Research Doctoral programs PON for the academic year 2021/2022 - XXXVII cycle

Theme title	Development of solutions for the generation of efficient FPGA-based computing architectures
Project title	Integration of Neural Networks on FPGA

Giulio Bianchini

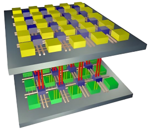
Year	II
Tutors	Mirko Mariotti, Daniele Spiga

5/10/2023, Bertinoro

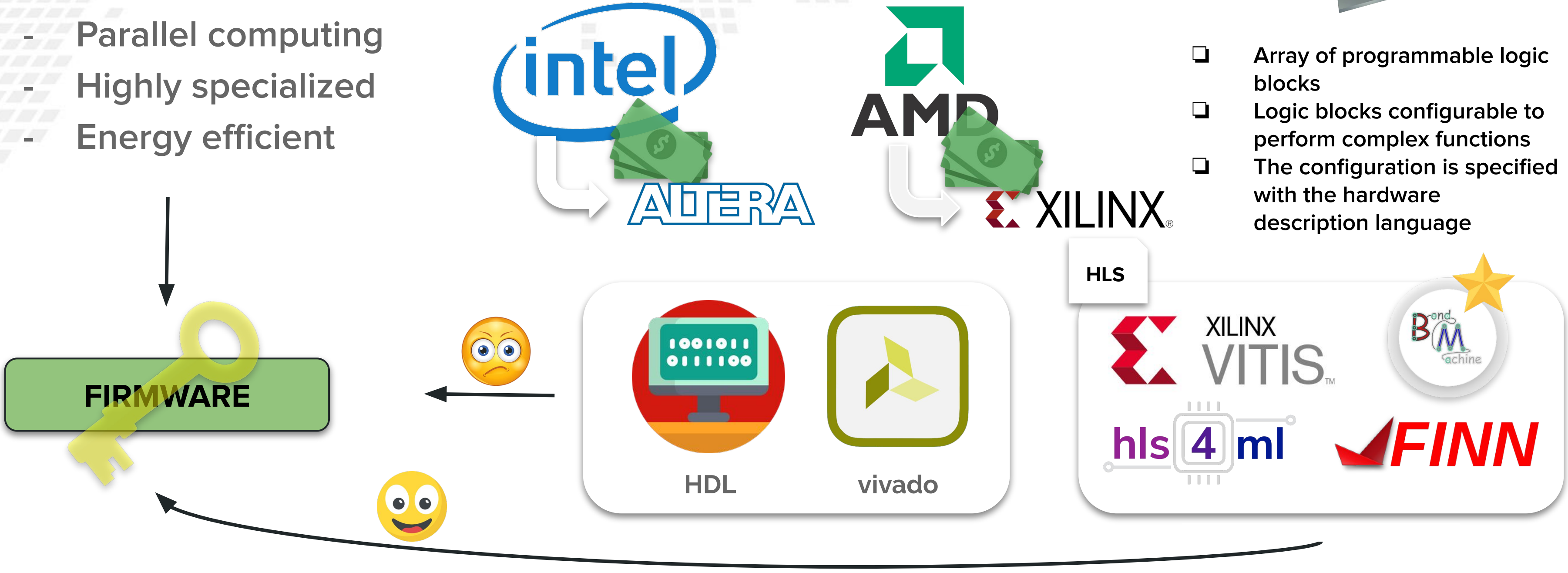
FPGA

A Field Programmable Gate Array (FPGA) is an integrated circuit whose logic is re-programmable.

- Parallel computing
- Highly specialized
- Energy efficient



- ❑ Array of programmable logic blocks
- ❑ Logic blocks configurable to perform complex functions
- ❑ The configuration is specified with the hardware description language



http://bondmachine.fisica.unipg.it/

BondMachine

A framework to build dynamical computer architectures

The BondMachine is an open source (https://github.com/BondMachineHQ) software ecosystem for the dynamical generation of computer architectures that can be synthesized on FPGAs.

- High level programming language (Golang) for both the hardware and software
- Functional style programming
- Architecture generating compiler
- Computational graph and Machine Learning Models

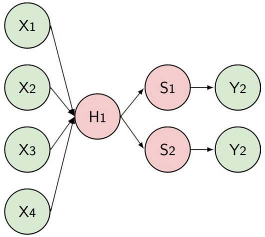
Article published on Parallel Computing, Elsevier 2022 DOI:10.22323/1.351.0020

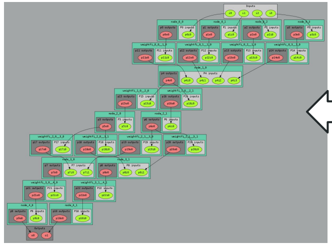
Research activity

Led to

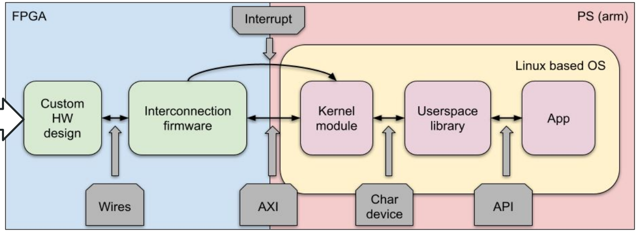
Firmware development

Low latency machine learning inference on FPGA





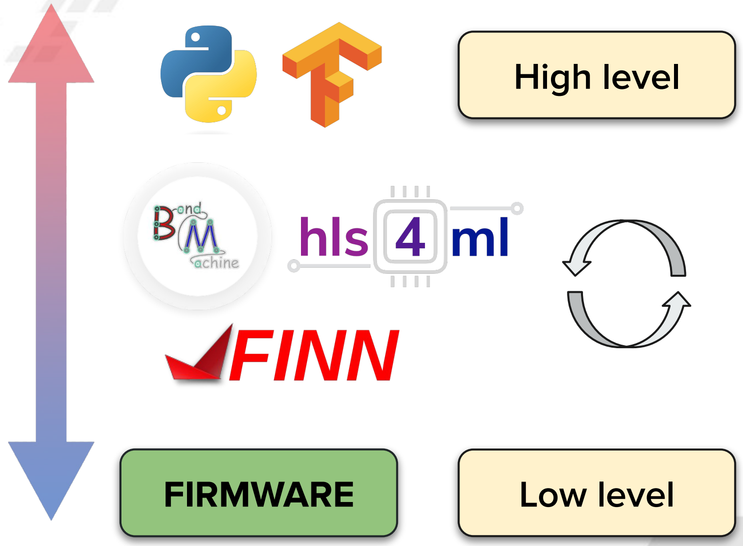
Development of accelerated systems on hybrid processors



Build firmware for low latency machine learning inference on FPGA

? Implement customized and parallel architectures tailored to specific ML models, resulting in faster processing speeds and lower power consumption compared to traditional CPUs or GPUs.

Starting from high-level code and standard ML framework, with HLS tools like HLS4ML or BondMachine, get the firmware implementations of machine learning algorithms



Why BondMachine?

The machine learning model is trained with standard frameworks and synthesized in FPGA as a graph of heterogeneous and interconnected processors.

- Optimized resource usage
- Highly customizable
- Available at a high level
- Vendor independent
- User-friendly
- Cloud extension

We have integrated this solution with Inference-as-a-Service cloud native systems

ML Inference on cloud

Integrate our system with cloud-native inference as a service system

Ease of usage and flexibility

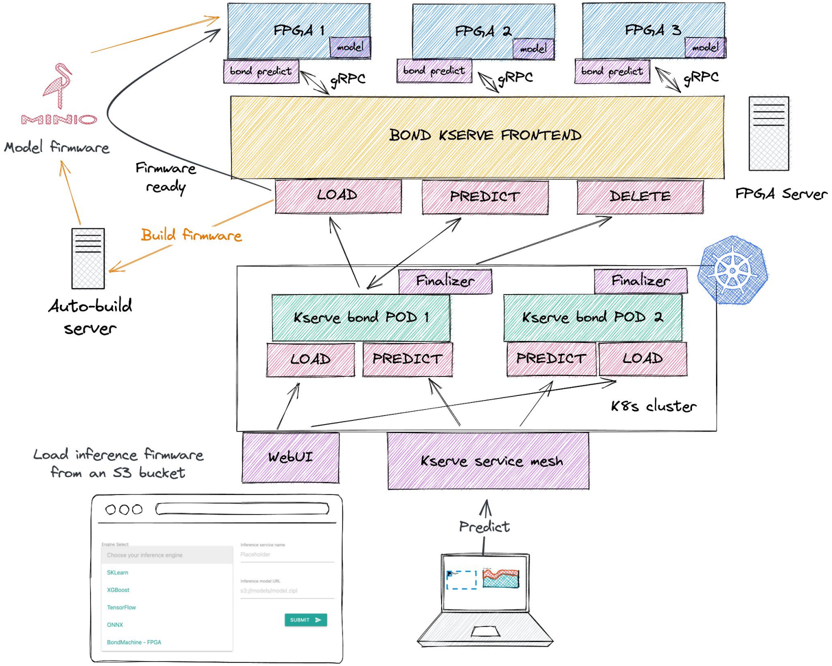
Being able to deploy an inference algorithm on FPGA without caring for "where" the resources are

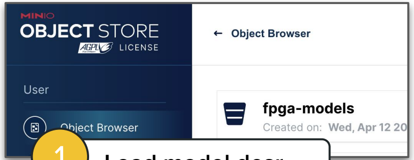
Democratic access and management

Leveraging cloud/k8s native tools, you can reuse a well established way to orchestrate the bookkeeping and distribution of the payloads

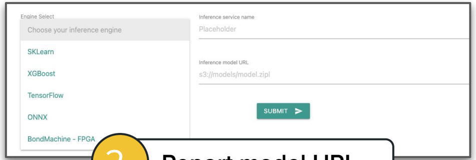
Easy Prototyping

Automation of the build and load process -> the framework take care of vendor specific details

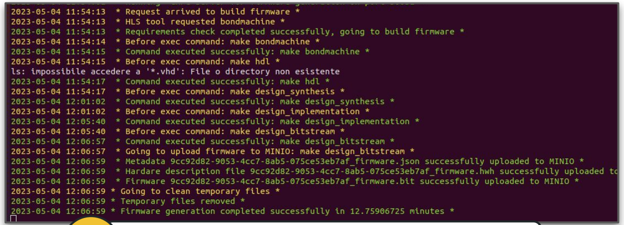




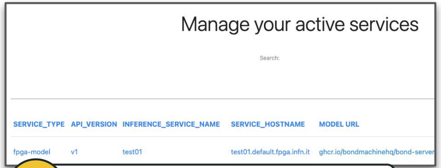
1 Load model desr.



2 Report model URL



3 Wait build of fw and load on FPGA



4 Get endpoint for predictions

5 ML inference

Thanks for the attention!

- ★ Website http://bondmachine.fisica.unipg.it/
- ★ Paper https://www.sciencedirect.com/science/article/pii/S0167819121001150
- ★ CHEP 2023 https://indico.jlab.org/event/459/contributions/11826/
- ★ CCR 2023 https://agenda.infn.it/event/34683/contributions/197368/
- ★ CCR 2022 https://agenda.infn.it/event/30202/contributions/168531/
- ★ InnovateFPGA 2018 Iron Award https://github.com/innovatefpga/2018-EM083

- ★ Main repo https://github.com/BondMachineHQ
- ★ ML inference on cloud repo https://github.com/BondMachineHQ/kserve-bond-extension.git

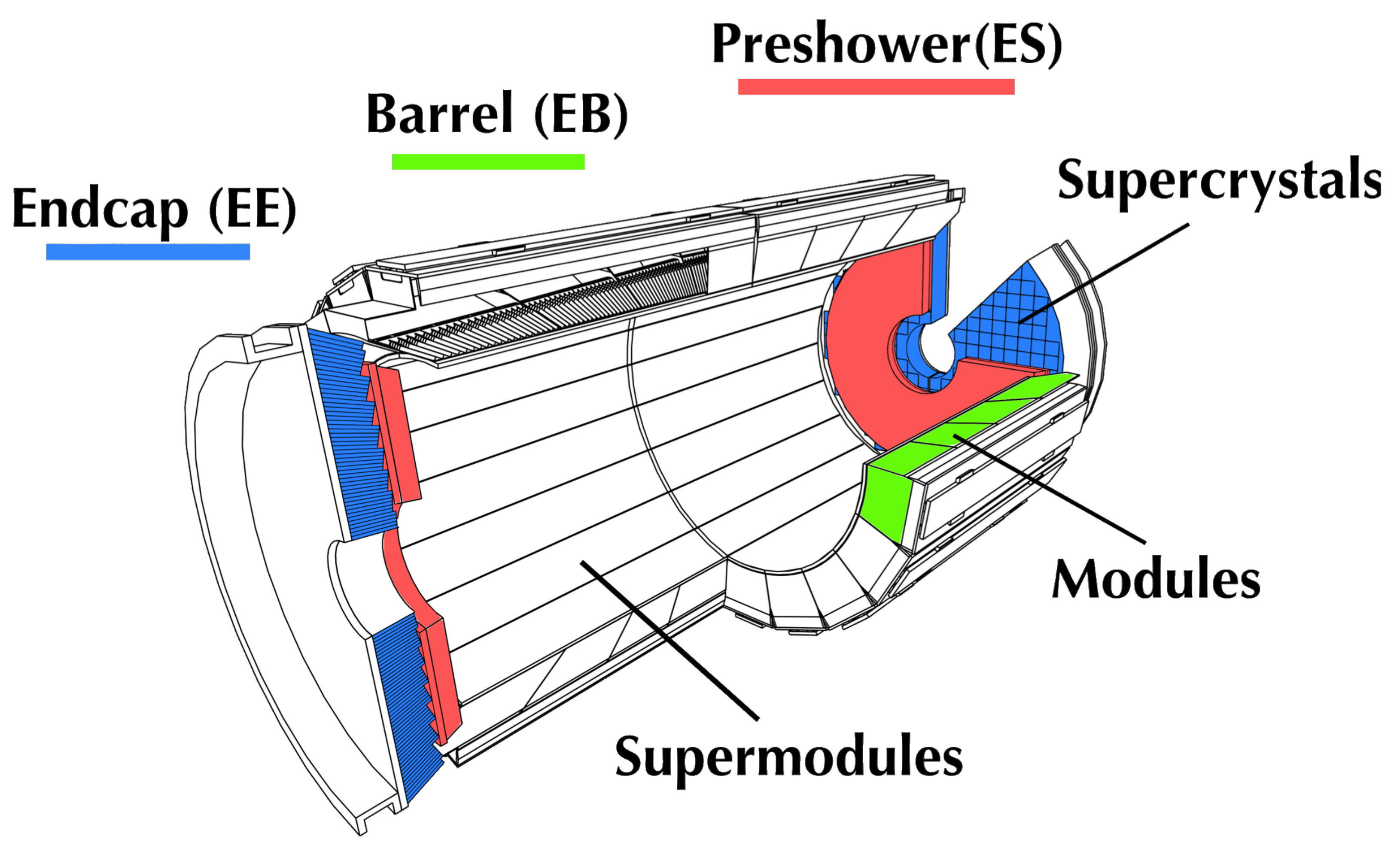
CMS ECAL Barrel Upgrade for HL-LHC

Lightning Presentation

XIV edition of ESC School on Efficient Scientific Computing (ESC23)

4–12 October 2023, Bertinoro (Italy)

CMS ECAL Barrel Upgrade



CMS EM Calorimeter (ECAL)

∗ 61200 lead tungstate (PWO_4) crystals in the barrel
∗ Fast decay scintillation light (25 ns)
∗ Small Molière radius (2.2 cm)
∗ Read by a pair of avalanche photo-diodes (APDs)

HL-LHC upgrade challenges for the ECAL barrel

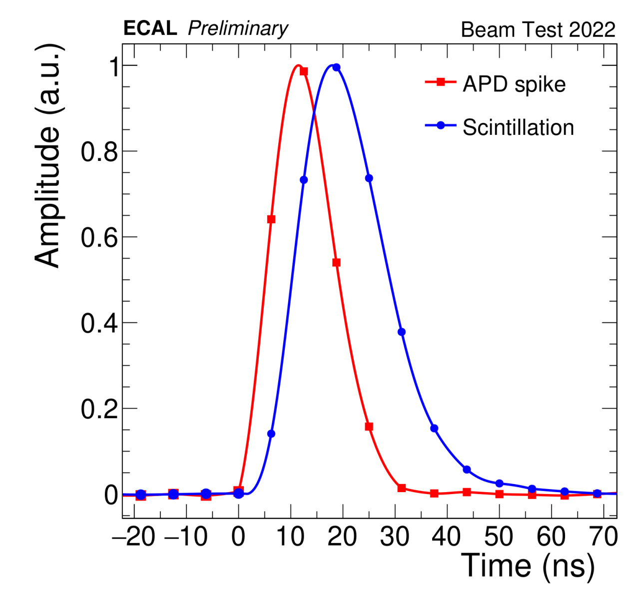
∗ Increased **pileup from 40~60 up to 200**

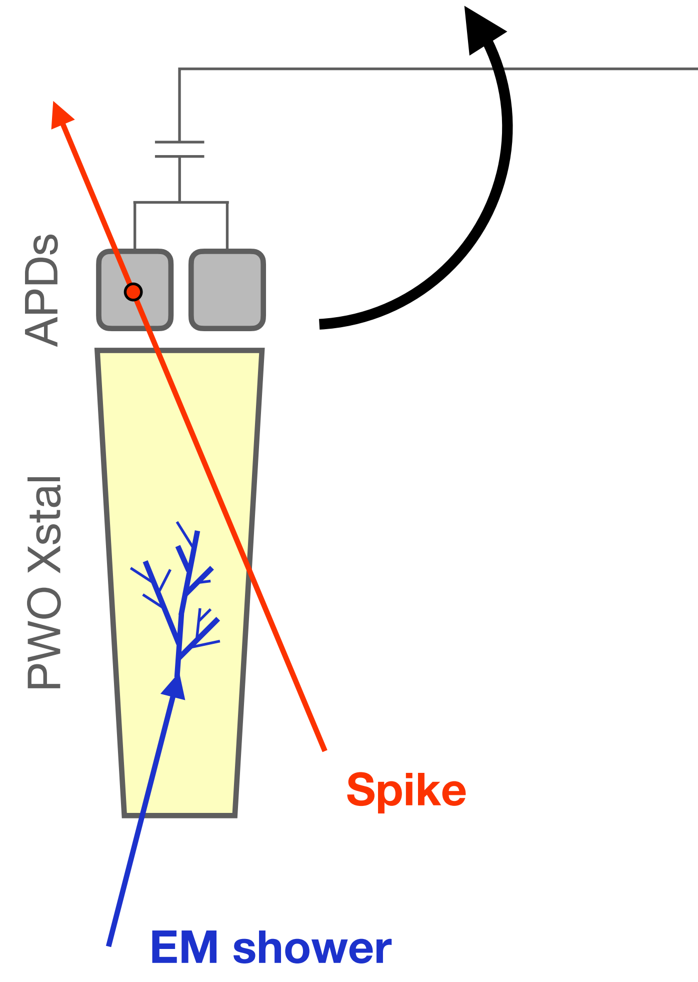
∗ **Increased noise** caused by detector and sensor **ageing**

Faster front-end (FE) electronics to

∗ **Improve time resolution** for better primary vertex reconstruction -> <30 ps for E>50 GeV

∗ **Discern scintillation signals from APD spikes** (signals from direct ionization of the APDs)

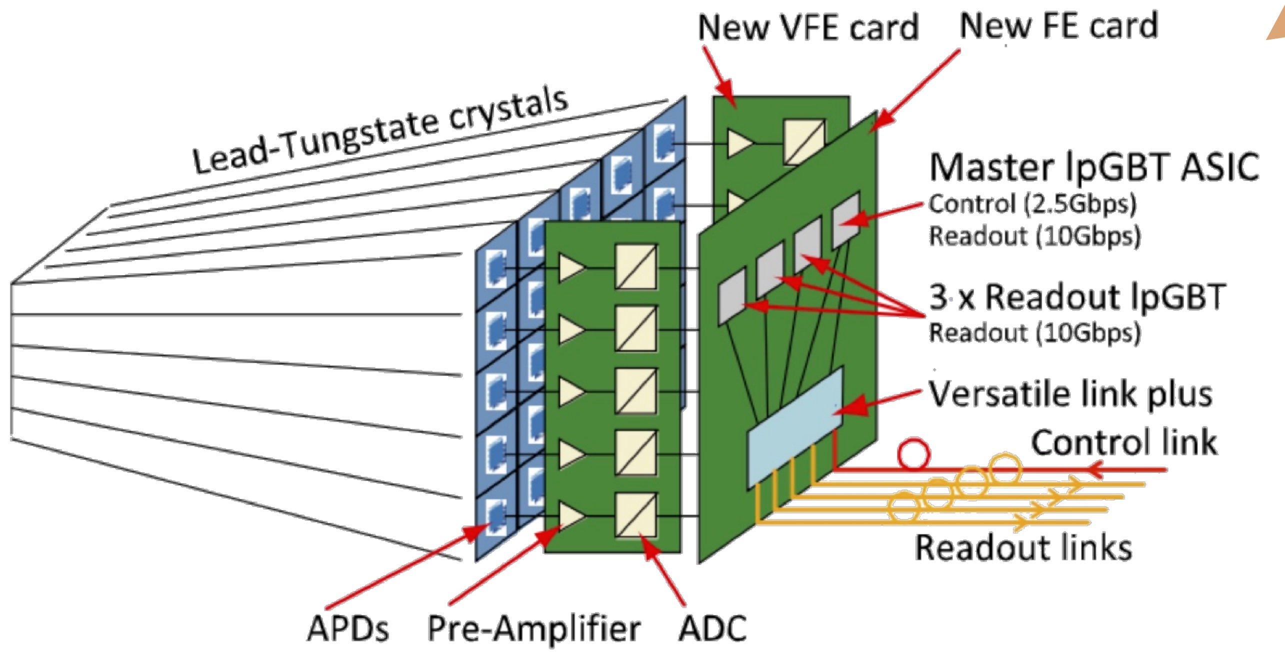




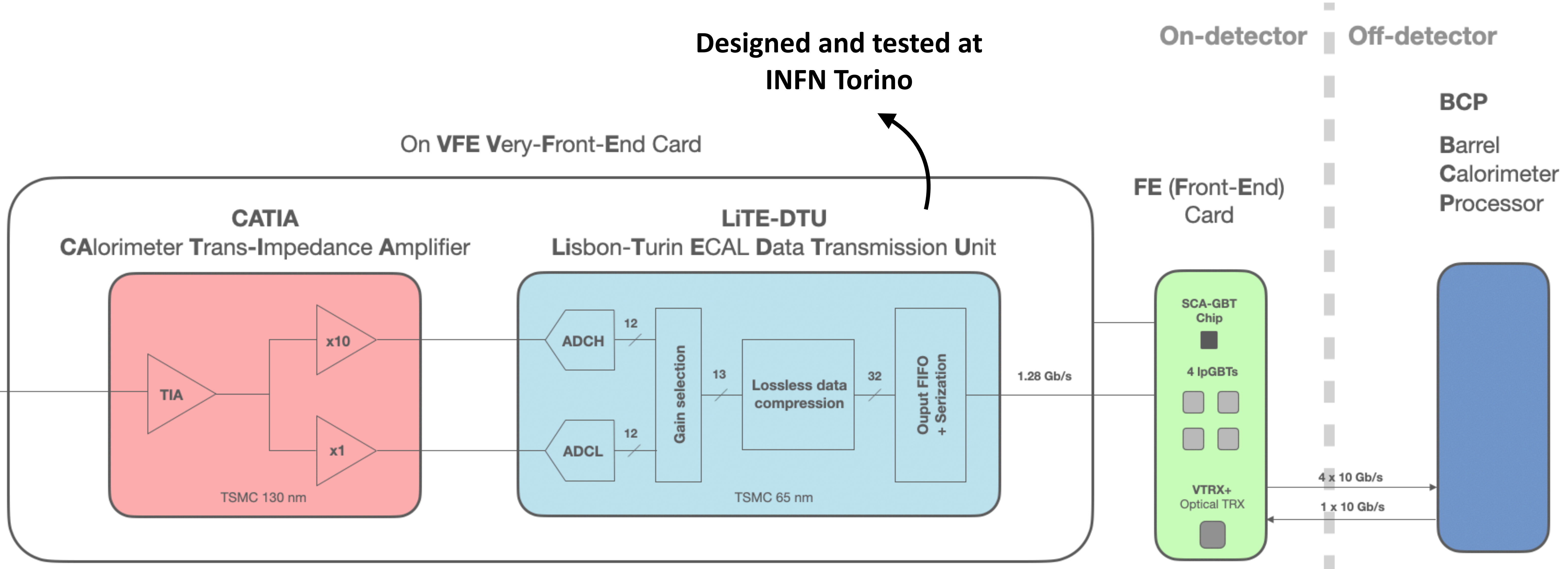
CMS ECAL Readout Upgrade

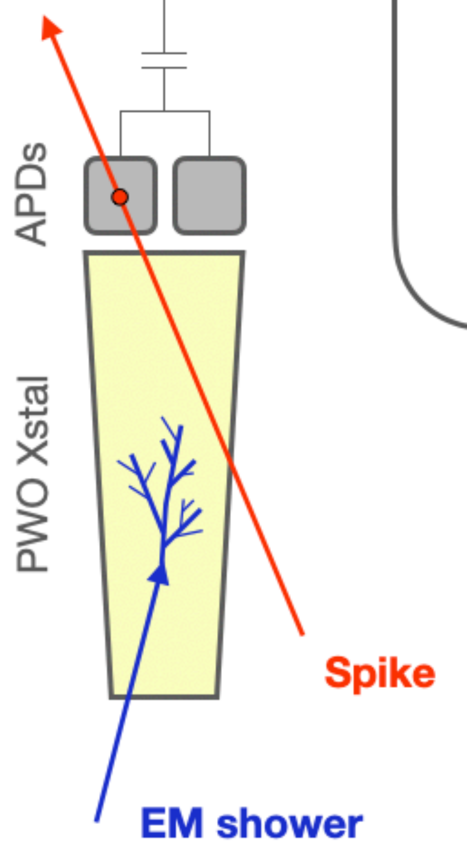
New back-end to to cope with new CMS trigger and DAQ requirements

* Rate: 100 -> 750 kHz
* Latency: 4.5 -> 12.5 us



The new front-end readout unit for HL-LHC





Unchanged in HL-LHC:

* Geometry
* Crystals
* APDs

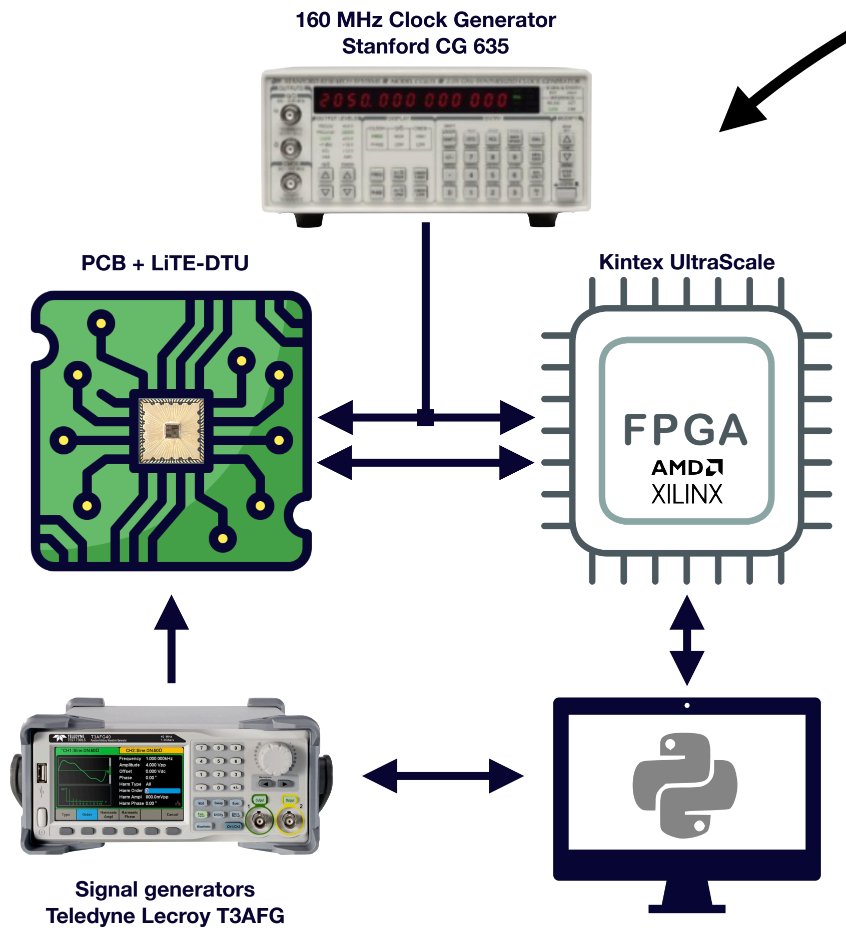
New faster dual-gain amplifier to cover a dynamic range from 50 MeV to 2 TeV

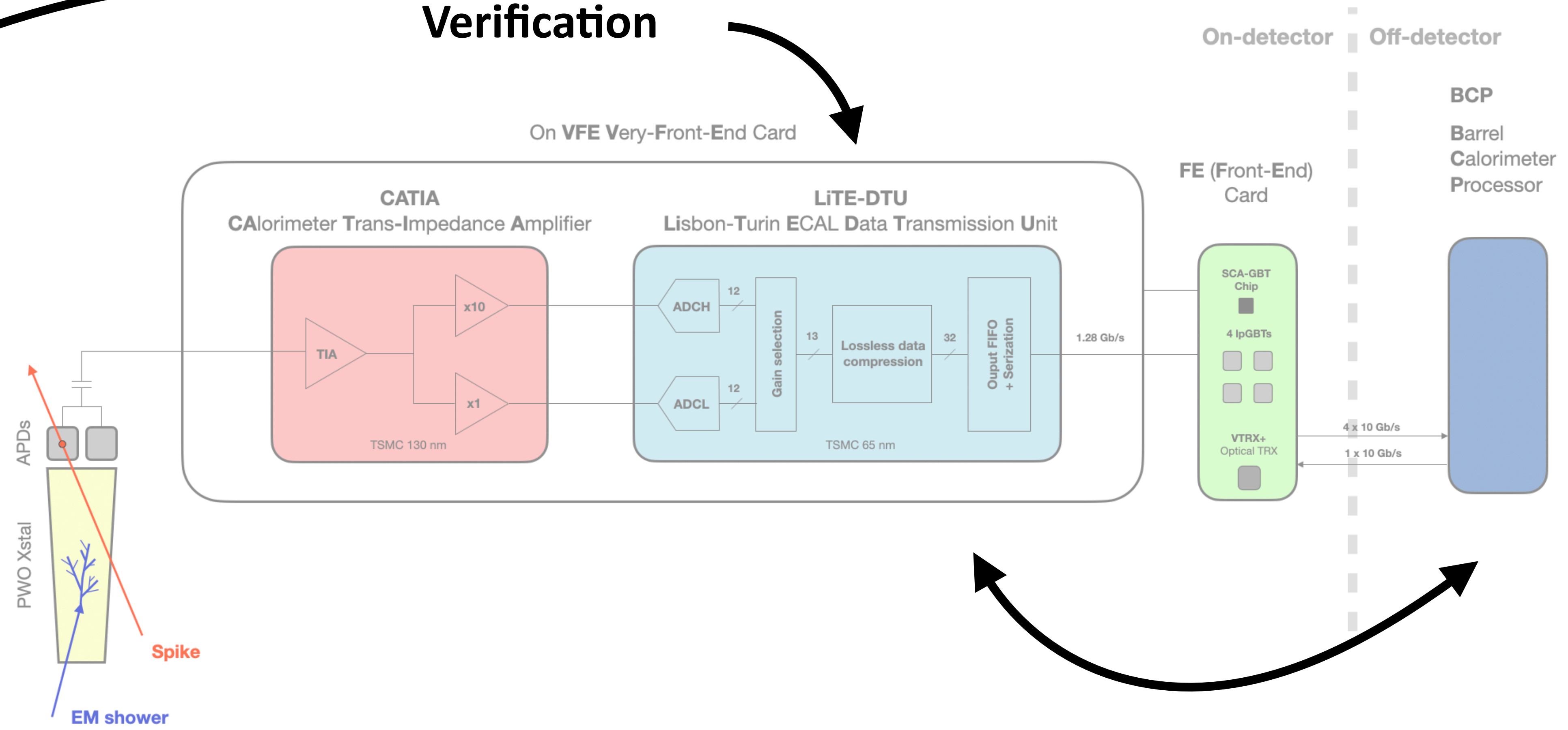
New 12-bit 160 MHz (legacy => 40 MHz) A/D conversion and digital processing ASIC to perfom gain selection, lossless data compression and serilisation

The new front-end will be triggerless: the trigger primitives will be generated in the BCP

My activities in the CMS ECAL Upgrade

(1) LiTE-DTU ASIC Verification

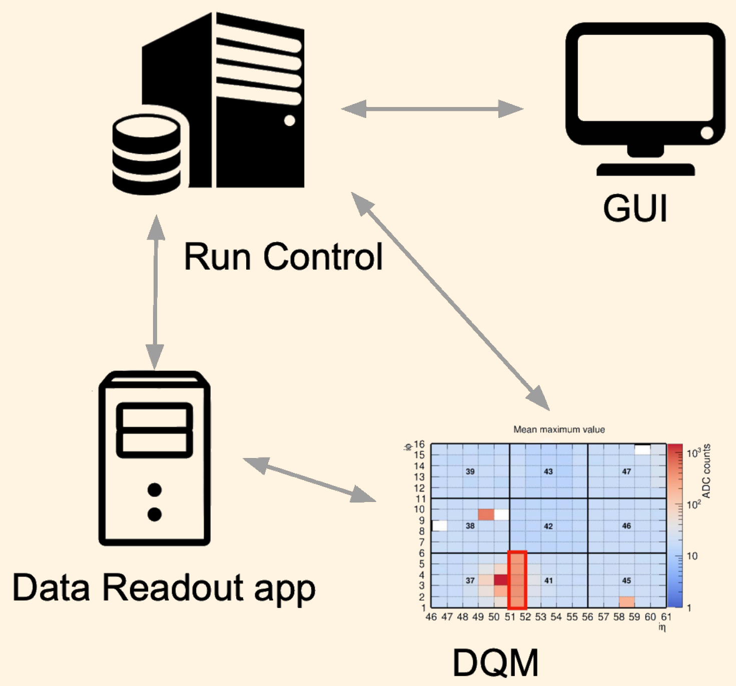




(2) System Test and Debug

(3) Beam Tests Preparation and Data Analysis

* Last successful test beam campaign at **CERN H4 facility in July 2023** with the full system (previous in 2022, 2021)
* Tested **200 channels** of a supermodule read by the full readout chain (front-end to back-end)
* I developed a data readout application to integrate the controller of the BCP (written in Python) with the existing C++ DAQ software



* Automatic on-bench verification software development
* The test sequence lasts **30/40 s** and the same Python-based software will be used to test the final production (~90k chips) by the testing company

* To make sure that the single devices work correctly in the system
* To make sure that the system is stable
* **To spot critical points before the production**

❗ Larger scale ==> efficiency is increasingly important

. . . for the future => integration into the CMS certral DAQ
=> large scale commissioning tests

A uniform and heterogeneous global event reconstruction at the CMS experiment

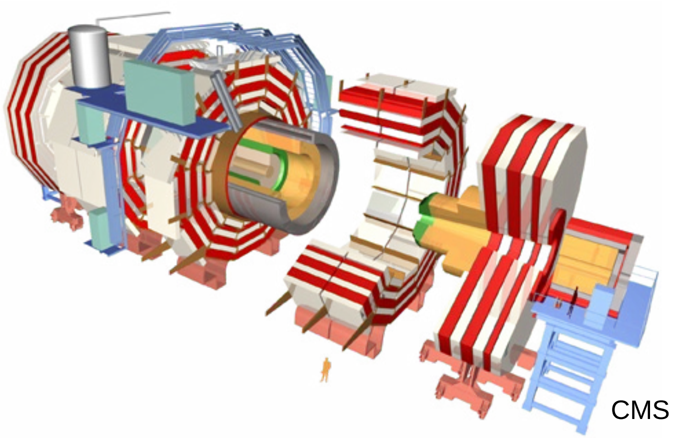
Alessandro Brusamolino, Karlsruhe Institute of Technology

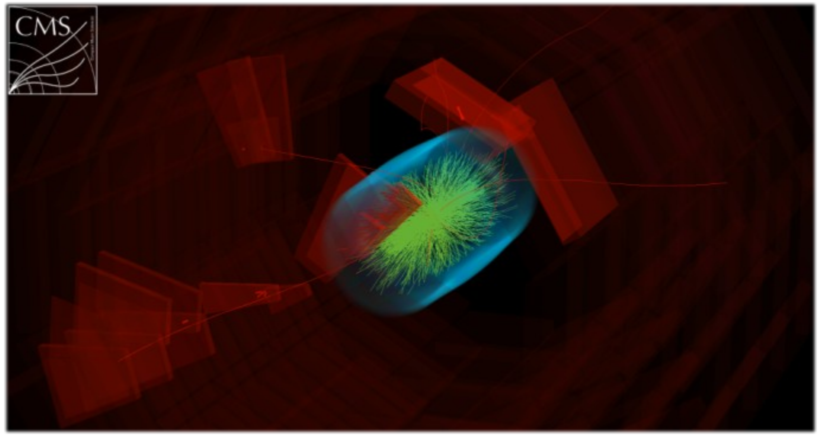
Who am I?

- Master’s degree in Physics at the University of Milano-Bicocca
  - **Data reduction techniques** for CMS Run 3
  - Started working on event reconstruction
- Currently PhD student at the Karlsruhe Institute of Technology
  - Working on the **CMS global event reconstruction** for Phase 2

The CMS Experiment for HL-LHC

- General purpose detector
  - **Tracker** → track charged particles
  - **MTD** → time of arrival of charged particles
  - **ECAL, HCAL, HGCAL** → energy measurement
  - **Superconducting magnet** → 3.8 T magnetic field
  - **Muon chambers** → identify muons
- ~ 40 MHz collision rate
  - **Level-1 Trigger** → 750 kHz
  - **HLT** → 7.5 kHz
- Up to 200 simultaneous collisions (**pileup**)

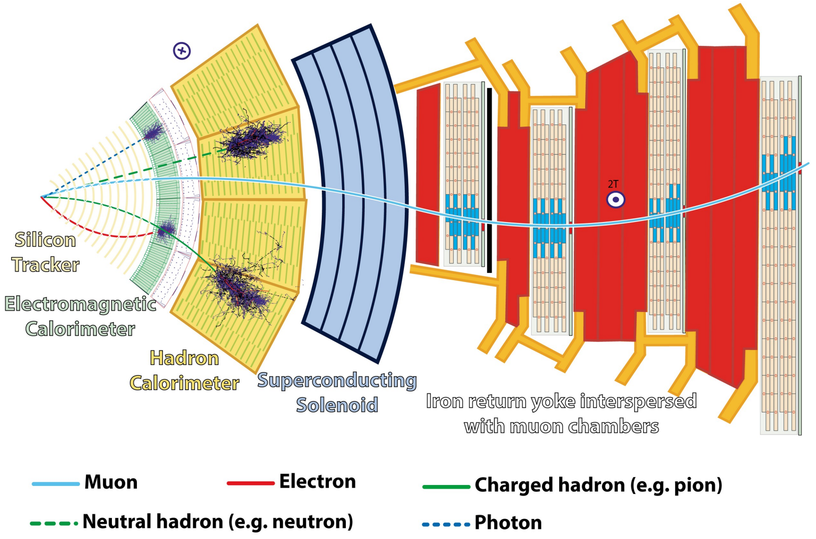

CMS detector until Run 3

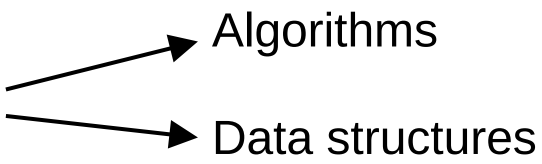


Global event reconstruction

- Combine information from the different parts of the sub-detectors
  - Identify final state particles
  - Reconstruct particle properties
- The current global reconstruction is **Particle Flow**
- **High pileup scenario during Phase 2**
  - Additional contributions from particles coming from pileup interactions

    → performance are degraded
  - Novel algorithms developed for Phase 2
    - **TICL (The Iterative CLustering)** developed for HGCAL
      - Start from hits in the calorimeter and return particle properties and identification probabilities
      - Developed with heterogeneous computing in mind

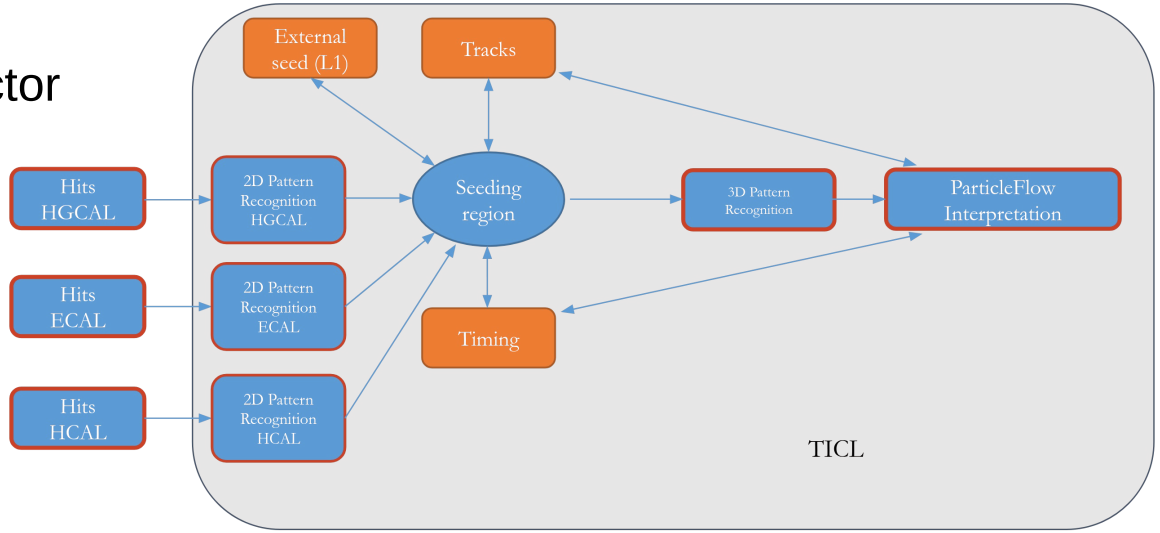




Extending TICL to the Barrel region

- Goal is to have a uniform event reconstruction across the entire detector
- TICL is easily generalizable to be run on different detectors

  → **extension to ECAL and HCAL**
- Reconstruction would benefit from such an approach
  - Exploiting other steps of the reconstruction offloaded to GPUs
  - Use same data structures and algorithms
  - Good performances in critical region of the detector



Alessandra Casale

Tecnologist

INFN LNGS Computing and Networking Systems Services

INFN
LNGS
Istituto Nazionale di Fisica Nucleare
Laboratori Nazionali del Gran Sasso

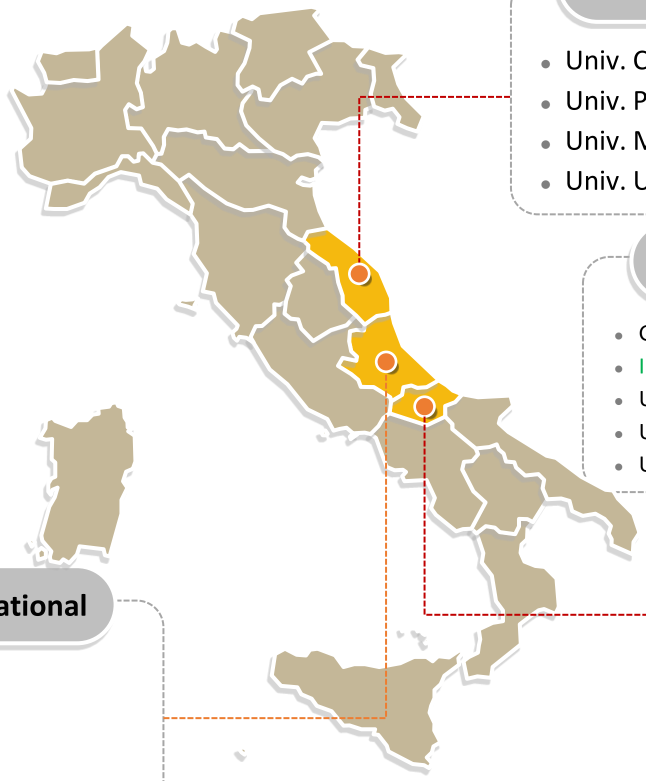
Education and career history

- Master's degree in Nuclear Physics (University of Genova, 2011)
- Ph.D. in Physics Underwater Acoustics (University of Genova, 2017)
- Patent submission, optical hydrophone (granted in 2020)
- Software developer (supply chain, machine learning) +8 years
- User Experience Design course (POLI.design Politecnico Milano, 2021)
- Tecnologist at INFN Laboratorio Nazionale del Gran Sasso since July 2023

My current activity

HPC4DR project

Started in 2020, involving the universities and research centers of the Abruzzo, Marche and Molise regions, hard hit by earthquakes and other catastrophic events in 2016/2017.



MARCHE

- Univ. Camerino
- Univ. Politecnica Marche
- Univ. Macerata
- Univ. Urbino

ABRUZZO

- GSSI
- INFN-LNGS: HPC Infrastructure
- Univ. Chieti Pescara
- Univ. L'Aquila
- Univ. Teramo

MOLISE

- Univ. Molise

Local vs. National

- INGV
- INFN – LNGS
- INAF
- CNR – ITC

Reduction of risks connected to disasters due to natural and human-made phenomena

High-performance computing (HPC) technological infrastructure

My current activity

HPC infrastructure @LNGS

On January 27, 2022, the agreement was signed between INFN and CINECA for the free transfer of:

- n. 5 GALILEO racks each containing 72 nodes
- n. 1 GALILEO rack containing 36 calculation nodes
- n. 1 GALILEO rack containing spare parts

The computing infrastructure is Lenovo NextScale
Each rack holds up to 6 Lenovo NeXtScale n1200 Enclosures (6RU)
Each enclosure contains 12 NeXtScale nx360 M5 Compute Nodes

- 2*Intel Xeon E5-2697 v4 @ 2.30GHz 18-core each (Broadwell)
- 128GB RAM/node, 3.5GB RAM/core

5 rack * 72 nodes, 1 rack * 36 nodes -> 396 nodes, 14256 cores

- Each server has a peak computing power of 1.3Tflop/s.
- The total computing power is approximately 0.5PFlops
- The compute nodes are interconnected by a 100Gb/s Intel Omnipath network and a 1Gb/s Ethernet network

- n. 1 node with 8 GPU from GSSI (Gran Sasso Science Institute)

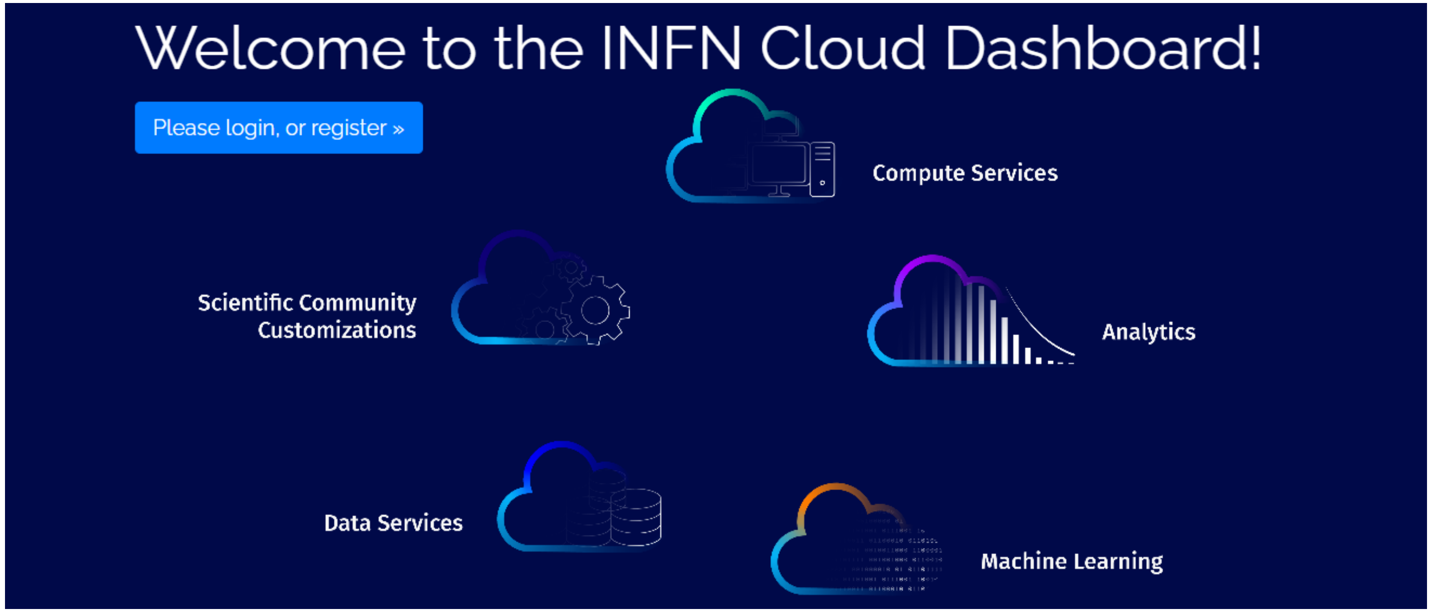
It's time to deepen my knowledge in parallel computing for testing and user support ☺

My current activity

INFN CLOUD project

The INFN Cloud is an internal project which aims to

- manage a (large) fraction of the INFN resources in a sustainable and optimized way;
- make different INFN communities able to access resources, regardless of the availability of local and dedicated hardware



The federation of the LNGS infrastructure will allow:

- To expand the INFN Cloud backbone (BARI, CNAF) with a scalable set of satellite sites,
- To share with the INFN community a portion of LNGS-HPC resources

Study of matter-antimatter asymmetry and strangeness equilibration at the LHC with ALICE

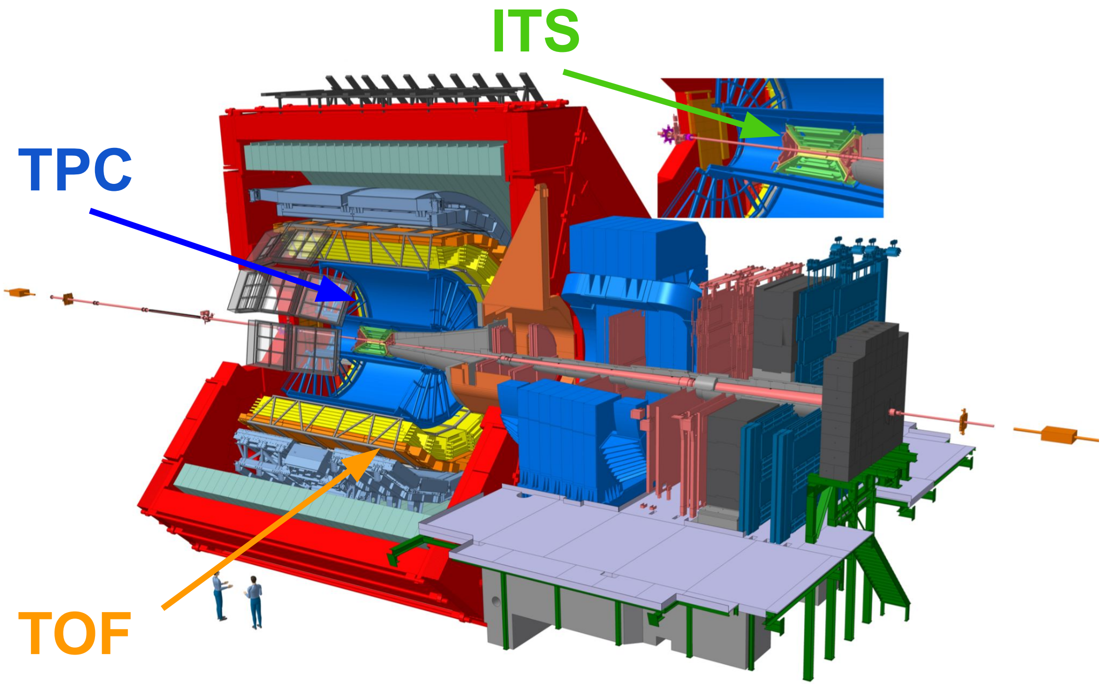
Mario CIACCO
PhD Student at Politecnico and INFN, Torino

ESC 2023 - Bertinoro (FC)

mario.ciacco@cern.ch

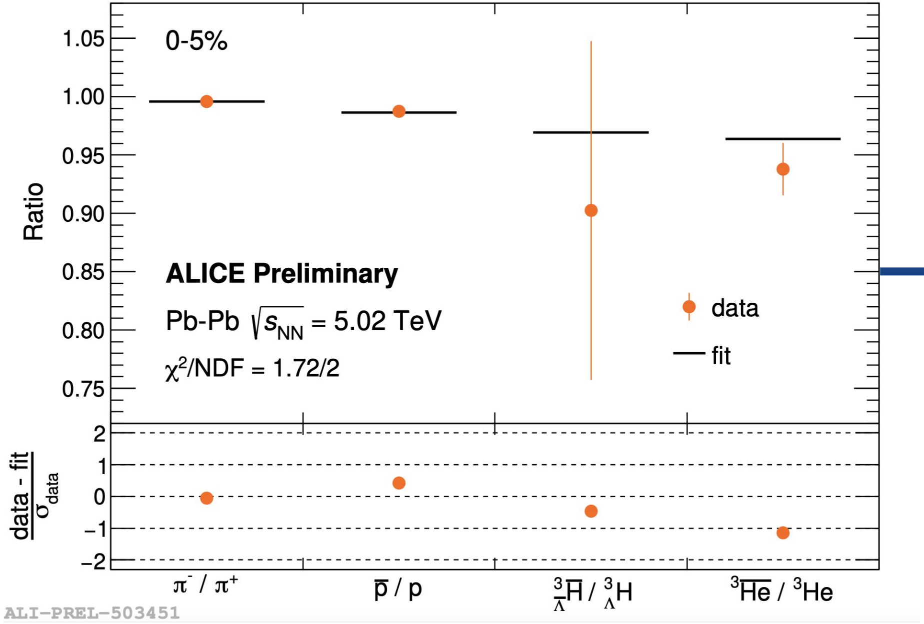
Matter-antimatter asymmetry at the LHC

- **LHC → antimatter factory**
  - Matter created in Pb–Pb collisions at TeV scale → (almost) exact equilibration of matter and antimatter
  - Similar to the conditions of the early Universe according to standard cosmology
- How close to perfect symmetry?
  - Measure the **antimatter-matter balance**
  - **ALICE** → outstanding particle identification performance

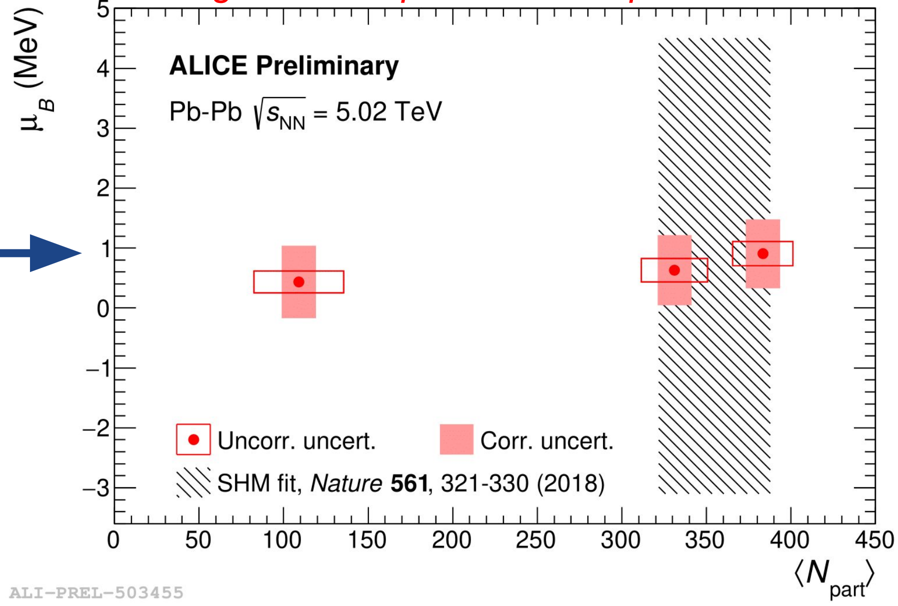


Matter-antimatter asymmetry at the LHC

- **Antimatter/matter yield ratios** measured for different identified species
  - Protons and light nuclei → sensitive to the baryonic asymmetry of the system
  - **Baryon chemical potential μ_B** → global variable providing the net-baryon content



significant improvement of precision

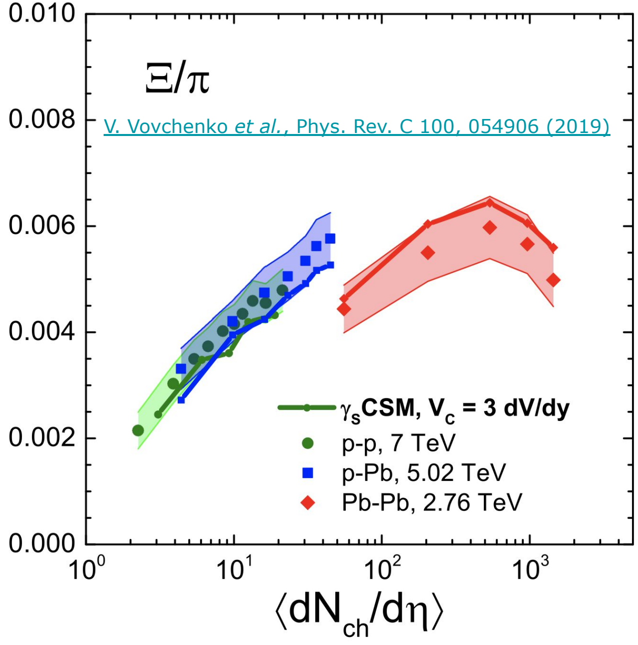


fits based on statistical hadronization model

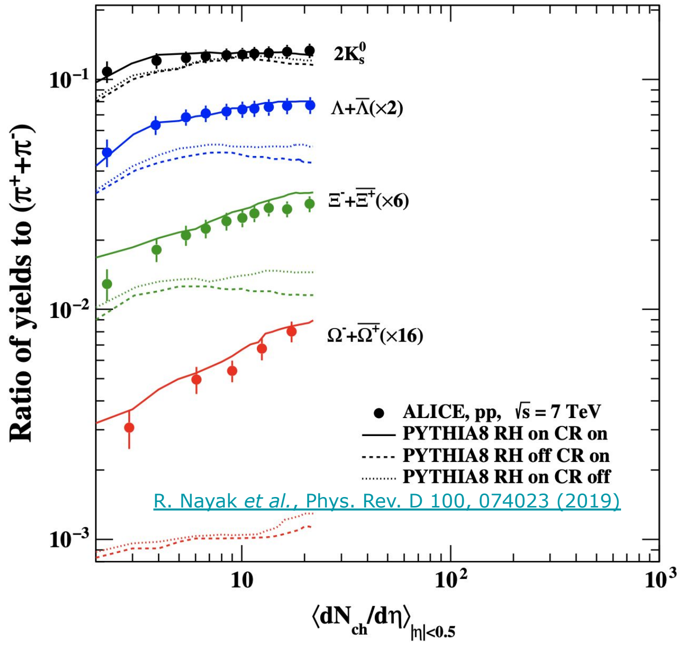
Strangeness production at the LHC

- Enhanced strangeness production in nuclear collision → bulk of thermalised partons
- Similar effects observed also in hadronic collisions
  - Does **strangeness equilibrate** also **in small collision systems?**

Canonical statistical hadronisation model



Lund string fragmentation (PYTHIA)



Strangeness conservation in (non-)thermal models

Canonical statistical model

- Charge conservation over a volume
- Symmetric correlation of **same-** and **opposite-charge** hadron pairs
- Large rapidity correlation for strangeness conservation

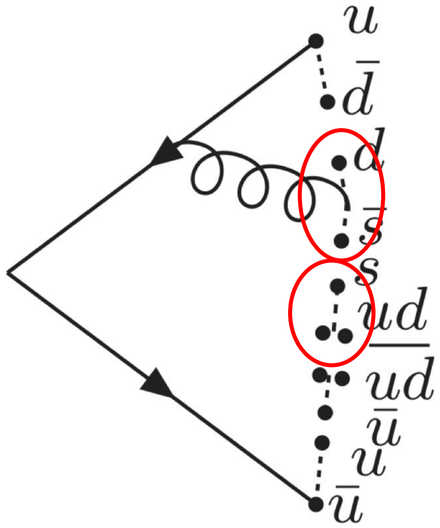
$$\mathcal{Z}(B,Q,S) = \int_{-\pi}^{\pi} \frac{d\phi_B}{2\pi} \int_{-\pi}^{\pi} \frac{d\phi_Q}{2\pi} \int_{-\pi}^{\pi} \frac{d\phi_S}{2\pi}\, e^{-i\,(B\phi_B + Q\phi_Q + S\phi_S)} \times \exp\left[\sum_j z_j^1\, e^{i\,(B_j\phi_B + Q_j\phi_Q + S_j\phi_S)}\right].$$

Conserved exactly over **correlation volume V_c**

V. Vovchenko *et al.* Phys. Lett. B 785, 171 (2018)

Lund string fragmentation (PYTHIA)

- **Local** conservation of charges
- Correlated production of **opposite** charges
- Small rapidity correlation for strangeness conservation

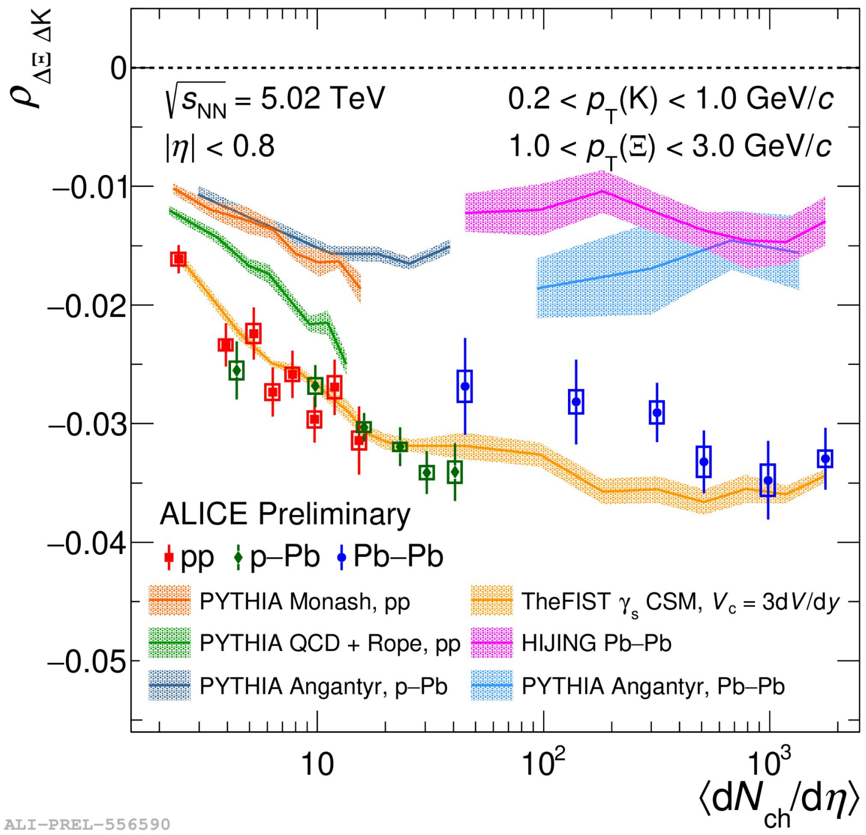


→ Can be explored via **event-by-event measurements!**

Net-Ξ net-kaon number correlation

- Same- and opposite-sign correlations → **2 different species** → **charged kaons** and **Ξ**
  - Negligible effect of heavy resonance decays
- Correlation of net-particle number
  - Contains effects of both charge pairings
  - "How much does ($\Xi^+ - \Xi^-$) number difference influence ($K^+ - K^-$) number difference due to strangeness conservation?"
- The results obtained can be described by **thermal models**
- **String fragmentation**, despite matching the measured yields, **fails to describe higher-order moments**

ALICE-PUBLIC-2023-003



UNIVERSITÀ DI TRENTO
UNIVERSITÀ DELLA CALABRIA
DIPARTIMENTO DI FISICA
MUR
Finanziato dall'Unione europea
NextGenerationEU
PhD SST
Space Science and Technology

DIPARTIMENTO DI

FISICA

ESC23 Scuola Internazionale INFN

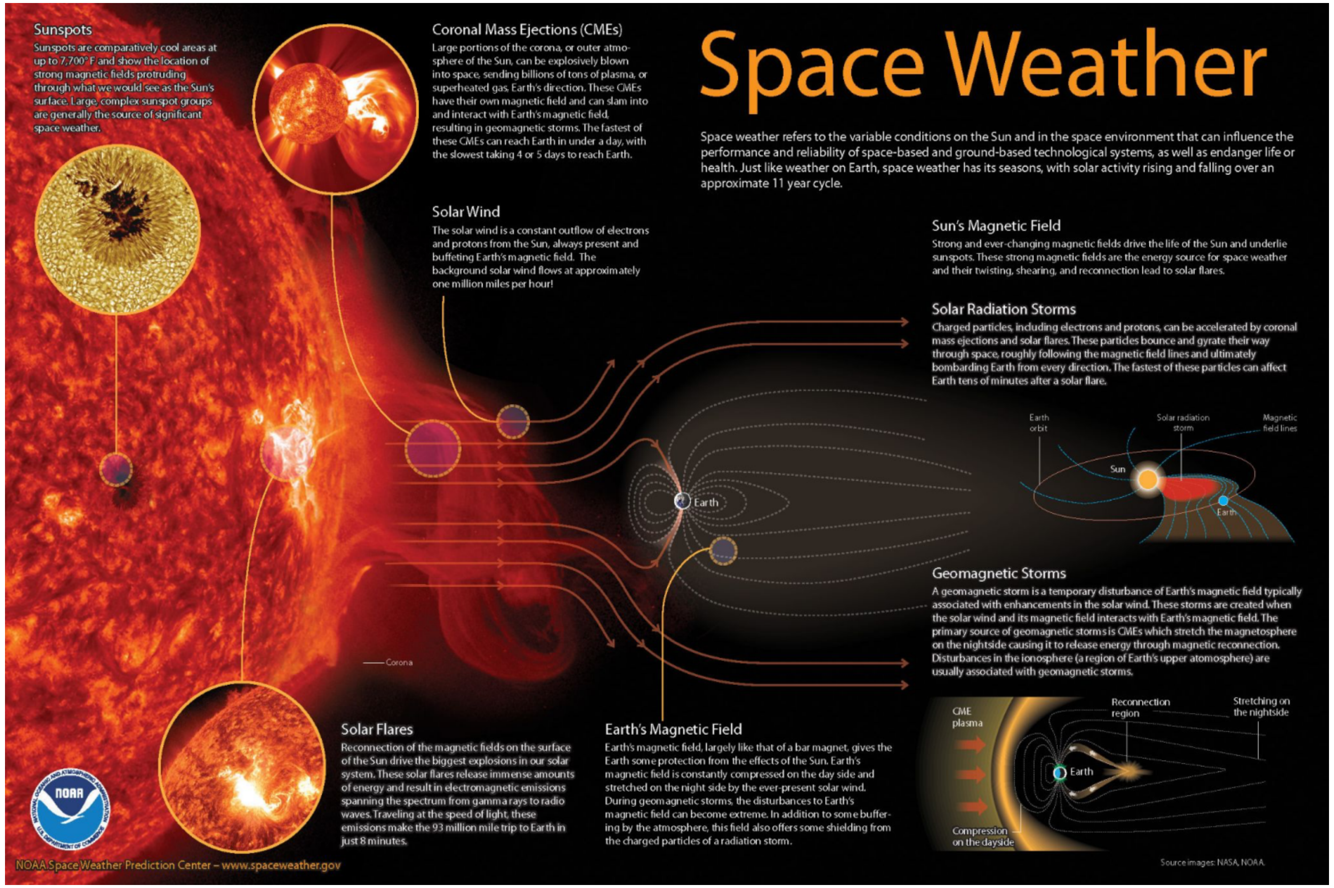
Architetture, strumenti e metodologie per lo sviluppo di applicazioni di calcolo scientifico su larga scala

Solar Flare Forecasting based on Deep Learning Models

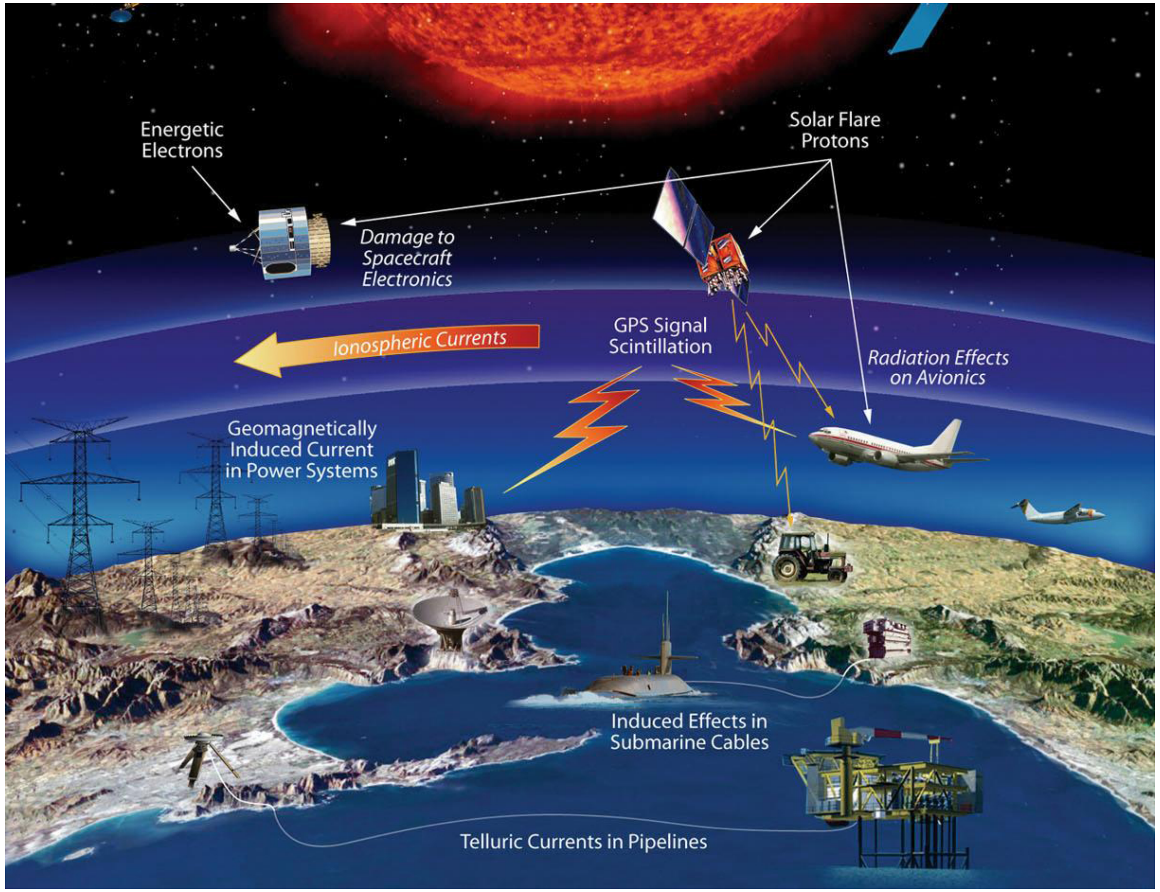
Elizabeth Doria Rosales

Supervisor: Prof. Vincenzo Carbone
Co-Supervisor: Prof. Fabio Lepreti

Impact of Space Weather on Earth and Human Technologies



NOAA space weather infographic.



Technological and infrastructure affected by space weather events.
Credits: NASA

UNIVERSITÀ DI TRENTO Dipartimento di Fisica

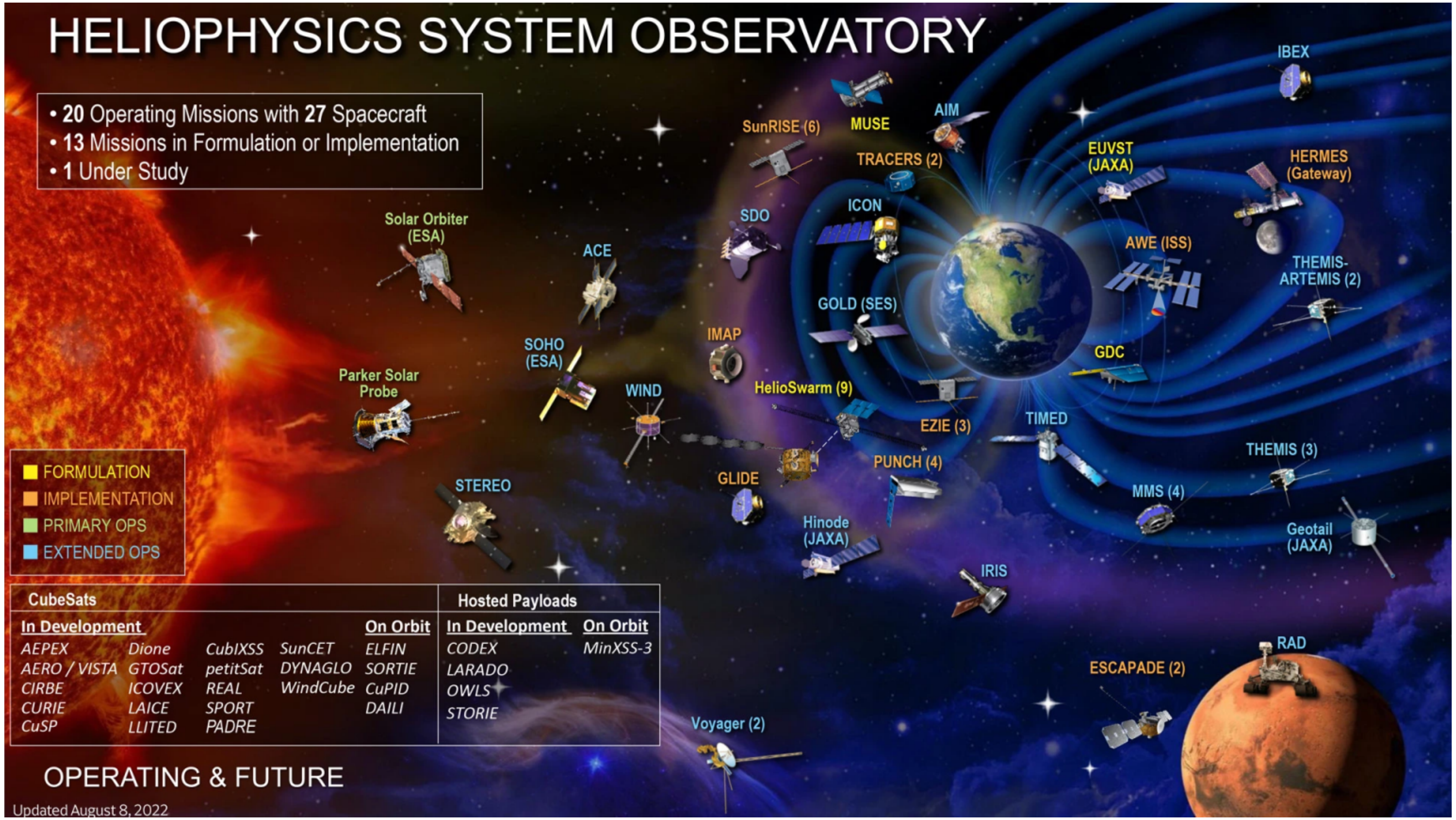
UNIVERSITÀ DELLA CALABRIA DIPARTIMENTO DI FISICA

MUR

Finanziato dall'Unione europea NextGenerationEU

PhD SST Space Science and Technology

Heliophysics Missions and Programs

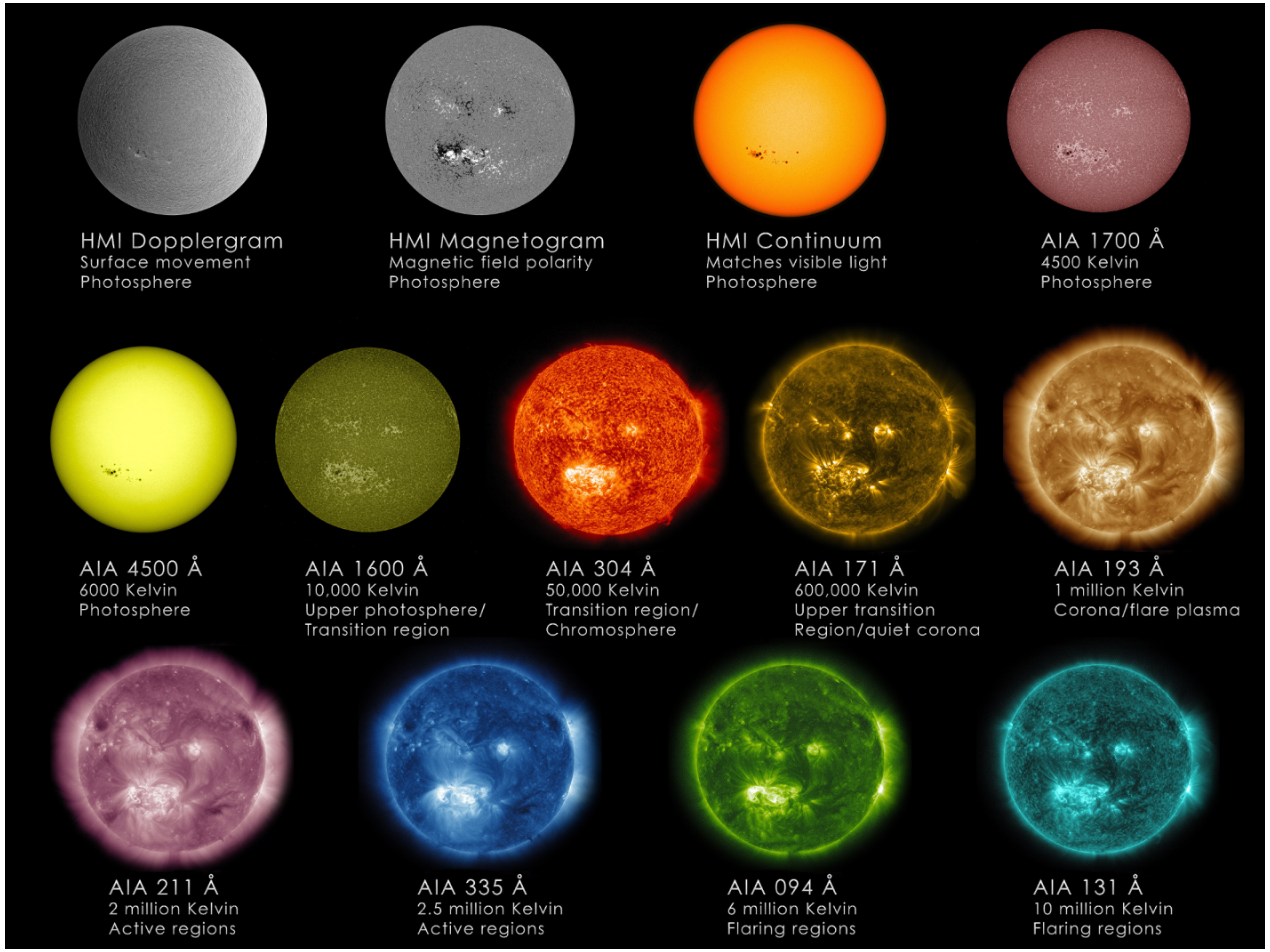


Credit: NASA's Goddard Space Flight Center

SDO launched on February 11, 2010, 10:23 am EST on an Atlas V from SLC 41 from Cape Canaveral.

After launch, the spacecraft was placed into an orbit around the Earth with an initial perigee of about 2,500 km (1,600 mi). SDO then underwent a series of orbit-raising maneuvers which adjusted its orbit until the spacecraft reached its planned ***circular, geosynchronous orbit at an altitude of 35,789 km (22,238 mi), at 102° West longitude, inclined at 28.5°.***

How SDO Sees the Sun

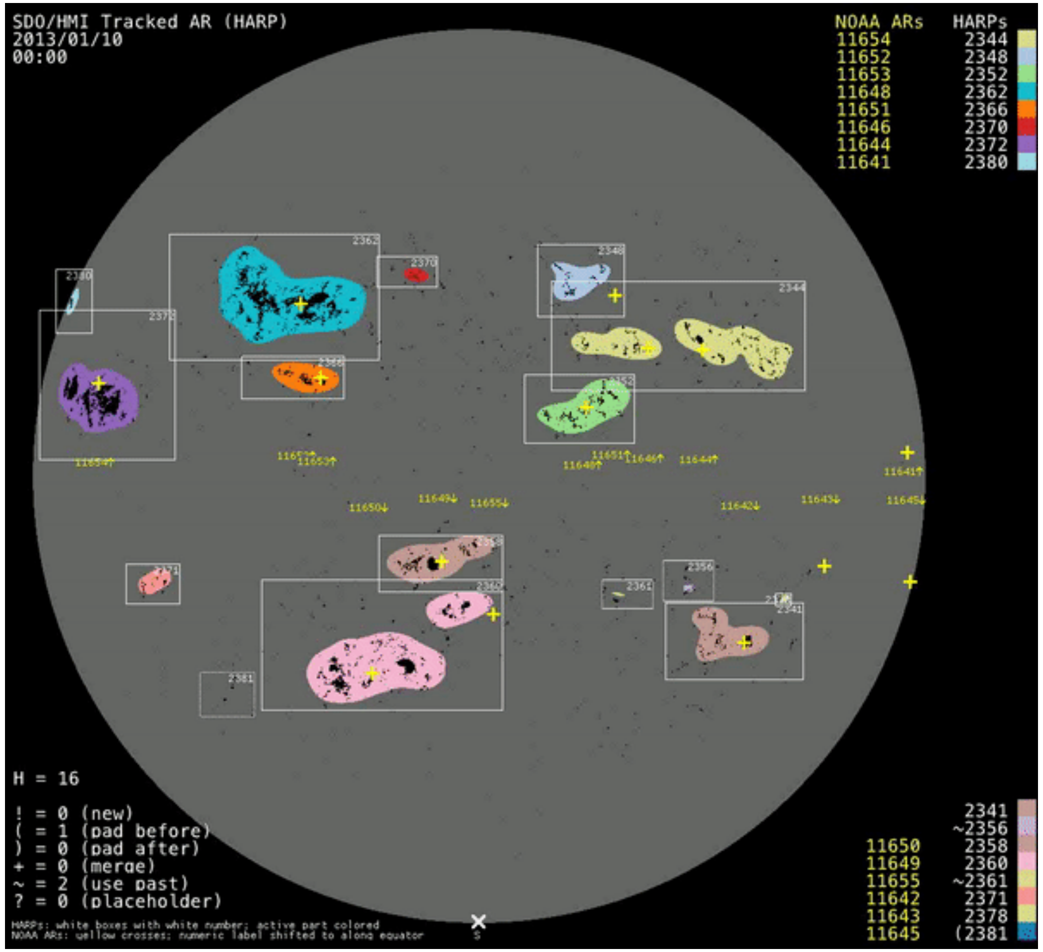


Each of the wavelengths observed by NASA's Solar Dynamics Observatory (SDO) was chosen to emphasize a specific aspect of the Sun's surface or atmosphere.

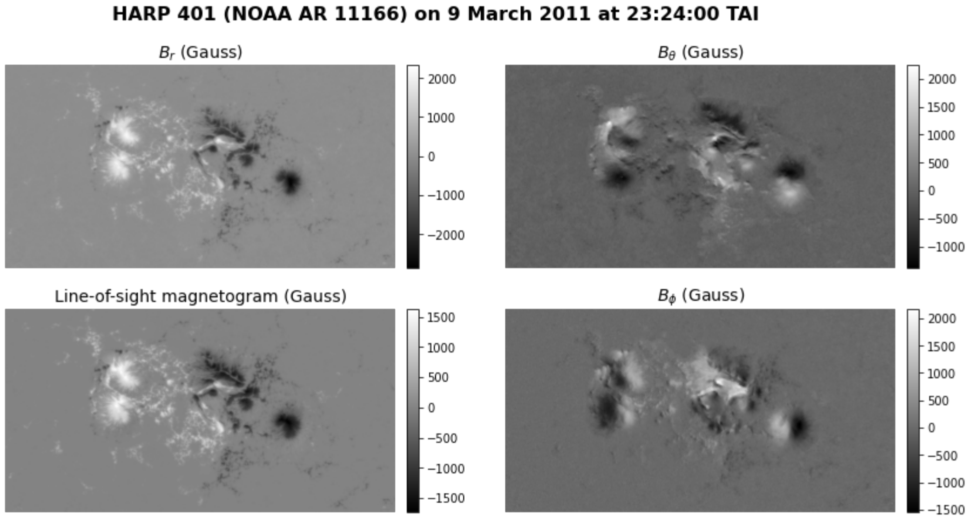
This image shows imagery both from the Advanced Imaging Assembly (AIA), which helps scientists observe how solar material moves around the Sun's atmosphere, and the Helioseismic and Magnetic Imager (HMI), which focuses on the movement and magnetic properties of the Sun's surface.

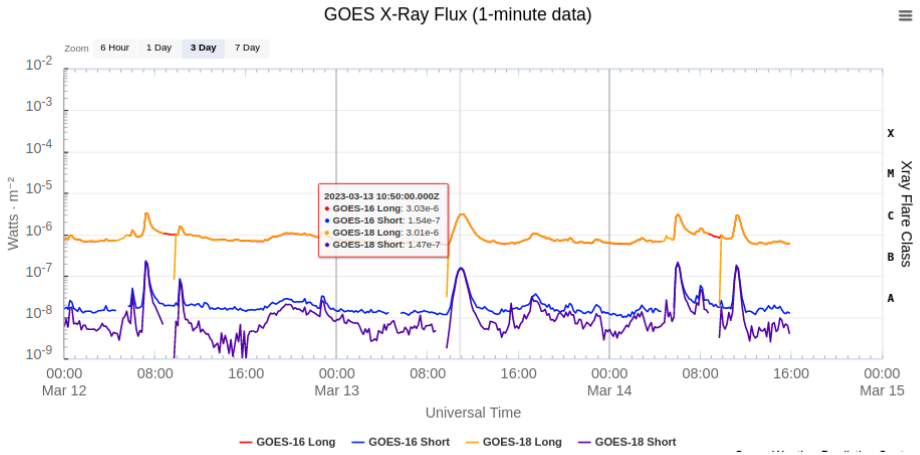
Credits: NASA/SDO/GSFC

Spaceweather HMI Active Region Patch (SHARP)



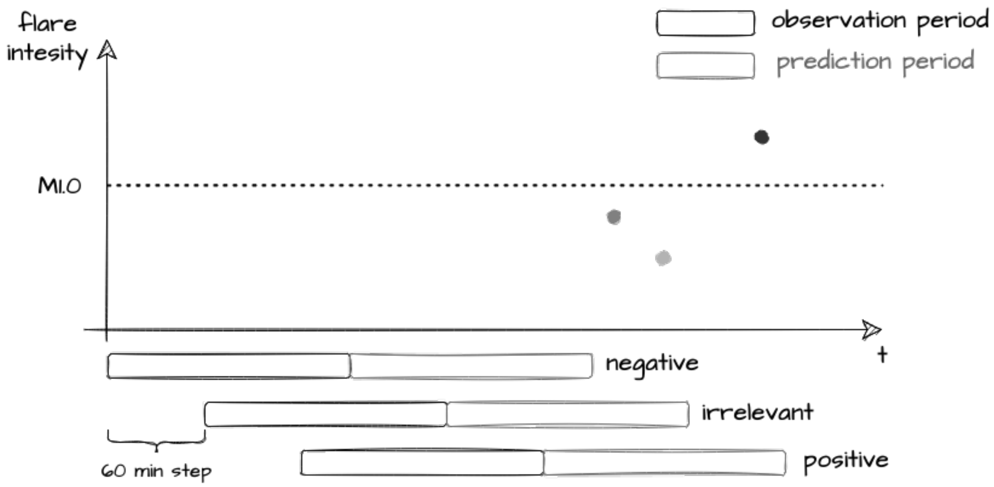
Joint Science Operations Center (JSOC) (http://jsoc.stanford.edu/)

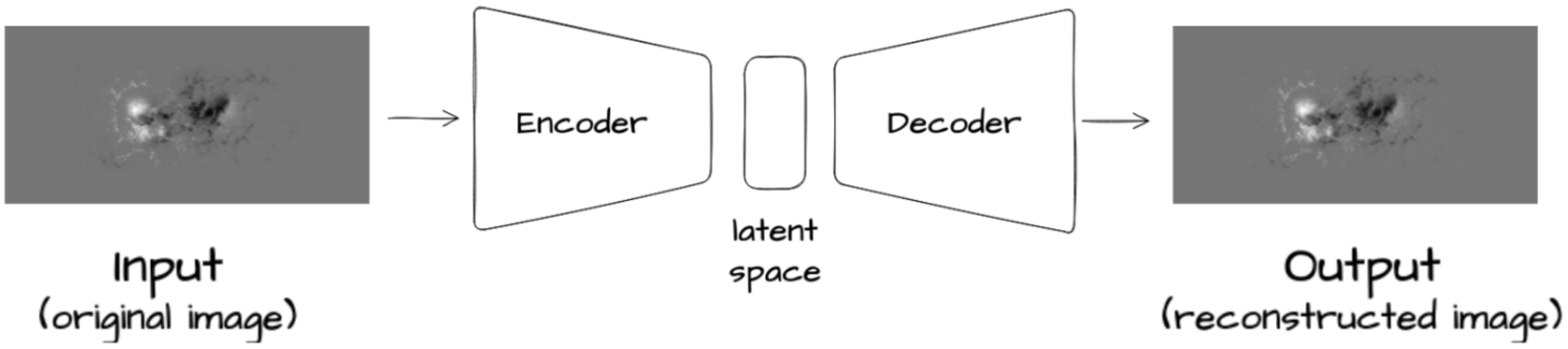




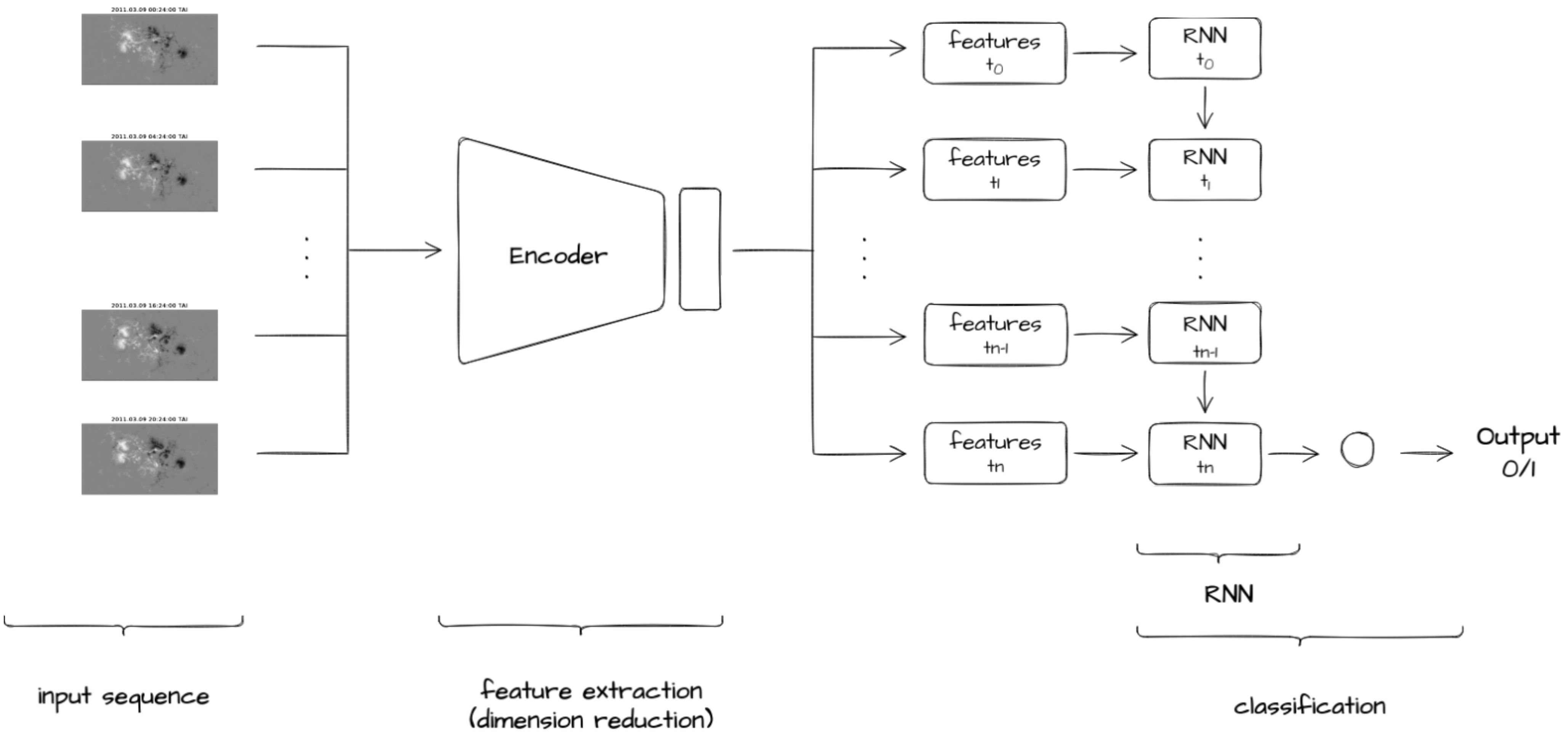
Geostationary Operational Environmental Satellite (GOES) (https://www.swpc.noaa.gov/)

First Part: Feature extraction





Second Part: Time series classification



Lightning Students Talk

XIV INFN International School on Efficient Scientific Computing (ESC23)
4-12 October, Bertinoro

Sabrina Giorgetti

Getting to Know Me

Academic Path

- Master's degree in Physics of Data at the University of Padua (2020-2022)
- Technical student at CERN (2022)
- Member of the CMS collaboration (since March 2022)
- PhD student in Physics at the University of Padua (since February 2023)
  sabrina.giorgetti@studenti.unipd.it
- INFN Padova associate (since February 2023)

Why I'm interested in the school?

- Learn tools and techniques to improve my programming skills
  - Memory usage and parallel programming
- Boost my current work, and potentially stimulate new projects
  - GP-GPU programming

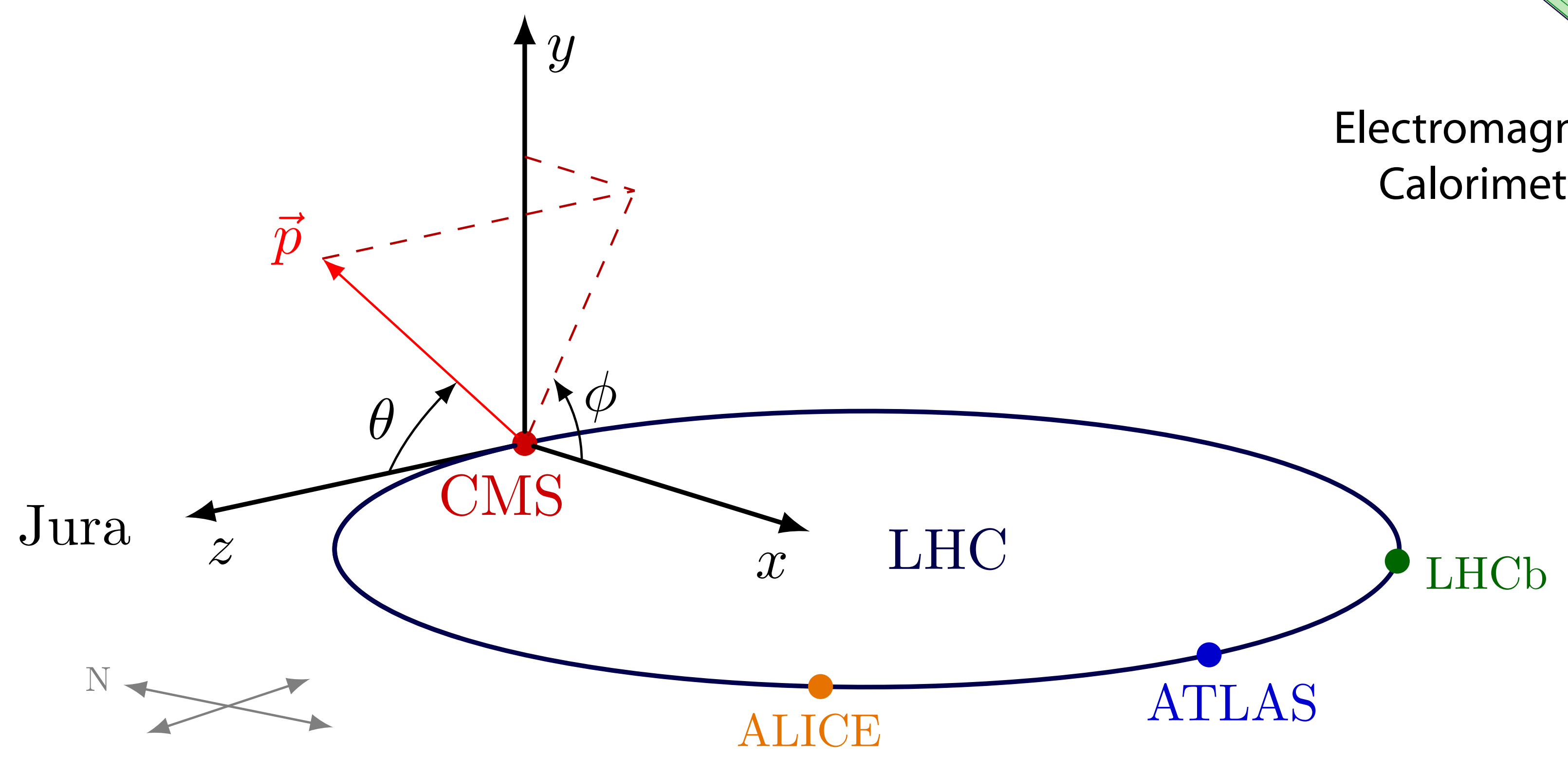
The CMS experiment

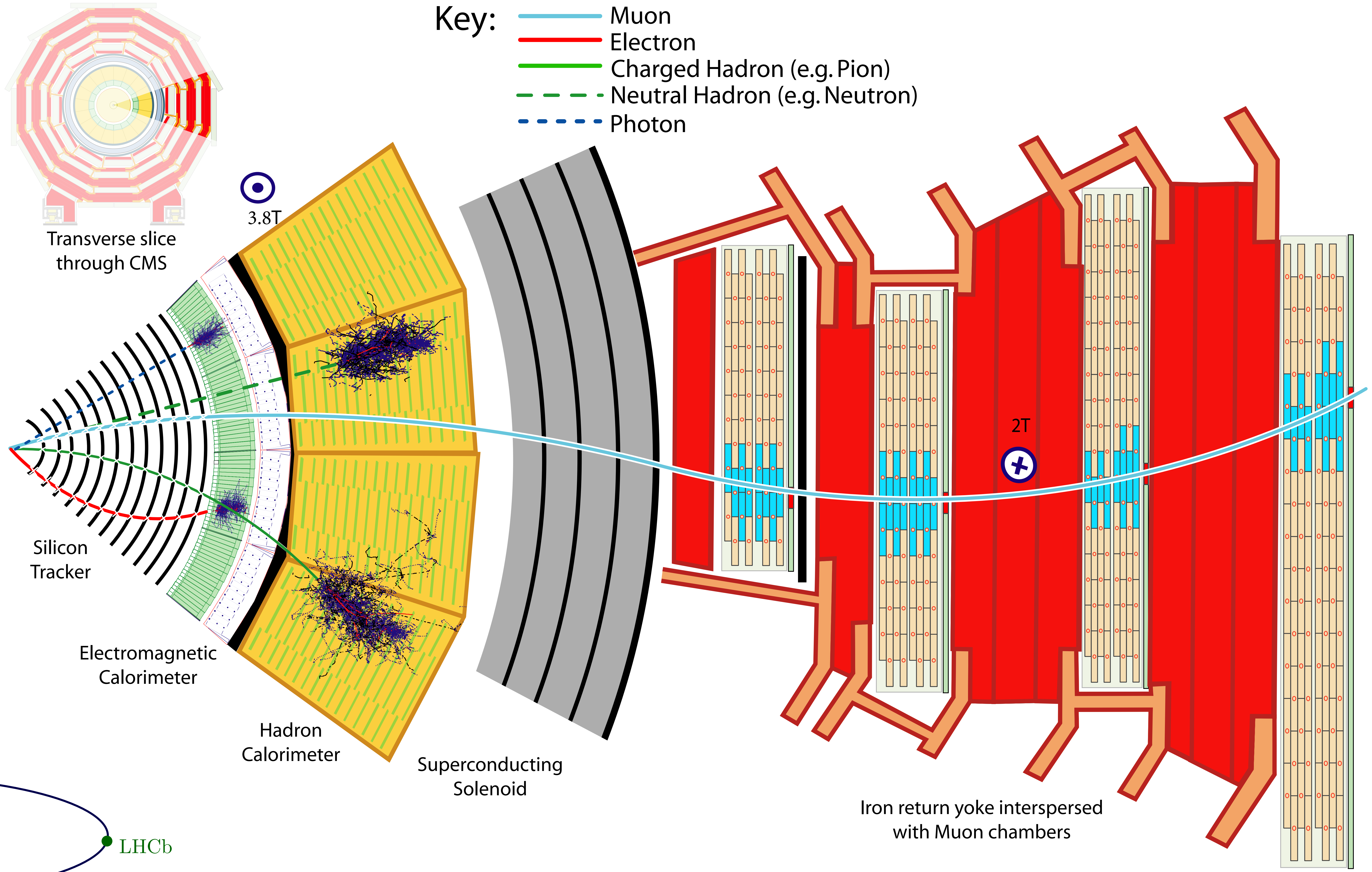
Large Hadron Collider (LHC)

- LHC delivers proton-proton collisions at ECM = 13.6 TeV since the start of Run 3 (July 2022)
- LHC bunches collide every 25 ns with a resulting event's rate of 40MHz

CMS

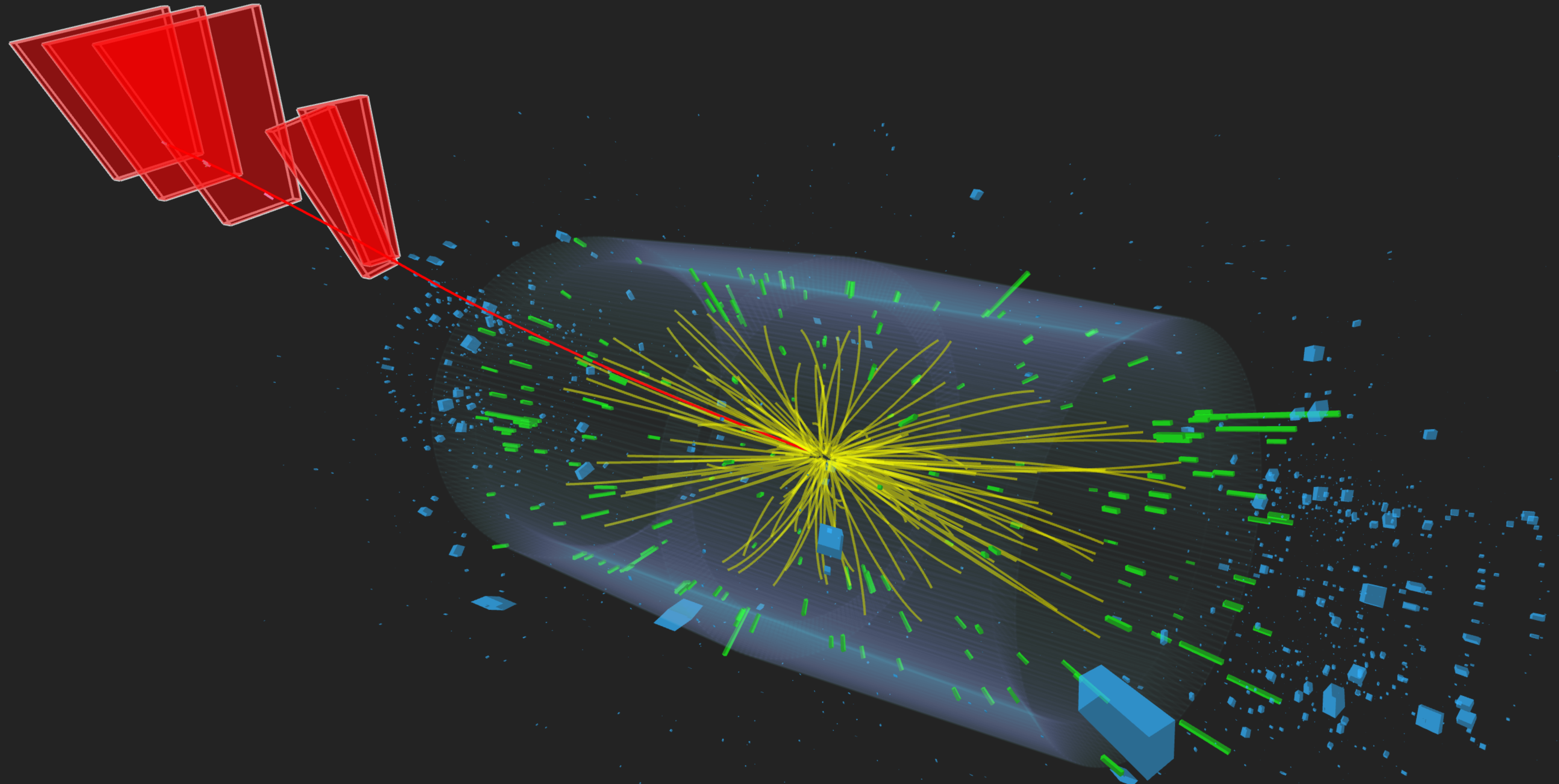
- General-purpose detector
- Measure the energy and momentum of collisions products



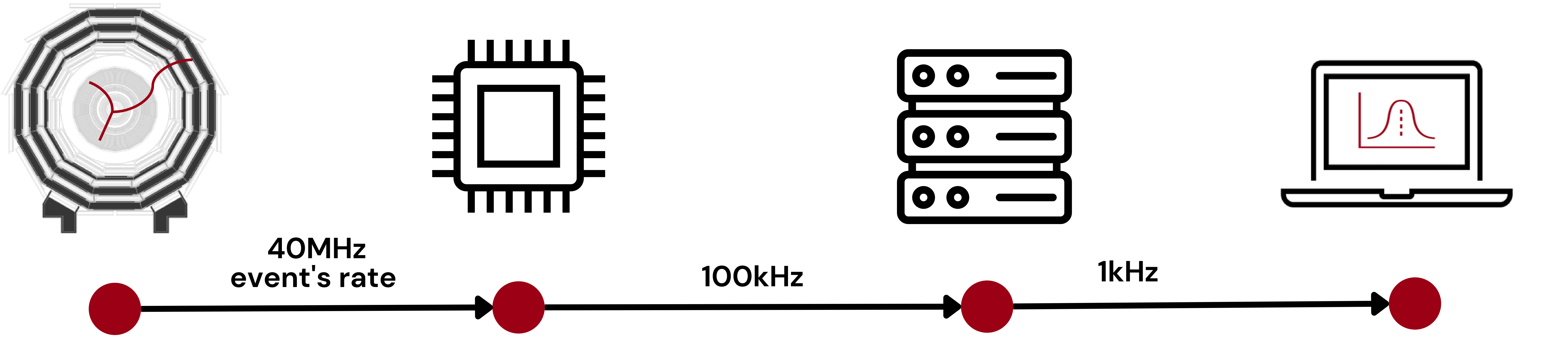


CMS Experiment at the LHC, CERN
Data recorded: 2023-Apr-21 17:14:17.991232 GMT
Run / Event / LS: 366403 / 108625285 / 113

FIRST COLLISIONS OF 2023 IN CMS



The CMS data flow



DETECTOR COLLISIONS

LEVEL 1 TRIGGER

- Coarser data from calorimeters and muon systems
- Data bandwidth O(100) TB
- Trigger latency O(μs)
- Implemented in hardware (ASICs and FPGA)
- **99.75% events rejected**

HIGH LEVEL TRIGGER

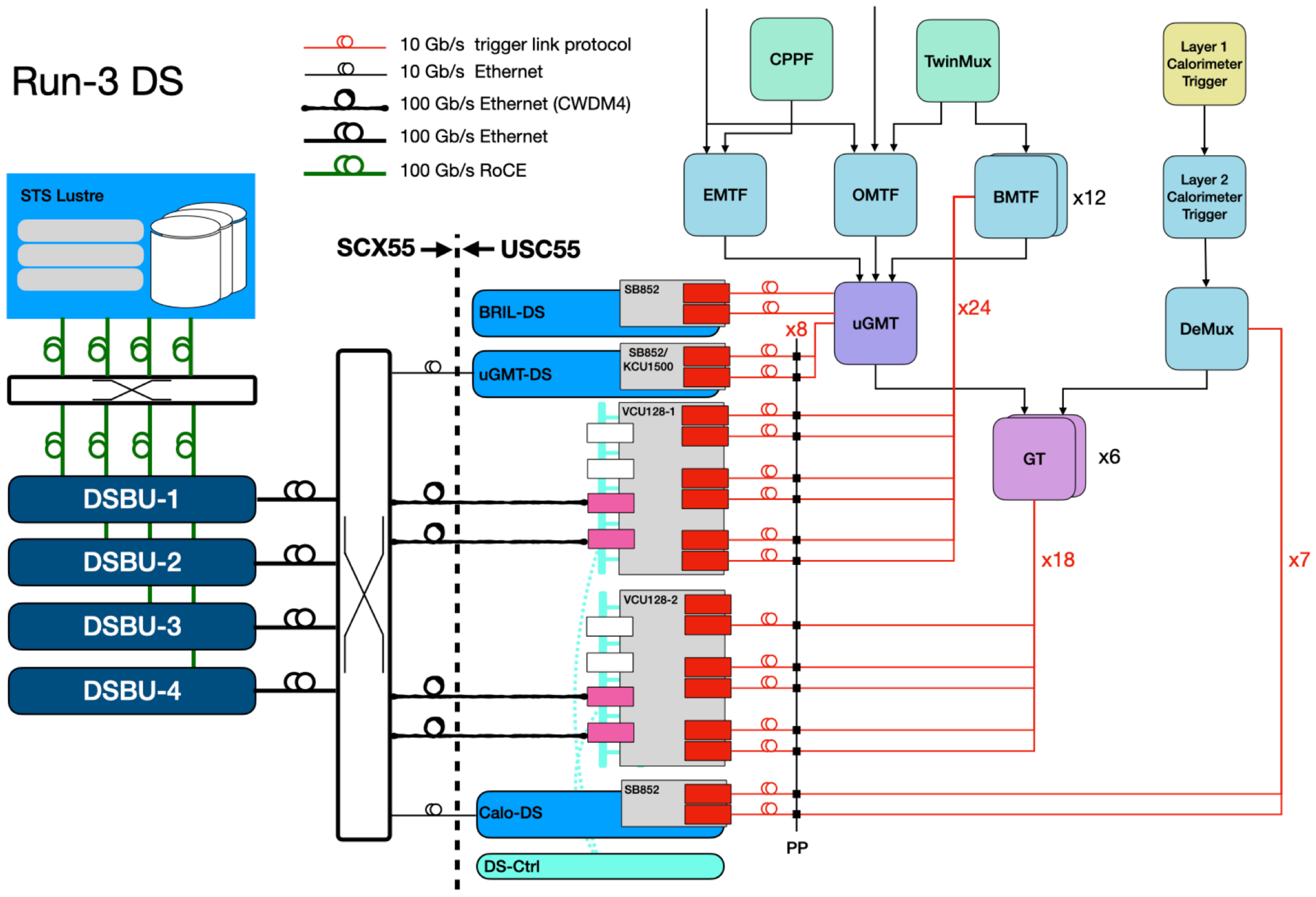
- Analysis of the full event with high-granularity data
- Latency O(100 ms)
- Implemented in software: runs on commercial computers
- **99% events rejected!**

EVENTS REJECTED ARE LOST!

OFFLINE DATA ANALYSIS

The Level-1 Data Scouting system

The Run 3 demonstrator



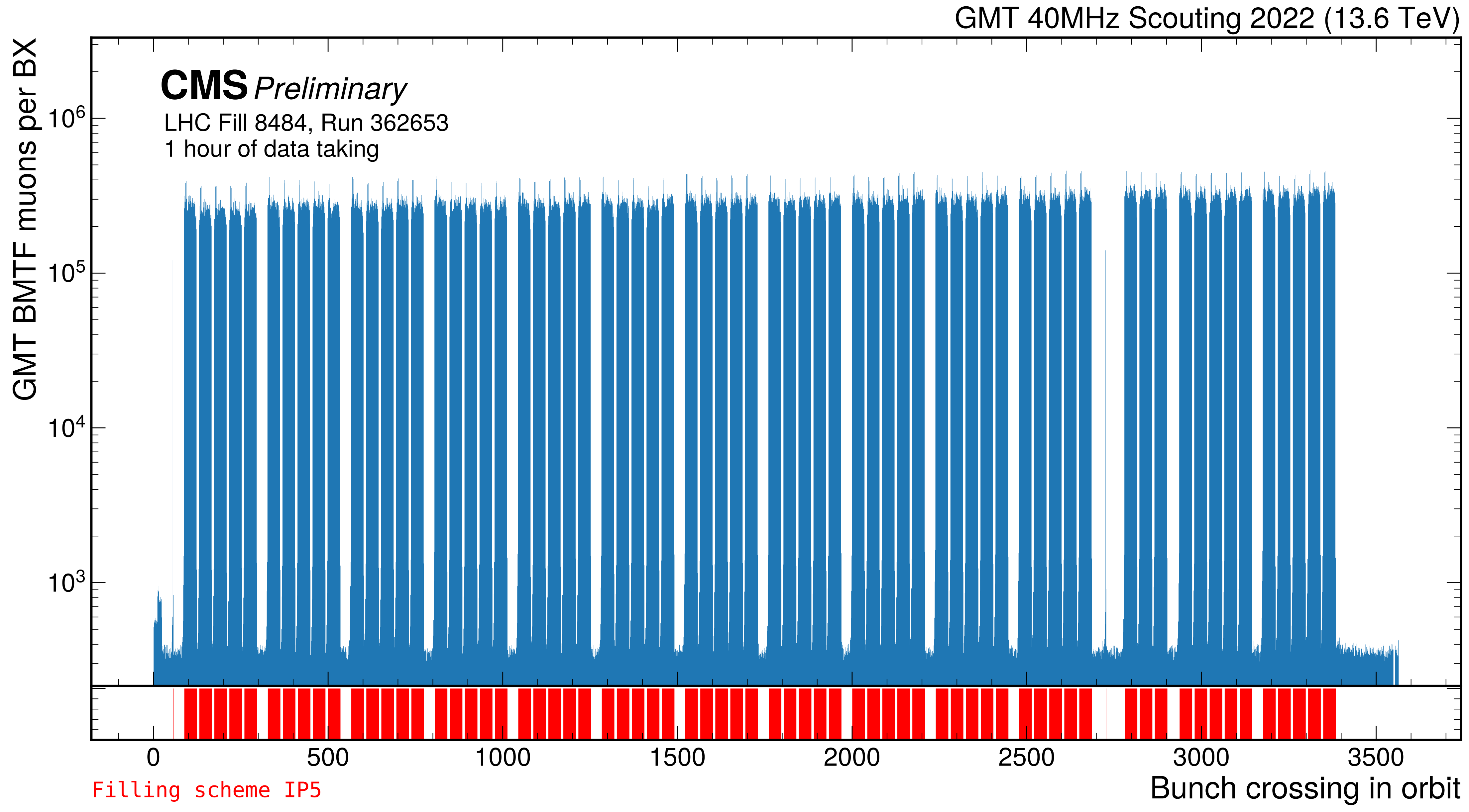
- Novel data acquisition system that stores data before the Level-1 (L1) accept
- The Level-1 Data Scouting (L1DS) system is designed for the CMS Phase-2 at HL-LHC
- A demonstrator, parallel and independent from the standard trigger chain, is currently being developed in Run 3
- Gathers L1 primitives at the full bunch crossing frequency from different sources:
  - *Global Muon Trigger (GMT)* : up to 8 muons
  - *Barrel Muon Track Finder (BMTF)*: stubs primitives
  - *Calorimeter Layer-2*: up to 12 each of e/γ, jets, τ missing transverse energy and energy sums
  - *Global Trigger (GT)*: trigger decision algorithms bits

The Level-1 Data Scouting system

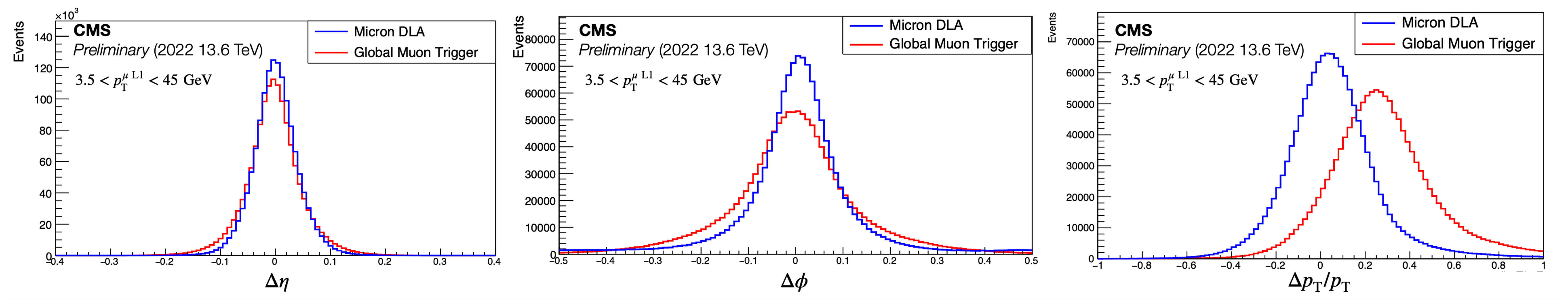
Analysis and ML applications

L1DS analysis

- Characterisation of data collected since the start of Run 3
  - Monitor the rates, occupancy over bunch crossings
  - Histograms of objects' parameters, study di-muon events
- Trigger-less data at 40MHz
  - 1 min of L1DS data taking corresponds to ~2.5GB of data formatted for analysis
  - Need to use distributed framework for processing (e.g Spark)



Global Muon Trigger (GMT) muons occupancy per bunch crossing (BX) within an LHC orbit for muons reconstructed by the Barrel Muon Track Finder (BMTF) [CMS-DP-2023-025].



The distribution of η, ϕ, p_T differences between the neural network (Micron DLA) prediction (or GMT) values, and the offline reconstructed muon tracks, for matched muons. A significant improvement in track parameter resolution is observed in the MDLA result (blue) when compared to the GMT output (red) [CMS-DP-2022-066].

ML applications

- Recalibration of L1 trigger primitives to improve resolution
- Implementation of algorithms in hardware (e.g `hl4ml`)

Presentation of Research Activity

@ ESC23

Jekatěrina Jaroslavceva[1,2]

Supervisor: Felice Pantaleo[1]

[1] CERN, European Organization for Nuclear Research, Meyrin, Switzerland
[2] Czech Technical University in Prague, Prague, Czech Republic

Who am I?

ABOUT ME

- First-year **PhD** student in Computer Science, since September 2023, at the **Czech Technical University in Prague**.
- Master thesis title: "A New Trackster Linking Algorithm Based on Graph Neural Networks for the CMS Experiment at the Large Hadron Collider at CERN".
- Currently, a DOCT student @ **CERN** working on improving reconstruction in **HGCAL** (new CMS sub-detector) using Graph Neural Networks.

RESEARCH ACTIVITIES:

>> Enhancing the CMS Event Reconstruction using **Artificial Intelligence**: efficient techniques for linking granular information, such as calorimetric tracksters, tracks, and MTD timing, in the Phase 2 CMS detector.

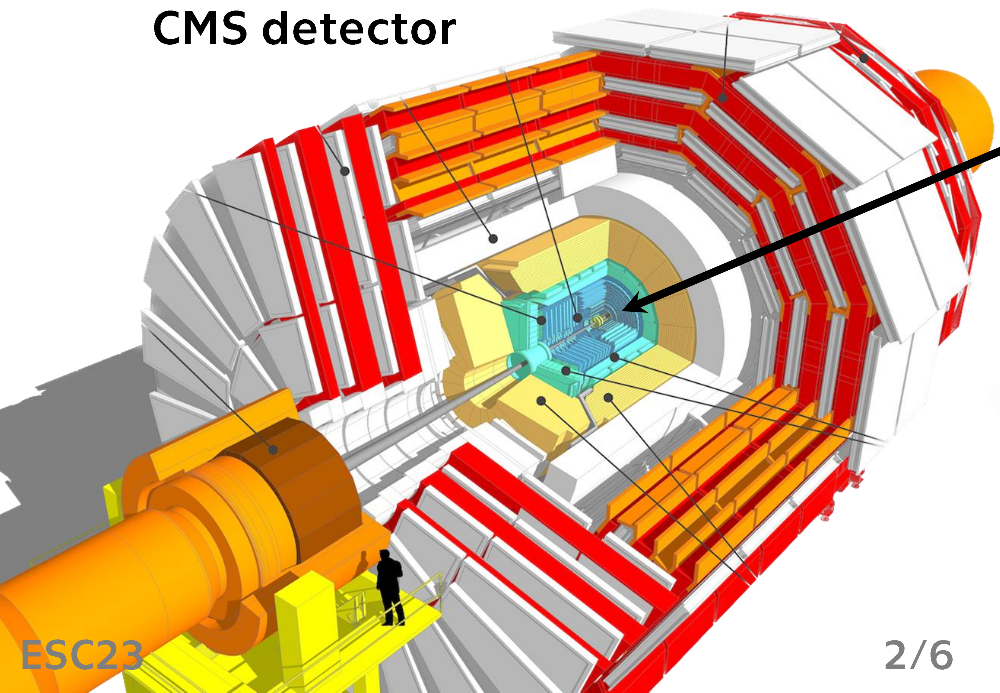
>> Designing a **heterogeneous** inference approach within the CMSSW framework, enabling seamless resource management and execution on GPUs when available, while gracefully falling back to CPU execution.

The CMS High Granularity Calorimeter

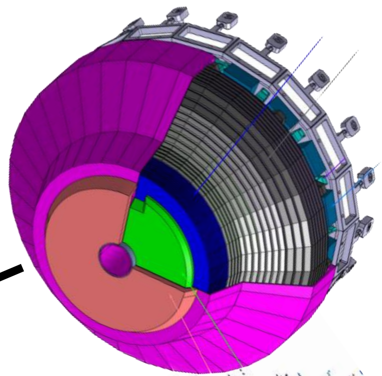
- HGCAL operation starts in 2029.
- Measures the energy of traversing particles like large 3D camera.
- Particles interact with the detector material and produce particle showers.
- Interactions captured in sensors as hits.

- Collision frequency of 40 MHz.
- ~60 simultaneous proton-proton collisions (pile-up).
- Rise to ~140 in 2029.

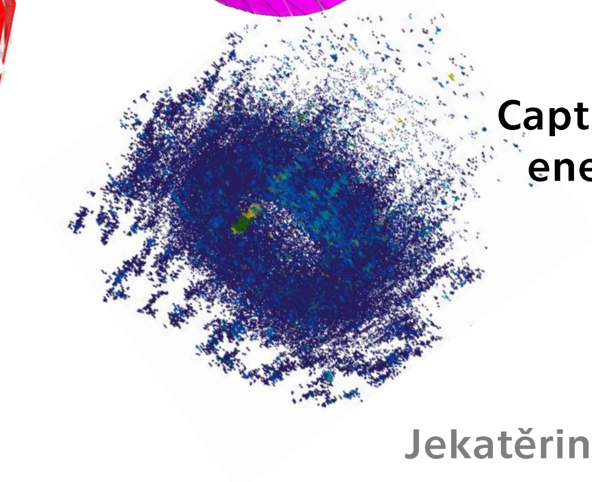
CMS detector



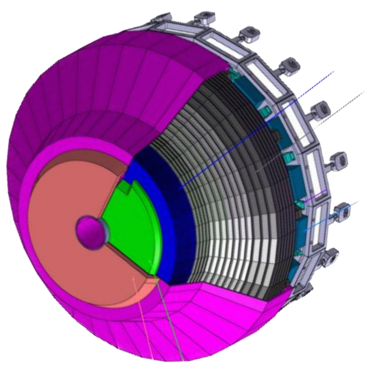
The High Granularity Calorimeter (under development)

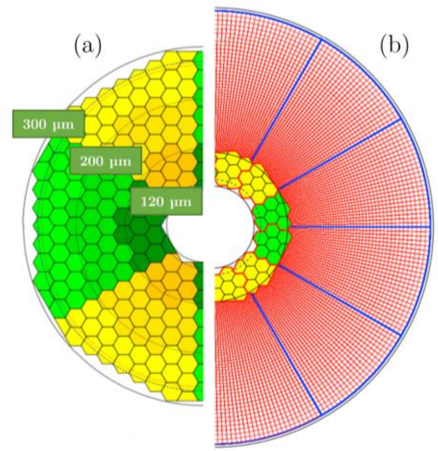


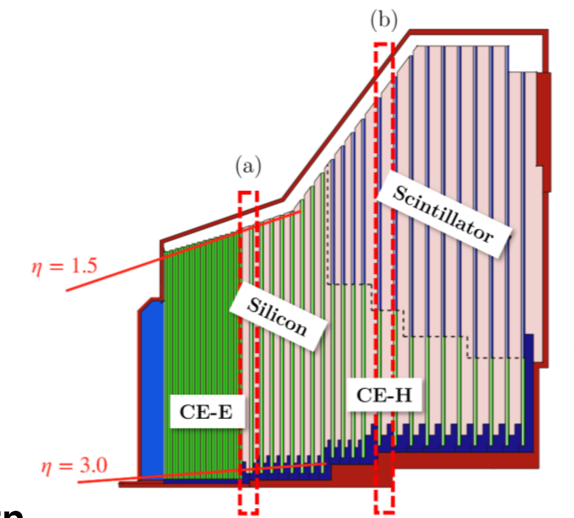
Captured cloud of energy depsits

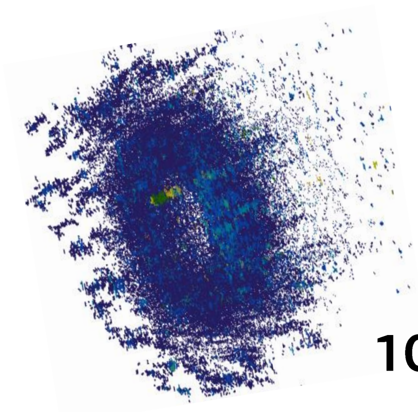


Reconstruction in HGCAL



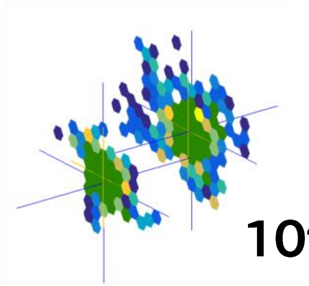






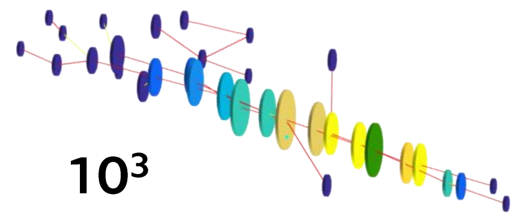
10^5

2D clustering



10^4

Pattern recognition



10^3

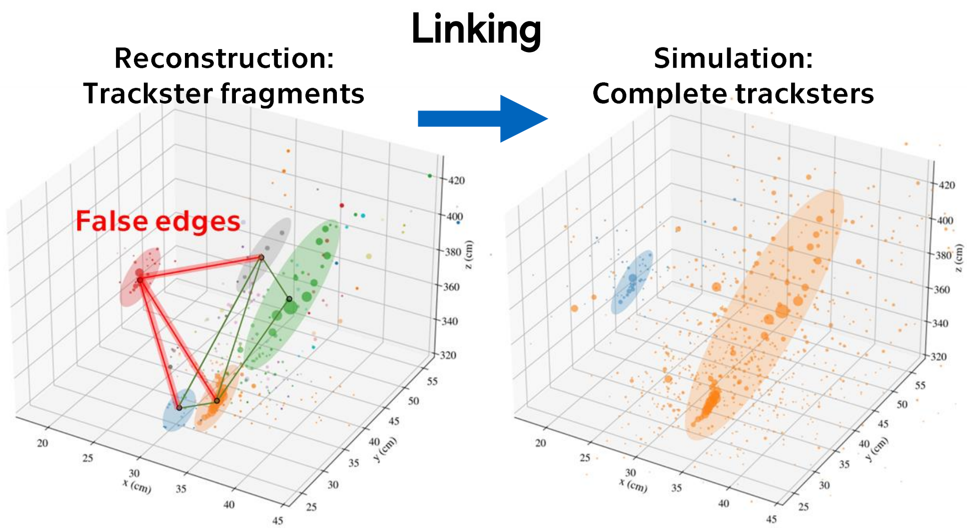
Hits (*x*, *y*, *z*, *t*, *E*): Energy deposits in the detector sensors

Layer-cluster (LC): cluster of hits from the same particle on a detector layer (2D)

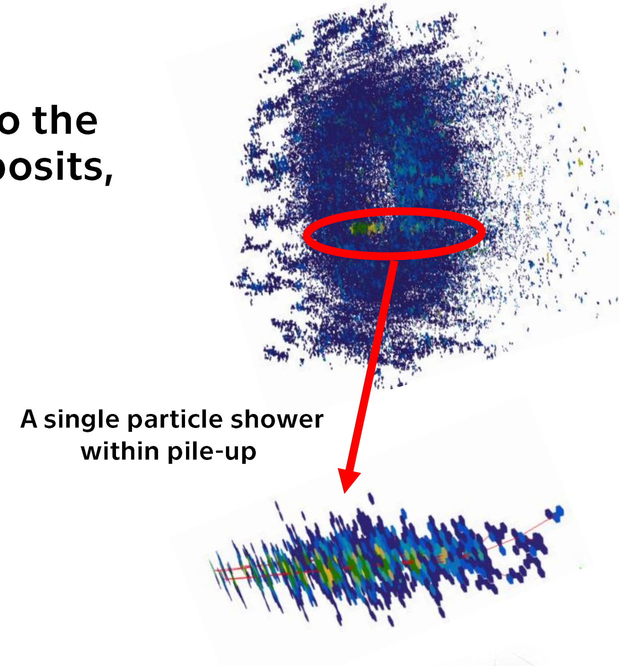
Trackster: collection of LCs from the same particle shower (3D)

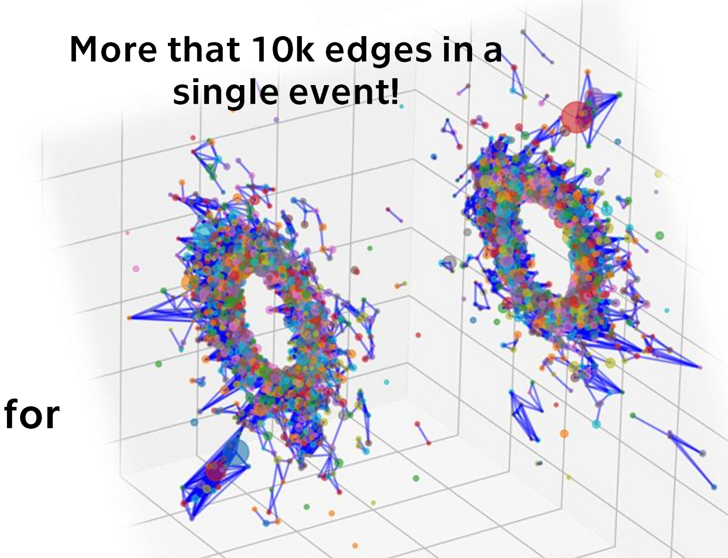
Trackster Linking

- Ideally, individual particle showers = individual tracksters.
- A clustering algorithm produces trackster fragments due to the detector's irregularity, distant and non-aligned energy deposits, tuning for high purity.
- **An additional trackster linking step is needed.**



Reduced example of a small window event graph of 2 particles.



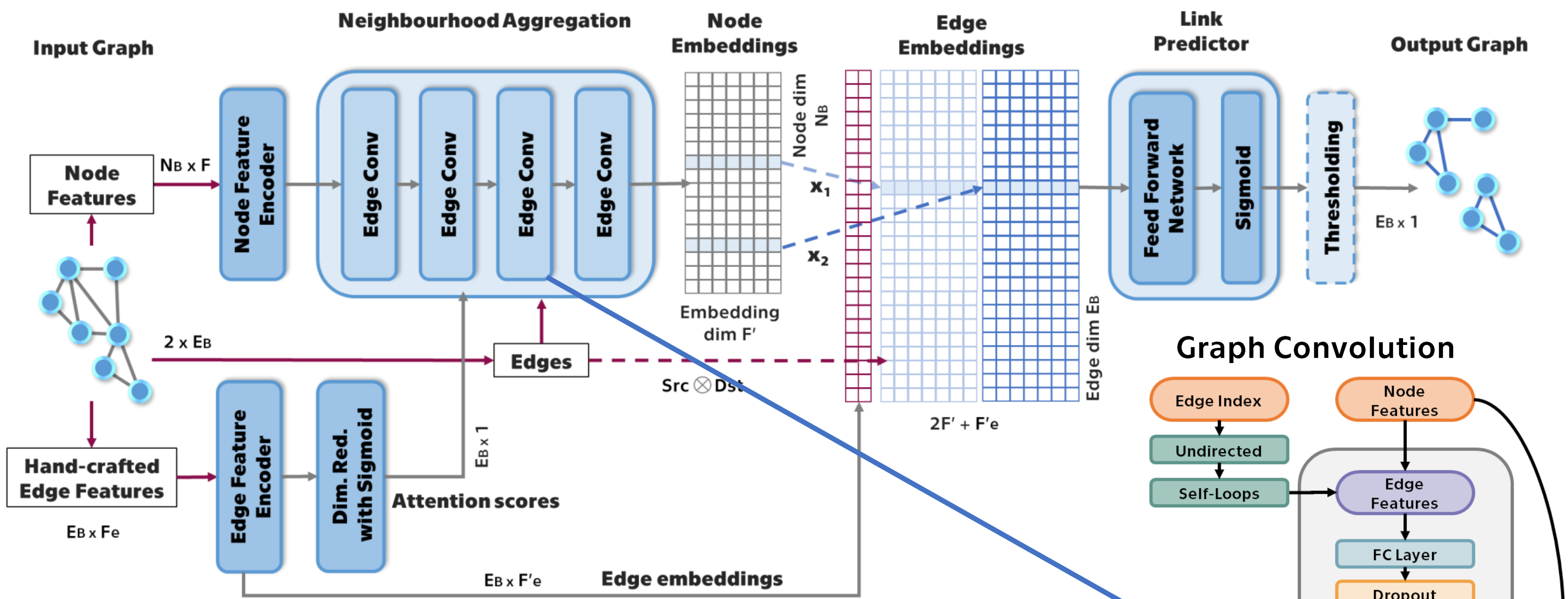


Project goal: recover complete particle showers.

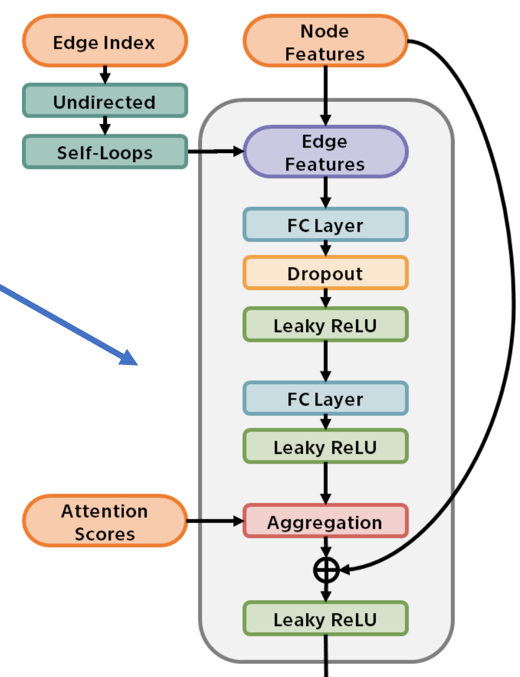
How: Develop a Graph Neural Network (GNN) approach for learning structures on a graph of nearby tracksters.

Graph Network Architecture

*** ONNX-Exportable for inference in production environment**



Graph Convolution



Node feature update:

$$\mathbf{h}_i^{t+1} = \mathbf{h}_i^t + \sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{ij} \cdot \mathrm{NN}(\mathbf{h}_i^t \,||\, \mathbf{h}_j^t - \mathbf{h}_i^t)$$

Attention scores:

$$\alpha_{ij} = \frac{1}{1 + \exp(-\mathbf{a}_\Theta^T (\mathbf{u}_{ij}^0)^T)} \in [0, 1]$$

Focal Loss function:

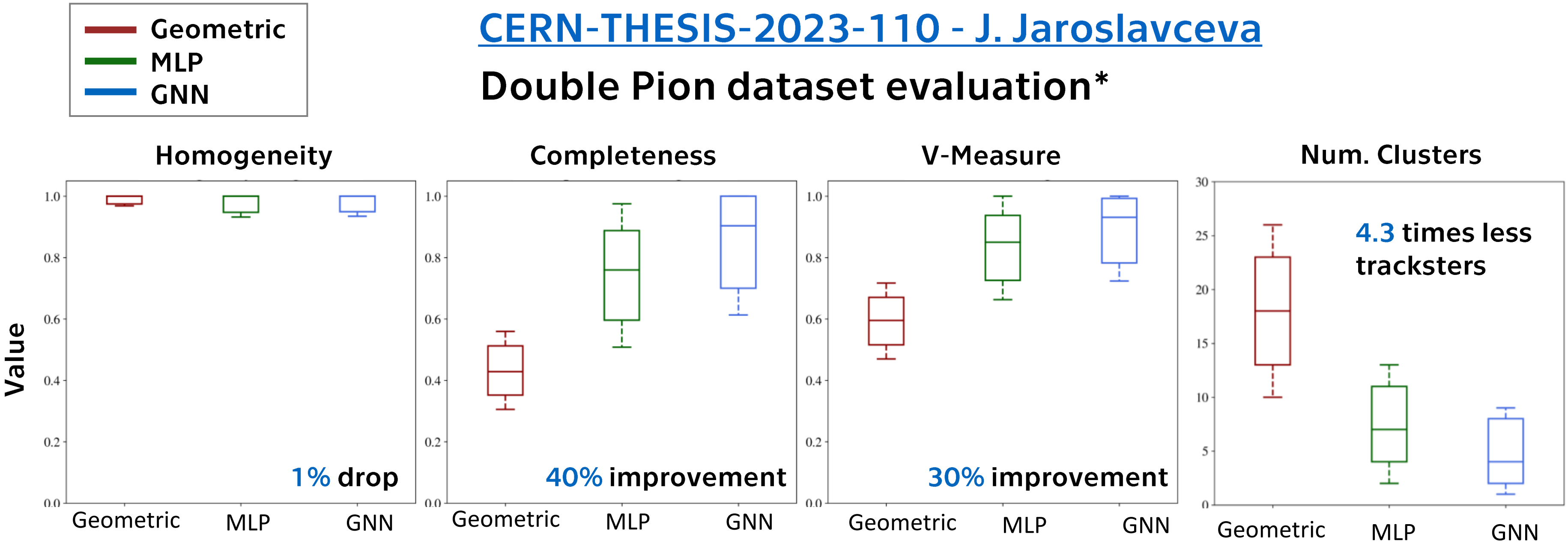
$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t)$$

$$p_t = \begin{cases} p & \text{if } y = 1, \\ 1 - p & \text{otherwise.} \end{cases}$$

Performance Evaluation

[CERN-THESIS-2023-110 - J. Jaroslavceva](#)

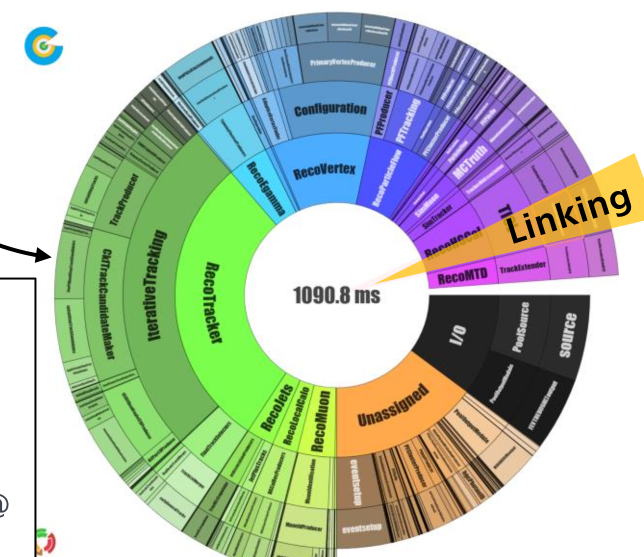
Double Pion dataset evaluation*



Why GNN?

- CLUE3D Trackster can be used as nodes in a **Graph**, which GNN takes as input
- Local information **aggregation**.
- graph the TICLGraph
- **Nodes Feature**: Trackster properties
- **Edge Features**: Geometrical, Energy and Time compatibility
- Fast CPU inference (ONNX)

GNN CPU timing in the production environment: ~3.2% of the total reconstruction



Single core CPU* inference time in CMSSW, in PU 200:
Geometric Linking 680 ms/ev
GNN: 1091 ms/ev

*CPU: Intel(R) Xeon(R) Silver 4216 CPU @ 2.10GHz

Trigger performance monitor at the LHCb detector

Bogdan Kutsenko, Dorothea vom Bruch, Anton Poluektov

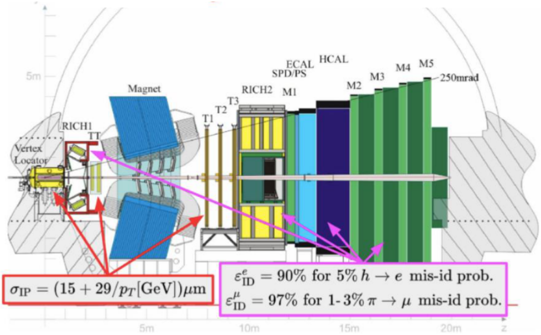
Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France

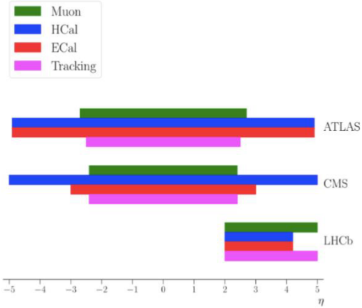
ESC23 - October 6, 2023

LHCb detector

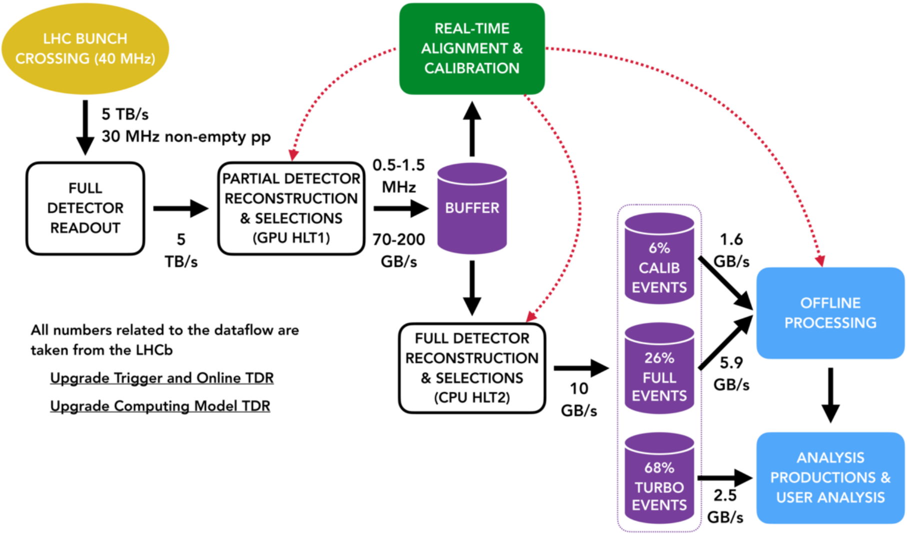
Single-arm forward spectrometer:

- Designed for the heavy flavor physics with $2 < \eta < 5$
- Coverage is complementary to ATLAS and CMS
- Excellent performance of tracking, powerful particle identification and the fact that B meson produced in the forward region allow to perform high precision measurement of semileptonic B decays





LHCb Dataflow Run 3



All numbers related to the dataflow are taken from the LHCb

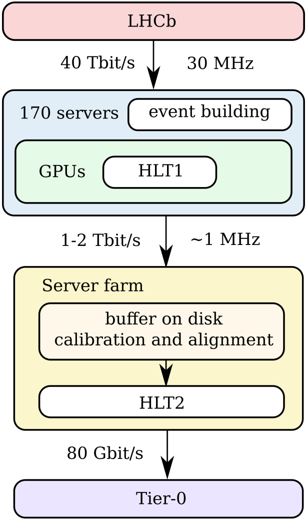
- Upgrade Trigger and Online TDR
- Upgrade Computing Model TDR

Allen - high-level trigger (HLT) application on GPUs. The HLT in the LHCb experiment at CERN's Large Hadron Collider (LHC) plays a crucial role in efficiently filtering data from the collisions that occur within the detector. It runs with the upgraded LHCb detector in the Run 3. Two new algorithms for the Allen project were developed on GPUs:

- Splitting of tracks into two different categories based on the defined selection criteria (right/left side selection, random selection, etc.)
- Two different sets of tracks are used to reconstruct two sets of primary vertexes (PVs)

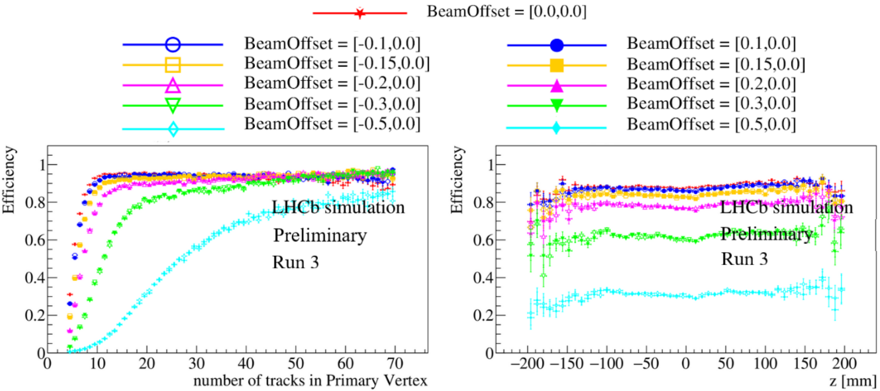
Motivation:

- Real-time reconstruction needs detailed monitoring of the physics performance of the algorithms
- Opportunity to split tracks, in general, can be useful for resolution estimation on data



Allen project

- The performance of the PV reconstruction algorithm heavily relies on good alignment. Therefore an additional PV performance monitor is required



- Resolution monitor is currently work in progress

Status and outlook

The performance of the PV reconstruction algorithm was studied:

- Motivation for an additional PV performance monitor become apparent
- To create monitor new algorithms for the Allen project were developed on GPUs and tested

Plans:

- Using new Allen algorithms make an additional performance monitor of the PV resolution, which will be shown in the control room during the data-taking period

Interests in the field of scientific computing:

- Programming with Threading Building Blocks
- GP-GPU Programming with CUDA

Thank you!