



Data analysis and hardware R&D at CMS

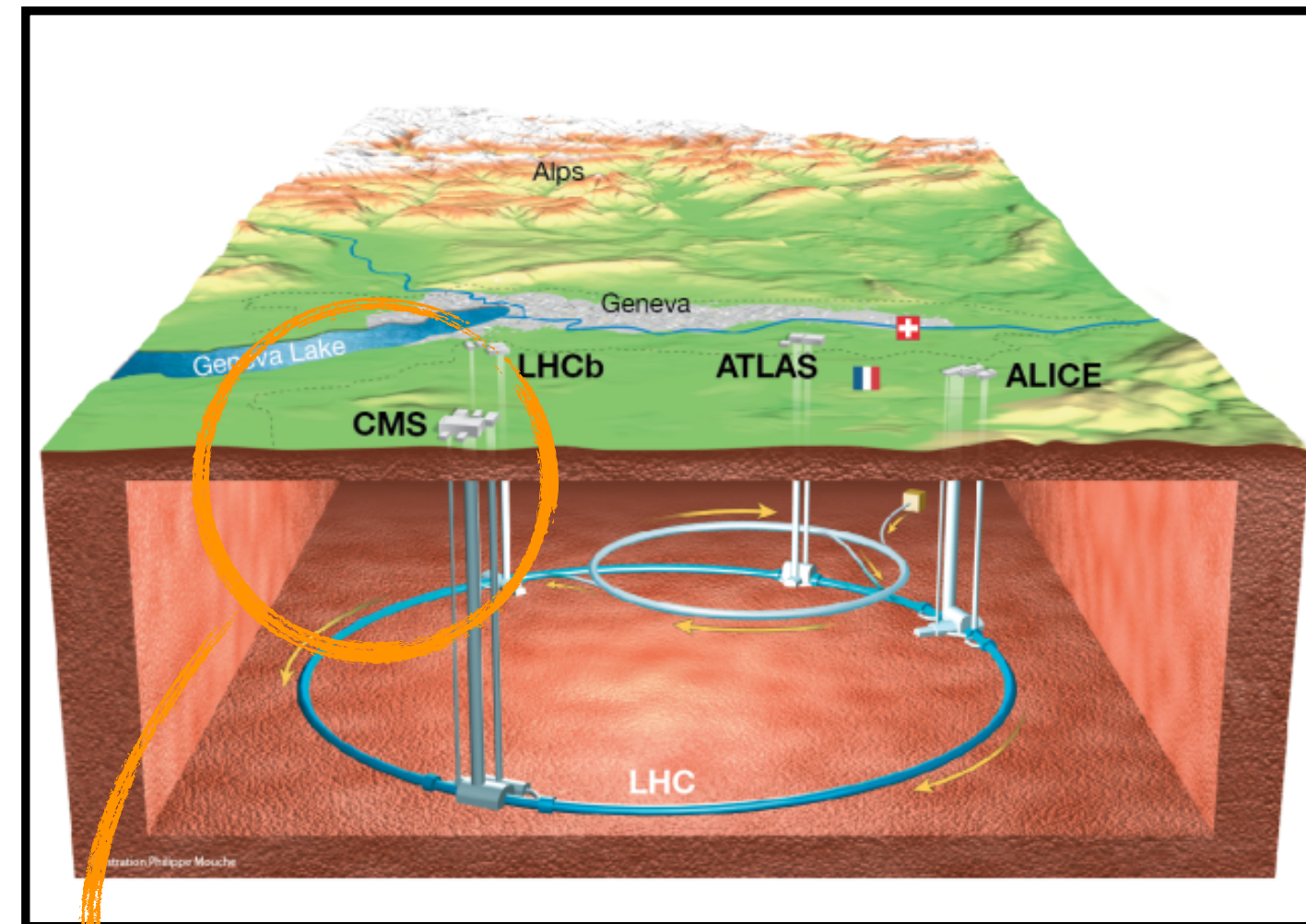
master thesis projects

22nd may, 2026

<https://indico.global/event/17991/>



The Large Hadron Collider



- CMS is one of the four major CERN experiments Located at Cessy (FR) along the **Large(st) Hadron Collider** (LHC)

- **LHC**: colliding two proton beams at the energy of 13.6 TeV

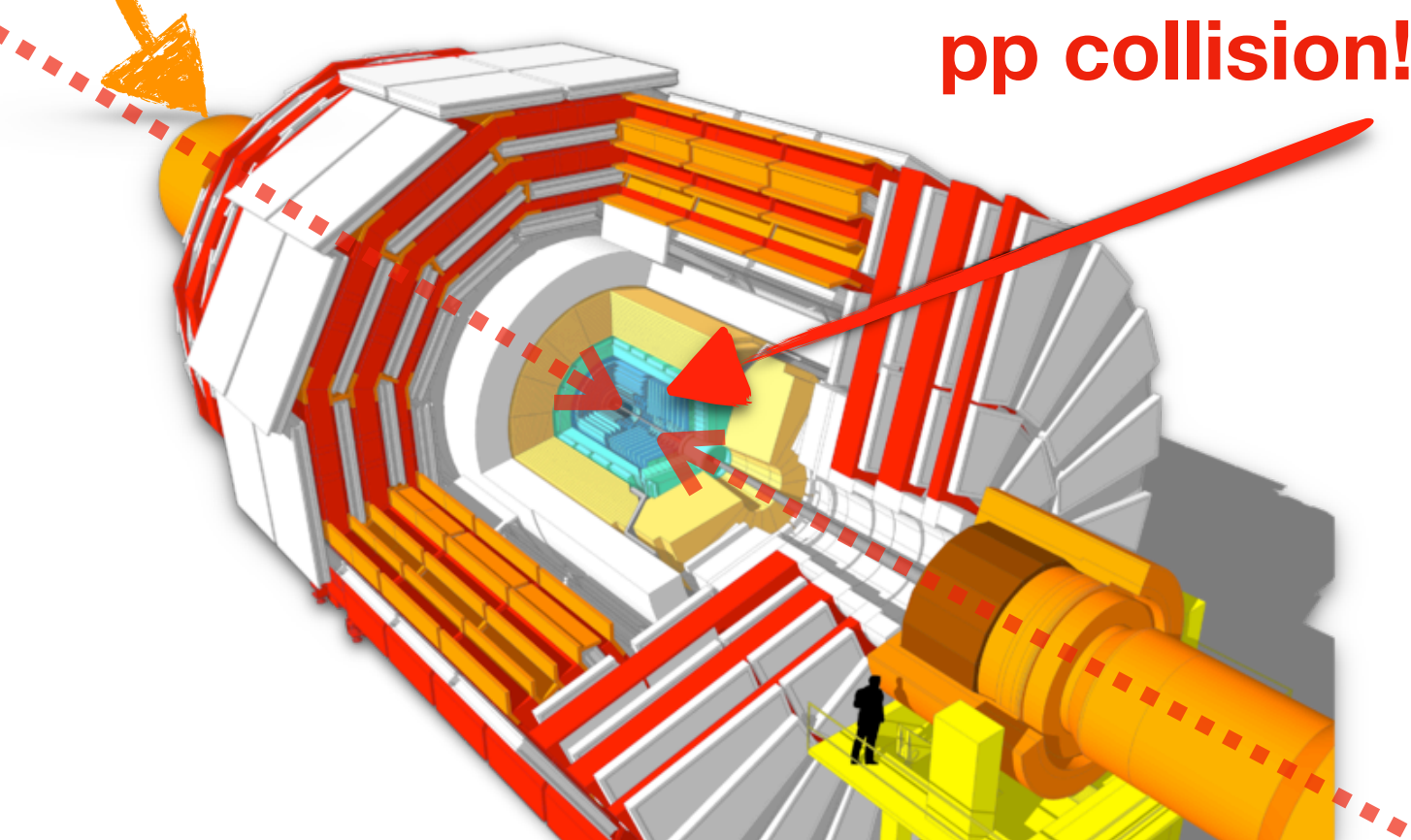
- ★ ~10 μm protons bunches colliding every 25 ns

- ★ 100 B protons per bunch, about 60 pp interactions per bunch-crossing:
highest luminosity every reached by a hadron collider (10^{34} Hz/cm²)

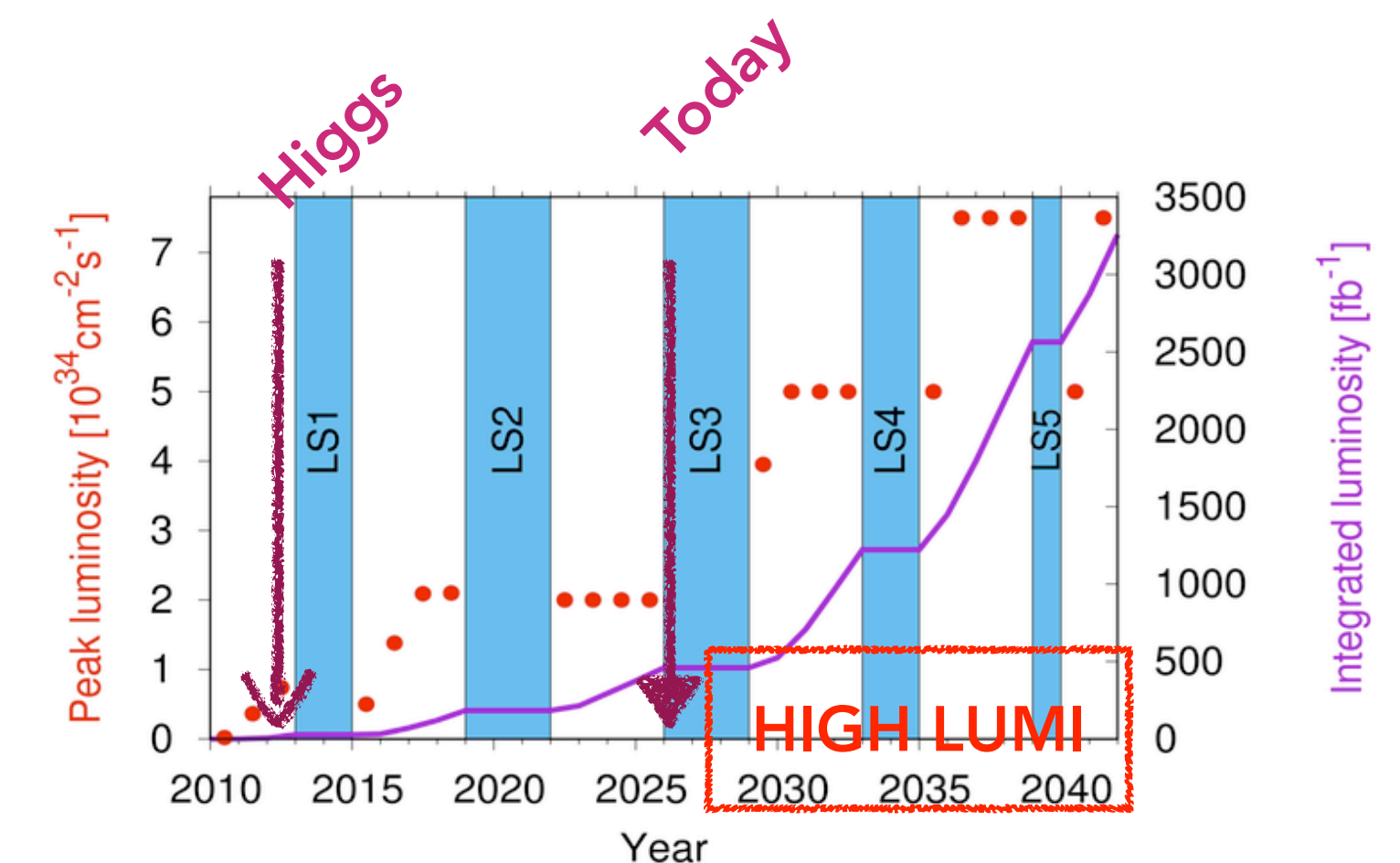
- ★ even higher (x5) in a few years: **high**

luminosity upgrade → opening new

possibilities in HEP research for the next 15 years!

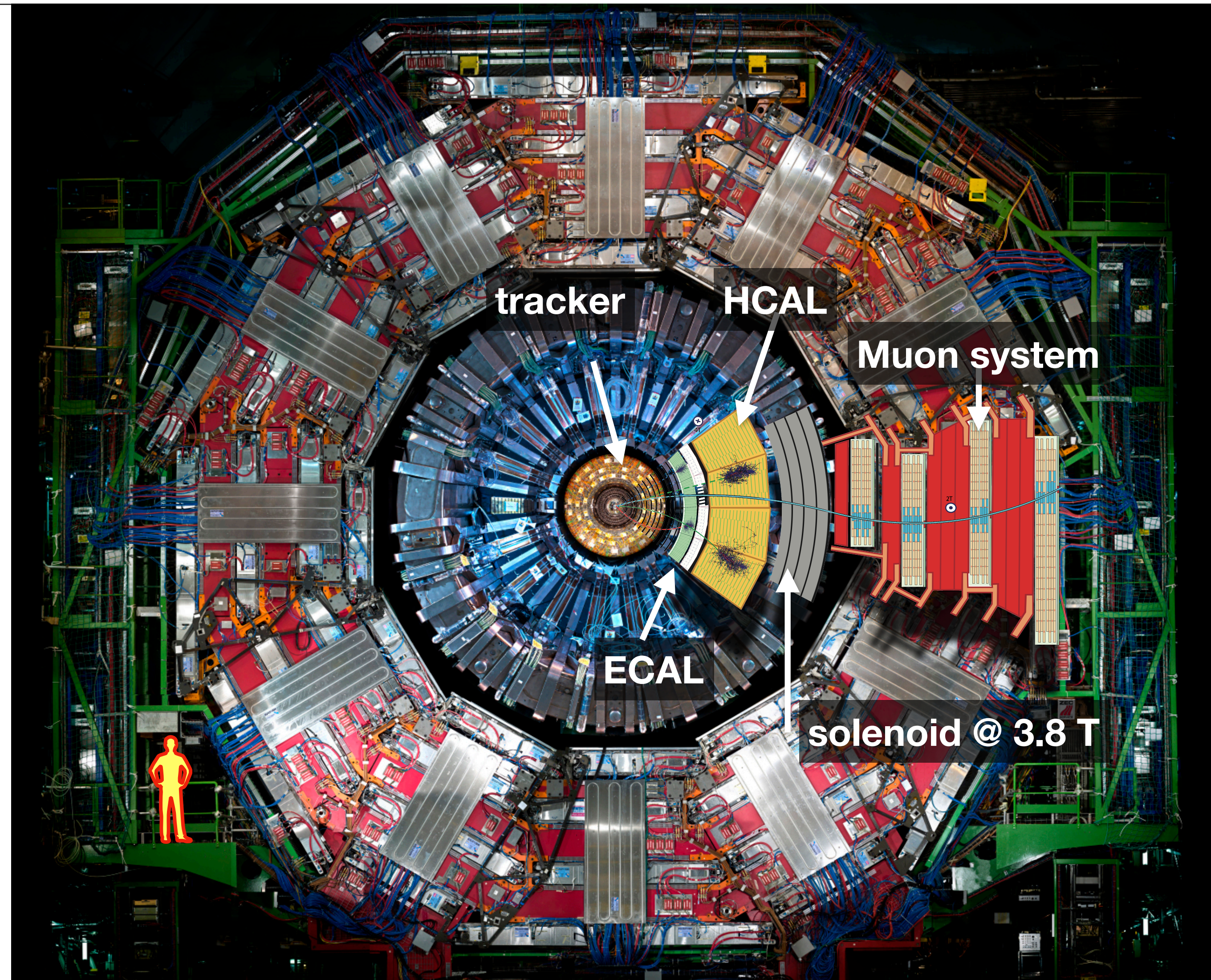


the proton beam pipe



The Compact Muon Solenoid

- 21m long, 15m high "cylinder" featuring a 3.8 T solenoid for bending charged particles
- **particles are detected when they traverse different detectors:** tracking planes, calorimeters, muon tracking planes
 - ➔ events (type of particles, energy, momenta, positions, ...) are reconstructed merging the information from the various detectors
- **general purpose experiment:** physics reach spanning from low-momentum physics to high-mass observables
 - ➔ higgs boson searches and studies, rare decays searches, precision physics, ...
 - ➔ **many of these fields (hardware, analysis, software) are covered in Milano Bicocca!**



Data analysis

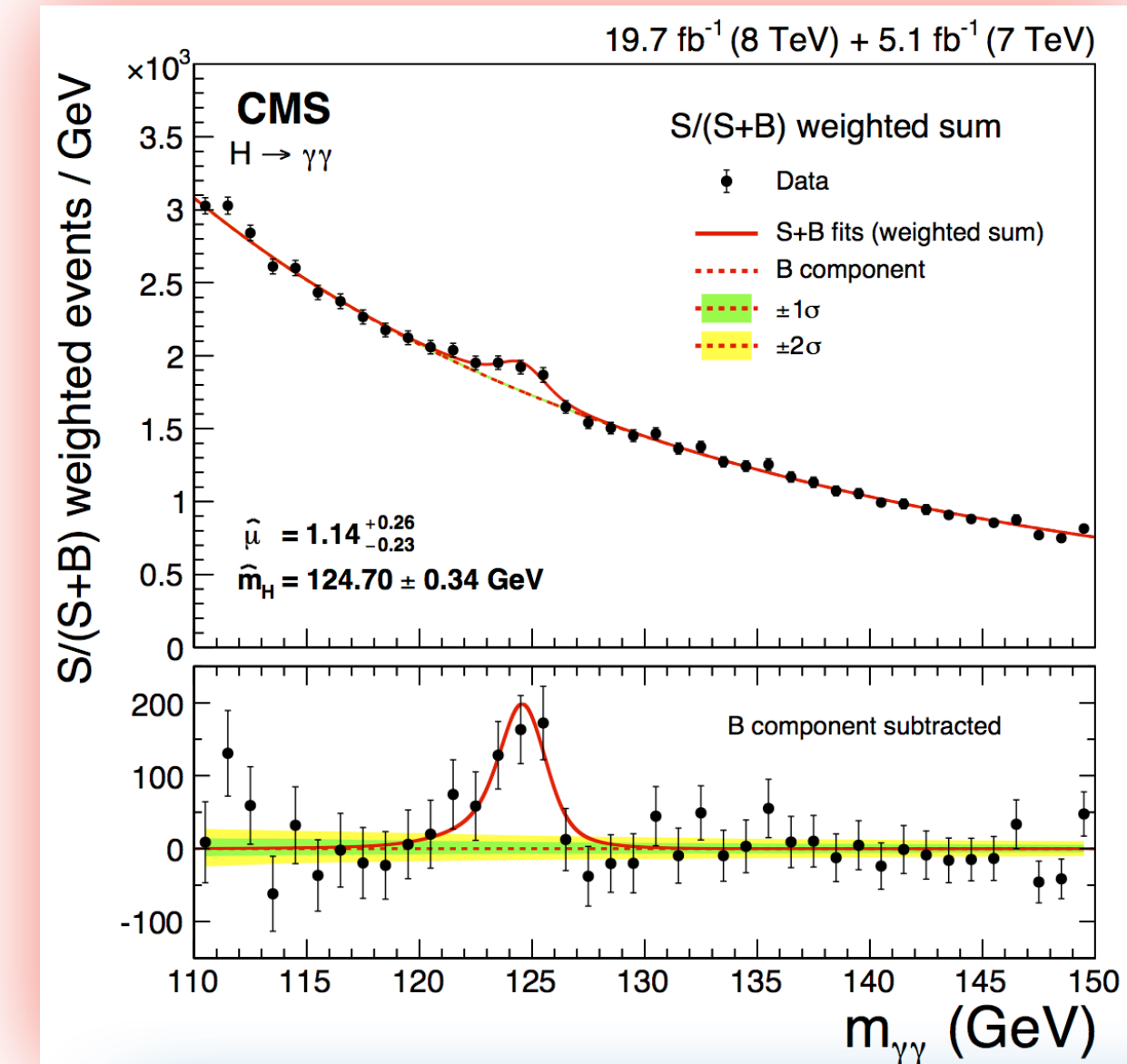
Standard Model and Beyond

- In 2012 ATLAS and CMS confirmed the existence of a **Higgs-like particle**. Is that the end?
 - do we know each of its properties? (no)
- additional **open questions remain**: where does the particles mass hierarchy come from? Is there a connection between quarks and leptons? What is dark matter? What is the SM validity? Does it hold at very high energies? ...?
- The goal of data analysis is to test the current knowledge of particle physics (the SM) and explore new possible scenarios (new physics)



Data analysis in a nutshell

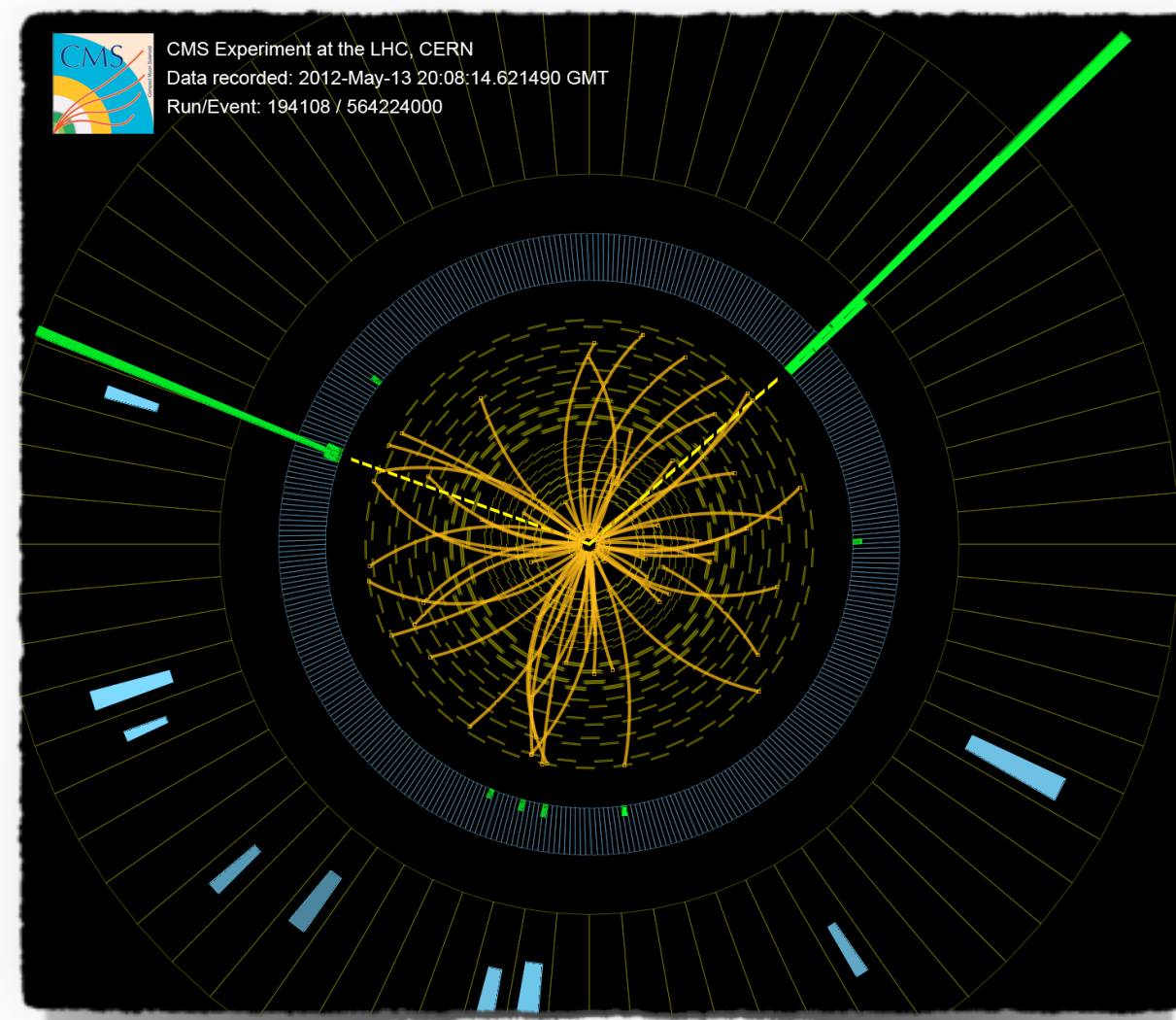
- Sub-detectors (trackers, calorimeters) generate an **electric signal when a particles interacts** within their volume
- several **algorithms are employed to reconstruct** and clean-up these signals
- **physics events are reconstructed** by identifying the number and type of particles produced during a collision, their energies and momenta
 - ➔ the observation can then be compared to several physics models (eg. SM)



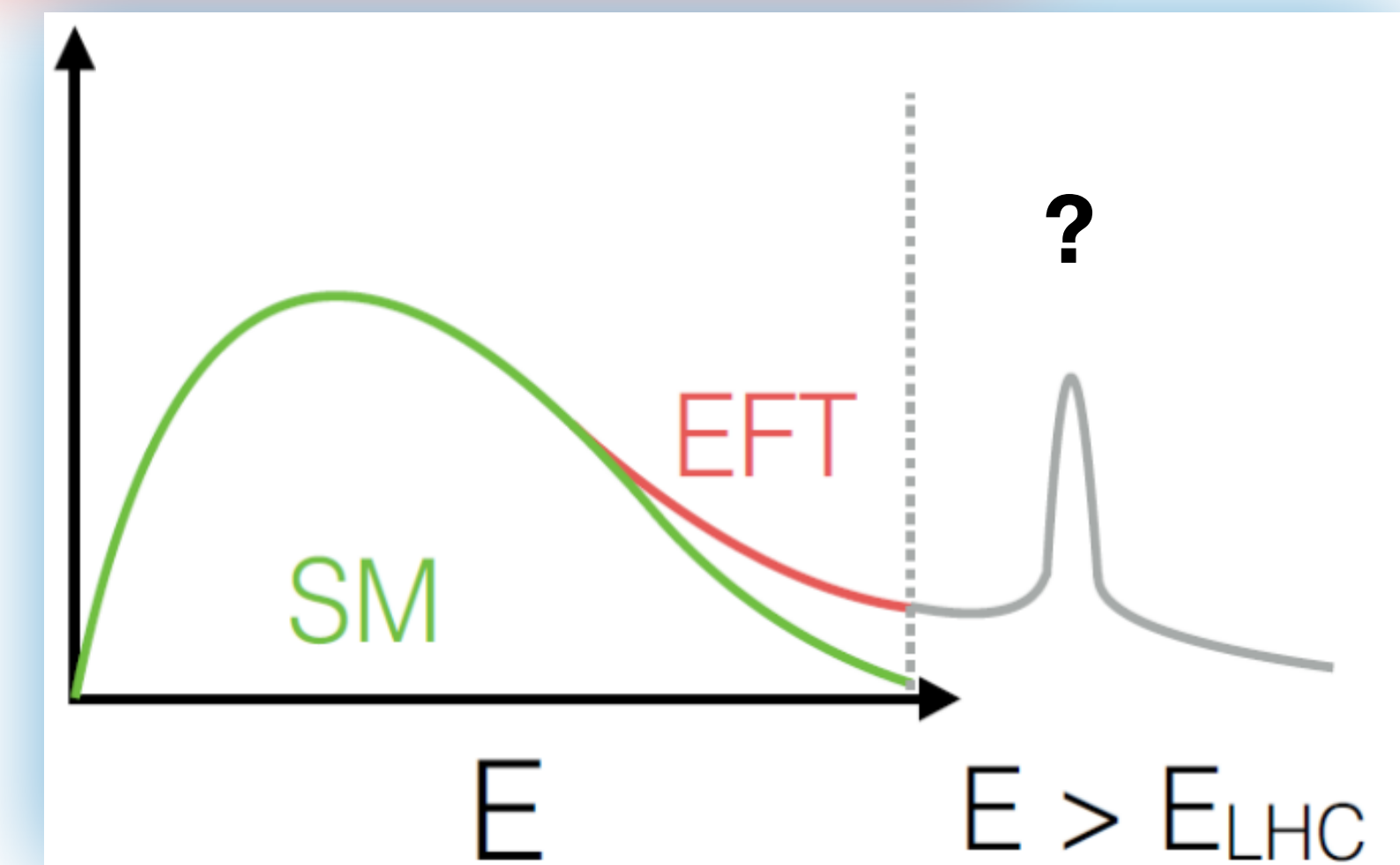
Higgs $\rightarrow \gamma\gamma$
invariant mass
plot

example: Higgs $\rightarrow \gamma\gamma$ event candidate (transverse plane projection)

two high energy deposits, compatible with a H event, are highlighted in green in this representation. The background of soft hadron tracks (yellow lines) and soft energy deposits (blue bars) are also shown



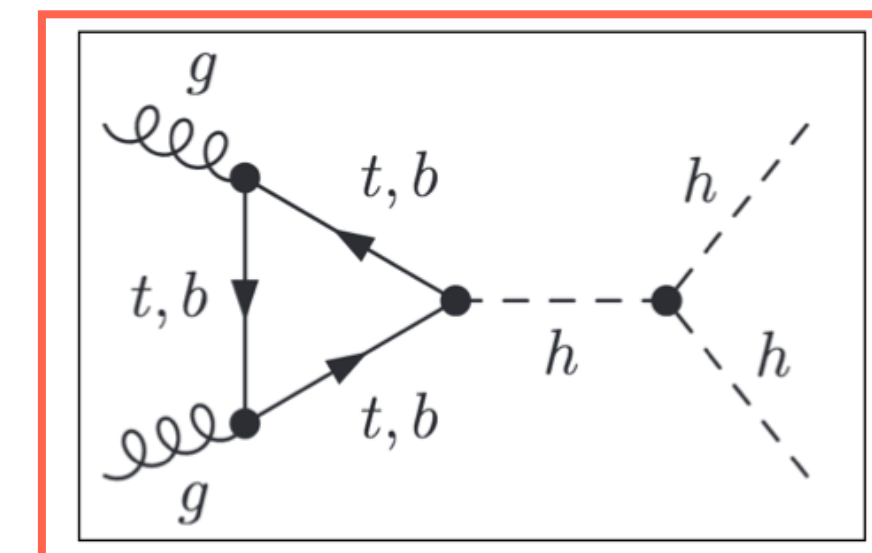
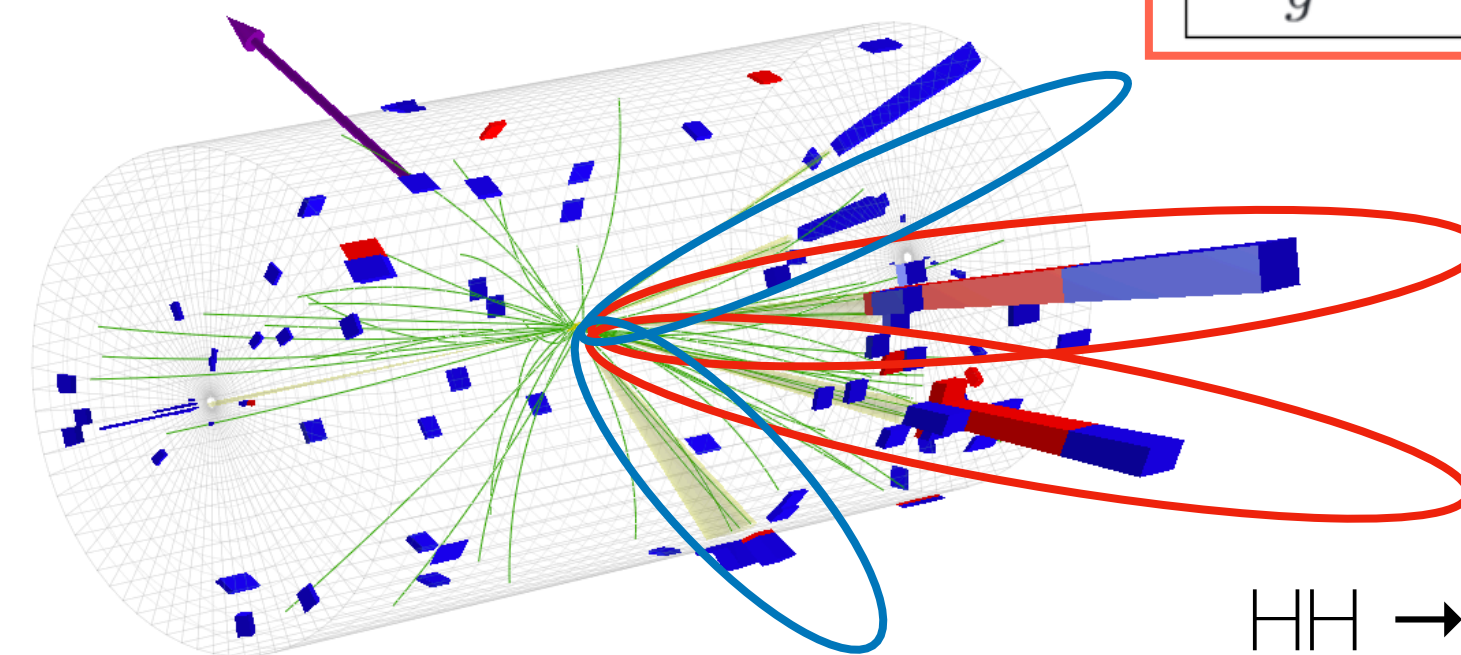
sketch: possible representation of new physics contribution as a distortion of the expected SM contribution



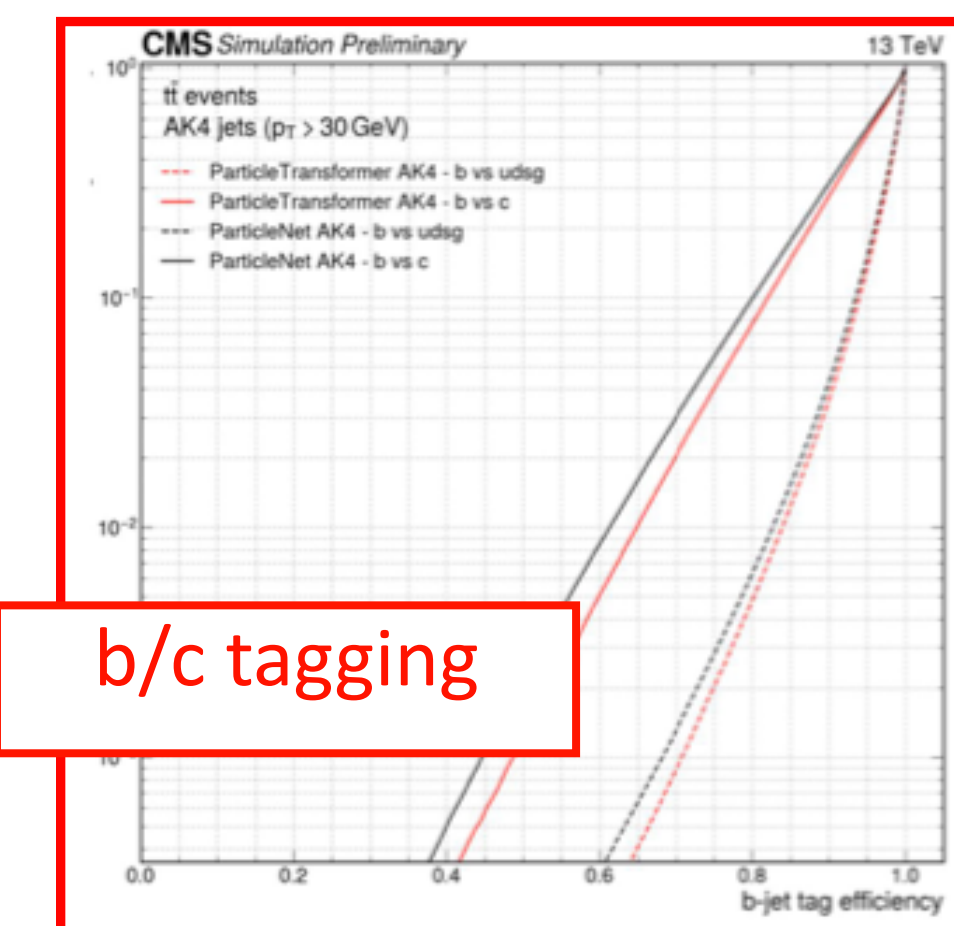
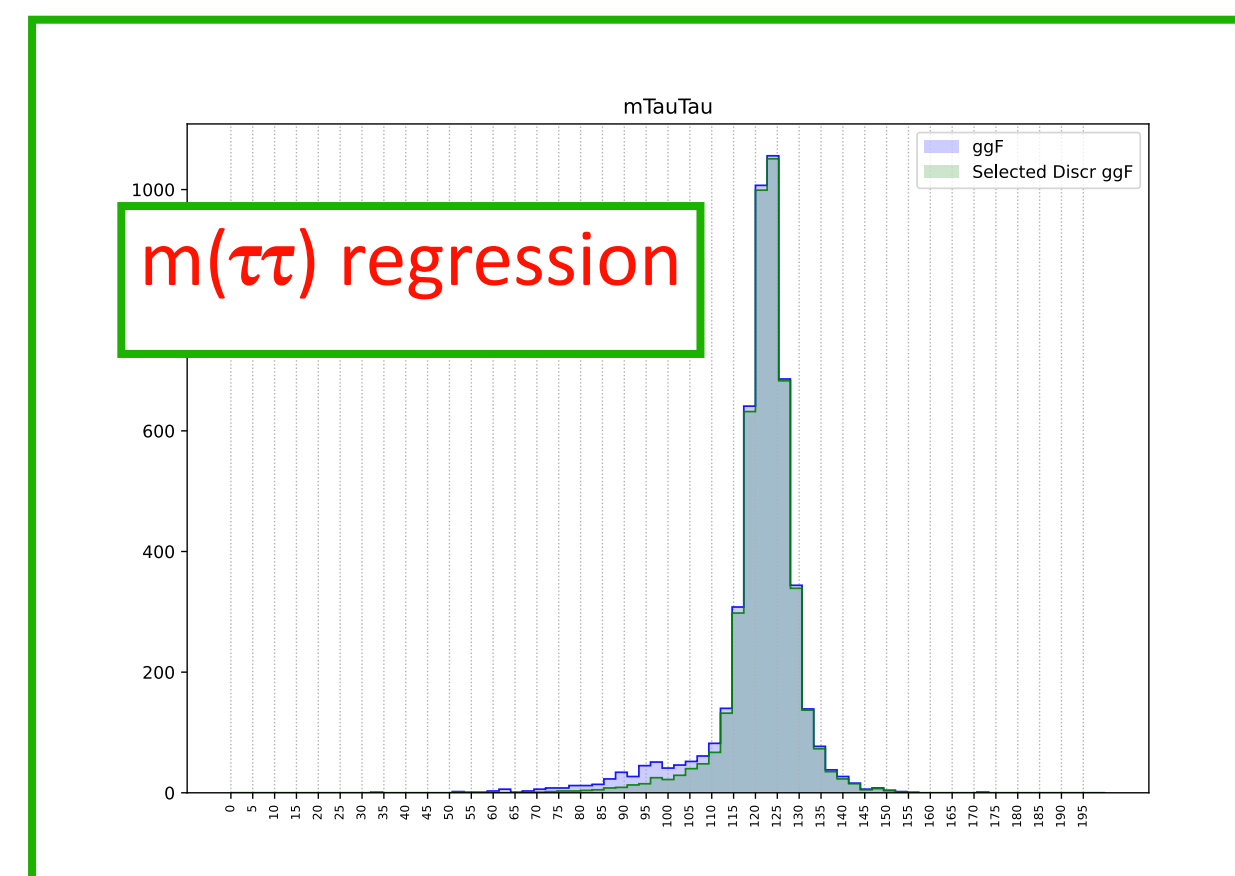
Higgs boson pair production

- measuring the **self-coupling (HH)** in CMS data
 - how does the Higgs boson interact with itself?
 - what can we learn about its potential?
 - is the observation compatible with the SM prediction?
 - search for **BSM physics contributions** (radions, gravitons, ...)
- in Milano Bicocca: $HH \rightarrow 4b$ and $HH \rightarrow bb \tau \tau$
 - development and ameliorations of the current selection and **reconstruction algorithms** for **background estimation and rejection** and **signal reconstruction**
 - mass regression** (via ML) for estimating the H mass from $\tau \tau$ events
 - jet tagging** with machine learning algorithms for increasing the purity of the selected signal candidates
 - EFT interpretation** of the observation

CMS Experiment at LHC, CERN
 Data recorded: Wed Oct 3 11:09:52 2018 UTC
 Run/Event: 323954 / 16341342
 Lumi section: 9
 Orbit/Crossing: 2209447 / 3295

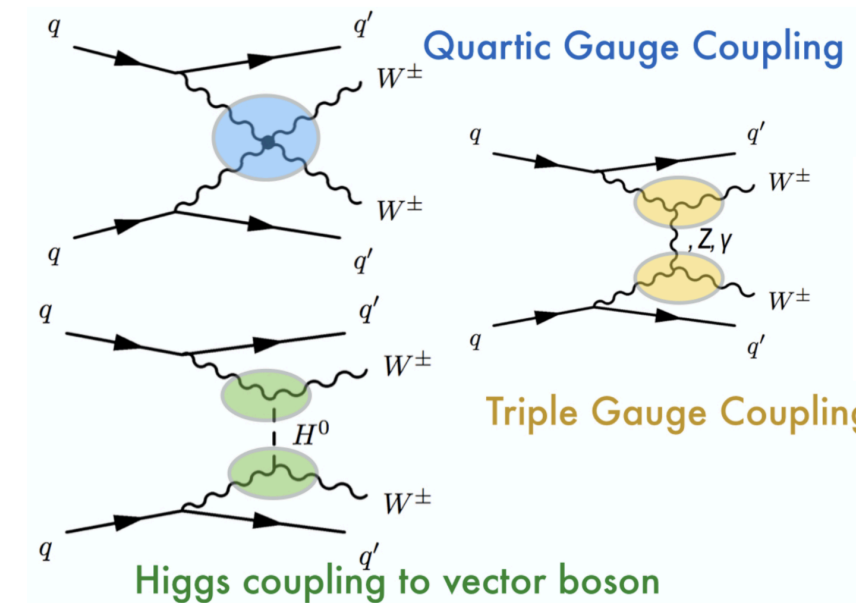


$HH \rightarrow bb \tau \tau$
 signal candidate in
 CMS data

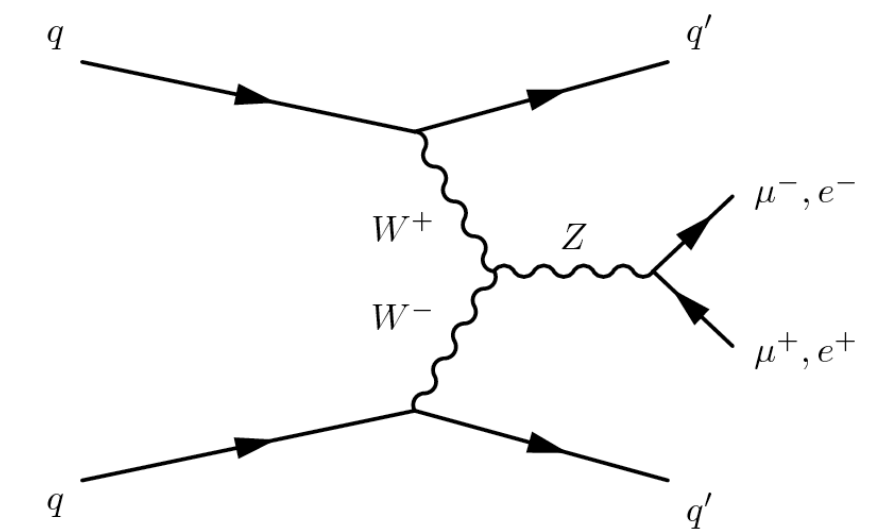


Vector Boson Scattering / Fusion

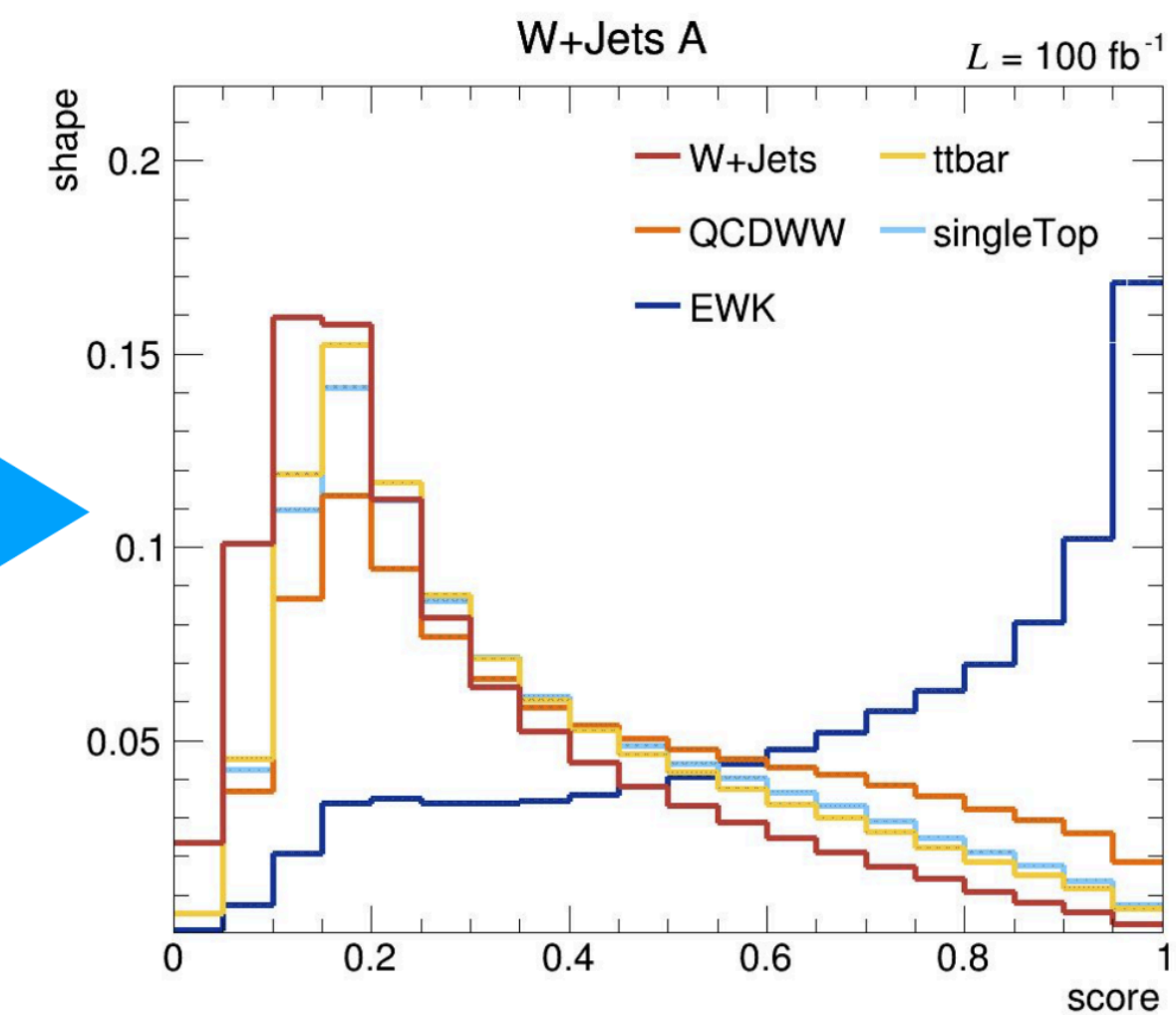
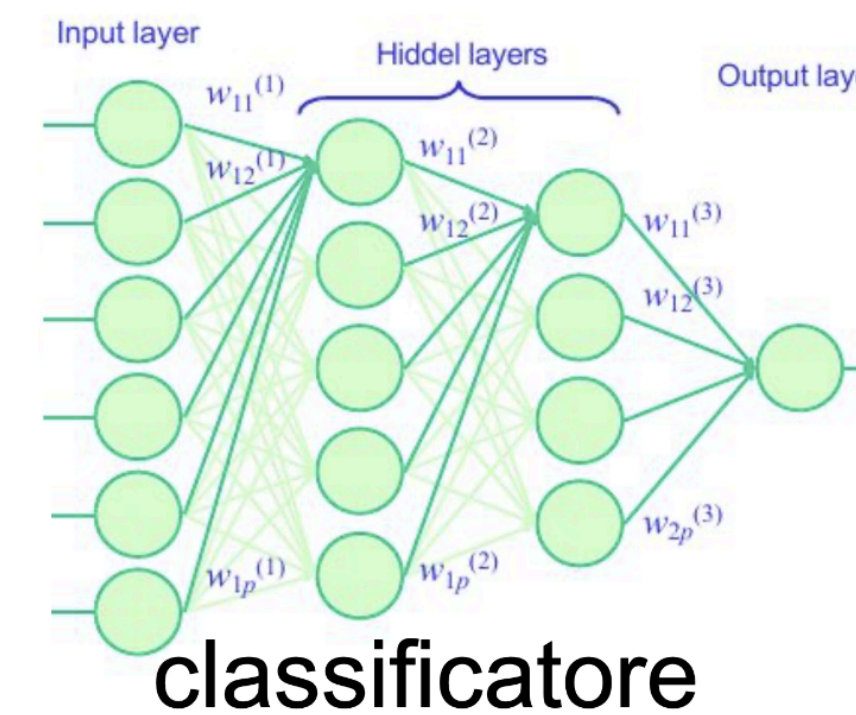
- Study the boson pair production in pp collisions (W^+W^+ , W^-W^+ , WZ)
 - rare processes where new physics can give a significant contribution to the dynamics of the production rate: **validity test for the SM**
 - **EFT interpretation** of the results
- ▶ optimisation of the **EFT fit**
- ▶ development of **ML techniques for isolating new physics contributions**



vector boson scattering

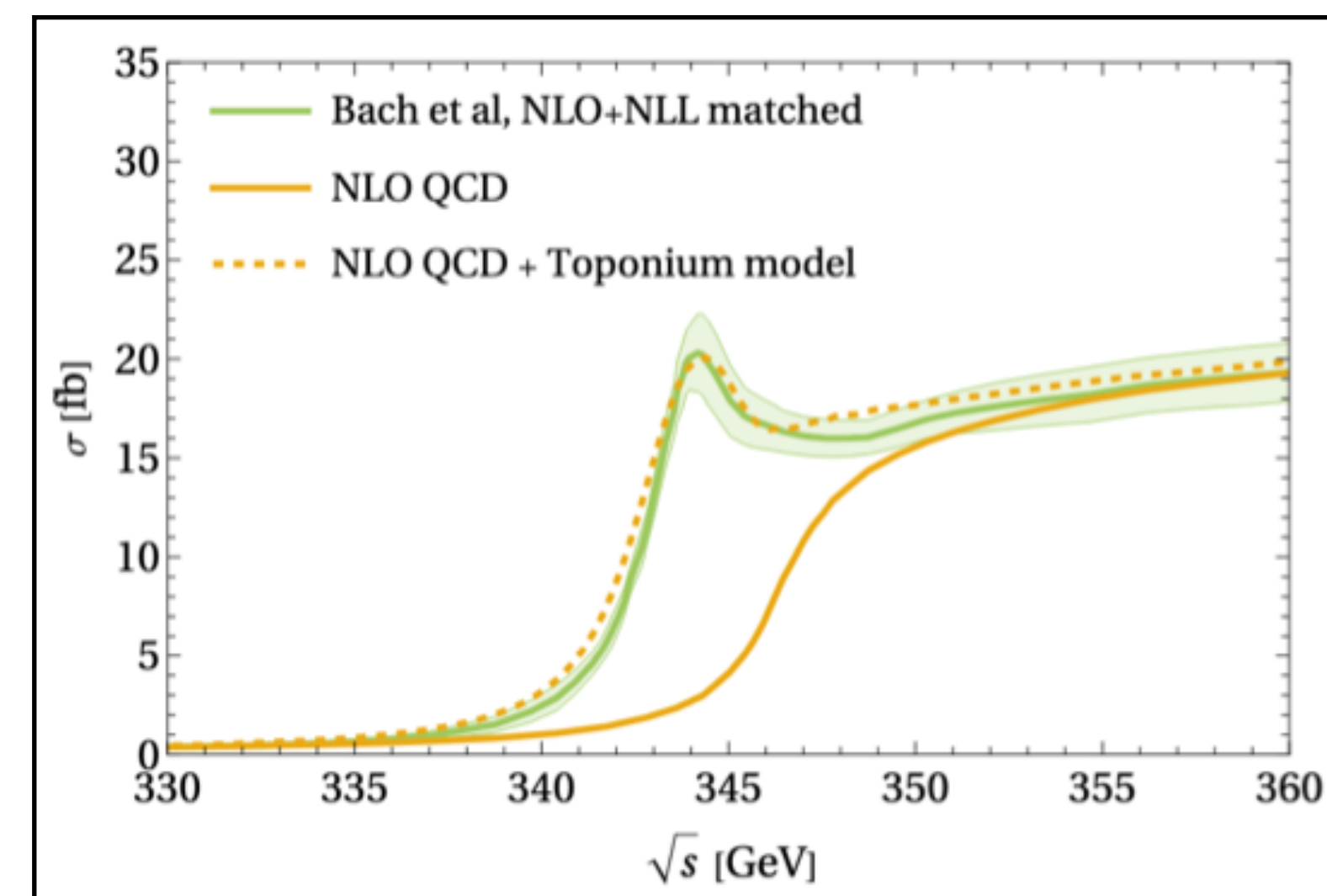


vector boson fusion



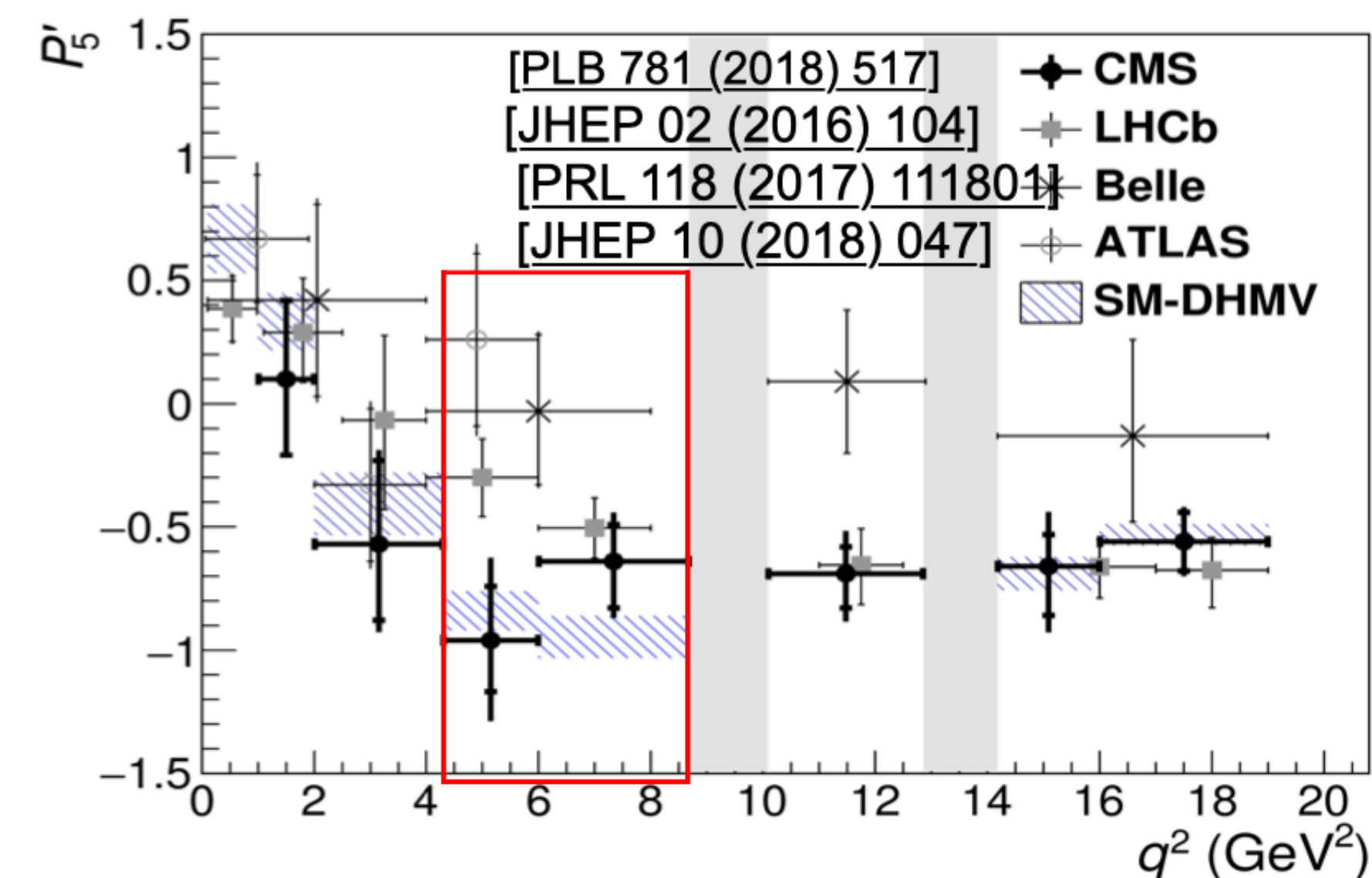
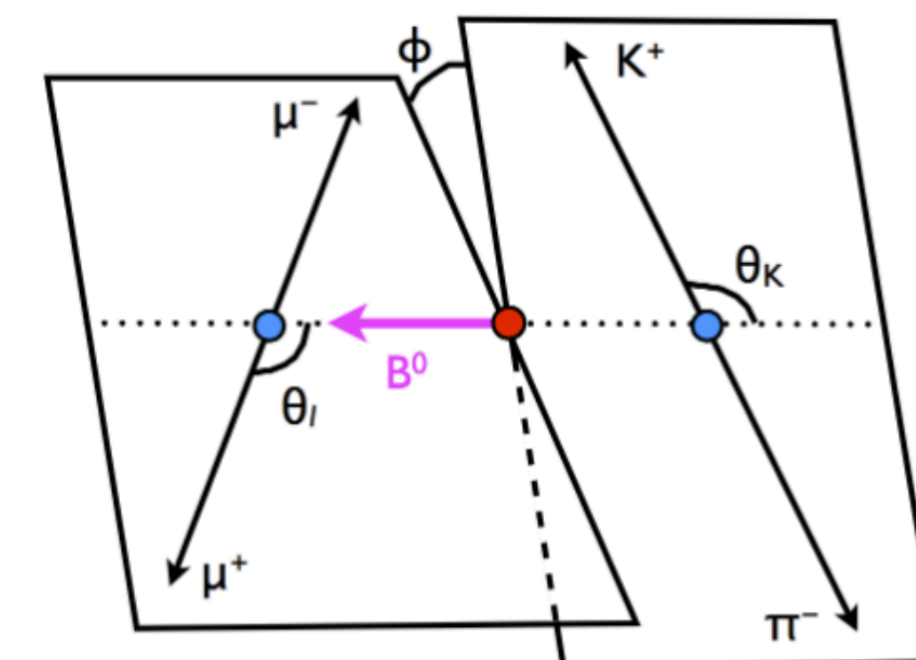
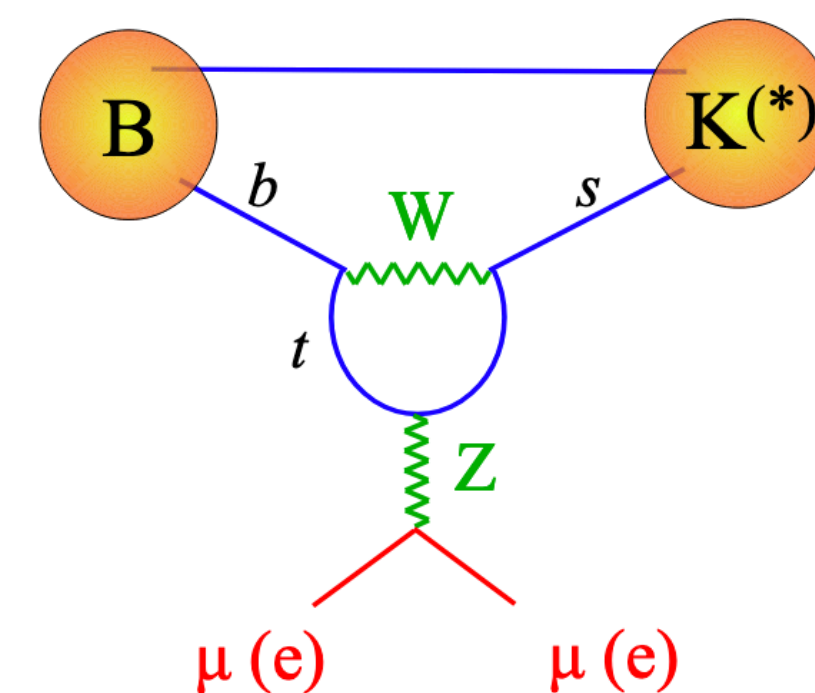
Toponium (η_t) characterisation

- recently, CMS observed an **excess of events close to the invariant mass threshold of top-top production** (~ 350 GeV)
 - possible quasi-bound state (10^{-25} s) of top quark pairs (toponium η_t)
- **this project:** confirm the existence of a real quasi-bound state studying the properties of its final states
 - upper limit on the rare processes:
 - $\eta_t \rightarrow \gamma\gamma$, $\eta_t \rightarrow \gamma H$, $\eta_t \rightarrow bb$ ll
 - estimation of the top mass corrected for the QCD binding energy



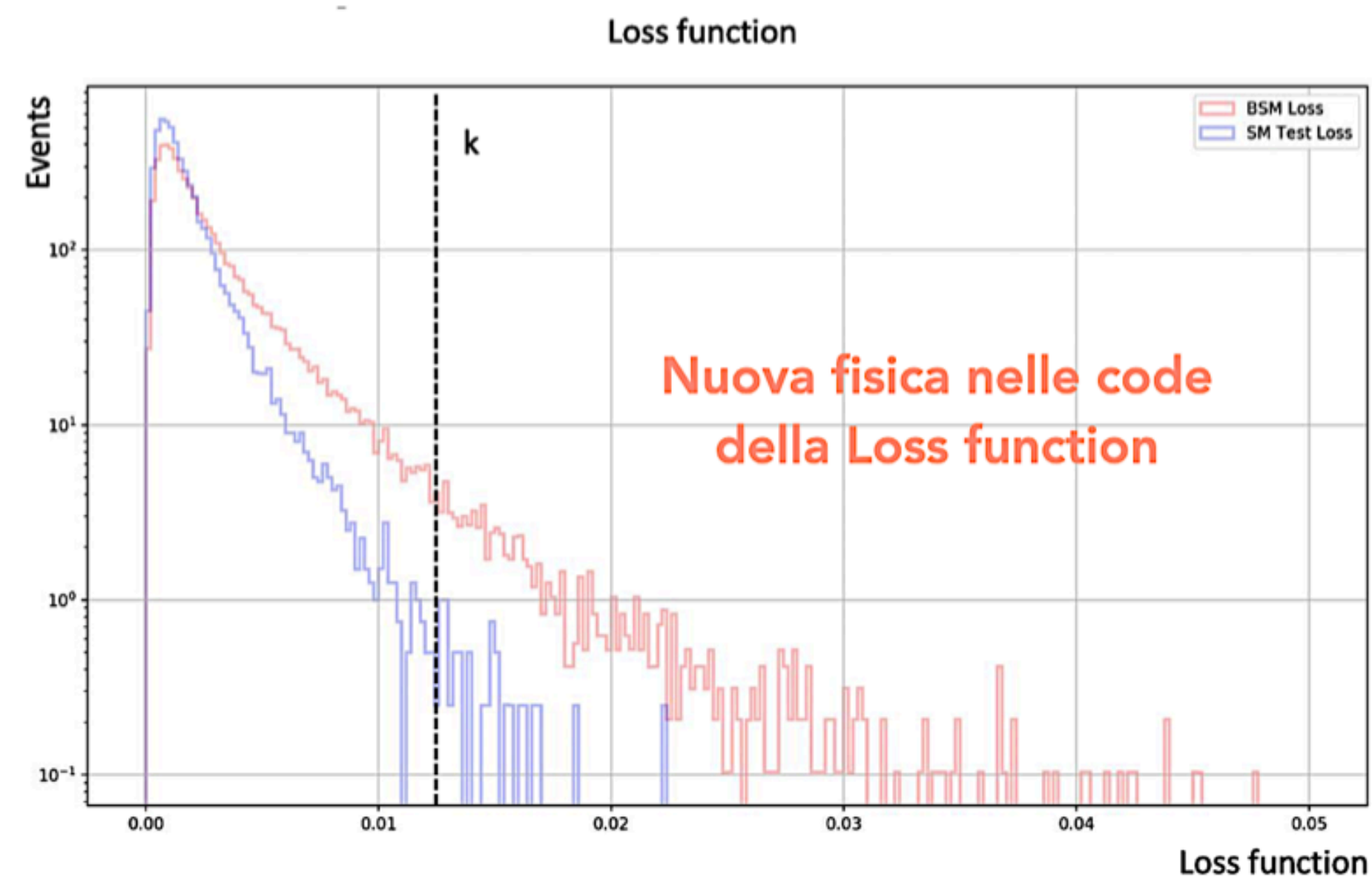
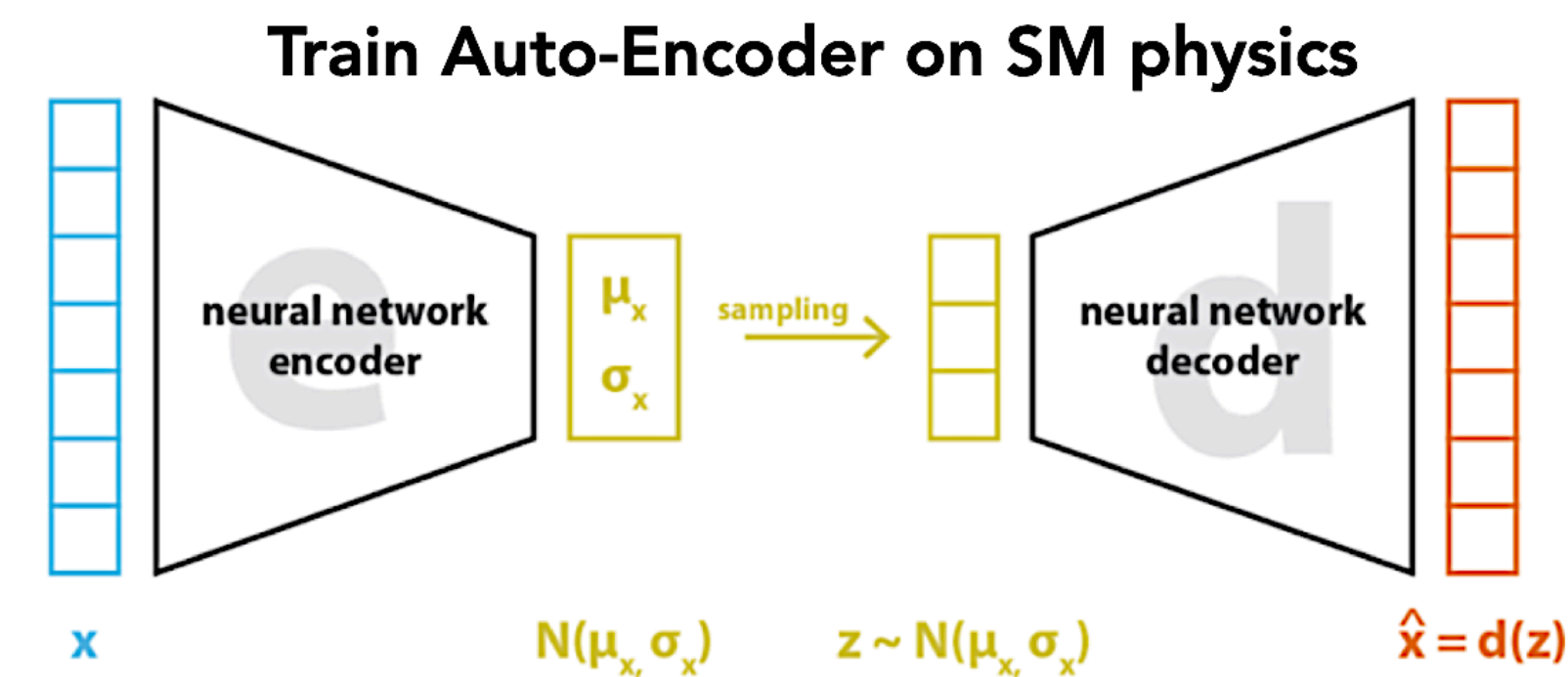
$b \rightarrow s$ in B^0 decays

- Study of **$b \rightarrow s$ angular distributions** to search for deviation from the SM predictions
 - tensions with the SM reported by LHCb
 - the feedback of many experiments is crucial
- Targeting the final state **$B^0 \rightarrow K^*(K \pi) \mu\mu$** and **$B^0 \rightarrow K^*(K \pi) \tau\tau$**
 - **precision analysis**: requires the understanding and correct estimation of all the contributing processes (signal + background)
 - employment of **machine learning** techniques for data correction, background rejection, signal reconstruction



Anomaly detection

- **Auto-encoders:** *unsupervised machine learning* technique for recognising whatever is not compatible with the given definition of "standard model" (anomaly detection)
 - alternative approach wrt. to classical analyses
 - search for deviation from the SM and test of EFT predictions

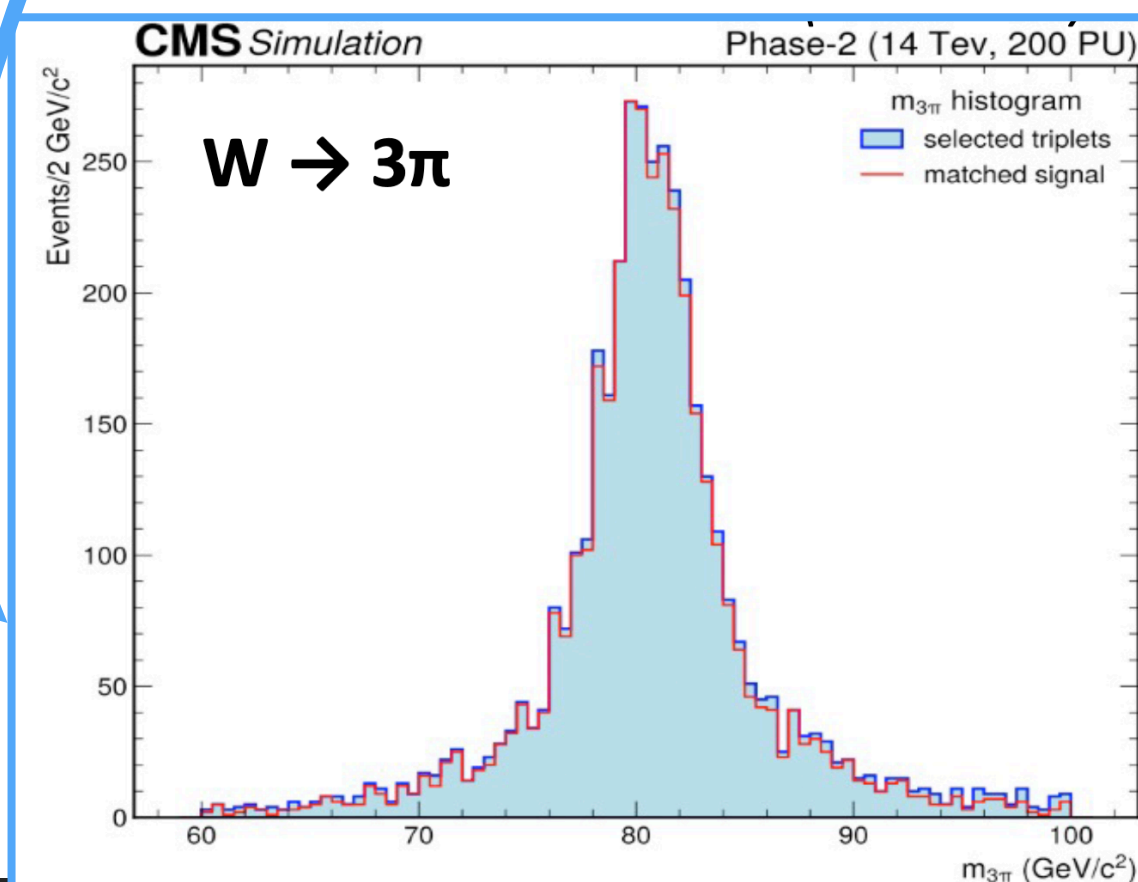


FPGA development

An FPGA cluster has been installed in Milano!

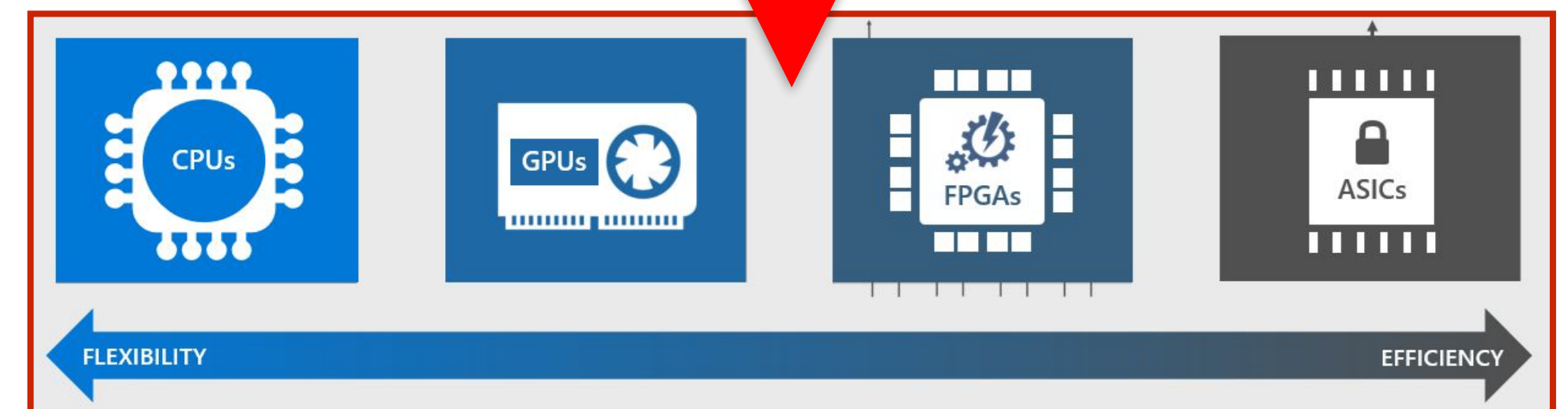
Development of new ML-based trigger strategies based on FPGA reconstruction

- The project
 - Dedicated L1 triggers for the $H \rightarrow \tau\tau$ analysis
 - mass regression for $H(\tau\tau)$
 - hadronic τ identification with CNN
 - Search for rare hadronic decays such as $W^\pm \rightarrow 3\pi$ entirely done at trigger level!



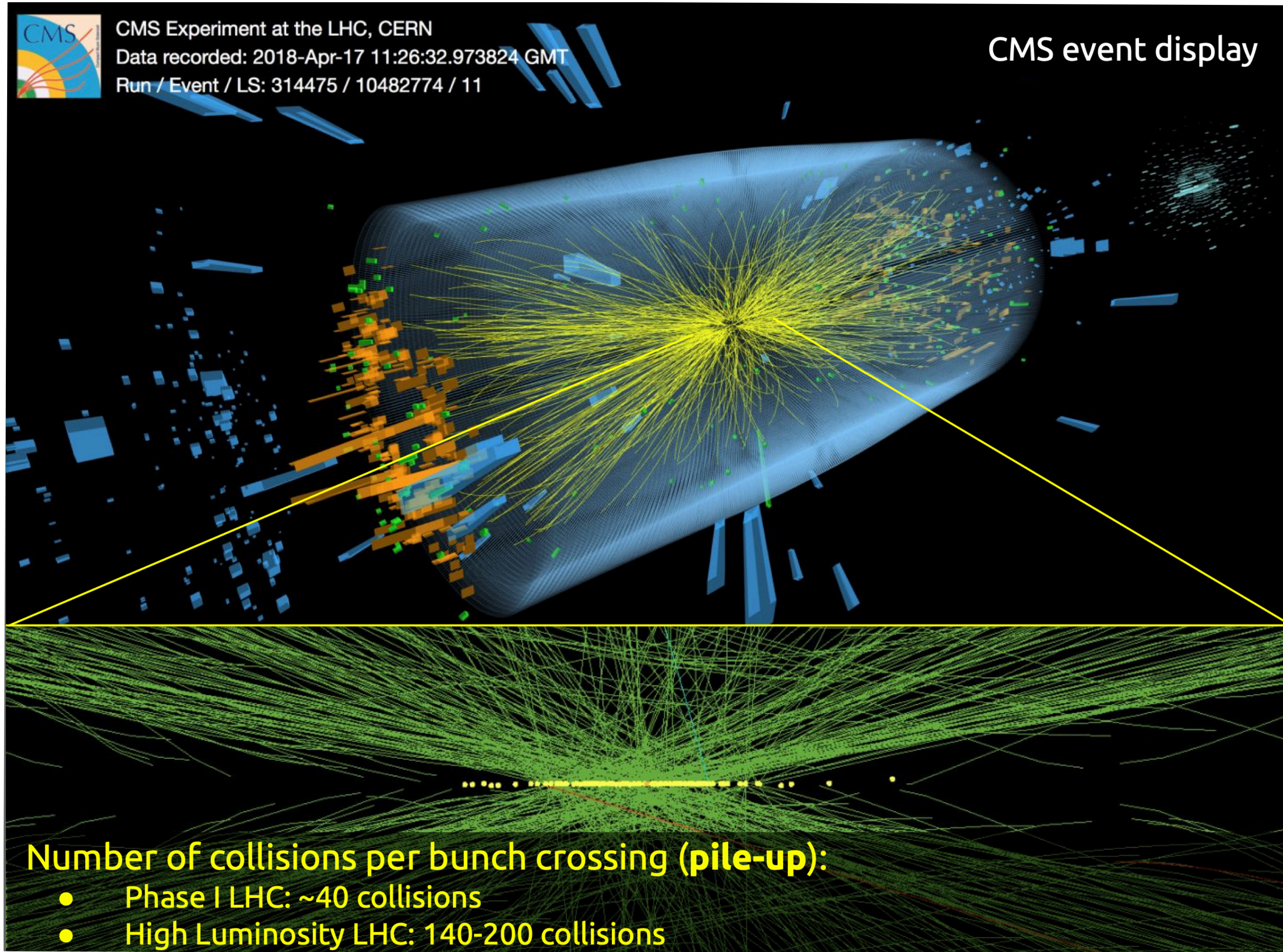
Ultra fast simulation on FPGA for HEP

- Producing a large amount of simulated data is crucial for HEP analyses
- ML can be used for simulating physics events \rightarrow Generative Adversarial Networks (GANs)
- FPGA are very fast devices
- can be used for ultra-fast simulations of physics events



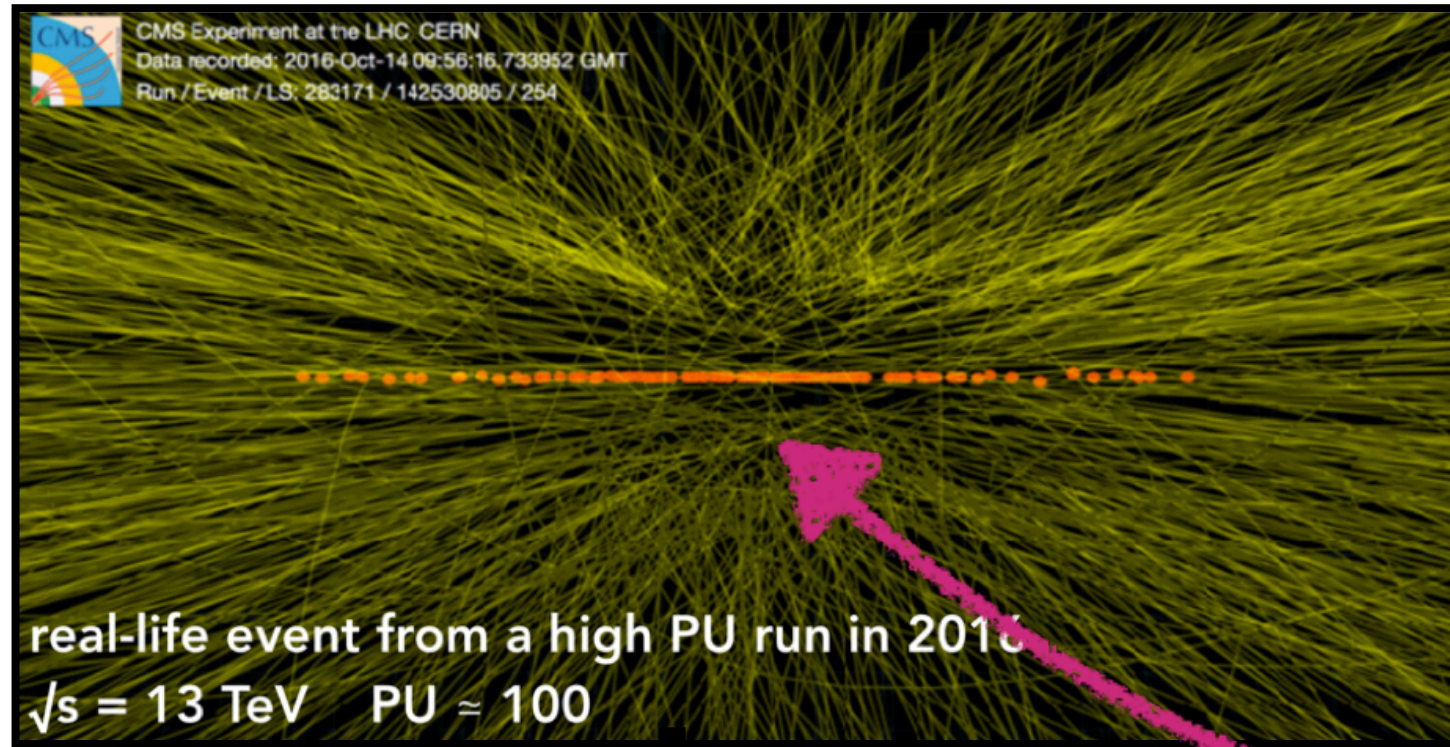
CMS upgrade

High luminosity

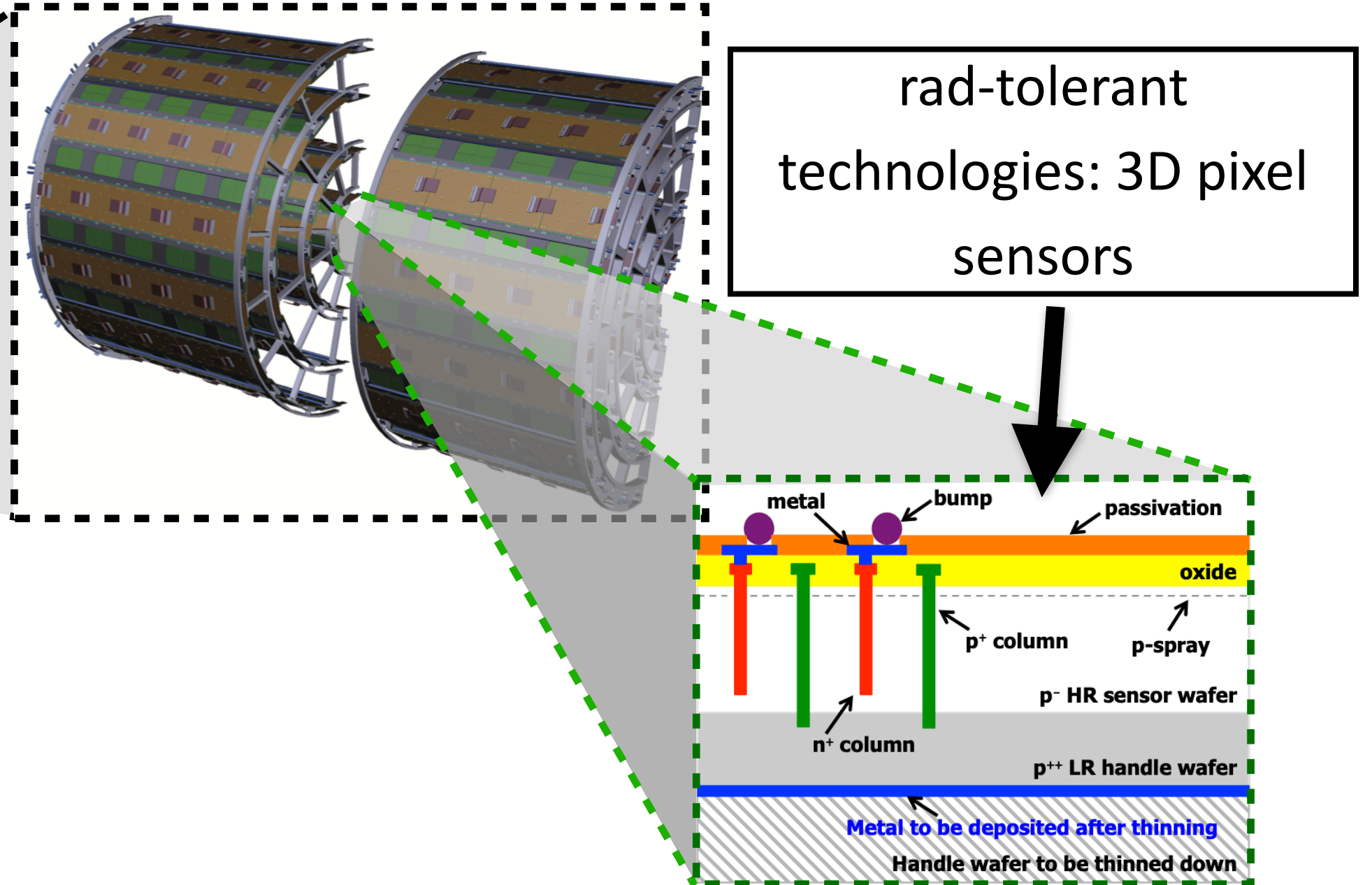
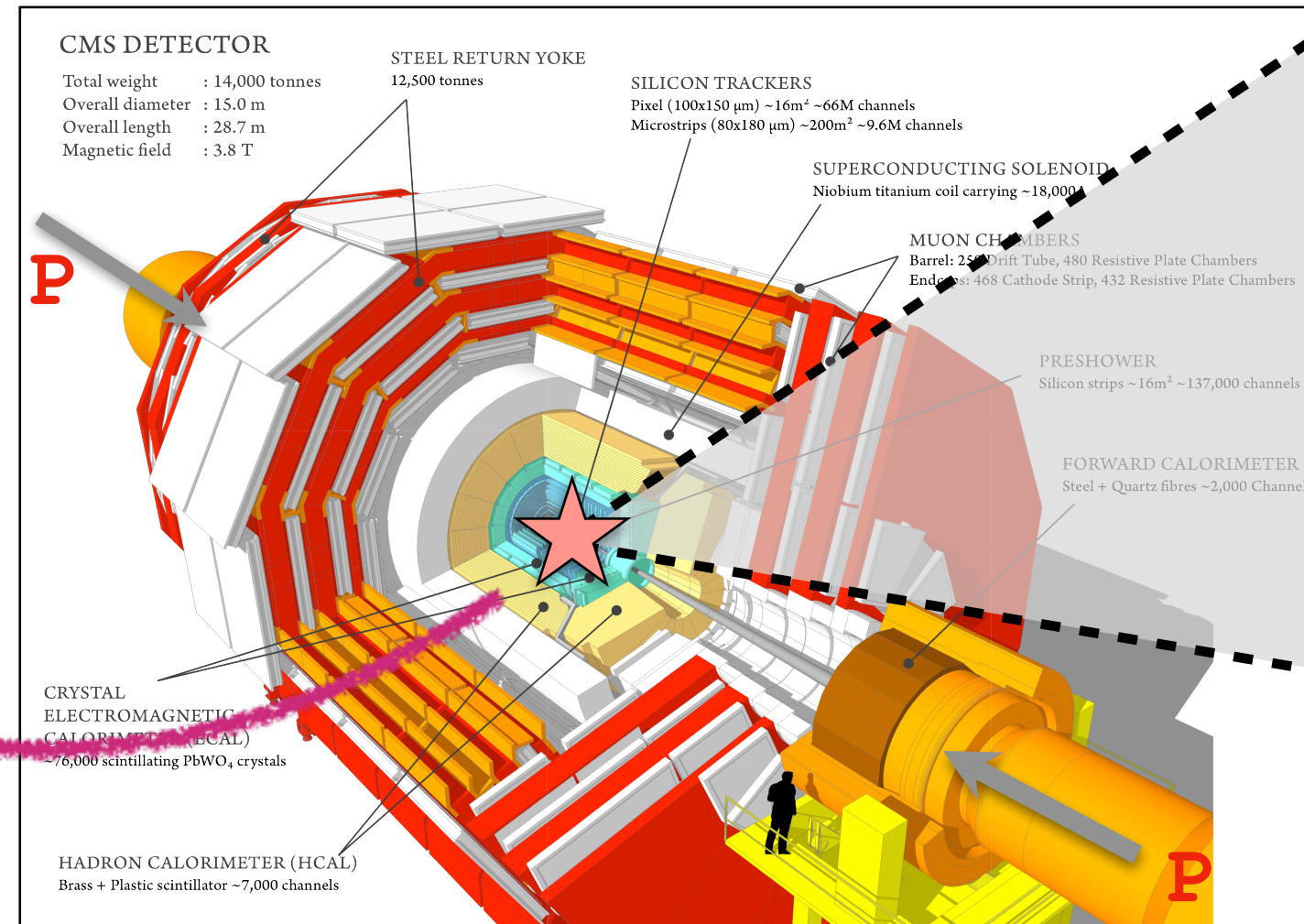


- **HL-LHC**: 3x the current luminosity
- **In-time pileup**: number of proton-proton interactions in a single bunch crossing
 - pileup now: ~60
 - pileup @ HL-LHC: 140-200
- ▶ large radiation fluency through the detectors
- ▶ major challenge for event reconstruction
- ▶ **new fast and rad-hard detectors are required**

The pixel tracker upgrade



HL-LHC: up to 200 pp interaction per event!



Project goal

- replace the current pixel tracker with one with a larger granularity (x6 smaller pixels), better coverage, more radiation tollerant
- **crucial steps happening now: we are getting close to the final production of the detector**
- characterisation of prototype modules and system tests and integration
- software development for data acquisition

Contatti: simone.gennai@mib.infn.it mauro.dinardo@unimib.it paolo.dini@mib.infn.it sandra.malvezzi@mib.infn.it

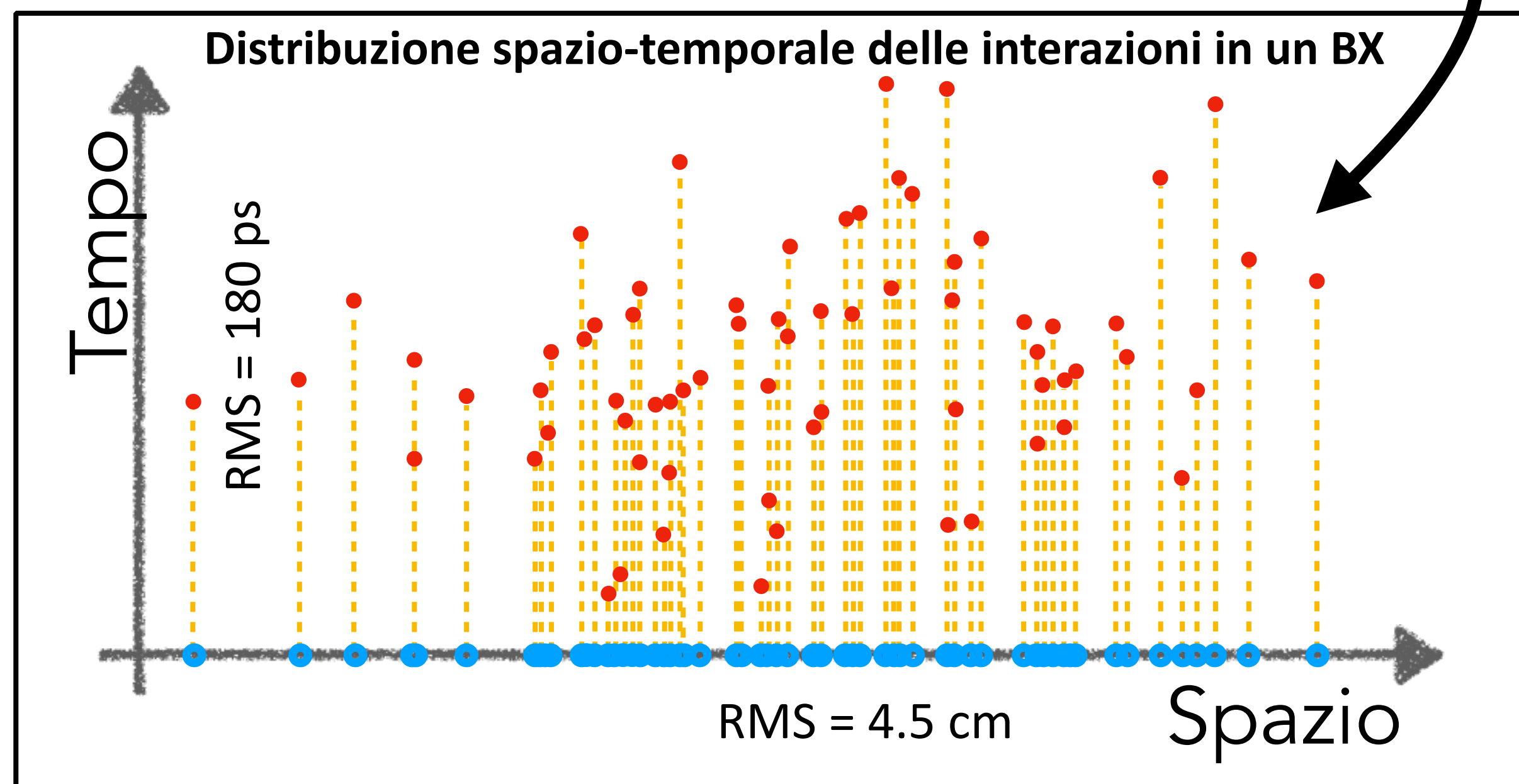
MIP timing detector (MTD)

MTD is a new detector thought for pileup mitigation in the dense HL-LHC environment

activity in Milano (and at CERN)

- **goal:** add the time information to charged particles produced in pp interactions $\sigma_t \sim 30\text{-}40$ ps
- **consequence:** 4D (timing + spatial) event reconstruction
- **MTD** represent an **intriguing challenge** in HEP!

- **Assembly, characterization and system-test** (cosmic rays, radioactive sources, lasers) of MTD modules in U2!
- Software development for **detector control** and **data acquisition**
- detector simulation and event reconstruction **optimization**



Il rivelatore sarà
realizzato in U2!



post LHC proposals

MAXICC: development of a new calorimeter

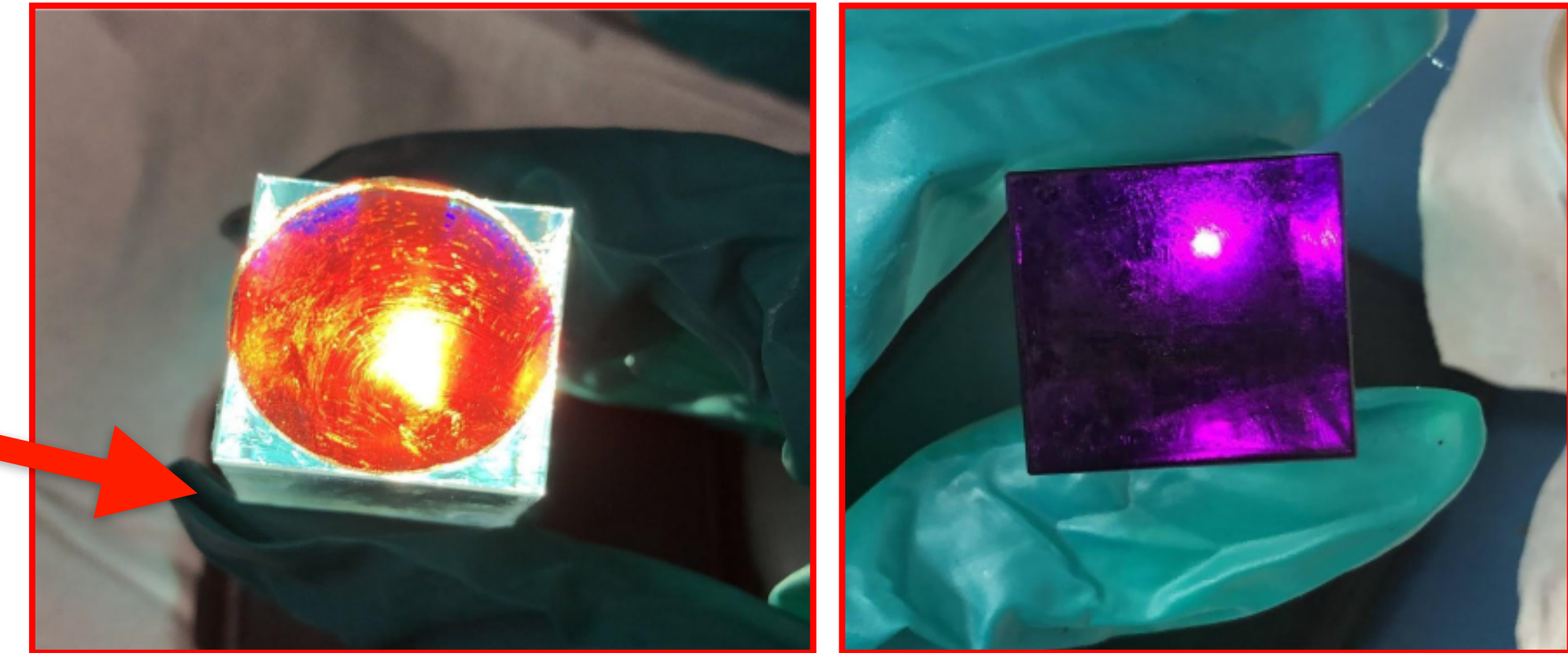
- **goal-1:** open new frontiers on precision SM measurements
- **goal-2:** reach higher energies

possible options

- Collider e^+e^- with $E_{CM} \sim 250$ GeV and $L = 100$ km
- Collider pp with $E_{CM} \leq 100$ TeV and $L = 100$ km

Milano Bicocca → Maximum information crystal calorimetry (MAXICC)

- A e^+e^- collider for precision Higgs measurements will be the priority
- Milano Bicocca is leading the development of a high-precision 6D calorimeter (position, energy, timing, particle identification) for increasing the reach of future experiments
- **Thesis opportunity:** contribute to the development and study of crystals and SiPMs:
 - Lab activity
 - Studies on test beams (CERN, DESY, Fermilab)
 - Development of dedicated simulations



Contatti: marco.lucchini@unimib.it andrea.benaglia@mib.infn.it

Opportunities in HEP

- **Stimulating international work environment**, opportunity for internships abroad (e.g. CERN, FermiLab, DESY, LLR, ...)
- Opportunity to make a significant contribution to complex and high-impact analyses
- post graduate: eventually acquire real **responsibility within the CMS experiment** and actively contribute to its operations
- Skills that can be acquired (for High Energy Physics and beyond!)
 - Ability to solve complex problems
 - Use of modern programming languages (e.g., C++ and Python)
 - Use of software for **data manipulation, data analysis, and machine learning**



CMS Experiment at the LHC, CERN

Data recorded: 2022-Jul-05 14:48:56.743936 GMT

Run / Event / LS: 355100 / 51596902 / 53

*Grazie per
l'attenzione!*

