

SMARTHEP

REAL-TIME ANALYSIS FOR
SCIENCE AND INDUSTRY

SMARTHEP Yearly Meeting

**ESR1: Machine learning and
Real-Time Analysis for
Higgs boson measurements
and fleet safety**

Lund, 27 November 2023

Patin Inkaew



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

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
but not much experience with particle physics analysis

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
but not much experience with particle physics analysis

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
~~but not much~~ experience with particle physics analysis

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

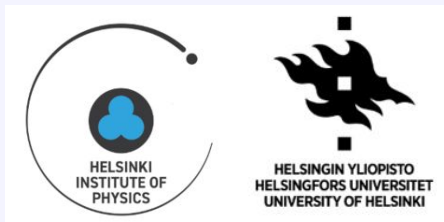
Stanford University, CA, USA (Thai Government Scholarship)

- **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
a bit more experience with particle physics analysis

ESR1: Machine learning and Real-Time Analysis for Higgs boson measurements and fleet safety



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

Secondment:

CERN, Switzerland

Secondment:

Verizon Connect, Italy

Supervisors:

Mikko Voutilainen,
Henning Kirschenmann

Collaborator:

Maurizio Pierini

Collaborators:

Leonardo Taccari,
Francesco Sambo

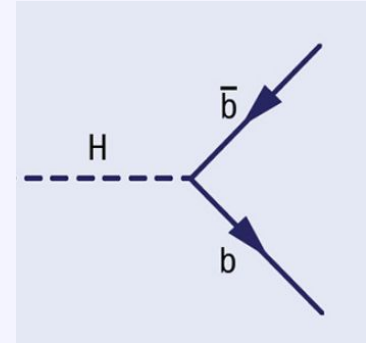


ESR1: Machine learning and Real-Time Analysis for Higgs boson measurements and fleet safety



Why Higgs?

- Higgs boson was one of the initial goal of LHC
- Higgs boson was discovered in 2012!
- Higgs boson gives mass to other particles
- **But where does Higgs' mass come from?**
- If Higgs gives its own mass, self-interaction (di-Higgs) can be observed
- Unfortunately, di-Higgs events are extremely rare → goal for HL-LHC
- Improve Higgs boson measurement
 - $H \rightarrow \gamma\gamma$ / $H \rightarrow ZZ^* \rightarrow 4l$ are rare, but have low background
 - $H \rightarrow bb$ has higher branching ratio, but suffers larger background
 - Real-time analysis can improve data acquisition rate in trigger system
 - Validated with frequent and well-studied $Z \rightarrow bb$



ESR1: Machine learning and Real-Time Analysis for Higgs boson measurements and fleet safety



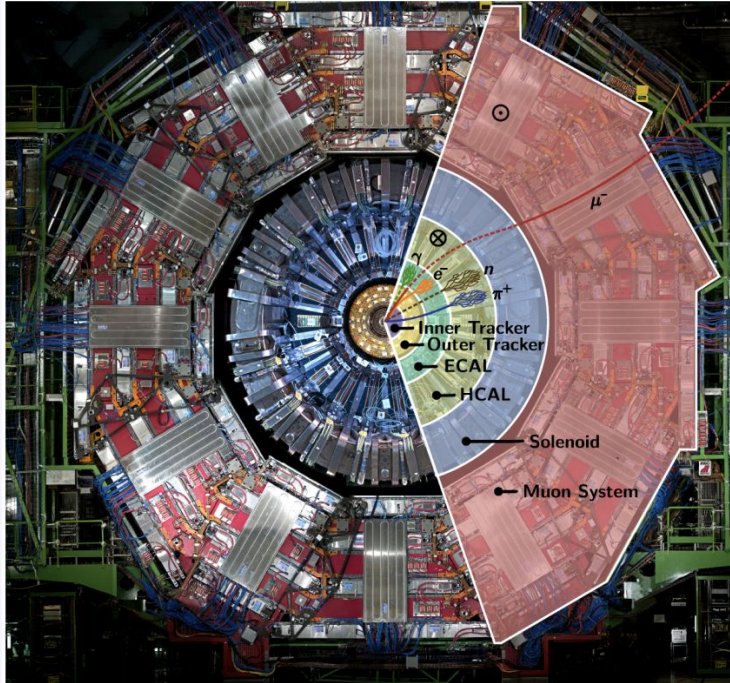
Data Scouting Jet for Run 3 at CMS



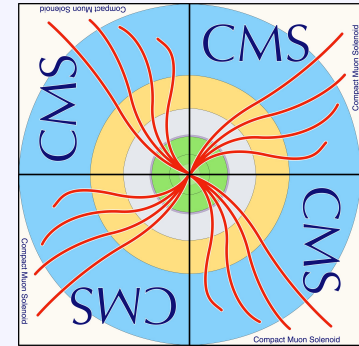
Data Scouting Jet for Run 3 at CMS



The Compact Muon Solenoid Experiment (CMS)

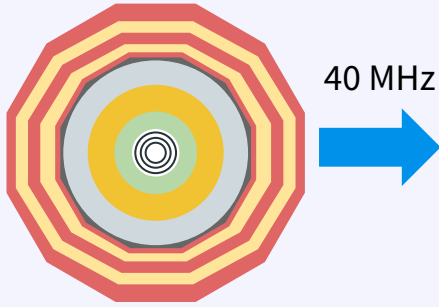


CMS is a general-purpose detector, one of the four large experiments at LHC. Focus: rare event search, precise measurement



CMS Trigger System

LHC collides pp every ~ 25 ns = 40 MHz

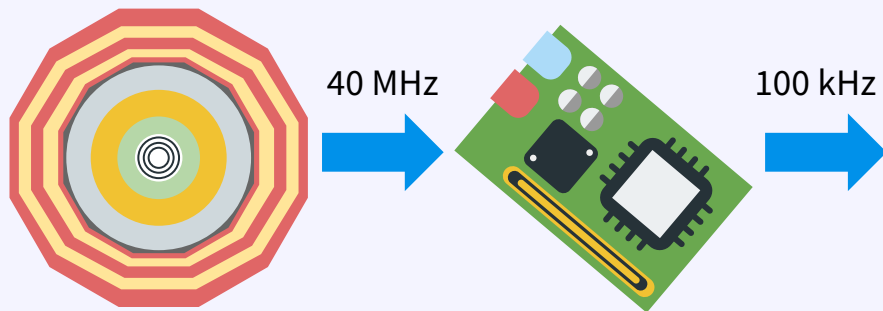


CMS Detector



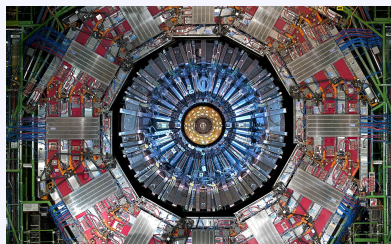
CMS Trigger System

L1 (Hardware based on FPGA) reduces rate to ~ 100 kHz



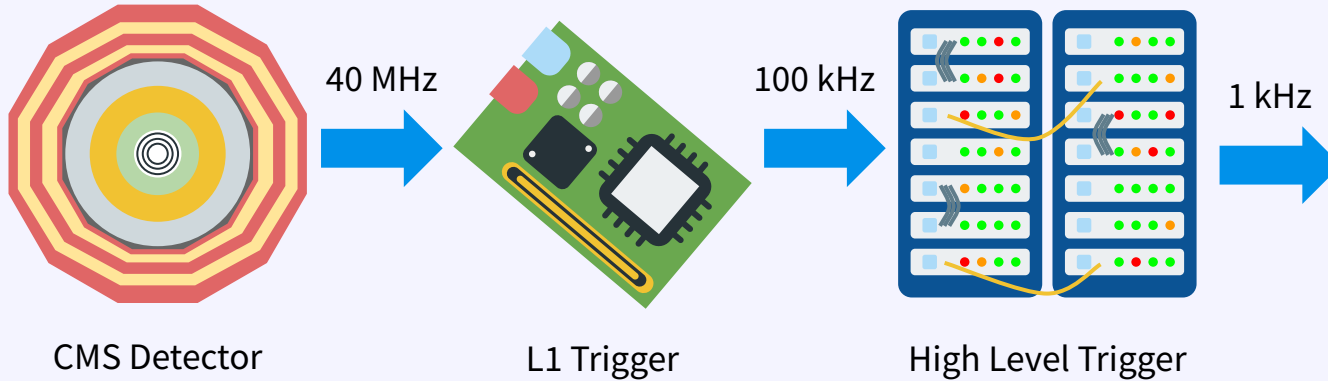
CMS Detector

L1 Trigger



CMS Trigger System

HLT (Computer farm) reduces rate to ~ 1 kHz



CMS Detector



L1 Trigger

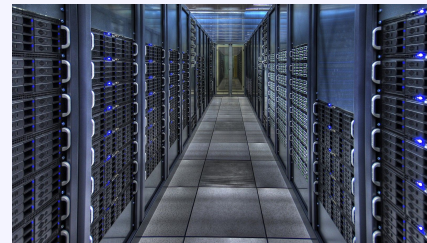
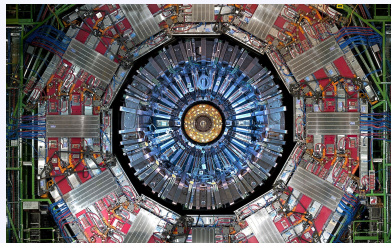
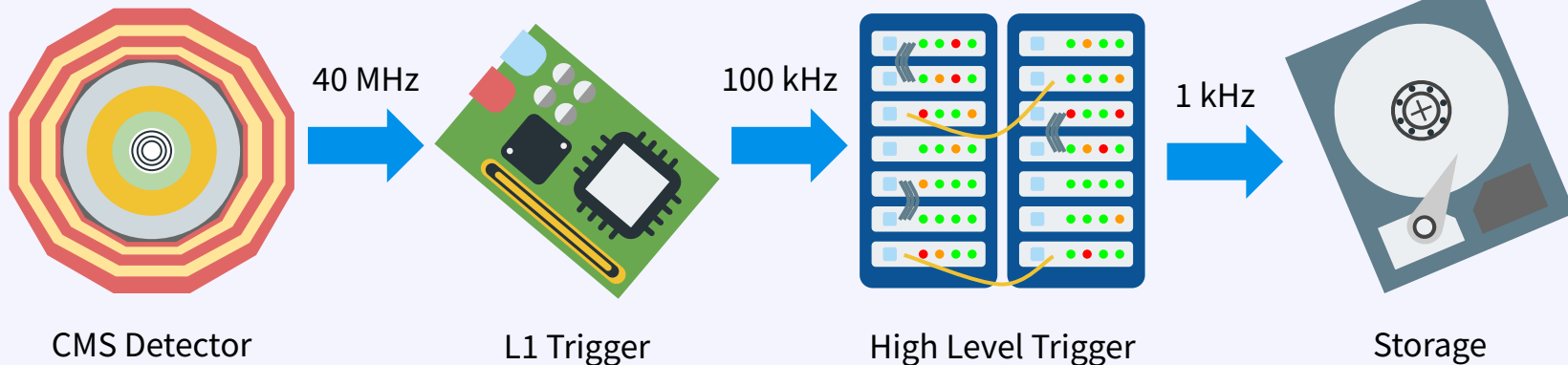


High Level Trigger



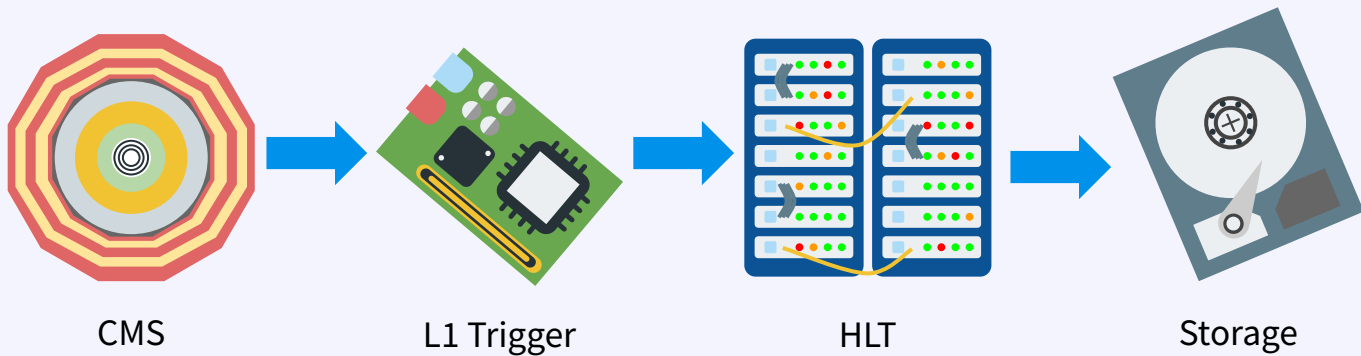
CMS Trigger System

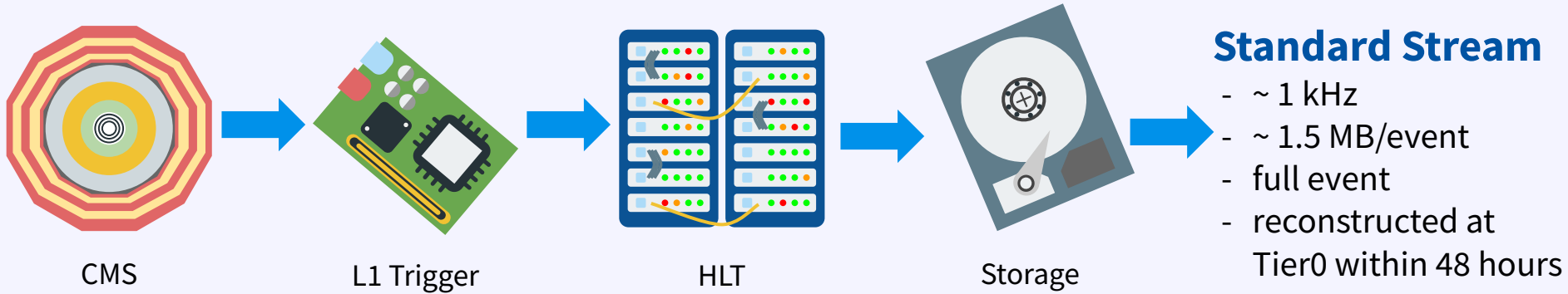
Data is transferred and stored at Tier 0

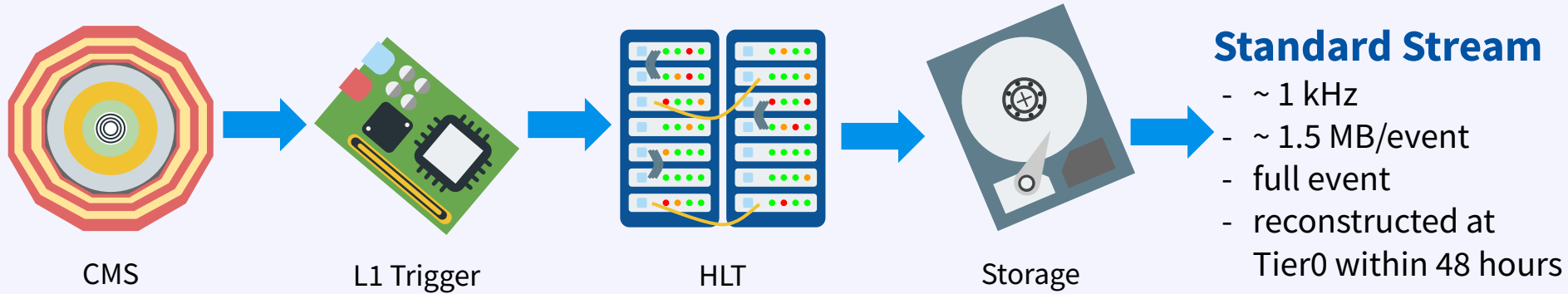


Data Scouting Jet for Run 3 at CMS

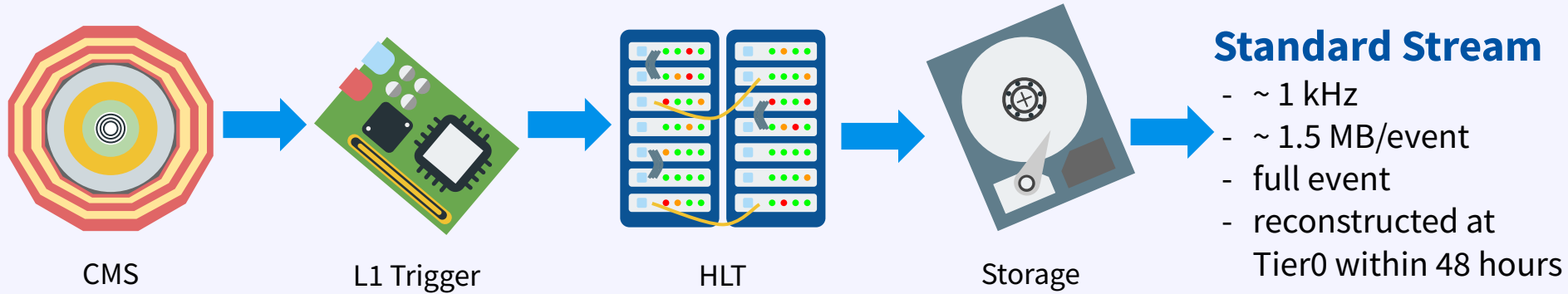






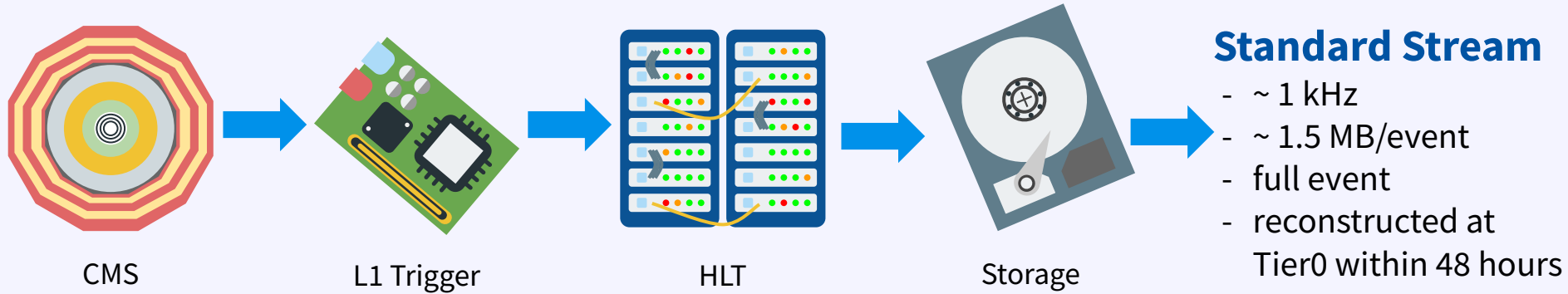


The real bottleneck is data recording rate (MB/sec), not event rate (event/sec)



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

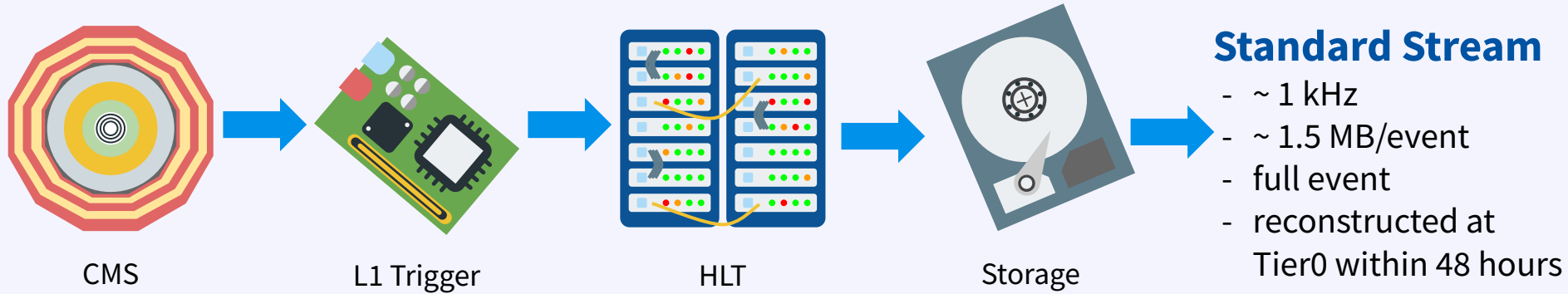
$$\text{Data writing rate (MB/s)} = \text{event size (MB)} \times \text{rate (Hz)}$$



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

$$\text{Data writing rate (MB/s)} = \text{event size (MB)} \times \text{rate (Hz)}$$

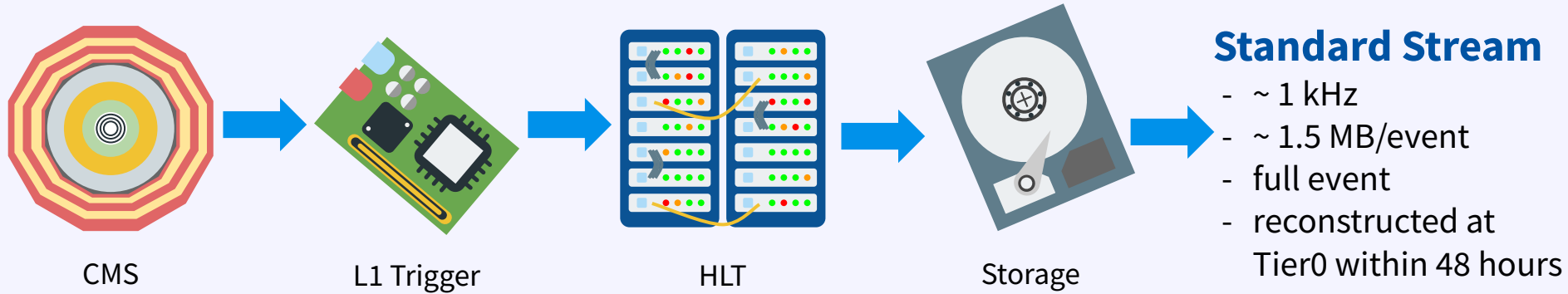




The real bottleneck is data recording rate (MB/sec), not event rate (event/sec)

$$\text{Data writing rate (MB/s)} = \text{event size (MB)} \times \text{rate (Hz)}$$

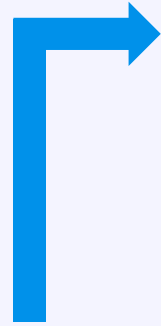




The real bottleneck is data recording rate (MB/sec), not event rate (event/sec)

$$\text{Data writing rate (MB/s)} = \text{event size (MB)} \times \text{rate (Hz)}$$

only keep smaller high-level description, and discard raw detector signals

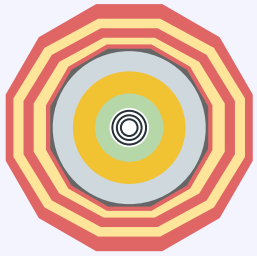


Scouting Stream

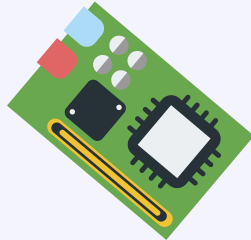
- reduced data format
- event size reduced by 100x
- event rate increased by 30x
- based on reconstruction at HLT (no offline reconstruction)

Standard Stream

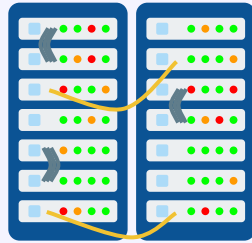
- ~ 1 kHz
- ~ 1.5 MB/event
- full event
- reconstructed at Tier0 within 48 hours



CMS



L1 Trigger



HLT



Storage



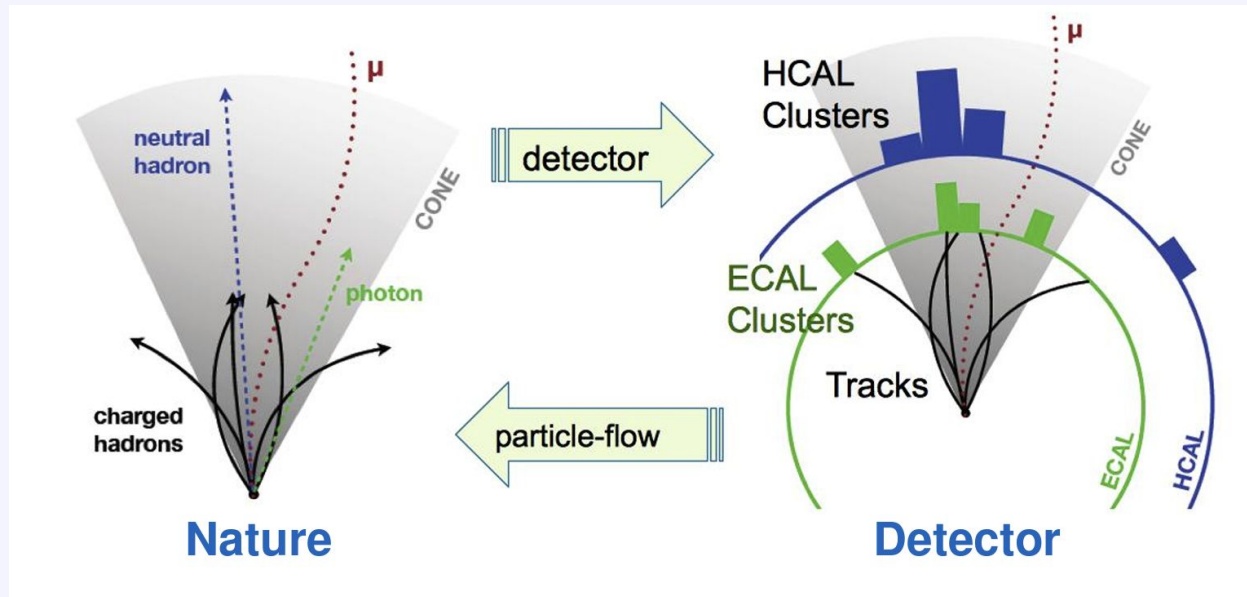
The real bottleneck is data recording rate (MB/sec), not event rate (event/sec)

$$\text{Data writing rate (MB/s)} = \text{event size (MB)} \times \text{rate (Hz)}$$



only keep smaller high-level description, and discard raw detector signals

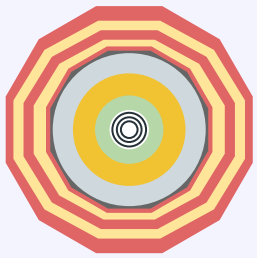
Interlude: Particle Flow (PF)



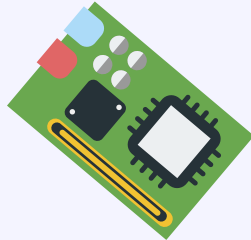
PF combines all sub-detector primitives (tracks, clusters) to produce physical objects: electron, photon, muon, neutral and charged hadrons.

Scouting Stream
= only PF particles
reconstructed at HLT

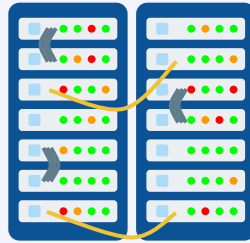
Standard Stream
= all raw detector
readout



CMS



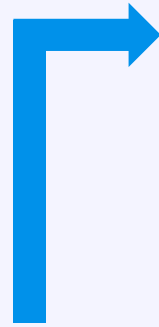
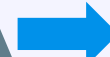
L1 Trigger



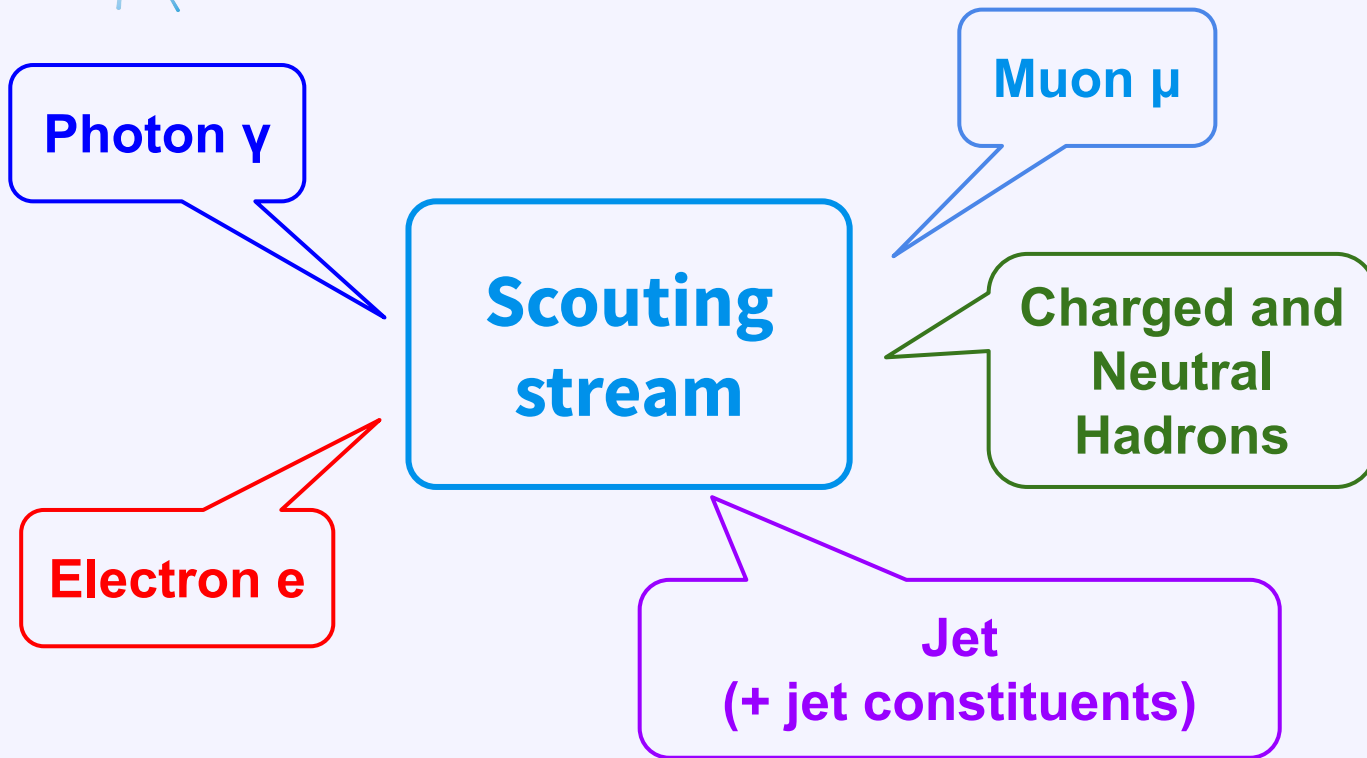
HLT



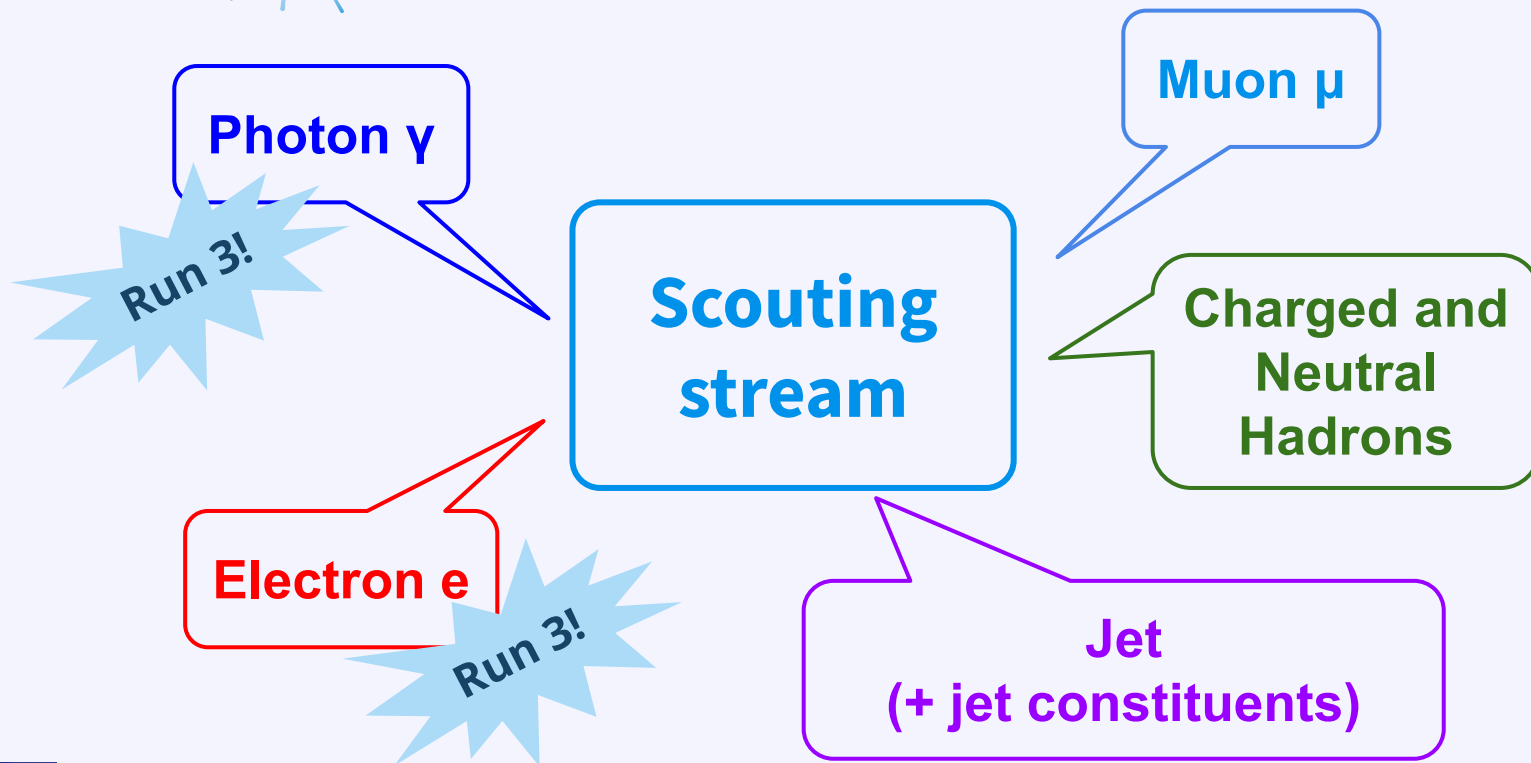
Storage



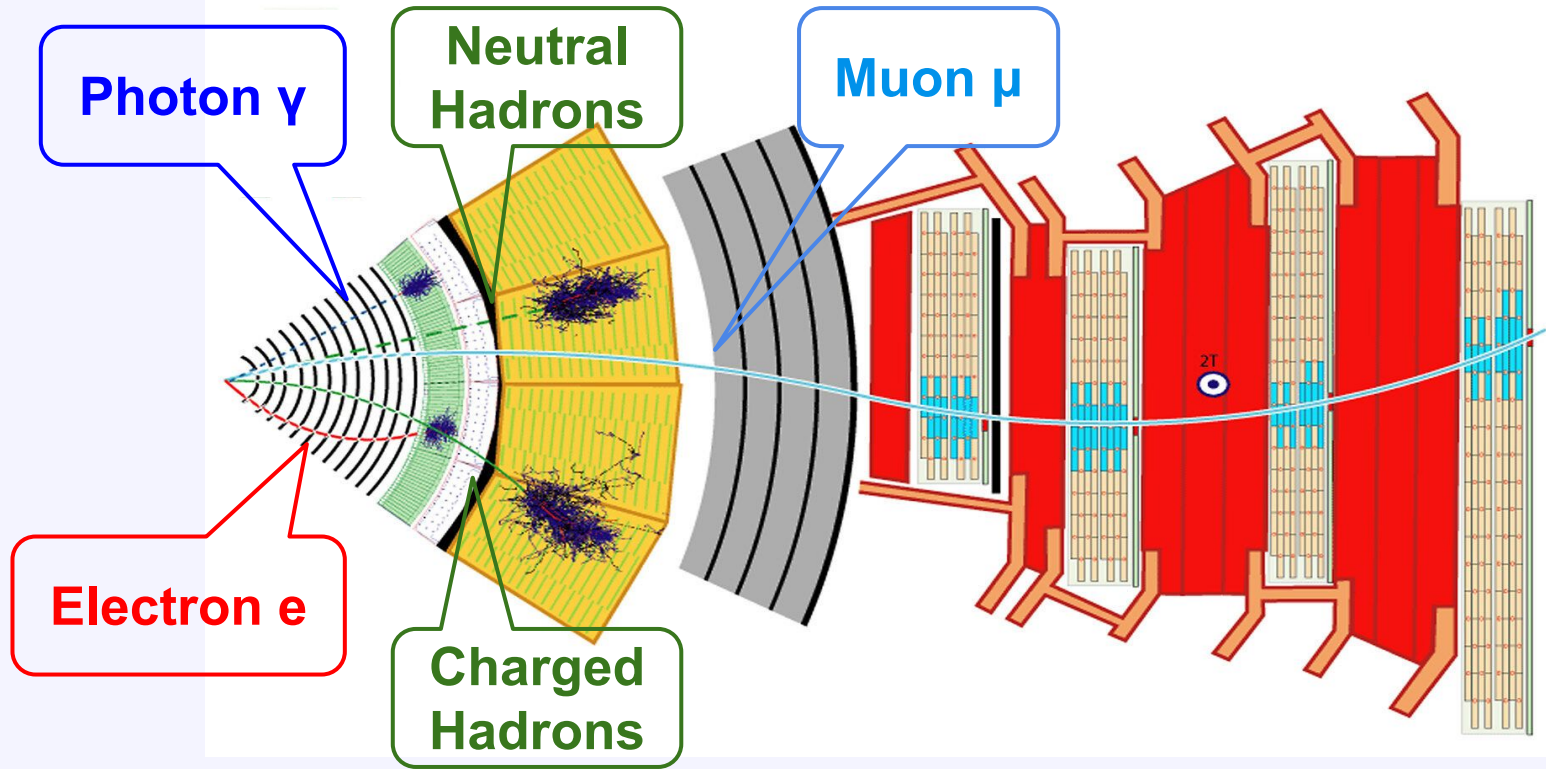
What's in scouting stream



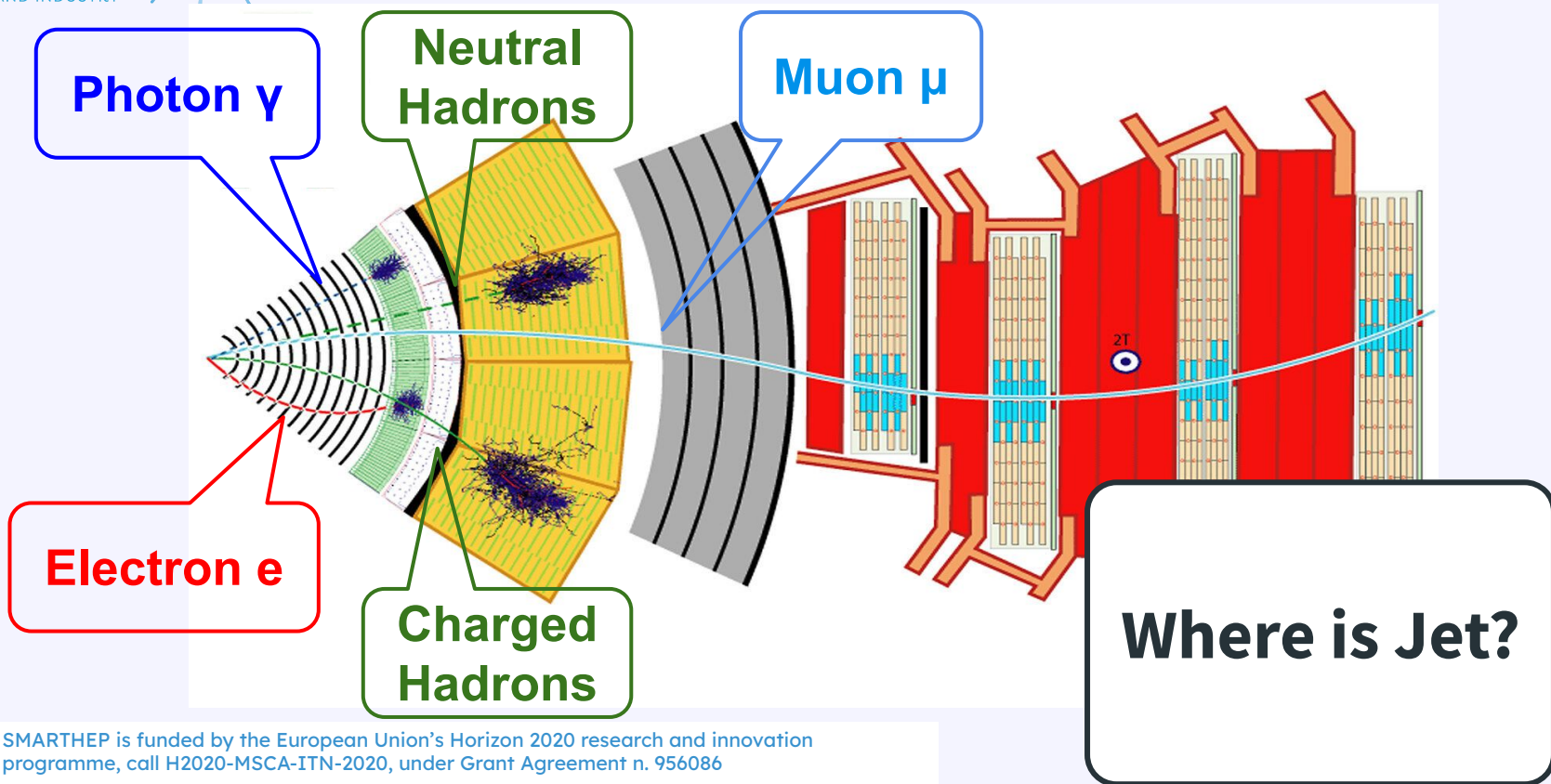
What's in scouting stream



What's in scouting stream



What's in scouting stream

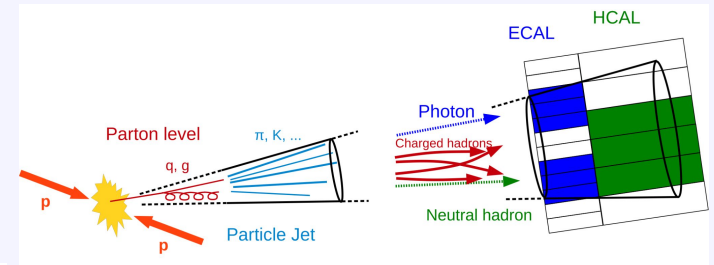


Data Scouting Jet for Run 3 at CMS



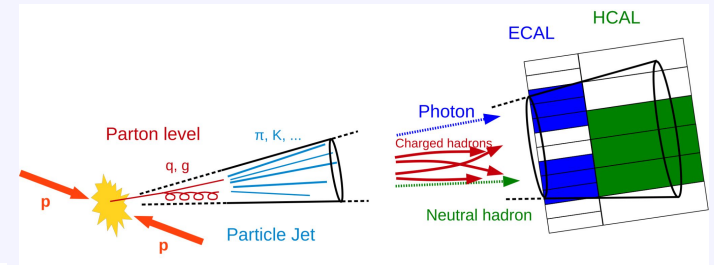
What is “Jet”?

- Jets are the signatures of quarks and gluons (?)
 - color confinement
 - hadronization: quarks and gluons quickly become a cone of particles “**Jet**”



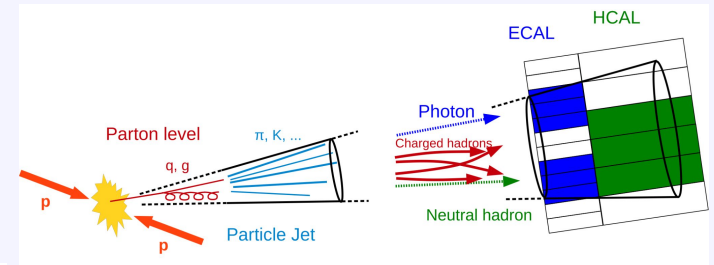
What is “Jet”?

- 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



What is “Jet”?

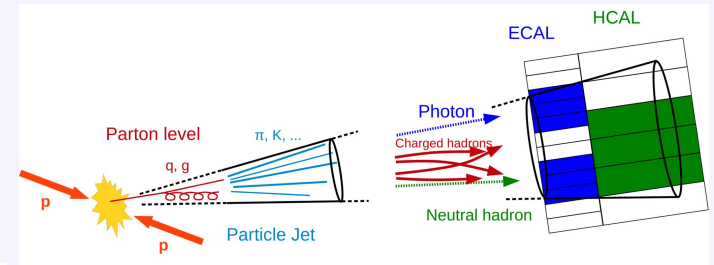
- Jets are the signatures of quarks and gluons (?)
 - color confinement
 - hadronization: quarks and gluons quickly become a cone of particles “**Jet**”
 - **Jet clustering**: group particles to a jet
 - In a more generalized picture, almost everything becomes a jet: gluons, quarks, top quarks, W/Z boson,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

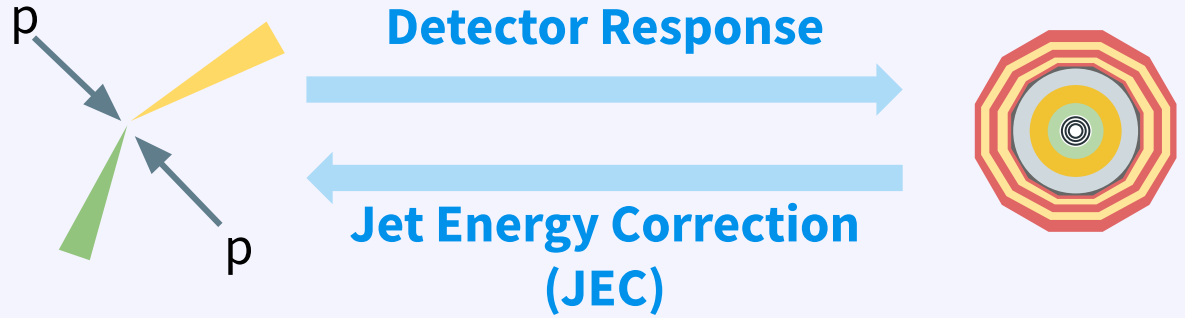


What is “Jet”?

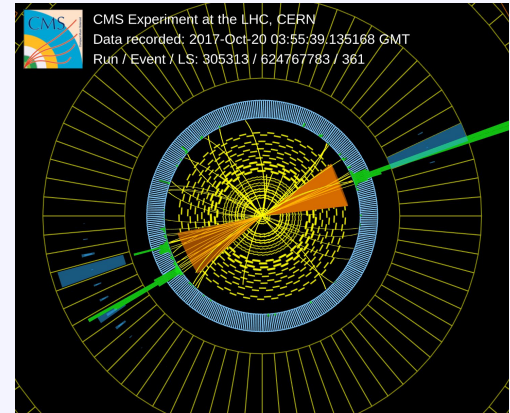
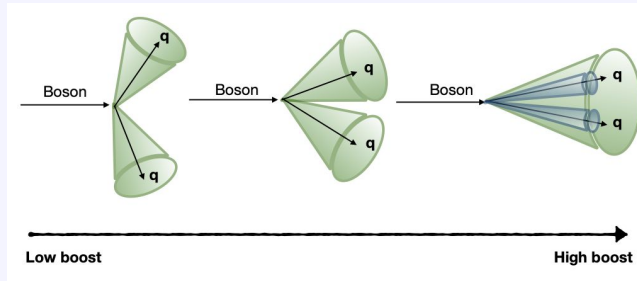
- Jets are the signatures of quarks and gluons (?)
 - color confinement
 - hadronization: quarks and gluons quickly become a cone of particles “**Jet**”
 - **Jet clustering:** group particles to a jet
 - In a more generalized picture, almost everything becomes a jet: gluons, quarks, top quarks, W/Z boson,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

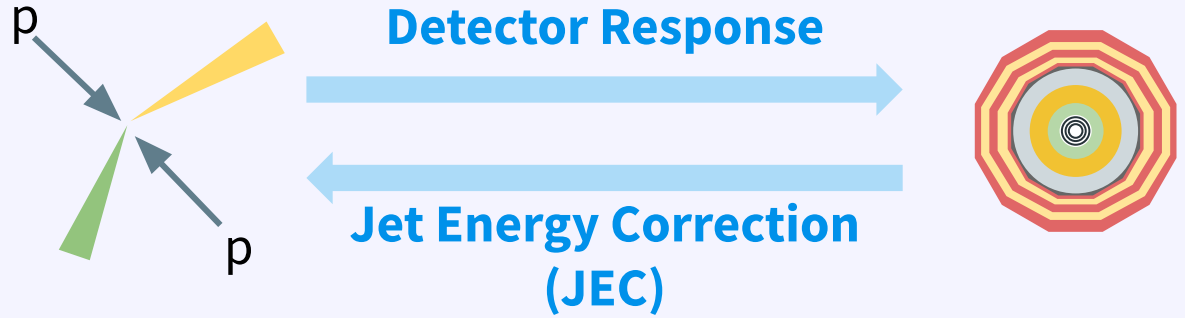
Lots of interesting stuff!



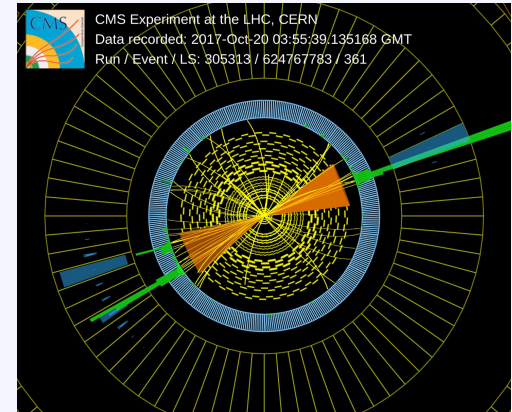
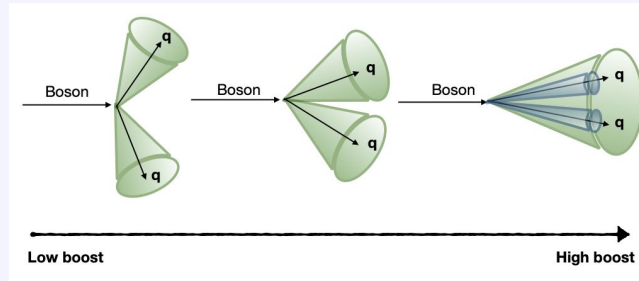


- 1. Jet Energy Correction (JEC) for scouting jets**
- 2. Boosted $H \rightarrow bb$ analysis using scouting data**



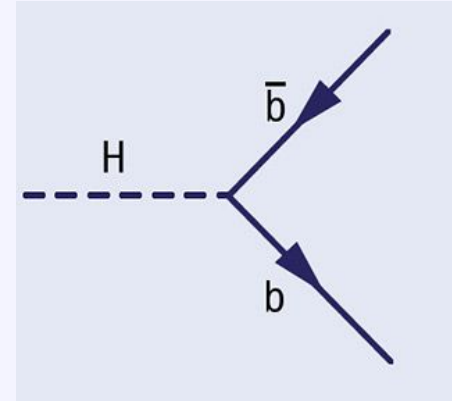
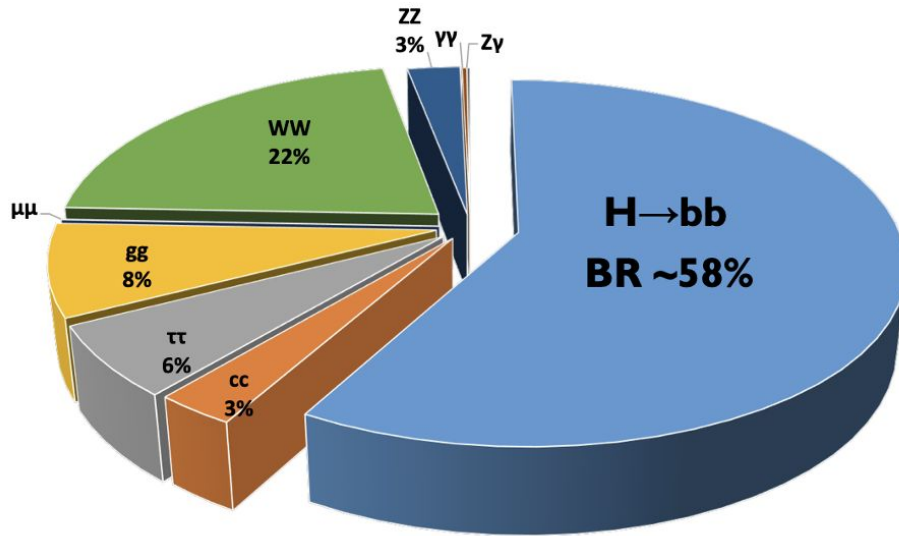


1. Jet Energy Correction (JEC) for scouting jets
2. Boosted $H \rightarrow bb$ analysis using scouting data



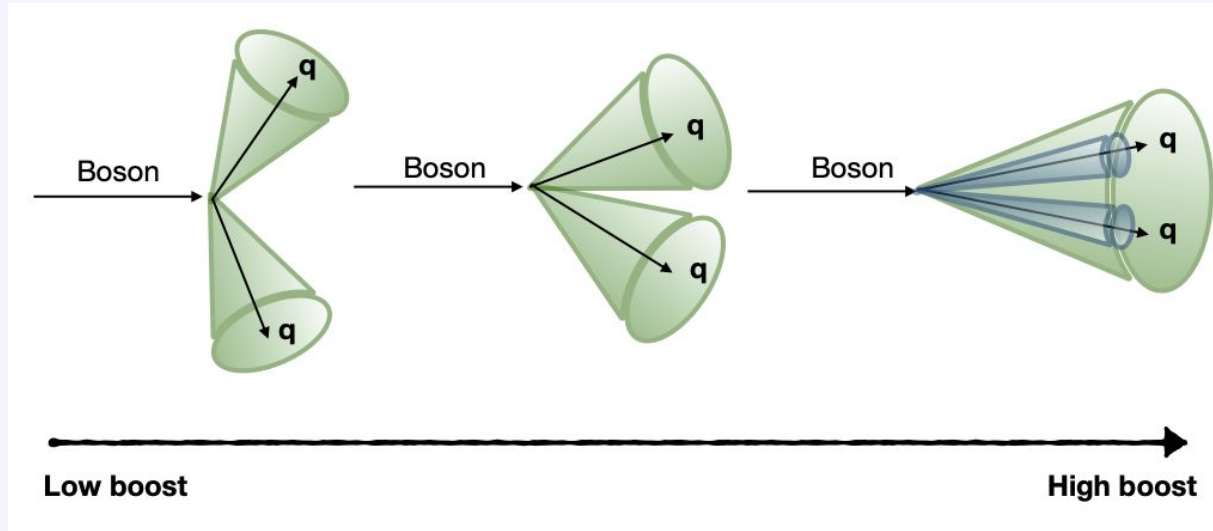
Boosted objects analysis exploiting jet substructure

Decay modes:



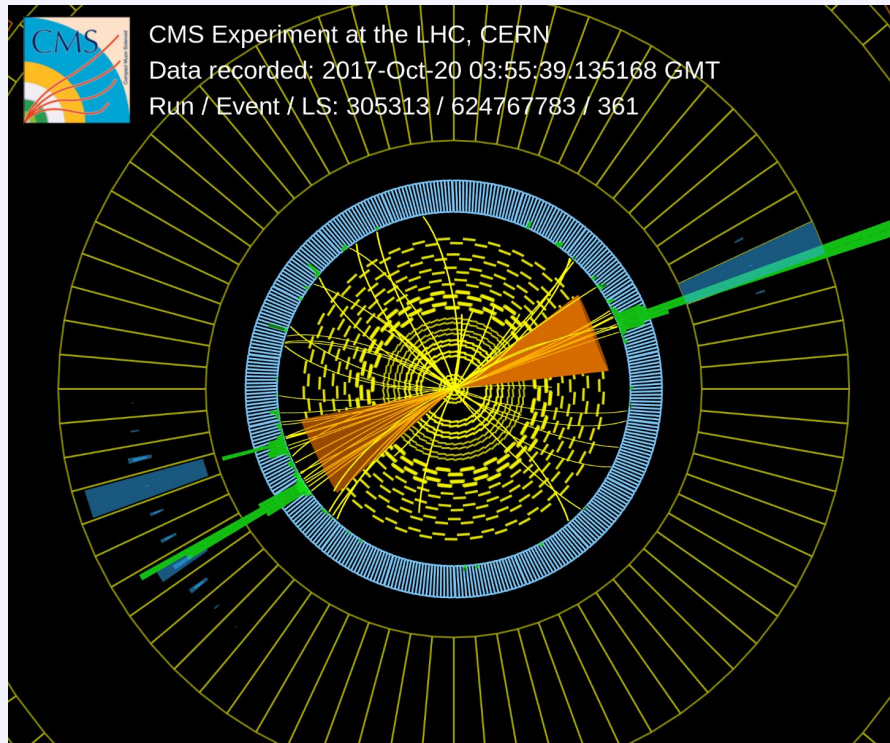
- $H \rightarrow b\bar{b}$ is the most probable decay mode
- However, suffer from enormous background (QCD)

Boosted objects analysis exploiting jet substructure



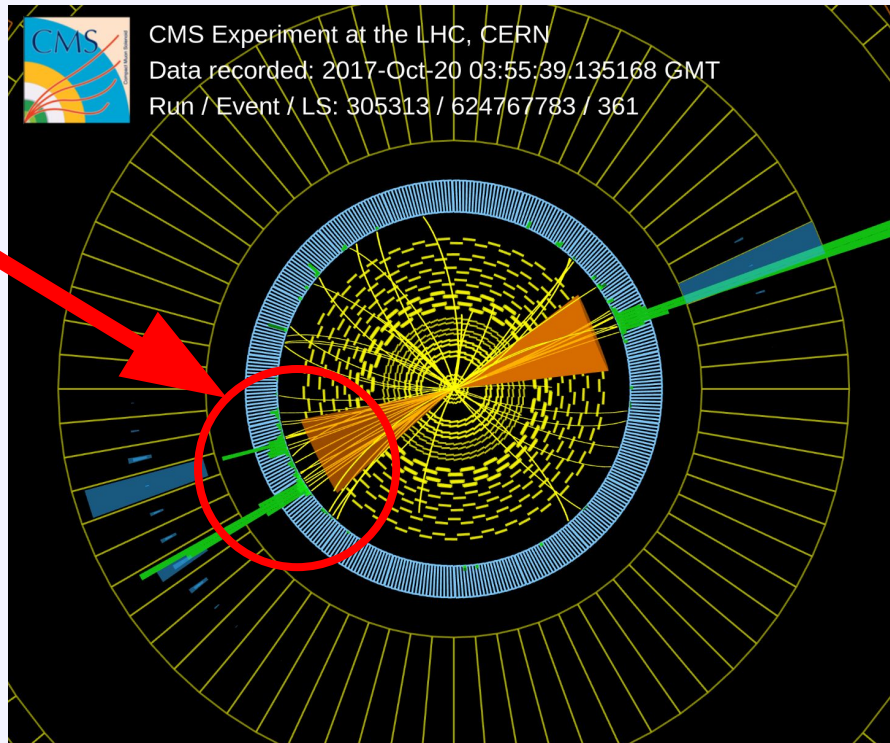
With high boost, two jets originating from single boson merge into single large jet.
Probing jet substructure can improve signal sensitivity from QCD background.

Boosted objects analysis exploiting jet substructure

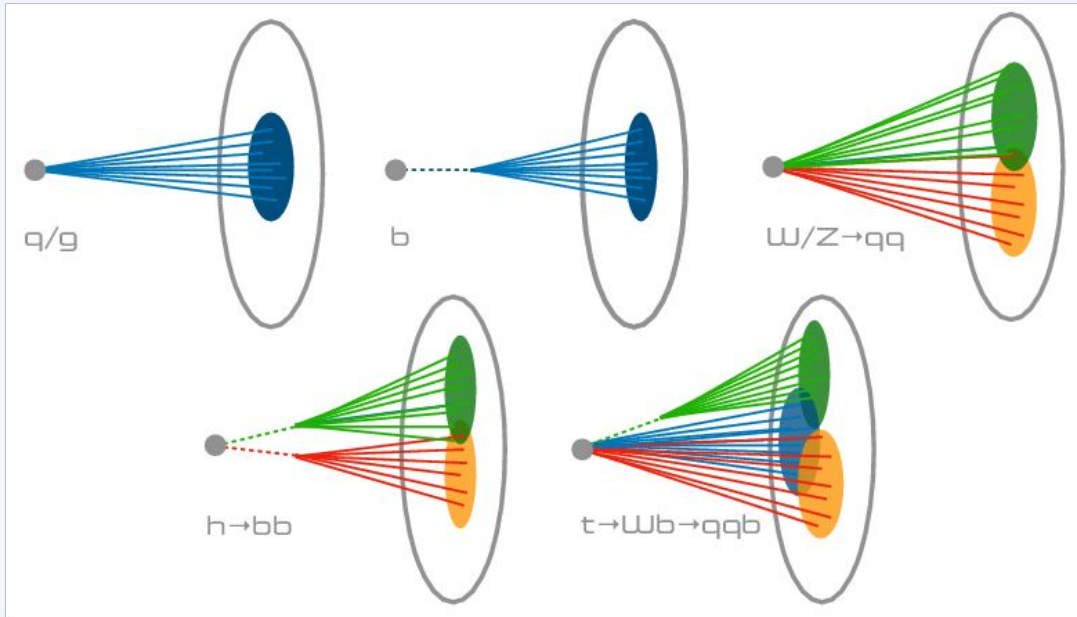


Boosted objects analysis exploiting jet substructure

2 jets



Boosted objects analysis exploiting jet substructure

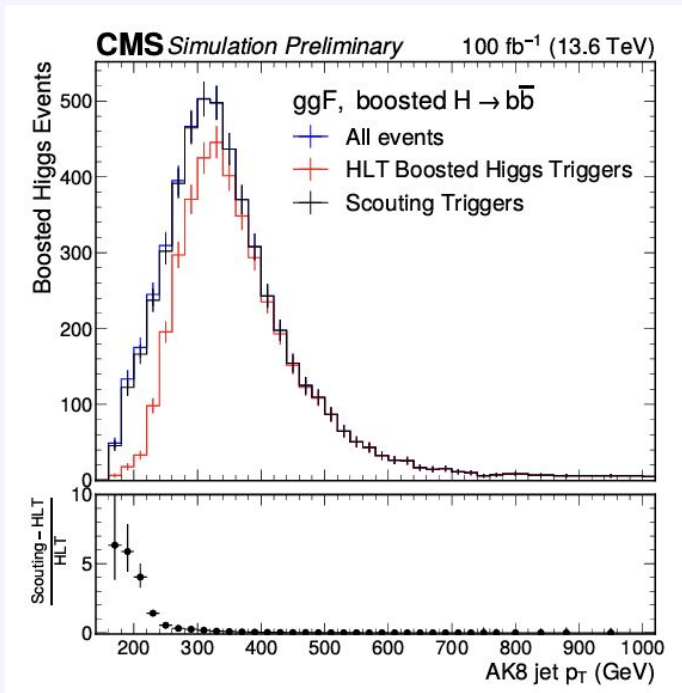
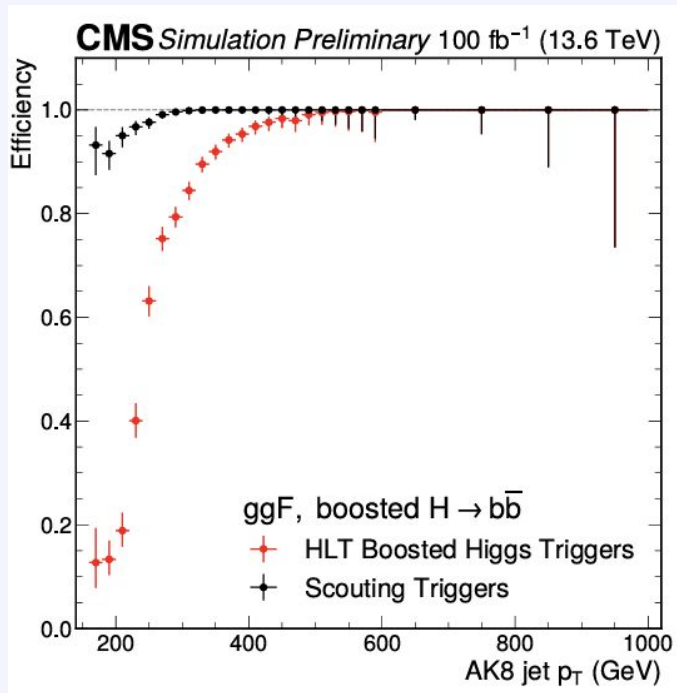


Jet structure indicates type
of original particles

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

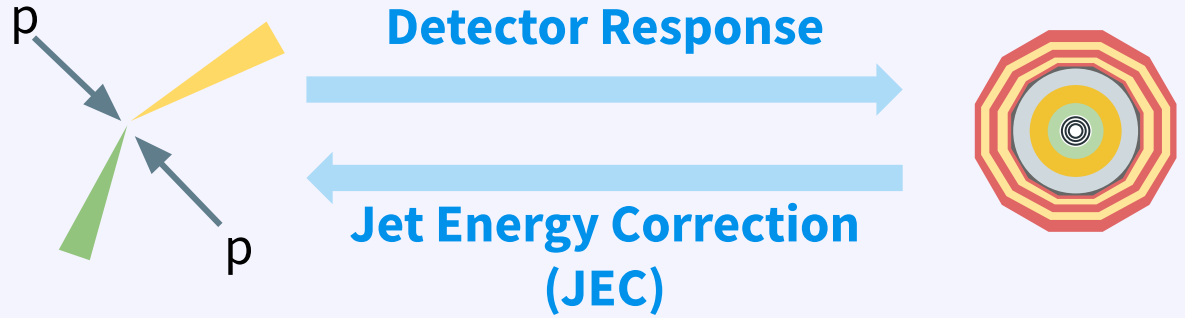
arxiv.org/abs/1909.12285

Trigger efficiency studies of the CMS Run-3 Data Scouting

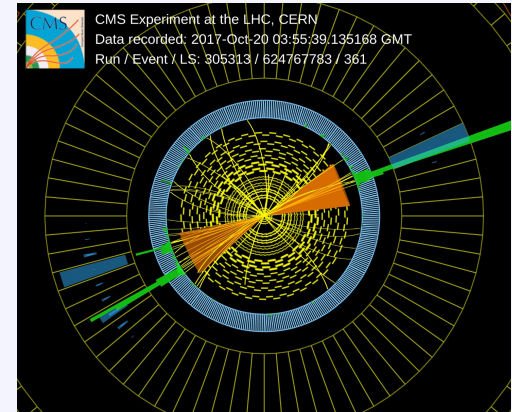
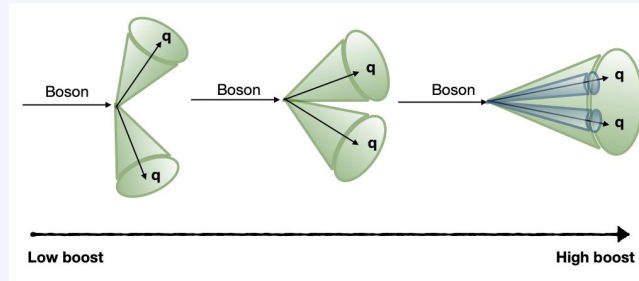


Overall number of events gain ~20%

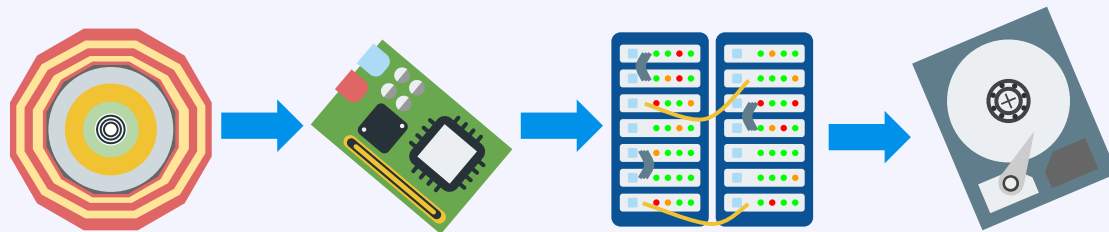
See gains particularly in low p_T region



1. Jet Energy Correction (JEC) for scouting jets
2. Boosted $H \rightarrow bb$ analysis using scouting data

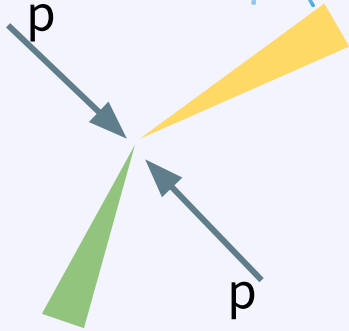


Jet Energy Correction (JEC)

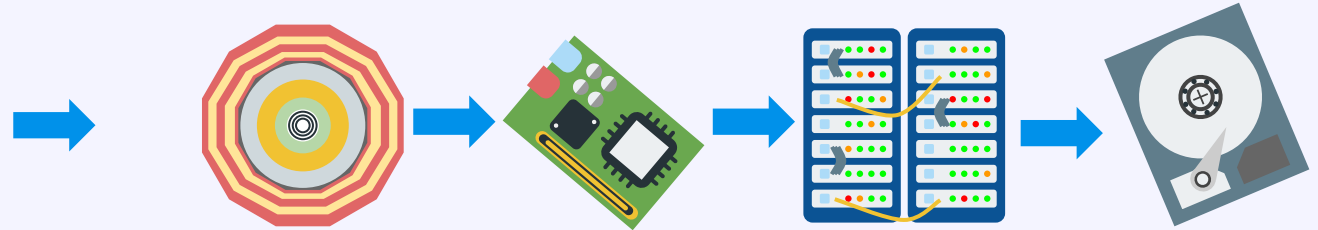


SMARTHEP

REAL-TIME ANALYSIS FOR
SCIENCE AND INDUSTRY

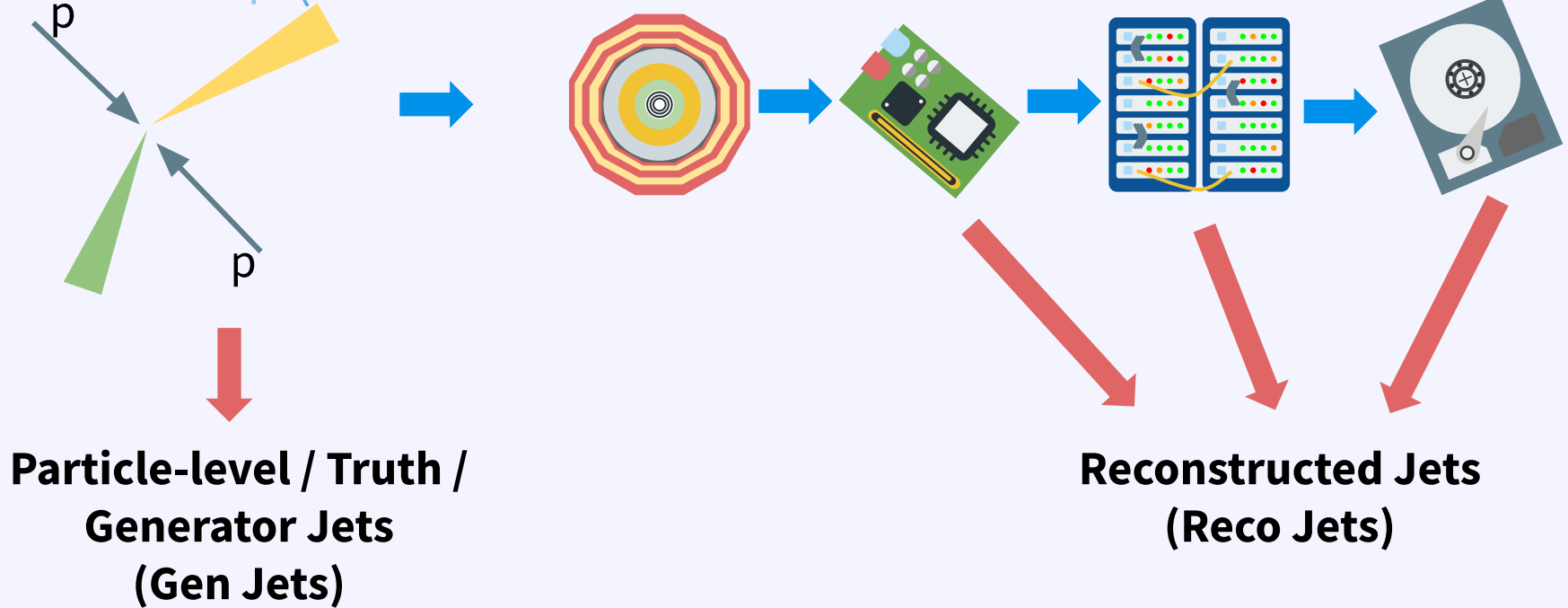


Jet Energy Correction (JEC)

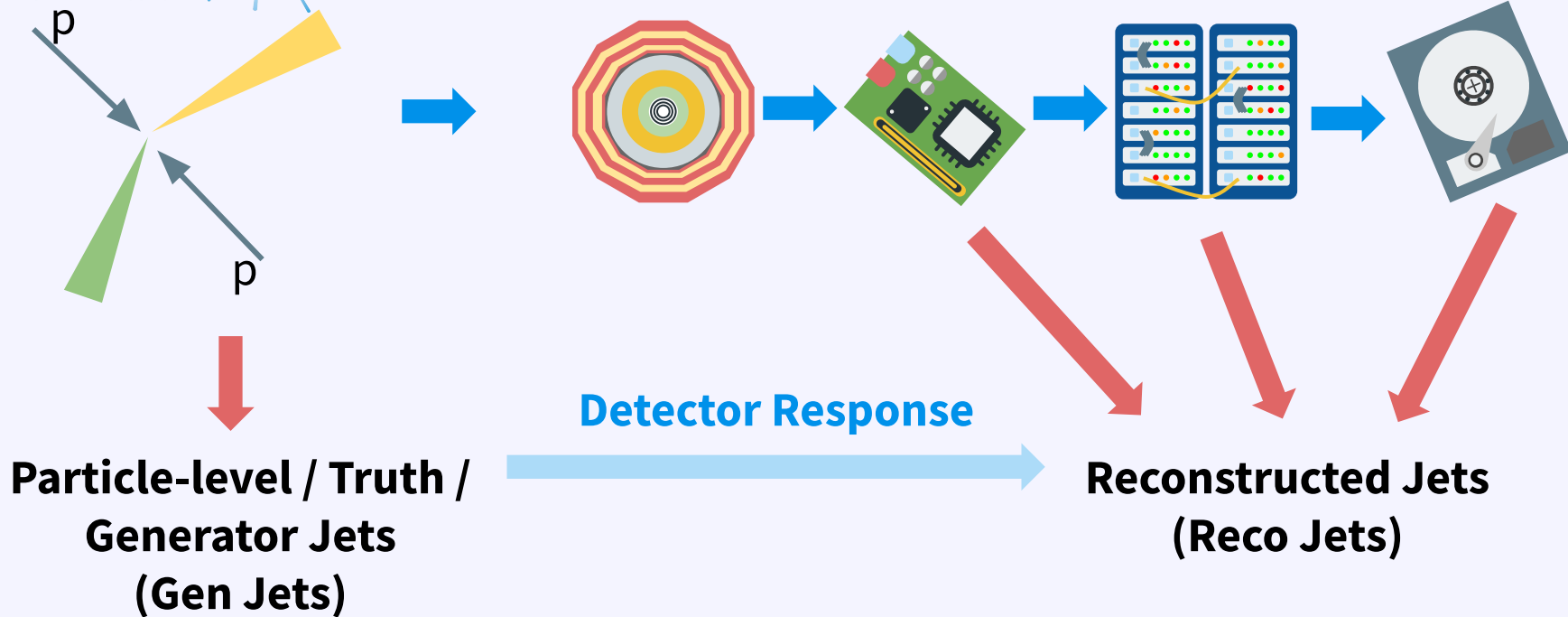


SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

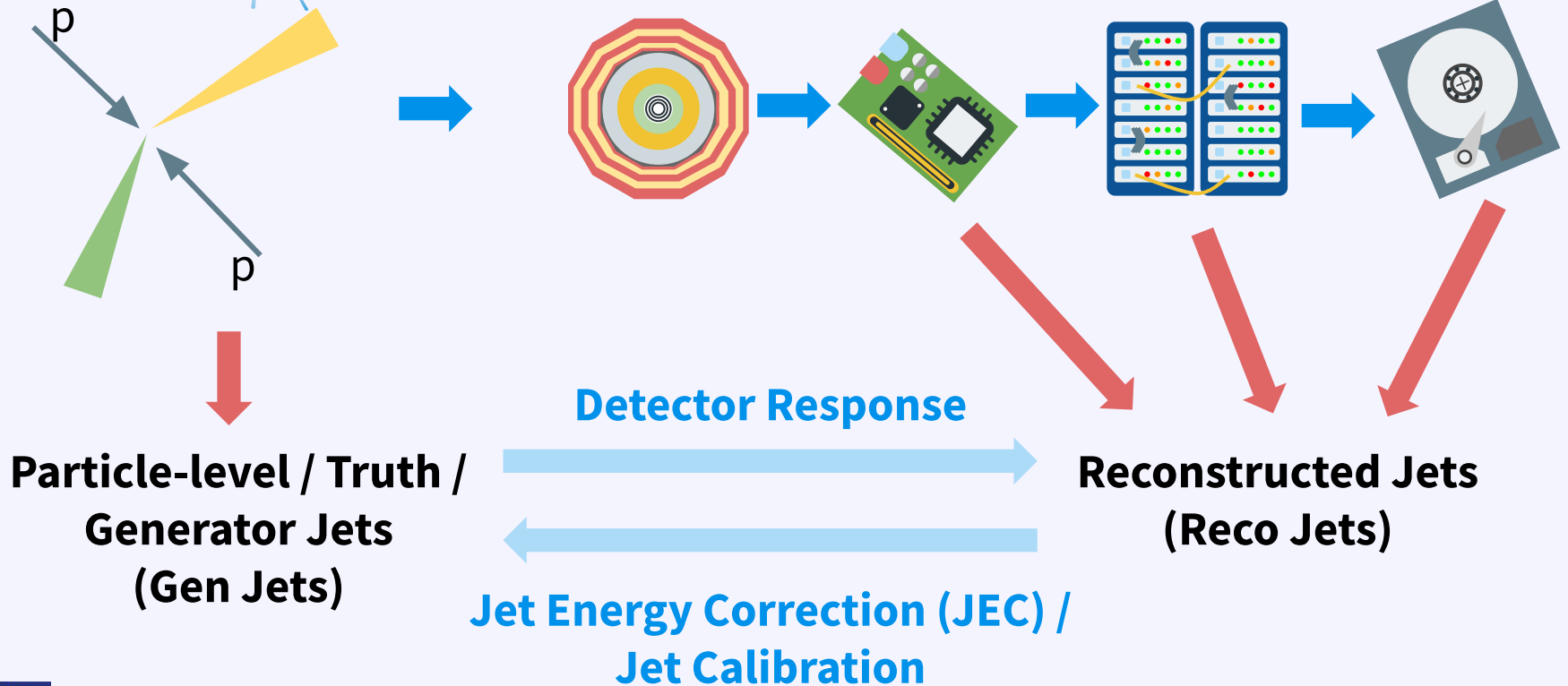
Jet Energy Correction (JEC)



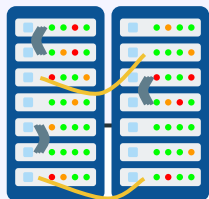
Jet Energy Correction (JEC)



Jet Energy Correction (JEC)

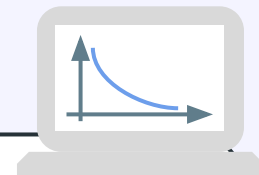


JEC for Scouting (+Online) Jet



HLT/Scouting 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)

JEC for Scouting (+Online) Jet

transfer low systematic uncertainties from offline to online reconstruction



HLT/Scouting 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)

JEC for Scouting (+Online) Jet

transfer low systematic uncertainties from offline to online reconstruction

HLT/Scouting Jet

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

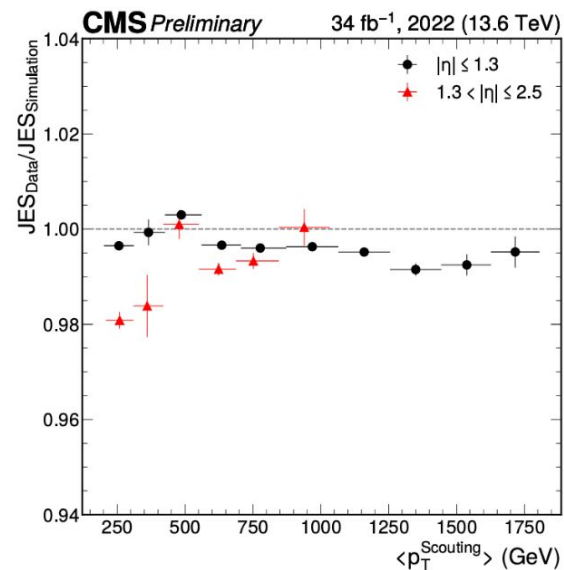
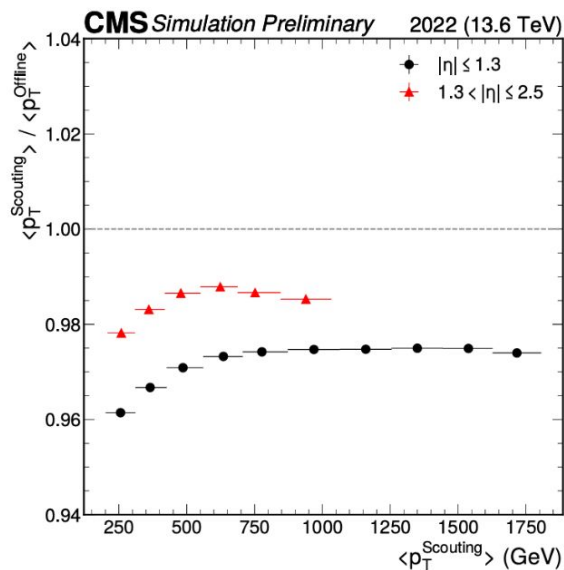
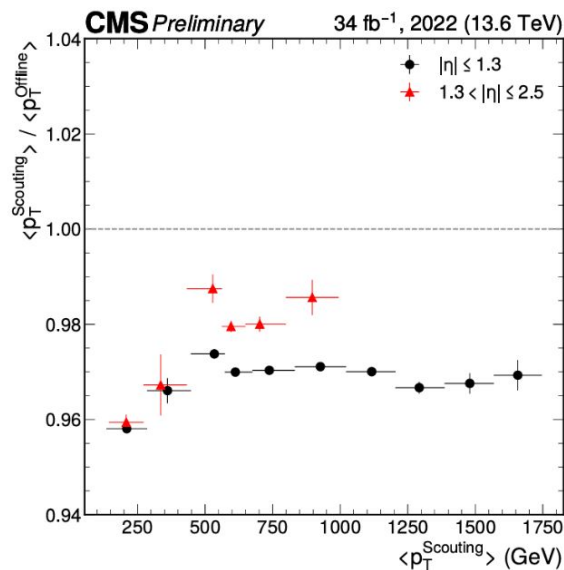
Offline jets

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

improve offline calibration with abundant HLT jet statistics



Studies by Adelina Lintuluoto



Other activities



Participation



Past

- PAPU Fall Seminar (22 November 2022): **lightning talk!**
- CMS Week December 2022 (5 - 9 December 2022)
- Spåtid 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)



Participation

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)



CMS Young Scientist Committee (CMS YSC)

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



Midsummer School in QCD 2024

24 June – 6 July 2024

Saariselkä, Finland



[indico](https://indico.cern.ch/event/qcd2024)

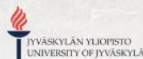
Register by
31 January 2024



indico.cern.ch/event/qcd2024

Organizing Committee

Christophe Royon, Henning Kirschenmann, Kati Lassila-Perini
Kenneth Österberg, Mikko Voutilainen, Tuomas Lappi



THE UNIVERSITY OF
KANSAS

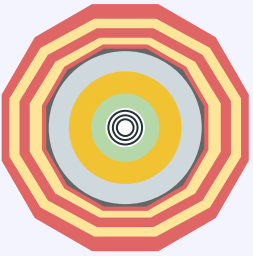
HL-LHC Upgrade: L1 Scouting



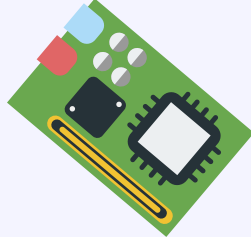
Scouting Stream
= only PF particles
reconstructed at HLT

Standard Stream
= all raw detector
readout

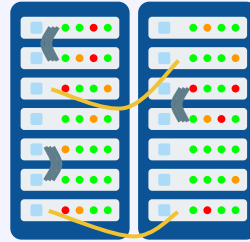
L1 Scouting (40 MHz)
= L1 Primitives



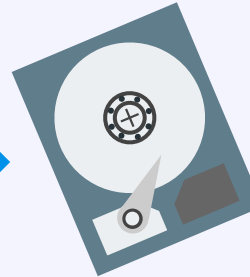
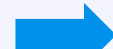
CMS



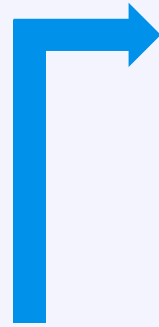
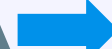
L1 Trigger



HLT



Storage



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)
- JEC for scouting jets
 - offline and online jets can support calibration of each other
- Boosted H → bb using scouting data
 - scouting data contains 20% more H → bb events overall, particular gain at low momentum region
 - studies on tagger performance on scouting jets
- Other activities
 - learning a lot, meeting a lot of new people, and a lot of memories

