UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

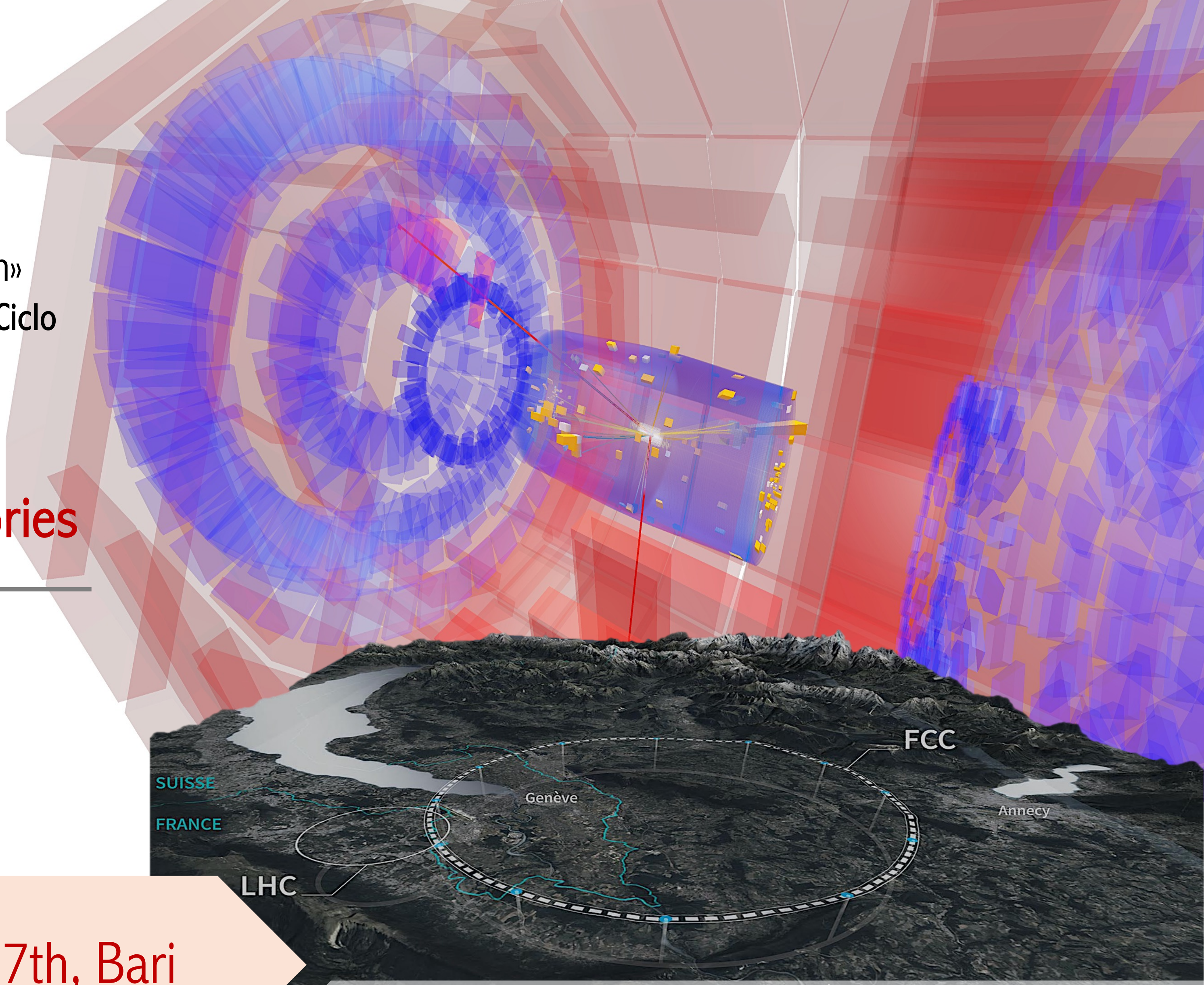
Dipartimento Interateneo di Fisica «Michelangelo Merlin»

Dottorato di Ricerca in Fisica Nucleare, Subnucleare, XXXVII Ciclo

# Search for HH→bbμμ with the CMS experiment and future Higgs boson factories

PhD Candidate: BRUNELLA D'ANZI

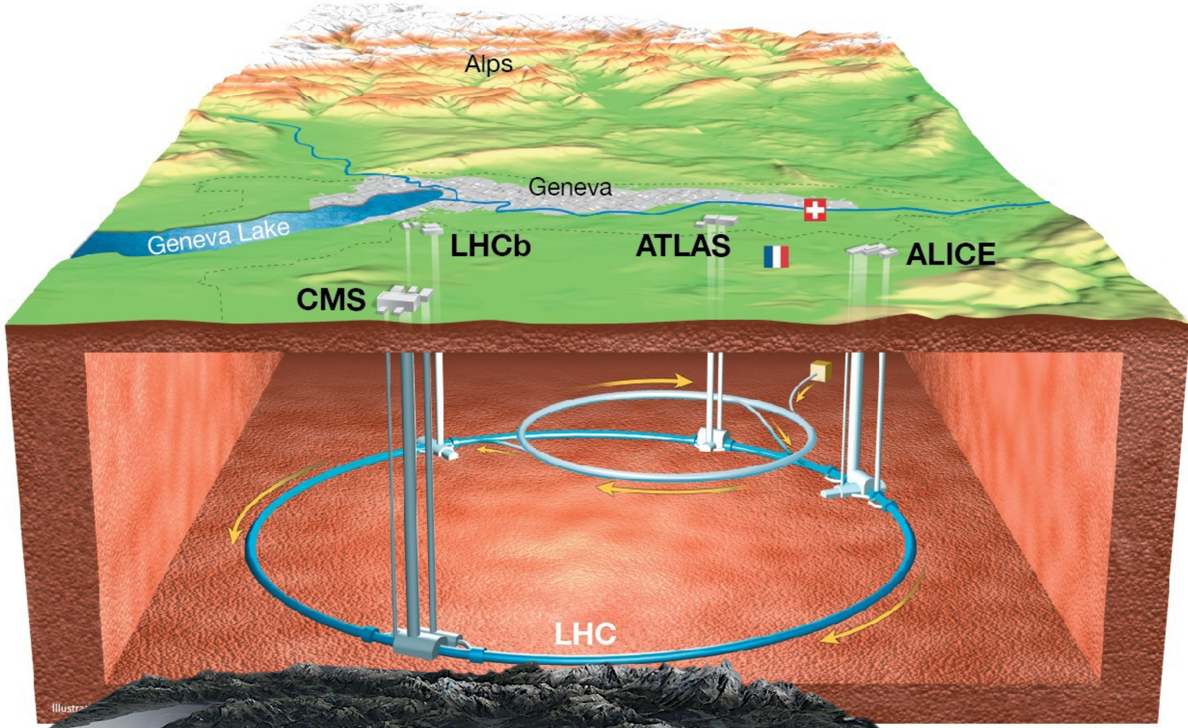
Advisor: Prof. NICOLA DE FILIPPIS



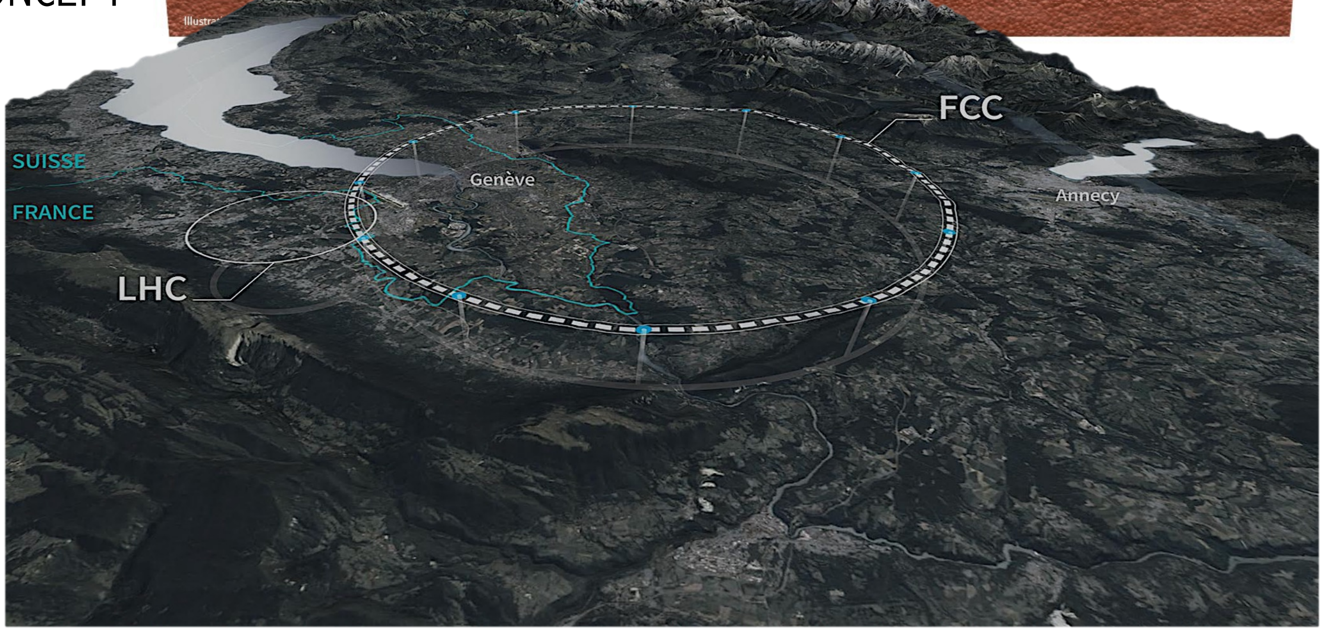

Doctoral Thesis Defense | 2025 April 7th, Bari

1

# Experimental introduction



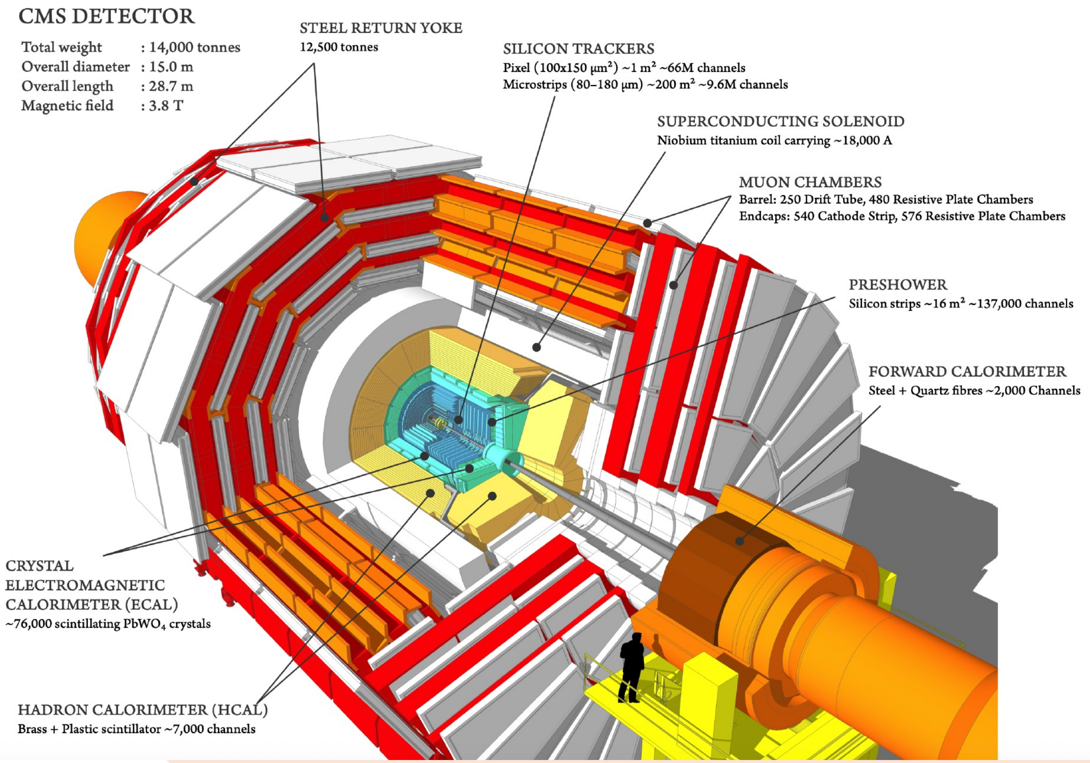

- THE LHC AND THE CMS EXPERIMENT
- THE FUTURE CIRCULAR COLLIDER AND THE IDEA DETECTOR CONCEPT

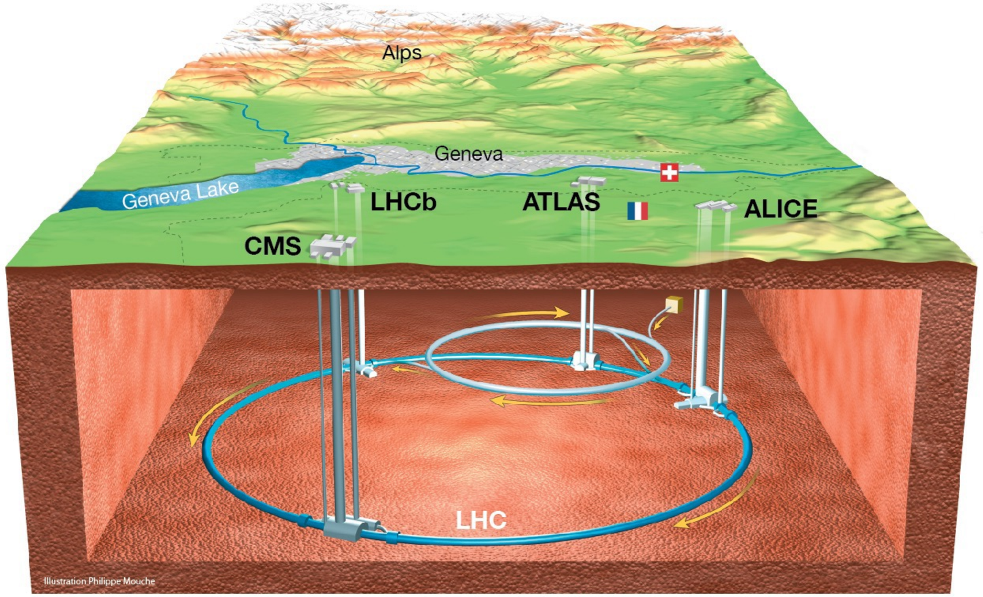



*"Detectors... are really the way you express yourself. To say somehow what you have in your guts. In the case of painters, it's painting. In the case of sculptures, it's sculpture. In the case of experimental physicists, it's detectors. The detector is the image of the guy who designed it."*
– Carlo Rubbia

# THE LHC AND THE CMS EXPERIMENT

- The **Large Hadron Collider (LHC)** is the largest and the most powerful accelerator in the world (**pp collisions**)
- It reached so far energies of $\sqrt{s}$ = **13.6 TeV** during Run 3 (2022-now), O(1500) charged particles/event
- In 2030, LHC will evolve to the **High-Luminosity phase (HL-LHC)**, reaching $\sqrt{s}$ = 14 TeV

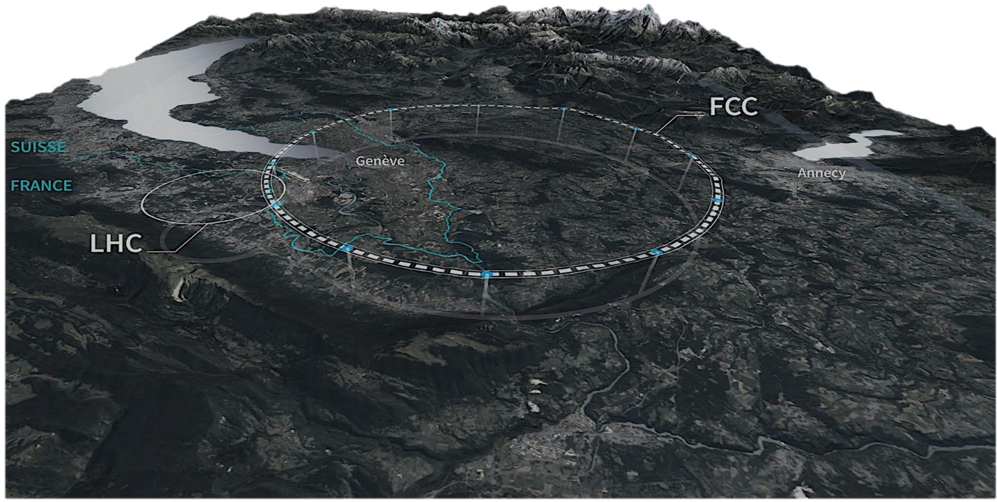


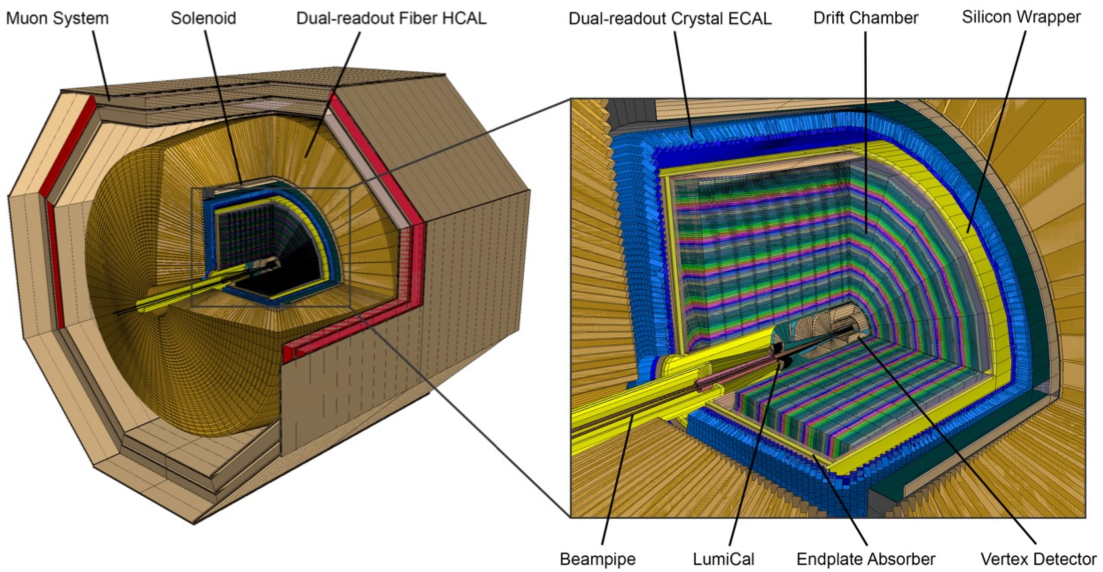




- The **Compact Muon Solenoid (CMS) experiment** is a general purpose experiment
- It collected 138 $fb^{-1}$ of data during LHC Run 2 (2016-2018)
- It will collect 3000 $fb^{-1}$ of data during HL-LHC

# THE FCC AND THE IDEA DETECTOR CONCEPT

- **Future Circular Collider (FCC)**, a circular collider with ~91 km circumference, is a post-LHC project proposed at CERN (CEPC in China, arxiv:2312.14363).
  - Precision tests of the SM, especially Higgs self-couplings, and searches for new physics (dark sector particles, extended Higgs sectors, and new gauge bosons).
- Designed as a multi-stage plan: FCC-ee (Higgs factory, $e^+e^-$ collisions) $\longrightarrow$ FCC-hh (energy frontier, pp collisions at $\sqrt{s} = 100$ TeV).



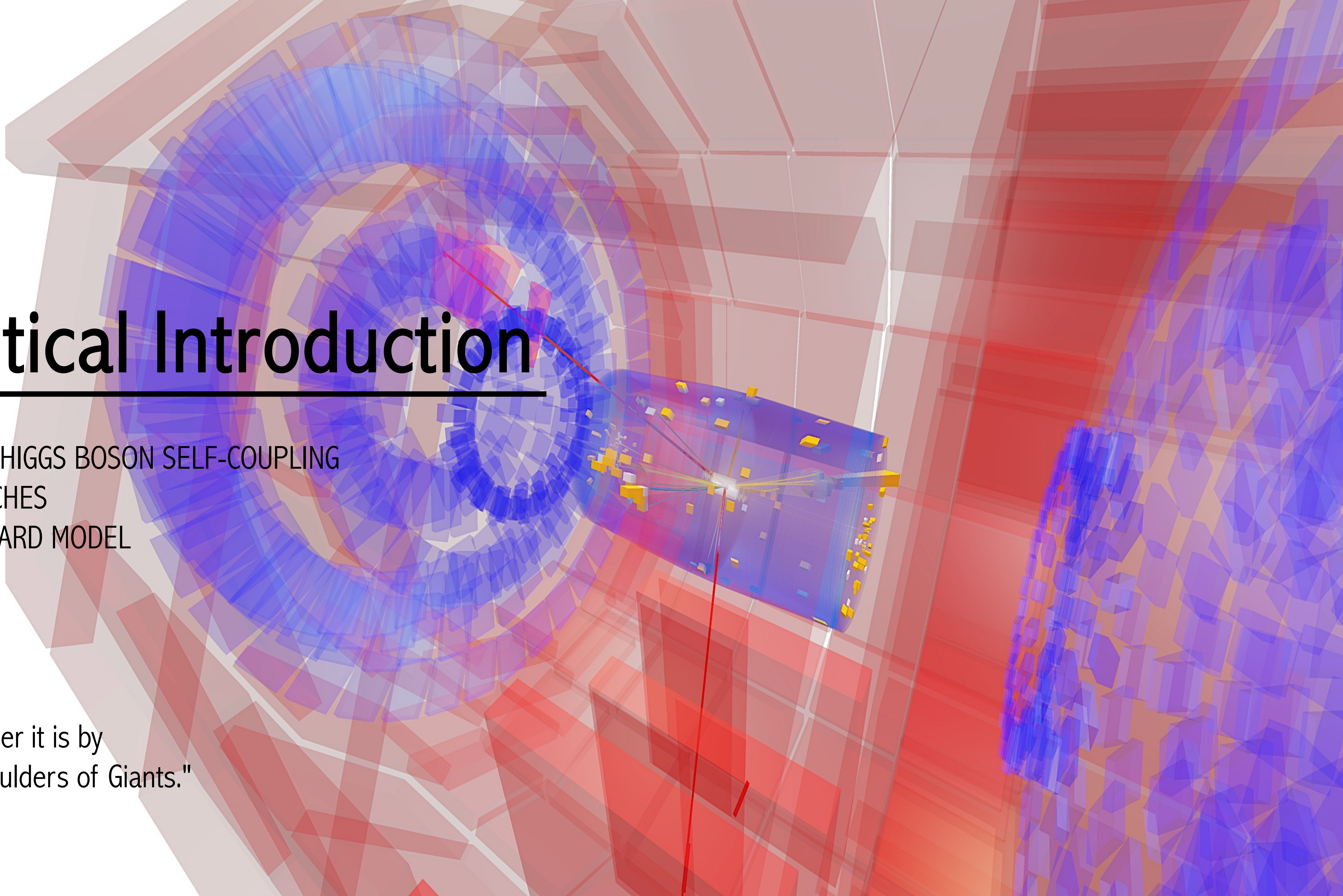




- **IDEA Detector** (Innovative Detector for Electron-positron Accelerator, arxiv:2502.21223) is a **general-purpose detector** concept optimized for **FCC-ee phase**:
  - Ultra-light drift chamber for precision tracking, and particle identification (along with Time of Flight - TOF - detectors), with minimal material budget.

# 2

# Theoretical Introduction

- THE TRILINEAR HIGGS BOSON SELF-COUPLING
- DI-HIGGS SEARCHES
- BEYOND STANDARD MODEL

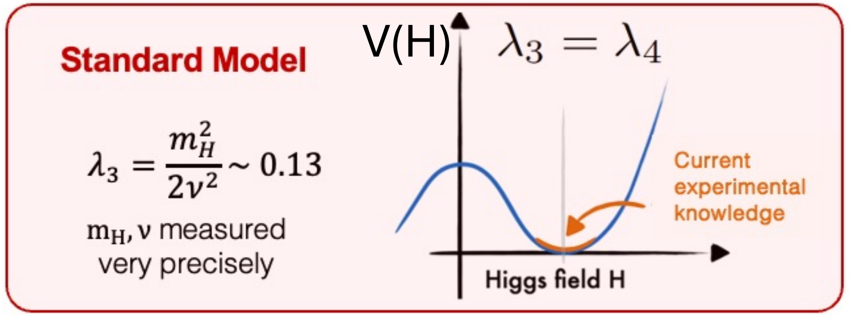


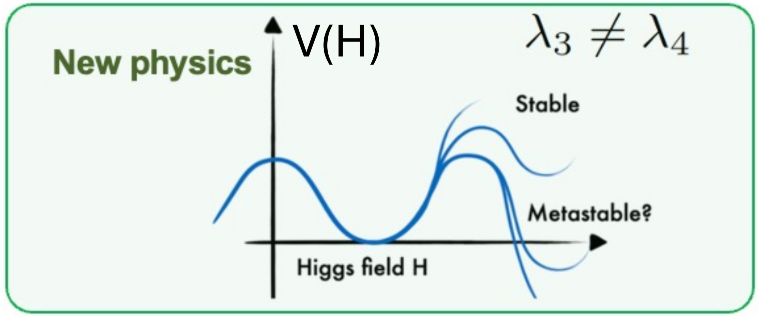
"If I have seen further it is by standing on the shoulders of Giants."
– *Isaac Netwon*

# HIGGS BOSON SELF-COUPLINGS

- The trilinear Higgs boson self-coupling $\lambda_3$ is a key parameter of the Standard Model (SM) of particle interactions:
  - precisely predicted theoretically
    - Scalar Higgs boson mass $m_H$
    - Vacuum expectation value (VEV), energy scale of the electroweak symmetry breaking $\nu$
  - experimentally directly unmeasured.

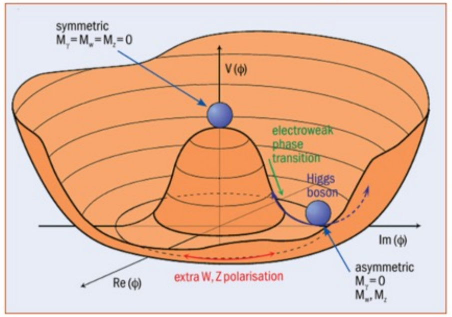
- Its observation would probe, for the first time, the only unobserved part of the SM Lagrangian, the energy potential V(H) of the Higgs boson.

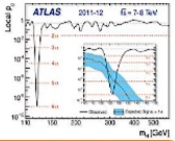


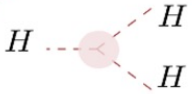

NOTE: The lifetime of the metastable vacuum is tens of orders of magnitude longer than the age of the Universe, making any transition to a different vacuum state irrelevant for the current cosmological timescale.

$$V(H) = \boxed{\frac{1}{2} m_H^2 H^2} + \boxed{\lambda_3 v H^3} + \boxed{\frac{1}{4}\lambda_4 H^4}$$

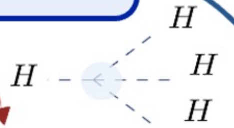



**Higgs discovery** (2012)




**Trilinear self-coupling $\lambda_3$ :** direct access via HH production (expected evidence at HL-LHC, promising results also at FCC-hh) (*)

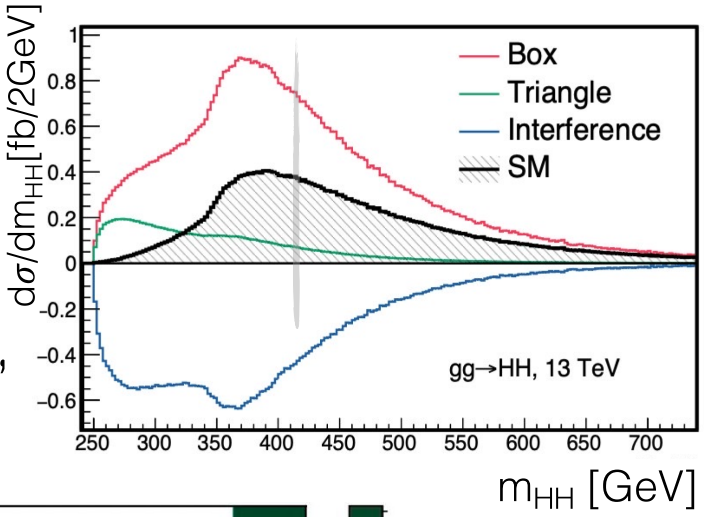



**Quartic self-coupling $\lambda_4$ :** direct access via production HHH (20??) (*)

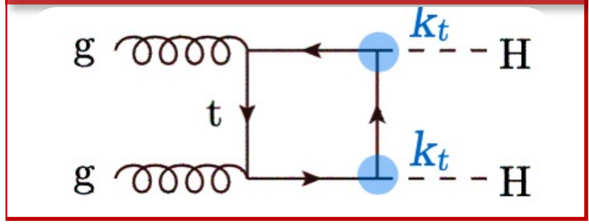
(*) $\lambda_3$, $\lambda_4$ can be constrainted indirectly for Next-to-Leading-Order (NLO) contributions to the single Higgs production cross section.

# DI-HIGGS SEARCHES

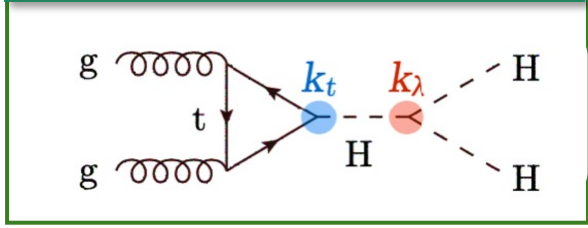
- At LHC, the HH pairs are mainly produced through gluon-gluon fusion (ggF) via fermionic loops
- A **very small cross section (XS)** $\sigma_{ggF,HH}$ (SM) ~ 31 fb, due to leading order (LO) destructive interference, ~**1000x smaller** than **single Higgs one** (pb → fb)

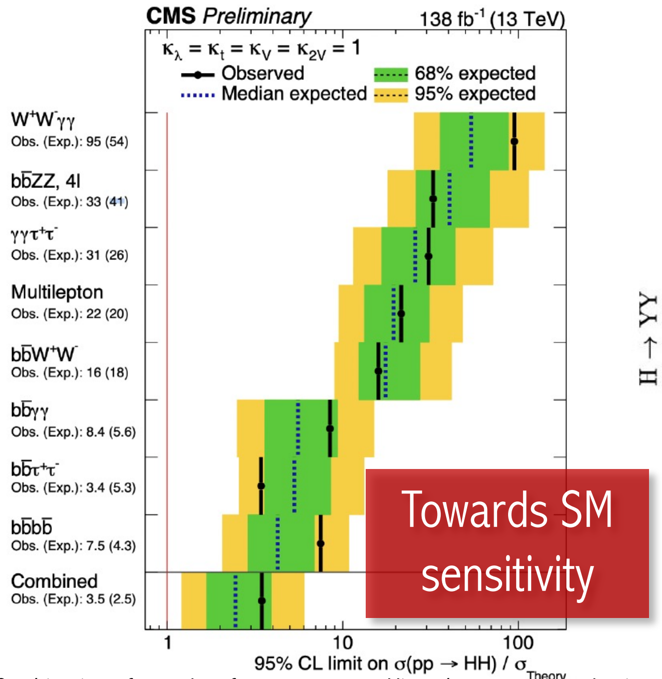



Box LO Feynman diagram

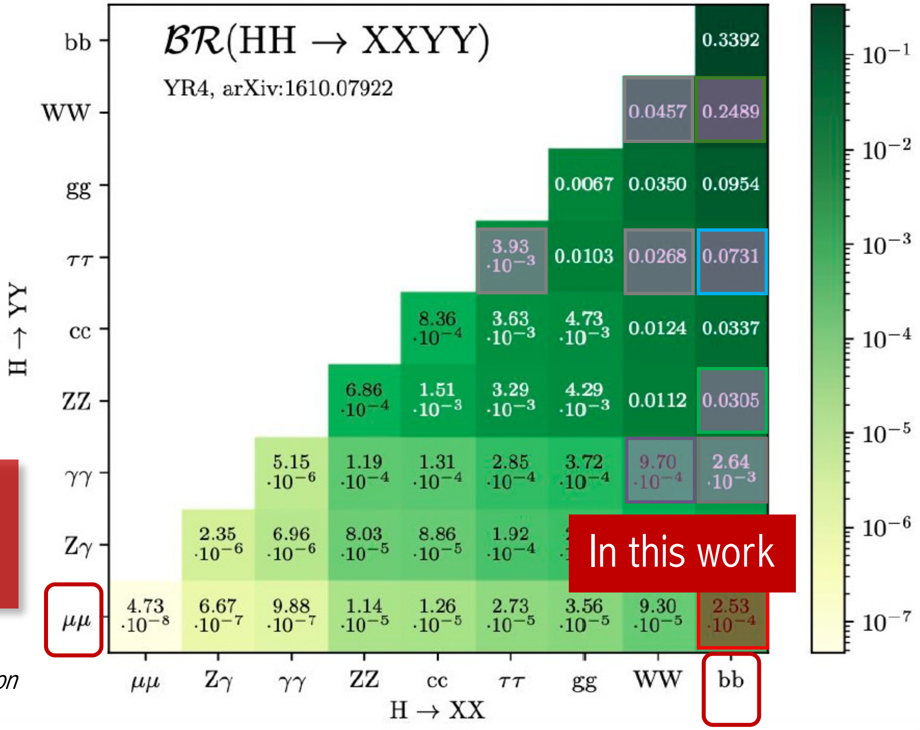



Triangle LO Feynman diagram



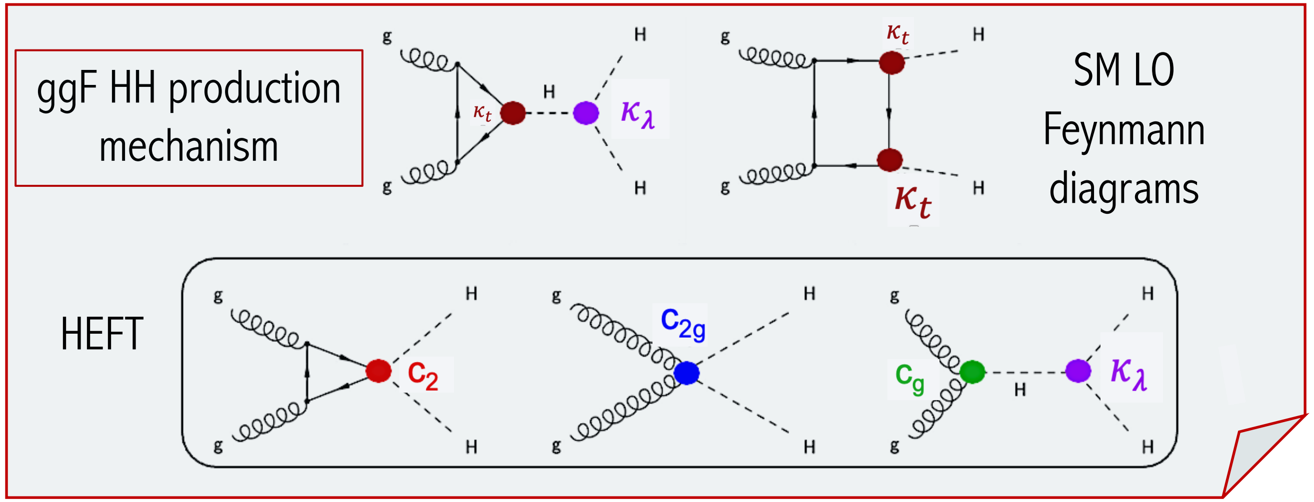

*Combination of searches for nonresonant Higgs boson pair production in proton-proton collisions at $\sqrt{s}$ = 13 TeV, cds record*

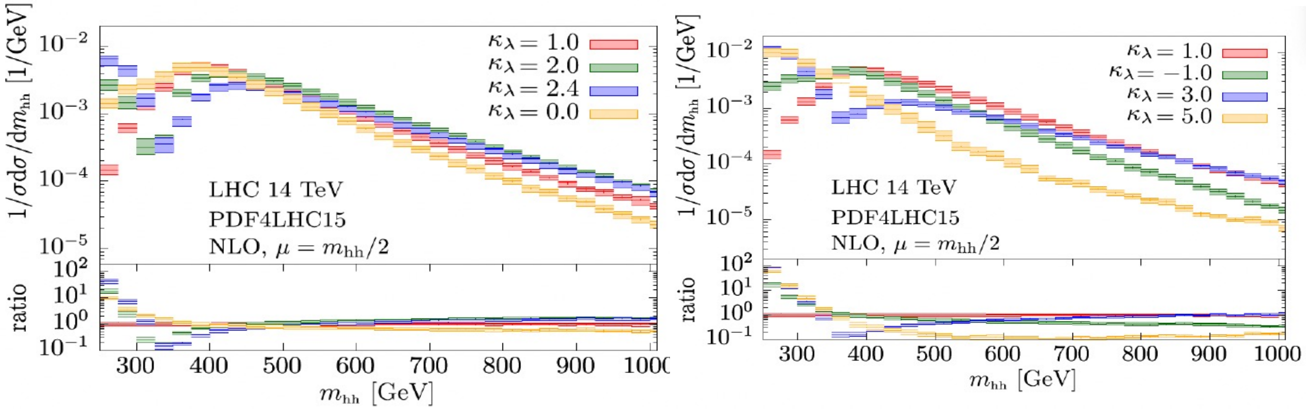



- At LHC and future HL-LHC, it is crucial to study all possible final states and to maximize sensitivity and observation potential, and rare ones need:
  - Improved reconstruction (charged) object techniques (Run 3), better particle identification (PID) at FCC
  - Machine learning techniques for signal to background discrimination
  - Analysis with high statistics (choice of Run 2)

# BEYOND STANDARD MODEL

- HH production is **highly sensitive to BSM effects**
  - via **new particles decaying into HH** (resonant)
  - through **loop-level modifications** (non-resonant, in this work).
- In CMS, current HH BSM interpretations use the Higgs Effective Field Theory approach (HEFT):
  - deviation from the SM **introduce anomalous couplings kappa** (e.g. $\kappa_\lambda = \lambda_3/\lambda_3^{SM}$ , $\kappa_t$ ), changing the **HH XS** and **kinematics HH properties**
  - additional gluon gluon fusion (ggF) couplings ($c_2$, $c_{2g}$, $c_g$)
- Complementary research in different decay channels ⟺ complementary sensitivity to couplings' variations
- So far, no BSM evidence has emerged, despite some early fluctuations later excluded with more data.



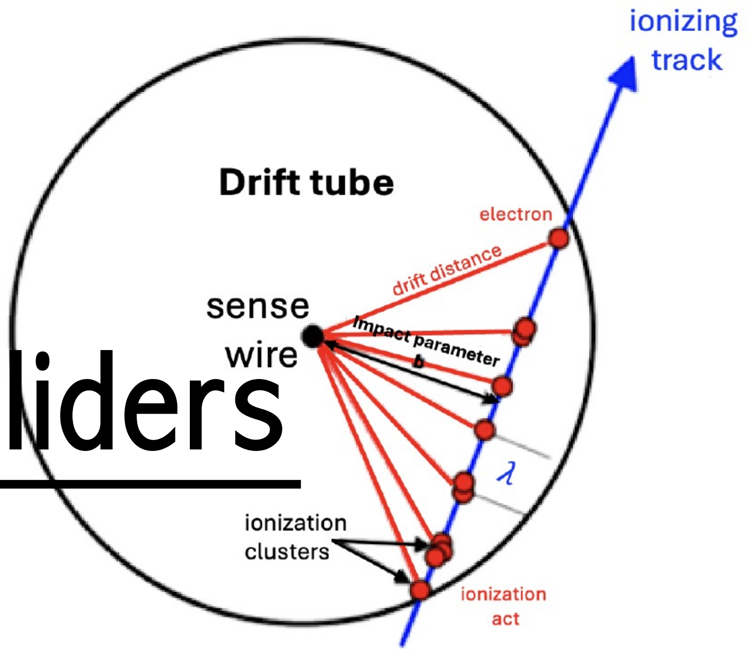

$\kappa_\lambda$ = 0: an enhanced cross-section due to the absence of destructive interference.
$\kappa_\lambda \approx$ 2.45: interference is minimized and the total cross-section reaches its minimum.
Higher values (e.g., $\kappa_\lambda$ = 5): increase the triangle contribution, altering the kinematics and suppressing events in the low-$m_{HH}$ region.

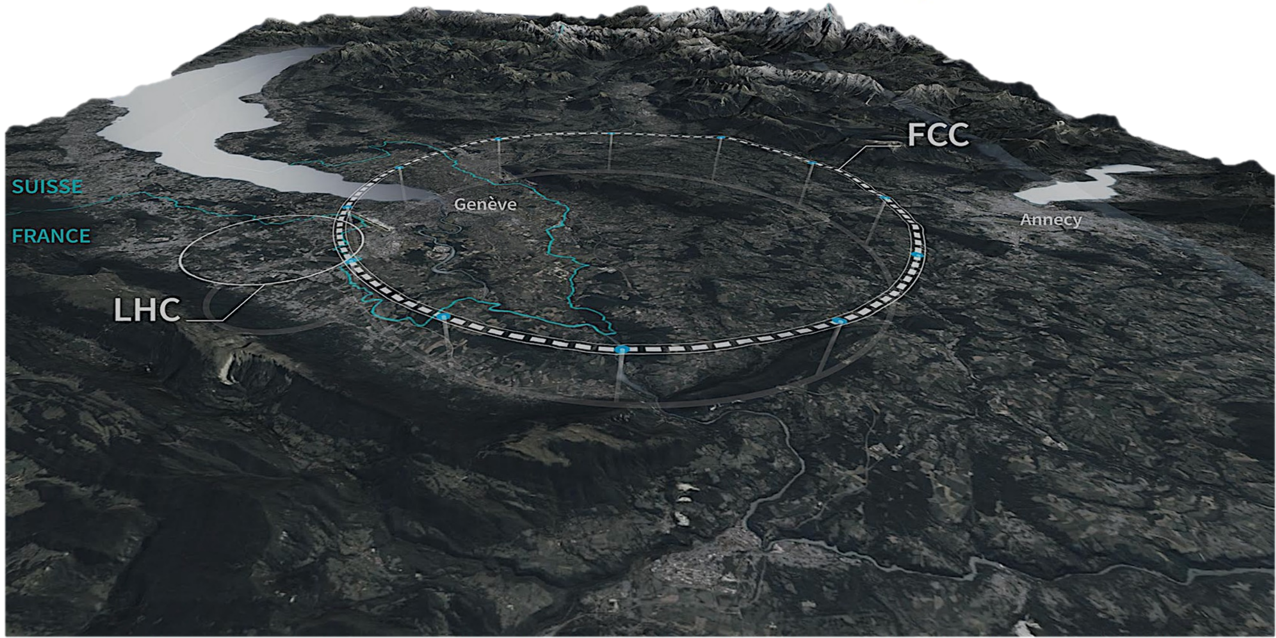
In this work

# 3

# Particle identification for future colliders

- A NOVEL PID: THE CLUSTER COUNTING METHOD
- BEAM TESTS OVERVIEW
- BEAM TEST ANALYSIS STRATEGY
- BEAM TEST RESOLUTION RESULTS



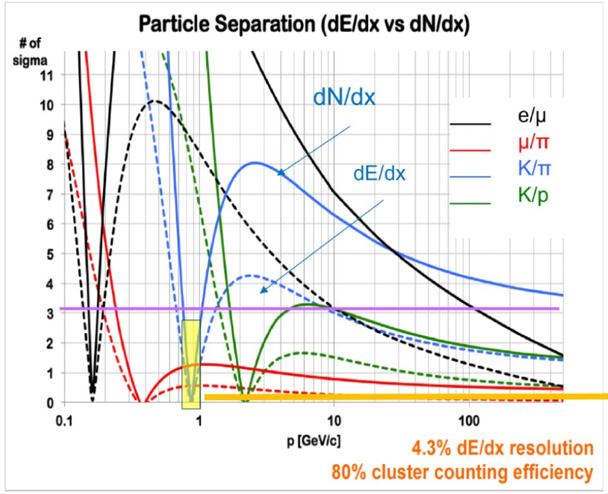




*"Science is a way of thinking much more than it is a body of knowledge."*
*– Carl Sagan*

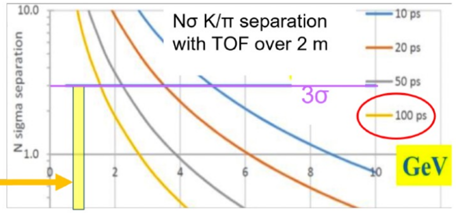
# A NOVEL PID: THE CLUSTER COUNTING METHOD

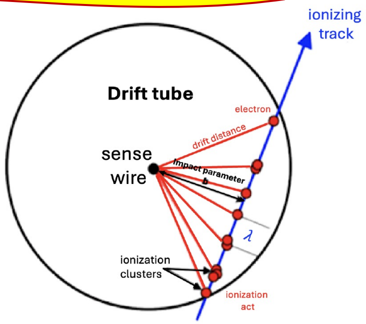
- **Truncated Mean method (dE/dx)** is the traditional PID technique based on the **measurement** of the **average ionization energy loss per unit length** by a charged particle in a medium predominantly via Coulomb interactions (Betha-Bloch formula).
  - **Large fluctuations** in energy deposition measurement (**Landau distribution** expected ⟶ Gaussian via truncation)
  - Dependence on gas gain and cluster size

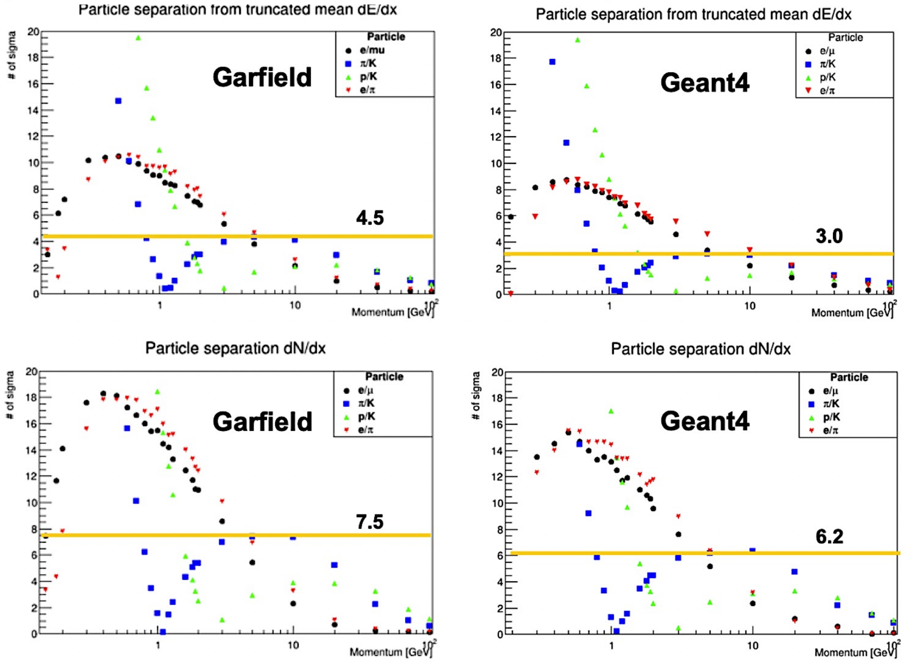
- **Cluster Counting (dN/dx)** is the innovative technique that **counts individual primary ionization clusters per unit length** along a particle's track.
  - It exploits the **Poisson statistics** of primary ionization clusters formation
  - **Robustness against fluctuations** in energy deposit and electronic noise
  - **Insensitive to high-ionizing backgrounds** like γ-rays.




**IDEA drift chamber configuration**: charge particle traversing 200 drift cells, 1 cm wide, filled with a mixture of 90% He - 10% $iC_4H_{10}$. PID inefficiency at ∼ 1 GeV is recovered thanks to TOF detectors.




**Analytical calculations** and **simulation studies** (Garfield++, GEANT4) show (with some discrepancies) that the **particle separation** is improved **by a factor of ∼ 2** with respect to the traditional dE/dx method.
⟶ Need to **validate with collection and analysis of real data** (main work of this PhD)!

*First beam test results published in Nucl.Instrum.Meth.A* 1048 (2023) 167969

# BEAM TESTS OVERVIEW

arxiv:2502.21223

## Goals of Beam Tests

Quantify **PID performance** with **dE/dx vs dN/dx**

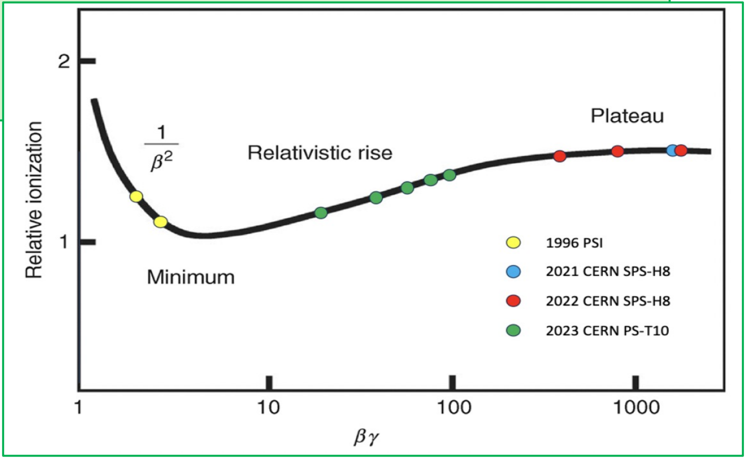
- Study **Poisson nature of the cluster counting (CC) technique** and its dependence on βγ
- Evaluate **CC algorithm efficiency** w.r.t. **geometry, gas gain, electronics (sampling rate)**
- Identify limiting factors: e.g. **space charge effects**

## Setup & Parameters Explored

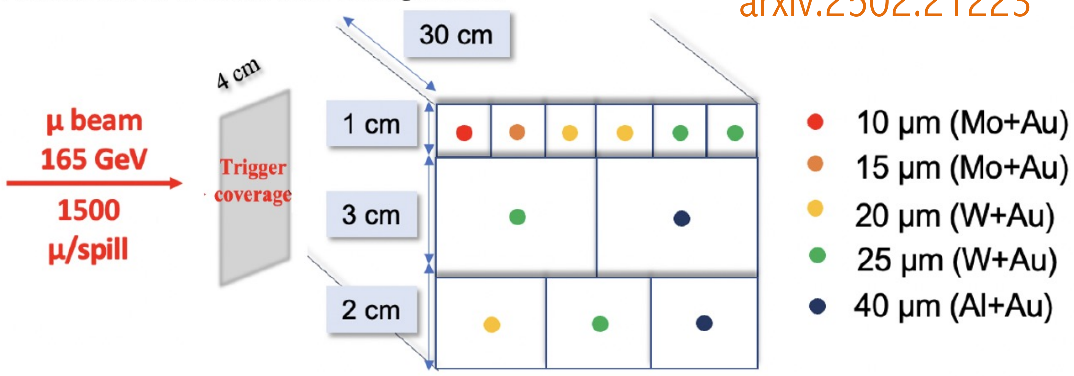
**Square section drift tubes:** 1.0–3.0 cm drift cell size, anode wires: 10–40 μm: identified better configurations for the available DAQ (1-1.5 cm)

- **Gas mixtures:** He–Isobutane (90/10, 85/15, 80/20)
- **Gas gains:** 1–5 × $10^5$

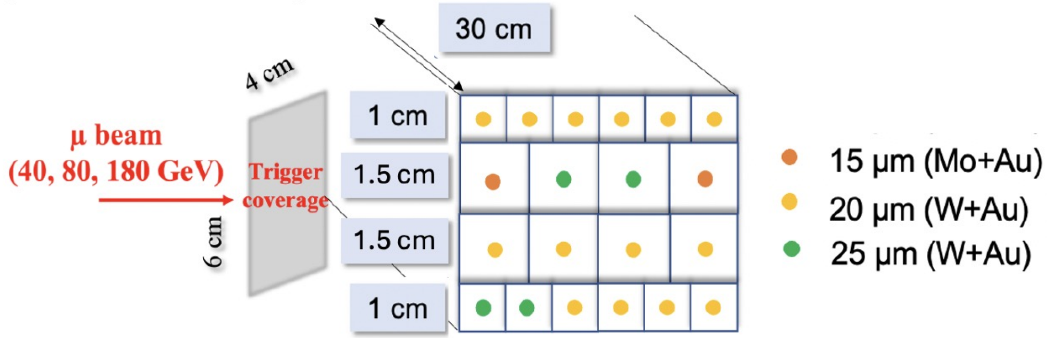
- SPS-H8 @ CERN (Nov 2021, July 2022): muons at 165–180 GeV/c **(Fermi plateau of Betha-Bloch), in this work**
- PS-T10 @ CERN (Jul 2023): muons at 2–12 GeV/c **(relativistic rise)**

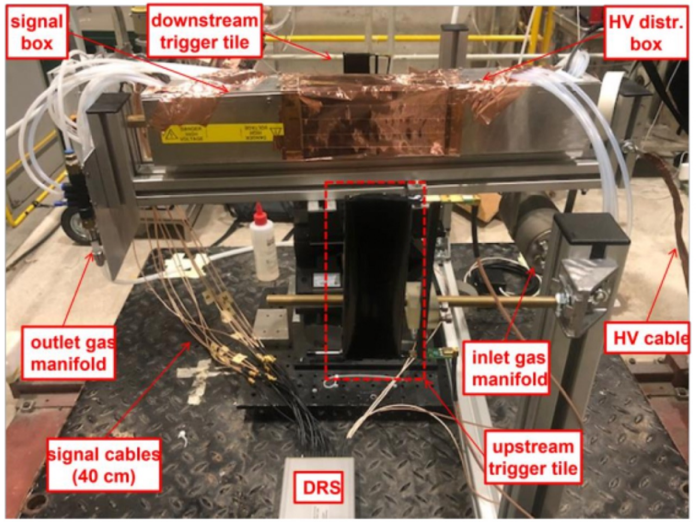



November 2021 Beam Test Configuration:




July 2022 Beam Test Configuration:



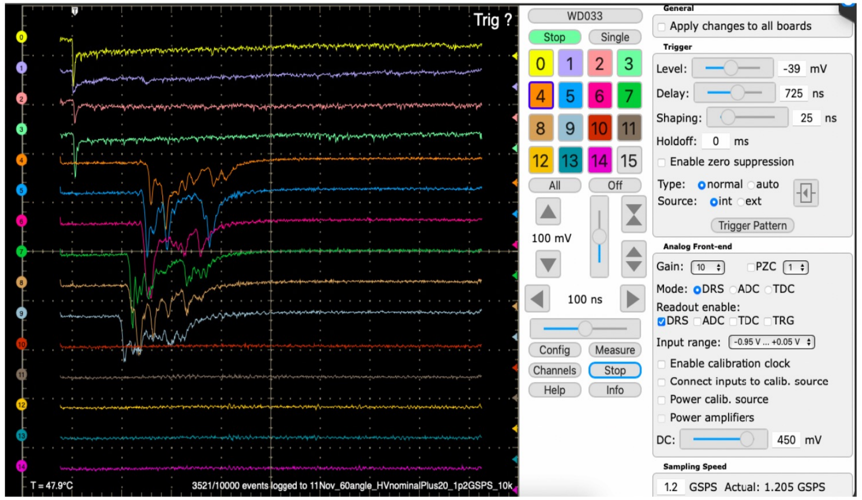
# BEAM TESTS ANALYSIS STRATEGY

*PoS* ICHEP2022 (2022) 335

## 0. DAQ

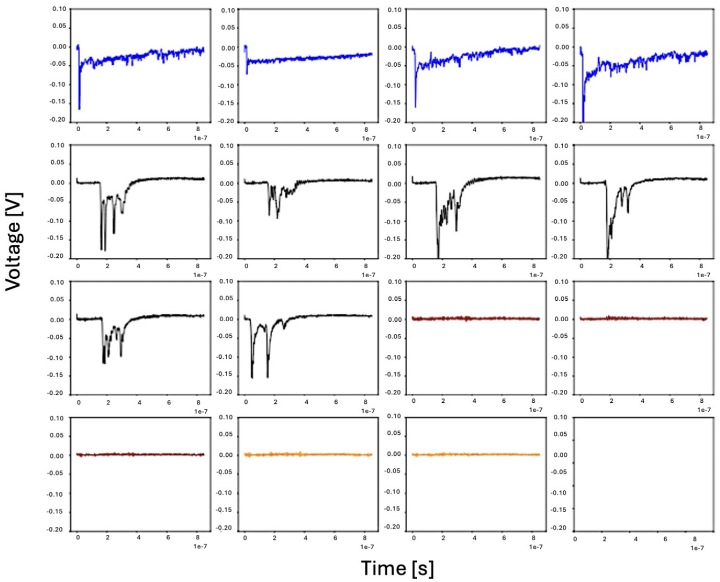
Trigger channels

Drift tubes Voltage signals




## 1. Preprocessing

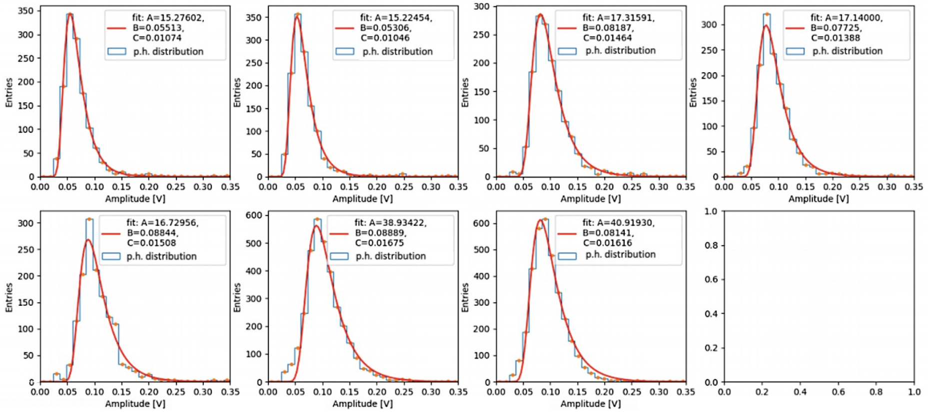
- **baseline normalization** (first 30 ns)
- waveforms < 3 mV **discarded as noise**
- Removal of **signal out of data acquisition window**




A muon passing through a row of 6 drift tubes of 1 cm cell size

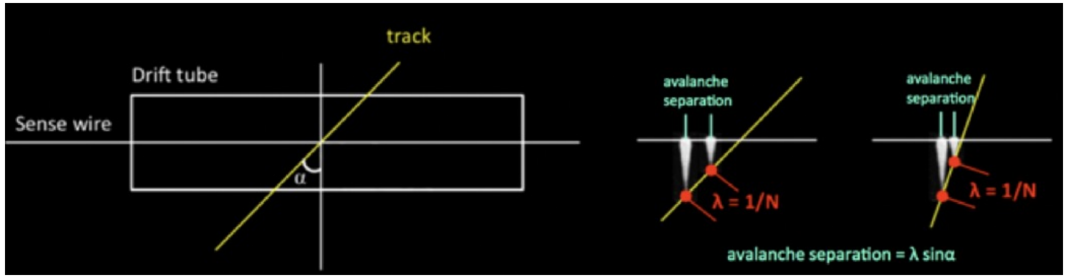
## 2. Gas Gain Study

- **Single electron pulse height distributions** (average maximum voltage per event) fitted with Moyal → used to extract **gas gain** (parameter B = Most Probable Value).




## 3. Charge effects study

- Estimate the **overlap of single electron avalanches** → assess impact on gas gain and CC performance.
- Study of **avalanche separation λ·sin(α)**, where **α** is the **angle between track and sense wire normal**: **space charge suppresses gain** due to **overlapping avalanches** increasing with **angle α** for inclined tracks: → **60–75% suppression** at 90/10 He-$iC_4H_{10}$ and few $\times 10^5$ gas gain.
- No strong dependence on gas gain (within tested range) and sense wire diameter → **Space charge corrections are necessary** for **accurate CC cluster reconstruction**.

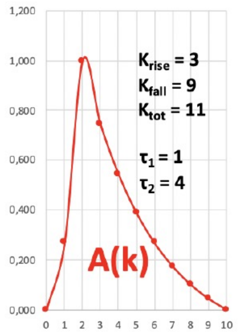


# BEAM TESTS ANALYSIS STRATEGY

## 4. Finding electron (red) peak algorithms development (comparable results) :

- **DERIV Algorithm:**
  - Based on **1st and 2nd derivatives** of waveform.
  - Peak identified by **local extrema, curvature, and amplitude criteria**.
- **RTA Algorithm:**
  - Uses **digitized exponential templates** (rising & falling).
  - Scans waveform and matches against templates.
  - Iteratively subtracts matched peaks.
- **Machine Learning method** also tested (with CEPC team), but **not optimal for FPGA implementation** (slow, complex) (ACAT, *Nucl.Instrum.Meth.A* 1046 (2023) 167734).

| k | A(k) |
|---|---|
| 0 | 0.0 |
| 1 | 0.269 |
| 2 | 1.0 |
| 3 | 0.744 |
| 4 | 0.545 |
| 5 | 0.390 |
| 6 | 0.269 |
| 7 | 0.175 |
| 8 | 0.102 |
| 9 | 0.044 |
| 10 | 0.0 |




## 6. Performance Scans & CC Efficiency

- Parameters varied:

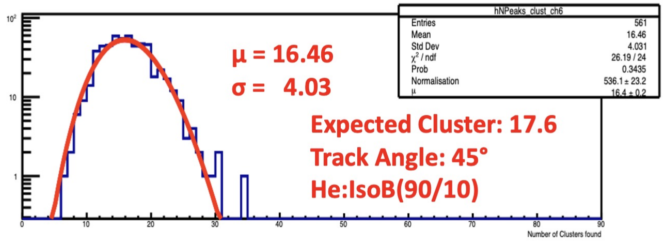
Gas mixture (He:$iC_4H_{10}$: 90/10, 85/15, 80/20), Track angle, gas gain (HV), cell size, wire diameter.

- Efficiency = measured clusters / expected clusters OR ratios between different drift cell size to be theory-independent
  - CC efficiency slightly varies with HV (~20% steps).

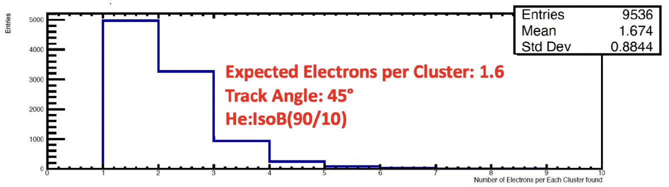
## 5. Clusterization algorithm: associate electrons → **primary ionization clusters, Poissonian.**

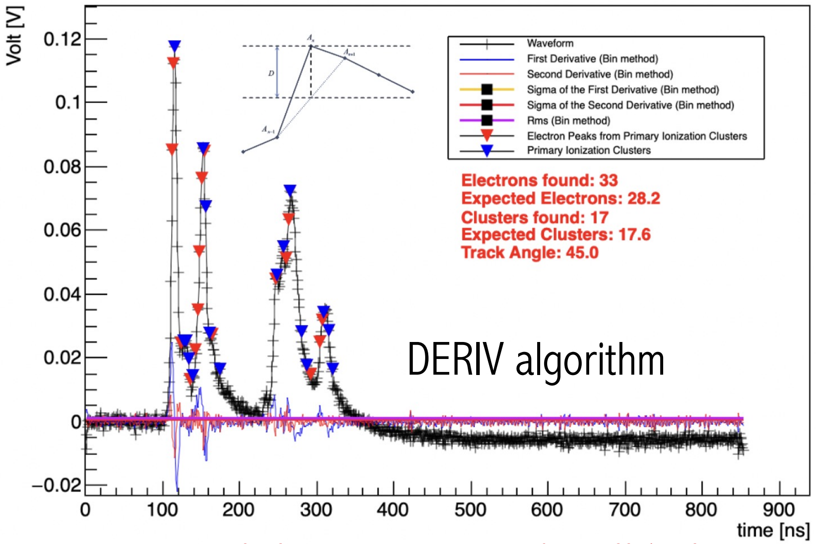
- Merge close electron peaks (reduce fakes).
- Associate **peaks within diffusion time window** (~ns, for Helium based drift chambers).
- Time spread ∝ √(electron drift time), depends on gas.
- Cluster position/amplitude: taken from the highest peak → results compatible with theoretical expectations!

**Number of Cluster Distribution**




**Electron per Clusters Distribution**




*Sense Wire Diameter 10 um – Cell Size 1.0 cm – Track Angle 45° – 1.2 GSa/s – Gas Mixture He:IsoB 90/10 – 165 GeV*

$$\text{Expected Clusters} = \delta_{cl/cm} \times \text{cell length[cm]} \times \text{relativistic rise factor} \times 1/\cos(\alpha)$$

*Nucl.Instrum.Meth.A* 1046 (2023) 167734

# BEAM TEST RESOLUTION STUDY

## Other Physics Observables

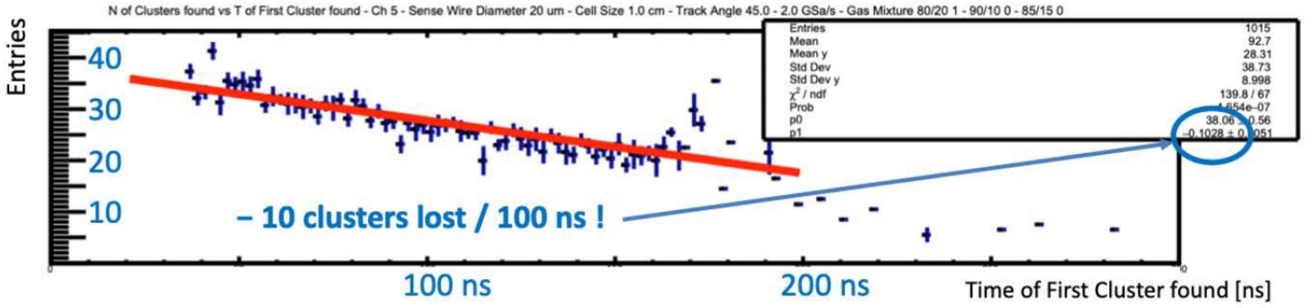
- **Drift velocity**: estimated from first cluster drift time window acceptance, compatible with expectations ($v_{drift} \approx 2.5$ cm/μs for Helium-based drift chambers)

## Space charge effects

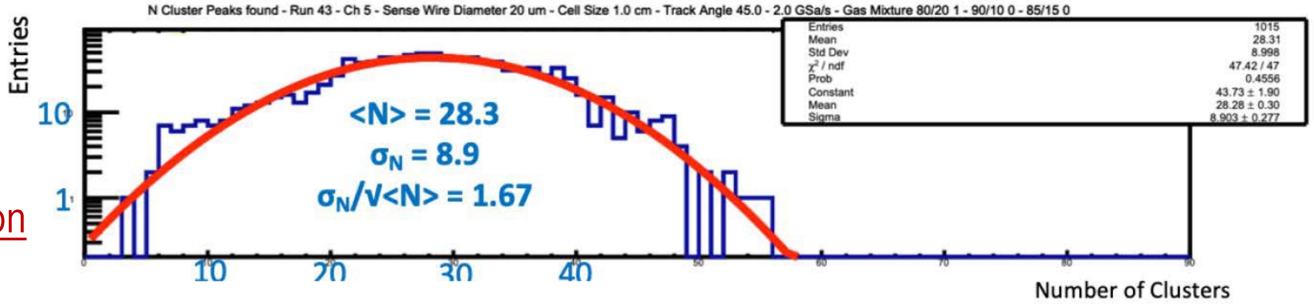
- **Suppress signal** at **small angles** ($\alpha = 0$).
- ~ **10** clusters lost per 100 ns at high drift times.

## Corrections Factor

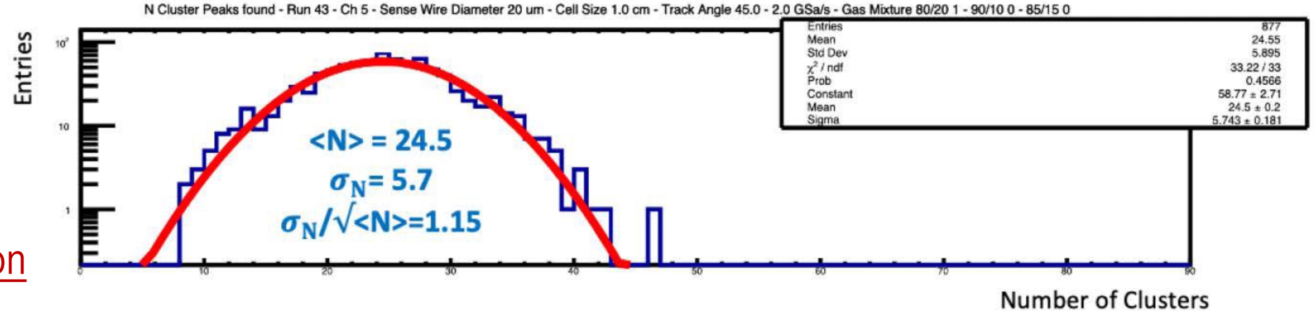
- Clustering efficiency corrected for:
  - Attachment, recombination, and space charge effects.
- **DERIV correction example (gas 80/20)**:
  - Linear slope fit applied to account for time-dependent cluster loss.




Before correction




After correction



# BEAM TEST RESOLUTION RESULTS

*PoS* ICHEP2024 (2025) 1067
*PoS* ICHEP2024 (2025) 1104

## Dataset

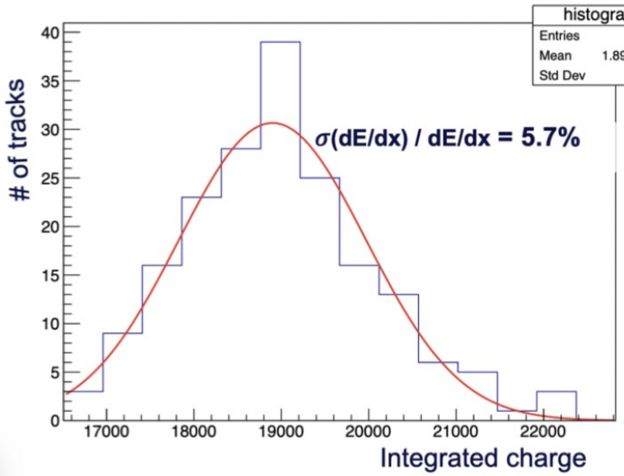
- Muons at **165 GeV** (CERN, 2021, Fermi plateau)
- **Drift tubes:** 1 cm cell, 20 μm wire (high statistics)
- He:IsoB 90/10 gas mixture, 2 m tracks
- **Same hits used for both dE/dx and dN/dx** ⟶ direct comparison

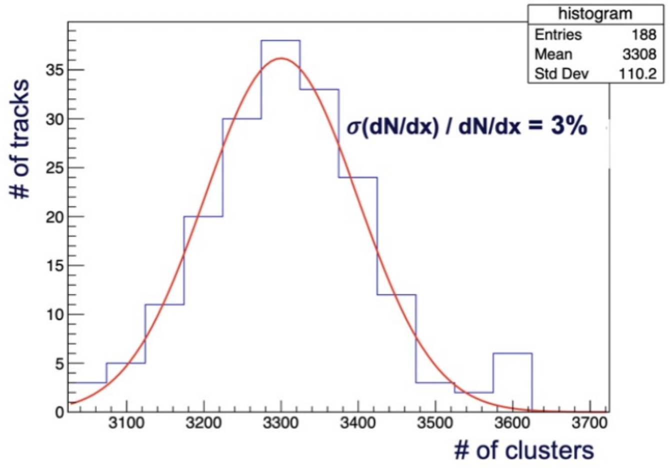
## Results

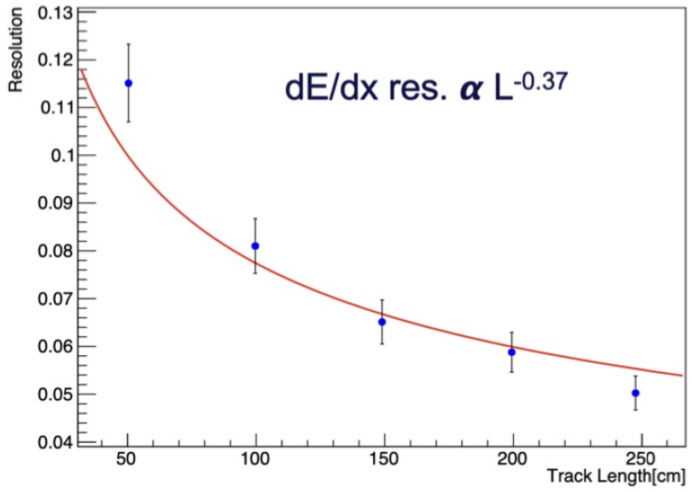
- dE/dx resolution: **~5.7%** (with 80% truncated mean)
- dN/dx resolution: **~3%** ⟶ ~2x better
- Resolution vs track length:
  - dE/dx $\propto \mathrm{L}^{-0.37}$
  - dN/dx $\propto \mathrm{L}^{-0.5}$ ⟶ better scaling

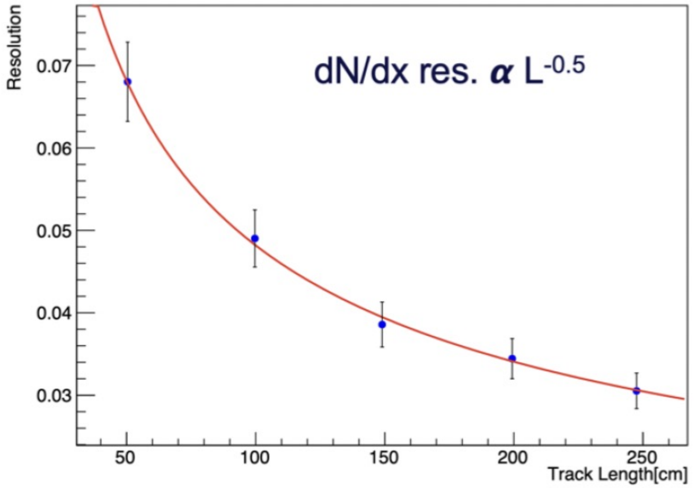
## Gas Mixture Trade-offs consideration:

- **Argon:** more ionization, better resolution, more multiple scattering
- **Helium:** less ionization, large time separation, faster drift, better for high-rate experiments



4

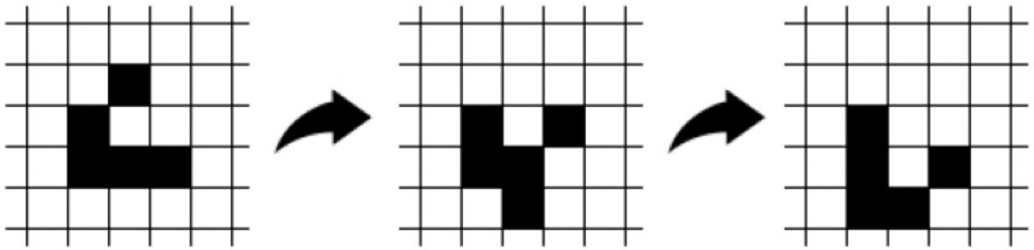
# Track seeding technique for the CMS experiment

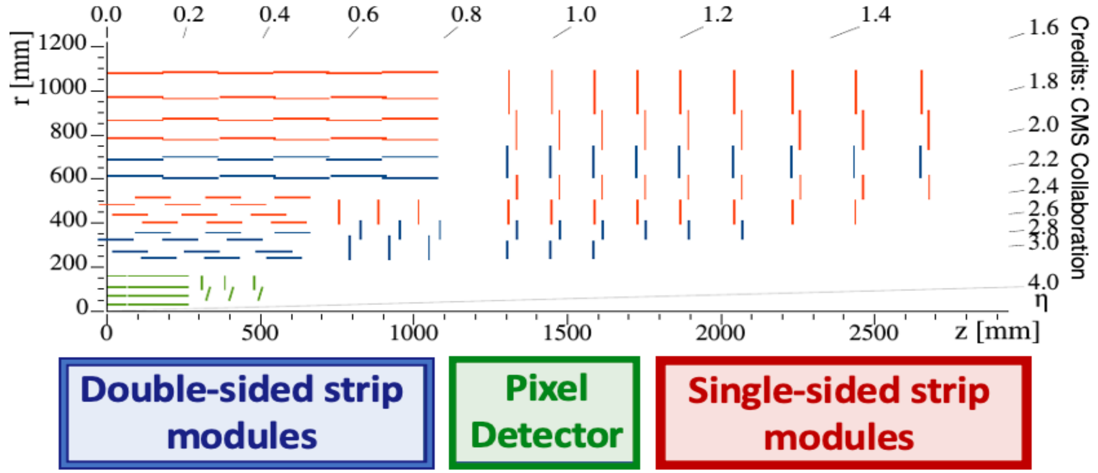
- THE CELLULAR AUTOMATON (CA) SEEDING ALGORITHM
- CA EXTENSION TO OUTER TRACKER
- CA EXTENSION PERFORMANCE

*"The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny...'" – Isaac Asimov*

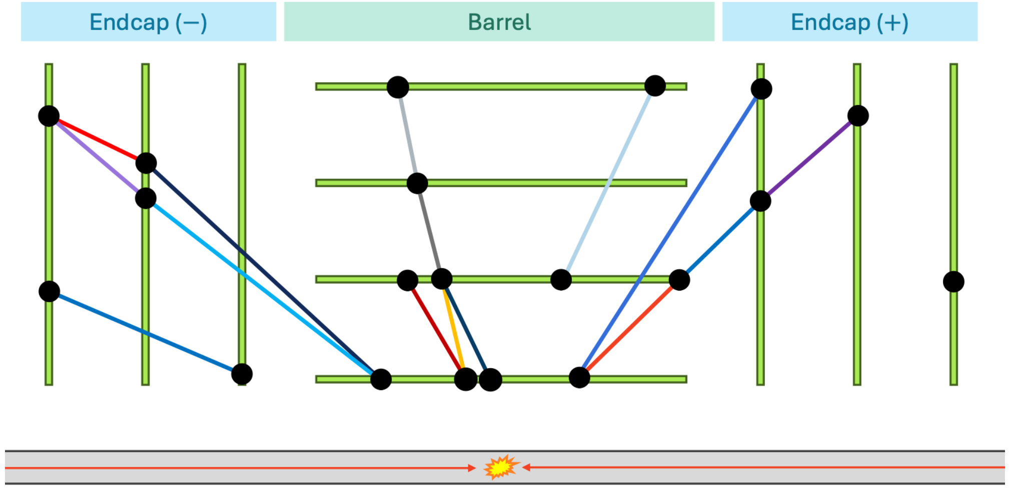
# THE CA SEEDING ALGORITHM

- The Cellular Automaton (CA) is an algorithm **designed for parallel architectures**, can be used either as the **main track finder algorithm** or as a **tracking seeding step** to another algorithm (offline or online reconstruction) ⟶ currently using pixel detector.
- CA is inspired by the **Conway's Game of Life** classic simulation where **each square cell evolves over time based on the state of its neighbors** — similar to how track signals **(hits)** evolve into track segments.






Run 3 and Run 2 (2017-) CMS tracker schema




- CA reconstructs tracks from hits using a **graph-based approach**.
- Tracks are seeded through **CA doublets** ⟶ **triplets** ⟶ **quadruplets**.
- **Cells (doublets)** evolve by connecting to **geometrically compatible neighbors** ($\Delta$ z, $\Delta\phi$, etc.).

# CA EXTENSION TO OUTER TRACKER

CMSSW PR #47090
CMSSW PR #47271

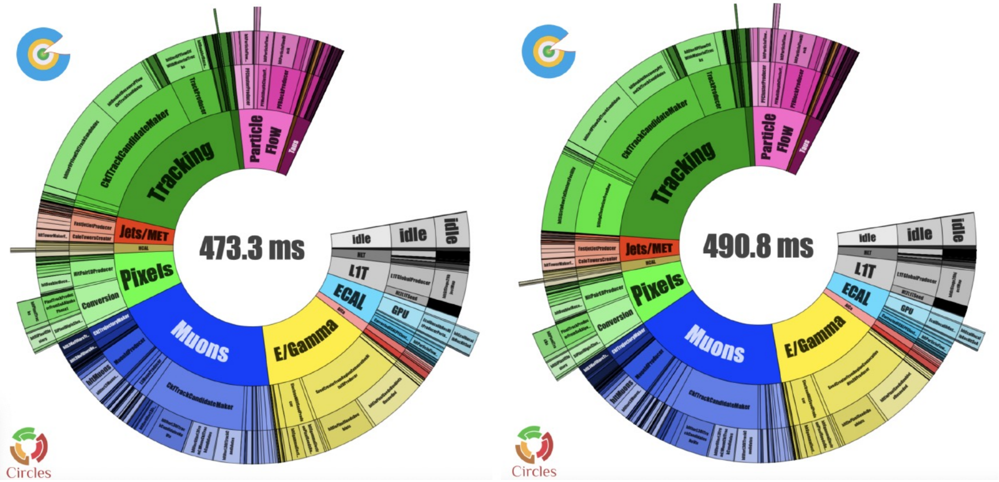
## Motivation

- **Momentum resolution** improvement
- **Fake tracks** reduction
- **Recover pixel off failures** → benefits for all analyses

## Data Flow and Run 3 (HLT) Context (running on hetherogeneous architectures)

- **Input:** Structure of Arrays (SoA) **reconstructed hits from pixel detector** and **2D matched strip hits** as merged collection for CA doublet search.
- **Global unpacking of strip clusters** (CPU running) required for the CA extension.
- Impact on **event processing throughput**: **4% decrease due to strip cluster unpacking** (other tracking users would benefit from it).

Timing event processing with (without) strip unpacking on the right (left)

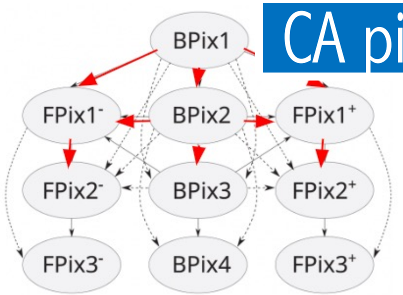


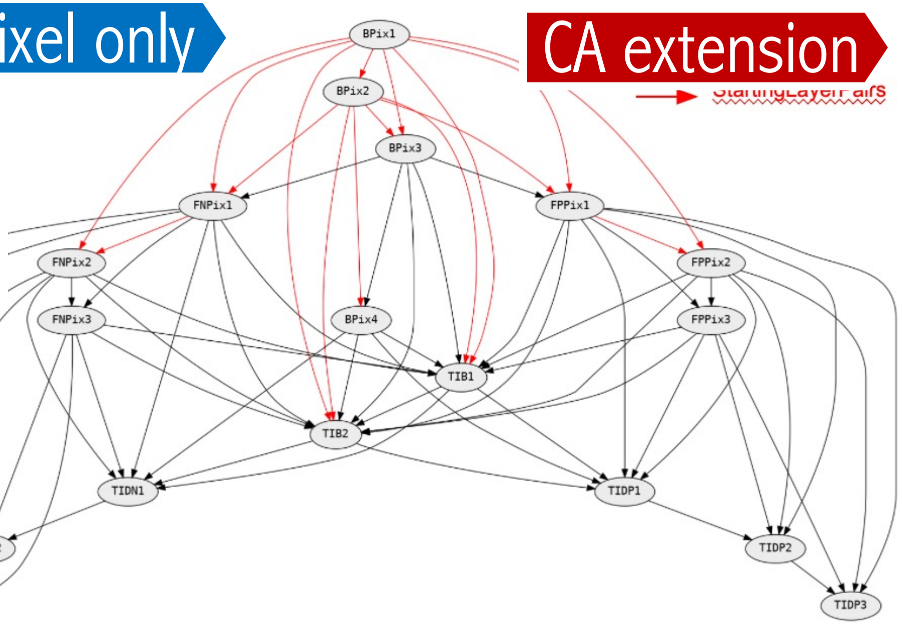
# CA EXTENSION TO OUTER TRACKER

CMSSW PR #47090
CMSSW PR #47271

Motivation

- **Momentum resolution** improvement
- **Fake tracks** reduction
- **Recover pixel off failures** → benefits for all analyses




**Data Flow and Run 3 (HLT) Context (running on hetherogeneous architectures)**

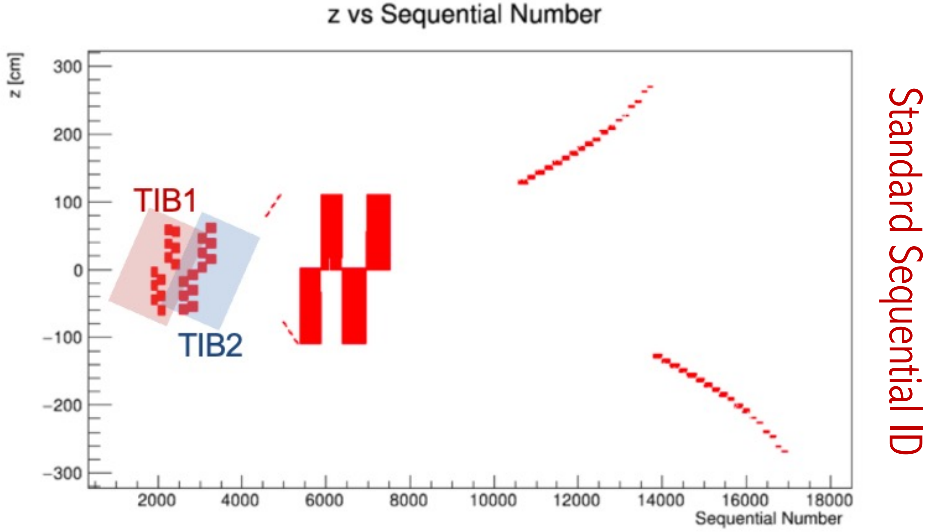
- **Input:** Structure of Arrays (SoA) **reconstructed hits from pixel detector** and **2D matched strip hits** as merged collection for CA doublet search.
- **Global unpacking of strip clusters** (CPU running) required for the CA extension.
- Impact on **event processing throughput**: **4% decrease due to strip cluster unpacking** (other tracking users would benefit from it).

**Graph Construction and Optimization**

- CA algorithm requires **layer adjacency graph for hit pairs** (18 → 50 pairs).
- Careful **selection of layer pairs** to optimize GPU containers sizes without exceeding memory limits (~2,000,000 doublets - acceptable in heavy-ion runs).

**Customized Sequential Layer IDs**

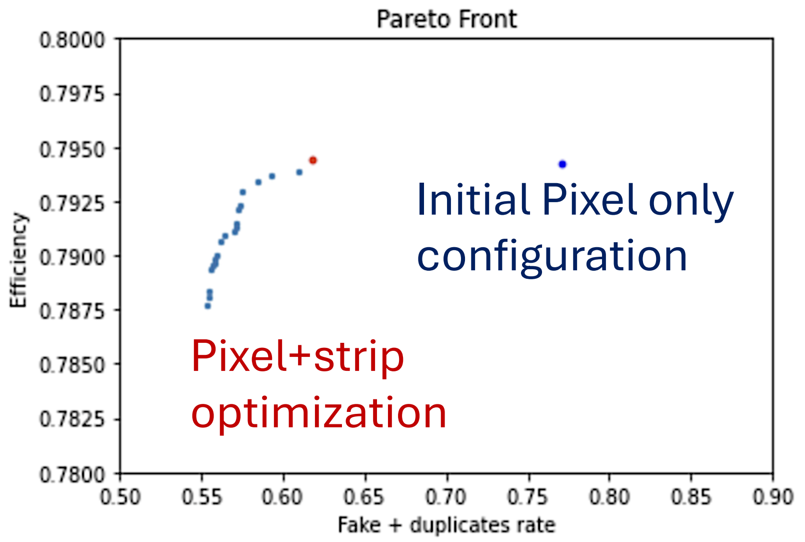
- Default detector ID leaves empty memory allocations for strips, especially for TIDs.
- Solution: **Custom sequential ID** mapping to efficiently process consecutive layers.



# CA EXTENSION PERFORMANCE

**Optimize CA parameters to maximize efficiency and minimize fake rates.**

- For the current pixel-only CA implementation, already integrated in **2025 data-taking** a slight **efficiency improvement in HLT full tracks** (simulated to reconstructed tracks association)
- For the CA+extension parameter optimization:
  - **First Approach: simulated track parameters distributions study**
  - **Second Approach: Multi-Objective Particle Swarm Optimization (MOPSO)** is a **swarm of agents** in the parameter space in multiple iterations exploring the solution space for the best solution found by the swarm **(pareto front)** and the local best solution of each agent



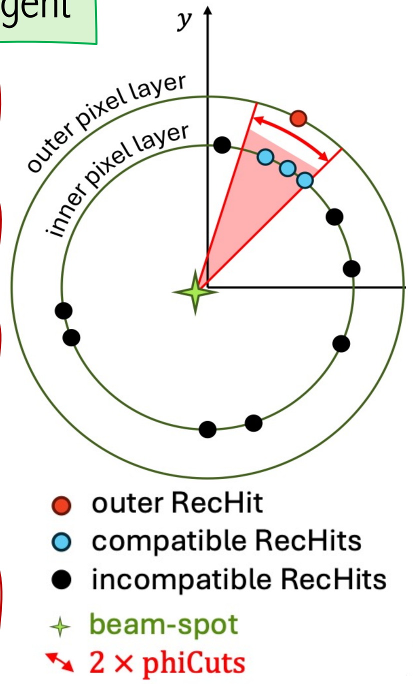

**Performance on simulated ttbar events with Run 3 pileup conditions.**

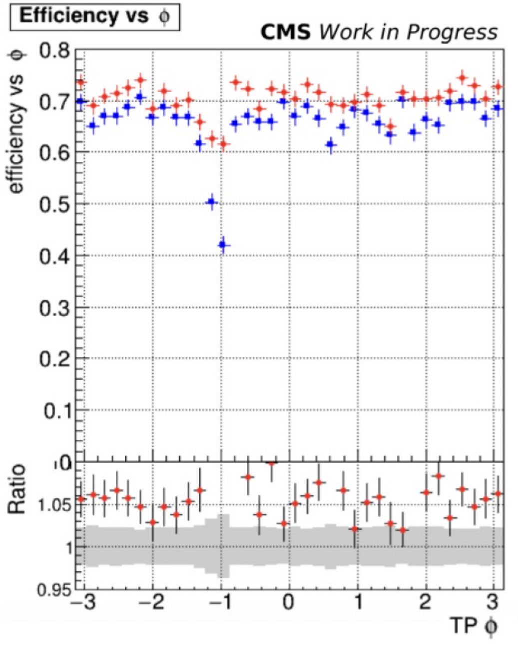
**First Configuration:** Require at least pixel triplets + pixel-strip quadruplets vs pixel only triplets for seeding:

- Fake+duplicate rate <u>decreased by 50%</u> for track seeding.
- Efficiency <u>decreased by 10%</u>.
- <u>Improved</u> momentum resolution <u>by 10%</u>.

**Second Configuration:** Require at least triplets (pixels-only, pixel+strip,strips-oly) vs pixel only triplets, most promising results (**no need of HLT doublet recovery tracking iteration**)!

- HLT efficiency of track seeds increases, along with <u>HLT full tracks by 20%</u> in pixel hole failure regions (plot on the right)



# 5

# CMS muon track reconstruction performance

- CMS OFFLINE TRACKING
- RUN 3 MUON TRACK RECONSTRUCTION PERFORMANCE
- TRACKING PERFORMANCE RESULTS

"Physics is about questioning,
studying, probing nature. You probe,
and, if you’re lucky, you get strange clues."
– Lene Vestergaard Hau

# CMS OFFLINE TRACKING

- **Precision tracking** is **essential** for **accurate charged particle momentum measurement**, which impacts various CMS analyses.
- **Track reconstruction** is used in **nearly every reconstruction element**, including **vertex reconstruction**, **flavor jet tagging**, and **pileup removal**.

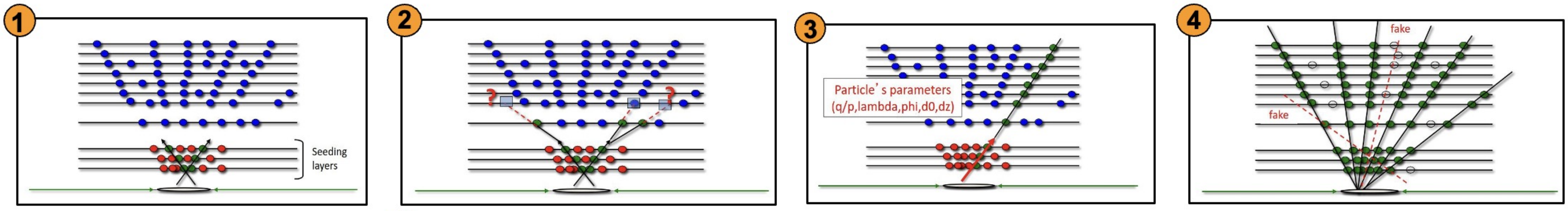
Run 3 offline CMS Tracking Strategy (same Run 2):

1. **Seed Generation:** Uses pixel hits for accurate seeding.
2. **Pattern Recognition:** Kalman filter-based propagation through tracker layers.
3. **Track Fitting:** Inside-out and outside-in Kalman filters with a smoothing stage.
4. **Track Selection:** Based on quality flags (hits, $\chi_2$, vertex distance).

**Iterative Tracking Process:** tracks with higher $p_T$ are reconstructed first, with more complex tracks (low $p_T$, displaced) tackled in later iterations.

**Run 3 Tracking Performance Developments:**

- **Muon Tracking Efficiency Framework**: 50% reduction in CPU time processing via Apache Spark-based approach (CMS-DP-2022-046), inclusion of systematics.
- **Muon Tracking Performance**: Assessed via the produced coously known resonance $Z \rightarrow \mu^+\mu^-$ using the data-driven tag-and-probe technique from 2022-2023 (CMS-DP-2024-054, PoS ICHEP2024 (2025) 931).
- **Detector Failures:** Observed in 2023, affecting tracking efficiency (~99% no longer consistent between data and Monte Carlo simulations, scale factors needed).



# RUN 3 MUON TRACK RECO PERFORMANCE

1. **Define and Classify Probes**: loose criteria, reconstructed with muon system (**standalone muon probe**).
2. Passing Probes: Verify **standalone probes** match tracker tracks in a cone (All Tracks or Tracker-only Seeded).
3. **Define Tags**: Tight identification selection, reconstructed using both muon chambers and the tracker.
4. **Pairing:** Opposite charge sign muons, invariant mass near Z. Exclude interchangeable pairs.
5. **Measured Efficiency** ($\epsilon$): Fraction of pairs with probes passing tighter selection. Background subtraction via simultaneous fit procedure.
6. **Simultaneous fit** to the **signal and the background tag-probe invariant mass** (for both passing and failing probes, data and Madgraph DY MC). Tracker momentum is used for passing probes and tag muons to improve mass resolution.
7. **Fake Matching Rate** $\epsilon_F$: Before probes classification, remove tracks near Z mass (40-200 GeV).

$\epsilon$ : **measured efficiency**
$\epsilon_M$ : **matching efficiency**
$\epsilon_T$ : **true tracking efficiency**
$\epsilon_F$ : **fake matching rate**

$$\epsilon = \epsilon_T\epsilon_M + (1-\epsilon_T\epsilon_M)\,\epsilon_F$$

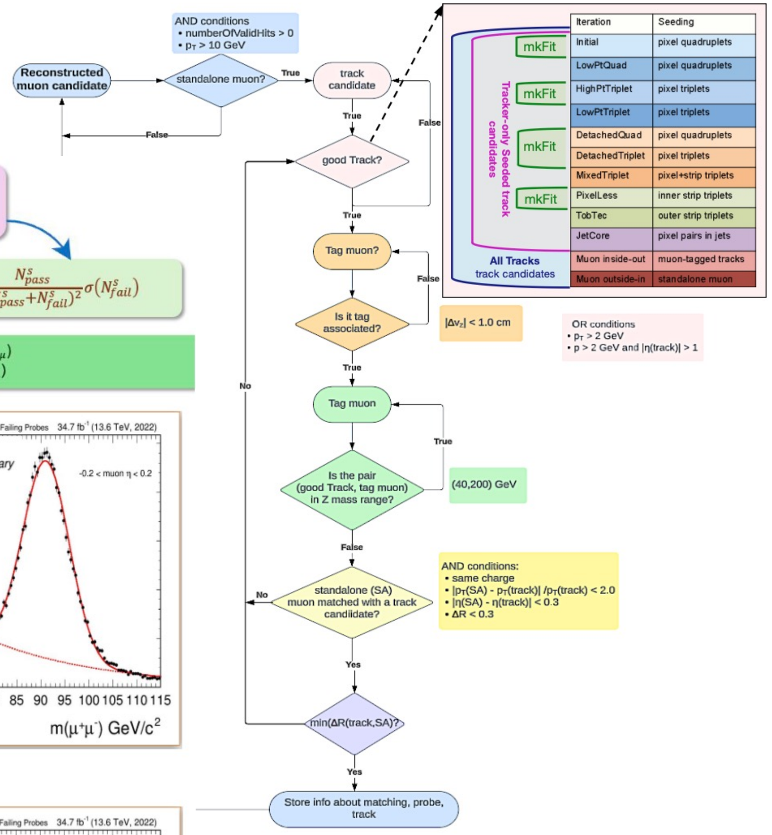
$$1-\epsilon_T\epsilon_M = \frac{1-\epsilon}{1-\epsilon_F}$$

$$N_{pass} = f^s_{pass}(m_{\mu\mu})N^s_{pass} + f^b_{pass}(m_{\mu\mu})N^b_{pass}$$
$$N_{fail} = f^s_{fail}(m_{\mu\mu})N^s_{fail} + f^b_{fail}(m_{\mu\mu})N^b_{fail}$$

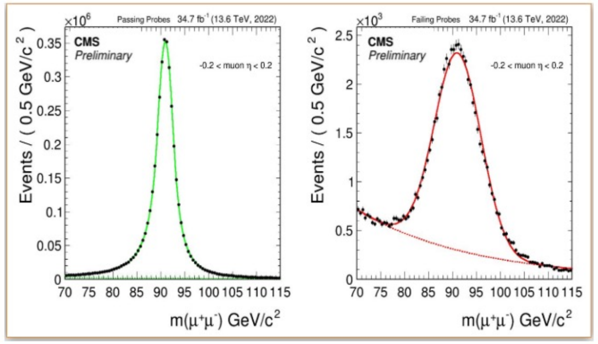
$$\epsilon_{Data}\ (\epsilon_{MC}) = \frac{N^s_{pass}}{(N^s_{pass}+N^s_{fail})}$$

$$\sigma(\epsilon) = \frac{N^s_{fail}}{(N^s_{pass}+N^s_{fail})^2}\sigma(N^s_{pass}) \oplus \frac{N^s_{pass}}{(N^s_{pass}+N^s_{fail})^2}\sigma(N^s_{fail})$$

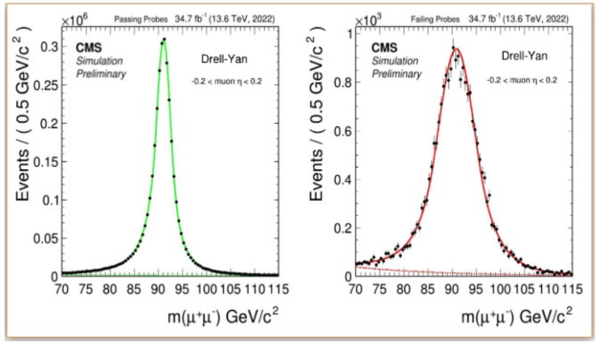
$f^s_{pass}$: fitting function for signal shape in **Passing Probes** = RooDCBShape(*)($m_{\mu\mu}$)
$f^s_{fail}$: fitting function for signal shape in **Failing Probes** = RooDCBShape(*)($m_{\mu\mu}$)




efficiency = (98.79 ± 0.01) %
$\chi^2$/ndof pass: 12.982, fail: 2.686
--- parameters
Passing Probes
- nBkgP = (156.01 ± 2.99) × $10^3$
- nSigP = (4046.16 ± 3.59) × $10^3$
Failing Probes
- nBkgF = (29.11 ± 0.21) × $10^3$
- nSigF = (49.51 ± 0.25) × $10^3$




efficiency = (99.38 ± 0.01) %
$\chi^2$/ndof pass: 8.260, fail: 1.530
--- parameters
Passing Probes
- nBkgP = (117.66 ± 2.30) × $10^3$
- nSigP = (3399.07 ± 2.92) × $10^3$
Failing Probes
- nBkgF = (1.34 ± 0.24) × $10^3$
- nSigF = (21.32 ± 0.28) × $10^3$



# TRACKING PERFORMANCE RESULTS

*PoS* ICHEP2024 (2025) 931

## 0. Tracking Matching Inefficiency ($1 - \epsilon_M$)

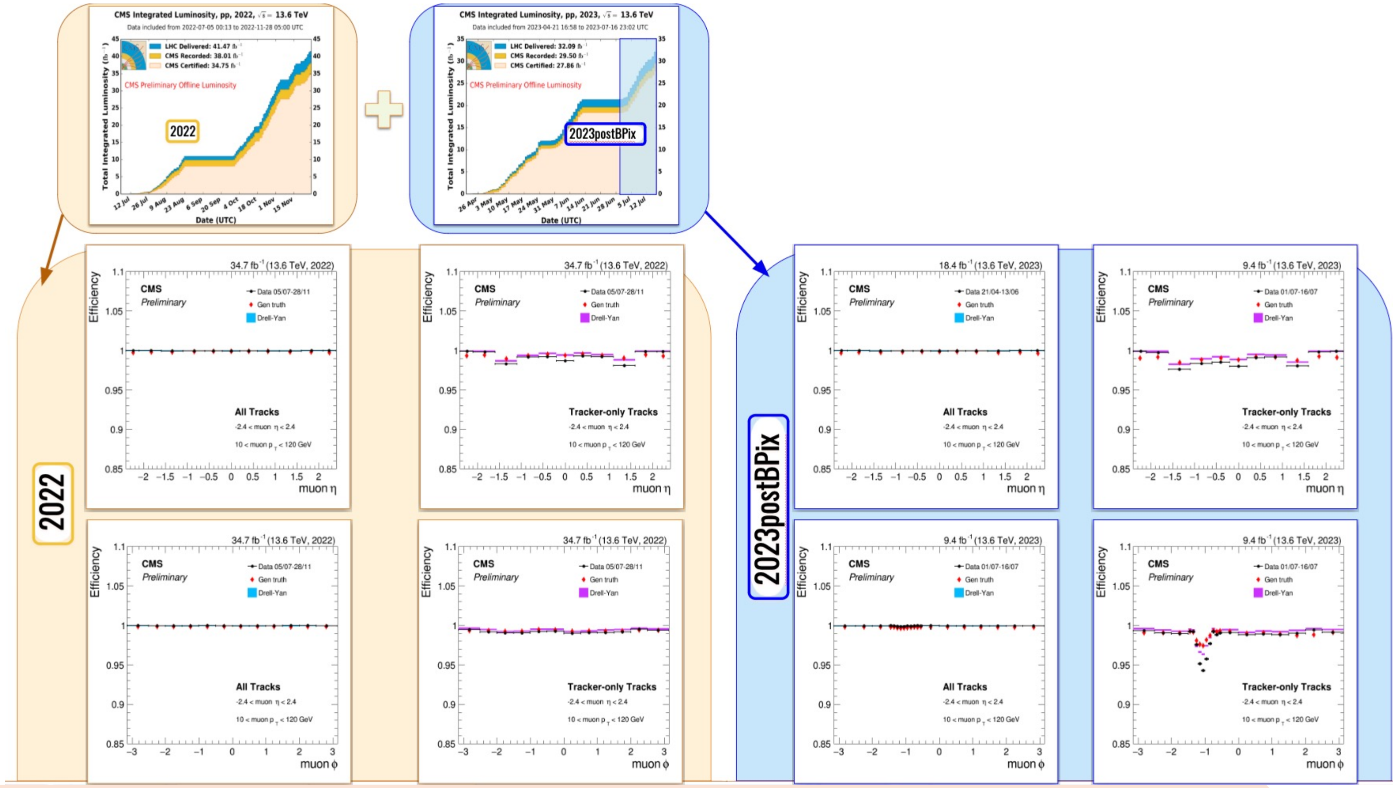
- Estimated from the **tail (upper 20%) of matching variable distributions**: $\Delta R$, $\Delta\eta$, and $\Delta p_T/p_T$ between standalone muons and inner tracks.
- Max inefficiency (~5%) observed in high-$p_T$ forward region [120, 200] GeV.

## 1. Generator Truth vs. Tag & Probe

- **Generator-level muons from $Z \to \mu^+\mu^-$** used as a cross-check.
- **Same selection strategy** applied as for reconstructed muons.
- Confirms T&P performance, **validating measured efficiencies** across 2022–2023 data periods.

## 2. Tracking Performance Results ($\varepsilon_T$)

- Efficiency ($\varepsilon_T$) measured vs $\eta$, $\phi$, $p_T$, and number of Primary Vertices.
- All Tracks (using muon system info) outperform Tracker-only seeded tracks, esp. in problematic $\eta$–$\phi$ regions since July 2023 ($\phi \sim -1$, $-1.5 < \eta < -0.2$).
- Displayed uncertainties include statistical and matching probability $\epsilon_M$ contributions, often smaller than marker size. Systematics related to fit functions can be significant.



6

# HH(bbμμ) Run 2 analysis

- EVENT PRESELECTIONS
- CORRECTIONS TO DATA AND SIMULATION
- DI-HIGGS SIGNAL REGION
- CONTROL REGIONS
- MACHINE LEARNING FOR DI-HIGGS SEARCH
- STATISTICAL ANALYSIS

*"It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong."*
*– Richard Feynman*

# EVENT PRESELECTION

- Collision events passed the **single-muon triggers** (online loose Identification & isolation trigger muon object found with $p_T \geq 24$ GeV (2016/2018) and $\geq 27$ GeV (2017)).
- The first step consists in **selecting only events of interest** → **Preselection**: events with **at least one good primary iteraction vertex** restricted along the z-axis and transverse plane ($z_{PV} < 24$ cm, $r_{PV} < 2$ cm, high number of degrees of freedom > 4).

1. Have two opposite-charge sign muons ($\ell_1, \ell_2$) with
   - Transverse momentum $p_T > 20$ GeV, $|\eta| < 2.4$, medium identification criteria, $|d_{xy}| < 0.5$, $|d_z| < 1.0$
   - Relative Particle-Flow Isolation requirement ($\Delta R = 0.4$) < 0.15 (**Final State Radiation** photons corrected)

2. At least one muon ($\ell_1$ or $\ell_2$) matched ($\Delta R < 0.1$) with the trigger object:
   - $p_T \geq 26$ GeV (2016/2018) and $\geq 29$ GeV (2017)to ensure that it lies on the single muon trigger plateau
   - Tight muon identification criteria
   - Event with exactly one tight muon pair

   To remove the low mass resonances, the dilepton invariant mass is $40 < m_{\ell\ell} < 200$ GeV ($p_T$ GeoFit correction applied)

3. Electrons veto: pT > 20 GeV, $|\eta| < 2.5$, MVA WP(90%), relative isolation ($\Delta R = 0.3$) < 0.15, $|d_{xy}| < 0.5$, $|d_z| < 1.0$

4. Particle-Flow reconstructed jets $p_T > 25$ GeV, $|\eta| < 2.5$ (2.4 for 2016), Loose pileup jet Identification criteria for jets $p_T < 50$ GeV
   - Loose (Tight) jet Identification criteria for 2016 (2017,2018), Cleaned wrt. muons/FSR: $\Delta R(j, \mu/\gamma) > 0.4$
   - Highest AK4 b-tagged (DeepJet algorithm) jets pair invariant mass – no stats for boosted category regime

# CORRECTIONS TO DATA AND SIMULATION

Corrections to data and Simulation **ensure simulation models signal & background accurately**
- Essential for analyses using **MC-driven backgrounds** like in this case

**Muons**

**Muon Efficiency Corrections (T&P Method)**
- Tag & Probe method applied on $Z \longrightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$ events
- Used to derive data-to-MC Scale Factors (SFs) as a function of $p_T$ and $\eta$ for Tracking Efficiency, Reconstruction & identification, Isolation and Trigger Efficiency ~1% correction typically, up to ~2% at high $|\eta|$ $\varepsilon=\varepsilon_{trk}\cdot\varepsilon_{ID|trk}\cdot\varepsilon_{ISO|ID}\cdot\varepsilon_{trig|ISO,ID}$

**Rochester Corrections**
- Applied to **both data and simulation** including **momentum scale correction** and **extra smearing in simulation** for resolution matching with data
- **Aligns Z peak** in di-muon invariant mass to **PDG value**

**GeoFit Correction based on $\eta$ and $d_{xy}$ for muon $p_T$**
- **Tracks** show **residual transverse displacement $d_0$** wrt primary vertex/beamspot leading to **degradation of mass resolution**
- The correction is the proportionality factor of $d_0 \propto \Delta p_T/R^2$
- **Charge-dependent**, opposite for $\mu^+$ vs $\mu^-$
- **Tuned for each Run 2 year** (2016, 2017, 2018)
- Introduced in Run 2 $H \longrightarrow \mu\mu$ CMS search
- Improves di-muon mass resolution **by 3–10%**

**Event weight**

**Pileup Reweighting**
- MC samples produced with approx. pileup profile
- Reweighting applied so MC matches data PU distribution

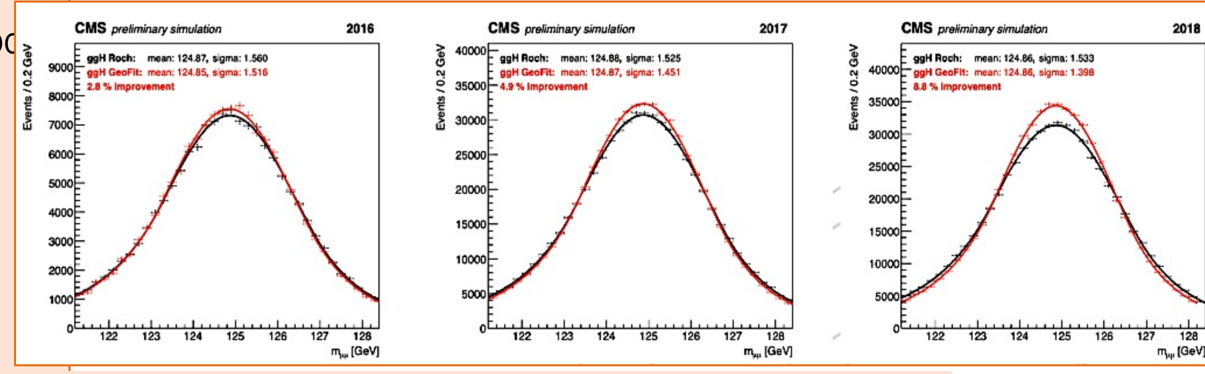
**Event Scaling MC to data luminosity**

**Jets**

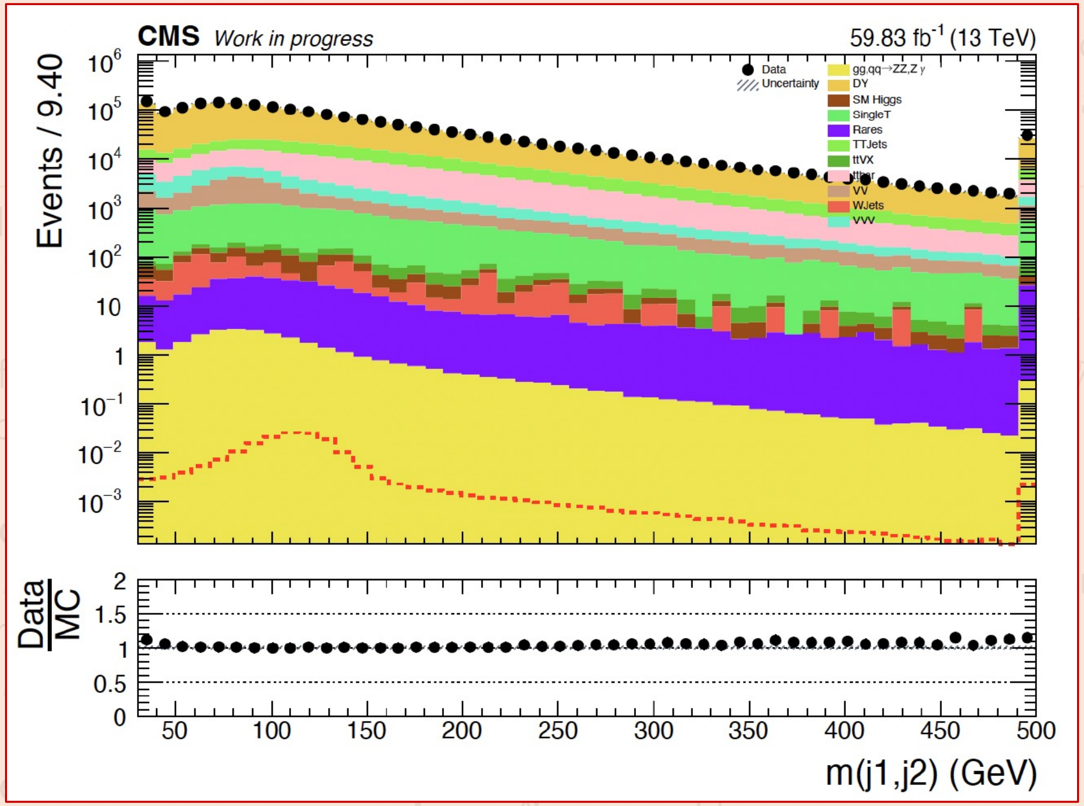
- **L1FastJet correction Jet Energy Resolution smearing** (simulation) : mitigates pileup effects (extra tracks & calorimetric energy)
- **Jet Pileup track removal:** discard tracks from pileup vertices

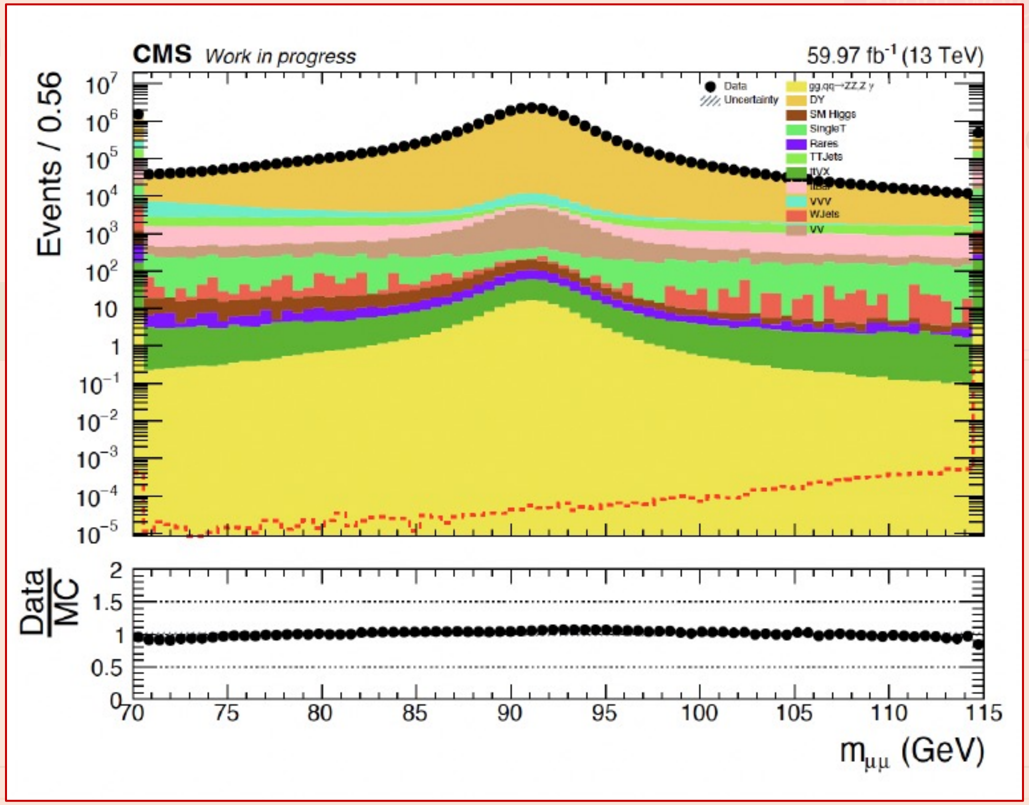
**Jet Energy Corrections (L2 + L3):** applied vs. $\eta$ and pT to match particle-level jets
- **Residual Jet Energy Corrections (data only)** from balance in:
  - Dijet, multijet, $\gamma$+jets, Z+jets events

**B-tag Score Reshaping (since MVA-based analysis)**



# CORRECTIONS TO DATA AND SIMULATION

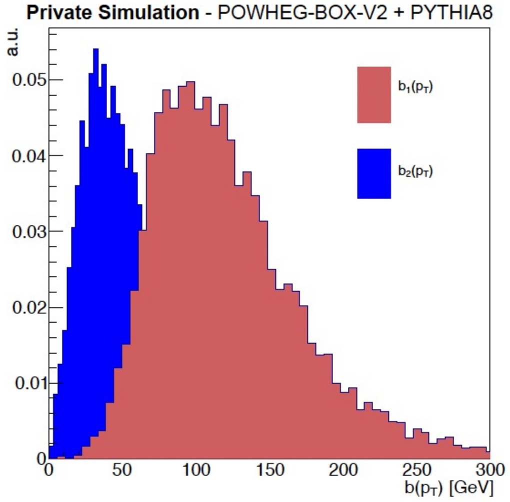


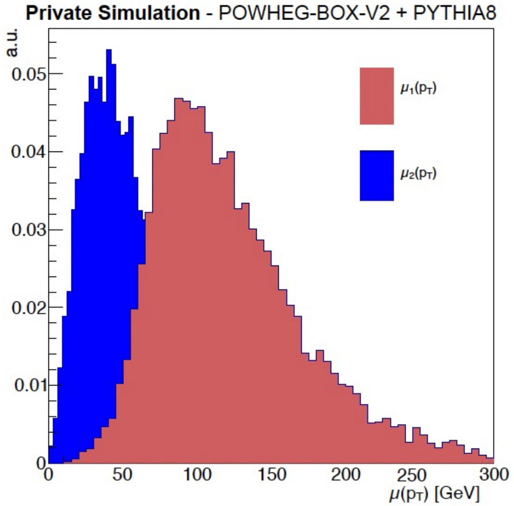




- **Backgrounds (bkgs)** are those SM processes whose **signature mimics the same of the signal**
- DY is the **most significant reducible background** (2 leptons, no resonant di-jets) **at this selection step**

# DI-HIGGS SIGNAL REGION

- Final state: **2 high-pT, isolated muons + 2 b-jets**
- Previous preselections are motivated by **generator level studies**: from Higgs decay at rest ($p_T^H \approx 0$) $\longrightarrow$ muons emitted back-to-back with expected $p_T \sim 60$ GeV



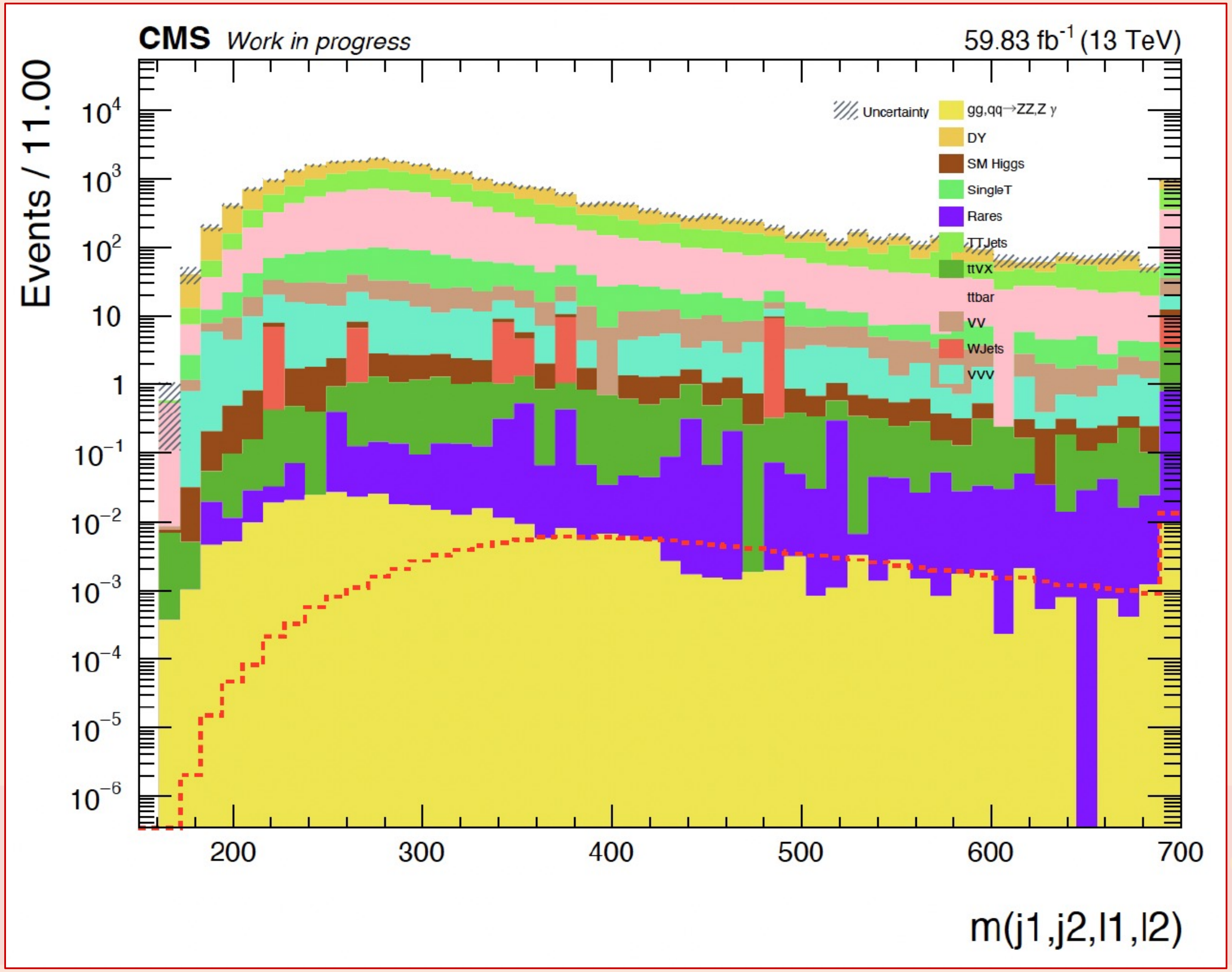




- **Signal Region (SR)** are the **selections where the signal is most likely to be present** and the SM background is reduced:

- To suppress some backgrounds like ttbar and non resonant dibosons: $70 < m_{bb} < 140$ GeV
- Di-muon Higgs candidate: $115 < m_{\ell\ell} < 135$ GeV to be decorrelated from bbZZ and bbWW searches (narrow width of approximately 2.49 GeV for Z boson and broad and countinous for W $\longrightarrow$ μν) .

| Background sample | Reaction | Generator + PS | Cross Section $\cdot BR$ (pb) |
|---|---|---|---|
| | **Drell-Yan** | | |
| DY+ jets (10 GeV $< M_{jj} <$ 50 GeV) | $Z \to ll + jets$ | aMCNLO (FXFX) + Pythia8 | 18610 |
| DY+jets ($M_{jj} >$ 50 GeV) | $Z \to ll + jets$ | aMCNLO (FXFX) + Pythia8 | 5954 |
| | **Top** | | |
| $t\bar{t}$ | $t\bar{t} \to +2l + 2\nu$ | Powheg + Pythia8 | 86.61 |
| | $t\bar{t} \to 2l + 2q$ | Powheg + Pythia8 | 358.57 |
| Single Top | tW, top (antitop) | Powheg + Pythia8 | 35.9 |
| | t-channel top | PowhegV2 + Madspin + pythia8 | 136.02 |
| | s-channel (lepton decays) | aMCNLO + Pythia8 | 3.40 |
| tZq | tZq | aMCNLO + Pythia8 | 0.0758 |
| ttV | $t\bar{t}W \to l\nu + jets$ | aMCNLO (FXFX) + madspin + pythia8 | 0.2001 |
| | $t\bar{t}Z \to 2l + 2\nu$ | aMCNLO + pythia8 | 0.2529 |
| $t\bar{t}$+X | TTTT | aMCNLO + Pythia8 | 0.0091 |
| | TTTW | aMCNLO + Pythia8 | 0.00073 |
| | TTWW | aMCNLO + Pythia8 | 0.0070 |
| | TTTJ | aMCNLO + Pythia8 | 0.000039 |
| | **SM Higgs Background** | | |
| ggF (M125) | $gg \to H \to \mu\mu$ | Powheg + Pythia8 | 0.01057 |
| VBF (M125) | $q\bar{q} \to H \to \mu\mu$ | Powheg + Pythia8 | 0.0008228 |
| $W^+H$ (M125) | $q\bar{q} \to W^+H \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001858 |
| $W^-H$ (M125) | $q\bar{q} \to W^-H \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001164 |
| ZH (M125) | $q\bar{q} \to ZH \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001923 |
| ggZH (M125) | $gg \to ZH \to \mu\mu + X$ | Powheg + Pythia8 | $2.0 \times 10^{-6}$ |
| ttH (M125) | $gg \to t\bar{t}H \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001103 |
| bbH (M125) | $gg \to bbH \to \mu\mu + X$ | amcatnlo + Pythia8 | 0.0001059 |
| THQ (M125) | $tHq \to \mu\mu + X$ | Madgraph + Pythia8 | $1.612 \times 10^{-5}$ |
| THW (M125) | $tHW \to \mu\mu + X$ | Madgraph + Pythia8 | $3.292 \times 10^{-6}$ |
| | **Boson, Diboson** | | |
| bbZ | $Z + b\bar{b} \to X + b\bar{b}$ | aMCNLO + pythia8 | 45.9 |
| Z + (0,1,2) jets | $Z \to l\nu + jets$ | aMCNLO + pythia8 | 45.0 |
| W + (0,1,2) jets | $W \to l\nu + jets$ | aMCNLO + pythia8 | 12.178 |
| VV | $WW \to 2l + 2\nu$ | Powheg + Pythia8 | 12.178 |
| | $WZ \to 3l + \nu$ | Powheg + Pythia8 | 4.102 |
| | WZ $WZ \to 2l + 2q$ 2l2q | aMCNLO (FXFX) + madspin + pythia8 | 5.606 |
| | $ZZ \to 2l + 2\nu$ | Powheg + Pythia8 | 0.5644 |
| | $ZZ \to 2l + 2q$ | aMCNLO (FXFX) + Madspin + pythia8 | 3.224 |
| | $ZZ \to 4l$ | Powheg + Pythia8 | 1.256 |
| $gg \to ZZ$ | $gg \to ZZ(2e2\mu, 2\mu2\tau, 2e2\tau)$ | MCFM701 + Pythia8 | 0.00318 |
| | $gg \to ZZ(2e2\mu, 2\mu2\tau, 2e2\tau)$ | MCFM701 + Pythia8 | 0.00318 |
| | $gg \to ZZ(4\mu, 4\tau)$ | MCFM701 + pythia8 | 0.00157 |
| | $gg \to ZZ(2\mu2\nu)$ | MCFM701 + pythia8 | 0.00149 |
| $q\bar{q} \to ZZ$ | $q\bar{q} \to ZZ$ | Powheg + Pythia8 | 1.256 |
| | **Triboson** | | |
| VVV | WWW | aMCNLO + Pythia8 | 0.2086 |
| | WWZ | aMCNLO + Pythia8 | 0.1651 |
| | WZZ | aMCNLO + Pythia8 | 0.05565 |
| | ZZZ | aMCNLO + Pythia8 | 0.01398 |

# DI-HIGGS SIGNAL REGION




Signal region blinded to real data due to CMS policy

| Background sample | Reaction | Generator + PS | Cross Section $\cdot BR$ (pb) |
|---|---|---|---|
| **Drell-Yan** | | | |
| DY+ jets (10 GeV $< M_{jj} <$ 50 GeV) | $Z \to ll + jets$ | aMCNLO (FXFX) + Pythia8 | 18610 |
| DY+jets ($M_{jj} >$ 50 GeV) | $Z \to ll + jets$ | aMCNLO (FXFX) + Pythia8 | 5954 |
| **Top** | | | |
| $t\bar{t}$ | $t\bar{t} \to +2l + 2\nu$ | Powheg + Pythia8 | 86.61 |
| | $t\bar{t} \to 2l + 2q$ | Powheg + Pythia8 | 358.57 |
| Single Top | tW, top (antitop) | Powheg + Pythia8 | 35.9 |
| | t-channel top | PowhegV2 + Madspin + pythia8 | 136.02 |
| | s-channel (lepton decays) | aMCNLO + Pythia8 | 3.40 |
| tZq | tZq | aMCNLO + Pythia8 | 0.0758 |
| ttV | $t\bar{t}W \to l\nu + jets$ | aMCNLO (FXFX) + madspin + pythia8 | 0.2001 |
| | $t\bar{t}Z \to 2l + 2\nu$ | aMCNLO + pythia8 | 0.2529 |
| $t\bar{t}$+X | TTTT | aMCNLO + Pythia8 | 0.0091 |
| | TTTW | aMCNLO + Pythia8 | 0.00073 |
| | TTWW | aMCNLO + Pythia8 | 0.0070 |
| | TTTJ | aMCNLO + Pythia8 | 0.000039 |
| **SM Higgs Background** | | | |
| ggF (M125) | $gg \to H \to \mu\mu$ | Powheg + Pythia8 | 0.01057 |
| VBF (M125) | $q\bar{q} \to H \to \mu\mu$ | Powheg + Pythia8 | 0.0008228 |
| $W^+H$ (M125) | $q\bar{q} \to W^+H \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001858 |
| $W^-H$ (M125) | $q\bar{q} \to W^-H \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001164 |
| ZH (M125) | $q\bar{q} \to ZH \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001923 |
| ggZH (M125) | $gg \to ZH \to \mu\mu + X$ | Powheg + Pythia8 | $2.0 \times 10^{-6}$ |
| ttH (M125) | $gg \to t\bar{t}H \to \mu\mu + X$ | Powheg + Pythia8 | 0.0001103 |
| bbH (M125) | $gg \to bbH \to \mu\mu + X$ | amcatnlo + Pythia8 | 0.0001059 |
| THQ (M125) | $tHq \to \mu\mu + X$ | Madgraph + Pythia8 | $1.612 \times 10^{-5}$ |
| THW (M125) | $tHW \to \mu\mu + X$ | Madgraph + Pythia8 | $3.292 \times 10^{-6}$ |
| **Boson, Diboson** | | | |
| bbZ | $Z + b\bar{b} \to X + b\bar{b}$ | aMCNLO + pythia8 | 45.9 |
| Z + (0,1,2) jets | $Z \to l\nu + jets$ | aMCNLO + pythia8 | 45.0 |
| W + (0,1,2) jets | $W \to l\nu + jets$ | aMCNLO + pythia8 | 12.178 |
| [illegible] | $WW \to 2l + 2\nu$ | Powheg + Pythia8 | 12.178 |
| | [illegible] | [illegible] | [illegible] |
| | WZ $WZ \to 2l + 2q$ 2l2q | aMCNLO (FXFX) + madspin + pythia8 | 5.606 |
| | $ZZ \to 2l + 2\nu$ | Powheg + Pythia8 | 0.5644 |
| | $ZZ \to 2l + 2q$ | aMCNLO (FXFX) + Madspin + pythia8 | 3.224 |
| | $ZZ \to 4l$ | Powheg + Pythia8 | 1.256 |
| $gg \to ZZ$ | $gg \to ZZ(2e2\mu, 2\mu2\tau, 2e2\tau)$ | MCFM701 + Pythia8 | 0.00318 |
| | $gg \to ZZ(2e2\mu, 2\mu2\tau, 2e2\tau)$ | MCFM701 + Pythia8 | 0.00318 |
| | $gg \to ZZ(4\mu, 4\tau)$ | MCFM701 + pythia8 | 0.00157 |
| | $gg \to ZZ(2\mu2\nu)$ | MCFM701 + pythia8 | [illegible] |
| $q\bar{q} \to ZZ$ | $q\bar{q} \to ZZ$ | Powheg + Pythia8 | [illegible] |
| **Triboson** | | | |
| VVV | WWW | aMCNLO + Pythia8 | [illegible] |
| | WWZ | aMCNLO + Pythia8 | [illegible] |
| | WZZ | aMCNLO + Pythia8 | [illegible]5565 |
| | ZZZ | aMCNLO + Pythia8 | 0.01398 |

# CONTROL REGIONS

Control Regions (CRs): the main bkgs are estimated in regions far away (or thogonal) from the SRs:

- High purity of the bkg $\longrightarrow$ high probability to find the bkg
- Low signal contamination $\longrightarrow$low probability to find signal
- From CRs scale factors (SFs) are (eventually) extracted and applied to SRs

Validation Regions (VRs): before applying SFs to the SRs, the estimate is validated$\longrightarrow$ VRs are kinematically between CRs and SRs

## 1. CR Higgs to Muon Pairs + at least two b-jets (check for lepton observables):

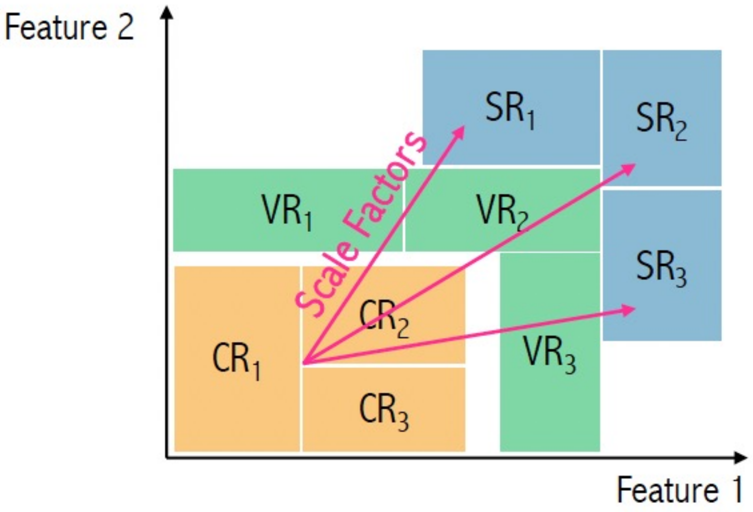
1. Di-muon mass in the Higgs window ($115 < m_{\ell\ell} < 135$ GeV ).
2. b-jet mass outside the Higgs range ($m_{bb} < 70$ GeV or $m_{bb} > 140$ GeV).
3. Main Background: Z + jets and Zbb production.

## 2. CR Higgs to b-jets (check jet observables):

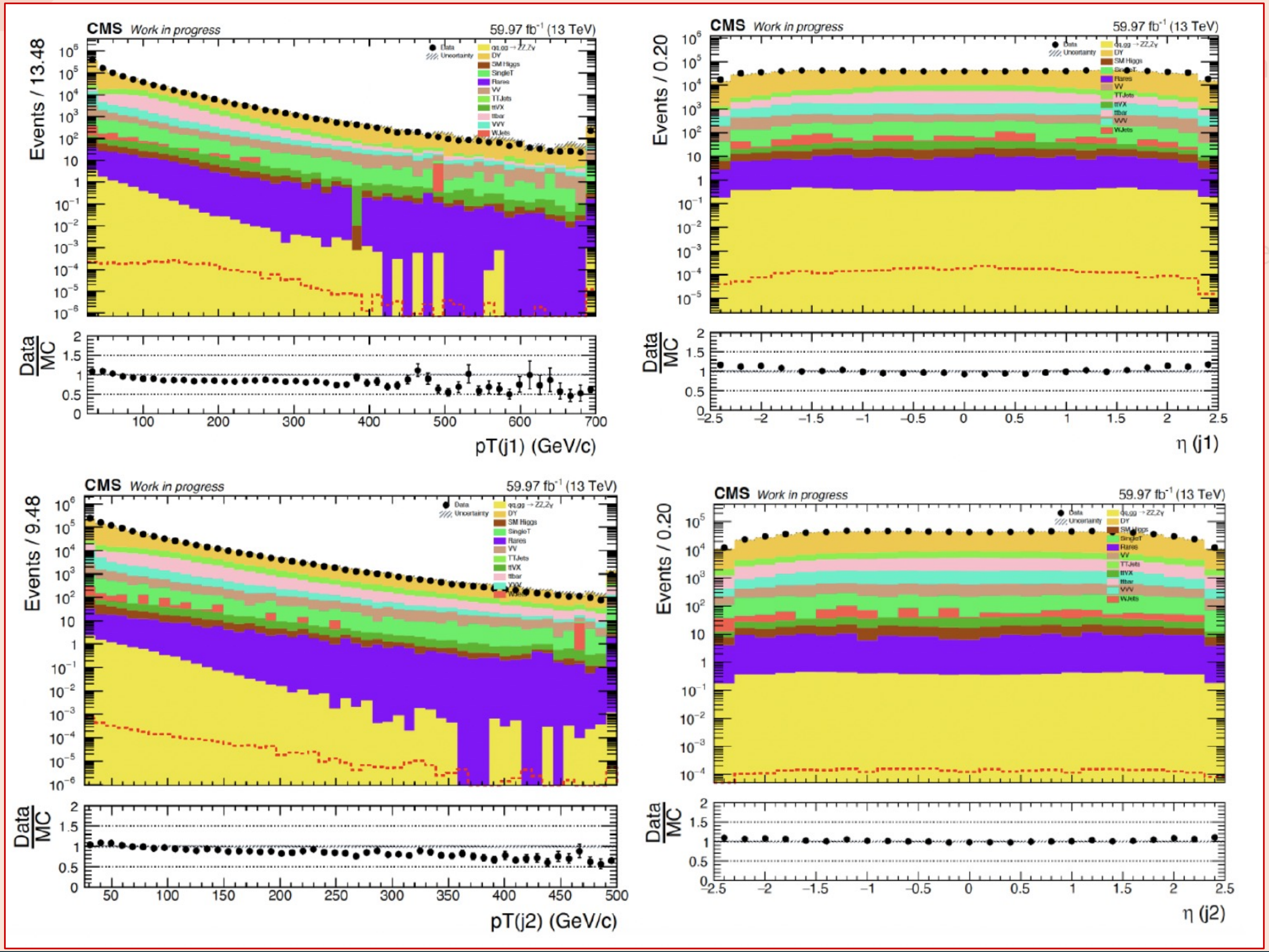
1. Di-muon mass outside Higgs mass window ($m_{\ell\ell} < 115$ GeV or $m_{\ell\ell} > 135$ GeV).
2. b-jet mass within the Higgs window ($70 < m_{bb} < 140$ GeV).

## 3. CR Muon Pairs and Two b-jets Sidebands:

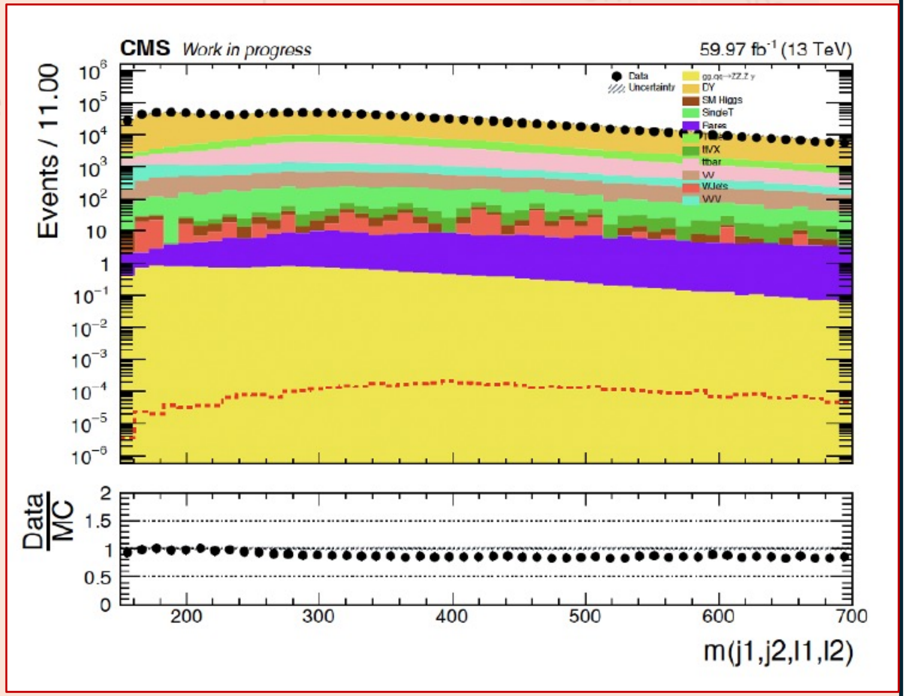
1. Di-muon mass outside the Higgs window ($m_{\ell\ell} < 115$ GeV or $m_{\ell\ell} > 135$ GeV).
2. b-jet mass outside the $H \longrightarrow bb$ window ($m_{bb} < 70$ GeV or $m_{bb} > 140$ GeV).
3. Double-check lepton and jet observables.



# CONTROL REGIONS




Sideband CR



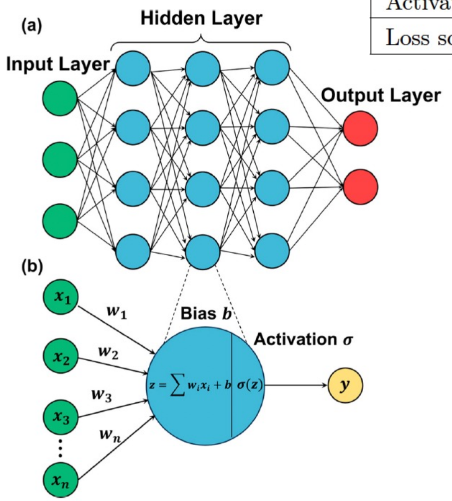

tagging score distributions, especially at high pT for b-jets.

- However, these discrepancies are ...zed by the ANN training that uses the ... set of inputs.

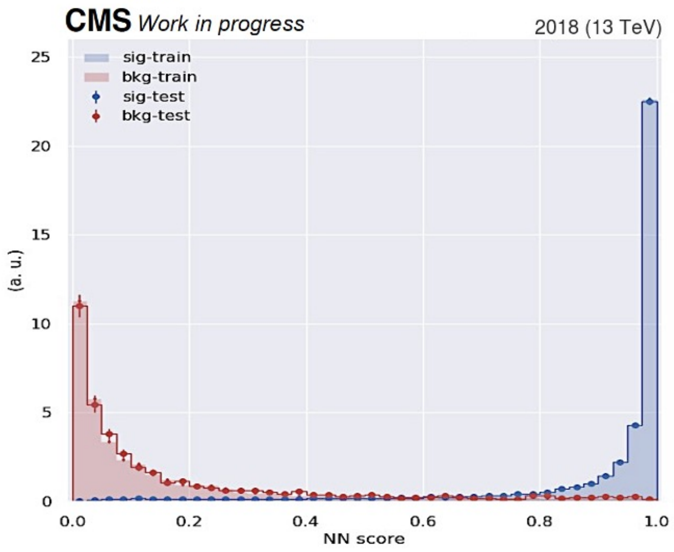
# MACHINE LEARNING FOR DI-HIGGS SEARCH

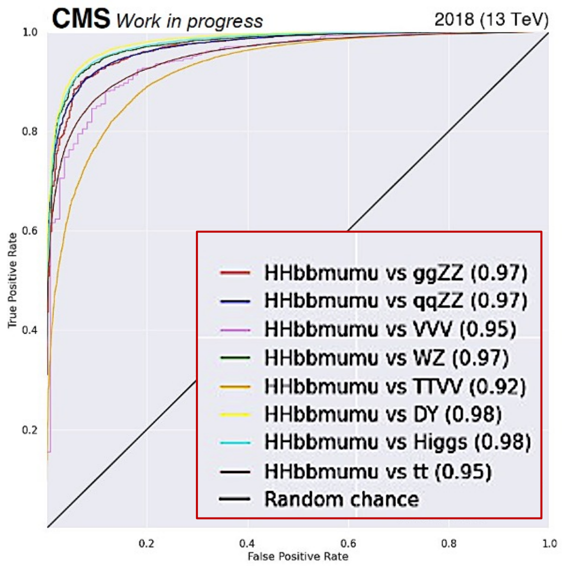
- Use **machine learning** to **enhance signal/background separation** in the HH signal region. **Several Algorithms Tested**: Neural Networks (NNs), Boosted Decision Trees (BDTs), Random Forests (RF).
- Strategy for **optimizing the hyper-parameter of binary classifiers for rare di-Higgs searches** in **Keras by choosing one or more metrics** (e.g. efficiency times purity) – Lighting talk at ACAT, J.Phys.Conf.Ser. 2438 (2023). SM signal optimization.

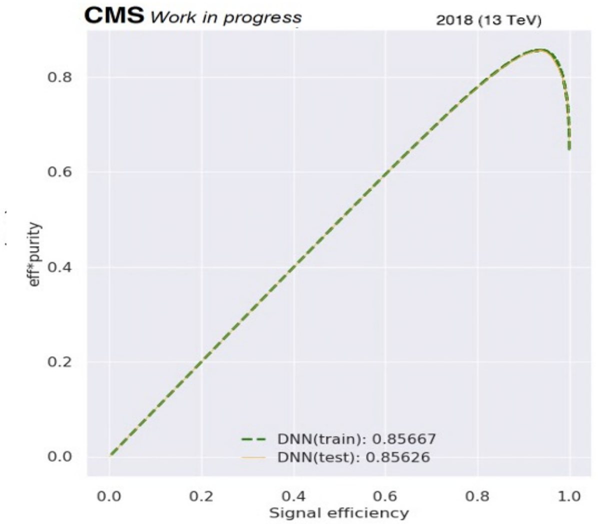
| Parameter | Explored Values |
|---|---|
| Inputs | Lepton/jet ($p_T$, $\eta$, b-tag score, b-tag WPs)/H/HH($p_T$,$m$) |
| Dropout rate | 10%, 20%, 30%, 40% |
| Topologies | 30:20:10, 21:13:8, 10:10:10:10, 50:50:50:50, ... |
| Early stopping | 50, 100, 600, 3000 epochs |
| Optimizers | SGD, Adam, Adagrad, AdaDelta, RMSprop |
| Batch size | 5, 32, 64, 128, 786 |
| Activation functions | ReLU, SeLU |
| Loss scaling | XS (process cross-section), event weights, none |

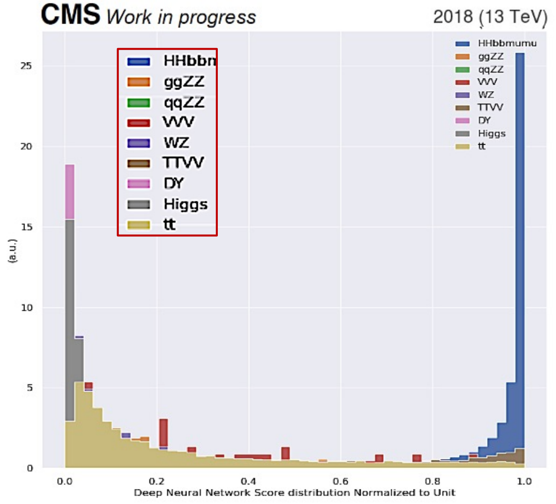



Artificial Neural Network (ANN):
- Best suited for **capturing complex, multidimensional correlations in a high-dimensional phase space**.
- No overfitting observed.
- Feature importance and correlation matrix for feature study.




- **Efficiency:** fraction of true signal events that are correctly classified as signal
- **Purity:** the fraction of true signal events in the total number of events predicted as signal

# STATISTICAL ANALYSIS

The analysis uses **binned distributions of NN outputs** per each physics process to **quantify the likelihood of an event being signal-like, optimizing signal-background separation**.

**Modeling the Hypotheses:**

1. **Background only hypothesis**: Based on SM simulations for background processes.
2. **Signal+background hypothesis**: Includes both signal (SM or HEFT) and background predictions.

**Binned Analysis:**

1. Data divided into bins (10 bins for statistical stability), each reflecting different signal-to-background ratios.

**Likelihood Construction:**

1. Compares **observed event** counts with **predictions under both hypotheses.**
2. **Asimov Dataset** used for test statistics in a blind analysis.
   - the median expected data for which the yields of all data samples are set to their expected values and the nuisance parameters are set at their nominal value
3. **Systematic uncertainties:** detector effects, theoretical model uncertainties, and other sources are incorporated as nuisance parameters in the likelihood.

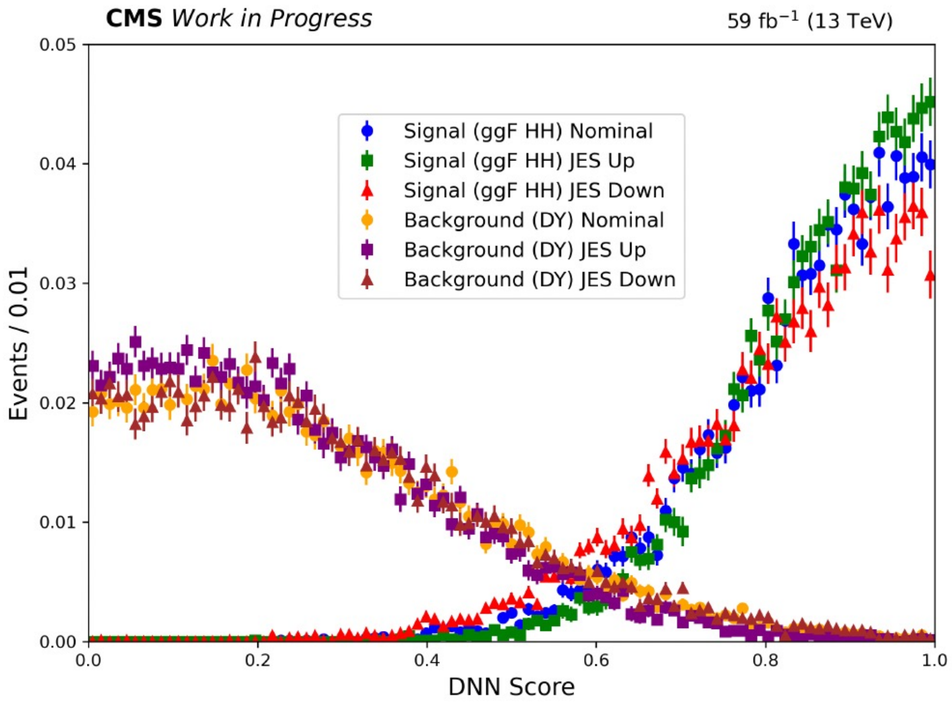
- The uncertainties are **propagated through the DNN score prediction** with shifts of 1σ from the nominal values for each input variable, resulting in a range of output distributions that for each event is **N(Inputs) [1 + 2 N(InputUncertainties)]**
- **Impact of the systematics uncertainty on the uncertainty on the parameter of interst (POI) checked**

## Theoretical and experimental systematics

| Process | Cross Section (pb) | Order | +QCD scale unc. (%) | -QCD scale unc. (%) | +(PDF+$\alpha_S$) unc. (%) | -(PDF+$\alpha_S$) unc. (%) |
|---|---|---|---|---|---|---|
| DY+jets | 18610 | NLO (QCD) + NLO (EWK) | +20.0 | -20.0 | +10.0 | -10.0 |
| tt | 83 | NNLO (QCD) | +20.5 | -30.0 | +21.0 | -21.0 |
| bbZ | 45.9 | NLO (QCD) | +20.0 | -20.0 | +10.0 | -10.0 |
| ggH | 48.58 | N3LO (QCD) + NLO (EWK) | +4.6 | -6.7 | +3.2 | -3.2 |
| VBF | 3.782 | NNLO (QCD) + NLO (EWK) | +0.4 | -0.3 | +2.1 | -2.1 |
| WH | 1.373 | NNLO (QCD) + NLO (EWK) | +0.5 | -0.7 | +1.9 | -1.9 |
| qq → ZH | 0.761 | NNLO (QCD) + NLO (EWK) | +0.5 | -0.6 | +1.9 | -1.9 |
| gg → ZH | 0.123 | NLO (QCD) + NLO (EWK) | +25.1 | -18.9 | +2.4 | -2.4 |
| ttH | 0.507 | NLO (QCD) + NLO (EWK) | +5.8 | -9.2 | +3.6 | -3.6 |
| bbH | 0.488 | NNLO (QCD) | +20.2 | -23.9 | — | — |
| tHq | 0.074 | NLO (QCD) + NLO (EWK) | +6.5 | -14.9 | +3.7 | -3.7 |
| tHW | 0.015 | NLO (QCD) + NLO (EWK) | +4.9 | -6.7 | +6.3 | -6.3 |

| Source | Production Process | Magnitude (%) |
|---|---|---|
| BR($H \to \mu\mu$) | All | 1.23 |
| BR($H \to bb$) | All | 1.5 |
| BR($H \to WW$) | All | 1.2 |
| BR($H \to \tau\tau$) | All | 1.6 |
| BR($H \to ZZ \to 4l$) | All | 1.5 |

| Uncertainty Type | 2016 (%) | 2017 (%) | 2018 (%) |
|---|---|---|---|
| Luminosity | 2.6 | 2.3 | 2.5 |
| Pileup reweighting | shape | shape | shape |
| Trigger | shape | shape | shape |
| Leptons ID,ISO, trig, reco efficiency | shape | shape | shape |
| Jet energy scale (JES) | shape | shape | shape |
| Jet energy resolution (JER) | shape | shape | shape |
| B-tagging score | shape | shape | shape |



# STATISTICAL ANALYSIS

The analysis uses **binned distributions of NN outputs** per each physics process to **quantify the likelihood of an event being signal-like, optimizing signal-background separation**.

**Modeling the Hypotheses:**

1. **Background only hypothesis**: Based on SM simulations for background processes.
2. **Signal+background hypothesis**: Includes both signal (SM or HEFT) and background predictions.

**Binned Analysis:**

1. Data divided into bins (10 bins for statistical stability), each reflecting different signal-to-background ratios.

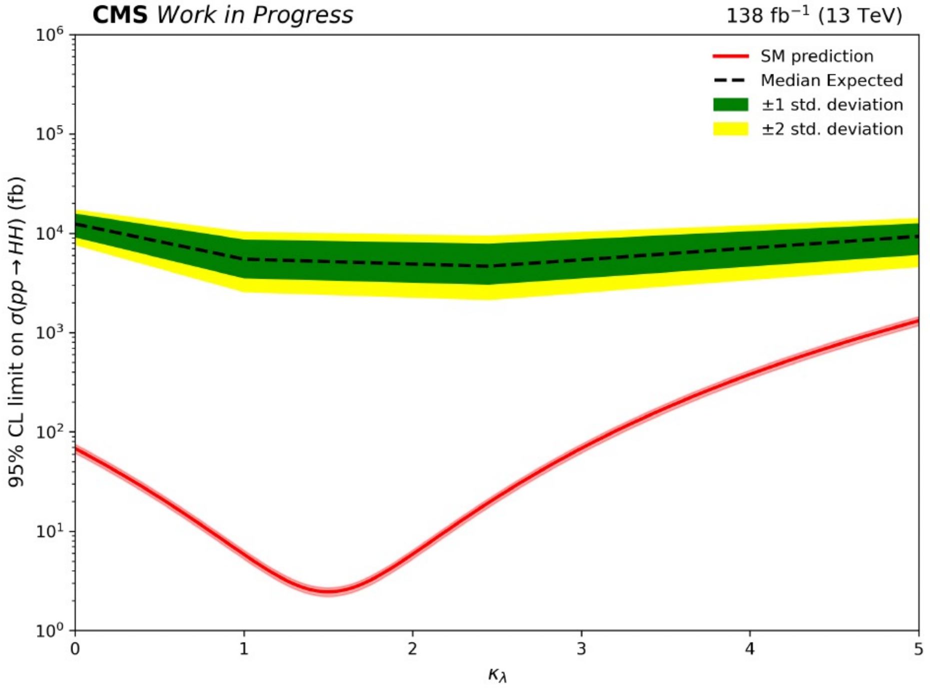
**Likelihood Construction:**

1. Compares **observed event** counts with **predictions under both hypotheses.**
2. **Asimov Dataset** used for test statistics in a blind analysis.
   - the median expected data for which the yields of all data samples are set to their expected values and the nuisance parameters are set at their nominal value
3. **Systematic uncertainties:** detector effects, theoretical model uncertainties, and other sources are incorporated as nuisance parameters in the likelihood.

**Test Statistic:**

profile likelihood ratio quantifies data agreement with each hypothesis and used to set a 95% Confidence Level (CL) upper limit on the di-Higgs signal strength $\mu = \sigma^{exp}(HH)/\ \sigma^{SM}(HH)$ via **Asymptotic Approximations:**

• **Wald's Theorem:** For large sample sizes, the test statistic follows a **non-central** $\chi^2$ distribution.



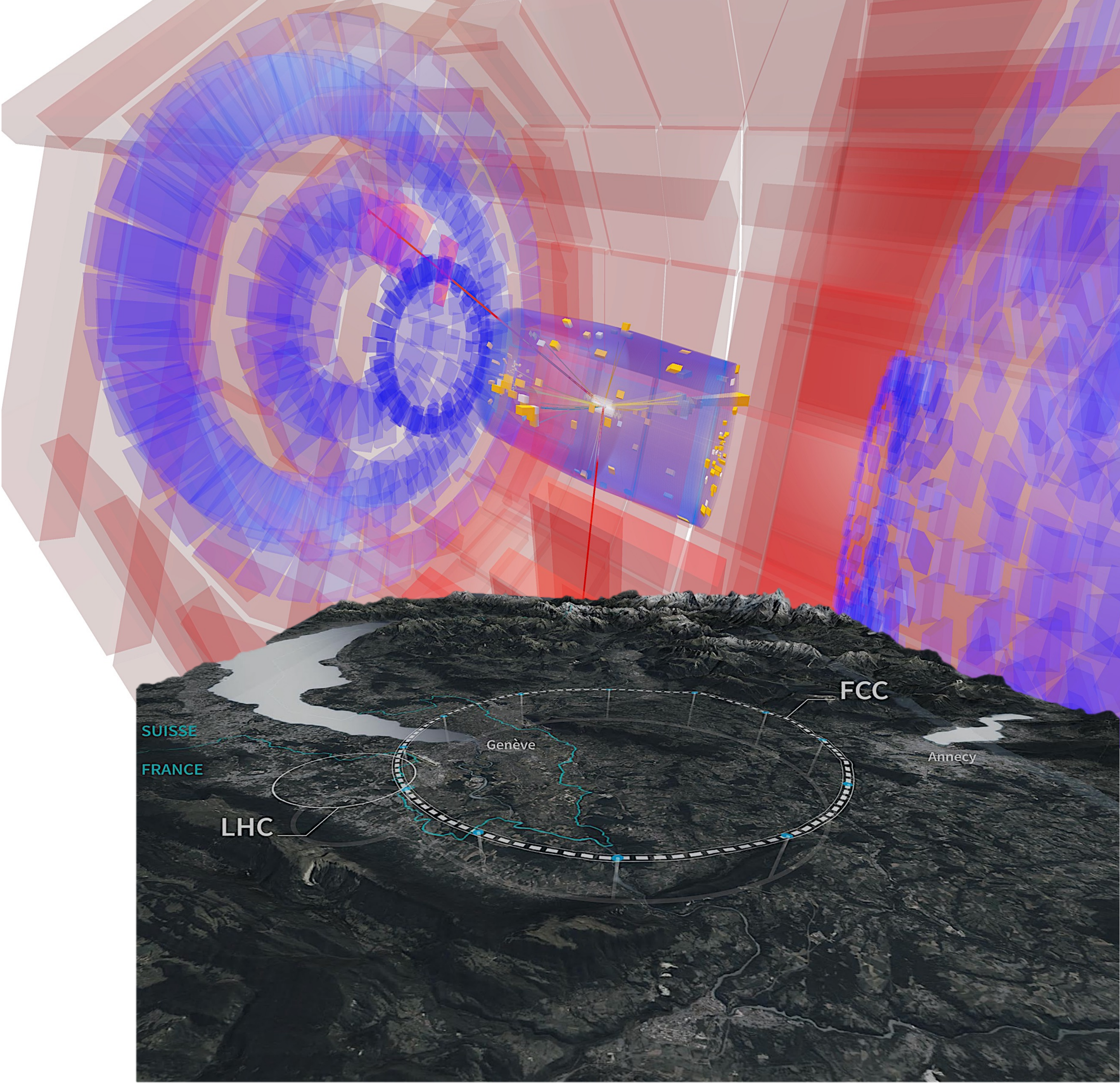

| Signal Strength ($\mu$) | Run 2 combined (137 fb$^{-1}$) | Run 3 (500 fb$^{-1}$) | HL-LHC (3000 fb$^{-1}$) |
|---|---|---|---|
| 2.5% quantile ($\mu - 2\sigma$) | 84 | 44 | 18 |
| 16.0% quantile ($\mu - 1\sigma$) | 116 | 61 | 25 |
| 50% quantile ($\mu$) | 177 | 93 | 38 |
| 84% quantile ($\mu + 1\sigma$) | 275 | 144 | 59 |
| 97.5% quantile ($\mu + 2\sigma$) | 329 | 171 | 70 |

The signal strength is typically assumed to improve according to the square root of the integrated luminosity $\sqrt{L}$. This is because the statistical uncertainty on the number of signal events scales with the square root of the number of events, and this is proportional to the luminosity.

7

# Conclusions

- SUMMARY AND FINAL CONCLUSIONS



# SUMMARY AND FINAL CONCLUSIONS

1. **Results of the gluon-gluon fusion non-resonant di-Higgs production in HH → bbμμ using $\sqrt{s}$ =13 TeV proton-proton collisions collected by the CMS experiment during Run 2 of the LHC:**

- For the first time inside an experimental collaboration, using full simulation and reconstruction algorithms.
- A multivariate approach using artificial neural networks was developed to improve the sensitivity
- Testing both SM and BSM scenarios (HEFT): 95% CL upper limit on signal strength with full Run 2 and projections for HL-LHC (3000 $fb^{-1}$) showing to be valuable for inclusion in a combination with other channels to improve the overall HH sensitivity.

2. **Results of tracking performance using $\sqrt{s}$ =13.6 TeV proton-proton collisions collected by the CMS experiment during Run 3 of the LHC and tracking seeding algorithm developments**

- Muon Tracking performance evaluation improved by including systematics, generator level truth study and new refined Tag and Probe procedure (50% CPU processing time scale)
- Reconstruction of charge particle track seeding algorithm, the CA, has been improved and more robust against detector aging, with the extension to outracker information, leading to a 20% efficiency gain, 50% reduction in fake tracks if using CA autotuner

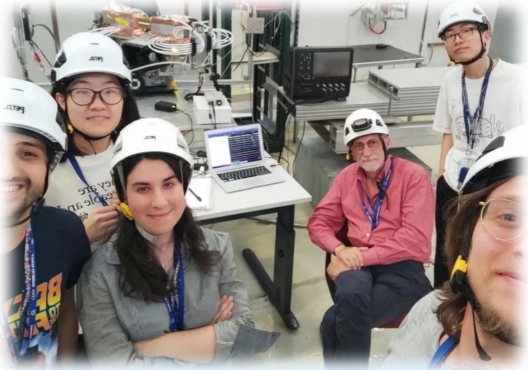
3. **A novel particle identification technique is proposed in a drift chamber for future collider detectors.**

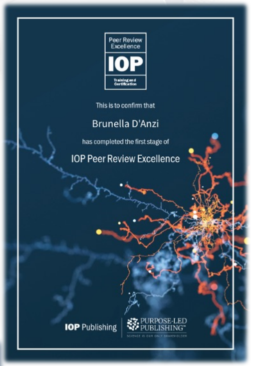
- A novel cluster counting technique for particle ID was developed for future colliders (e.g., FCC, CEPC), showing a 2 times improvement in charged particle separation (muons) under high pileup conditions

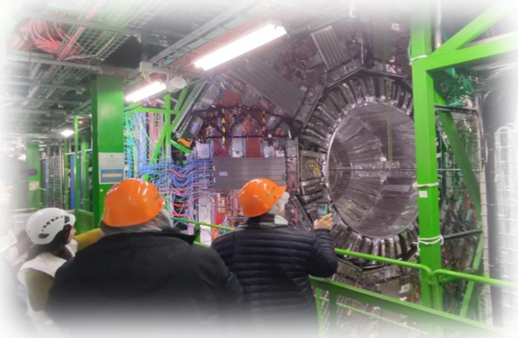
# Thank you for the attention

SPECIAL THANKS TO THE CMS DI-HIGGS, CMS TRACKING AND R&D FCC BARI & LECCE WORKING GROUPS!

THANKS TO UNIBA AND INFN FOR THE MANY EXPERIENCES RELATED TO THIS RESEARCH!



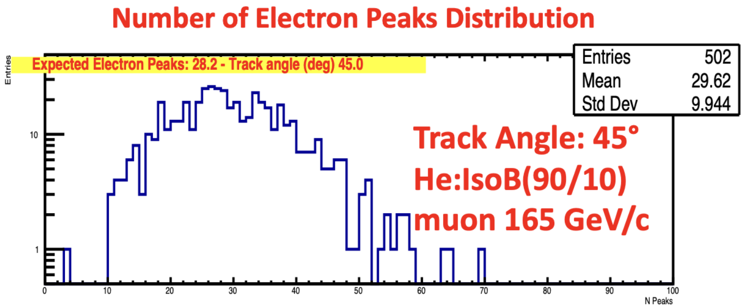



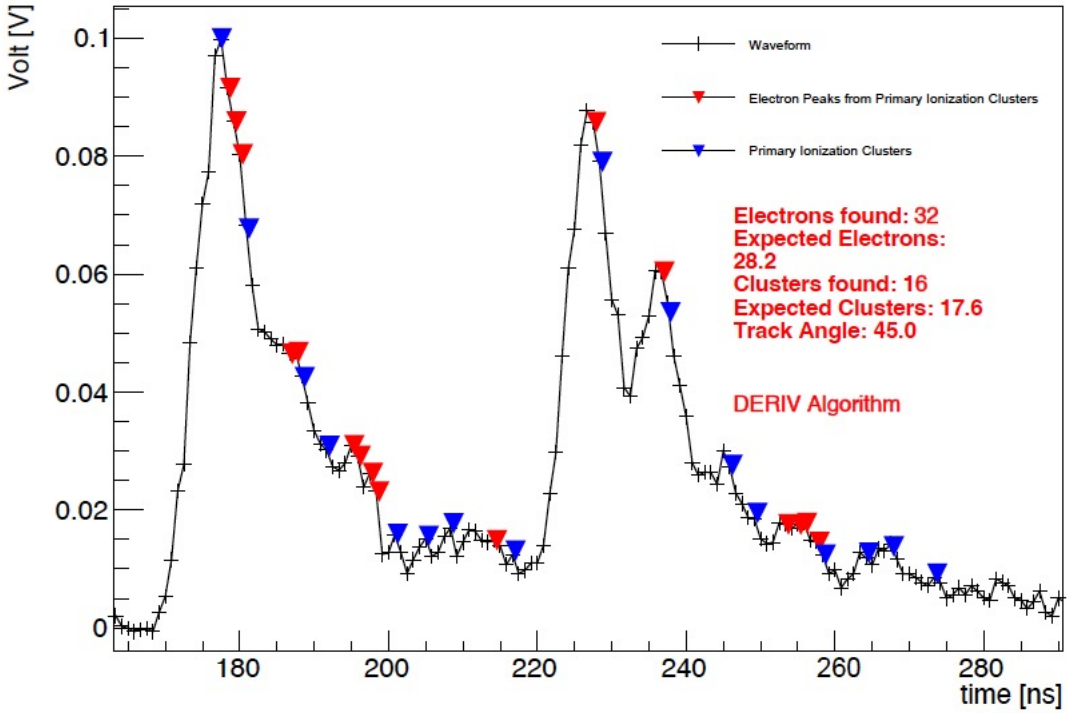



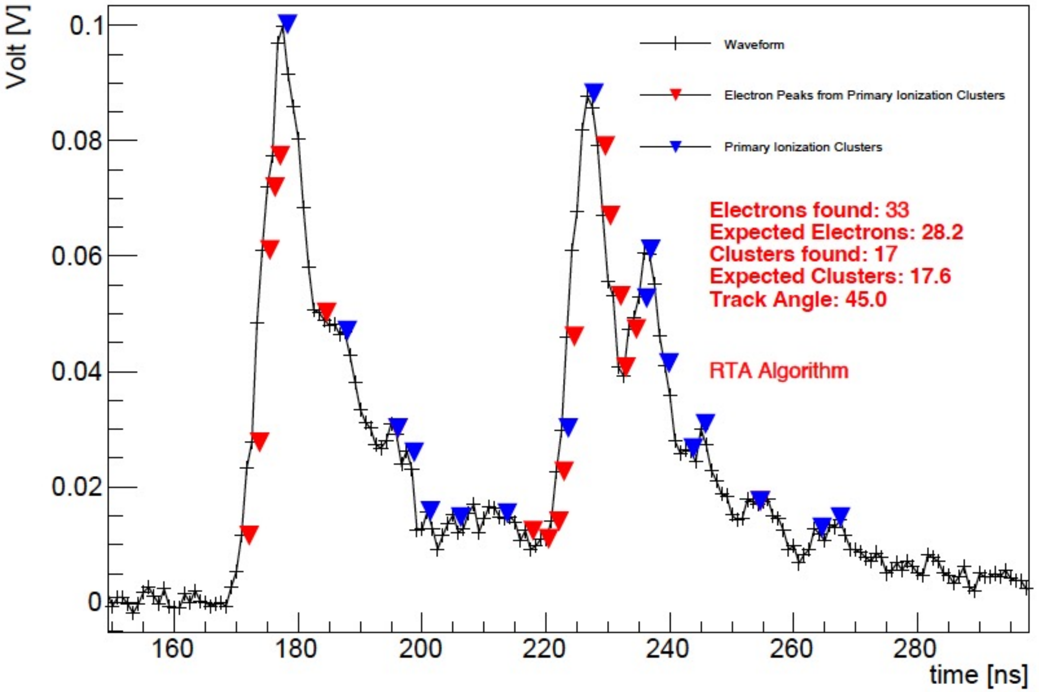



# BACKUP

# DERIV AND RTA PERFORMANCE

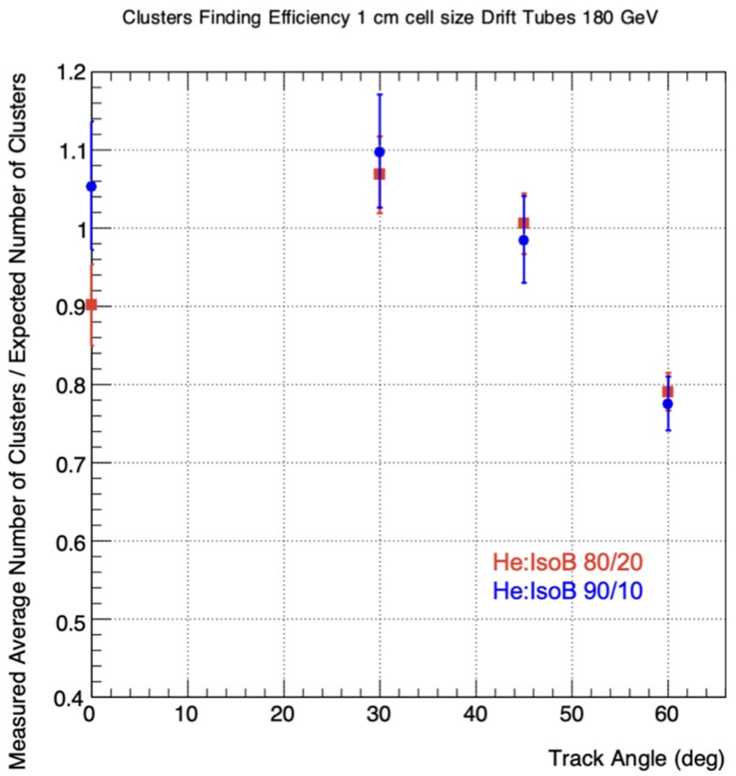


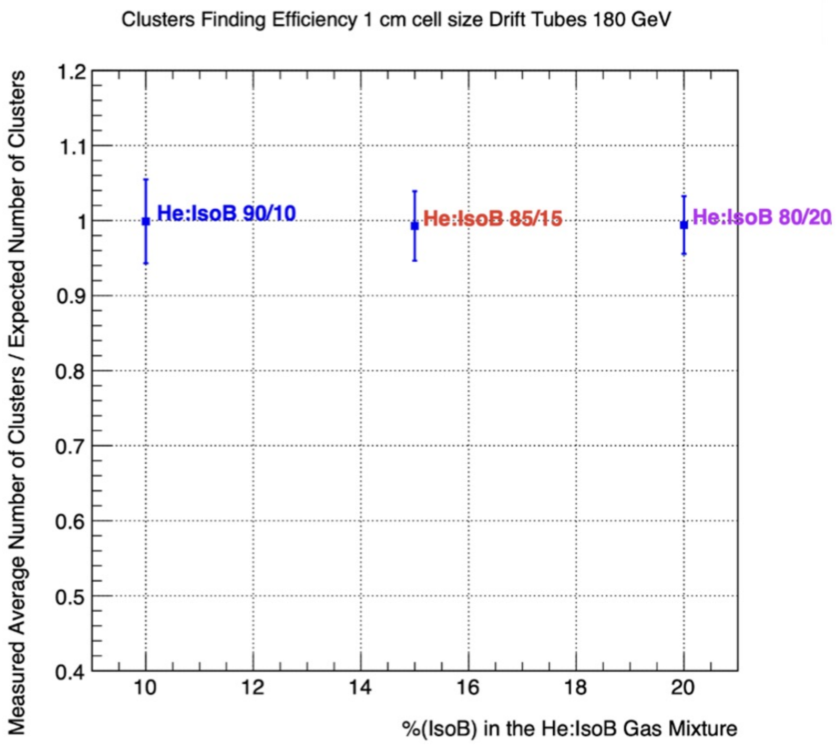



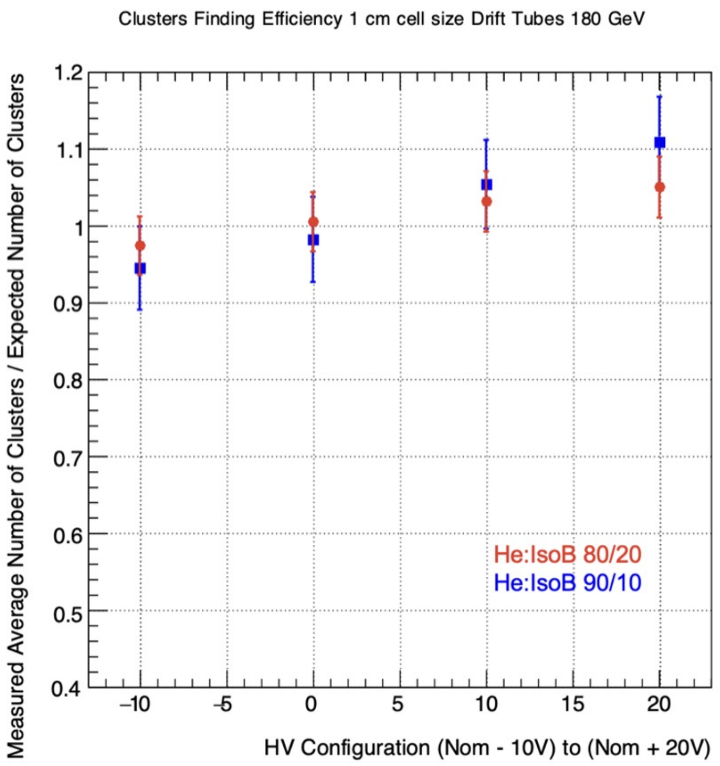




Data waveform recorded for a 1 cm drift cell with a 10 μm sense wire diameter, a 45° track angle, a sampling rate of 2 GSa/s, and a He/iC4H10 90/10 gas mixture for muon beam momentum at 165 GeV/c . The DERIV algorithm (left) and RTA algorithm (right) are used for electron peak identification, with cluster peaks indicated by blue arrows and electron peaks by red arrows. Line shapes are showing behaviours of DERIV parameters.

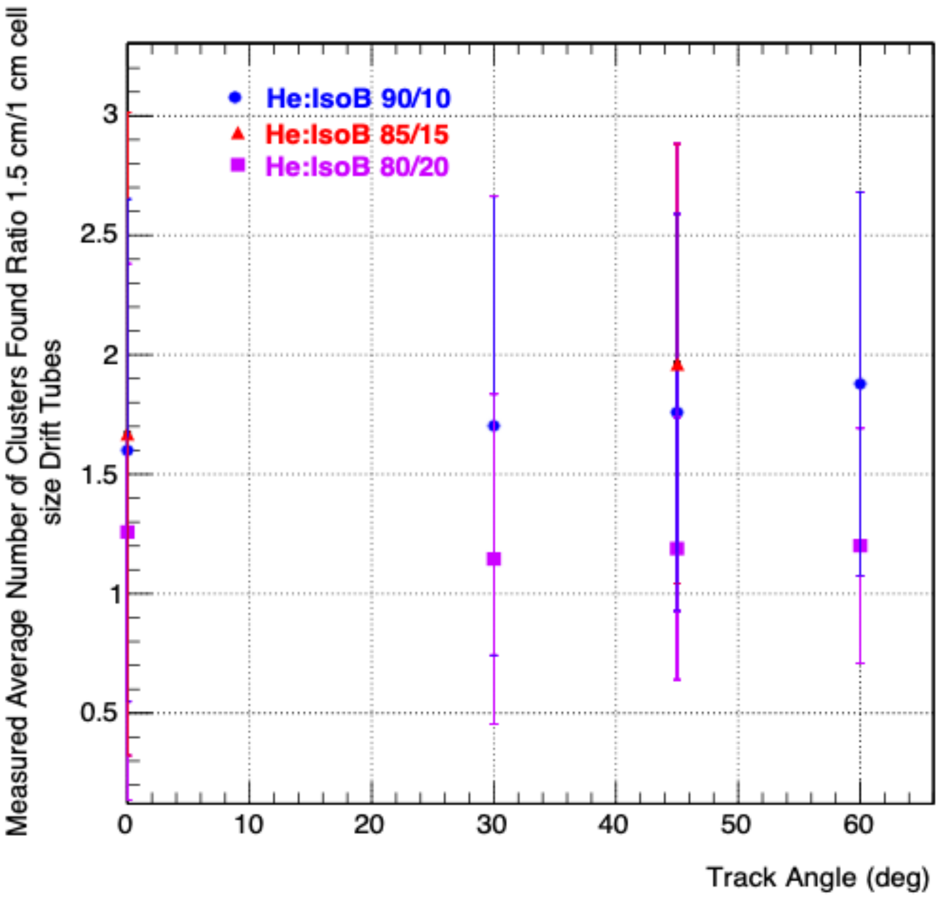
# PERFORMANCE SCANS

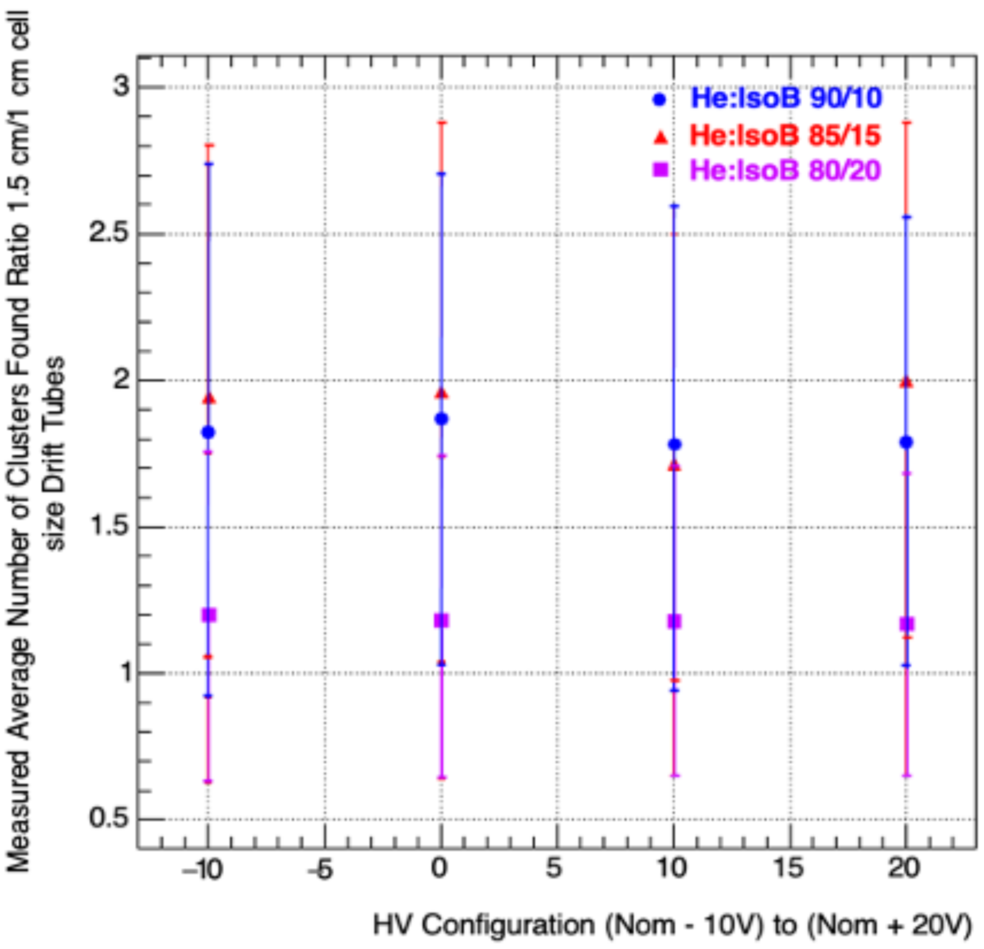


# PERFORMANCE SCANS



# SEPARATION POWER DEFINITION

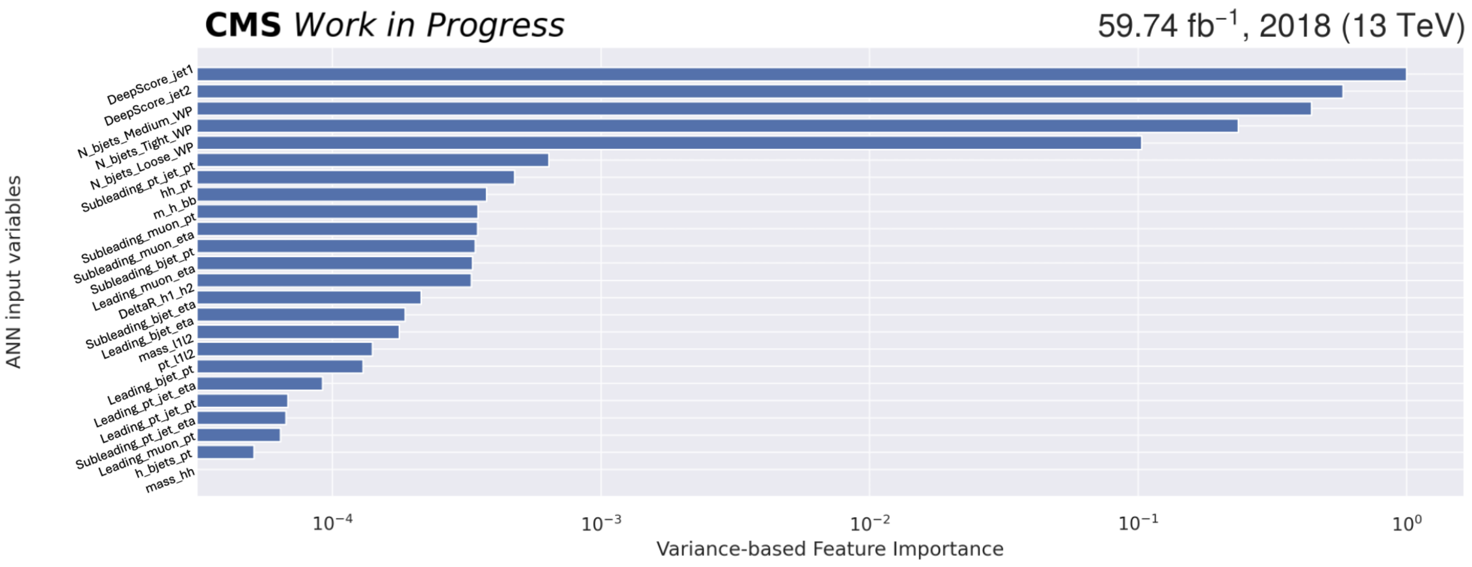
In Helium-based gas mixtures, the signals generated by ionization events in a gas detector can be spread in time on the order of a few nanoseconds. Utilizing fast read-out electronics facilitates efficient identification of these signals. By counting the number of ionization events per unit length (dN/dx), particles can be separated and identified with superior resolution compared to the conventional method based on ionization per unit length (dE/dx). The separation power for two particles, labeled $p_1$ and $p_2$, with different masses but the same momentum, can be evaluated using the following relation:

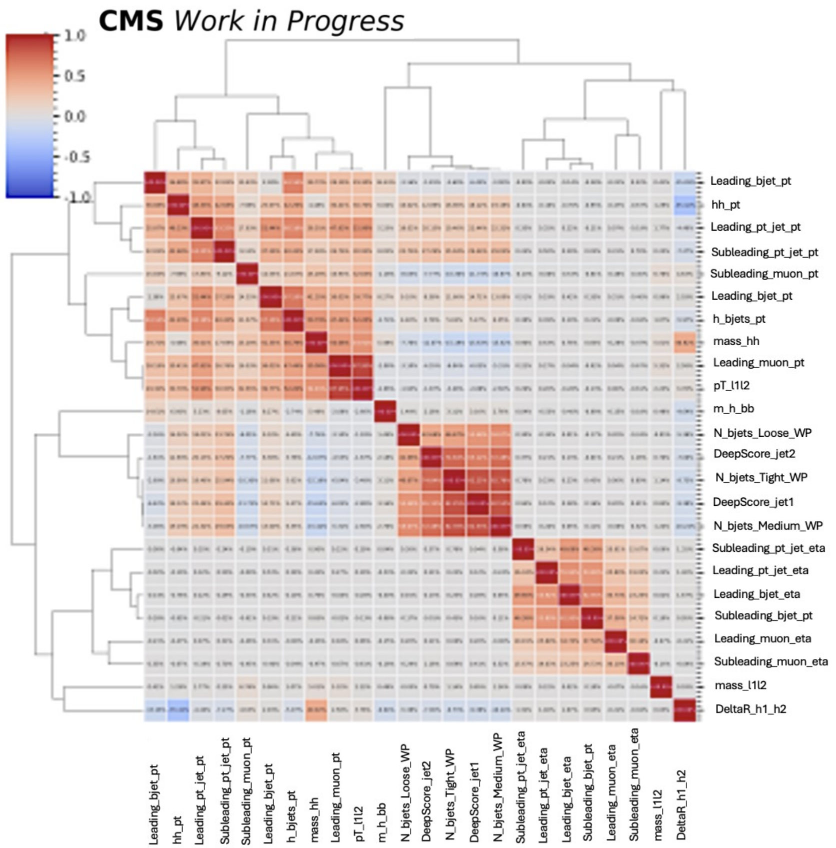
$$n_\sigma^E = \frac{\Delta p_1 - \Delta p_2}{\langle \sigma_{p_1,p_2} \rangle},$$

where $\Delta p_1$ and $\Delta p_2$ are the measured deposited energies, and $\sigma_E$ is the resolution in the ionization measurement (energy resolution), represented by the variance of the Gaussian distribution of the truncated mean values. The term $\langle \sigma_{p_1,p_2} \rangle$ is the average of the two resolutions, given by:

$$\langle \sigma_{p_1,p_2} \rangle = \frac{\sigma_{E,p_1} + \sigma_{E,p_2}}{2}.$$

# FEATURE IMPORTANCE AND CORRELATION MATRIX

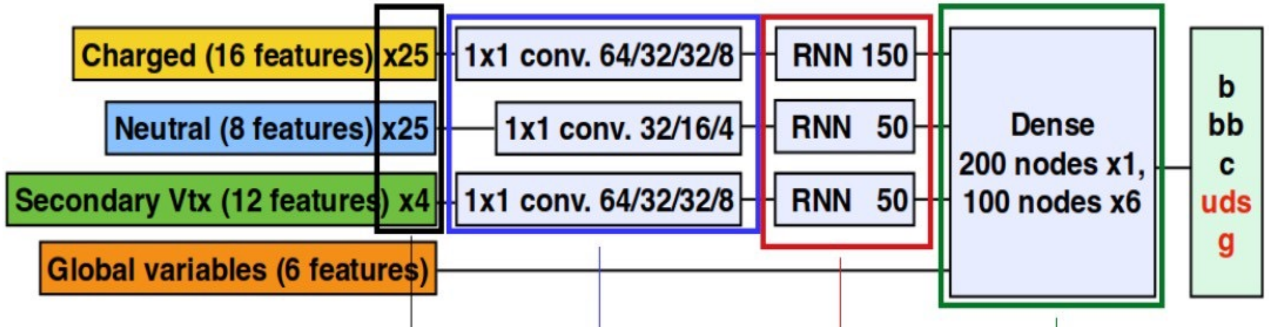


# B-TAGGING ALGORITHMS IN CMS

DeepJet (small R, AK4)

- Low-level information directly in a DNN to tag jets
- Jet as a list of particles lista di particelle (CNN-1D)
- Tagging heavy quarks and separate quark-gluons in one go

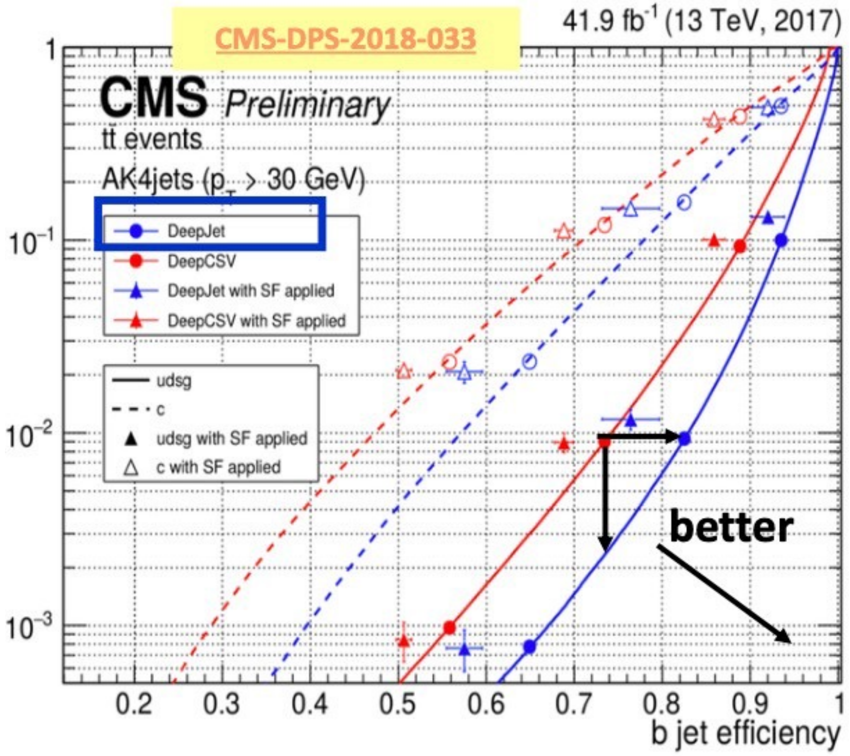



Numer of particles and secondary vertice

Feature extractor convolution made on every particle/SV[1x1]

Summary of information extracted in every set of features

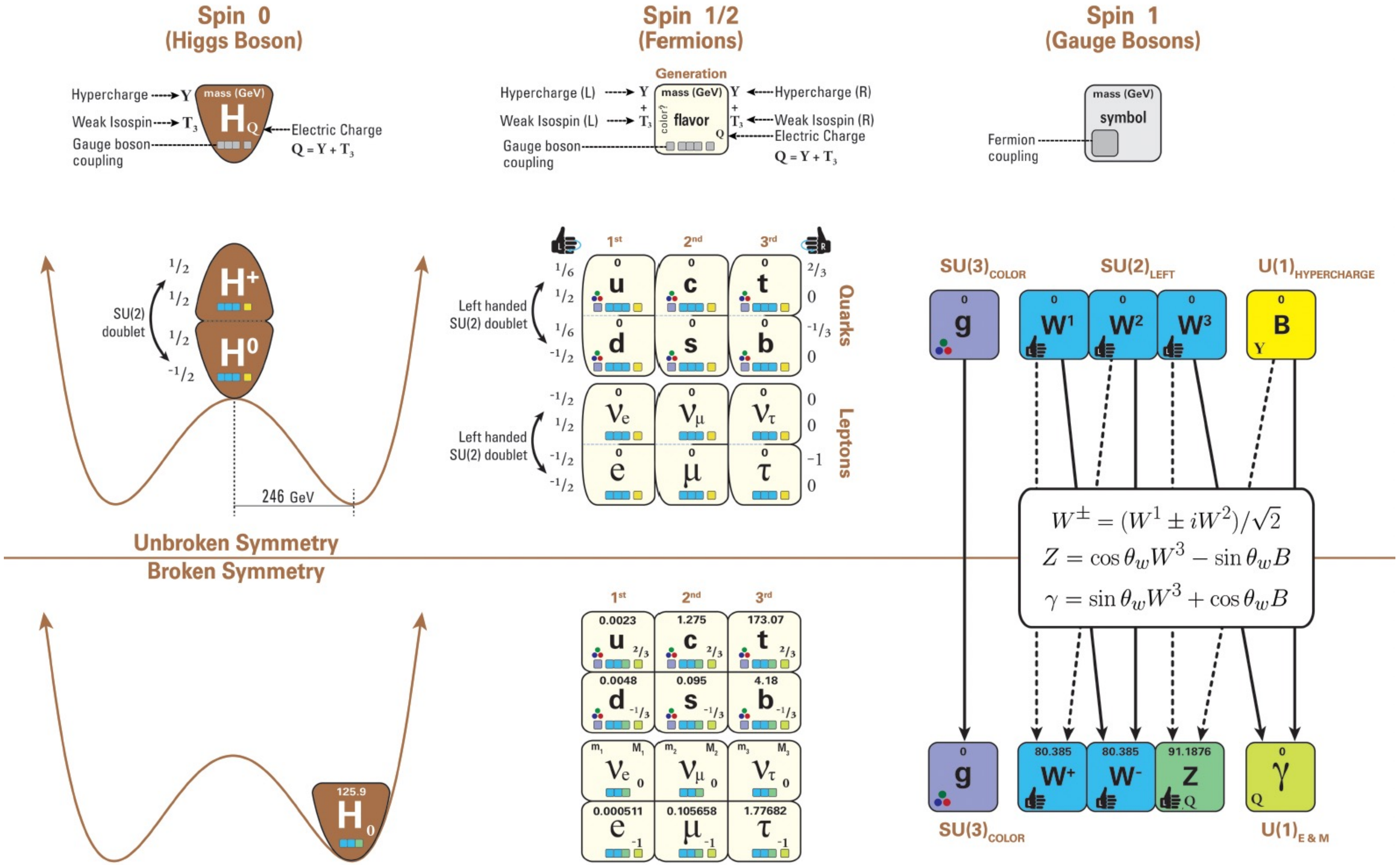
Correlation and classification



# STANDARD MODEL OF PARTICLE PHYSICS

Symmetric high-energy phase

Broken-symmetry low-energy phase



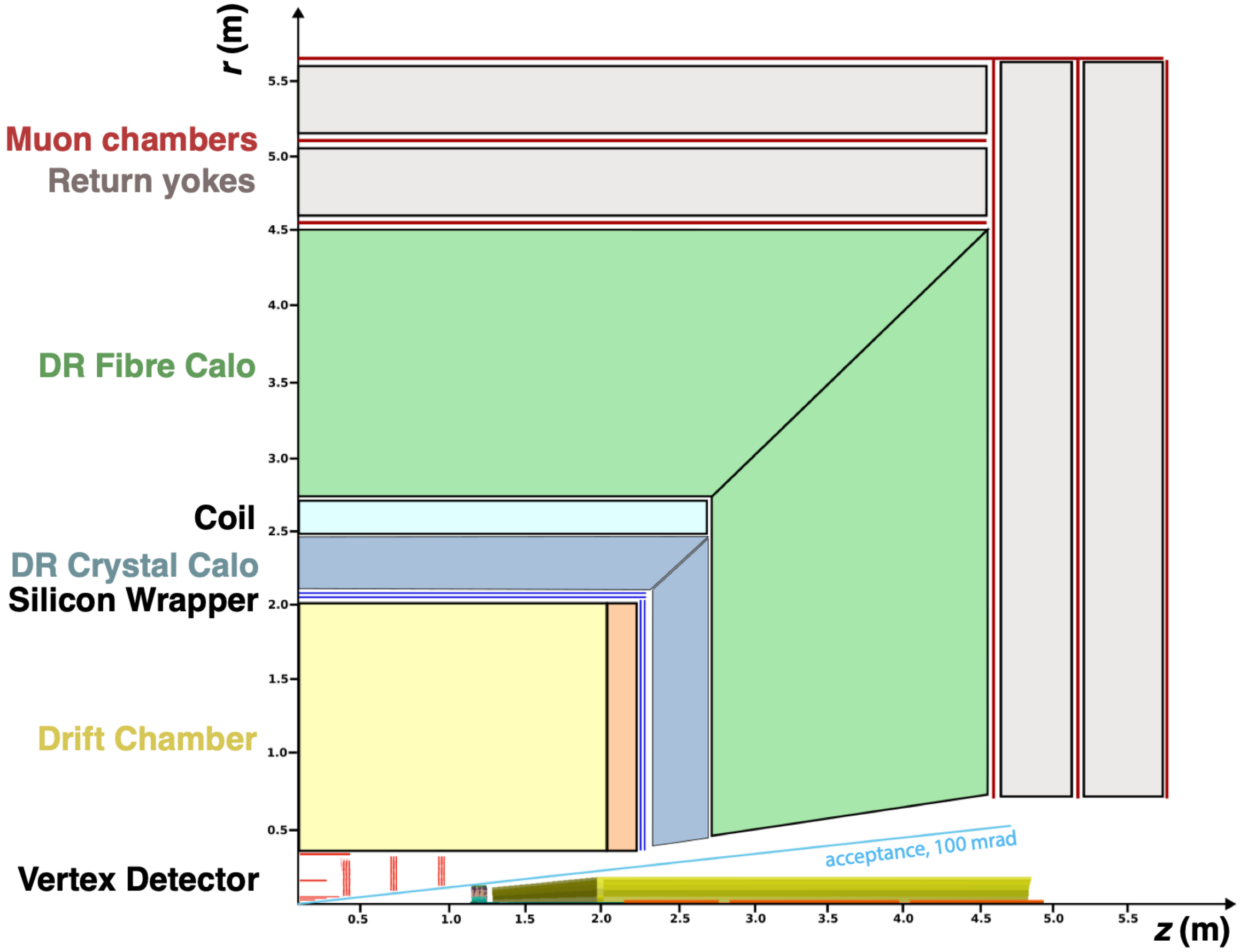
# THE LHC TIMELINE

# IDEA DETECTOR CONCEPT

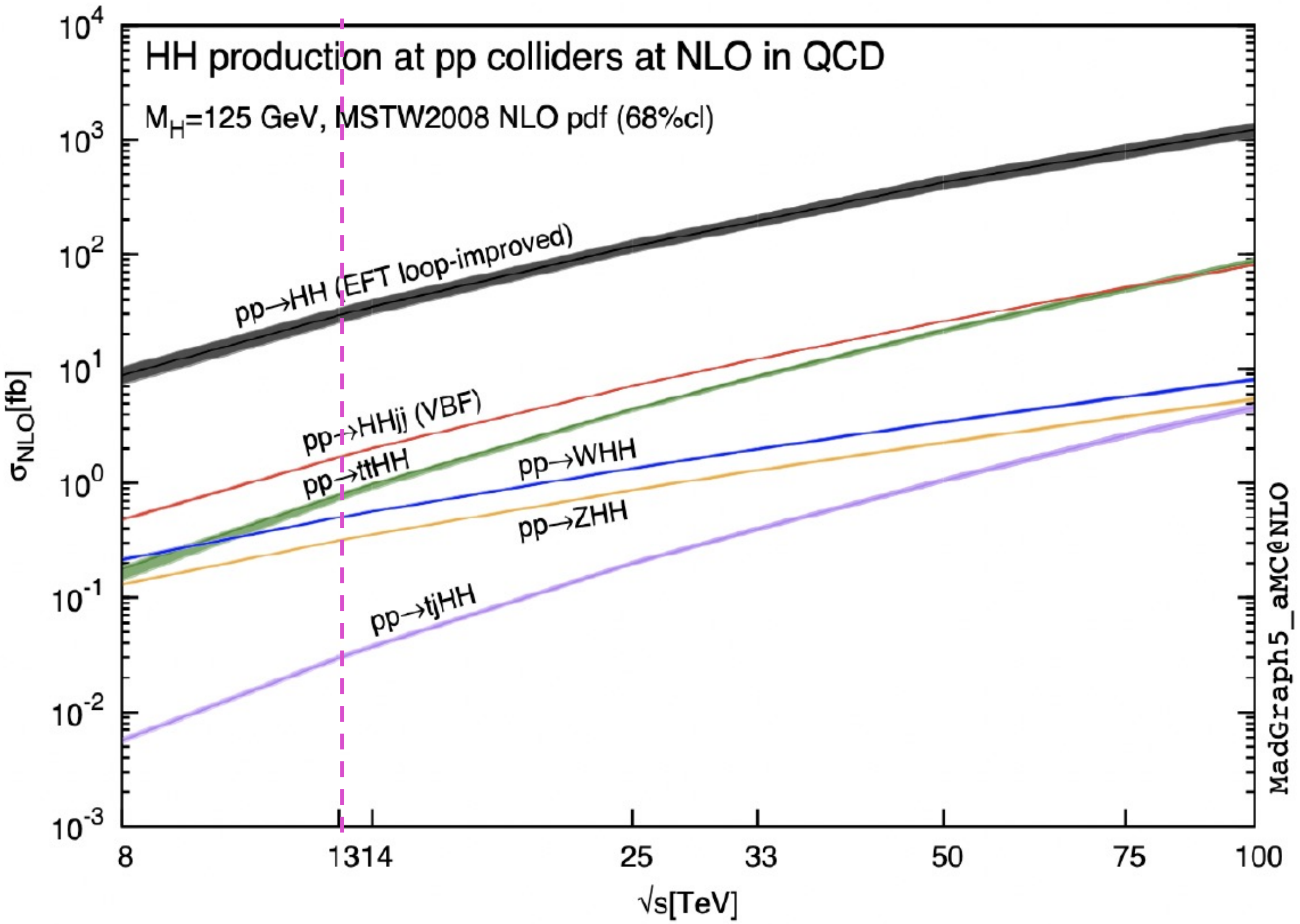
It would operate at four center of mass energies:

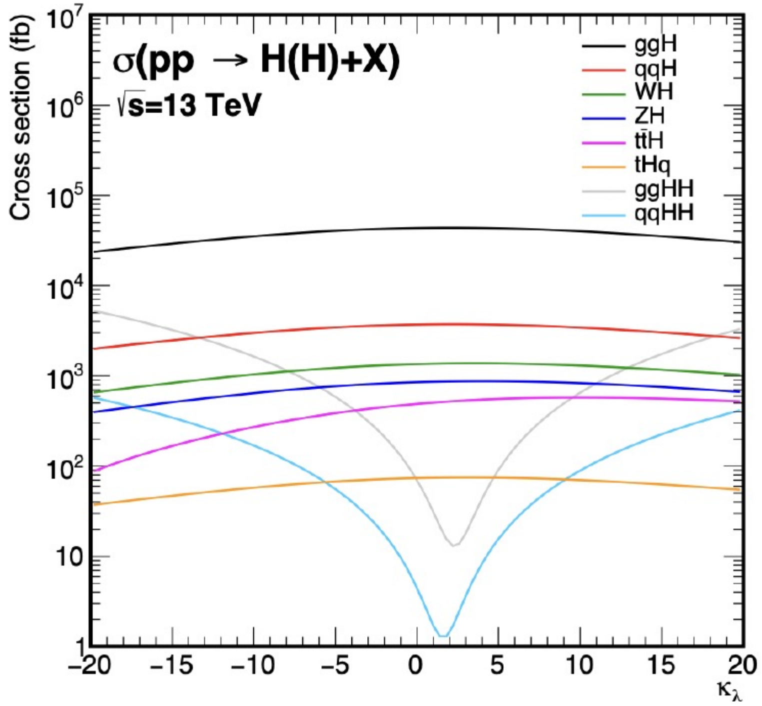
- Z boson mass,
- the WW production threshold,
- the Z-Higgs associated production
- the top quark pair production threshold.

At the Z resonance, the bunch spacing will be of 20 ns at FCC-ee.



# SM AND BSM HIGGS/DI-HIGGS CROSS SECTIONS




| Signal Samples | | | |
|---|---|---|---|
| **HH Sample $\kappa_\lambda$** | **Reaction** | **Generator + PS** | **XS × BR (pb)** |
| $\kappa_\lambda = 1$ | $gg \to HH \to b\bar{b}\ \mu^+\mu^-$ | POWHEG + PYTHIA8 | 7.82e-06 |
| $\kappa_\lambda = 0$ | $gg \to HH \to b\bar{b}\ \mu^+\mu^-$ | POWHEG + PYTHIA8 | 1.513e-05 |
| $\kappa_\lambda = 2.45$ | $gg \to HH \to b\bar{b}\ \mu^+\mu^-$ | POWHEG + PYTHIA8 | 2.841e-06 |
| $\kappa_\lambda = 5$ | $gg \to HH \to b\bar{b}\ \mu^+\mu^-$ | POWHEG + PYTHIA8 | 1.986e-05 |

# CONTROL REGIONS

**Transverse Momentum Discrepancies:**

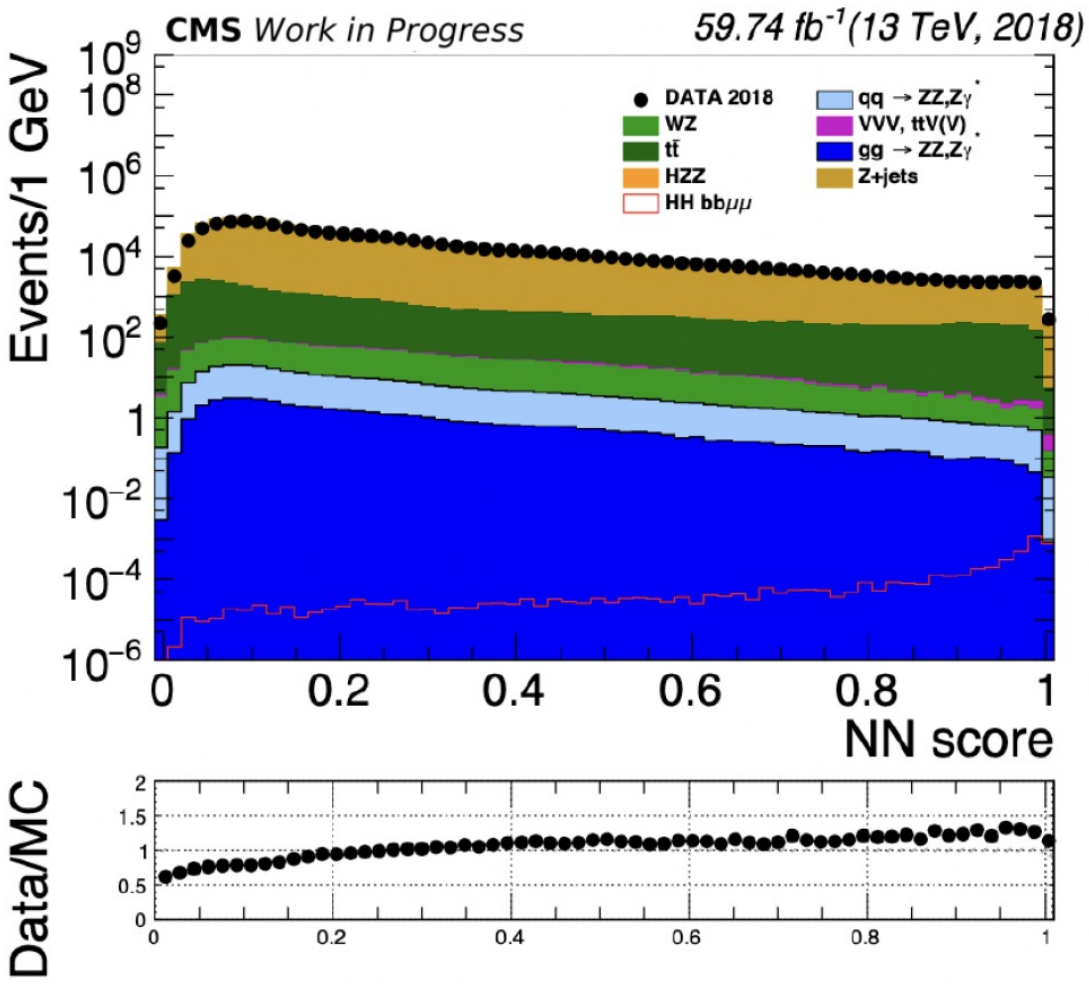
- Observed discrepancies in $p_T$ for leading muons and b-jets, especially at high pT values.
- High $p_T$ jets are overestimated in simulations due to missing higher-order corrections and parton shower modeling.

**b-tagging Discrepancies:**

- Differences between data and MC in b-tagging score distributions, especially at high $p_T$ for b-jets.
- However, these discrepancies are minimized by the ANN training that uses the full set of inputs.

**Impact on the DNN in the Sideband CR:**

- the data and MC distributions exhibit deviations from 50% maximum to 30%-40%, primarily in the extreme regions of the ANN score (near 0 and 1). - systematic uncertainties in background modeling contribute; the extreme score regions correspond to highly signal- or background-like events, making them more sensitive to imperfections in the modeling of input variables.




The examination of individual input feature distributions suggests that mismodeling effects in variables such as jet and lepton transverse momentum spectra and b-tagging discriminators propagate to the ANN output, amplifying the observed differences in these regions.

# UNCERTAINTIES

**1. Statistical Uncertainties:**

1. Depend on data size (e.g., Poisson fluctuations).
2. Can be reduced with more data.

**2. Systematic Uncertainties:**

1. Theoretical Uncertainties: Associated with model assumptions and predictions.
2. Experimental Uncertainties: Result from detector performance and measurement precision.

**Systematic Uncertainty Breakdown**

**•Theoretical:**

- Uncertainties in QCD scale, PDF, $\alpha_S$, and Higgs branching ratios.
- Affects both signal and background models.

**•Experimental:**

- Luminosity: 2.3%-2.6% across years.
- Pileup Re-weighting: Shape uncertainties for MC contributions.
- Lepton and Jet Identification: Shape uncertainties for muons, electrons, and jets.
- Jet Energy Scale/Resolution (JES/JER): Shape uncertainties affecting event reconstruction.
- B-tagging: Shape uncertainties for b-tagging efficiencies.

**Impact on Results**

**•Normalization vs. Shape Variations:**

- **Normalization:** Affects overall event yield, treated as log-normal distribution.
- **Shape:** Alters event kinematics, affecting distributions non-uniformly.

**•ANN Discriminant Shape Uncertainties:**

- Modifies NN output and leads to event migration between categories.
- Shifts in binned analysis based on input uncertainties (e.g., JES, JER, b-tagging).

# SR CUTFLOW

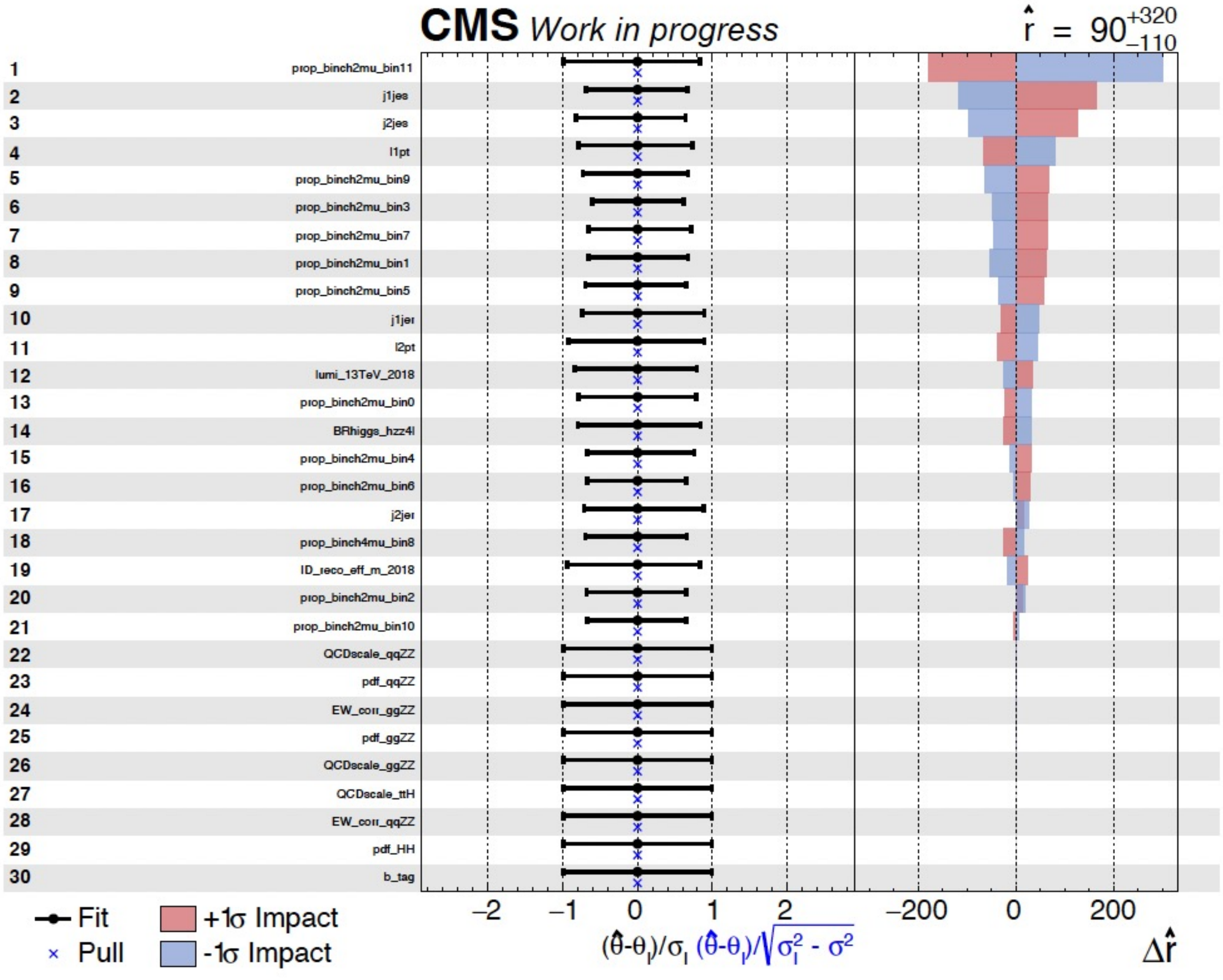
| Process | Yields 2016 | Yields (2017) | Yields(2018) |
|---|---|---|---|
| **Drell-Yan (DY)** | $11700 \pm 220$ | $13000 \pm 250$ | $18000 \pm 300$ |
| **Single Higgs production via ggF (ggH)** | $0.002 \pm 0.001$ | $0.002 \pm 0.001$ | $0.002 \pm 0.001$ |
| **Top quark pair + vector boson (TTV)** | $4.5 \pm 0.1$ | $5.5 \pm 0.1$ | $4.5 \pm 0.1$ |
| **Top quark pair production (TT)** | $8400 \pm 150$ | $9900 \pm 180$ | $15400 \pm 250$ |
| **Triboson (VVV)** | $3.0 \pm 0.4$ | $3.5 \pm 0.4$ | $3.0 \pm 0.4$ |
| **Diboson (VV)** | $25.0 \pm 1.0$ | $30.0 \pm 1.0$ | $25.0 \pm 1.0$ |
| **Single Higgs with vector boson (VH)** | $0.007 \pm 0.001$ | $0.008 \pm 0.001$ | $0.007 \pm 0.001$ |
| **Others** | $0.0003 \pm 0.0002$ | $0.0004 \pm 0.0002$ | $0.0003 \pm 0.0002$ |
| $\sum$ **backgrounds** | $20133 \pm 280$ | $22939 \pm 340$ | $33433 \pm 420$ |
| **Signal** (HH, $m_H = 125GeV$) | $0.160 \pm 0.005$ | $0.186 \pm 0.006$ | $0.146 \pm 0.005$ |
| **Total expected** | $20133 \pm 280$ | $22939 \pm 340$ | $33433 \pm 420$ |
| **Observed** | 20309 | 22984 | 33839 |

**Table 8.4:** Cut flow table after signal region selection with expected and data event yields for different processes in 2016, 2017, and 2018. The expected background yield is computed as the sum of all background contributions. The vector boson V can be both a Z boson and W boson.

# SYSTEMATIC UNCERTAINTIES IMPACTS

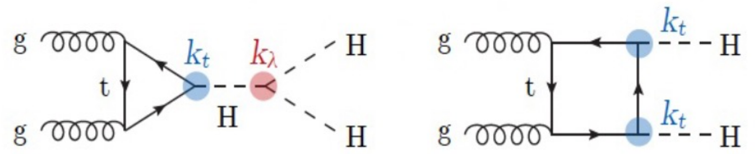
- The impact of a nuisance parameter θ is defined as the shift Δr induced on the value of the signal strength r by fixing θ to its best fit value and brought to its 1σ best-fit values, with all other parameters profiled as normal values and are allowed to vary freely to find the minimum of a profiled likelihood for the given fixed value of θ.

- The most important source of systematic uncertainty that affects the measurements is the systematics on the **statistical error for the last bins of the DNN** (called improperly prop_binch2mu_bin* ), followed by the jet energy scale for the leading and subleading jets (j1jes, j2jes) and the transverse momentum of leading and subleading muons (l1pt, l2pt).




Full Run 2 Impacts

# THEORETICAL EXPECTATION CURVE




$$A = k_t\, k_\lambda\, T + k_t^2\, B$$

$$\sigma(k_t, k_\lambda) \sim |A|^2 = k_t^2\, k_\lambda^2\, |T|^2 + k_t^4\, |B|^2 + k_t^3\, k_\lambda\, |T^*B + B^*T|$$

$$= \underbrace{k_t^2\, k_\lambda^2\, t}_{\textit{triangle}} + \underbrace{k_t^4\, b}_{\textit{box}} + \underbrace{k_t^3\, k_\lambda\, i}_{\textit{interference}}$$

- In symbols
  $\sigma(\kappa_\lambda, \kappa_t) = \mathbf{c}\,(\kappa_\lambda, \kappa_t) \cdot \mathbf{v}$ , where
  $\mathbf{c} = (\kappa_\lambda^2\, \kappa_t^2, \kappa_t^4, \kappa_\lambda\, \kappa_t^3)$ is the vector of the couplings and
  $\mathbf{v} = (t, b, i)$ is the vector of the components
- We have the full simulation of 4 NLO samples : $\kappa_\lambda = 0, 1, 2.45, 5$, $\kappa_\lambda = 1$
  Choosing 3 of them, we know their $\sigma$ from the generator and can compute their $\mathbf{c_i}$ (example : for $\kappa_\lambda = 5$, $\sigma_{NLO} = 79.03$ fb and $\mathbf{c} = (25, 1, 5)$ )
  So this relation holds:

$$\begin{pmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\end{pmatrix} = \begin{pmatrix} c_1^1 & c_1^2 & c_1^3\\ c_2^1 & c_2^2 & c_2^3\\ c_3^1 & c_3^2 & c_3^3\end{pmatrix}\begin{pmatrix}t\\ b\\ i\end{pmatrix} \qquad \text{or, in condensed form,} \qquad \boldsymbol{\sigma} = \mathbf{C}\mathbf{v}$$

- Therefore, by inverting the matrix

$$\mathbf{v} = \mathbf{C}^{-1}\boldsymbol{\sigma} \quad \Longrightarrow \quad \boxed{\sigma(\kappa_\lambda, \kappa_t) = \mathbf{c}^T(\kappa_\lambda, \kappa_t)\mathbf{C}^{-1}\boldsymbol{\sigma}}$$

Total cross section, or shape ($d\sigma/dx$)

Couplings to model

Known from the couplings of the input samples

3 input samples that are scaled and summed