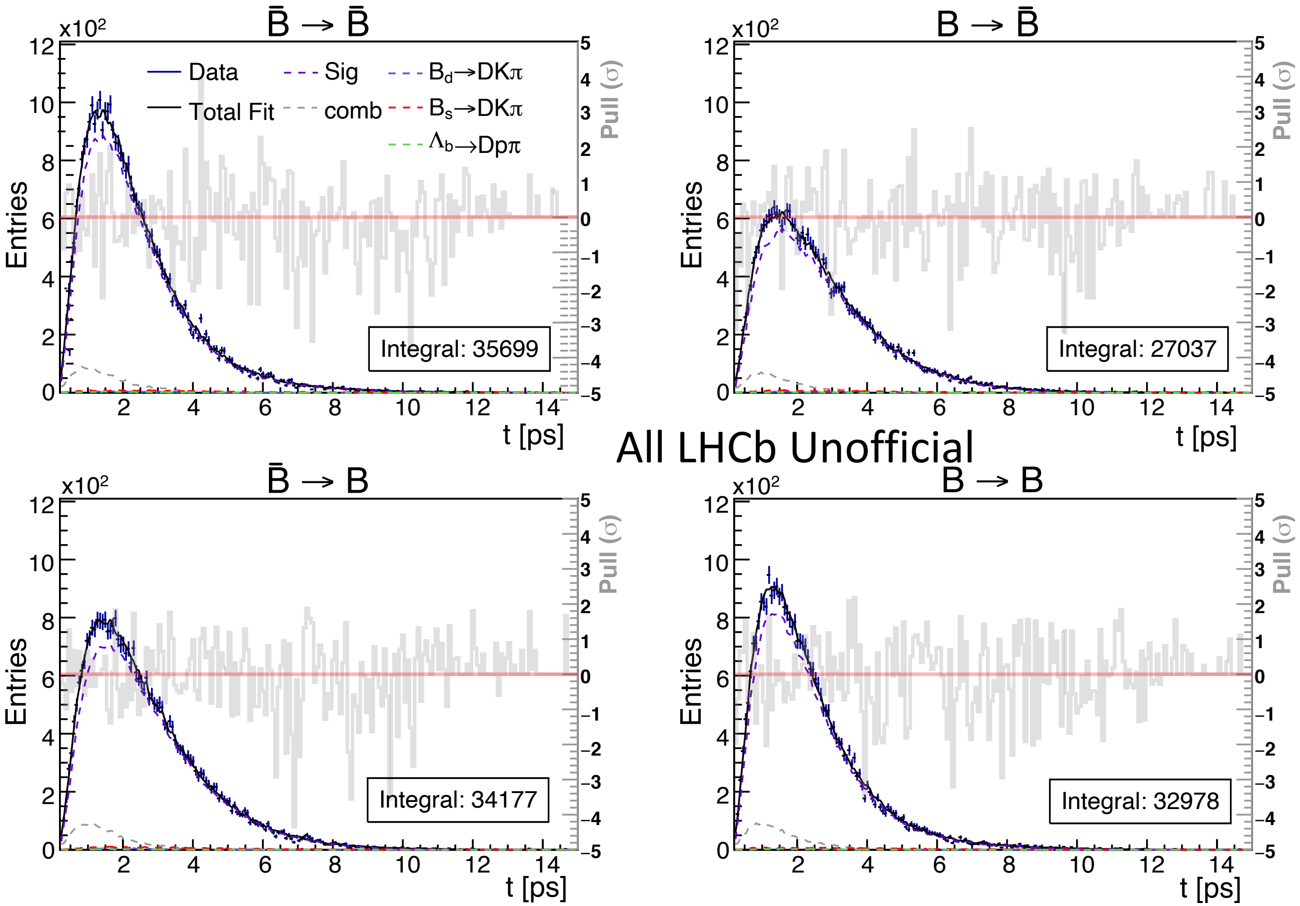
- To speed this up further, in a similar vein as is currently done in non-TD Laura, caching is implemented as much as possible and normalisations are only recalculated if Minuit makes an adjustment to a relevant variable
- In some use cases, execution time was reduced by >80%



Conclusion

- As a result we are able to model time-dependent amplitude analyses



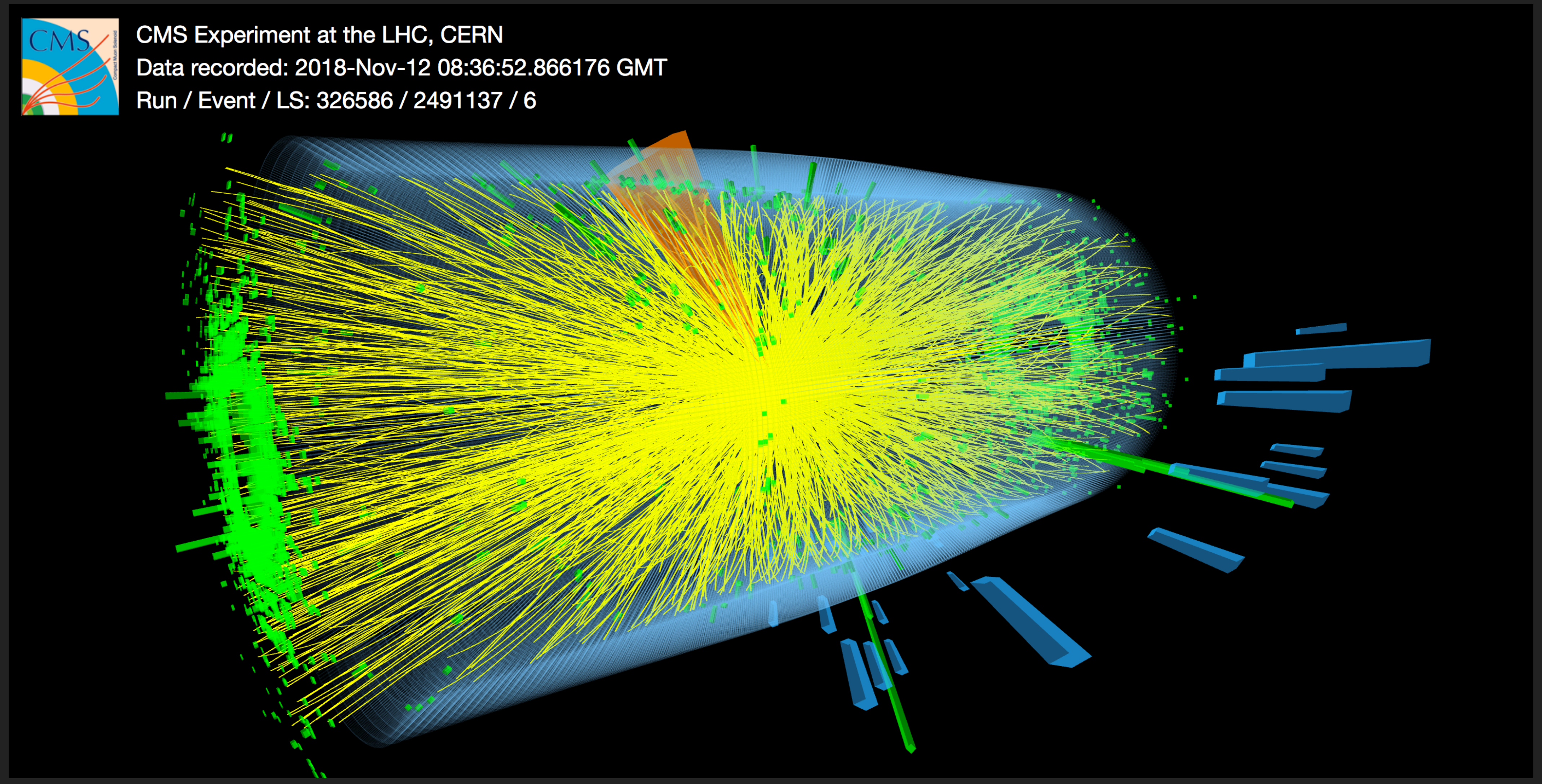


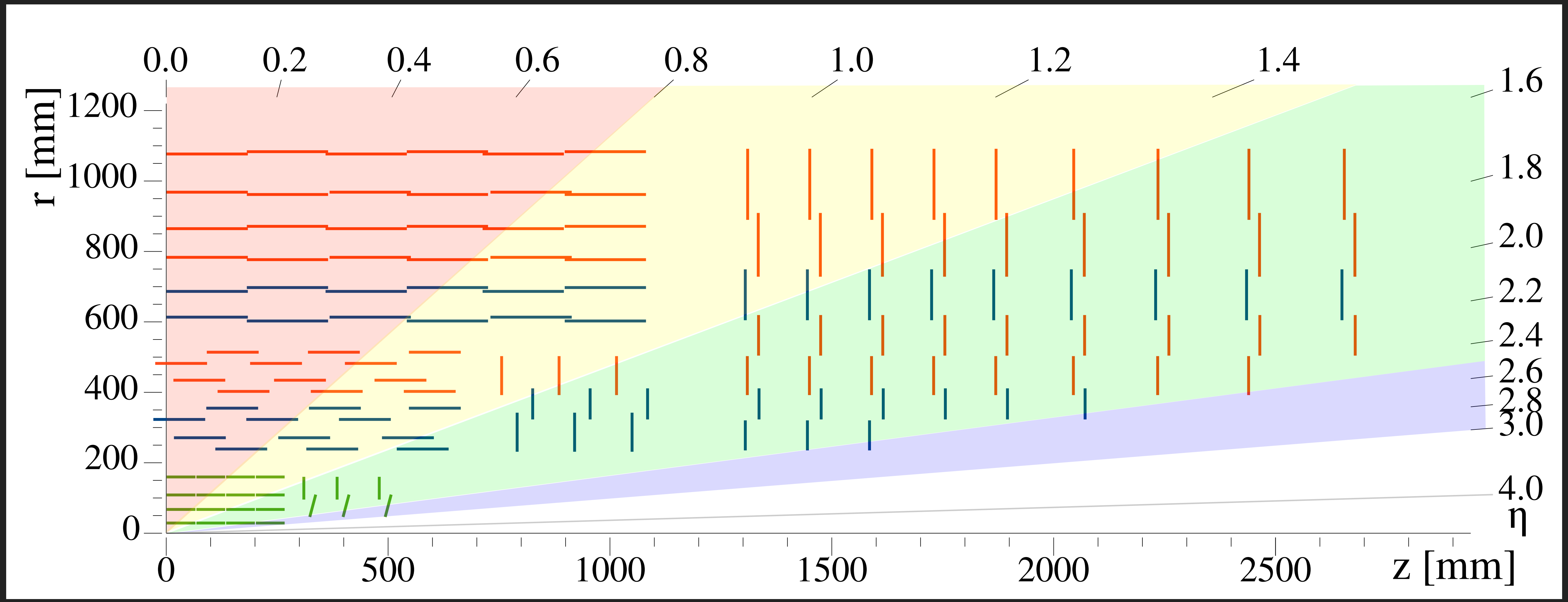
ALESSANDRO PASCOLINI

- Master degree in Pysics at Università degli studi di Perugia
  - “Measurement of Momentum Resolution using Low-mass Resonances”

MEASUREMENT OF MOMENTUM RESOLUTION USING LOW-MASS RESONANCES

- Run-2 environment
  - High Pile-Up & $\sqrt{s}$



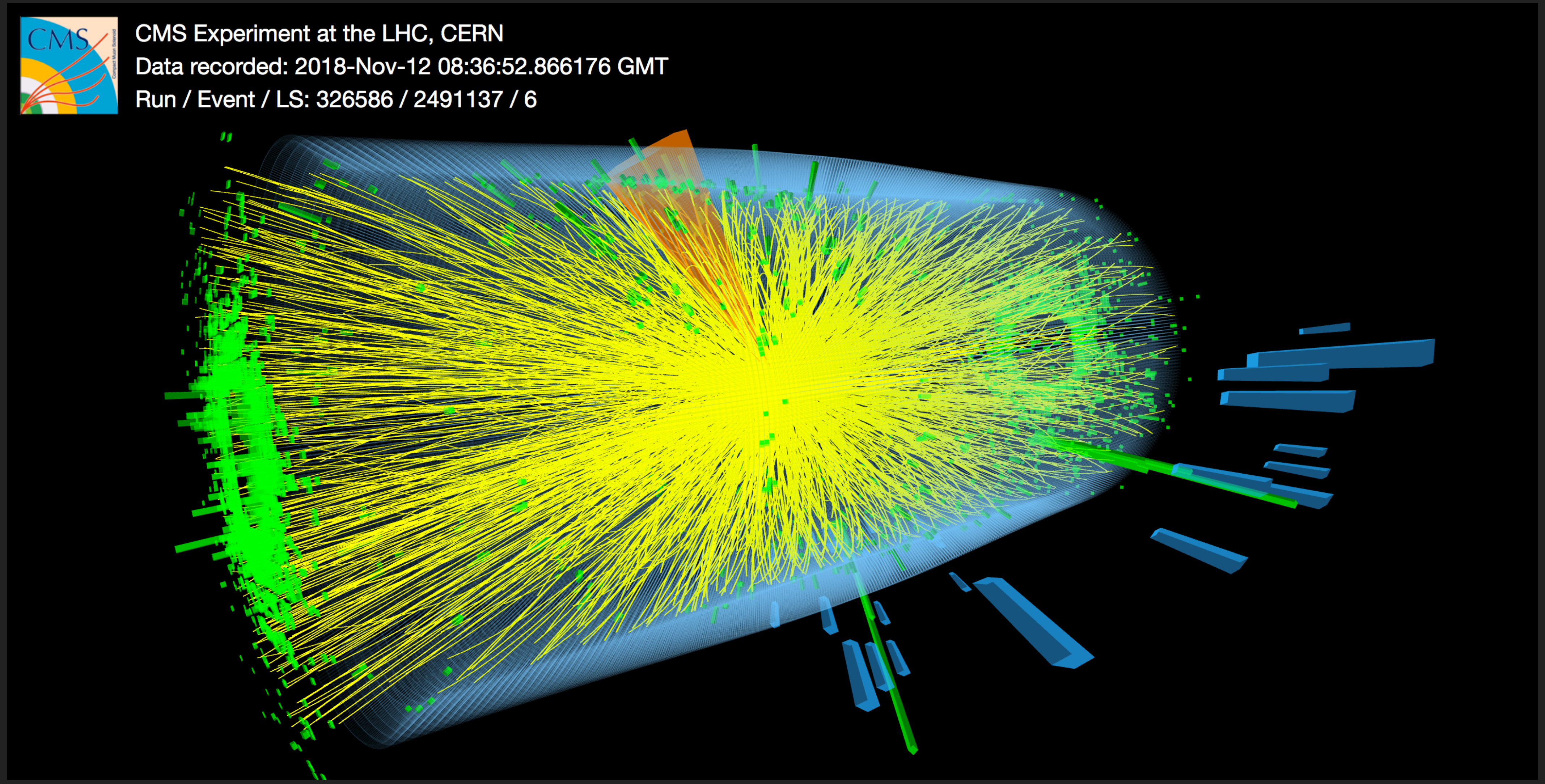


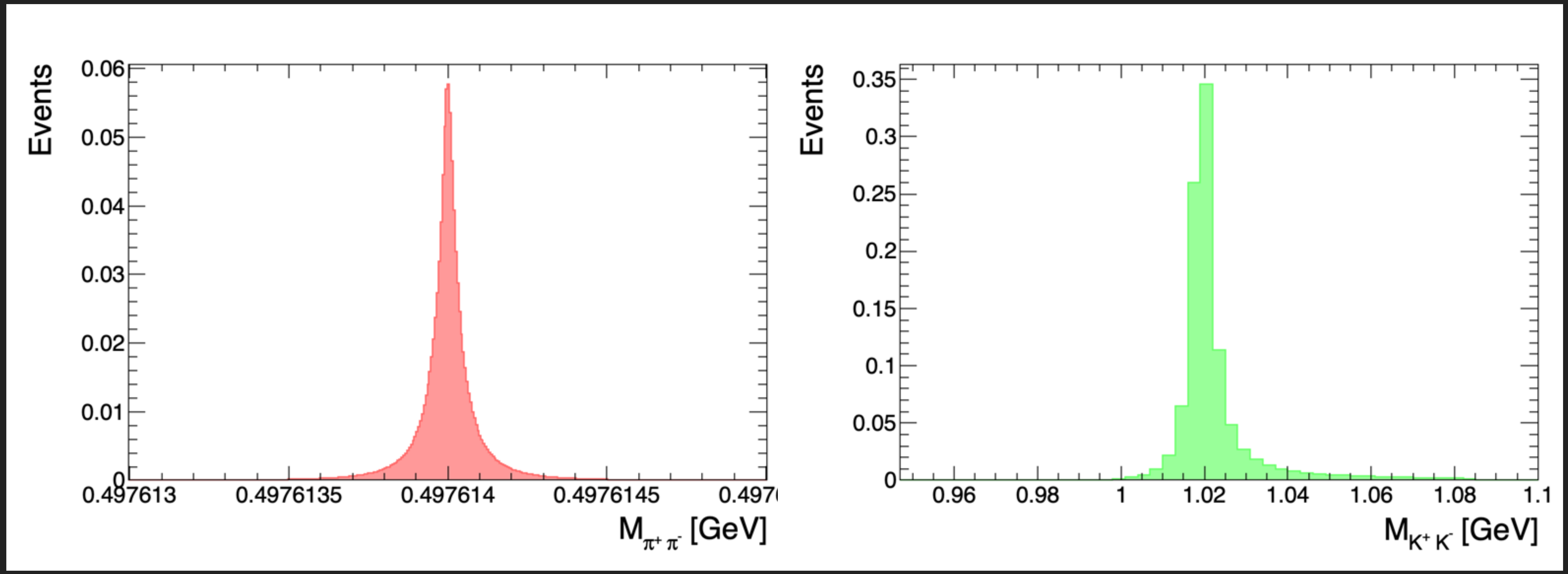
CMS Tracker Structure

MEASUREMENT OF MOMENTUM RESOLUTION USING LOW-MASS RESONANCES

- Run-2 environment
  - High Pile-Up & $\sqrt{s}$

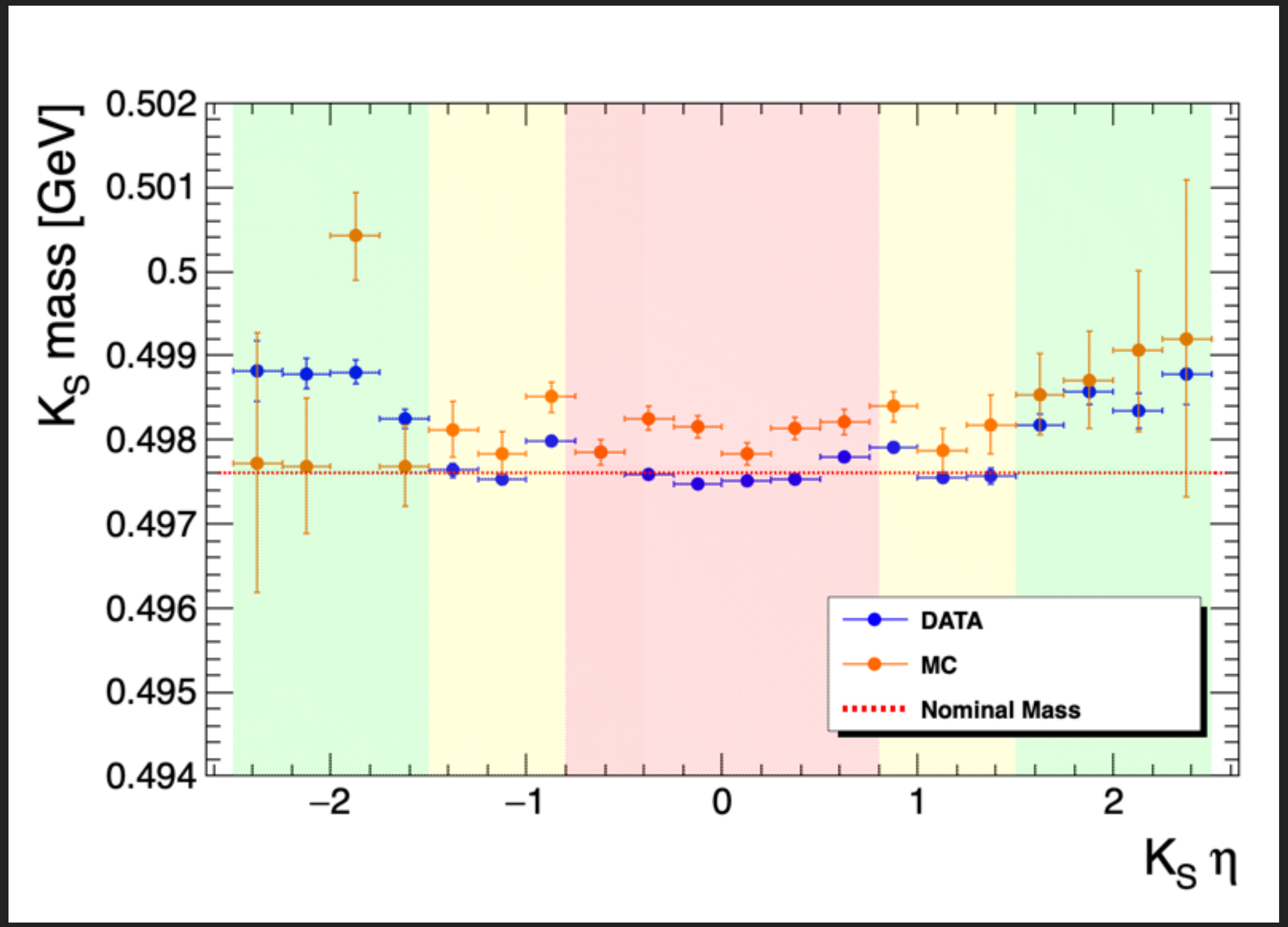
$K_S \to \pi^+\pi^-$ $\phi \to K^+K^-$



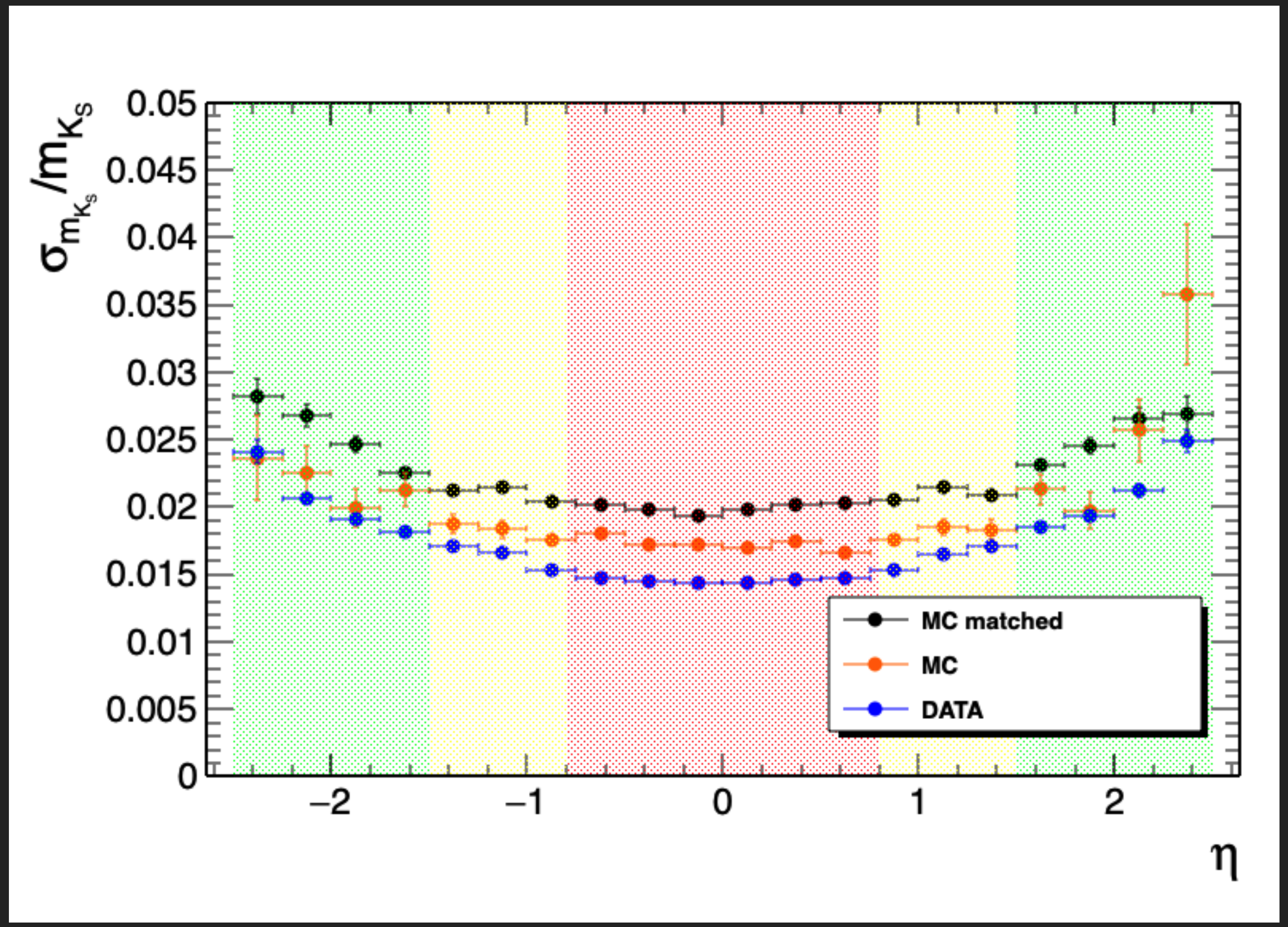


MEASUREMENT OF MOMENTUM RESOLUTION USING LOW-MASS RESONANCES

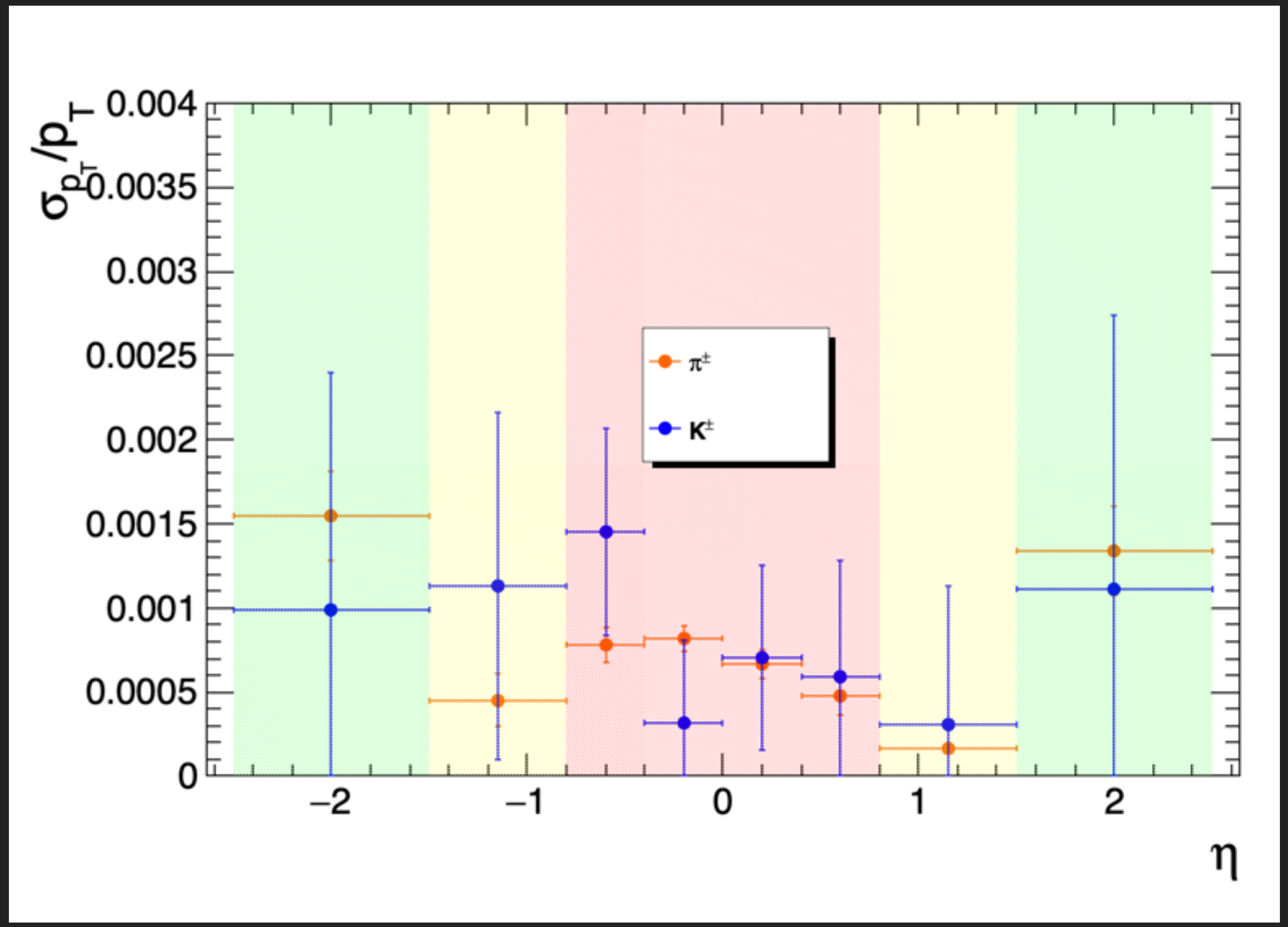
Invariant Mass Reconstruction



Invariant Mass Resolution



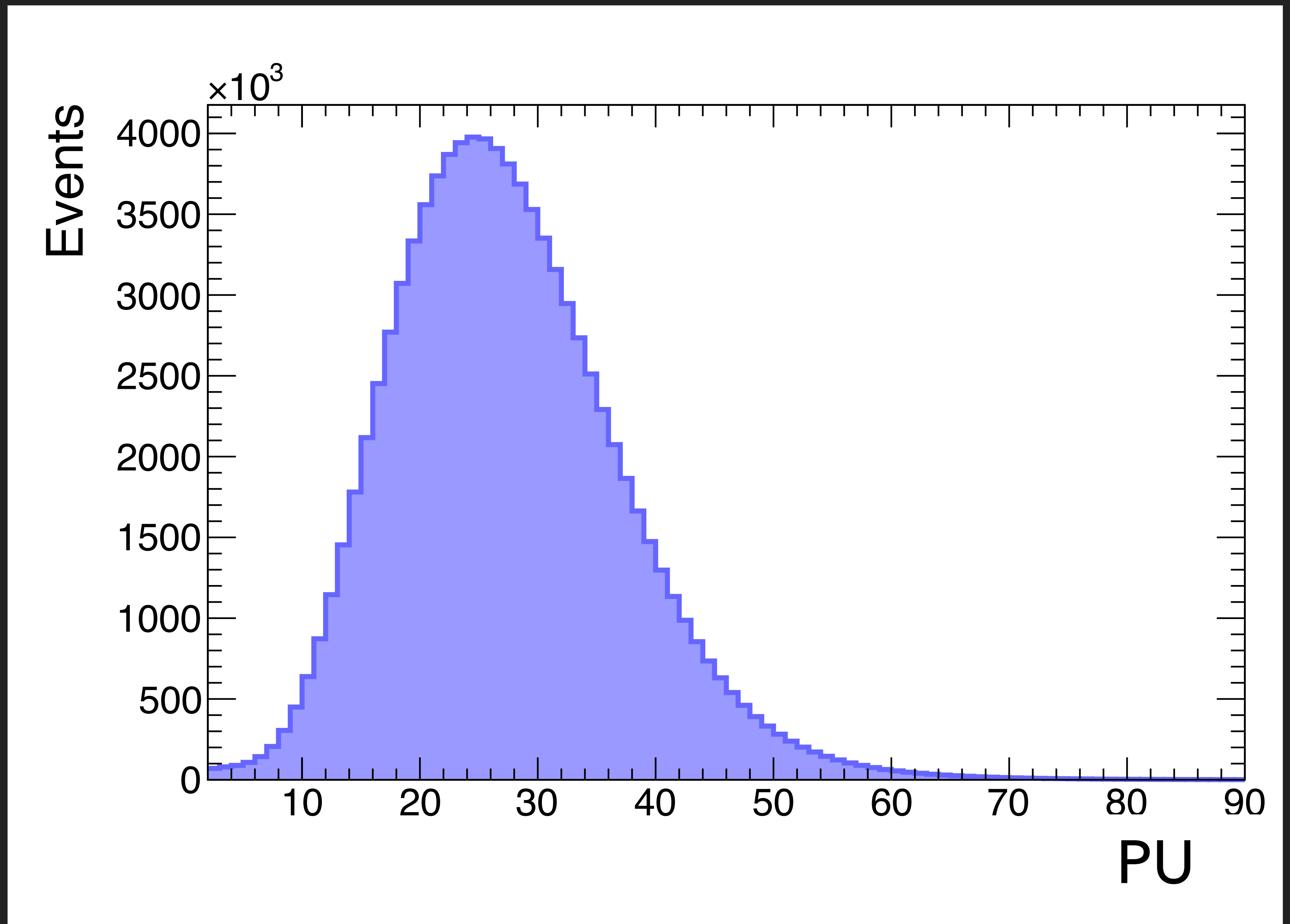
Transverse Momentum Resolution



Evaluate dependences from cinematic variables (p_T , η , ϕ)

MEASUREMENT OF MOMENTUM RESOLUTION USING LOW-MASS RESONANCES

- Future studies:
  - Pile-Up dependence
  - Higher energy decays $J/\psi \to \mu^+\mu^-$
  - Displaced decays (b-decays)



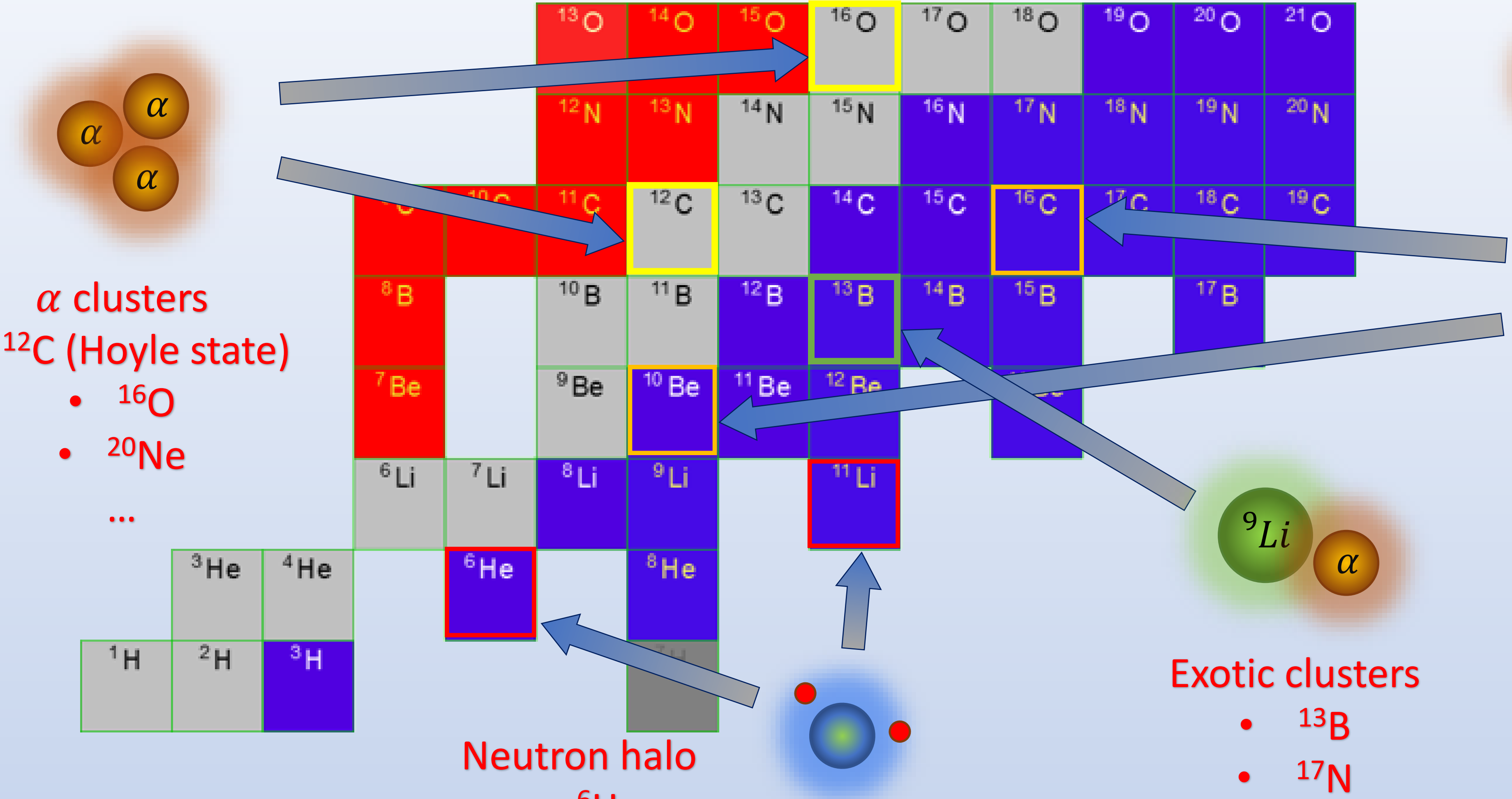
Pile-Up distribution

ALESSANDRO PASCOLINI

- Master degree in Pysics at Università degli studi di Perugia
  - "Measurement of Momentum Resolution using Low-mass Resonances"
- INFN - CNAF
  - User-supporter (mainly CMS)
  - Administrator of HPC
  - Researcher in HPC applications performance evaluation and optimization

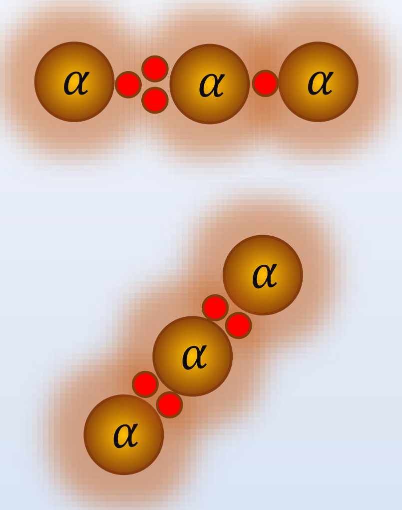
Study of cluster structures in light nuclei

- **CLIR** (**C**luster in **L**ight **I**on **R**eactions);



α clusters

- ^{12}C (Hoyle state)
  - ^{16}O
  - ^{20}Ne
  - …



Cluster chains

- ^{16}C
- ^{10}Be
- ^{22}O
- …

Exotic clusters

- ^{13}B
- ^{17}N
- …

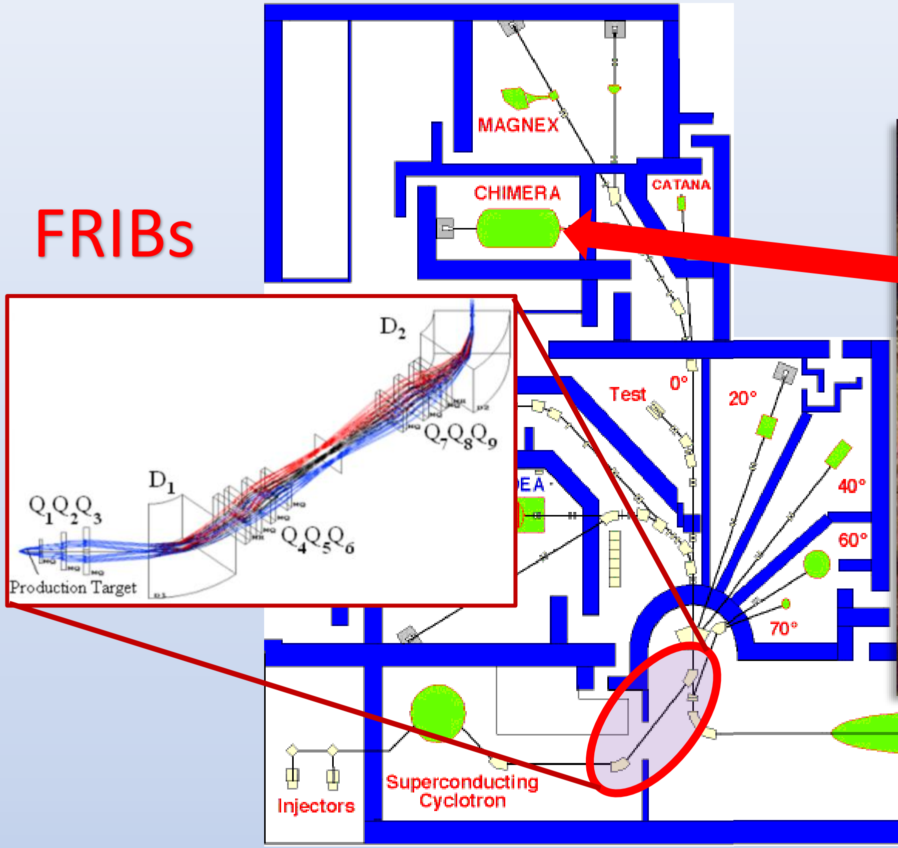
Neutron halo

- ^{6}He
- ^{11}Li

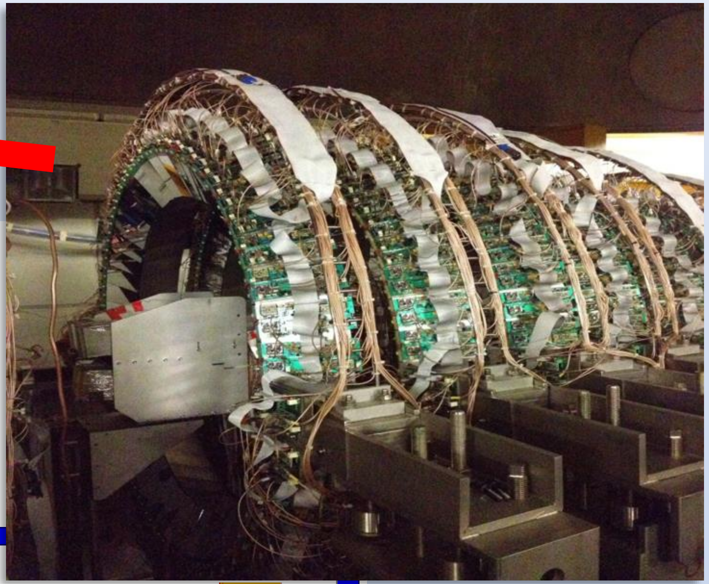
LNS: Facility FRIBs & Experimental apparatus

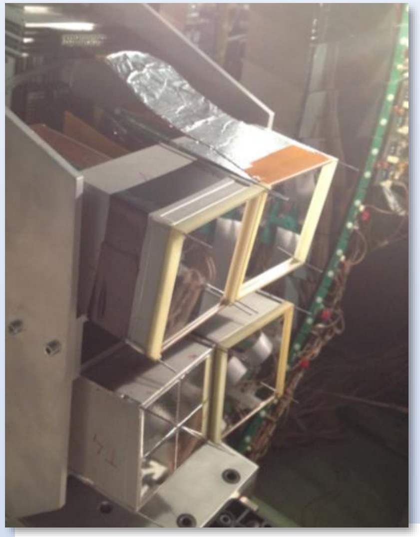
- Cocktail beam produced with In-Flight fragmentation technique:
  - Primary beam $^{18}O^{7+}$ (55 MeV/u), many neutron rich isotopes produced;
- CHIMERA (1192 units) + 4 FARCOS + Tagging system for identification and TOF.

FRIBs

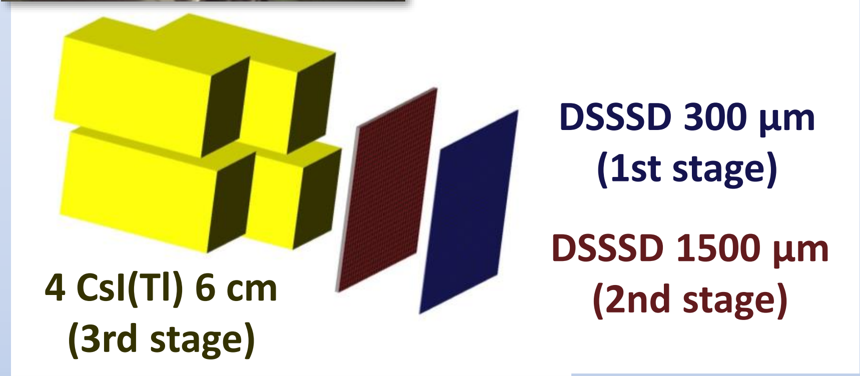


CHIMERA 4π MULTIDETECTOR



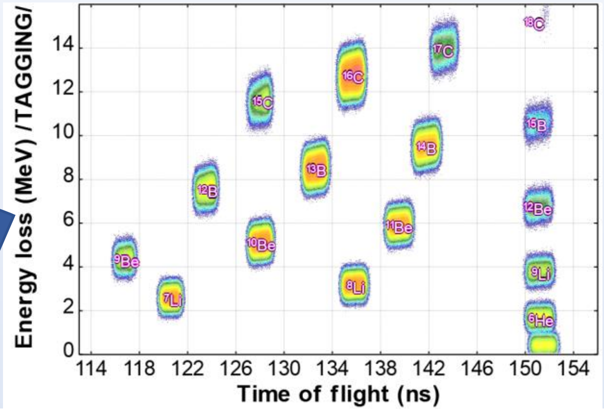


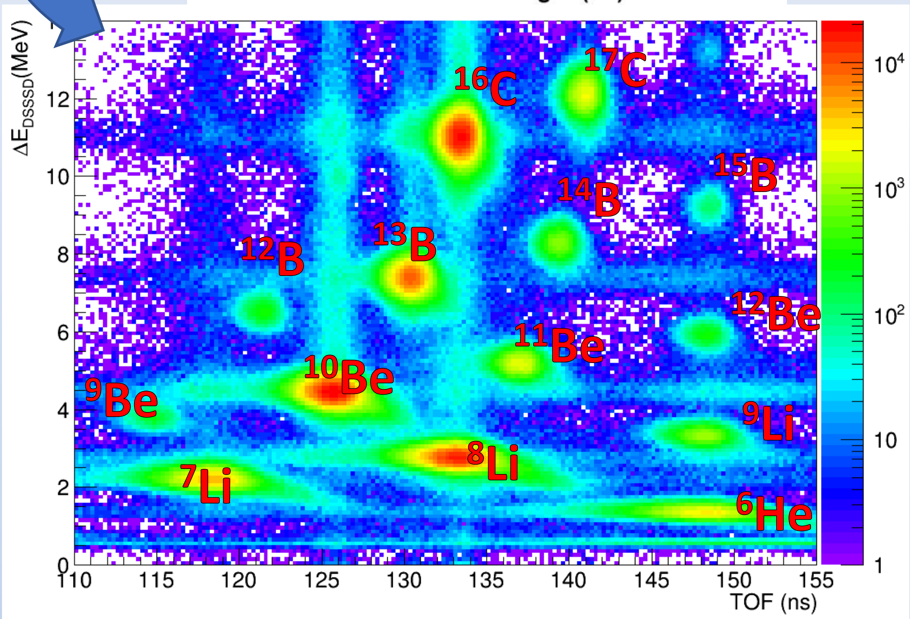
FARCOS
- 4 telescopes
- 132 total ch/tel
- Works like a magnifying glass



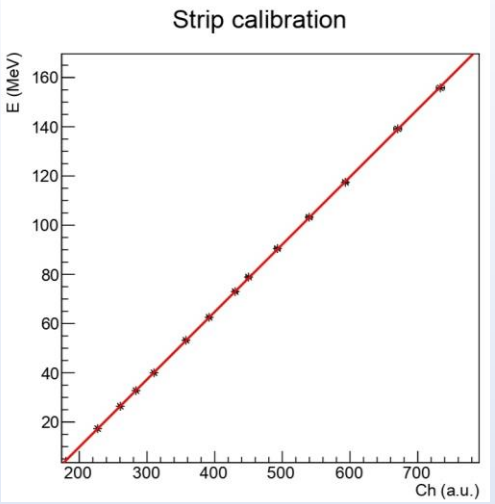
Some results…

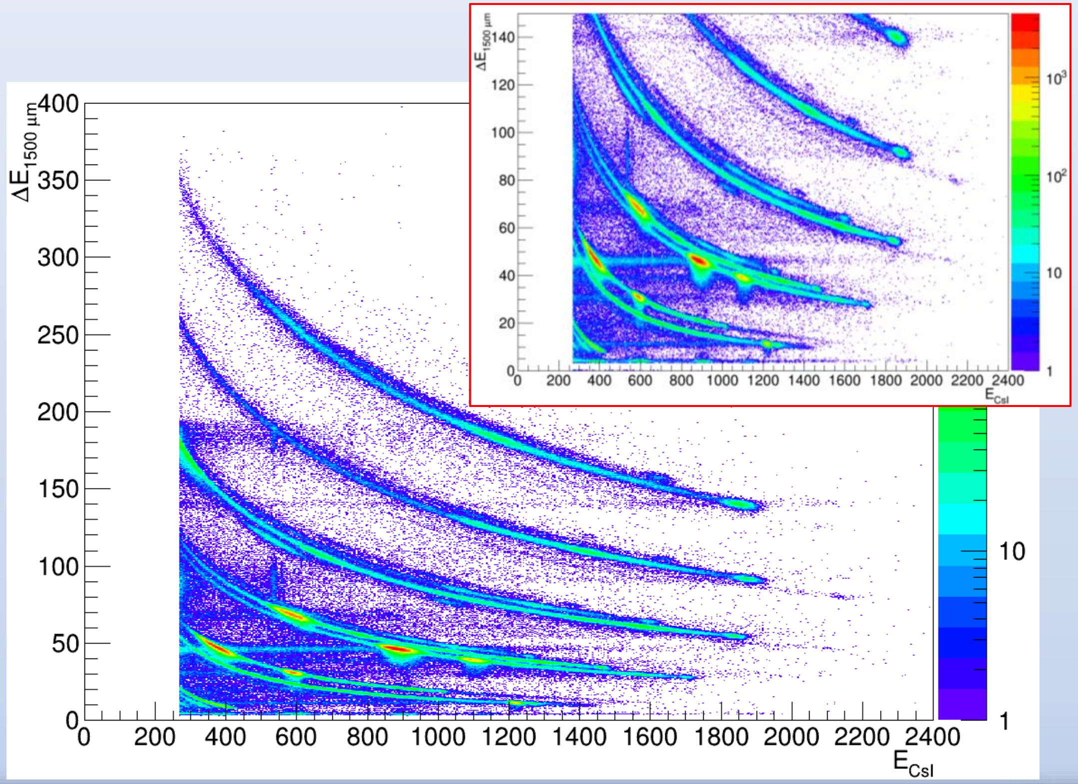
- Tagging calibration by simulations on LISE++

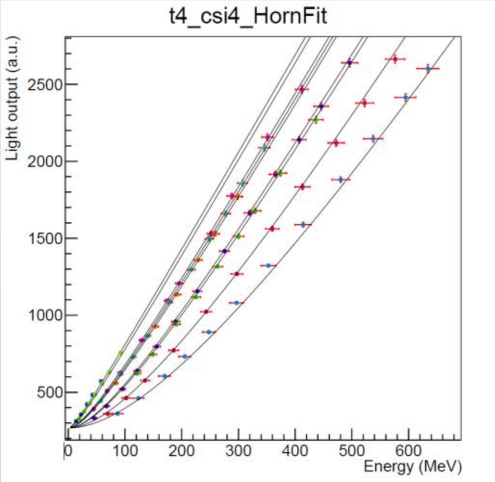




- Simulation of energy loss (KaliVeda framework based on ROOT)
- Energy calibration (MINUIT minimization) of FARCOS stages







Data analisys

- ROOT (C++)

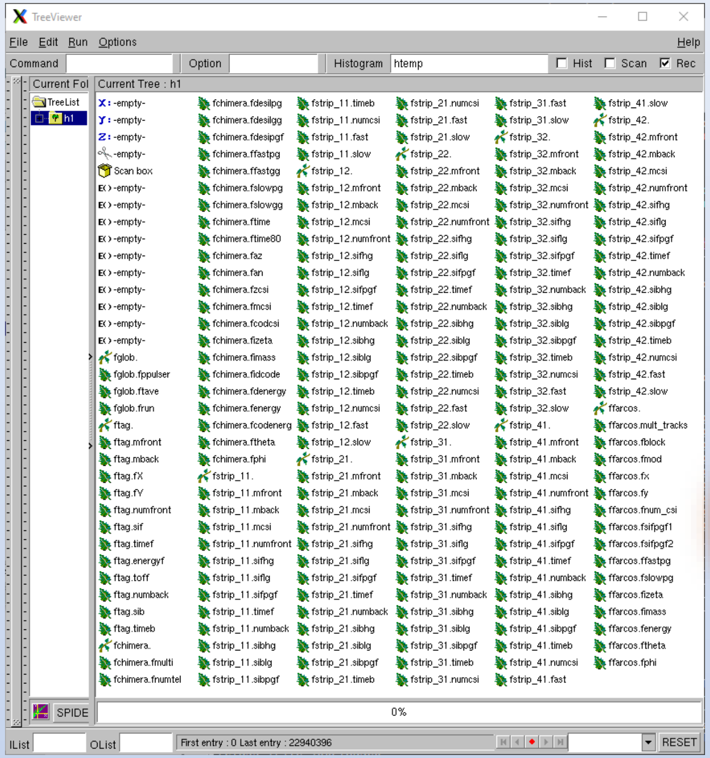
ROOT Data Analysis Framework

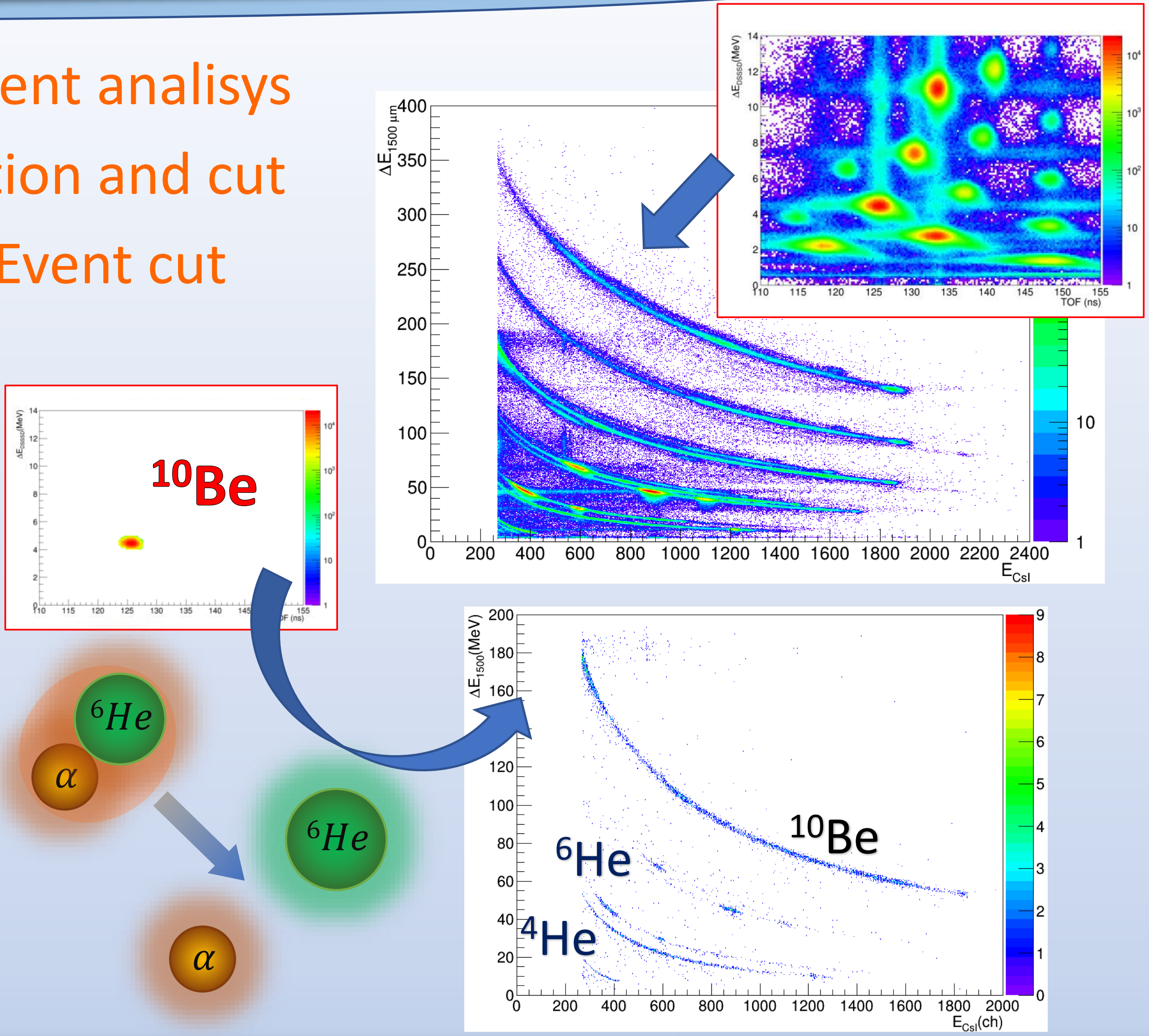
- Online
- Offline

analisys

- Custom libraries
- Data structure in TTree

- Event-by-event analisys
- Event selection and cut
- Graphical / Event cut

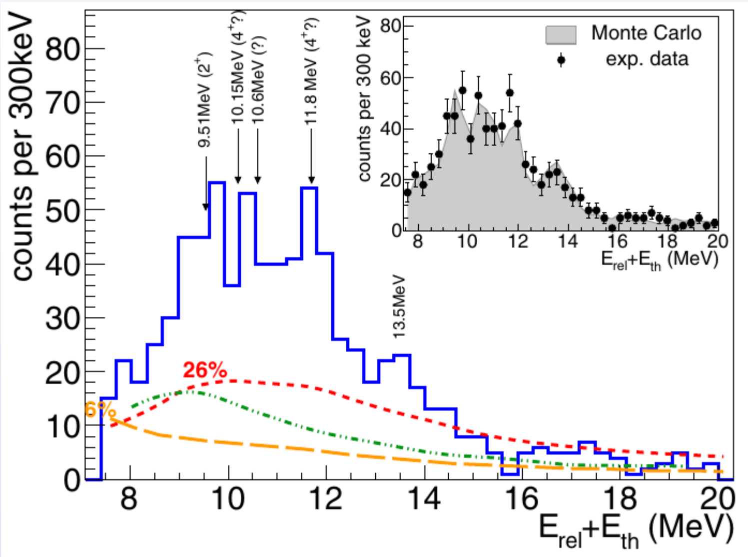


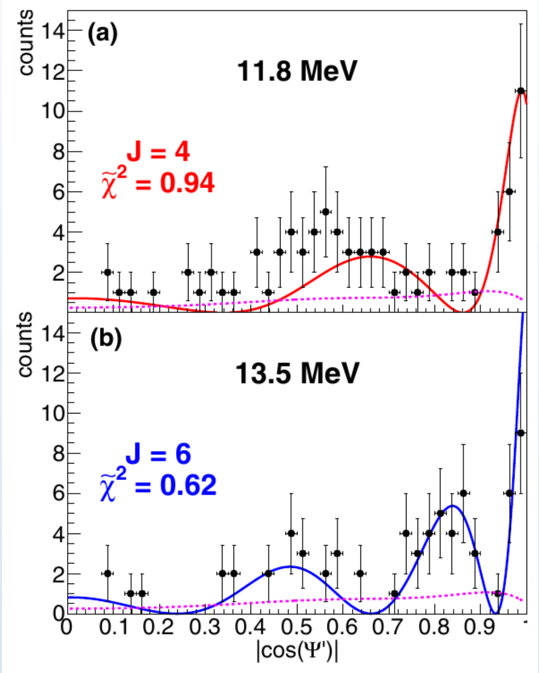


Future perspectives and interests

- Monte Carlo simulation of theoretical efficiency of cluster state production
- Main issues until now:
  - Poor memory management
  - Quite low processing speed
- Main expectation from the School

Past accomplishments [1]





[1] Dell'Aquila D. et al., Phys. Rev. C, 93 (2016) 024611.

- Learning and sharing of smart techniques to improve the speed and optimization of calculations
- New programmi tools and methodologies
- Multi-parallel programming (CPU, GPU?)

NVIDIA CUDA

THANK YOU

Università
degli Studi
di Padova

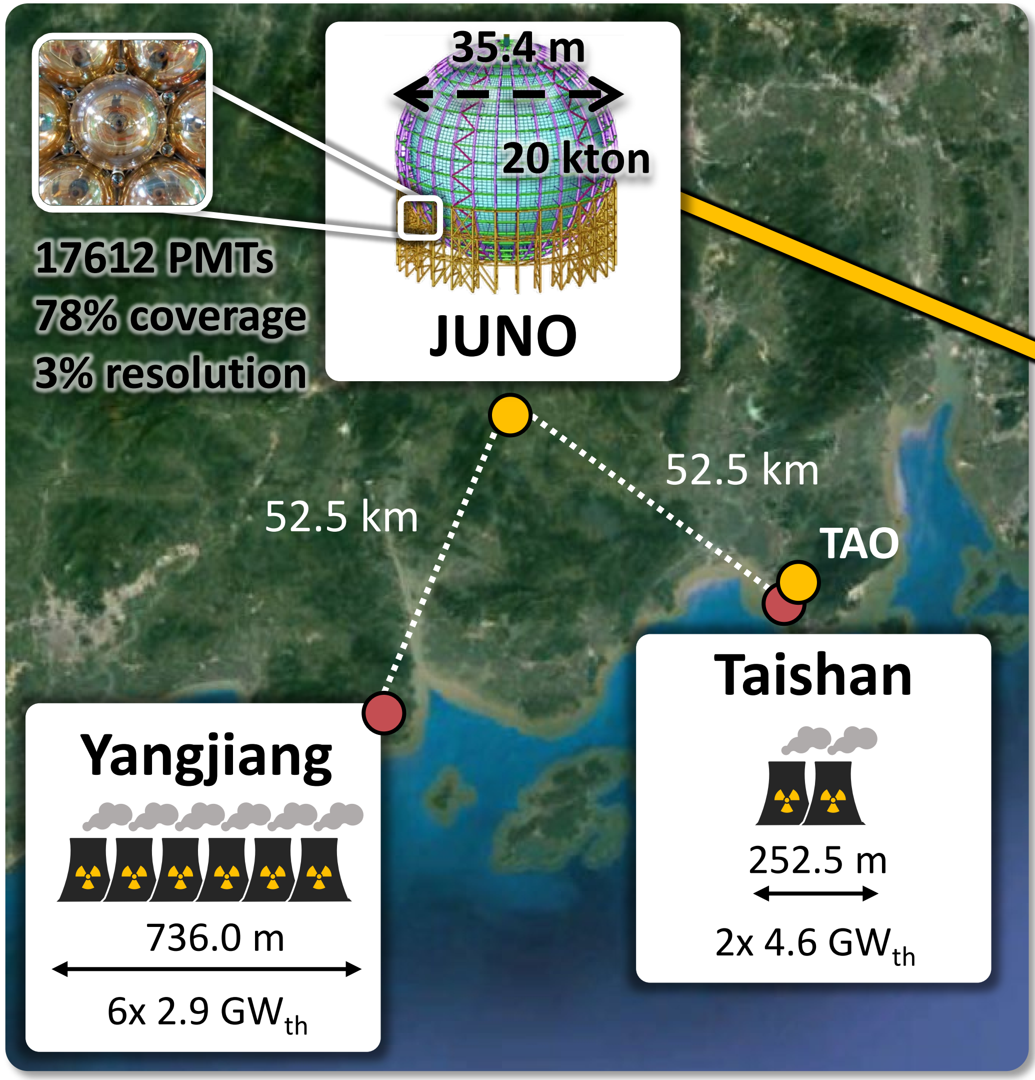
My research activity in JUNO

and interests related to the field of scientific computing ☺

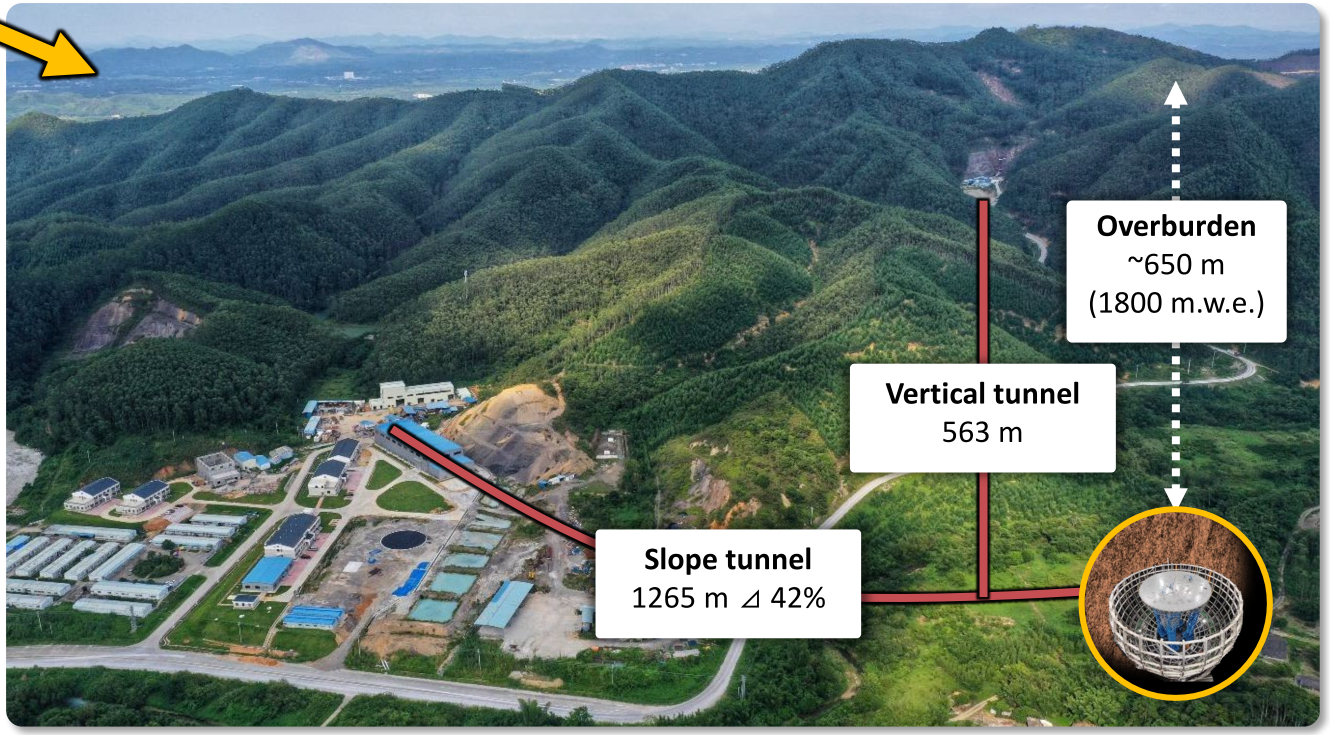
October 2022

Andrea Serafini

The Jiangmen Underground Neutrino Observatory



JUNO is a **20 kton** multi-purpose underground **liquid scintillator** detector currently under construction at a baseline of about **52.5 km** from eight **nuclear reactors** in the Guangdong Province of South China.



Oscillation physics with JUNO

JUNO aims at determining the neutrino mass ordering @ >3σ in 6 years, and will be the first experiment to:

- **simultaneously** observe **fast** and **slow oscillations** (Δm^2_{21}, θ_{12}, Δm^2_{31} and θ_{13})
- place **<1% precision** on Δm^2_{21}, θ_{12}, Δm^2_{31}

$$\Delta_{ij} = \frac{\Delta m^2_{ij} L}{4E}$$

$$\boldsymbol{P_{\bar{e}\bar{e}} = 1 - P_{21} - P_{31} - P_{32}}$$

SLOW

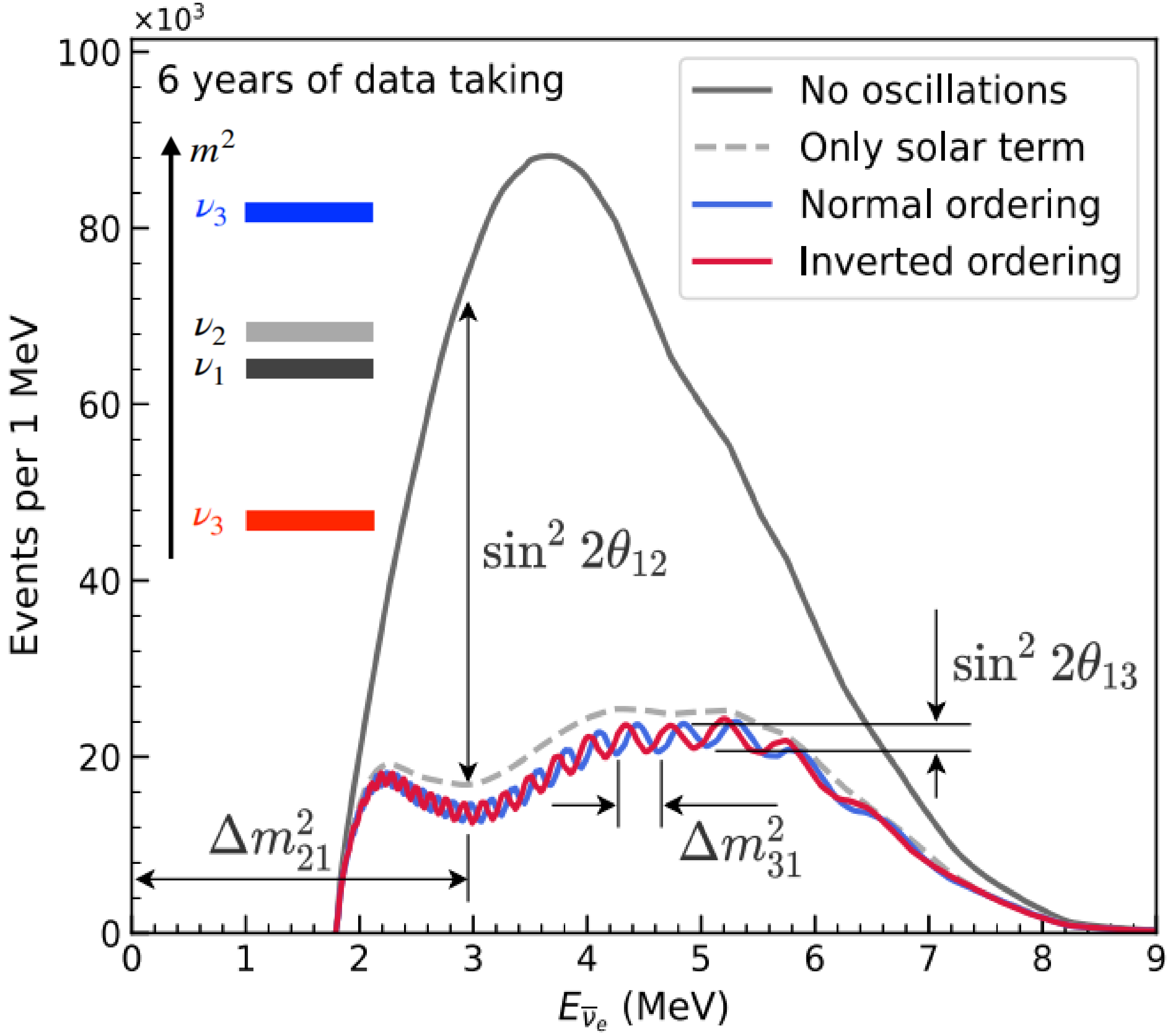
$$\boxed{P_{21} = \cos^4\theta_{13}\sin^2 2\theta_{12}\sin^2\Delta_{21}}$$

FAST

$$P_{31} = \cos^2\theta_{12}\sin^2 2\theta_{13}\sin^2\Delta_{31}$$

$$P_{32} = \sin^2\theta_{12}\sin^2 2\theta_{13}\sin^2\Delta_{32}$$

→ **probability does not depend on $\boldsymbol{\delta_{CP}}$ and $\boldsymbol{\theta_{23}}$**

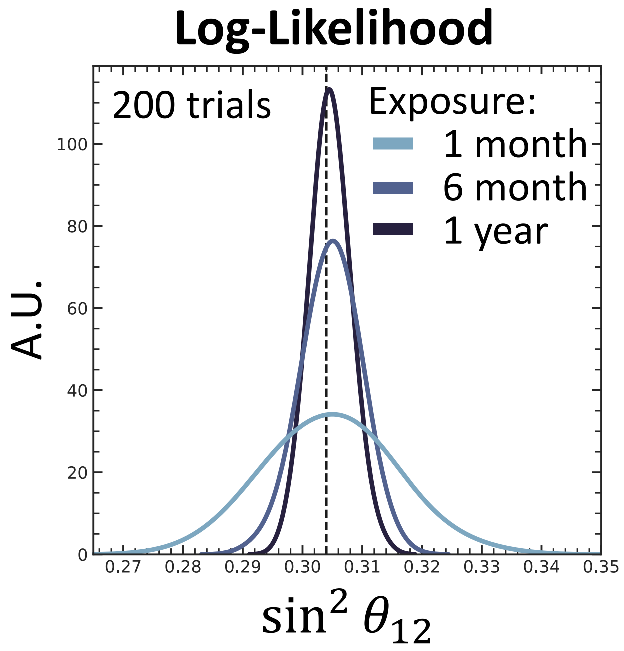


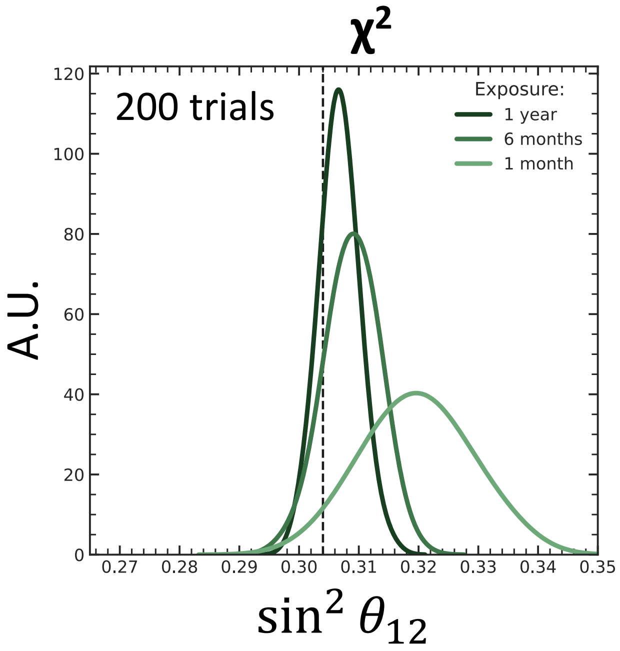
Possible biases with realistic datasets

Statistics-dependent biases

- In the first data taking period (<1 y), the choice of the descriptive statistics will have an impact in the estimation of the physics parameters

→ standard **χ^2 is statistically biased**

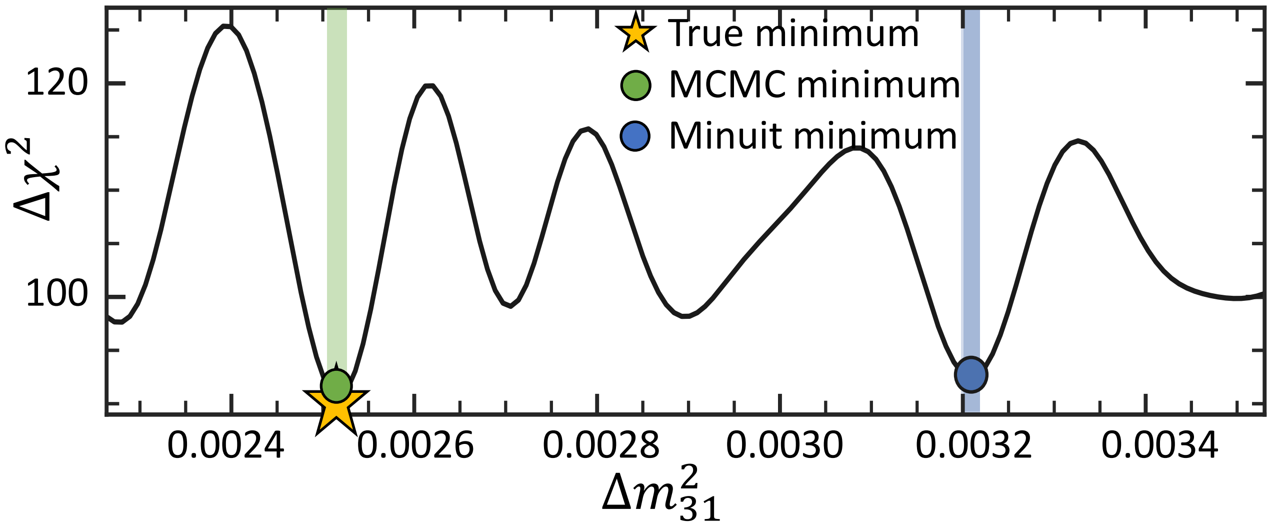




Minimizers limitations

- Local minimizers (e.g. Minuit) are not designed and optimized for high-dimension and complex parameter spaces. Bayesian samplers (e.g. MCMC) are instead not prone to local-minima biases.

→ **Minuit**-like minimizers **can be biased** in JUNO

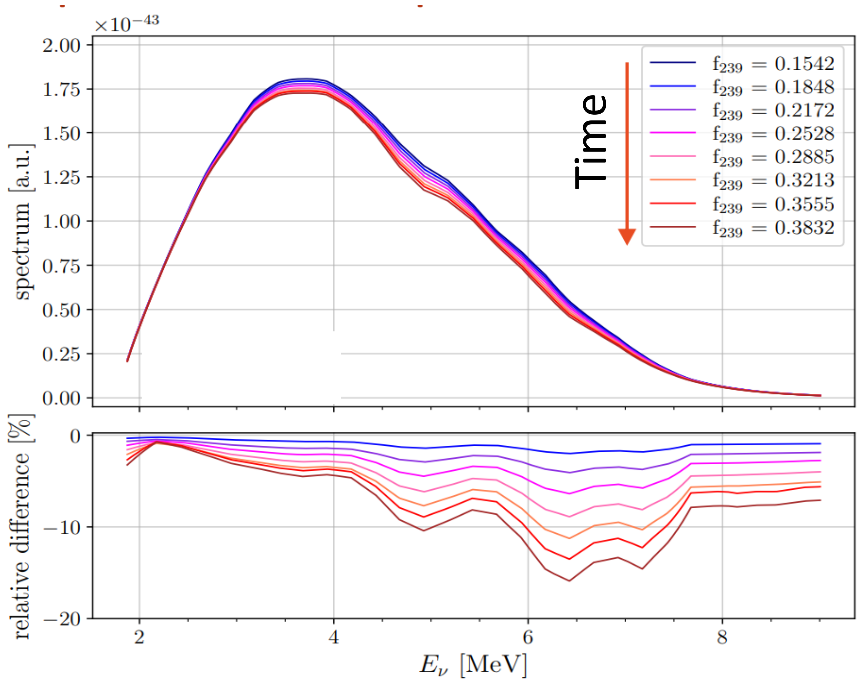


Possible information loss in the fit due to binning

Temporal information

The reactor antineutrino spectrum changes with time as fuels evolve.

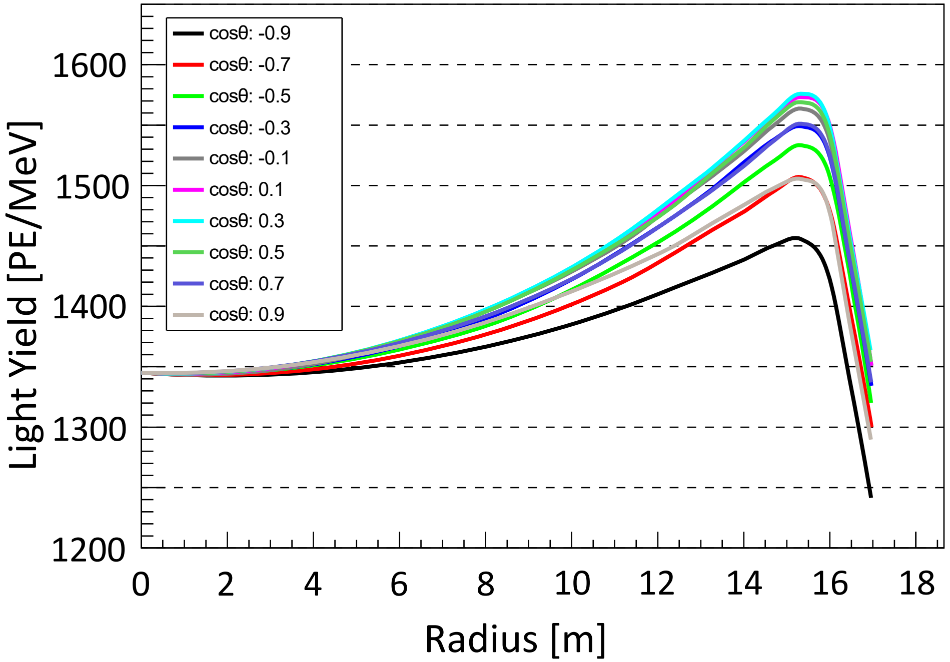
Binned spectra contain events and quantities averaged over a time period.



Spatial information

Detector non-uniformity, light yield, resolution have spatial/radial dependence

Binned spectra contain events and quantities averaged over the entire detector.



My goal: treat space and time as continuous variables → **unbinned likelihood!**

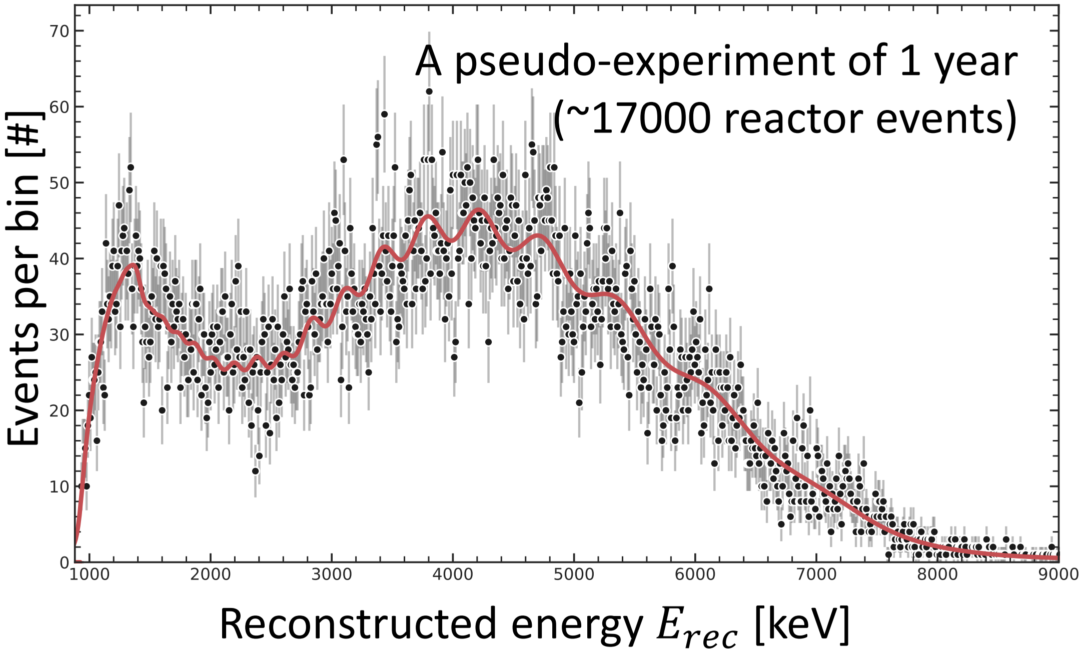
Development of a fitting framework

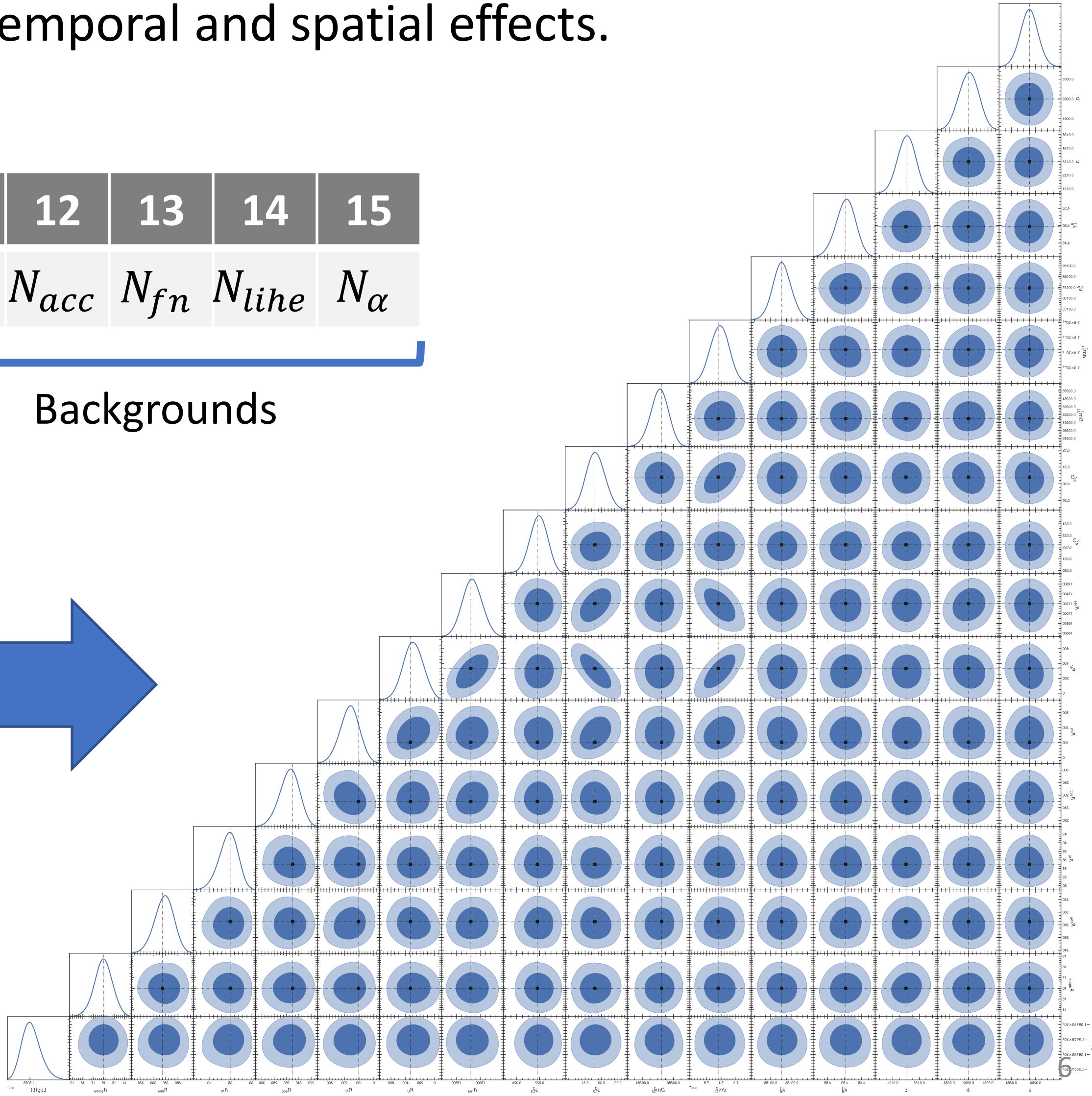
I am developing a MCMC-based fitter for estimating physics and nuisance parameters and posteriors based on an unbinned extended likelihood incorporating temporal and spatial effects.

Model parameters:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
a	b	c	k_B^1	k_B^2	$\boldsymbol{\Delta m^2_{21}}$	$\boldsymbol{\Delta m^2_{31}}$	$\boldsymbol{\theta_{12}}$	$\boldsymbol{\theta_{13}}$	$\boldsymbol{N_{rea}}$	$\boldsymbol{N_U}$	$\boldsymbol{N_{Th}}$	N_{acc}	N_{fn}	N_{lihe}	N_α

Resolution Non-linearity Osc. parameters Backgrounds





Research activities

Eugenia Spedicato[1]
University of Bologna,
[1]espedica@bo.infn.it

The MUonE experiment

Topic of research

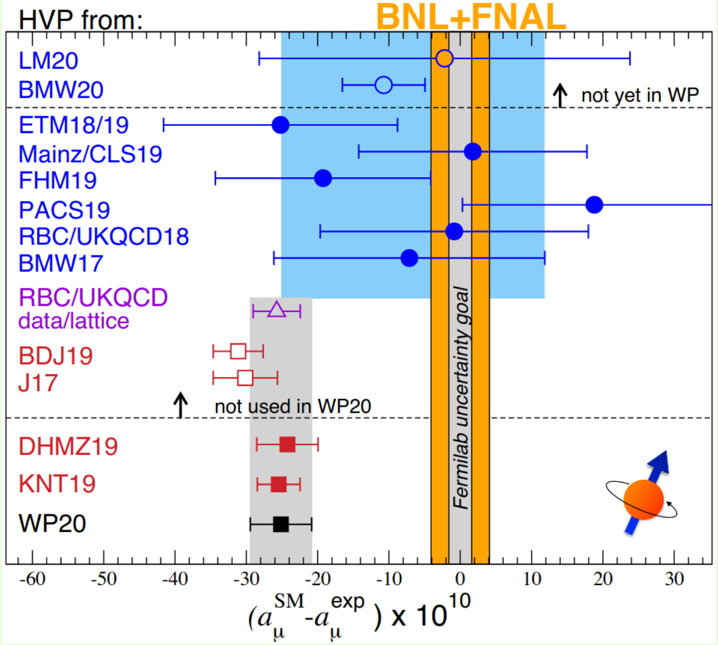
Anomalous magnetic moment of the muon a_μ and the discrepancy between its theoretical estimate and experimental measurement $\Delta a_\mu(th - exp)$

Aim of the experiment

Give a new independent estimate of the leading order hadronic vacuum polarization term a_μ^{HVP} which dominates a_μ theoretical uncertainty

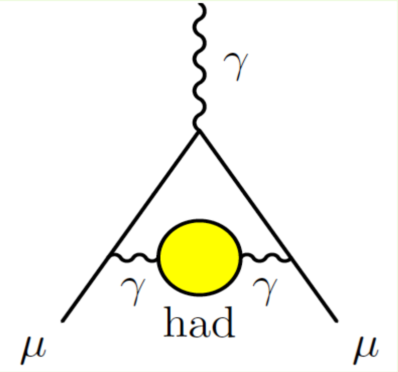
State of the art

It's a new proposal, every aspect is under development, from MC simulations, to detector components production and test.
Computing skills are fundamental to make up the building blocks of the whole infrastructure.



$$a_\mu^{SM} = a_\mu^{QED} + a_\mu^{EWK} + a_\mu^{had}$$

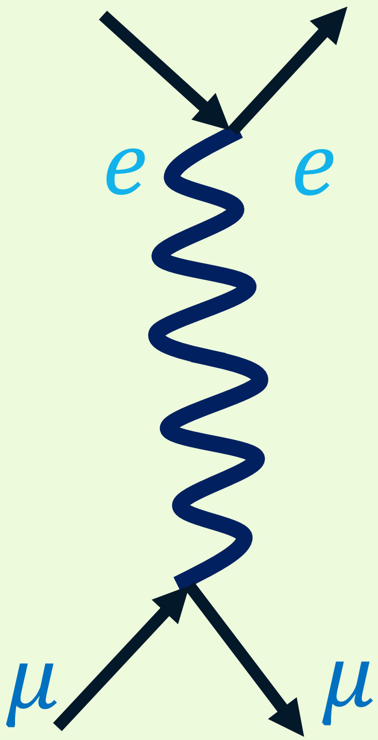
$$\boxed{a_\mu^{HVP} + higher\ terms}$$



The MUonE experiment

Main research activity

- ▷ Implementation of the **calorimeter response** in the *fast* simulation code;
- ▷ **Decoding and analysis** of the *raw data* collected during Test Beams;
- ▷ Check the **efficiency of the reconstruction algorithm** for $\mu - e$ elastic events (our signal) in MC and real data. It is implemented in a FairRoot environment based on **C++;**
- ▷ Collaborate for the implementation of a **CLOUD infrastructure** for the experiment**.**



I'm interested in learning how to write in a **more efficient way C++ code**, be more conscious on what's *behind* my coding decisions

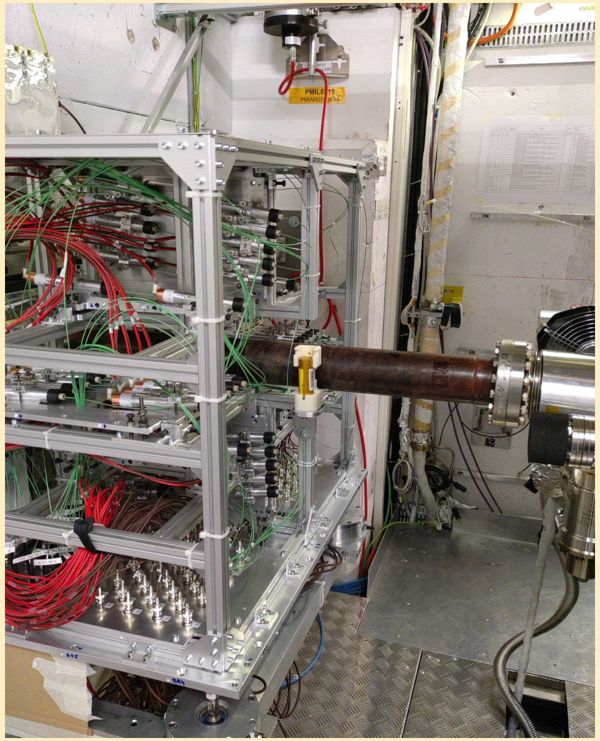
The LHCb experiment: PLUME project

Aim of the project

Provide a new detector for measuring the istantaneus luminosity (online and offline) and visible interactions per bunch at LHCb, which is fundamental for cross section measurements.

State of the art

It’s a new detector, it has been installed upstream the LHCb collision area and it’s now working during the Run 3 data taking. It is an on-going project, computing skills are needed to implement the decoding and the High Level Trigger line which at LHCb is done on GPUs, thus CUDA is required.



The LHCb experiment: PLUME project

Main research activity

Implement the **High Level Trigger 1** line for PLUME in the LHCb software called **Allen**

It runs on GPU-s thus **CUDA knowledge is needed**

MAIN TASK: decoding raw data

Clear interest in **learning** that language!

Bernardino Spissso
INFN Napoli

Summary of the main activities carried out

- Activities within the ATLAS experiment computing group of the Naples INFN
  - Administrator of the ATLAS resources in Naples
  - Participation in the DIOS project of the ESCAPE program
- Activities within the KM3Net experiment
  - Classification and regression via Deep Learning on GPU
  - Job containerization and submission via DIRAC of simulation atmospheric showers (ConCORDIA ESCAPE)
- Activities within the Hyper-KamioKande experiment
  - Background noise reduction via Machine Learning (BDT)
  - Participation in the deployment of the experiment DB

Activities within the ATLAS experiment

Administration and implementation of new resources

- For the management, configuration and production of the computing resources I have entirely created an automatic provisioning system from bare metal via PXE based on the software tools Kickstart, Puppet and Foreman, which is already currently in use for the machines:
  - of ATLAS experiment
  - of BELLE II e Km3NeT experiment
  - of the new INFN resources of the IBiSCo project
- I am also responsible of the storage systems (SANs) that currently we are migrating from DPM storage manager to dCache.

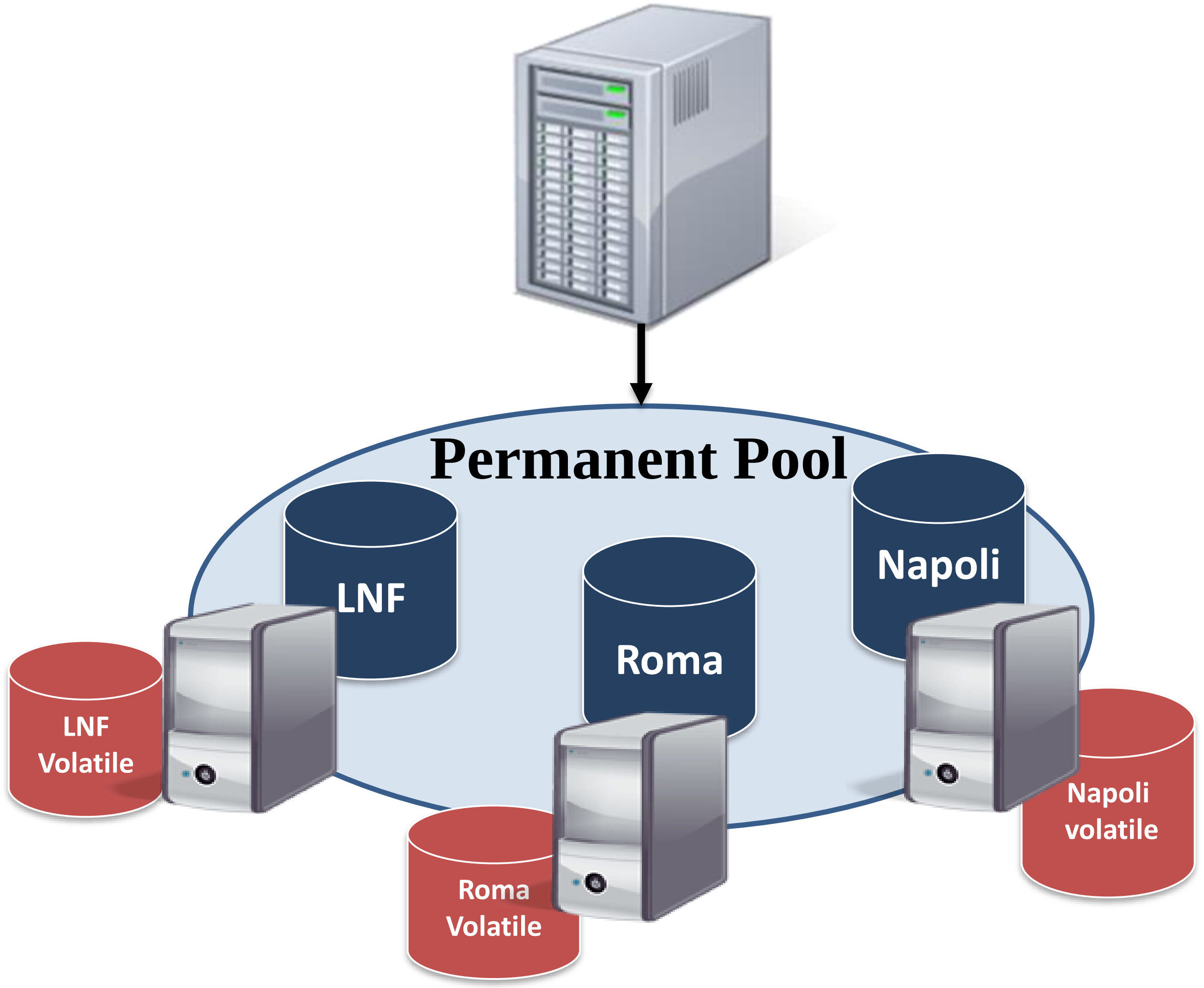
Participation in the DIOS project of the H2020 ESCAPE program

DIOS (Data Infrastructure for Open Science) ESCAPE consists of a European experimentation for the creation of a geographically distributed storage system on the model of data lakes.

- **Geographically distributed storage resources in ATLAS Tier2:**
  - **Napoli**
  - **Roma1**
  - **LNF**

.

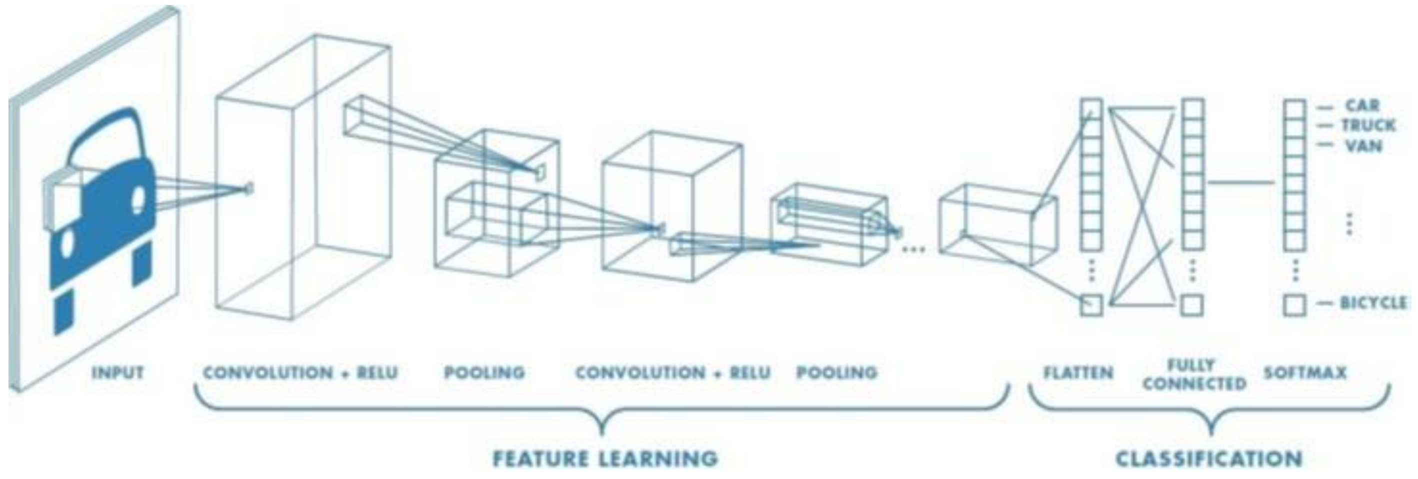
DPM head node Napoli

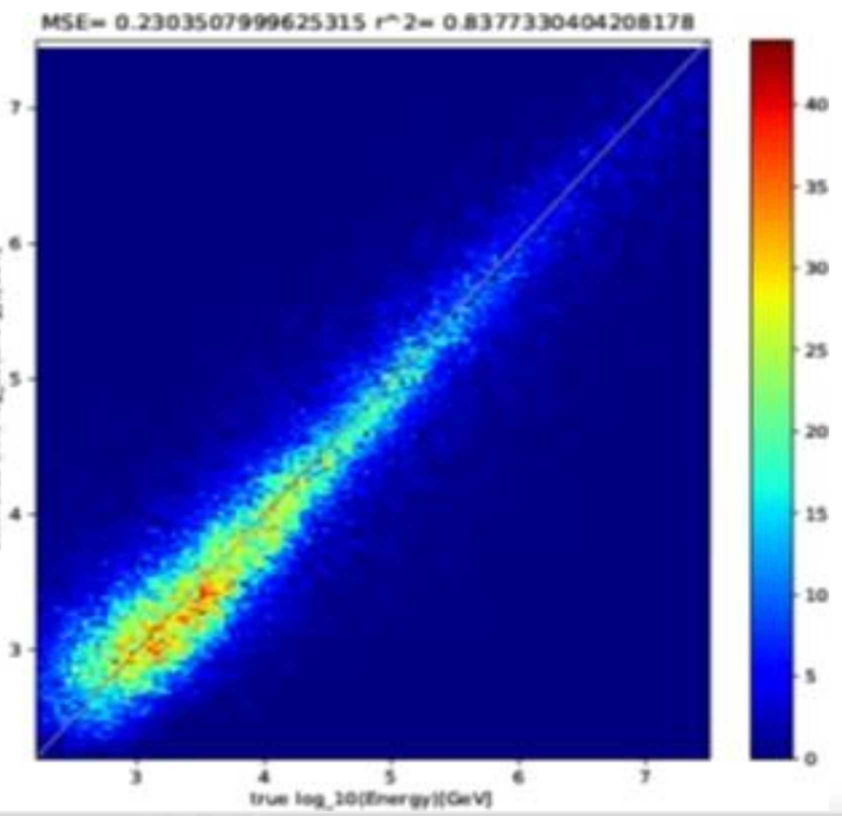


Activities within KM3NeT experiment

Classification and data regression via ML on GPU

- Classifications : signal/noise, particle identification
- Regressions: energy (graph on the right), incoming particle direction, interaction vertex
- Networks used: CNN, ResNet and GNN





IT resource admin

- Administrator of storage resources in the Naples Data Center.
- Experiment database mirror

OrcaNet + TensorFlow + Keras + NVIDIA CUDA

ConCORDIA

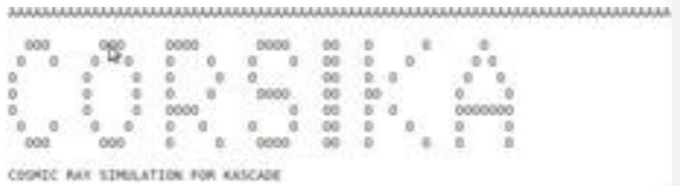
- Sub-project of the ESCAPE program which consists of a framework for the submission of CORSIKA simulations via DIRAC using container technology.

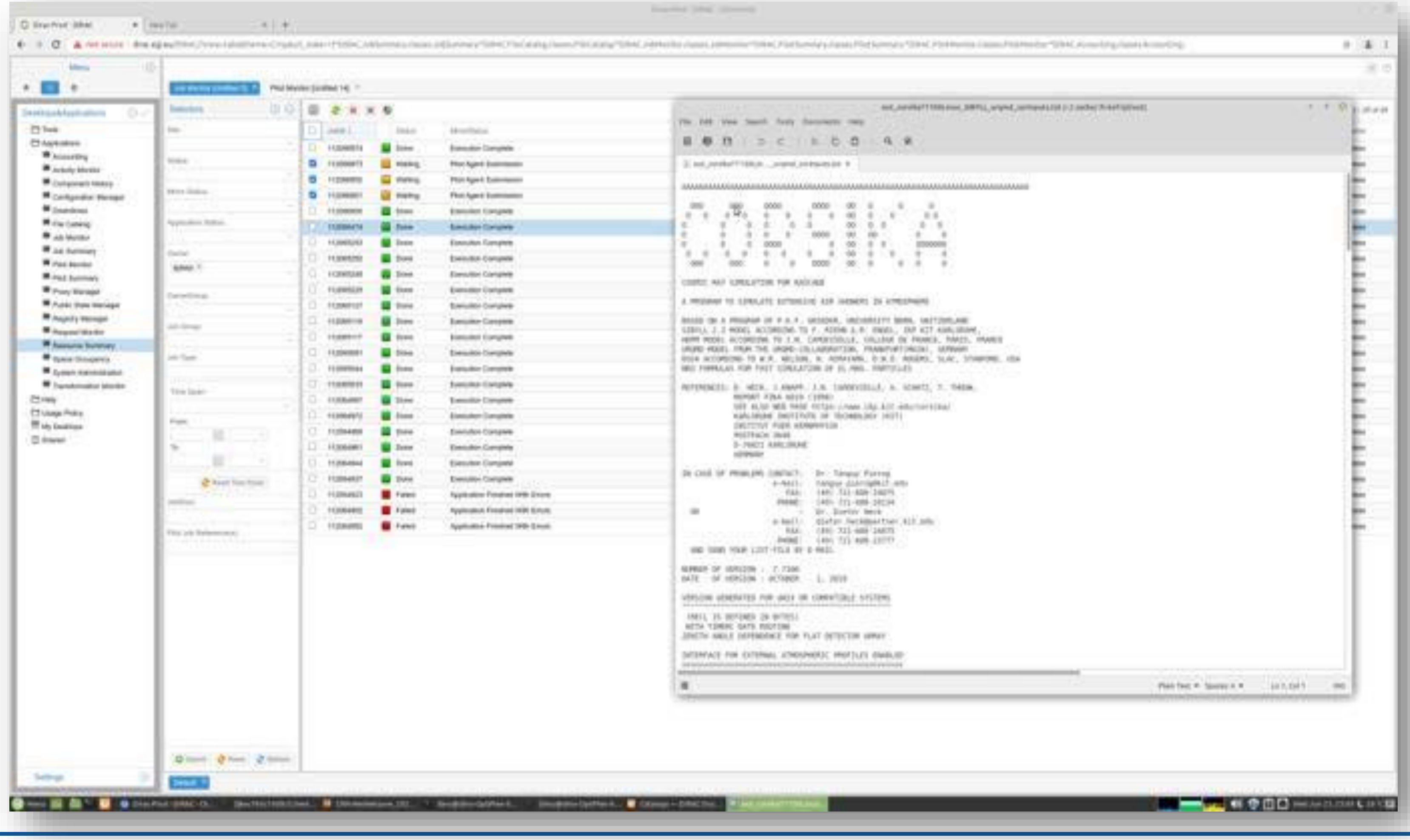
General architecture:

- Containerized CORSIKA (Singularity)
- Registration of VO km3net.org on EGI-Dirac
- Containerized Dirac CLI client (Docker)
- EGI-Dirac web portal https://dirac.egi.eu
- X509-based authentication (web and CLI client)

DIRAC THE INTERWARE

docker

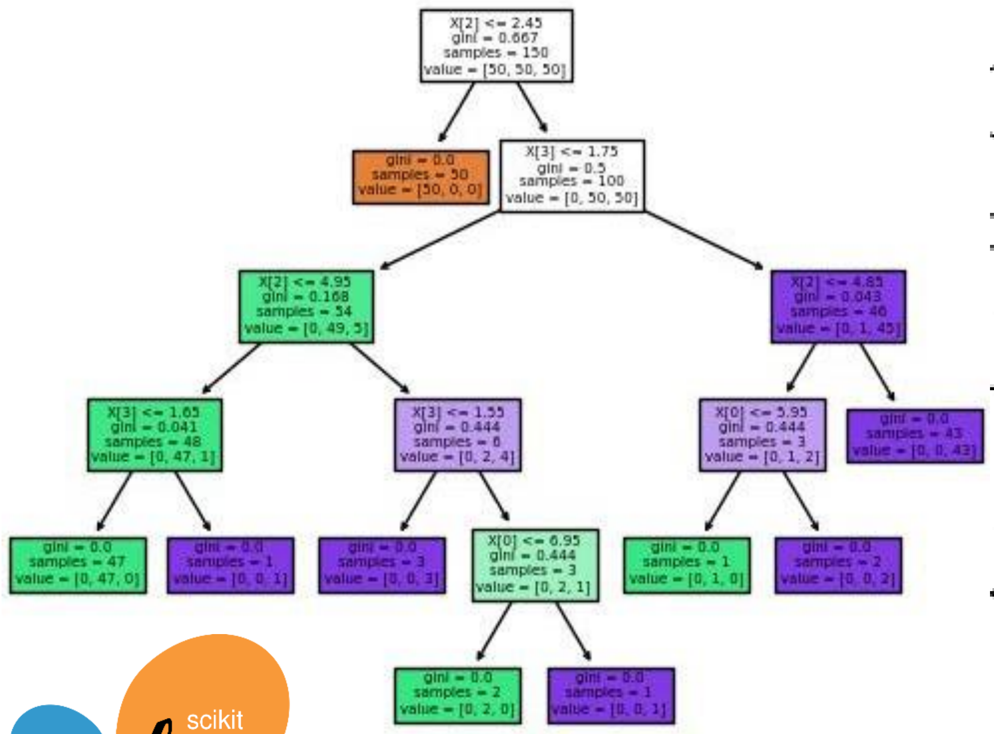


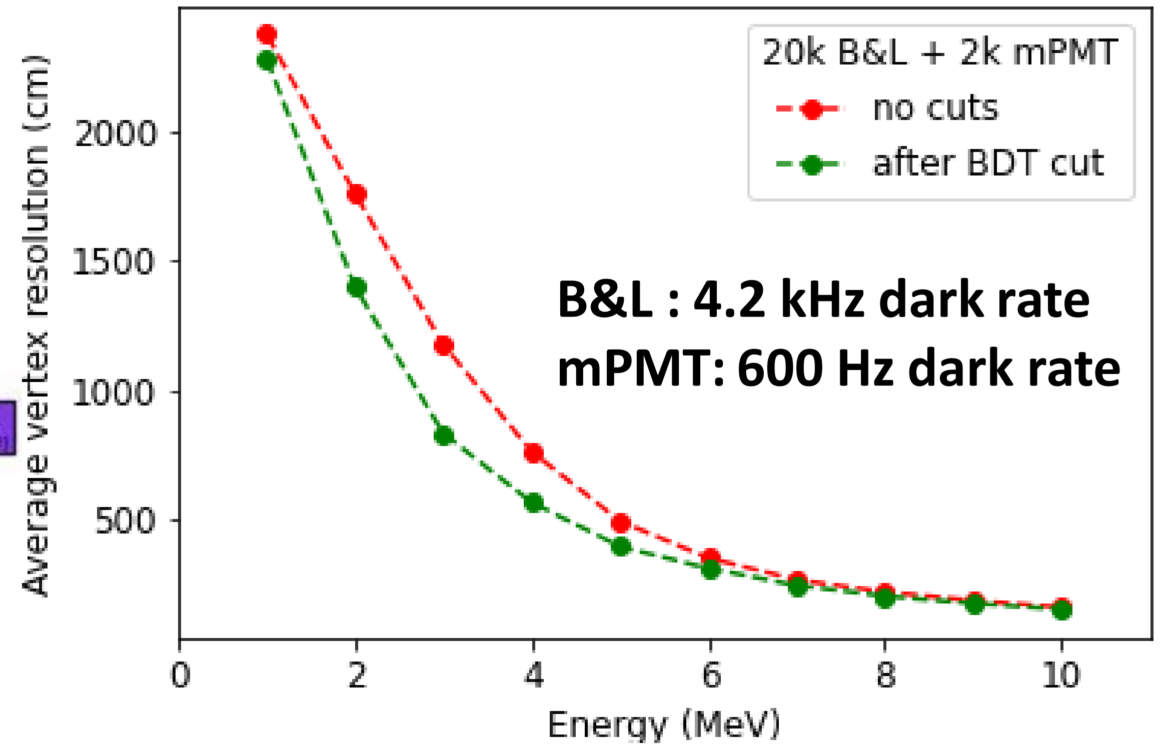


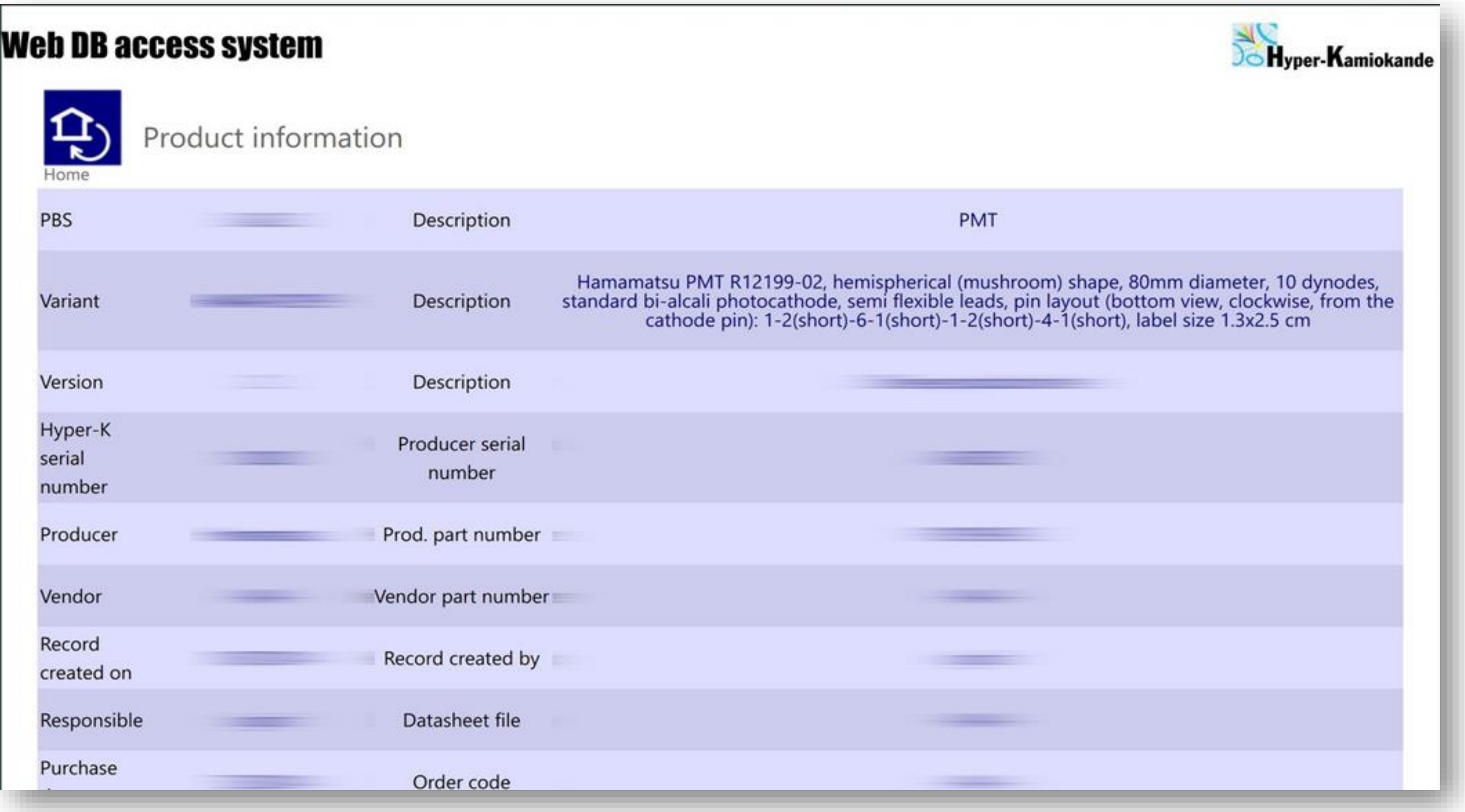
Activities within the Hyper-K experiment

Dark rate reduction through ML

- software reduction of the background noise (Dark Rate) for the Hyper-Kamiokande-FD detector using Machine Learning.
- Currently, the reduction of the Dark Rate takes place through the Boosted Decision Tree method, implemented through the Python Scikit-learn libraries specialized for Machine Learning.







Web DB access system

Hyper-Kamiokande

Product information

Home

PBS		Description	PMT
Variant		Description	Hamamatsu PMT R12199-02, hemispherical (mushroom) shape, 80mm diameter, 10 dynodes, standard bi-alcali photocathode, semi flexible leads, pin layout (bottom view, clockwise, from the cathode pin): 1-2(short)-6-1(short)-1-2(short)-4-1(short), label size 1.3x2.5 cm
Version		Description	
Hyper-K serial number		Producer serial number	
Producer		Prod. part number	
Vendor		Vendor part number	
Record created on		Record created by	
Responsible		Datasheet file	
Purchase		Order code	

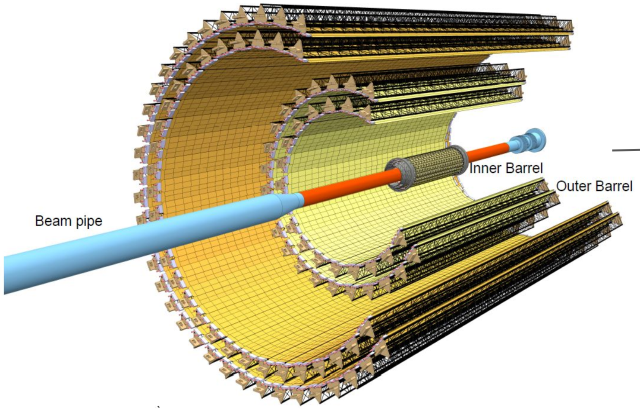
Activities in the Computing group

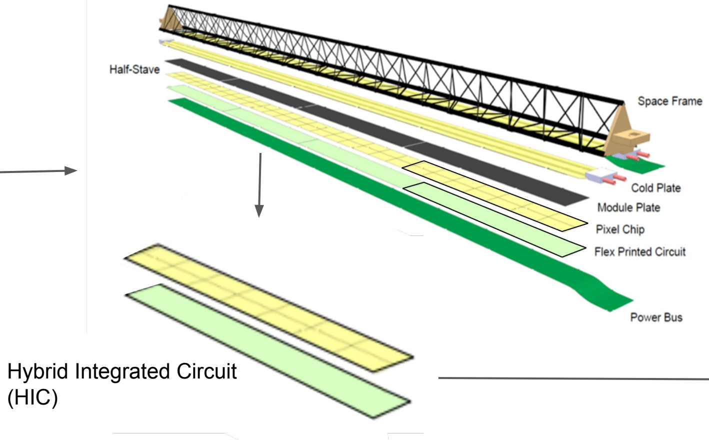
- HPC Compute Resource Administrator:
- Installation and configuration of all the necessary software (Python, Scikit-learn, WCSim, LEAF) for the aforementioned activities
- To optimize resources for the required workload
  - Parallelization of the simulation and analysis tools (WCSim, LEAF) of the HK experiment
- Member of the task force for the deployment of the experiment database borrowed from the KM3NeT DB

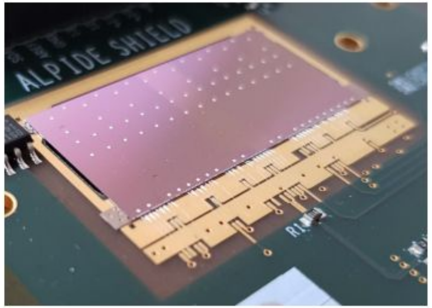
Calibration of the Upgraded ALICE Inner Tracking System

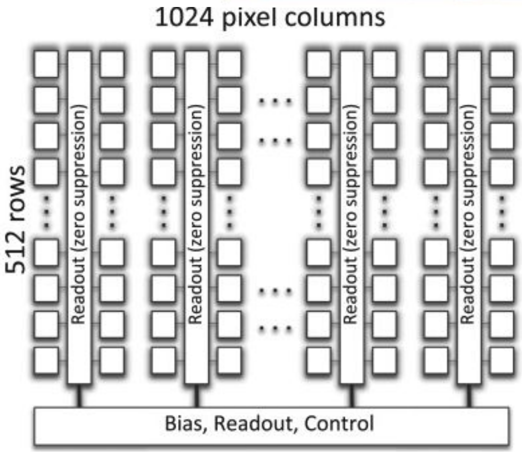
Andrea Sofia Triolo

ALICE ITS Overview

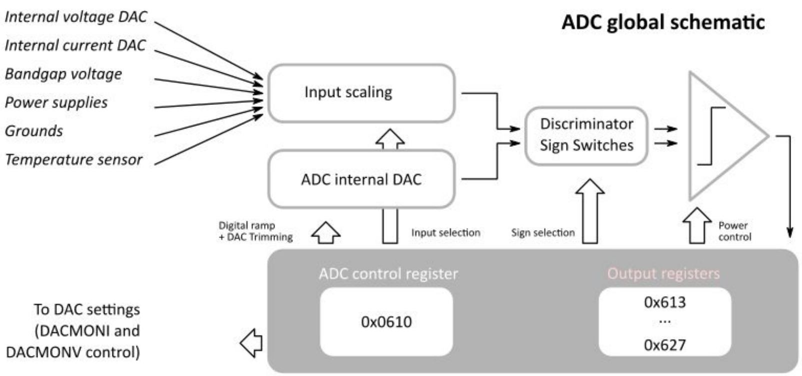








- Innermost ALICE detector
- 7 layers (Inner Barrel, Middle Layers, Outer Layers)
- 192 staves
- > 24000 monolithic active pixel sensor ALPIDE chips grouped in “staves”



ALPIDE application-specific integrated circuit includes a **monitoring ADC** to measure quasi-static voltages and currents generated on chip, that also features a **temperature sensor**.

Temperature sensor calibration:

$$T(°C) = (m_{chiller} \cdot T_{ADC}(AVDD_{ref}) + q_{chiller}) \cdot (1 + C_{heating})$$

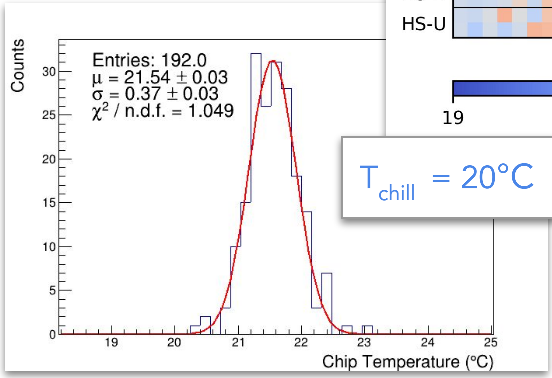
Chiller temperature dependence

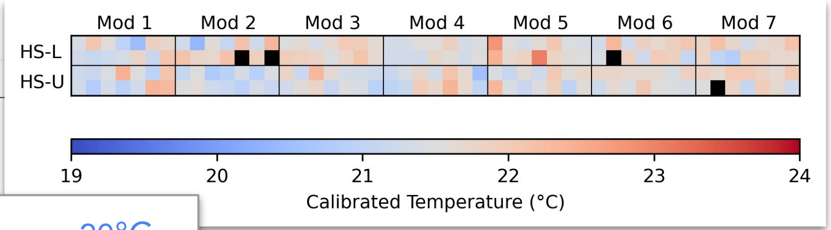
AVDD dependence

Heating up of the chips

Data taking performed through **ITSComm**, a C++ middleware between the DCS UI and the software that communicates with the CRU.
It maps all the detector functionality.

ITSComm functionalities have been expanded in order to perform data taking and calibration.





$T_{chill} = 20°C$

$T(°C) > T_{chiller}$

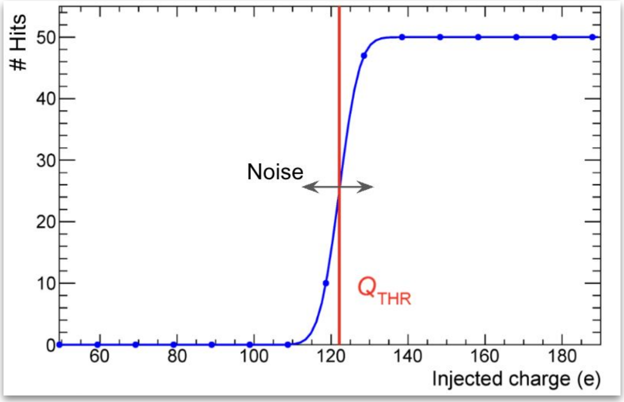
ALPIDE chips Threshold calibration and Threshold evolution

Full Detector calibration procedure:

- ➔ Digital/Analog scan: find and mask noisy double columns
- ➔ Threshold tuning
  - ◆ Tune VCASN and ITHR to set a threshold of 100 e-
- ➔ Threshold scan: verification of tuned thresholds
  - ◆ Fit of the derivative of the erf → THR and noise
- ➔ Noise calibration: find and mask noisy pixels

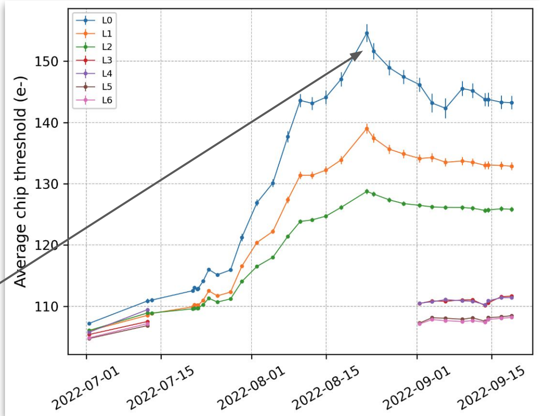
How:

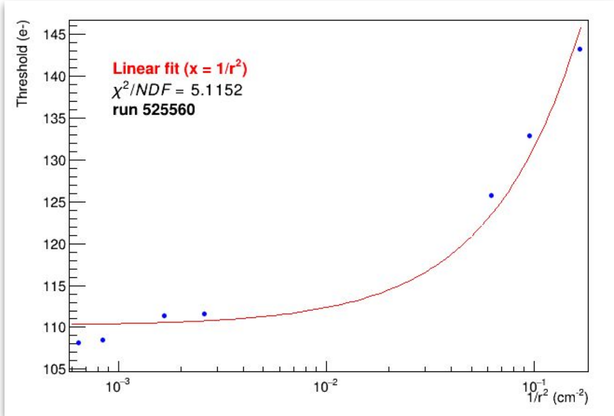
- ➔ Workflow on **ALICE O2** project:
  C++ functions that convert raw data (hits) in ROOT TTrees and in CCDB objects that can be analyzed



Threshold changes over time because of radiation effects.

24-08-2022
LHC stop





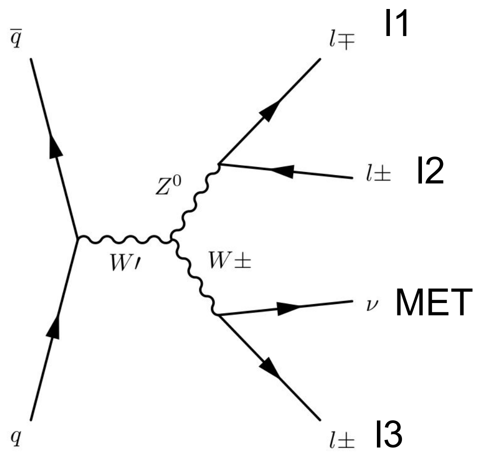
Threshold varies over layers with the $1/r^2$ law

Search for narrow-width diboson resonances with RunII data at CMS

Andres Vargas
The Catholic University of America
ESC22 - Oct/22

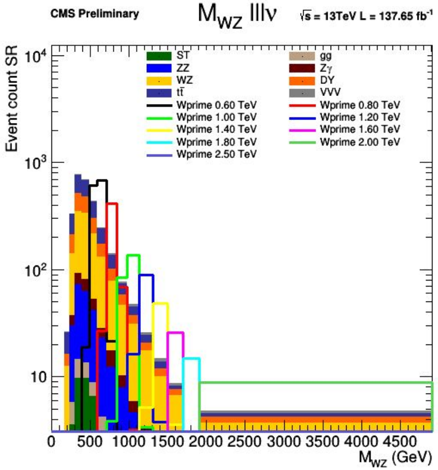
Analysis Strategy

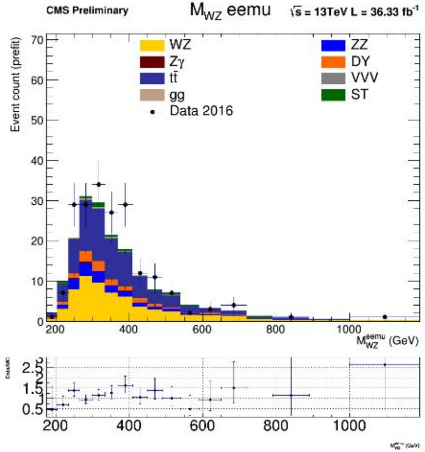
- Analysis is performed as a *bump-hunt search* on the spectrum of the invariant mass for the diboson resonance candidates. W' samples were generated in the 0.6 to 4.5 TeV range
- Analysis is splitted into 4 channels:
  - 3e = Z->ee W->ev
  - 2e1μ = Z->ee W->μv
  - 1e2μ = Z->μμ W->ev
  - 3μ = Z->μμ W->μv
- Background estimation is made purely from MC simulation for SM processes with similar signature in the final state:
  - Z+Jets
  - ttbar + TTV
  - WZ3LNu
  - ZGTo2LG
  - ZZTo4L
  - VVV
  - ST
- Two dedicated regions are used to evaluate the quality of the MC simulations.

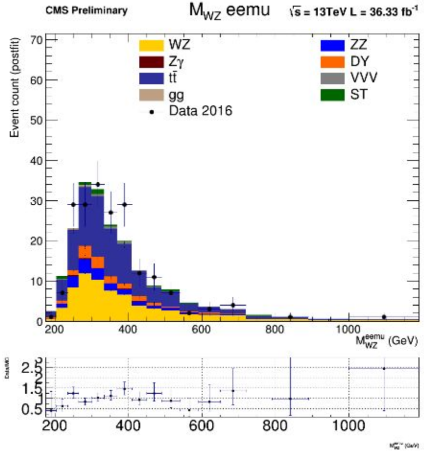


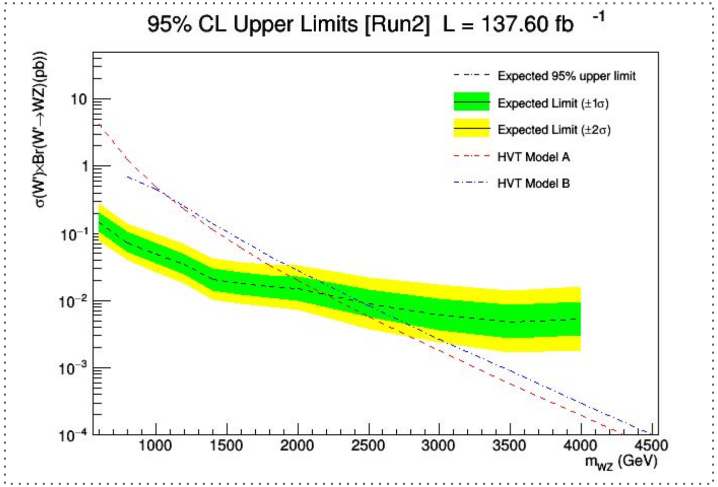
bump-hunt search

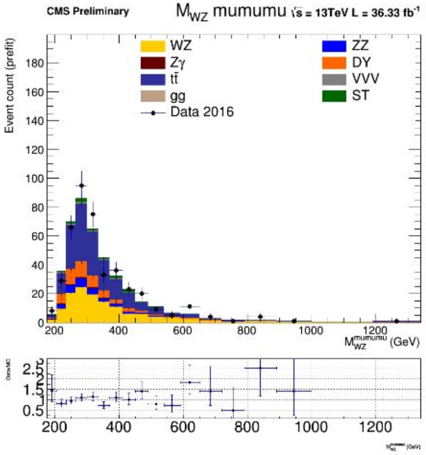
Stolen from Andrea Malara's slides

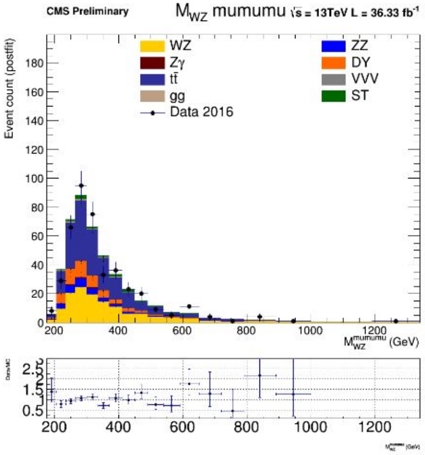








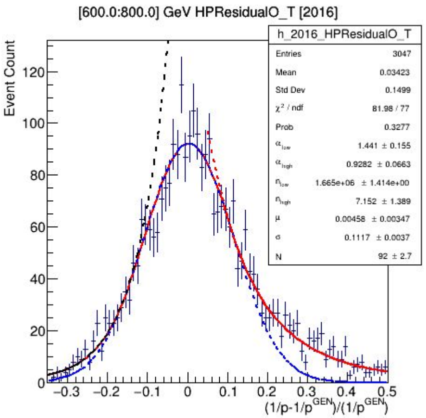


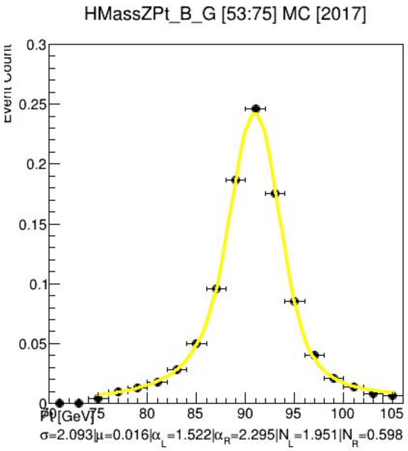


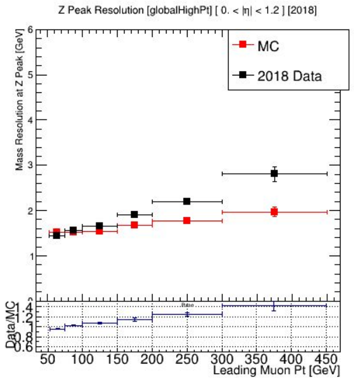
Some plots to get the gist of the analysis… Tools Used: NanoAOD, ROOT, PROOF, LPC-CondorHTC (FNAL), Higgs Combine

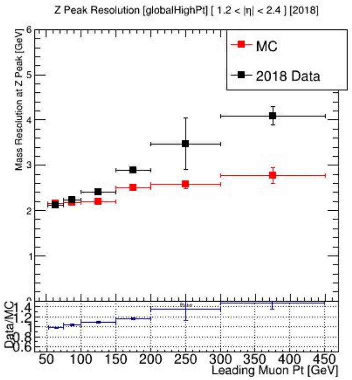
Contributions from the Analysis

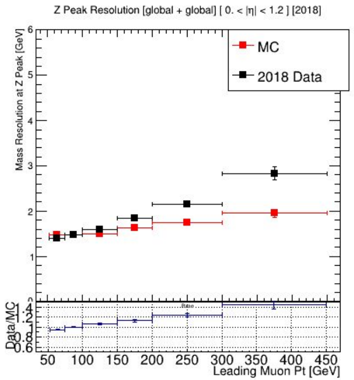
If you ever used muons at CMS…

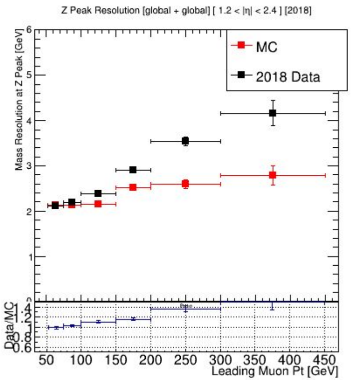








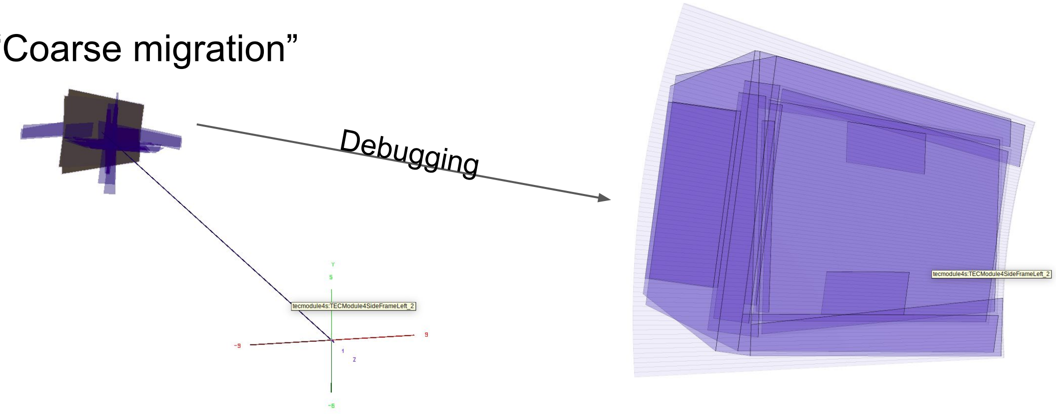


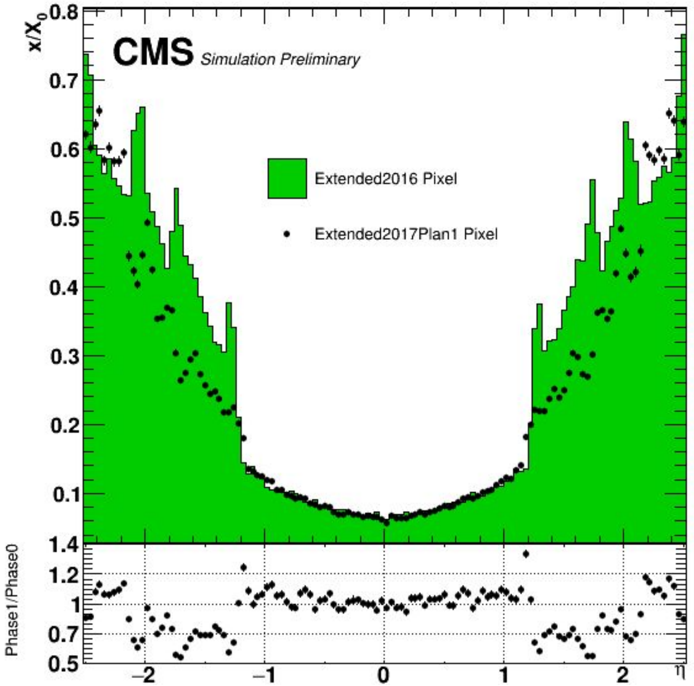


Contributions from the PhD

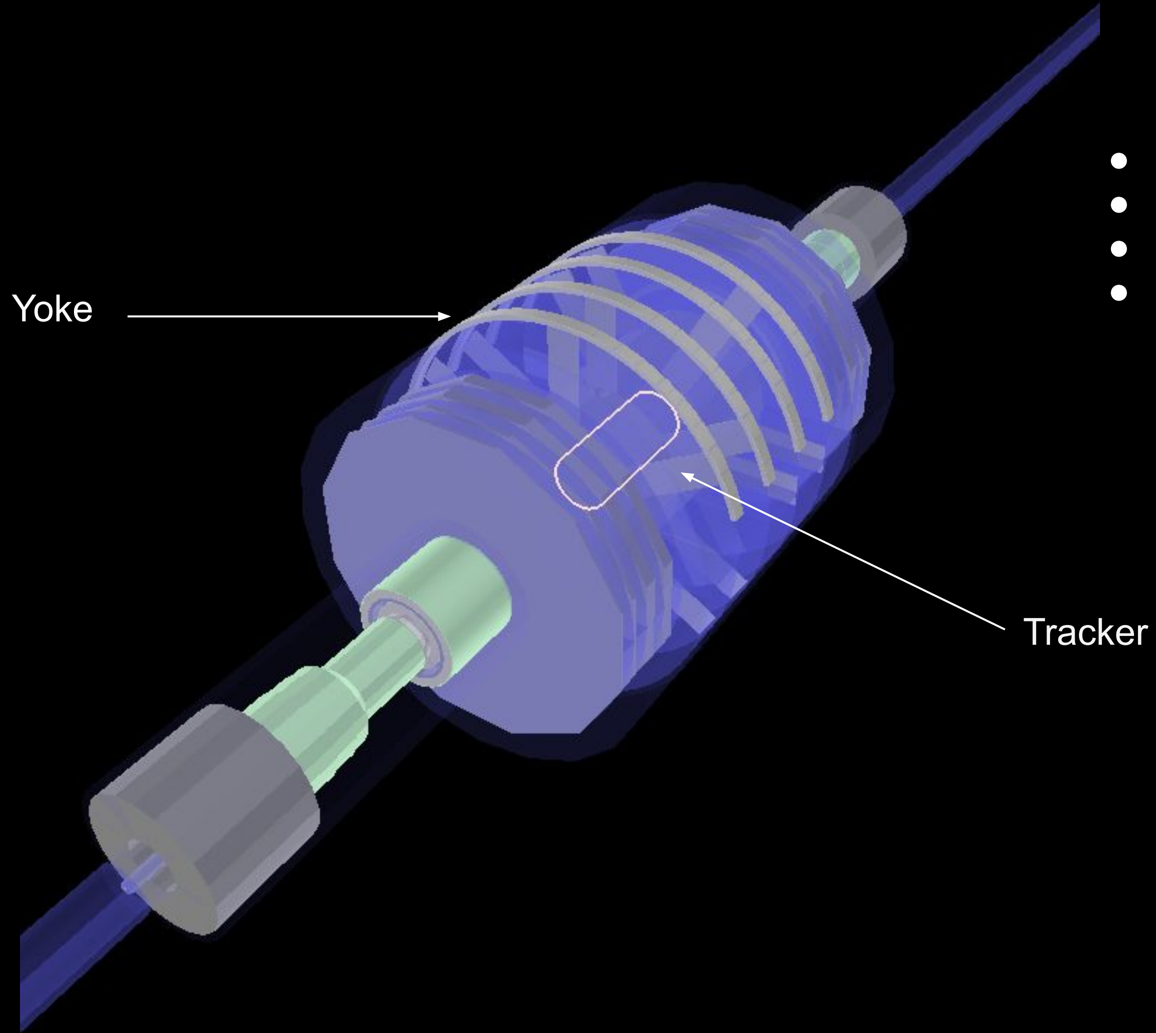
If you ever used MC simulation at CMS…
… or if you ever wrote a thesis after 2019

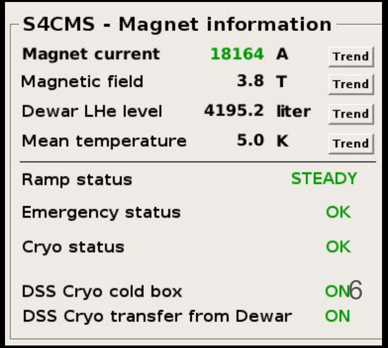
"Coarse migration"





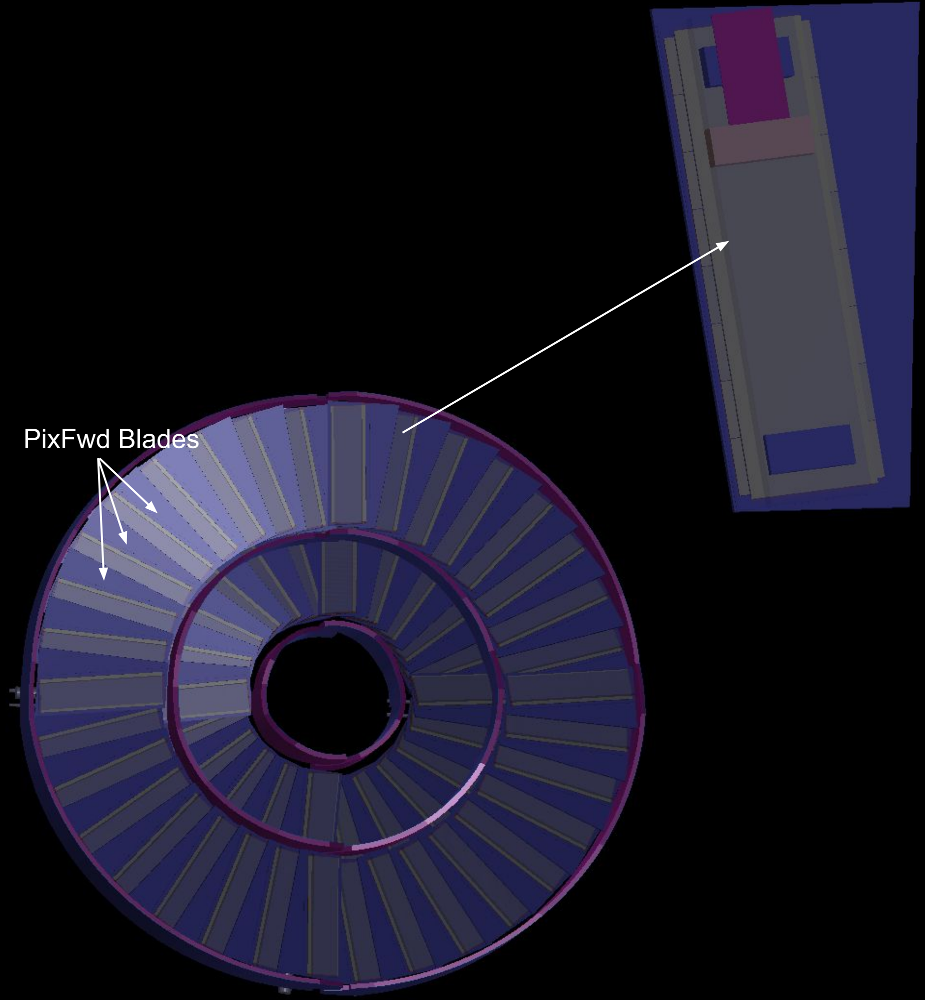
- Muon Detector
- HCal
- Ecal
- Tracker

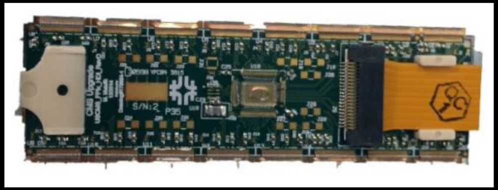


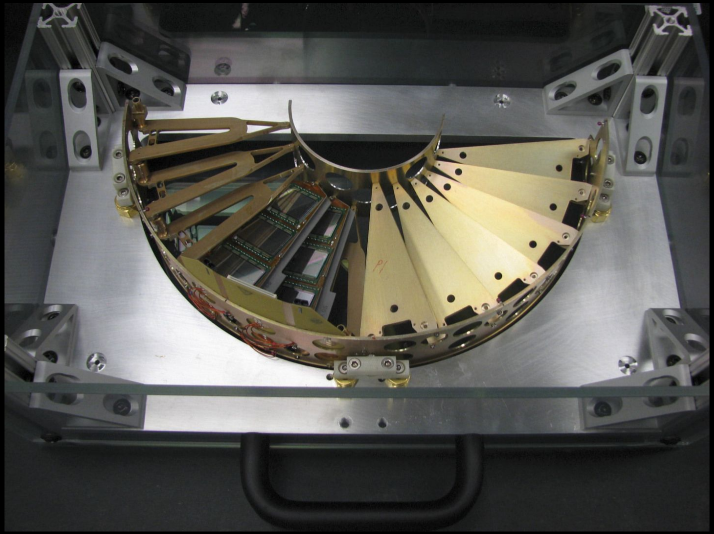


S4CMS - Magnet information

Magnet current	18164	A	Trend
Magnetic field	3.8	T	Trend
Dewar LHe level	4195.2	liter	Trend
Mean temperature	5.0	K	Trend
Ramp status		STEADY	
Emergency status		OK	
Cryo status		OK	
DSS Cryo cold box		ON	
DSS Cryo transfer from Dewar		ON	


PixFwd Blades





Summary

- Performing a physics search at CMS
- Contributing to computing activities
- Came for the physics, stayed for the computing
- Graduating in May/23 🤞Currently looking for "the next step"

Resume:
QA Version: https://bit.ly/3Rvmzu8
Dev Version: https://bit.ly/3rqiHjr

Parallel Event Interpretation

An Introduction to Me and My Doctoral Project

Zenny Wettersten (IT-GOV-INN)
zenny.wettersten@cern.ch

Introduction

- Double BSc from Lund University, Sweden (maths and TPP)
- MSc from Lund University, Sweden (TPP)
- Leaned more towards HEP-PH over my studies
- Now doctoral student at TU Wien, seated at CERN

Project Description

Madgraph5_aMC@NLO is an event generator to simulate the initial particle collisions of HEP experiments, such as those currently in operation at the Large Hadron Collider at CERN. With the forthcoming data taking periods, e.g. High Luminosity LHC (HL-LHC), the performance of all data simulation and processing software packages need to be drastically increased, including those of event generators. A first step into this direction for Madgraph5_aMC@NLO has been undertaken already by leveraging on CPU features such as vector instructions and parallel processing but also offloading the most compute intensive part of the Madgraph5_aMC@NLO event generator onto GPUs. The complexity of the physics processes that can be simulated with this improved software are currently limited to "leading order (LO)" calculations. For experiments to take full advantage of the improvements the software needs to be also made available for "next to leading order (NLO)" calculations.

The PhD project concerns GPU development for event generation within the Madgraph5_aMC@NLO framework, particularly with respect to NLO calculations. A scheme for parallel processing will be developed and implemented using the CUDA API, and possibly other programming models, with the aim of speeding up scattering element calculations.

Practically, the work will be to improve the performance of Madgraph5_aMC@NLO by implementing the workflow for computing platforms designed for acceleration such as GPUs; the optimization of the CPU execution; developing a strategy for the heterogeneous execution of the overall workflow; and investigating possible further performance improvements by reducing the required calculation precision of the algorithmic part of the workflow. The applicability of the new architecture will be cross checked and discussed in exchange with the stakeholder LHC experiments. In collaboration with the Madgraph5_aMC@NLO software engineering team the developed software will be discussed and made available for integration into the upstream project.

TL;DR: Let's port MGaMC to GPUs.

Current projects

- Optimising GPU port of MG5aMC in before production release
- Porting event reweighing functionality to GPUs
- Taking NLO HEP calculations to GPUs
- *(Developing HEP tools for new mathematical formalisms)*

Scientific computing goals

- GPU programming (particularly CUDA)
- Efficient memory allocation and structures
- Precision necessities, when to use floats vs doubles
- Optimising calculations and I/O

Bayesian multiscale mixture models via Hilbert-curve partitions

Daniele Zago

Department of Statistical Sciences, University of Padova

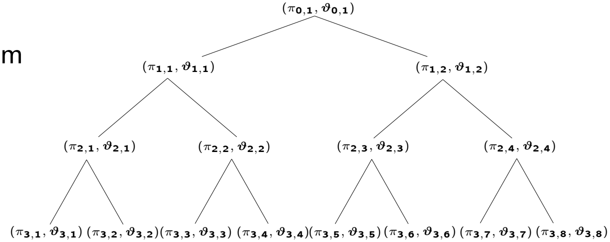
September 29, 2022

Multiscale models

- A Bayesian nonparametric multiscale mixture model is a mixture of the form

$$f(y_i) = \sum_{s=0}^{\infty} \sum_{h=1}^{2^s} \pi_{s,h} \mathcal{K}(y_i; \boldsymbol{\vartheta}_{s,h}).$$

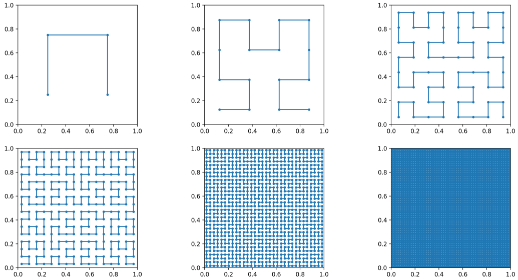
- The kernels $\mathcal{K}$ become more concentrated as the depth s increases and more spread out as h varies.



Truncated binary tree representation of the multiscale model

Hilbert curve

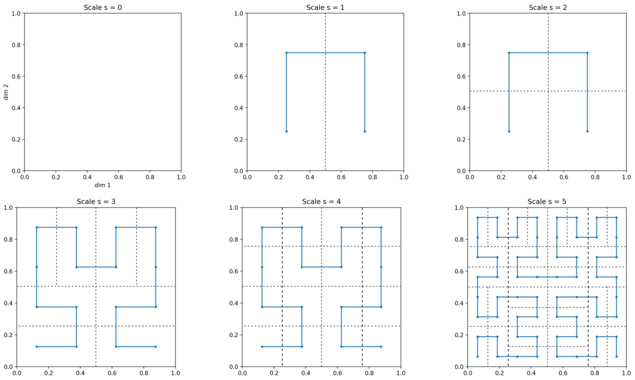
- Bayesian estimation of the above model **requires**:
  1. Partitioning a d-dimensional space into cubes (easy)
  2. Index these cubes using a binary tree (not easy)
- **Solution**: use the Hilbert curve to both define and index the partition.



First six approximations to the Hilbert curve.

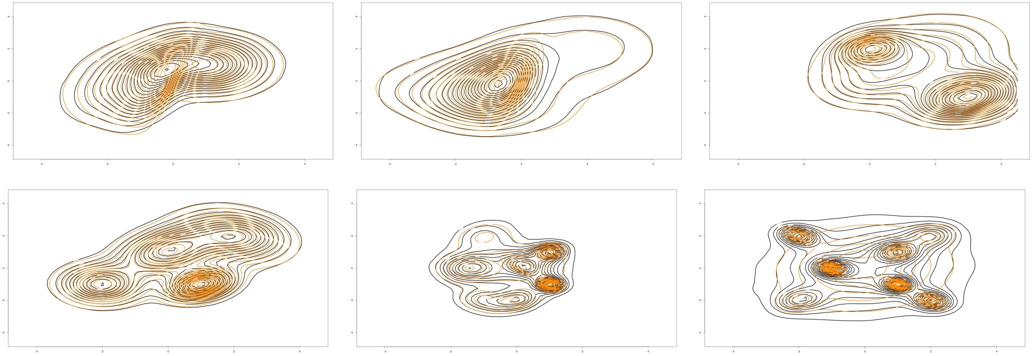
2D example

- The Hilbert curve partition has to be computed only **once**.
- Efficient `C++` code freely available (`shorturl.at/cFJUZ`).



Binary partition of $[0,1]^2$ obtained by using the Hilbert curve partition.

Example of application



True densities (*black*) and reconstructed densities (*orange*) from observed data using the proposed model.

Computational issues

- The model is **exponential in the number of components**, hence efficient `C++` code is required to run the estimation procedure.
- Estimation is **easily parallelizable** since each node can be treated separately.
- Some mathematical tricks (optional stopping theorem, ...) could be used to collapse the binary tree and speed up computation.
- For mathematical reasons, generating from a **truncated multivariate normal distribution** is crucial for applying this model. This is a hard topic with very recent contributions[1].

[1] Genz and Bretz, 2009; Botev, 2017; Botev and Belzile, 2020.

Botev, Z. I. (2017). “The Normal Law Under Linear Restrictions: Simulation and Estimation via Minimax Tilting”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.1, 125–148. arXiv: 1603.04166.

Botev, Z. I. and Belzile, L. (2020). *Truncated Multivariate Normal and Student Distributions*.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. en. Lecture Notes in Statistics. Berlin Heidelberg: Springer-Verlag.