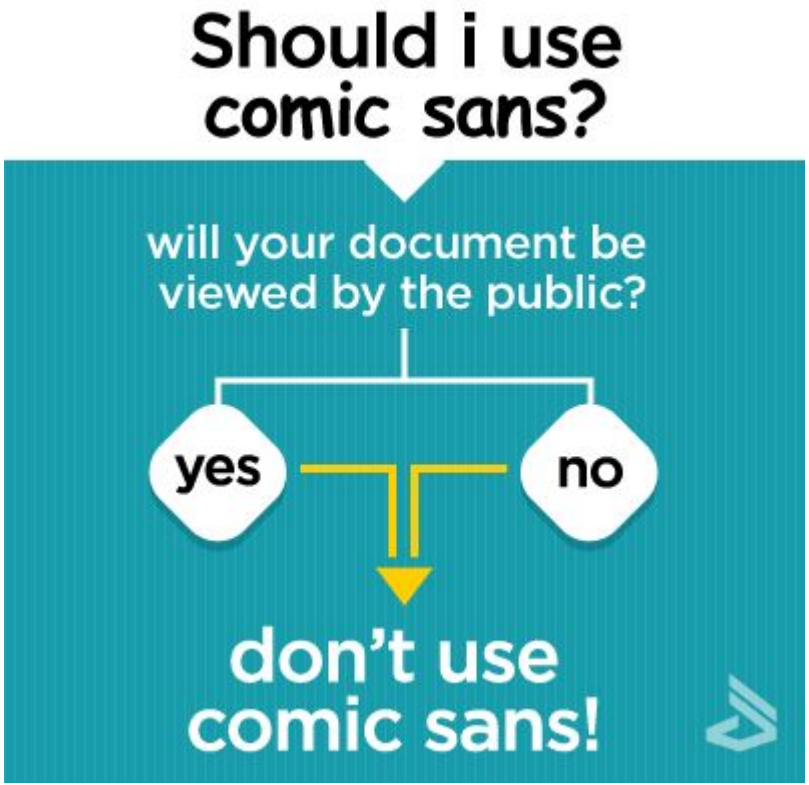
H → bb measurement exploiting data scouting during run 3 at CMS

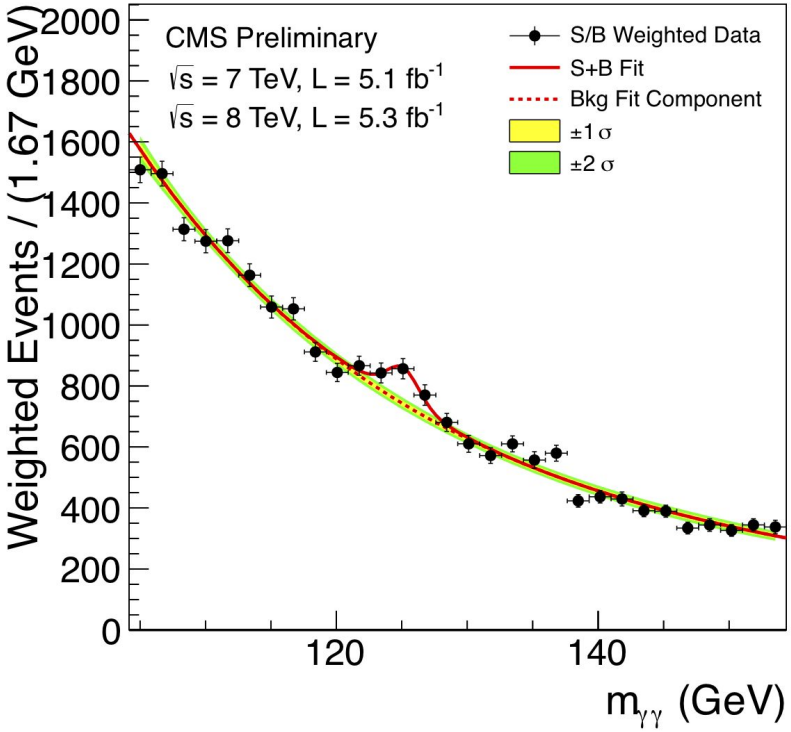
Patin Inkaew
Helsinki Institute of Physics
SMARTHEP Yearly Meeting (01.12.2023)
Lund University, Sweden

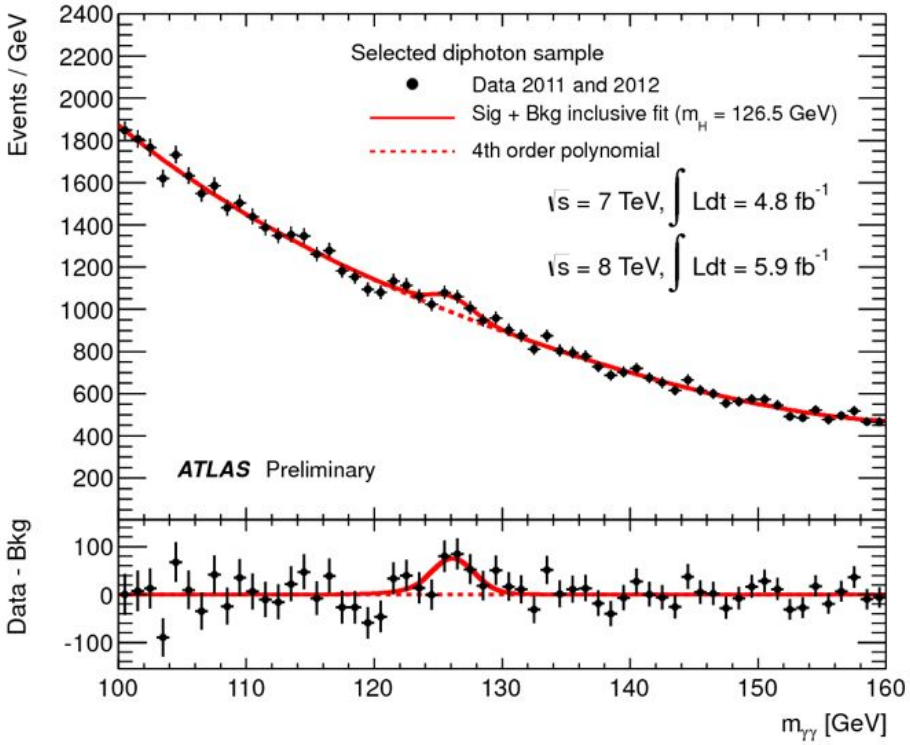
4 July 2012…

The Higgs announcement as open science



10 Hilarious Comic Sans Meme to Light Up Your Day





[Latest update in the search for the Higgs boson]

Real-time analysis for Higgs measurement!

The next steps …

ATLAS plans to submit a paper based on the data presented today at the end of July, at the same time as CMS and to the same journal

$H \to WW^{(*)} \to l\nu l\nu$ channel: plan is to include results in the July paper
$H \to \tau\tau$, $W/ZH \to W/Z\, b\bar{b}$: first results with 2012 data expected later in the Summer

MORE DATA will be essential to:
- ❑ Establish the observation in more channels, look at more exclusive topologies
- ❑ start to understand the nature and properties of the new particle

This is just the BEGINNING !

We are entering the era of "Higgs" measurements
First question: is the observed excess due to the production of a SM Higgs boson ?

Note:
- ❑ we have only recorded ~ 1/3 of the data expected in 2012
- ❑ the LHC and experiments have already accomplished a lot and much faster than expected

ATLAS: Status of SM Higgs searches, 4/7/2012 49

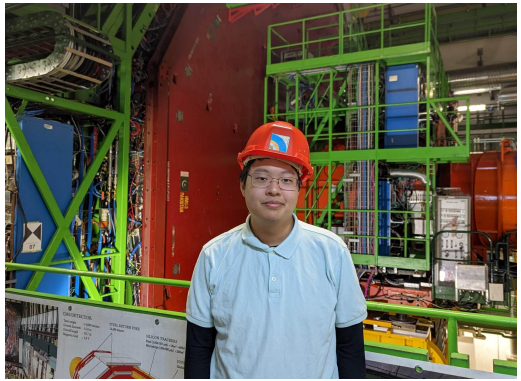
Latest update in the search for the Higgs boson

Outline

- Introduction
- Analysis: H → bb measurement exploring data scouting during run 3 at CMS
- Analysis: Jet Energy Correction (JEC) studies on scouting jets
- Other activities
- Conclusion

About Me

Name: Patin Inkaew (PI ~ 3.14)
Nickname: Earth
Birthday: 22 July 1998 (22/7 ~ 3.14)
Hometown: Bangkok, Thailand
Institution: University of Helsinki (UH), Helsinki Institute of Physics (HIP)
Contract start: 01/10/2022



Education

Stanford University, CA, USA (Thai Government Scholarship)
Coterminal program (Joint BS+MS) in 4 years
BS: Physics, Minor: Mathematics, East Asian Studies (Japan subplan)
MS: Computer science (AI track)
Research: Many things: laser, detector design, ML, CV, CG, ComBio
with a little experience in particle physics analysis

ESR1:

Machine learning and Real-Time Analysis for Higgs boson measurements and fleet safety

PhD:
University of Helsinki (UH) & Helsinki Institute of Physics (HIP), Finland

Supervisors:
Mikko Voutilainen,
Henning Kirschenmann

Secondment:
CERN, Switzerland (2025)

Collaborator:
Maurizio Pierini

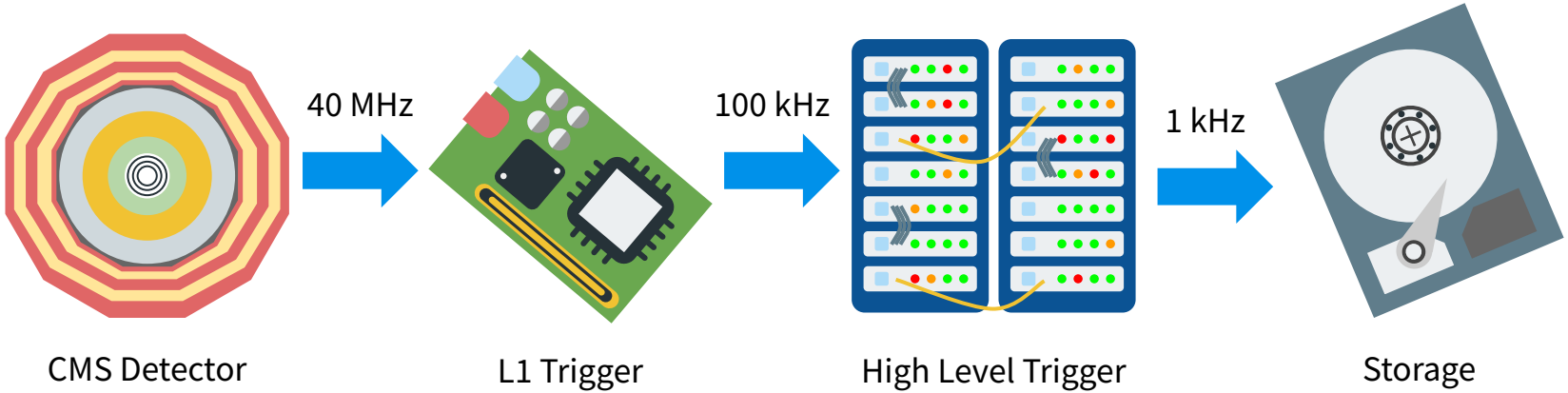
Secondment:
Verizon Connect, Italy (September 2024)

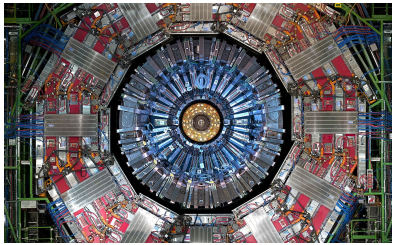
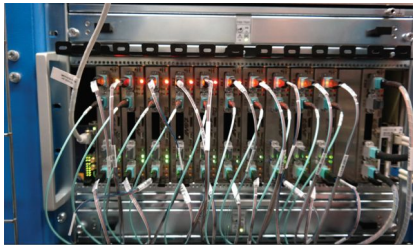
Collaborators:
Leonardo Taccari,
Francesco Sambo

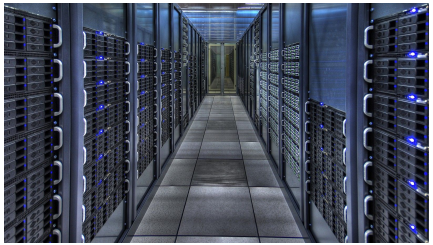
Introduction

CMS Trigger System → Data Scouting

CMS Trigger System



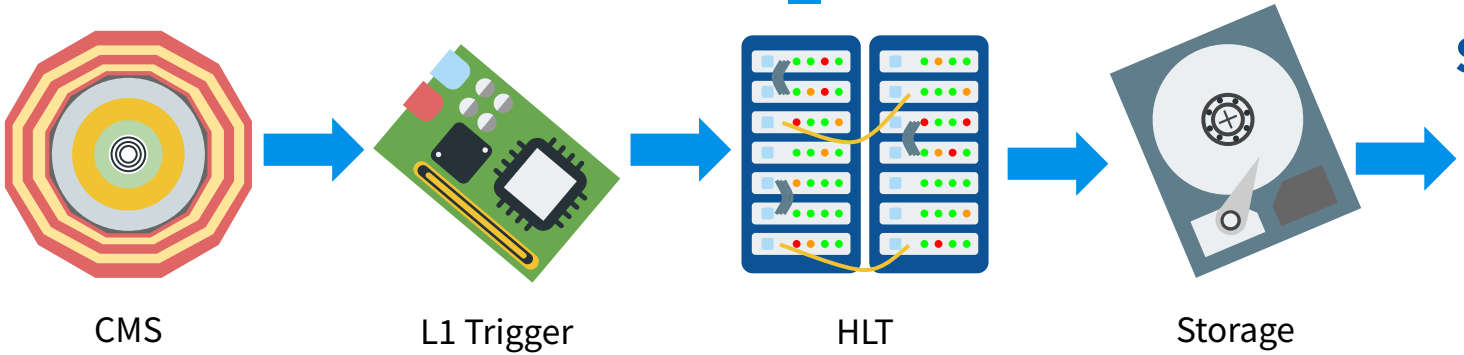


Data Scouting (HLT Scouting)

The real bottleneck is data recording rate (MB/sec), not event rate (event/sec)

Scouting Stream

- reduced data format
- event size reduced by 100x
- event rate increased by 30x
- based on reconstruction at HLT (no offline reconstruction)



CMS → L1 Trigger → HLT → Storage

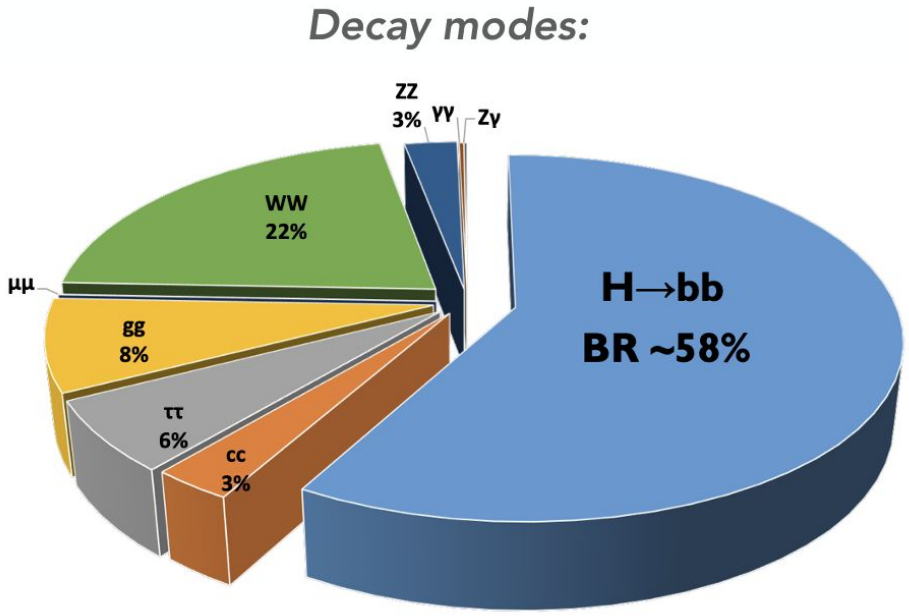
Standard Stream

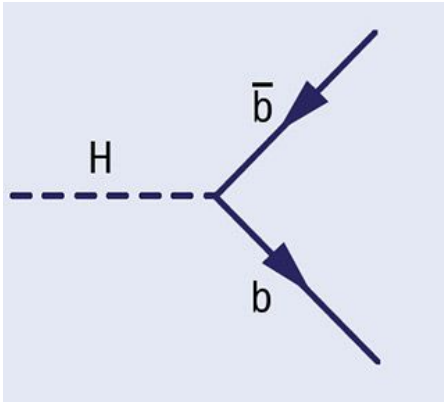
- ~ 1 kHz
- ~ 1.5 MB/event
- full event
- reconstructed at Tier0 within 48 hours

H → bb measurement exploring data scouting during run 3 at CMS

Motivation → Current progress (Trigger studies)

Motivation: Higgs decay modes





- H → bb is the most probable decay mode
- However, suffer from enormous background (QCD)

ATL-PHYS-SLIDE-2022-013

Motivation: Higgs production modes

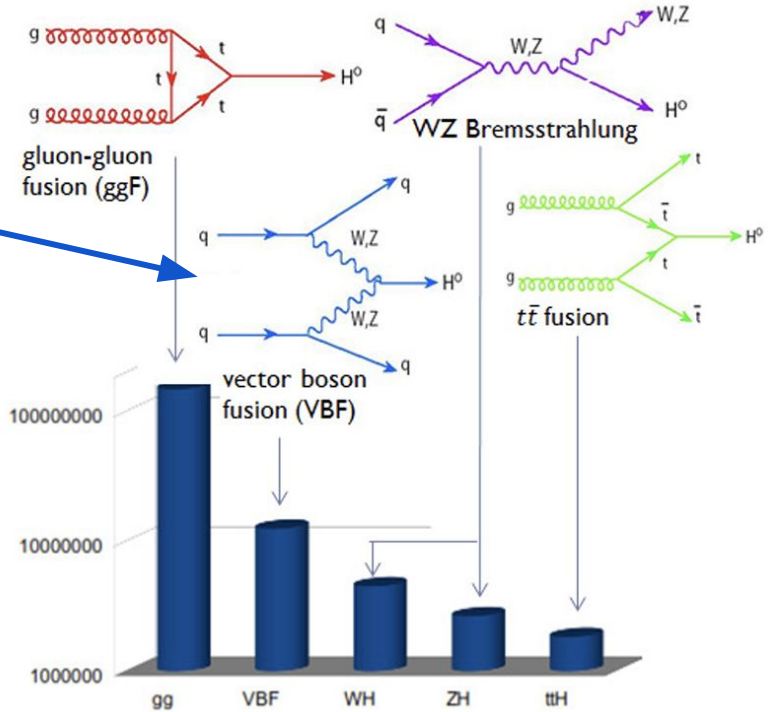
Gluon Fusion (ggF) and

Vector Boson Fusion (VBF)

→ most probable at LHC

ggF/VBF H → bb is full hadronic search

→ challenging to observe due to large QCD background



ATL-PHYS-PROC-2014-205

Motivation: latest result

Search for boosted Higgs bosons produced via vector boson fusion in the $H \to b\bar{b}$ decay mode using LHC proton-proton collision data at $\sqrt{s} = 13$ TeV

The CMS Collaboration

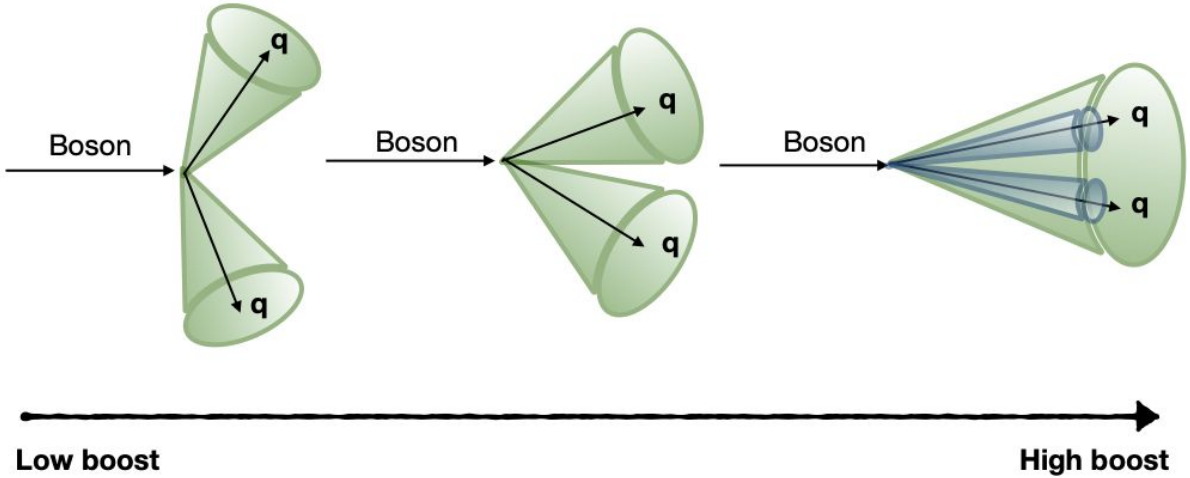
pT > 450 GeV?

Abstract

A search is conducted for Higgs bosons produced with high transverse momentum ($p_T > 450$ GeV) via vector boson fusion at the LHC proton-proton collider operating at center of mass energy $\sqrt{s} = 13$ TeV. The result is based on the 138 fb^{-1} data set

CMS-PAS-HIG-21-020

Motivation: boosted jets



With high boost, two jets originating from single boson merge into single large jet.

Probing jet substructure can improve signal sensitivity from QCD background.

CMS-PHO-EVENTS-2022-018

Motivation: latest result

Search for boosted Higgs bosons produced via vector boson fusion in the $H \to b\bar{b}$ decay mode using LHC proton-proton collision data at $\sqrt{s} = 13$ TeV

The CMS Collaboration

pT > 450 GeV?
Can we lower that?

Abstract

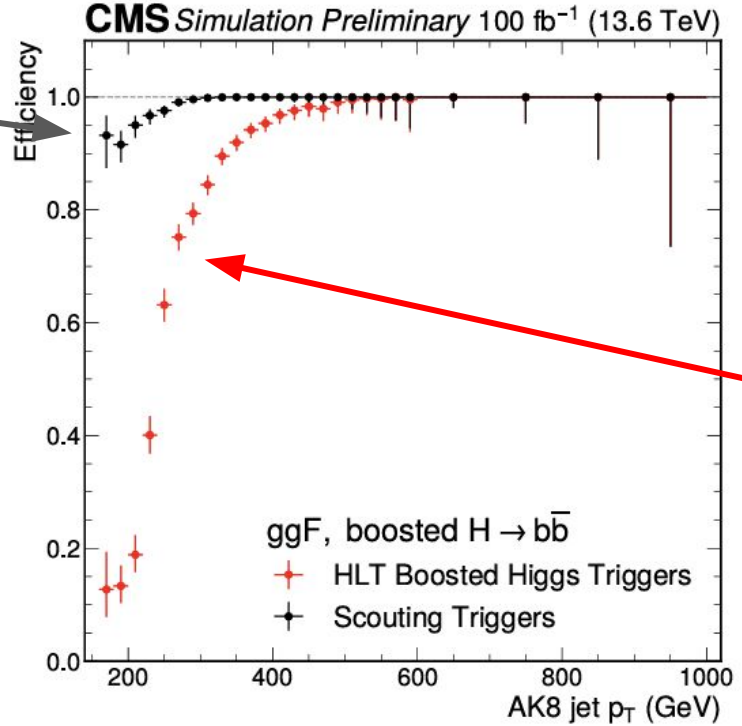
A search is conducted for Higgs bosons produced with high transverse momentum ($p_T > 450$ GeV) via vector boson fusion at the LHC proton-proton collider operating at center of mass energy $\sqrt{s} = 13$ TeV. The result is based on the 138 fb^{-1} data set

CMS-PAS-HIG-21-020

Trigger efficiency studies of the CMS Run-3 Data Scouting

Scouting Triggers (Run 3)
mostly pass through for L1 decisions

Faster turn-on
→ can lower pT requirement



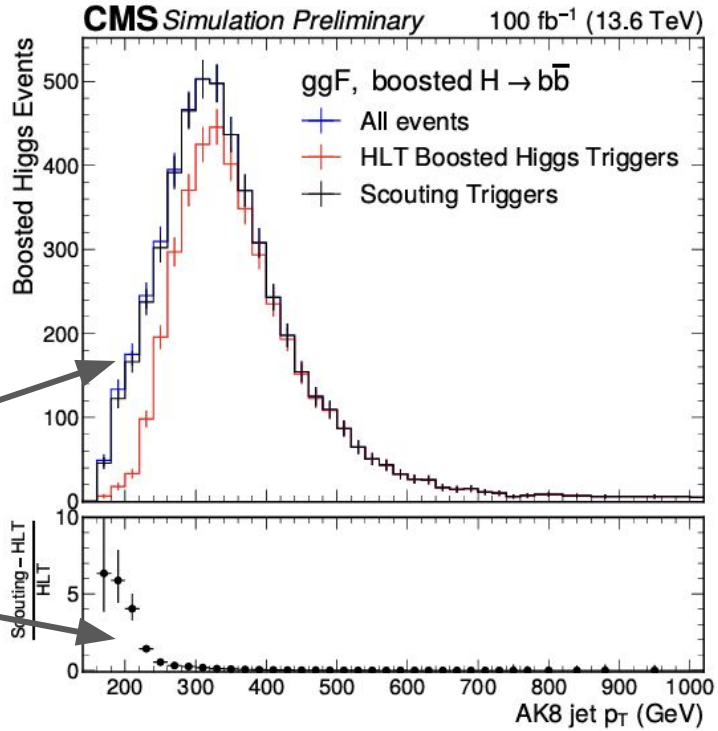
Preliminary study on simulation (ggF, VBF)

HLT triggers targeting boosted H→bb
based on ParticleNet

arxiv.org/abs/1902.08570

CMS-DP-2023-076

Trigger efficiency studies of the CMS Run-3 Data Scouting



Overall number of events gain ~20%

See particularly gains in low pT region

CMS-DP-2023-076

Trigger efficiency studies of the CMS Run-3 Data Scouting

Note: in this study:

- HLT triggers efficiency include tagger performance (ParticleNet)
- However, scouting triggers do no include tagger performance
- Differences in reconstruction → different tagger performance
- Training ParticleNet for scouting jet is in progress

If we now have a tagger, we would need to know kinematics of (b-)jets

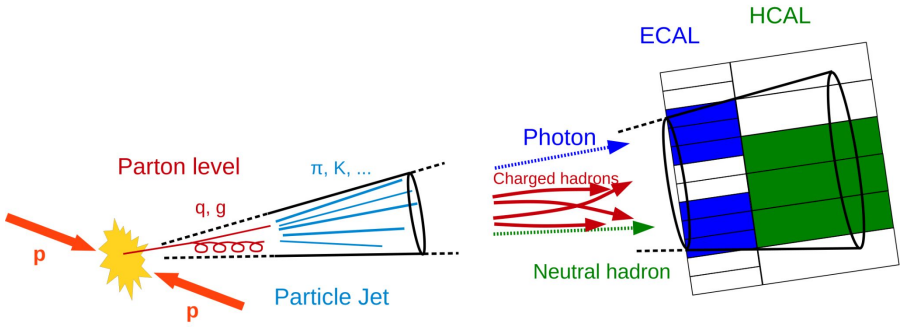
Jet Energy Correction (JEC) studies on scouting jets

Introduction: JEC → Current progress

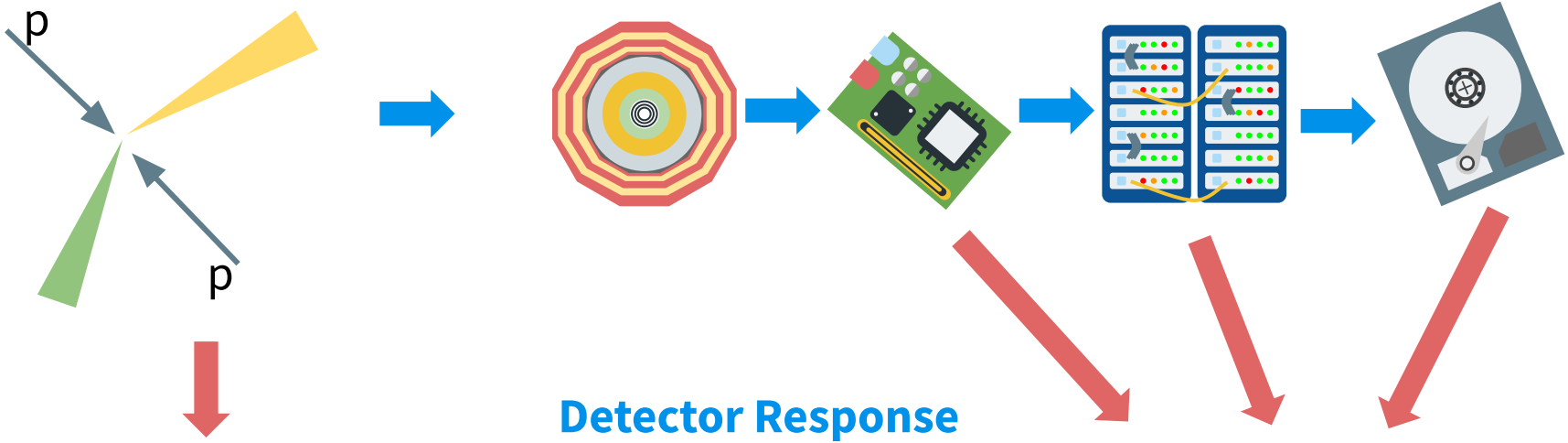
Introduction: What is a jet?

- Jets are the signatures of quarks and gluons (?)
  - color confinement
  - hadronization: quarks and gluons quickly become a cone of particles **“Jet”**
  - **Jet clustering:** group particles to a jet
  - In a more generalized picture, almost everything becomes a jet: gluons, quarks, top quarks, W/Z boson, Higgs boson

Lots of interesting stuff!



Introduction: JEC



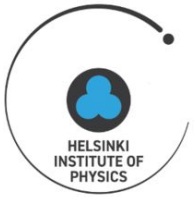
Particle-level / Truth / Generator Jets (Gen Jets)

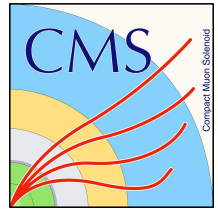
Detector Response

Reconstructed Jets (Reco Jets)

Jet Energy Correction (JEC) / Jet Calibration

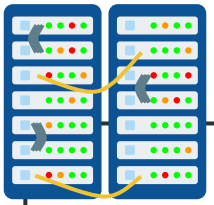
At CMS, the online reconstruction is “almost like” the offline, largest difference is simplified tracking



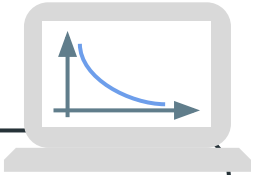


How can the offline and HLT reconstruction support each other?

JEC for scouting jets



transfer low systematic uncertainties from offline to online reconstruction



Scouting/HLT Jet

+ more statistics (exposed to full incoming data streaming before triggering)
- uses simpler reconstruction due to speed constraint

Offline jets

+ more sophisticated reconstruction
- contain less statistics (constructed from stored data after HLT)

improve offline calibration with abundant HLT jet statistics

Other activities

Other activities

Past

- PAPU Fall Seminar (22 November 2022): **lightning talk!**
- CMS Week December 2022 (5 - 9 December 2022)
- Spåtind 2023: Nordic Conference on Particle Physics (3-8 January 2023): **talk!**
- JetMET Workshop (15 - 17 May 2023)
- Stay at CERN (1 June - 20 August 2023): **shifts + summer project supervision!**
- CMS Data Analysis School (5 - 10 June 2023)
- CMS Week June 2023 (12 - 16 June 2023)
- 13th Patatrack Hackathon (26 - 30 June 2023)
- Advanced Artificial Intelligence for Precision High Energy Physics (16 - 28 July 2023)
- CERN School of Computing (20 August - 2 September 2023): **lightning talk!**
- Researcher Night (29 September 2023): **outreach!**
- Particle Physics Day (12 October 2023)
- ML4Jets (4 - 6 November 2023)

Other activities

CMS Young Scientist Committee (CMS YSC)

- CMS Welcome message translation (→ Thai)
- LHC Mentorship program
- LHC Job Matching Event
- Regional Representative (new initiative!)

Future

- ML@L1 Workshop (11 - 15 December 2023)
- Physics Day (4-6 March 2024): **organisation!**
- Midsummer school in QCD (24 June - 6 July 2024)
- Spåtind 2025: Nordic Conference on Particle Physics (Early 2025)

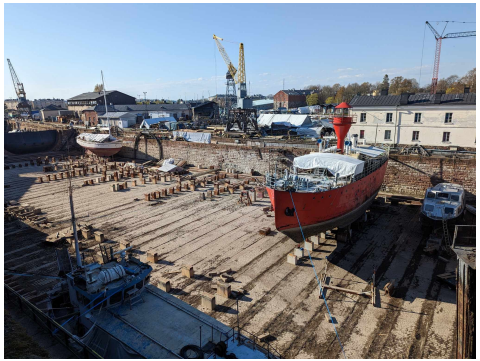
Midsummer School in QCD

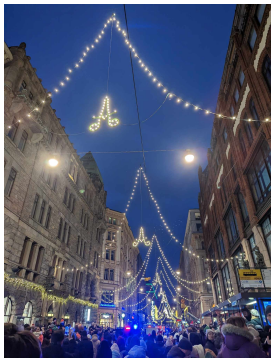
indico.cern.ch/event/qcd2024

Conclusion

Conclusion

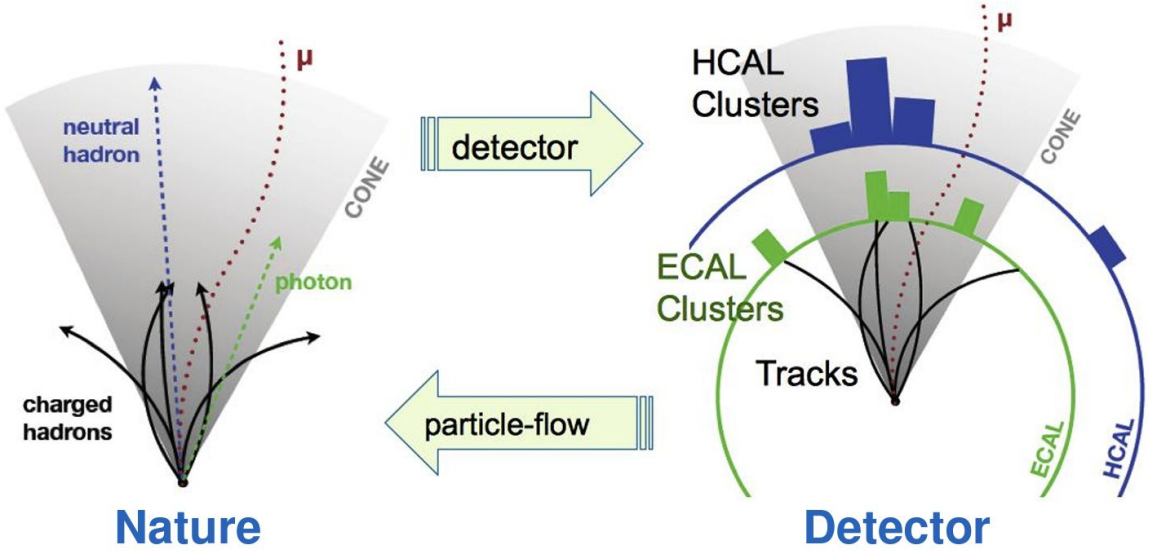
- **Introduction**
  - CMS trigger system: L1 and HLT to reduce data taking rate
  - Data scouting: save only PF candidates at HLT
    → increase event rate
  - Jets are signature of quark + gluons
    + other interesting new physics (e.g. boosted objects)
- **Boosted H → bb exploiting data scouting**
  - scouting data contains 20% more H → bb events overall, particular gain at low momentum region
  - ongoing studies on tagger performance on scouting jets
- **JEC for scouting jets**
  - offline and online jets can support calibration of each other
- **Other activities**
  - learning a lot, meeting a lot of new people, and a lot of memories





Backup

Interlude: Particle Flow (PF)

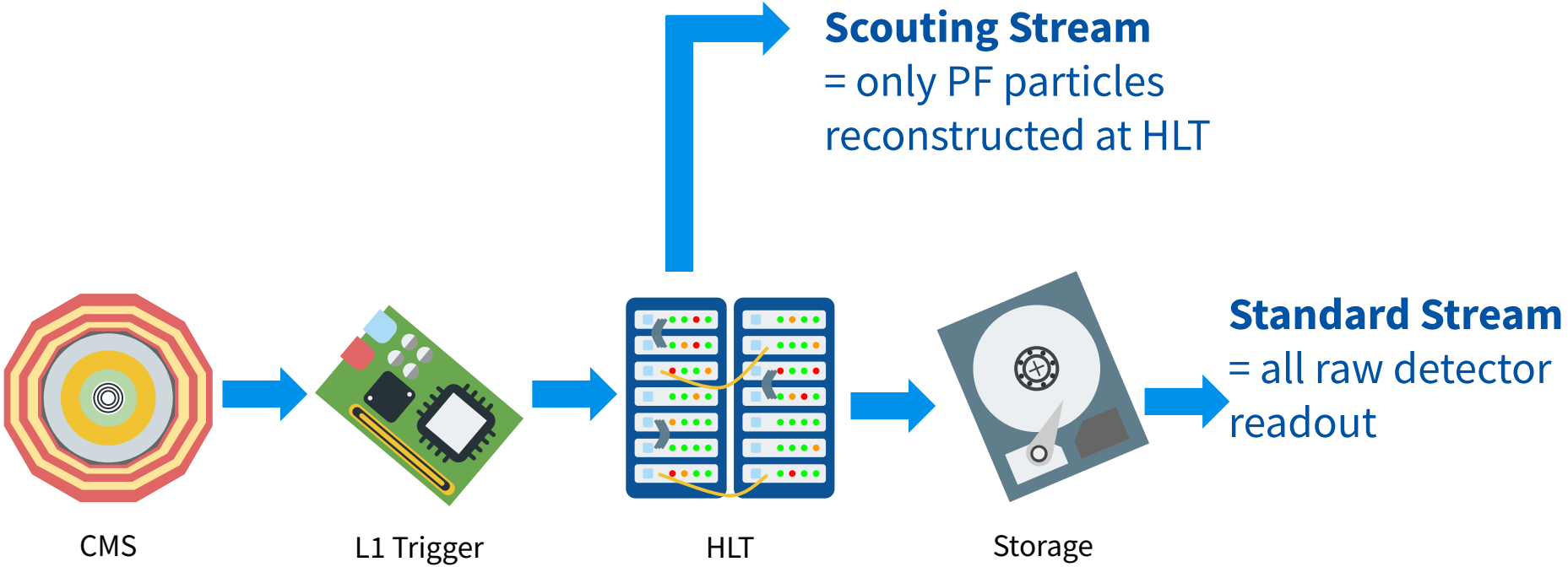


Tracks, Calorimeter Clusters

Particles:

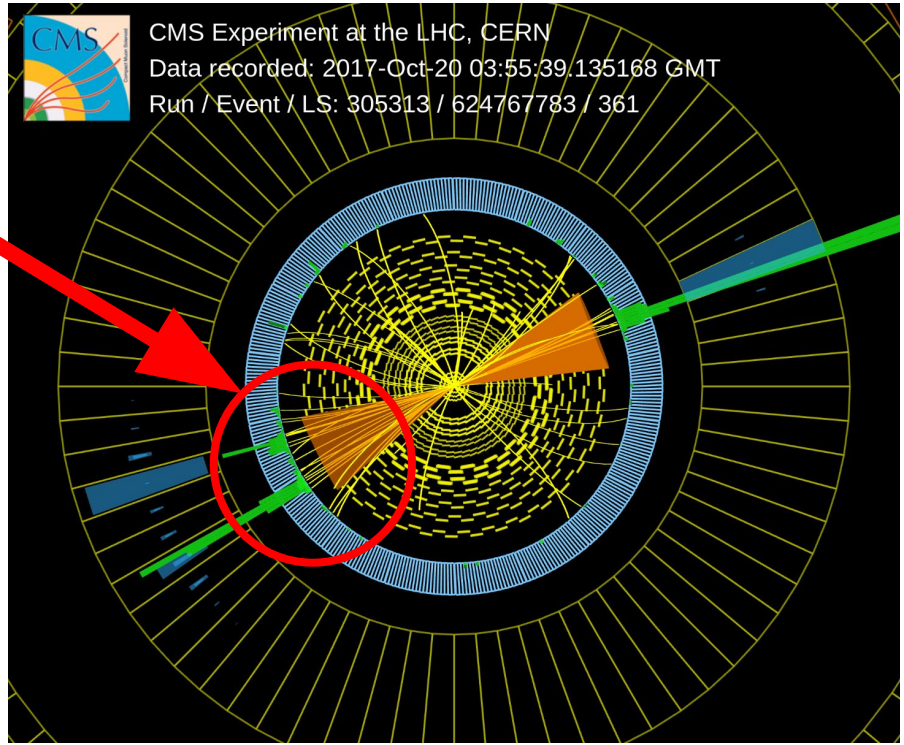
Electron, Photon, Hadron, Muon

Data Scouting (HLT Scouting)



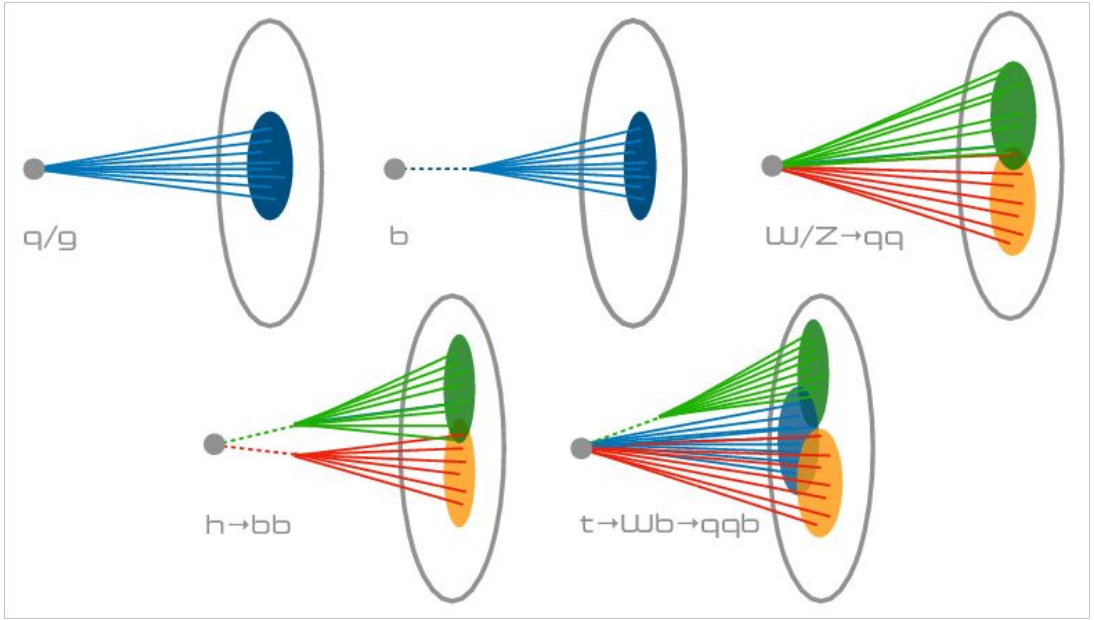
Motivation: boosted jets

2 subjets



CMS-PAS-HIG-19-003

Motivation: jet substructure

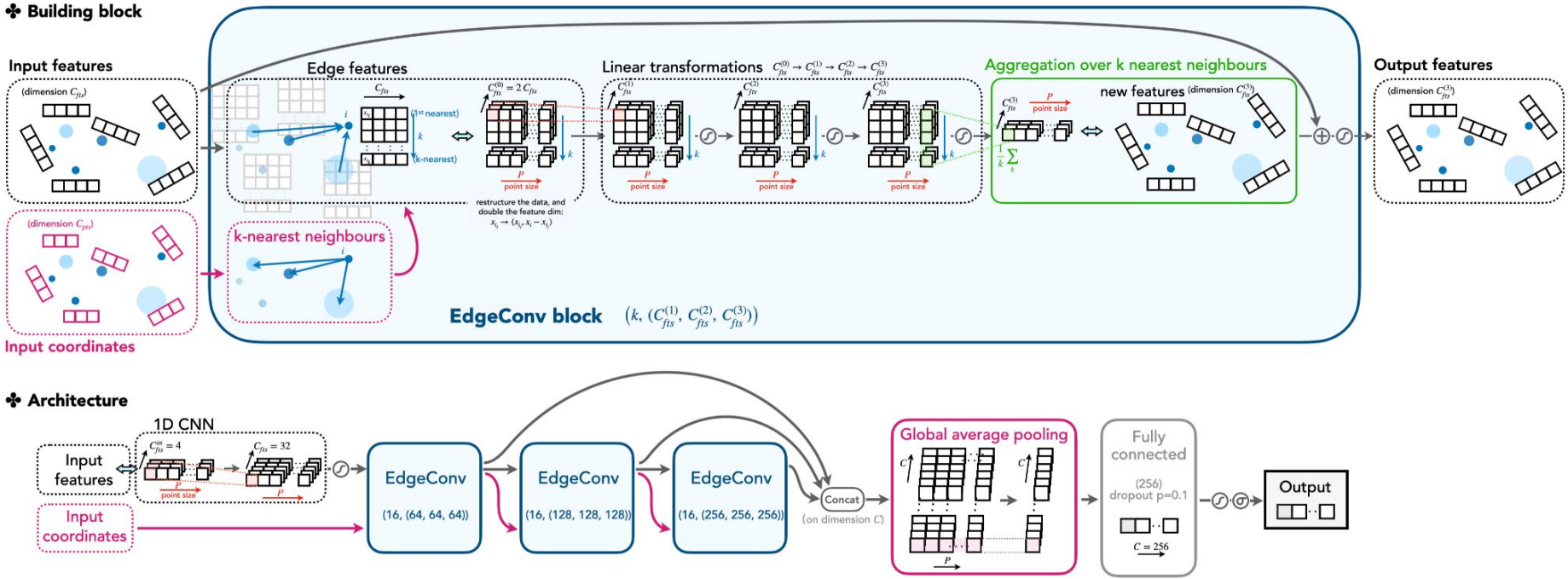


arxiv.org/abs/1909.12285

Jet structure indicates type of original particles

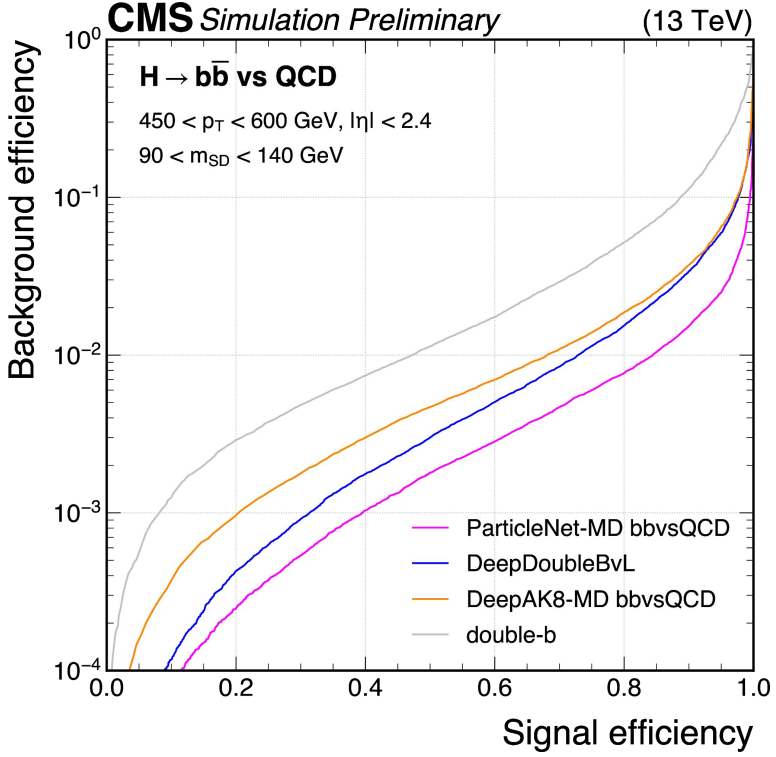
→ **jet tagging,**
e.g. with neural network (ParticleNet, ParT, etc.)

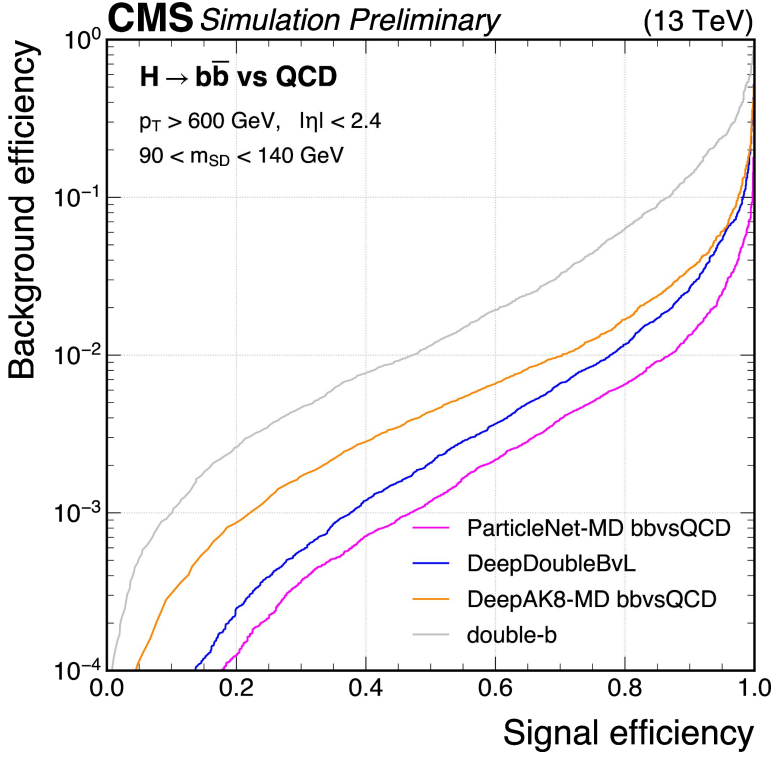
ParticleNet



[CMS Machine Learning Documentation - ParticleNet]

B-tagging performance (Run 2)

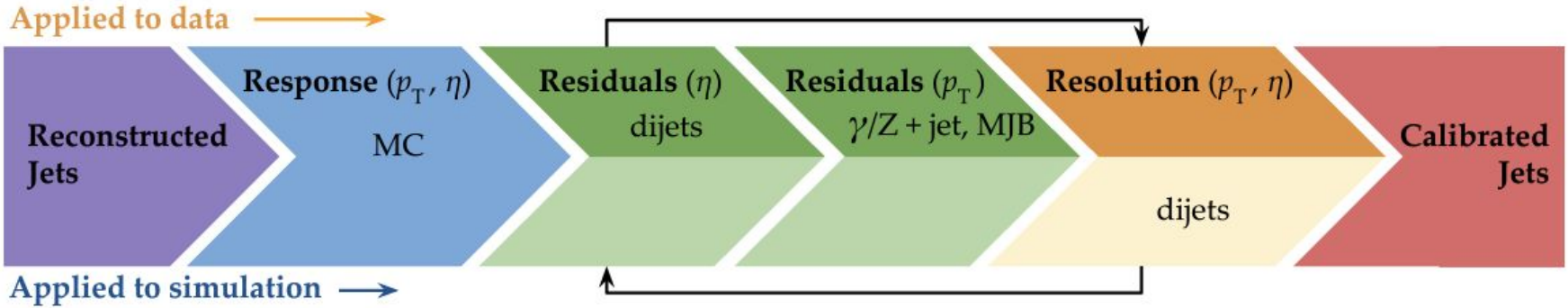




CMS-PAS-BTV-22-001

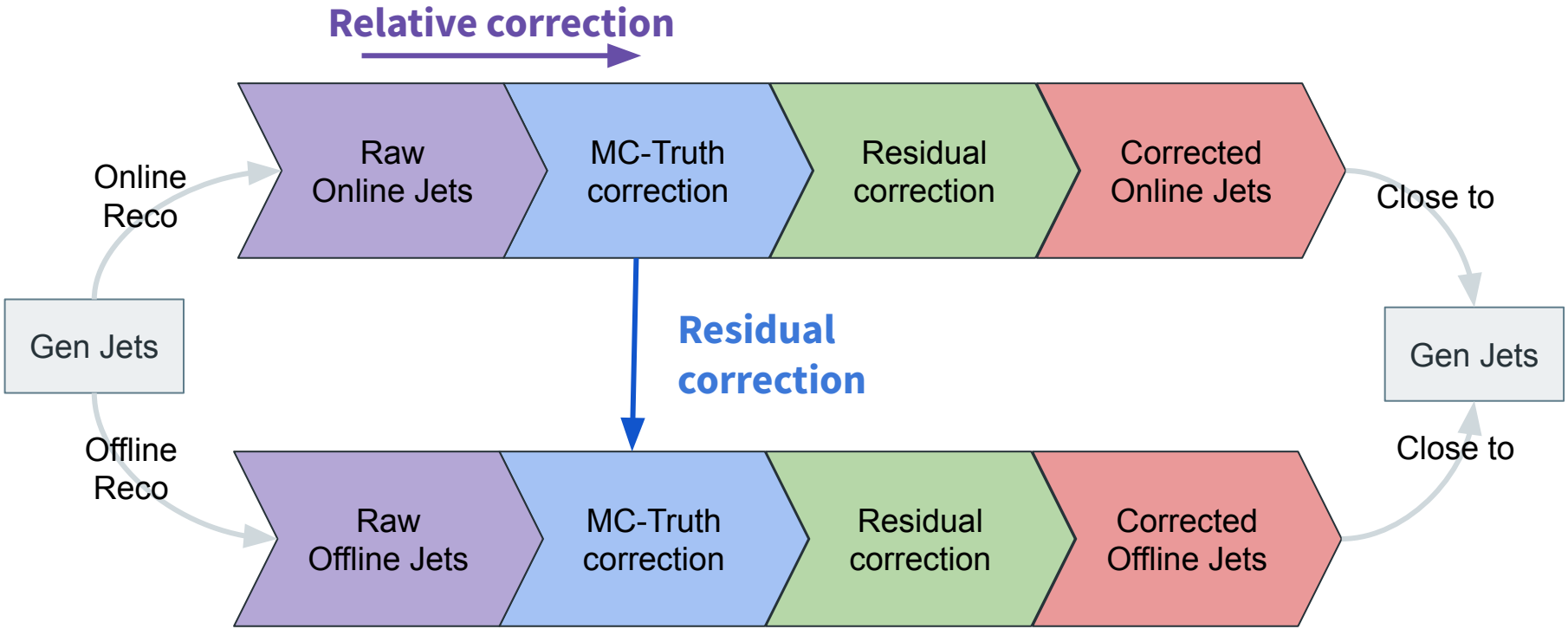
JEC in CMS Run 3

- Jet is clustered from PF candidates by **anti-kt algorithm** with R=0.4 or R=0.8
- **PUPPI (PileUp Per Particle Identification)** is applied to mitigate effects from pileup
- JEC is then applied: factorized approach - each step aims to correct specific effect



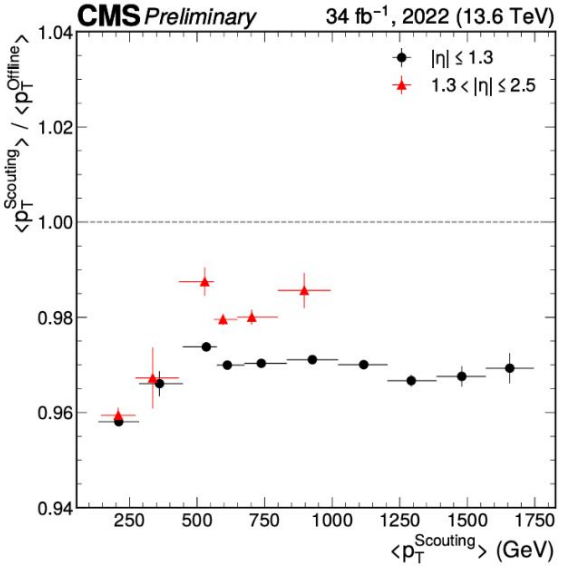
CMS-DP-2022-054

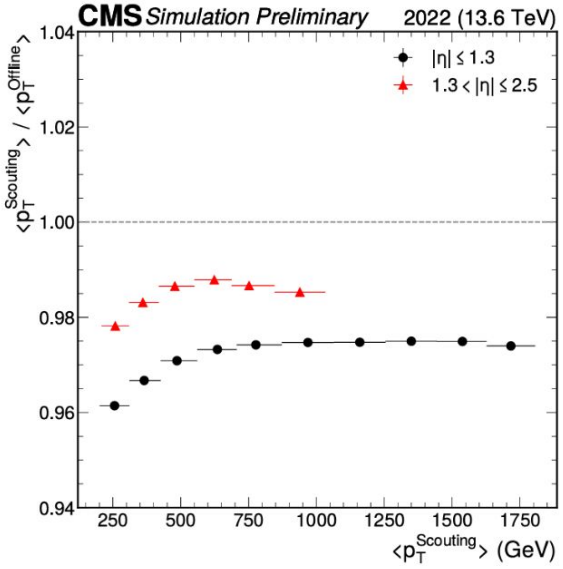
JEC for scouting jets

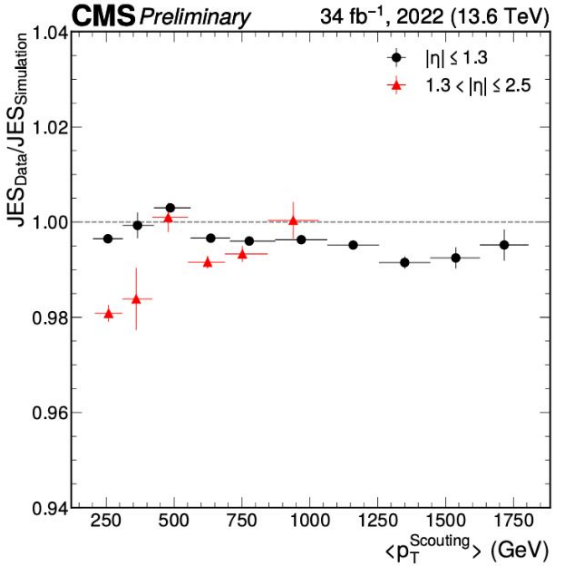


JEC for scouting jets

Studies by Adelina Lintuluoto



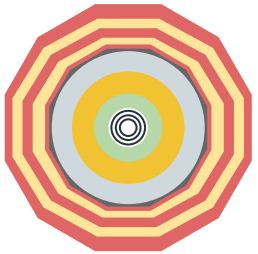


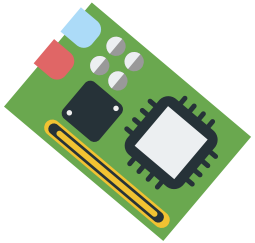


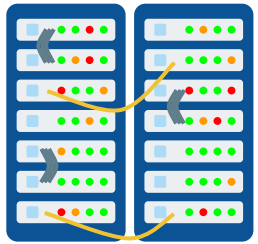
CMS-DP-2023-072

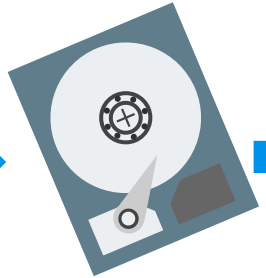
L1 Scouting

Scouting Stream
= only PF particles reconstructed at HLT









Standard Stream
= all raw detector readout

CMS

L1 Trigger

HLT

Storage

L1 Scouting (40 MHz)
= L1 Primitives