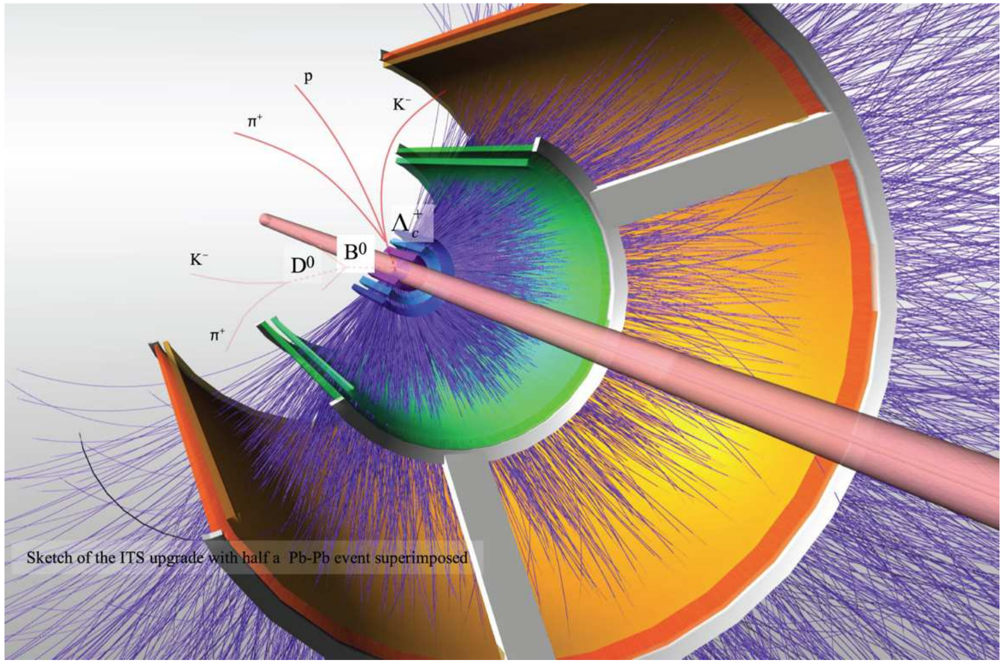
Design of high-performance vertexing and tracking algorithms optimized for large data throughput

Stefania Bufalino
Università and INFN Torino

Seventh INFN International School on Architectures, tools and methodologies for developing efficient large scale scientific computing applications

The Project Goal

- **Developement of an online reconstruction algorithm for the ALICE vertex detector at the LHC (CERN) after the Upgrade planned in 2018 and hence in a high luminosity and multiplicity environment**
- Key point for the project success: **parallel computing approach** exploiting the parallel architectures available on the market



-Primary vertex reconstruction
-Tracking in a high multiplicity environment (about 10^4 particles in one Pb-Pb collision)
- **Challenging aspect of the project:**
  - -Online selection of interesting decay topologies having displaced vertices w.r.t. the primary vertex
  - → Weak decays of particles containing heavy quarks (i.e. charm and beauty)

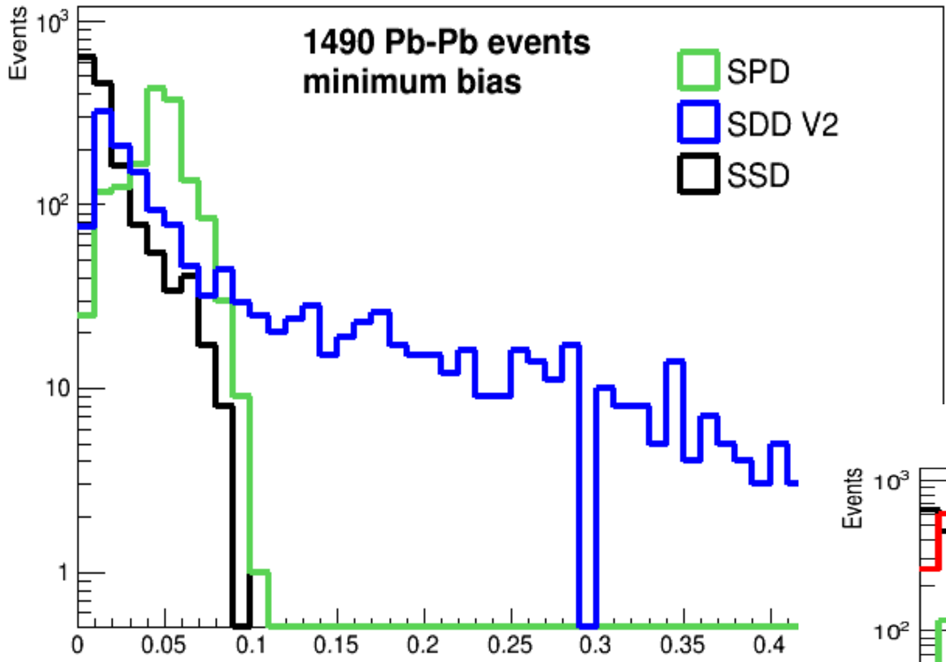
Now offline (time and memory consuming) → in the future it must be done online

Milestones for the next two years

- **First Year**
  - optimization of the local reconstruction and the online vertexing algorithm used in the ALICE vertex detector
    - Improvement in CPU and memory usage (partially done)
  - tests with different computing architecture and of different H/W solutions according to different memory handling
    - Different approaches will be tested like multi core CPU architecture, GPUs and Intel MiC

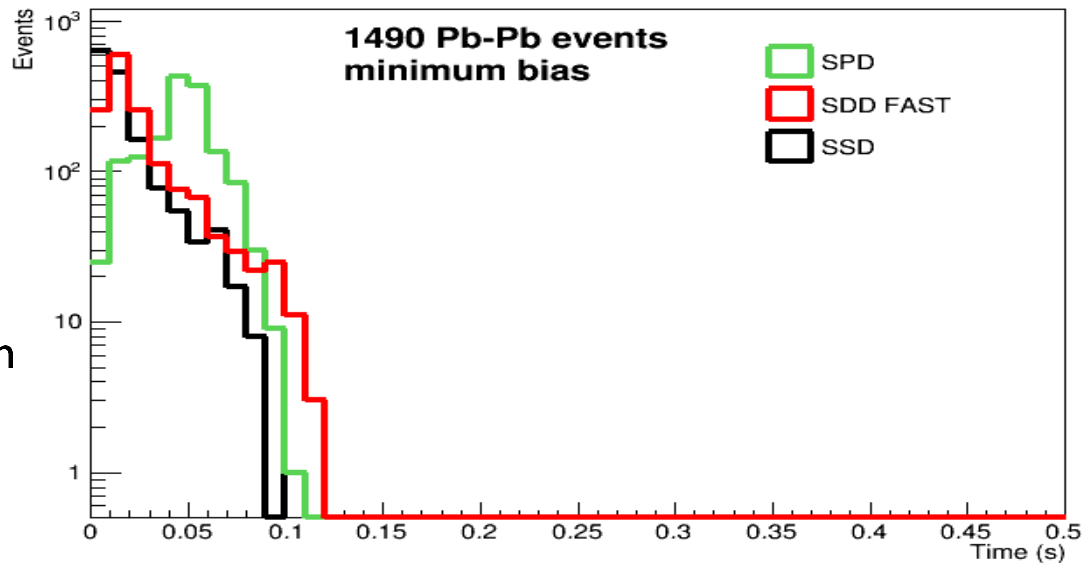
- **Second Year**
  - validation of the vertexing algorithm on parallel approach solutions defined during the first year of the project
  - validation of the online tracking code
  - algorithms optimization within the parallel approach chosen after the tests performed during the first year of the project.
  - define event topologies, for instance containing heavy flavour particles, that can be tagged with the reconstruction code developed in the previous milestones.

First results



Execution time
→FAST algorithm is comparable with the others

Optimizing the data structure in a standard C++ reconstruction program for the present Inner Tracking System we obtain a great improvement in the cluster finder execution time



INFN Bari

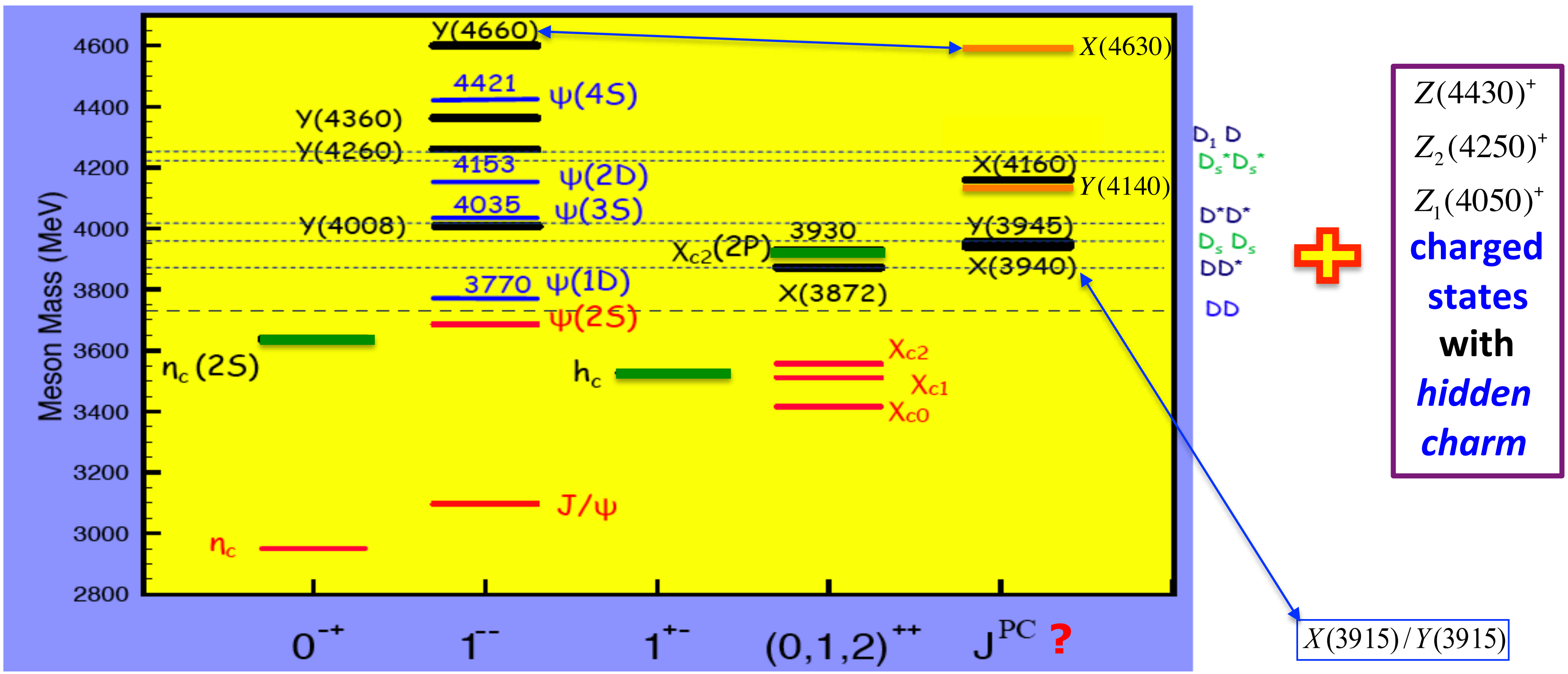
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Search and study of exotic *charmonium-like* states in CMS

Leonardo Cristella

Exotic charmonium

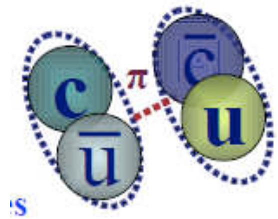
- In the last 10 years ~20 new states have been observed (known as *X, Y, Z* states) above the *open-charm* threshold $D\overline{D}$ *(~3.8 GeV)*
- They are inconsistent with expected charmonium spectrum
- Quantum numbers still not experimentally established for most of them
- In particula the *Z(4430) hidden-charm* charged state has generated a great interest since it must have a minimum quark content of $c\bar{c}d\bar{u}$ and thus it would represent an unequivocal manifestation of a 4-quark state meson!



Interpretative models for exotic states

To explain their nature alternative models have been introduced:

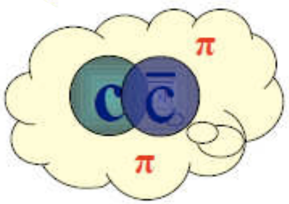
- **Hadron molecules**:
  weakly bound states formed by 2 (or more) hadrons
  conserving their hadronic entirety inside the molecule.



- **Hybrids**:
  bound states of quarks and gluons
  (i.e. charmonium + excited gluons)

- **Tetraquarks**:
  bound states made of a diquark-antidiquark pair
  (charged and doubly charged states are foreseen)

- **Hadro-charmonium**:
  binding a compact charmonium state inside an excited state
  of light hadronic matter (QCD analogous of the Van der Waals force)



Perspectives in CMS

- CMS can perform both an inclusive and exclusive search:
  - No signal hint with the Run-I data in the mass spectrum:
    - $\psi'\pi^-$ with $\psi' \to \mu^+\mu^-$ (*ψ' inclusive triggers*)

    The *Z prompt* production cross section in *pp* collisions is unknown and no theoretical predictions available… still to look into:

    - $\psi\pi^-$ with $\psi \to \mu^+\mu^-$ (*J/ψ inclusive triggers*)
  - exclusive searches from *B* decays:
    - $B^0 \to \psi'\pi^- K^+$ with $\psi' \to \mu^+\mu^-$
    - $B^0 \to J/\psi\pi^- K^+$ with $J/\psi \to \mu^+\mu^-$ would be the control channel
- In the exclusive search a high purity and statistical sample is needed in order to perform a *full amplitude analysis* of the 4-body decay which takes into account the correlation between the ψ' decay (and its polarization) and the mass and angular dependence of the $K\pi^-$ system.
- The *full amplitude analysis* requires:
  - an high purity sample which can be achieved through some ML techniques
  - a very time consuming fitting procedure which can be speed up by using GPUs (through the GooFit framework)
- Hence I am currently interested in the approaches that can be more effective in parallelizing scientific applications

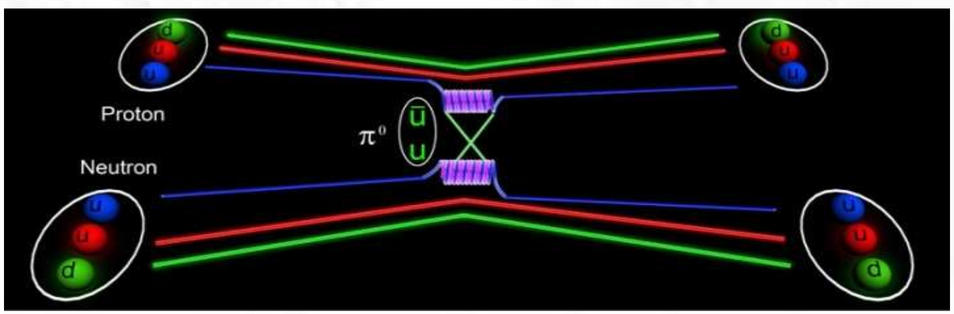
ESC 15

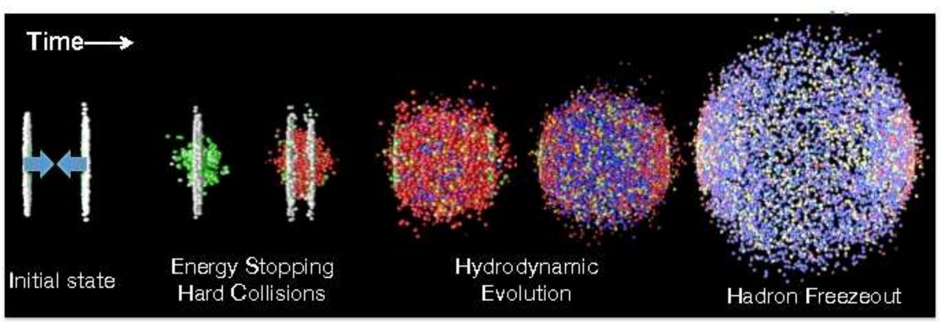
Giovanni Eruzzi

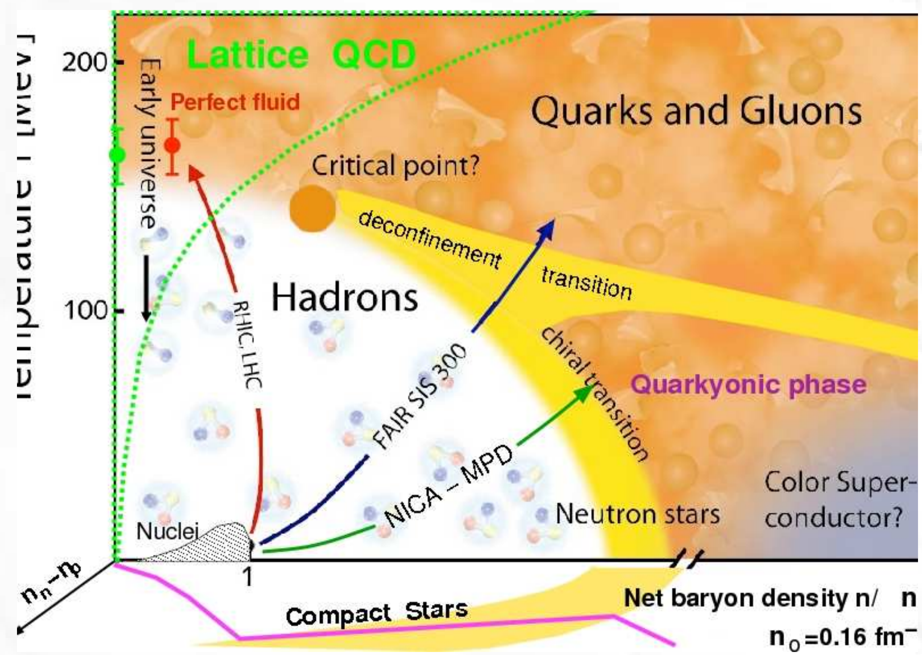
PhD in Physics - 2nd year

University of Parma and INFN

Strong interactions, Lattice field theories and Monte Carlo simulations







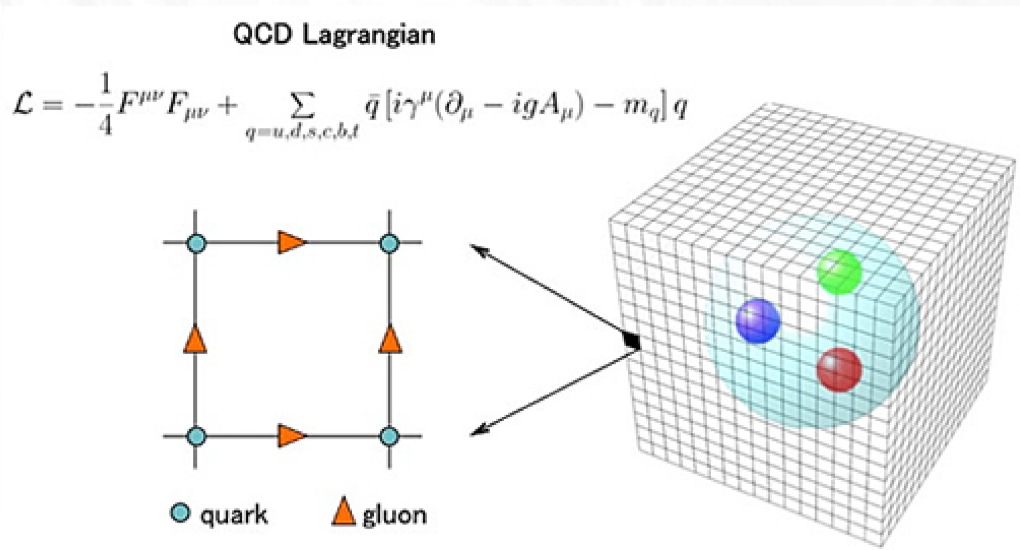
Ultimate goal ⟶ Understanding quark-gluon plasma, neutron stars...

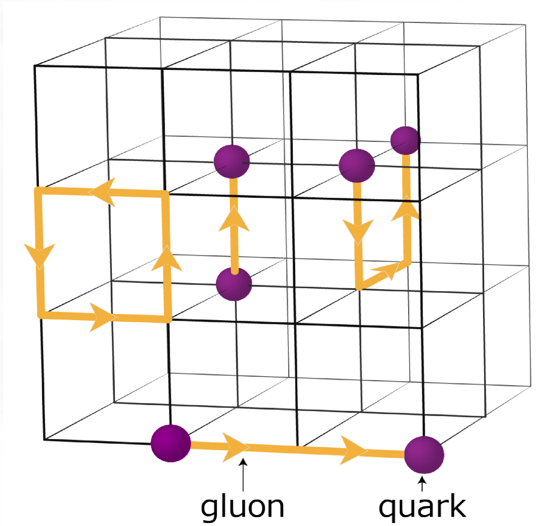
ESC 15

QCD → Space-time dynamics determined by a local Lagrangian

Discretize on a **lattice**! ↔ **Nearest-neighbour** interactions

The ideal setting for **parallel computations**!





→ Produce field configurations with the desired probability distribution

→ Expectation value of observables

→ Predictions and comparison with experimental data

Current research activity

Fabio Ferrari

Current research activities

- Search for CP violation in $\Lambda_b^0 \to pK^-$ and $\Lambda_b^0 \to p\pi^-$ decays at the LHCb experiment
- Challenging measurement since:
  - Λ_b^0 production asymmetry could be present at LHC
    - This effect originates from the fact that initial state contains only protons
    - Production asymmetries could mimic CP violation effects → one needs to measure this spurious asymmetries with high precision
  - Particles/anti-particles in the final state interact with different probabilities within the LHCb detector
    - Need to measure proton detection asymmetry with very high precision

Interests in scientific computing

- Maximum likelihood fits are CPU intensive processes
  - Computation of 1^{st} and 2^{nd} order derivatives to minimise logL function
  - Parallelization is fundamental to perform such fits
    - High number of parameters
    - High number of events
  - Often perform simultaneous fits to different quantities and/or datasets
    - Also benefits from parallelization over different CPUs

Sergio Fonte

National Institute for Astrophysics

ESC15 26-31 October 2015

Sergio Fonte in a Nutshell

Major in Physics at the University of Rome *La Sapienza*

***Thesis**: involved the development of a system dosimetric passive to monitor of the Martian environment under the Aurora project, the mission ExoMars / Pasteur ESA.*

Ph.D. in Space Sciences, Technologies and Measurements at the University of Padua

***Thesis**: implementation of a simulator and validator for the sequences of the flight instrument VIR (Visible and Infrared Spectrometer) aboard the NASA mission Dawn*

Technological researcher at the Institute for Space Astrophysics and Planetology (INAF-IAPS)

***Task**: management of the instrument VIR during the Dawn activities and planning of his observations*

Research activities and interests

Numerical simulations applied to (in the past)

- ***Solar System dynamics*** *during its formation*
- ***Optimization*** *of acquisitions of spectral imaging data in terms of illumination conditions*

Cybernetics applied to (present)

- ***Automatic controls theory*** *during the simulation and the execution of the observational sequences of space missions*
- ***Systems analysis*** *to verify the spacecraft telemetries*

Artificial Intelligence applied to (for the future)

- ***machine learning*** *to program an A.I. able to process a feasibility analysis and to assess the limits of the flight instrument during an observation campaign*
- ***parallel data mining*** *of space missions data during and after the space mission observational campaign*

... I love C/C++ since late 1993!

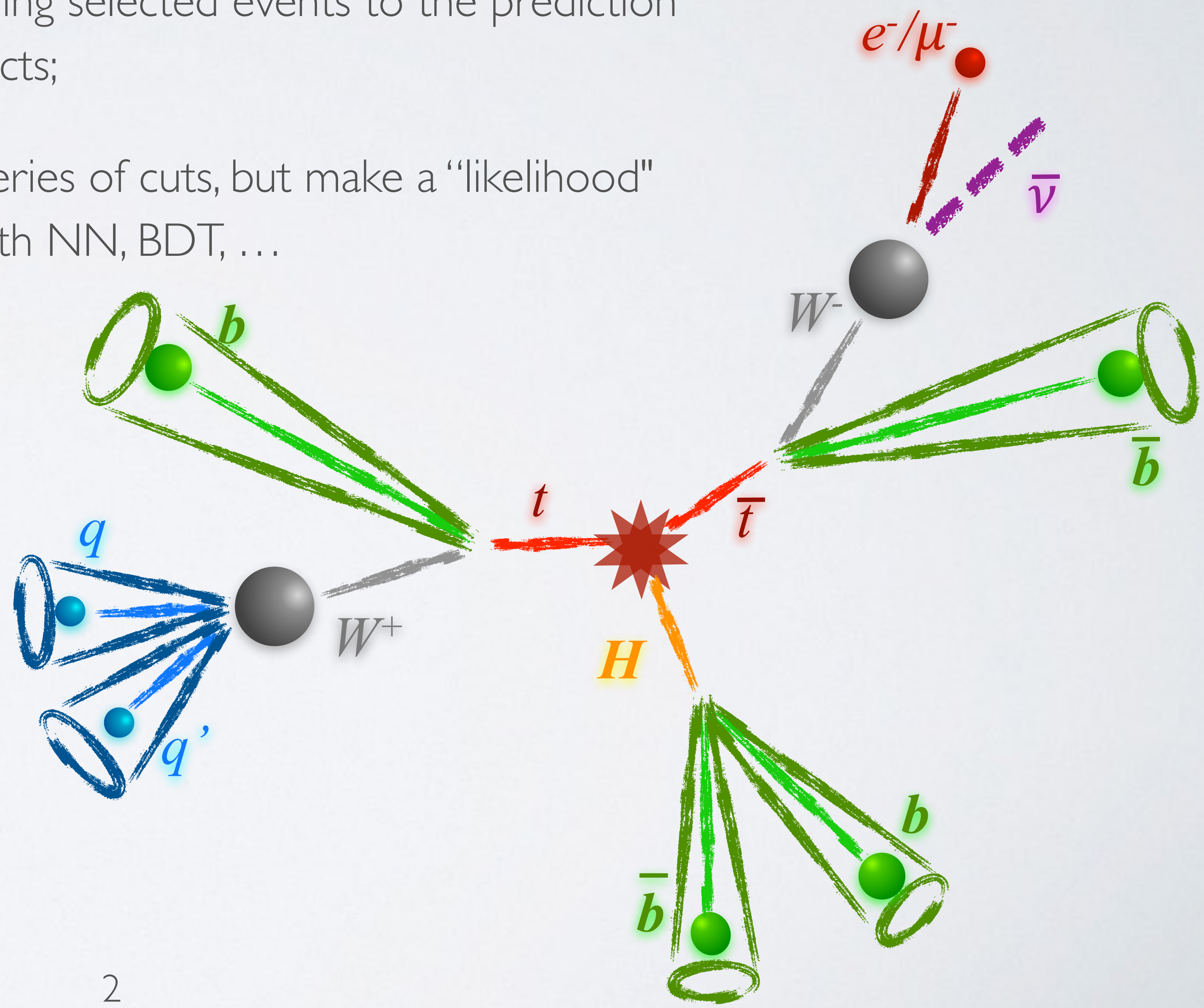
ttH boosted
and
TOM tagging

Biondi S., Franchini M., Spighi R.

Top quark and ttH analyses

- Working on **ttbar differential cross section** measurement and **ttH(H->bb)** cross section; both channels by selecting very high pT **(boosted)** objects;
- **Main challenges** (apart to treat the whole bunch of MC and real data)**:**
  - **Unfolding:** deconvolution techniques to bring selected events to the prediction level, accounting for detector migration effects;
  - **Multivariate selection:** not a sequential series of cuts, but make a "likelihood" dependent on many variables, optimised with NN, BDT, …

- **Boosted jet topologies** are not efficient with standard reconstruction due to cluster overlap;
- **New algorithms needed** like Template Overlap Method;



Top quark and ttH analyses

- Working on **ttbar differential cross section** measurement and **ttH(H->bb)** cross section; both channels by selecting very high pT **(boosted)** objects;

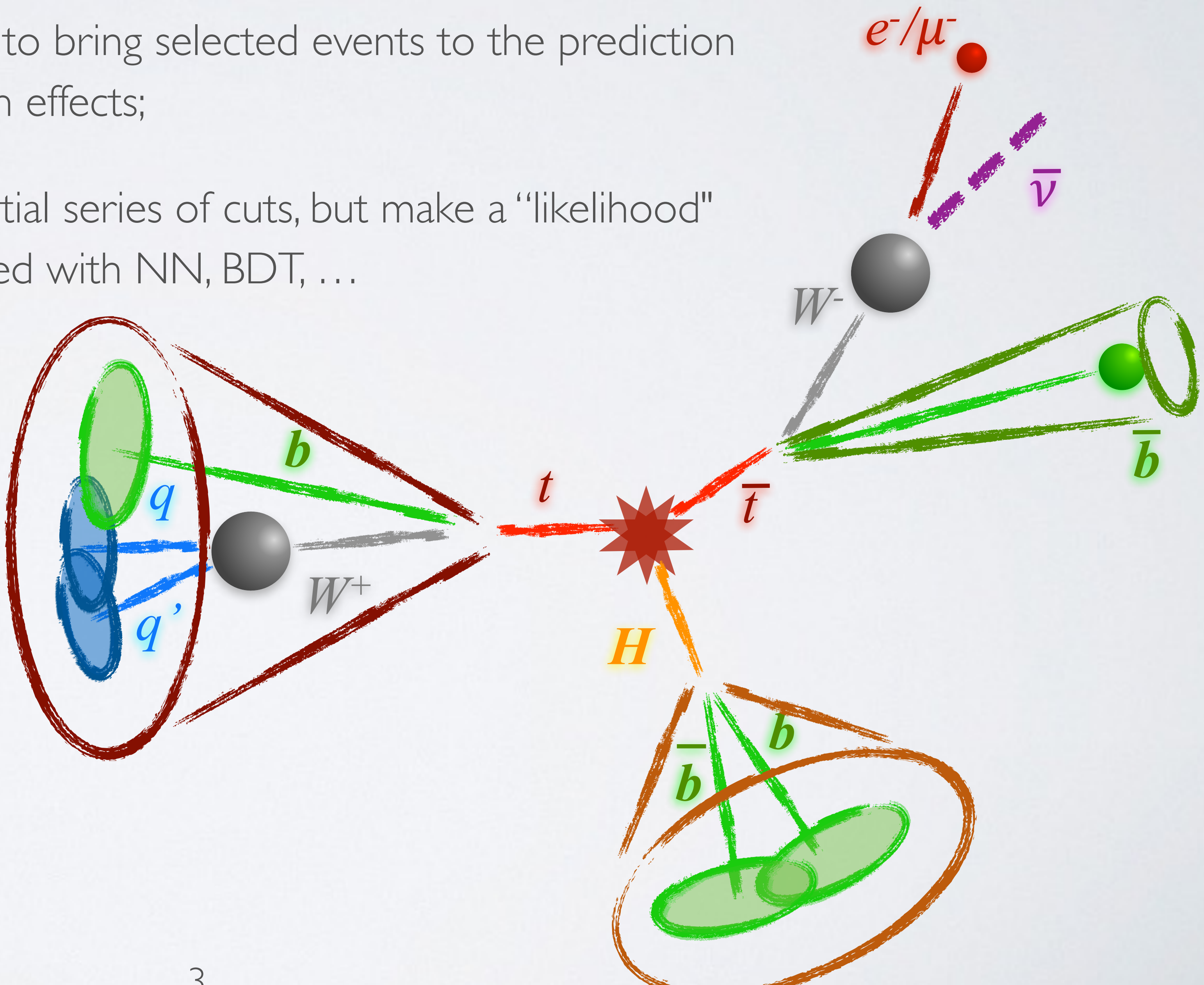
- **Main challenges** (apart to treat the whole bunch of MC and real data)**:**

    - **Unfolding:** deconvolution techniques to bring selected events to the prediction level, accounting for detector migration effects;

    - **Multivariate selection:** not a sequential series of cuts, but make a "likelihood" dependent on many variables, optimised with NN, BDT, …

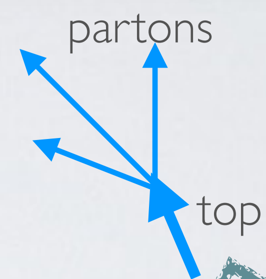
- **Boosted jet topologies** are not efficient with standard reconstruction due to cluster overlap;

- **New algorithms needed** like Template Overlap Method;



Top Template Overlap Method (TOM)

Reconstructed large-R jets
(working with any jet collection)
[AntiKt10]



Template Sample
Top quarks generated with uniform momentum.
Quite big data file ~50MB.

Comparison of the p_T (or energy) flow of **jet calorimetric clusters** with each parton of the **top Template**.

Define an overlap estimator

$$ov = \exp\left[-\frac{1}{2\sigma^2_{p_{T,p}}}\sum_{p=1}^{3}\left(p_{T,j}-p_{T,p}\right)^2\right]$$

$p_{T,j}$ = cluster pT
$p_{T,j}$ = parton p_T
$\sigma_{p_{T,p}}$ = 0.33

Complete disagreement 0 —— 1 Perfect agreement

- For each jet you have to run over all the simulated template of top decays to find the best one;
- Time consuming, some optimisation done so far: jet energy range, read the input only once, double->float, pointers…

Marco Fumana INAF-IASF Milano

Data reduction and data analysis software for Astronomical Spectroscopy

VIMOS MOS spectrograph Paranal Observatory	Cosmological surveys: **VVDS, zCosmos, VIPERS, VUDS, Vandels**
LBT Large Binocular Telescope	**LUCI**: MOS (infrared) spectrograph **MODS**: MOS spectrograph (visible band)
EUCLID NISP: infrared slitless spectrograph	ESA mission to map the geometry of the dark Universe. Euclid will cover the entire period over which dark energy played a significant role in accelerating the expansion

Software architecture

C/C++ Core	• Data structures • Class definitions • Algorithms
Wrapping them in python classes	• SWIG • Numpy C/API
Python like a glue	• Algorithm interactions • GUI • Shell Scripting Console

Marco Giardino

SAPIENZA UNIVERSITÀ DI ROMA

TECHNOLOGICAL RESEARCHER @ INAF - IAPS

iaps Institute for Space Astrophysics and Planetology
Istituto di Astrofisica e Planetologia Spaziali

INAF

EXOMARS

NASA DISCOVERY MISSION
Vesta
Ceres
Dawn
UCLA • JPL • ORBITAL • DLR • MPS • ASI • INAF • LANL

Europlanet
European Planetology Network

Current Activity

- Data Archiving manager for the VIR imaging spectrometer onboard NASA/DAWN mission.
  Processing pipeline based on Java/IDL technology.

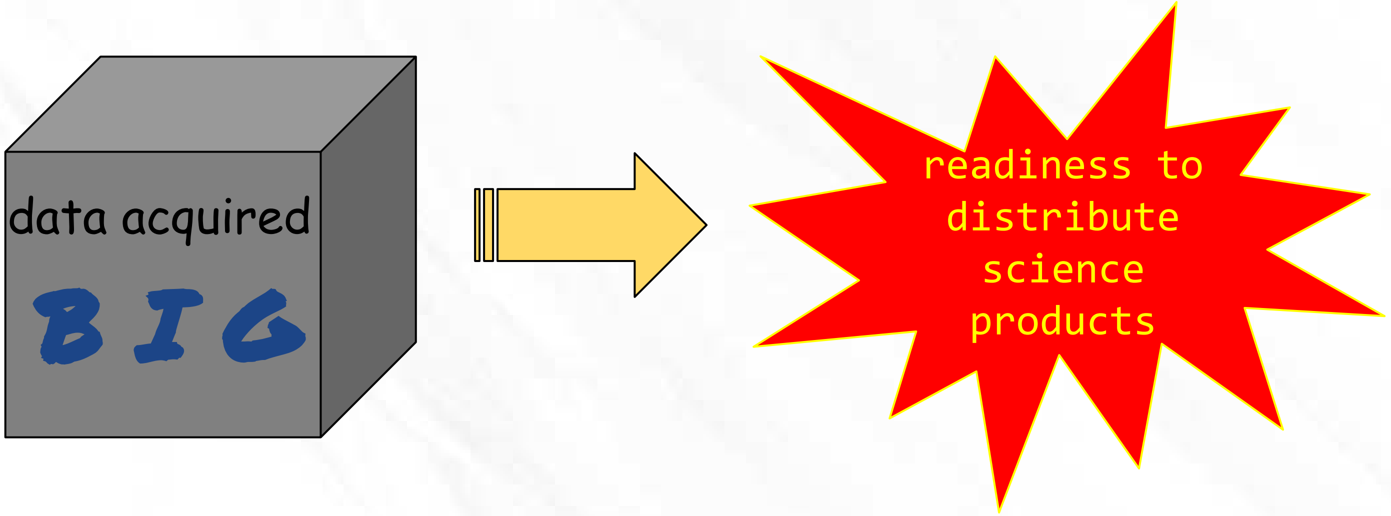
- Data Archiving manager for the Ma_Miss spectrometer onboard ESA/EXOMARS 2018 mission.

- Small Bodies and Dust Node technical manager for Europlanet.

- Virtual Observatory for planetary science (EPN-TAP) as a member of the Europlanet 2020 RI project.

Area of interest

- ★ Parallel computing to improve efficiency of data processing pipeline.
- ★ Efficient processing of large amount of data to be ingested into virtual observatories.
- ★ Vector computing and code optimisation methodologies.



Reconstructing tracks with displaced vertices with the $\overline{\text{P}}$ANDA detector at FAIR

Walter Ikegami Andersson

Uppsala University
for the PANDA collaboration

October 26, 2015
Bertinoro, Italy

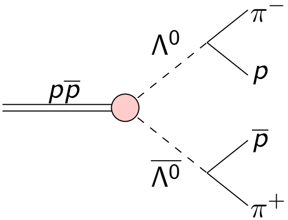
Physics motivation

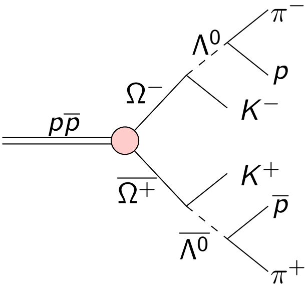
Why antihyperon-hyperon production?

- Hyperons produced at scales where QCD is poorly understood
- $\mathcal{CP}$ violation - needed to describe matter in the universe
- Never-before measured hyperon states
- Measure properties *e.g.* spin of hyperons

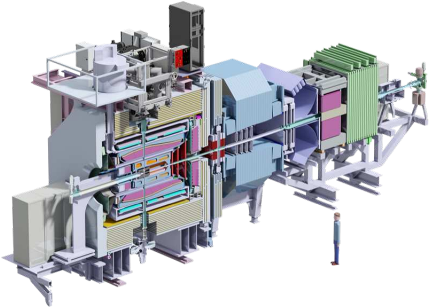
Challenge

Require large number of hyperons measured with high precision and accuracy!





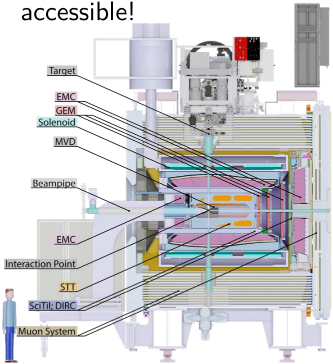
$\overline{\text{P}}$ANDA - antiProton ANnihilation at DArmstadt



- Target- and forward spectrometer provides a near 4π coverage
- $\overline{p}$ beam momentum of 1.5 - 15 GeV/c
- Wide range of hyperons accessible!

Target spectrometer used to detect particles with $|\theta| < 5^{\circ}$

- Several detectors *e.g.* STT, MVD, EMC
- Solenoid provides uniform $\vec{B}$ field
- High resolution measurement of momentum and PID



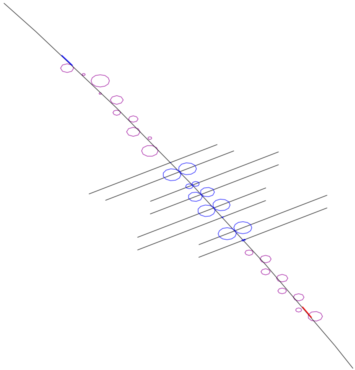
My work

Detecting hyperons poses several challenges:

- Long life-times of hyperons gives rise to displaced vertices
- Charged particles can be created outside the MVD

Development of tracking tool for the straw tube tracker:

- Reconstruct momentum of tracks from hyperon decay particles
- Heavily object-oriented C++ implementation, to be part of the PANDARoot framework
- High performance essential for large scale data processing



LIA LAVEZZI

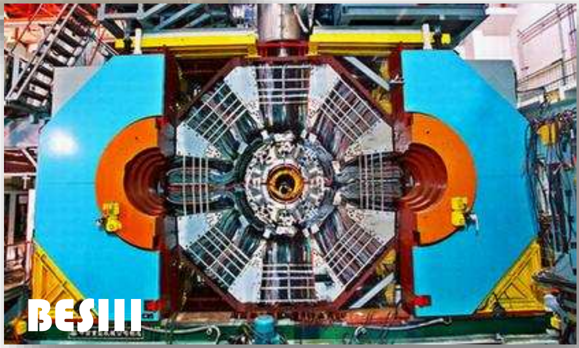
❖ Degree and Ph.D. in physics
@ **University of Pavia**

❖ Post Doc position (till the end of the week)
@ **University of Torino**

❖ Member of the **$\bar{\text{P}}$ANDA** collaboration since 2004,
- Computing group
- Tracking group

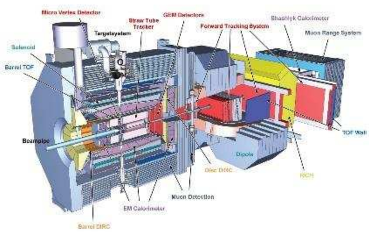
❖ Member of the **BESIII** collaboration since 2015

MAIN ACTIVITIES:
- ❖ Monte Carlo simulations:
  - ❖ geant3, geant4, Virtual Monte Carlo
  - ❖ ROOT
- ❖ Charged track reconstruction:
  - ❖ track finding
  - ❖ track fitting

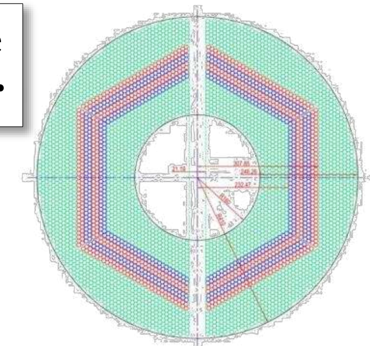


PANDA

@ FAIR, GSI, Darmstadt
$p_{BAR}p$ annihilation
$2.25 < \sqrt{s} < 5.47$ GeV

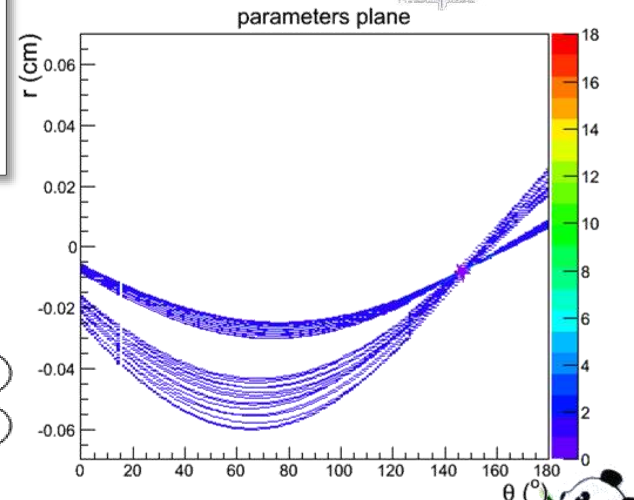


Simulation & reconstruction of the **Straw Tube Tracker**



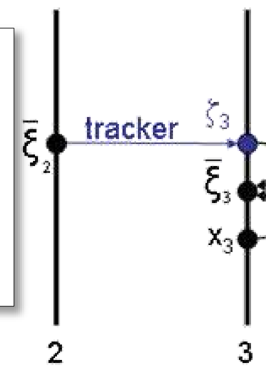
Track finding

- add hits from GEM to tracks from PR
- PR for secondary tracks, from decay vertices of long living neutral particles



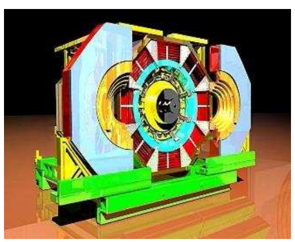
Track fitting

Kalman filter: iterative procedure which needs a track follower to propagate the state vector & cov matrix from a plane to another → GEANE



BESIII

@ BEPCII, Beijing
e^+e^- annihilation
$2 < \sqrt{s} < 4.6$ GeV



Cylindrical GEM

will substitute the existing drift chambers:

- test beam data analysis
- tracking algorithms

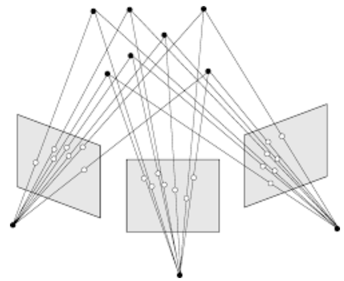
My work has just begun!

Pattern recognition is a topic where speed is crucial and parallelization may help a lot!

Alberto Madonna

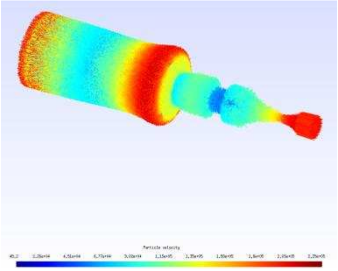
alberto.madonna@pd.infn.it

Background

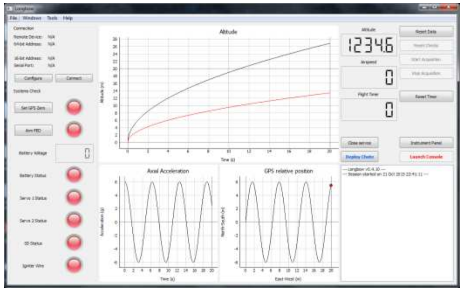


B. Eng.: Sparse Bundle Adjustment

GPU Computing



M.Eng.: Plasma Thruster Simulation



PyQt control platform for sounding rocket

Alberto Madonna

alberto.madonna@pd.infn.it

Current Research

Data Reduction and Analysis Pipeline for the ASTRI Prototype IACT

- Targets datacenter and low-power platforms
- Written in C++, runs on x86, ARM and GPU
- Exceeds predicted real-time requirements
- My tasks:
  - ➢ ADC Calibration
  - ➢ Instrument Response Functions
  - ➢ Code parallelization

Numerical Relativity

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu}$$

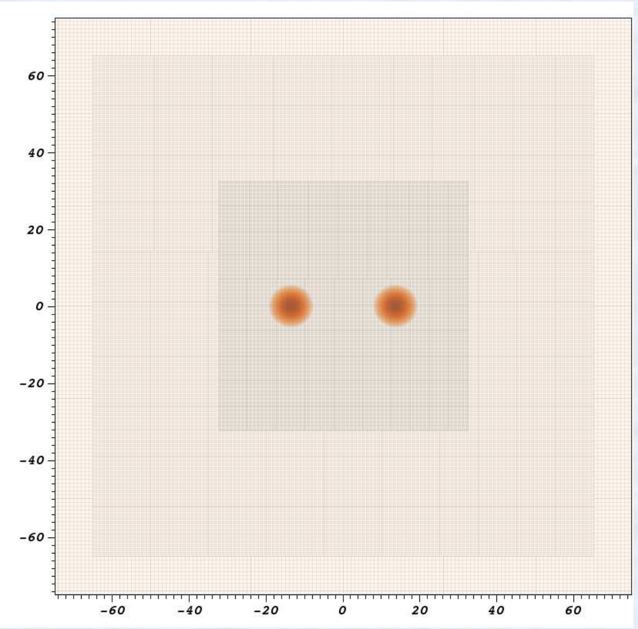
$$\nabla_\mu T^{\mu\nu} = 0$$

$$\nabla_\mu (\rho u^\mu) = 0$$

3+1 Einstein equations + (magneto)hydrodynamics discretized on a 3D mesh and evolved in time.

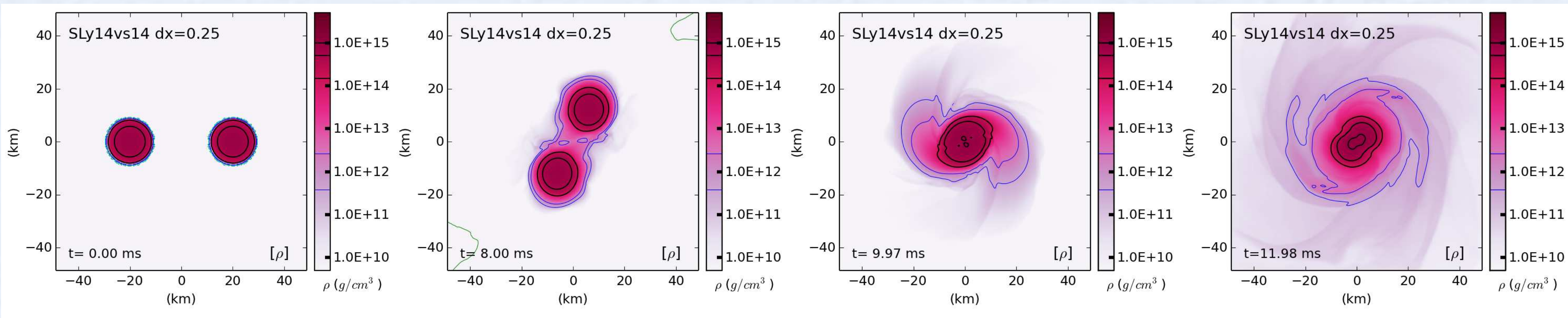
At least 24 spacetime+gauge variables
plus (8)5 (magneto)hydrodynamical variables
are stored and evolved at each grid point.
(+fluxes computation +signals extraction +buffer zones +ghost zones...)

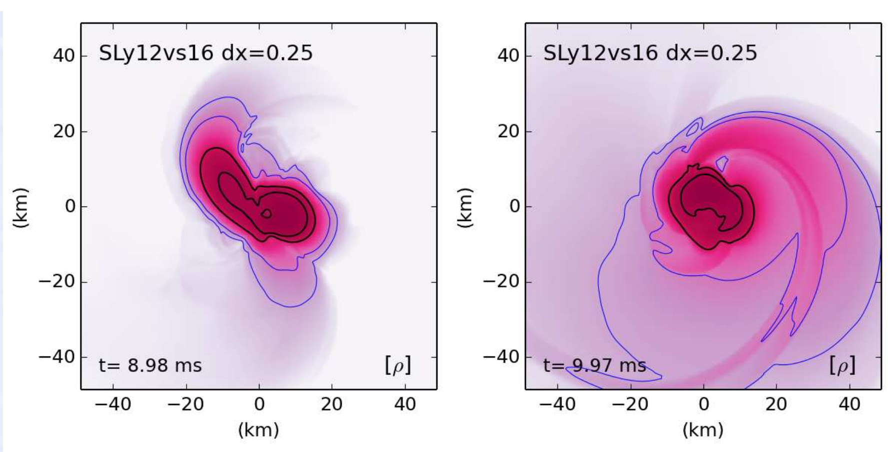
- (Adaptive) Mesh Refinement
  [ex. 6 levels of 120^3 points each with refinement factor 2]
- Code parallelization
  [MPI domain decomposition + OpenMP multithreading]
- Code running on parallel supercomputers for days!
  [Cineca Galileo cluster up to 8 nodes (256 cores)]
  [Cineca Fermi cluster up to 512 nodes (8192 cores)]

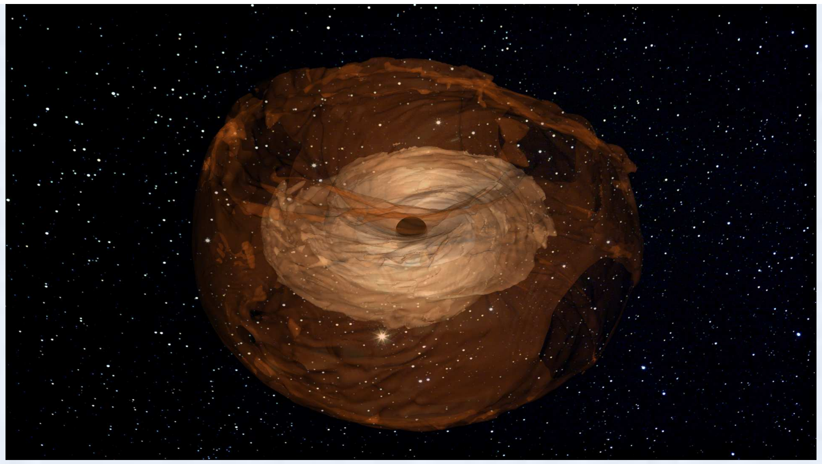


Binary neutron star mergers

[From: R. De Pietri, A. Feo , **F. Maione** and F. Löffler,
Modeling Equal and Unequal mass Binary Neutron Star Mergers Using Public Codes, arXiv:1509.08804]







- ➢Everything done using only public, open source software!

- ✓ The Einstein Toolkit (used for the dynamical evolution).
- ✓ The Lorene library (for initial data generation).
- ✓ Our Open Source PyCactus post-processing routines (based on Python-numpy-matplotlib).