III. Physikalisches
Institut A

RWTH AACHEN UNIVERSITY

Experimental Techniques in Particle Physics (WS 2020/2021)

# From Detector Data to Physics Results

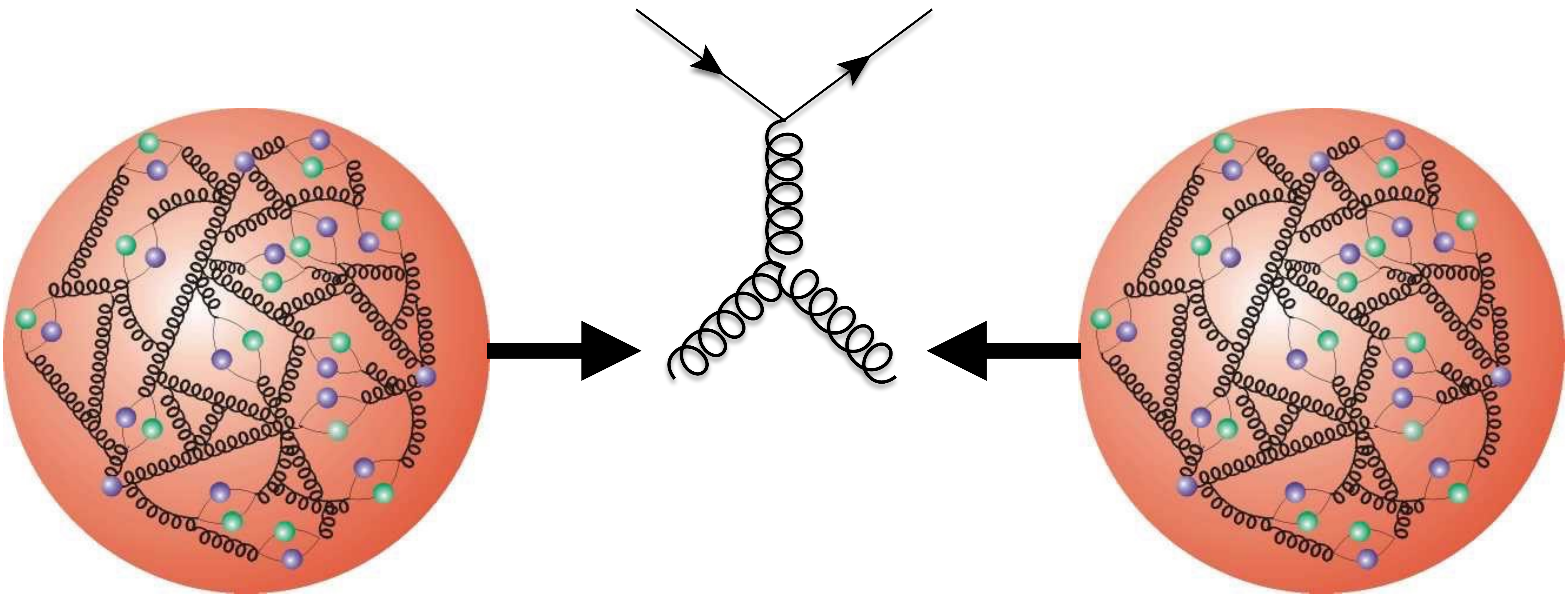
Prof. Alexander Schmidt

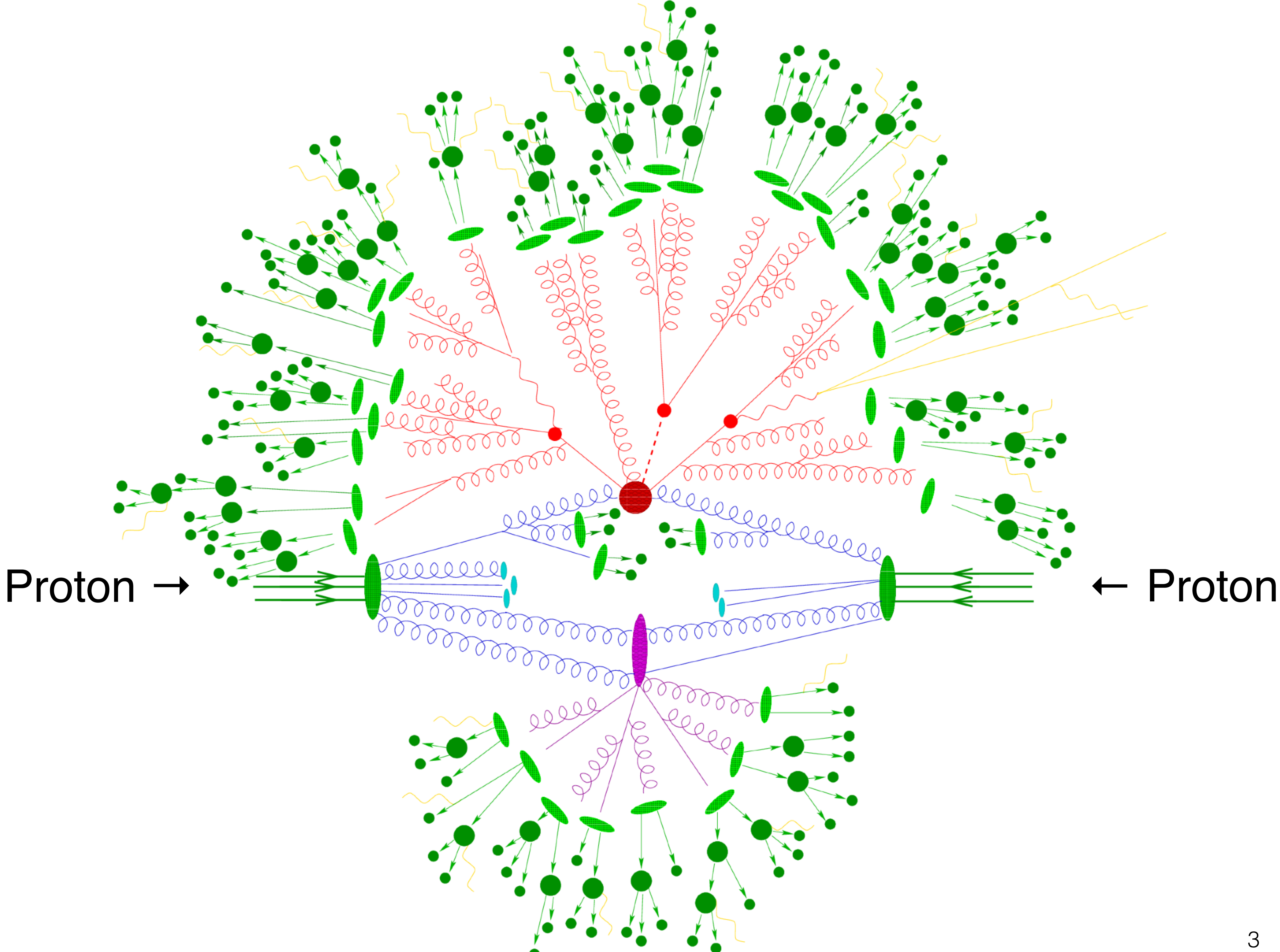
26.01.2021

# basics

- particles (e.g. top quarks) are produced in LHC collision
- most of them decay immediately (including top, Higgs,…)

- only the **final state products** after the decay chain enter the detector

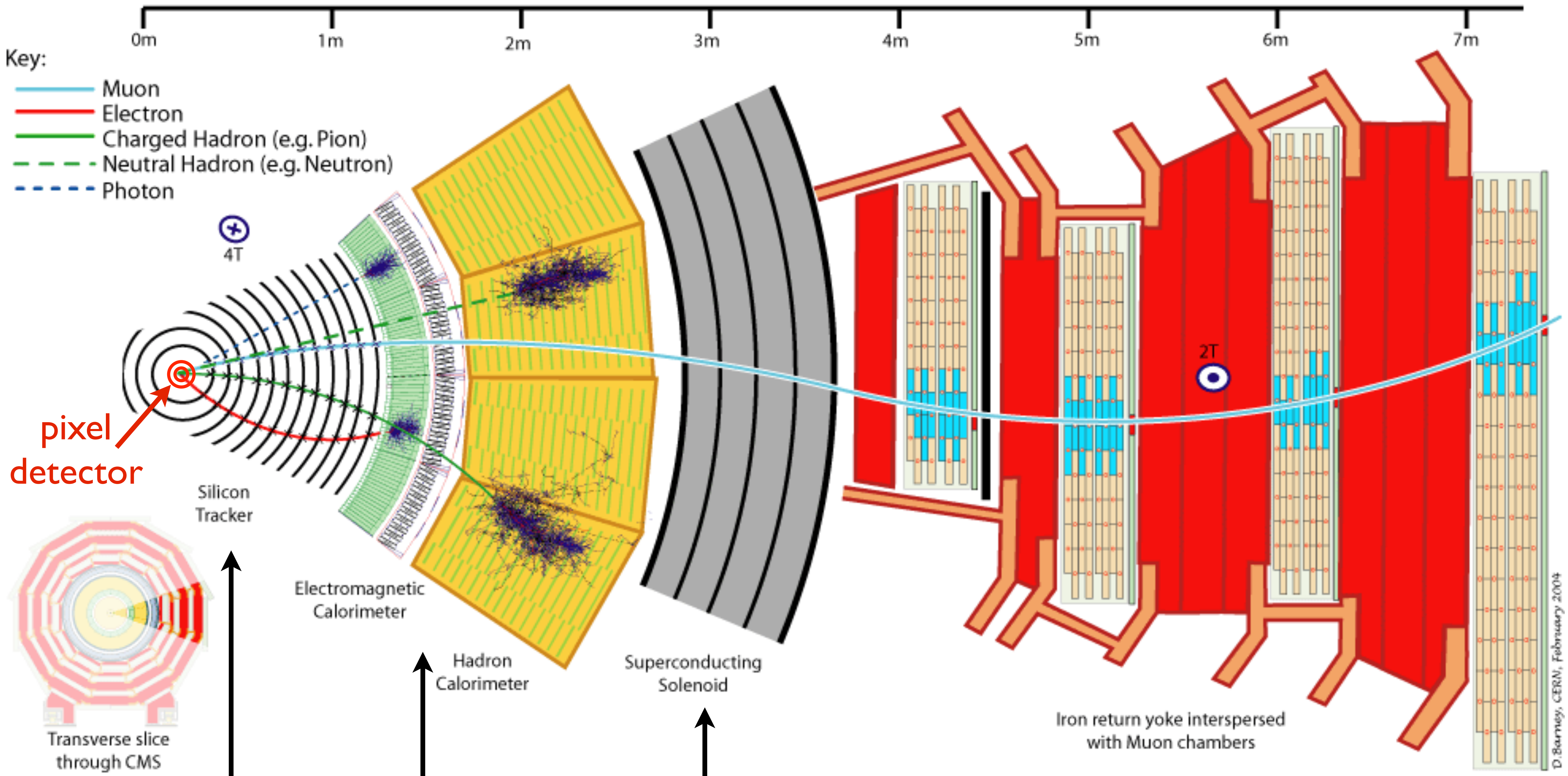



Proton →
← Proton

# basics

- particles interact with the detector and deposit energy (see details in lecture and GEANT course)
- the detector devices give a signal according to the deposited energy (see lecture)
- the first step is to TRIGGER interesting events (see FPGA course)
- based on the detector signals, events are **reconstructed** (e.g. silicon detector hits, tracks, jets)

# basics




tracks of charged particles

jets and electrons

curvature for momentum measurement

muons, e.g. from semileptonic B decays

# reconstruction

Raw data

GEANT4
detector simulation

**Detector reconstruction**
Tracker, ECAL, HCAL, Muon

**Detector calibration**
Before installation, every ~day,
end of year

**Object reconstruction**
Electrons, Muons, Taus, Hadrons,
Jets, Missing $E_T$, Jet taggers

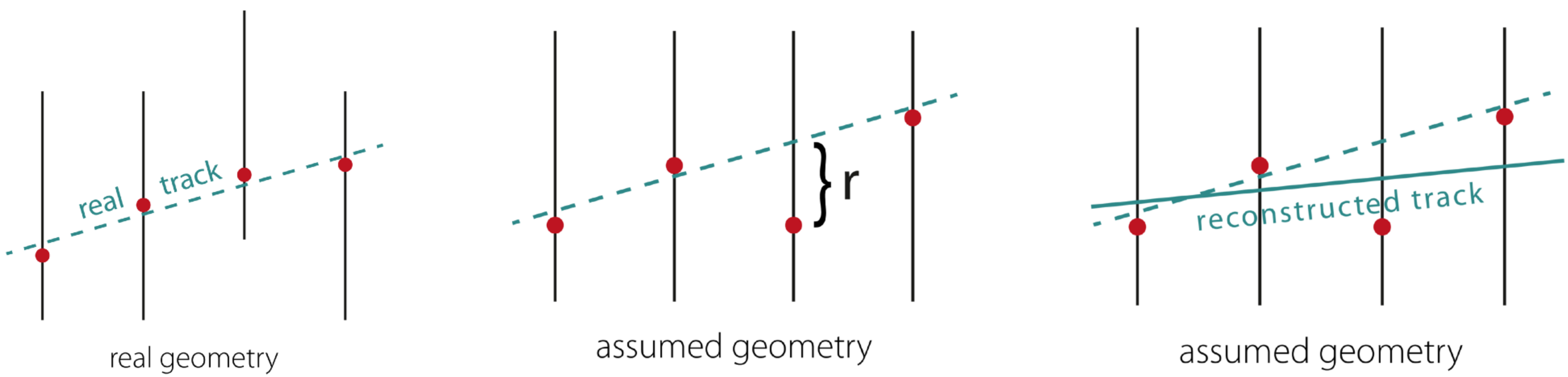
**Object calibration**
end of year

**Physics analysis**

**Physics result**

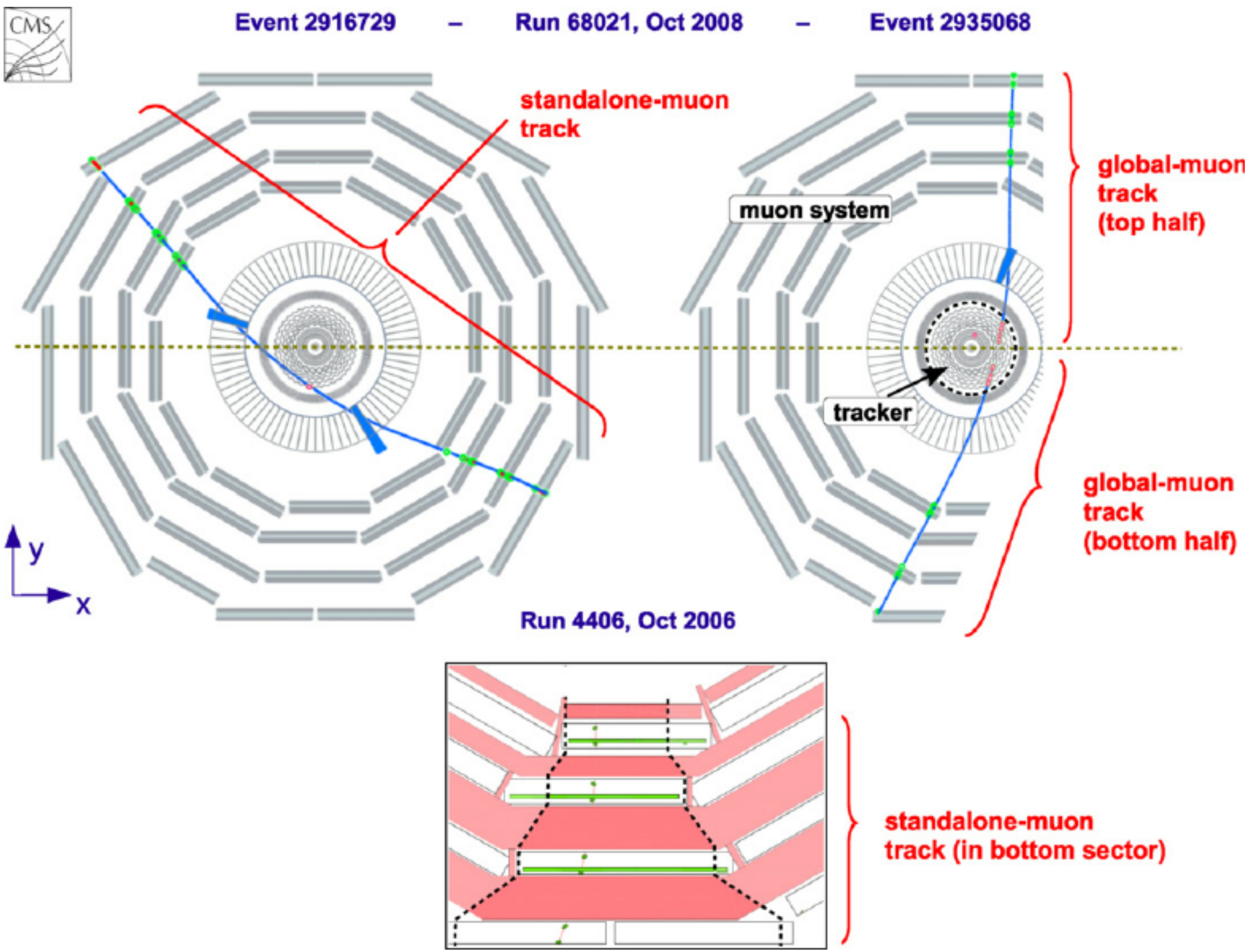
# track reconstruction and alignment

- we already discussed reconstruction of silicon tracker hits from charge signals
- we already discussed track reconstruction with Kalman filters
- next step: tracker alignment
- tracker detector modules can get slightly misaligned (e.g. from switching on-off the magnetic field, temperature dependent movements)



# track alignment

- CMS uses cosmic muon tracks, as well as collision tracks for alignment
- cosmic muons were the only possibility before the start of LHC



# track alignment

- standard alignment algorithms follow the least squares approach and minimise the sum of squares of normalised residuals from many tracks

$$\chi^2(\mathbf{p},\mathbf{q}) = \sum_{j}^{\text{tracks}} \sum_{i}^{\text{measurements}} \left( \frac{m_{ij} - f_{ij}\left(\mathbf{p},\mathbf{q}_j\right)}{\sigma_{ij}} \right)^2$$

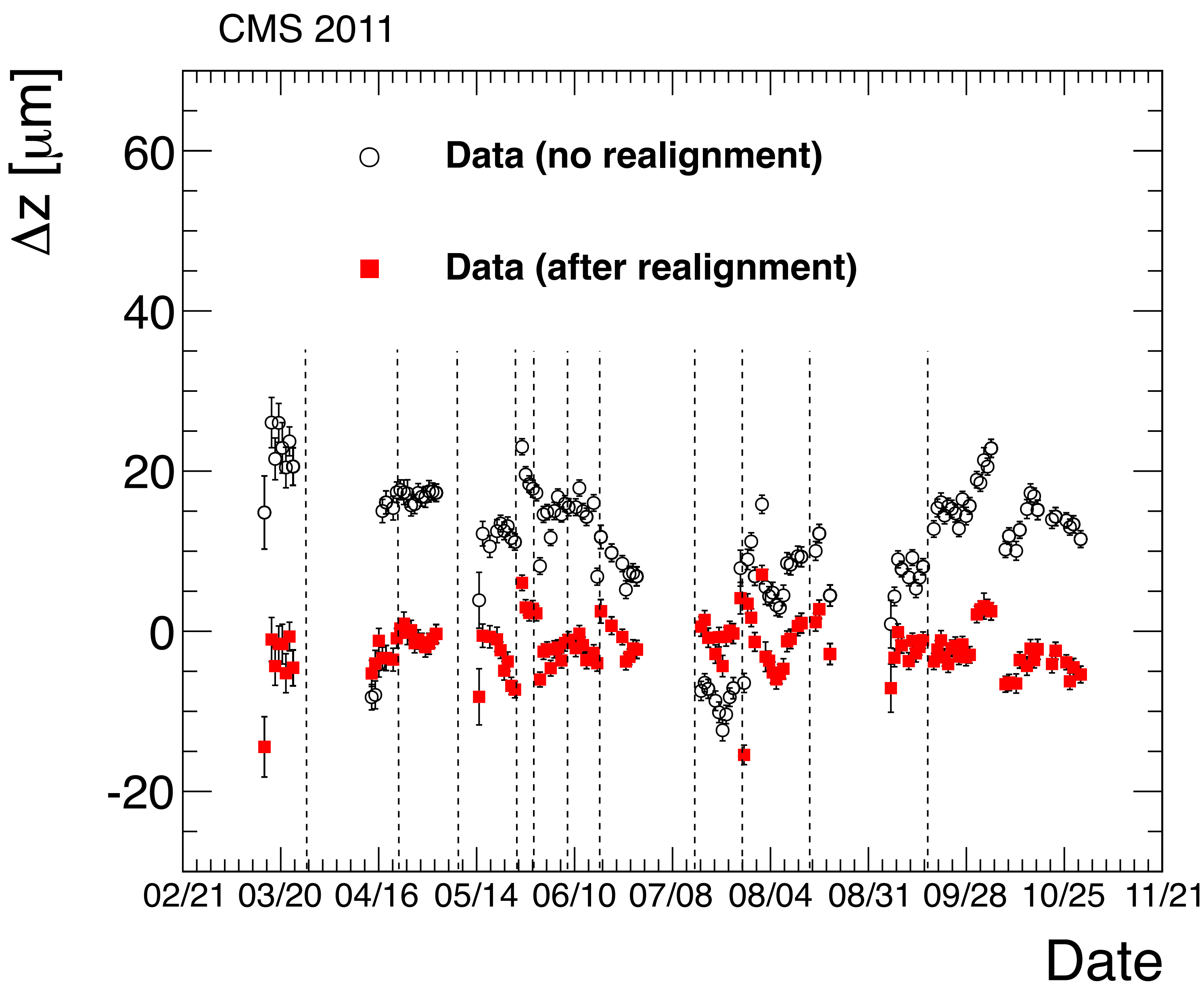
- $m_{ij}$ is the measured hit position
- $f_{ij}$ is the trajectory prediction at the position of the measurement
- $\sigma_{ij}$ is the hit position uncertainty
- **p** is the set of tracker geometry parameters
- $\mathbf{q}_j$ is the set of track parameters of track j

- $\chi^2$ is minimized, so that the updated tracker geometry **p** is found

- with 200 000 alignment parameters this is an excessively large system of equations that is not trivial to solve
- dedicated numerical strategies and softwares are used

# track alignment

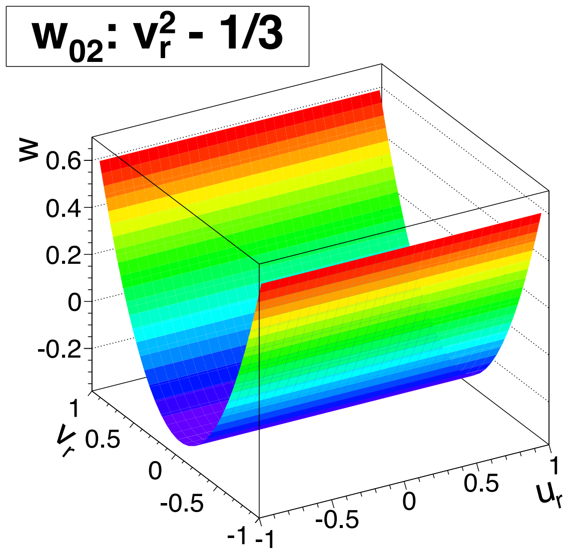
- relative longitudinal displacement of the two barrel pixel half-shells as function of time

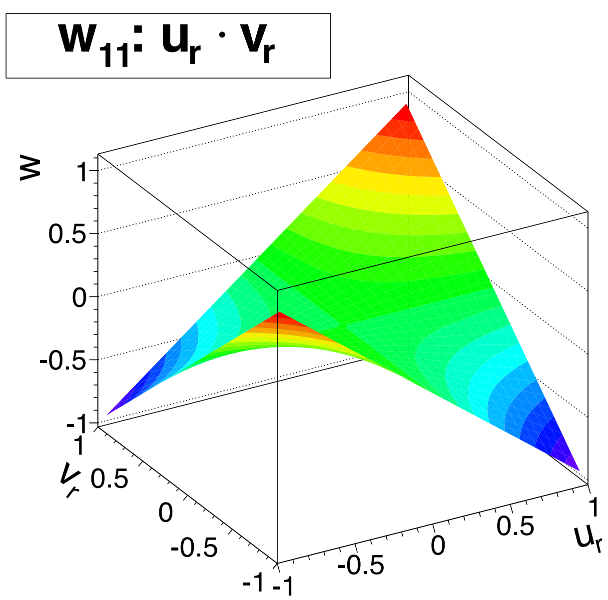


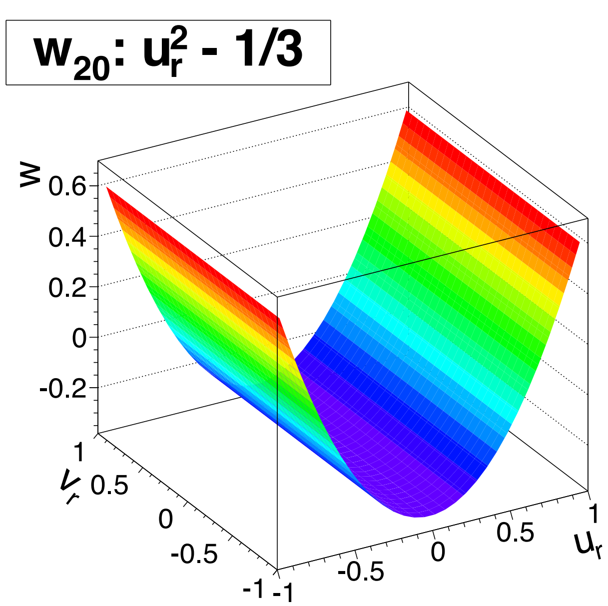

➡time-dependent alignment constants necessary

# tracker alignment

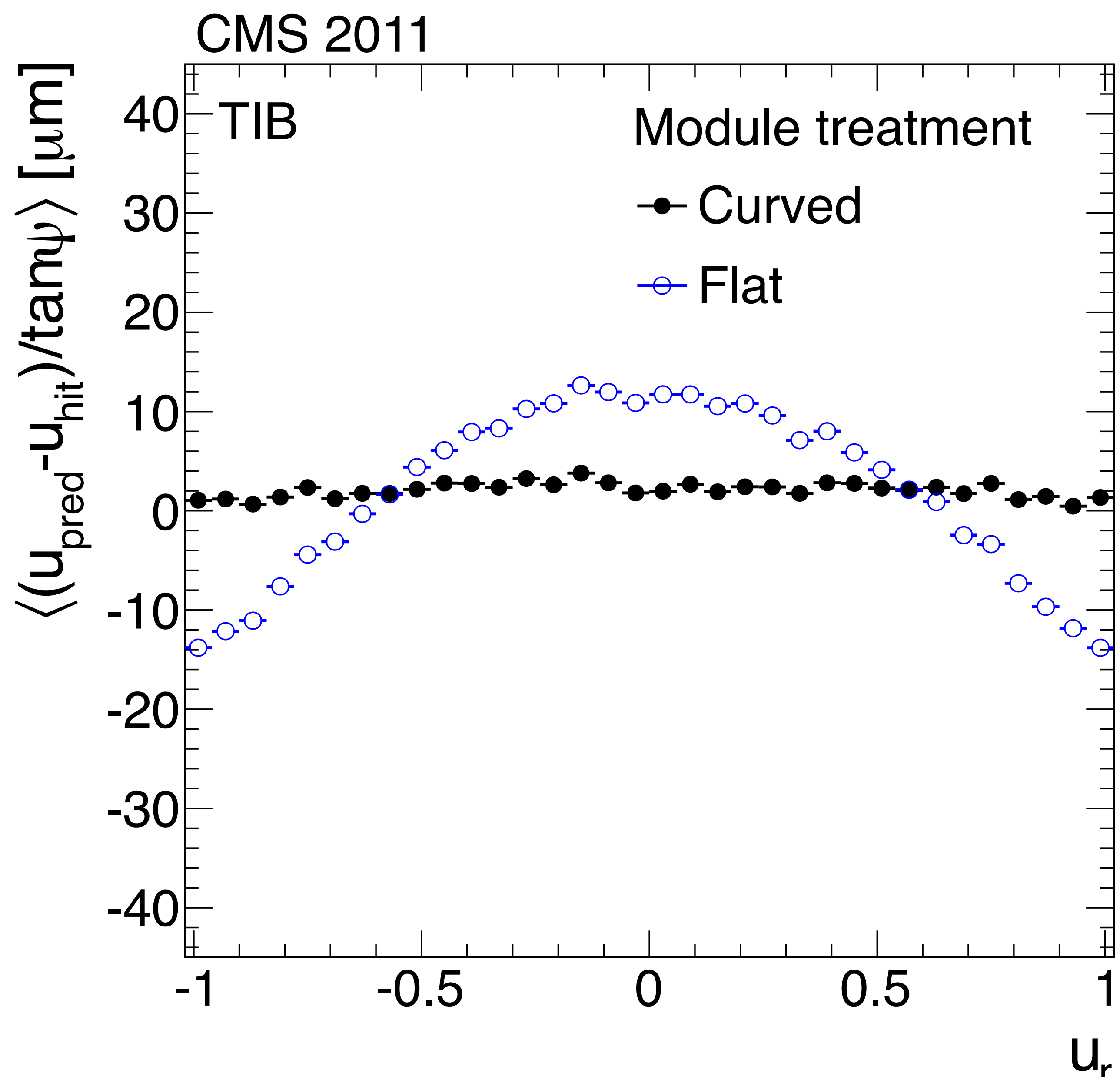
- deformations of the flat sensor modules are observed as well
- implemented as slight parabolic bending



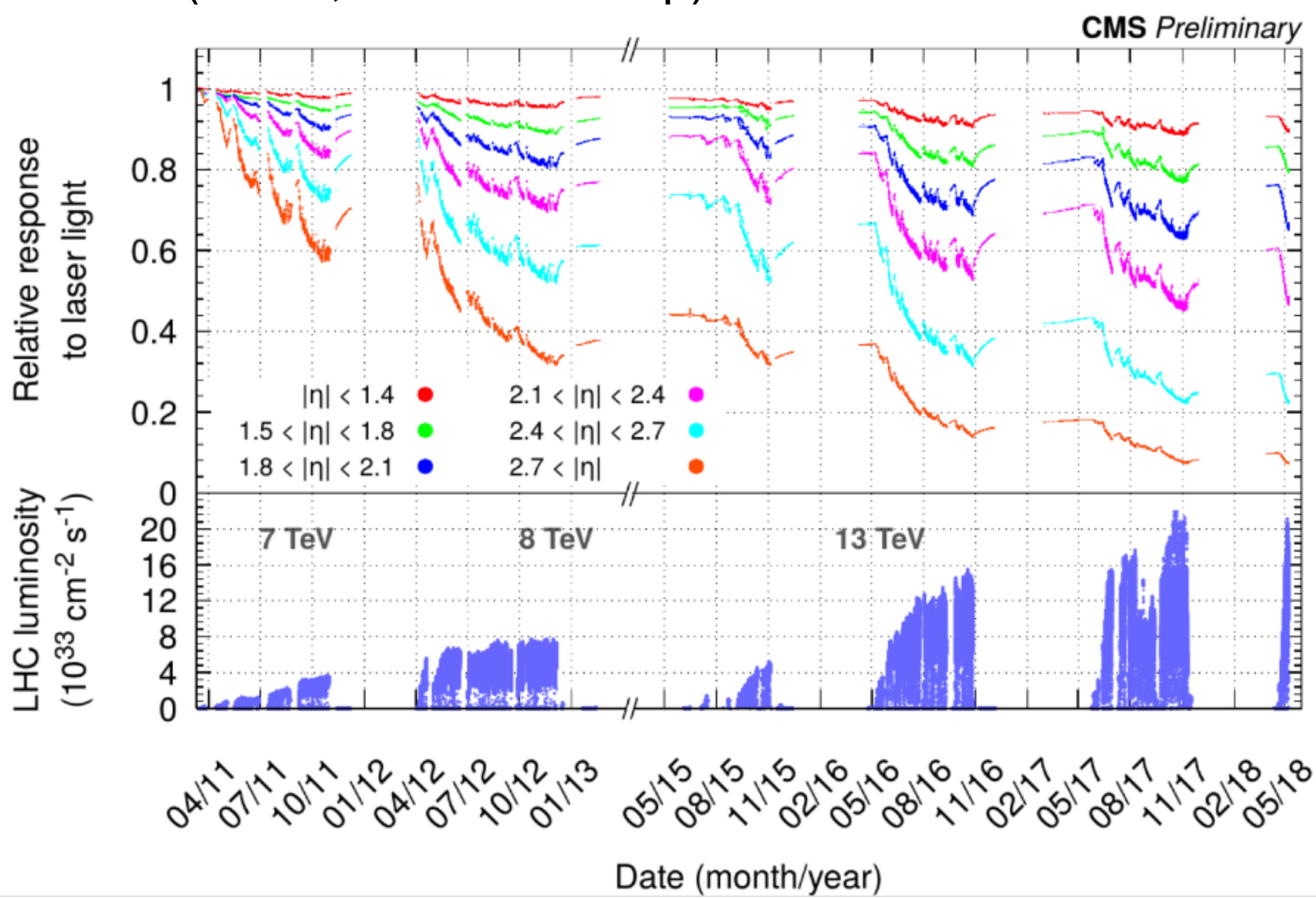




- **average hit residuals with/without curvature:**



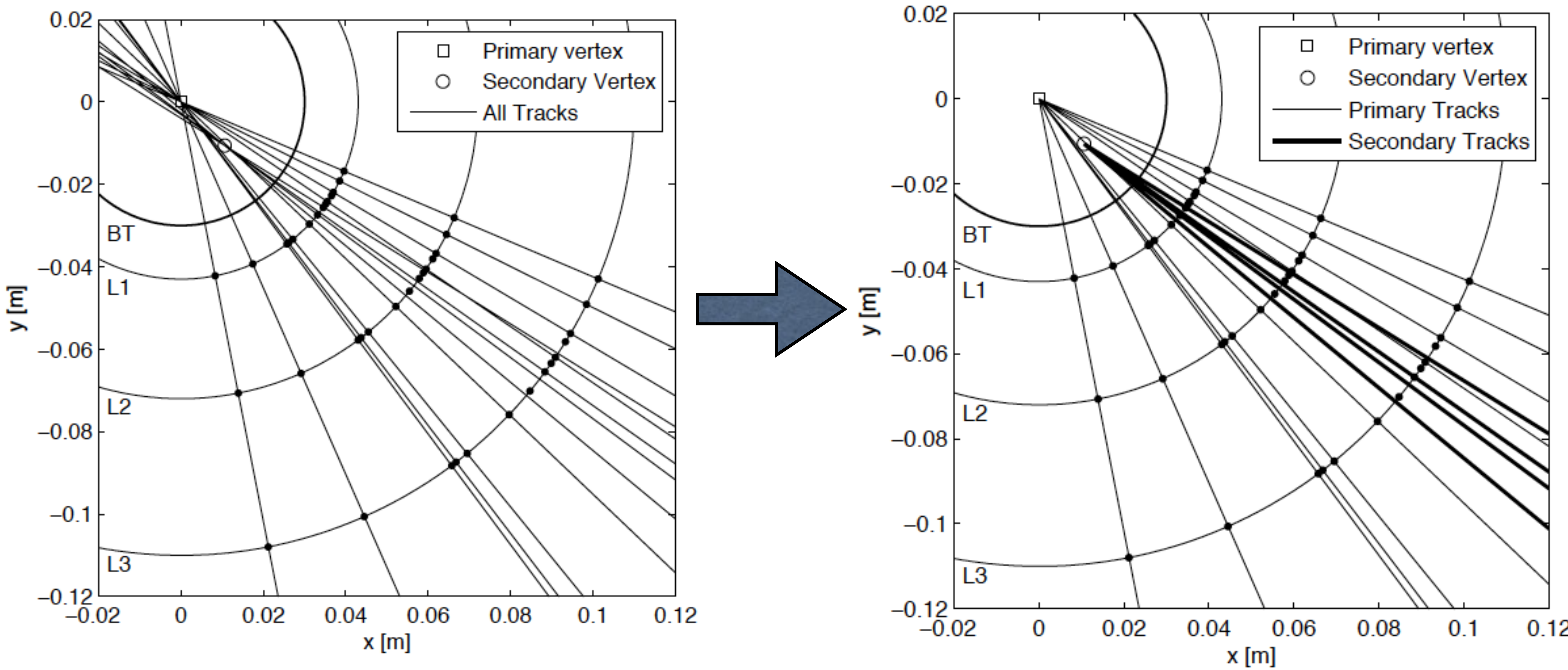
# calorimeter calibration

- Radiation damage to scintillators (transparency loss) corrected by laser measurement (ECAL, HCAL-endcap)




- Energy scale for ECAL also from data through Z→ee, $\pi_0 \to \gamma\gamma$, E/p($e^-$)

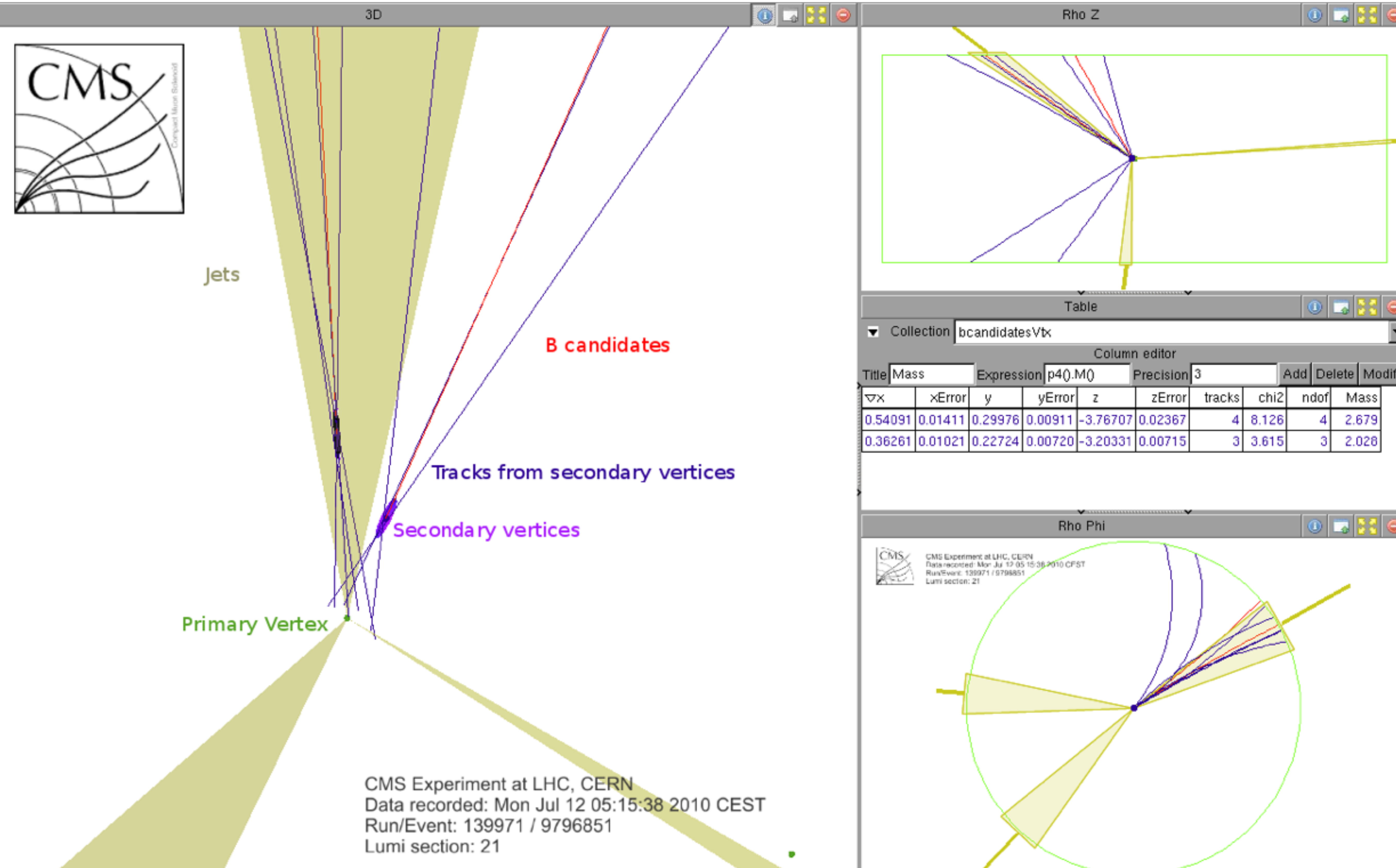
# vertex finding




## again a problem of “classification” or pattern recognition

- find points in space where tracks originate (and associated uncertainties)
- example: proton collisions, decays of long-lived particles

# secondary vertex



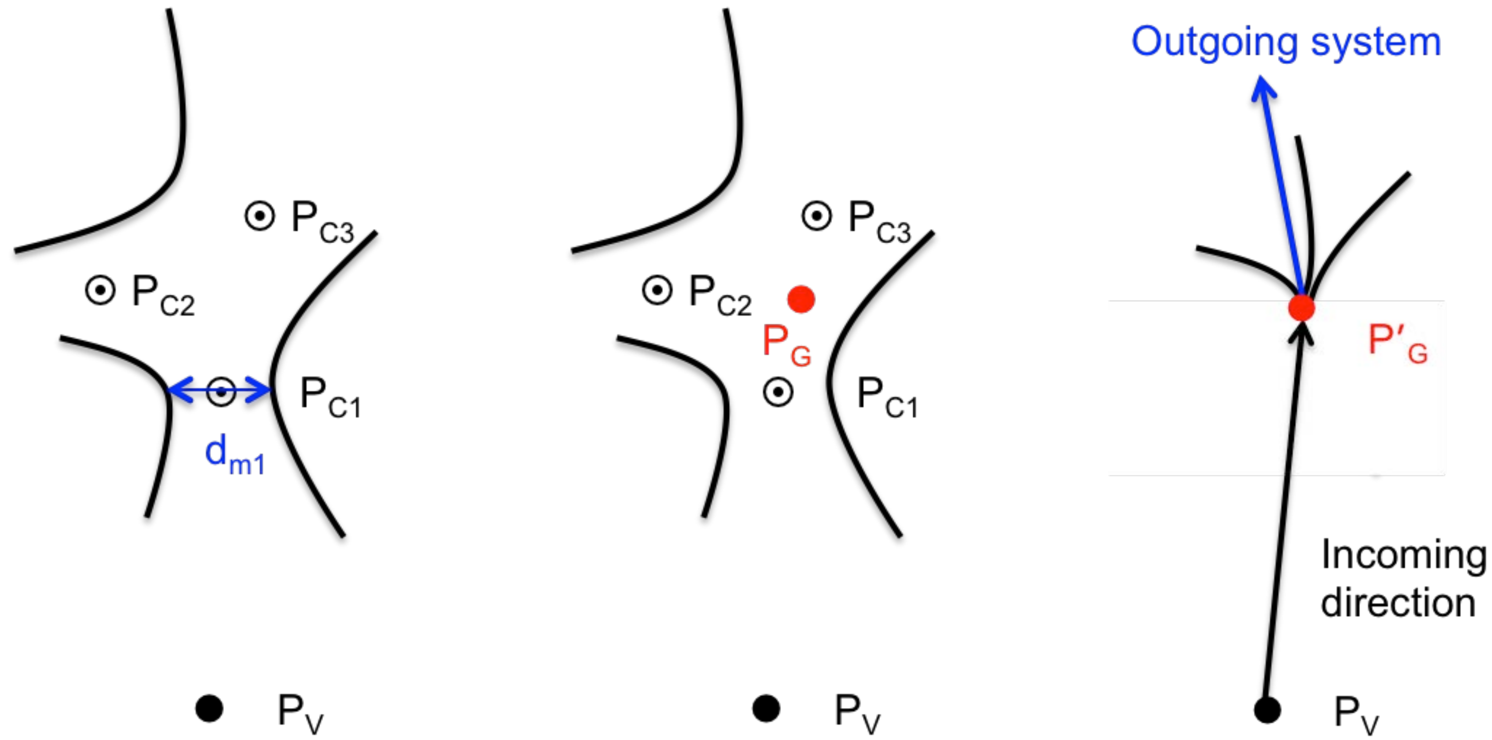

| ▽x | xError | y | yError | z | zError | tracks | chi2 | ndof | Mass |
|---|---|---|---|---|---|---|---|---|---|
| 0.54091 | 0.01411 | 0.29976 | 0.00911 | -3.76707 | 0.02367 | 4 | 8.126 | 4 | 2.679 |
| 0.36261 | 0.01021 | 0.22724 | 0.00720 | -3.20331 | 0.00715 | 3 | 3.615 | 3 | 2.028 |

# secondary vertex reconstruction

- secondary vertices arise from displaced particle decays
- example: b-hadrons (topic of b-jet identification, see later)

- also interaction of particles with detector material (photon conversions, bremsstrahlung or **nuclear interactions**) can give rise to displaced processes

- nuclear interaction identification in the inner tracker, example with 3 tracks:

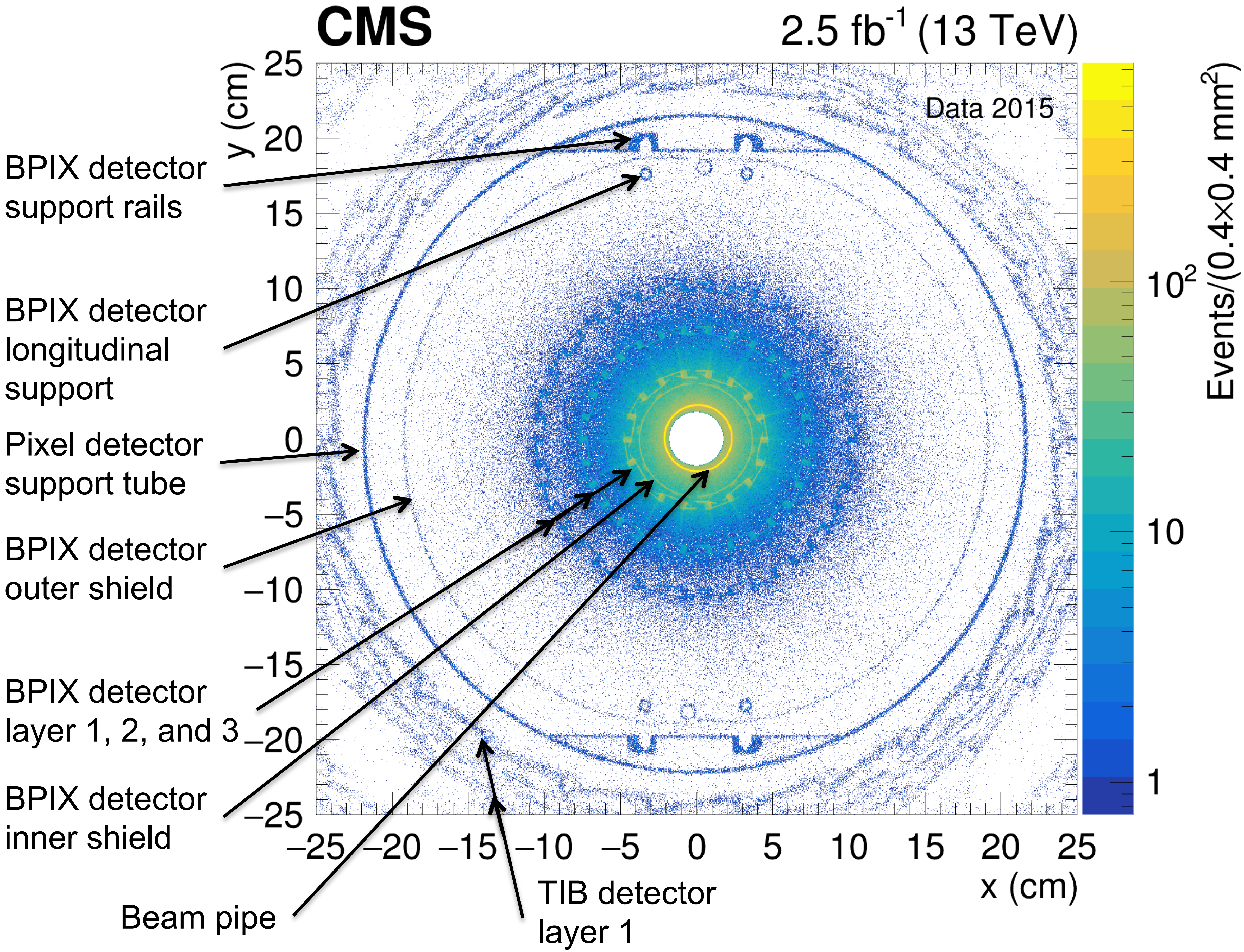



calculate pair-wise distance of closest approach $d_m$ . The center of the segment $P_C$ is a vertex position estimate

calculate barycenter $P_G$

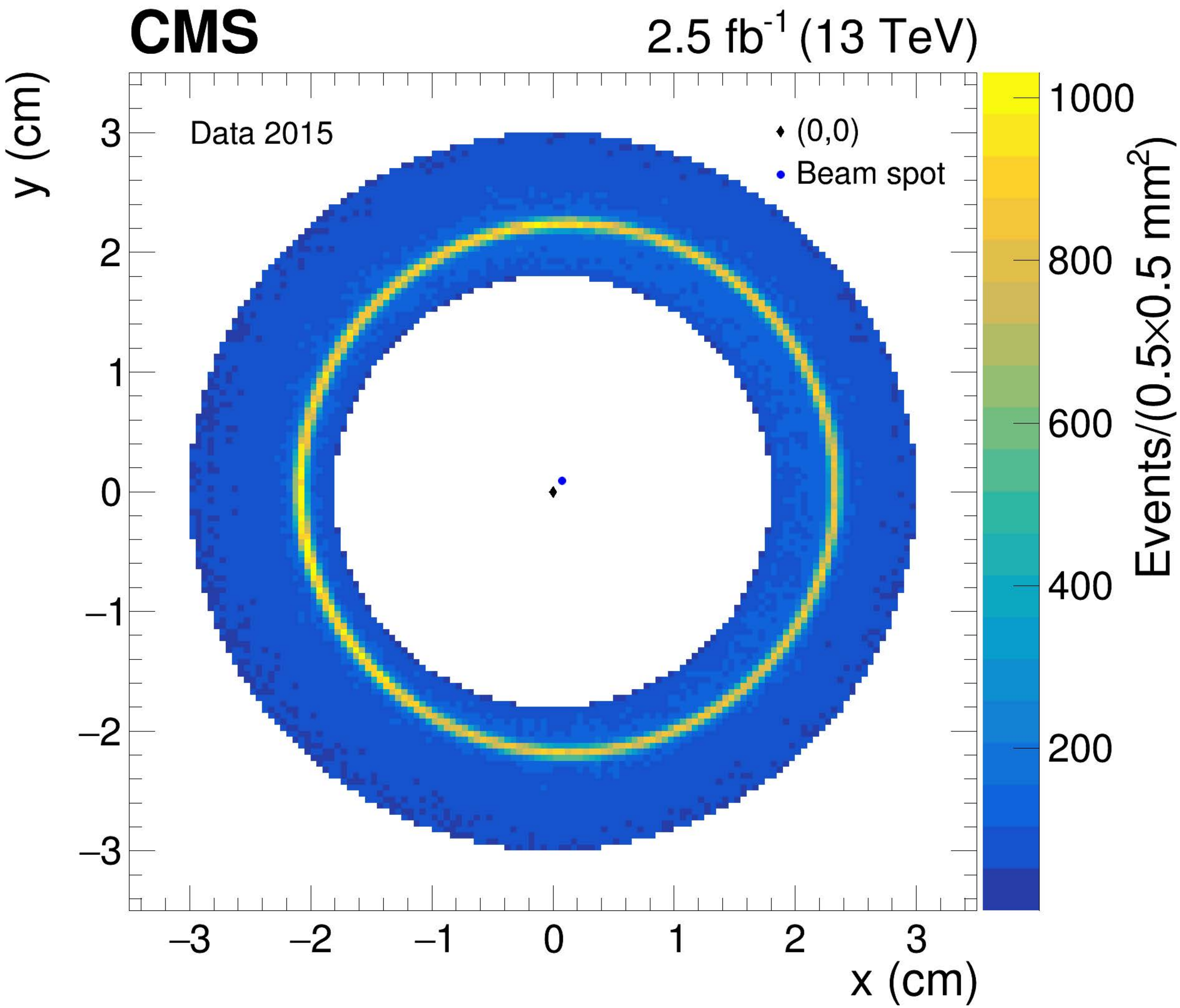
refit tracks to originate from this $P_G$ vertex

# radiography from secondary vertices



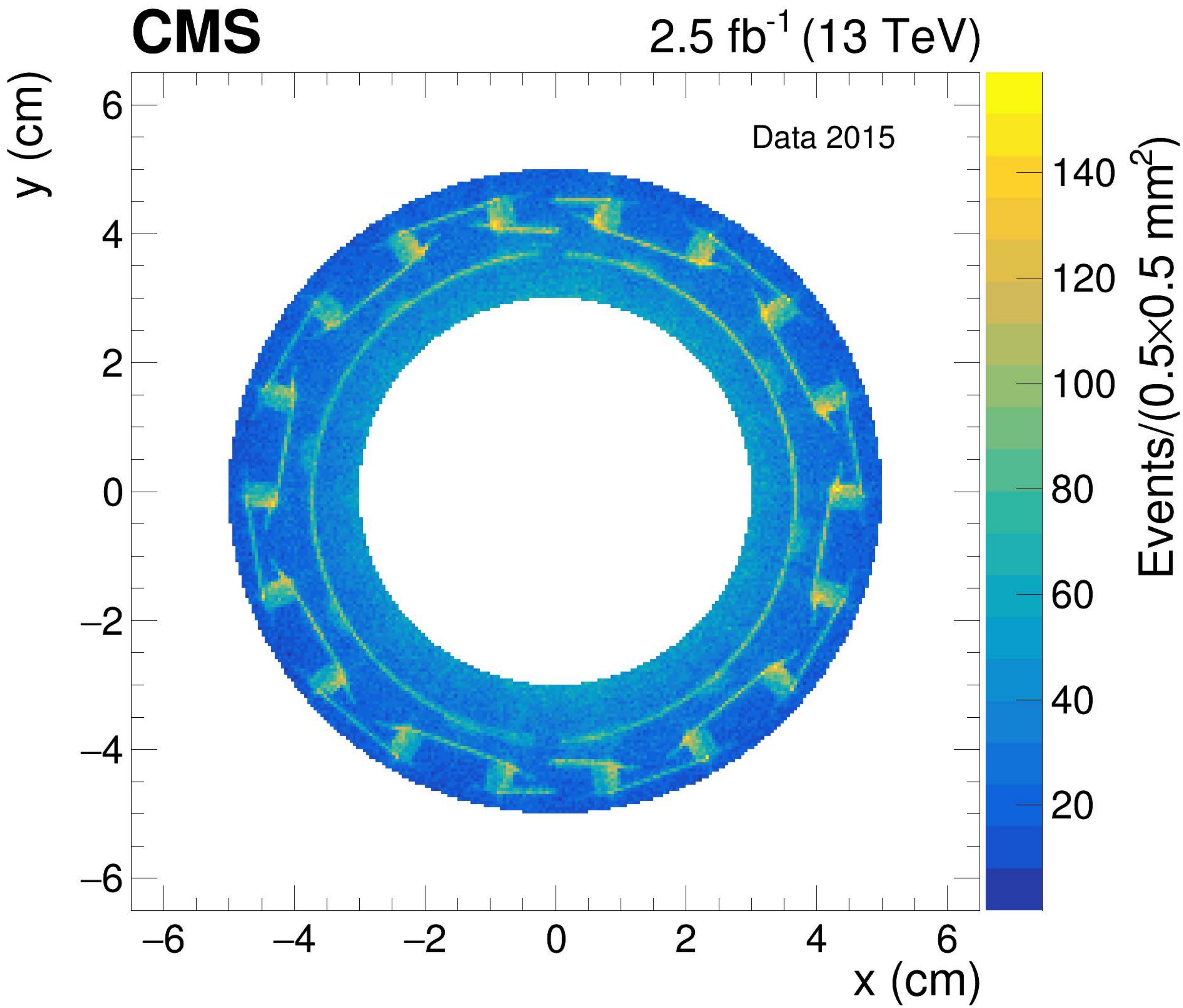
# radiography from secondary vertices

- zoom to the beam pipe region
- the pipe is displaced by 1mm with respect to the origin (0,0)



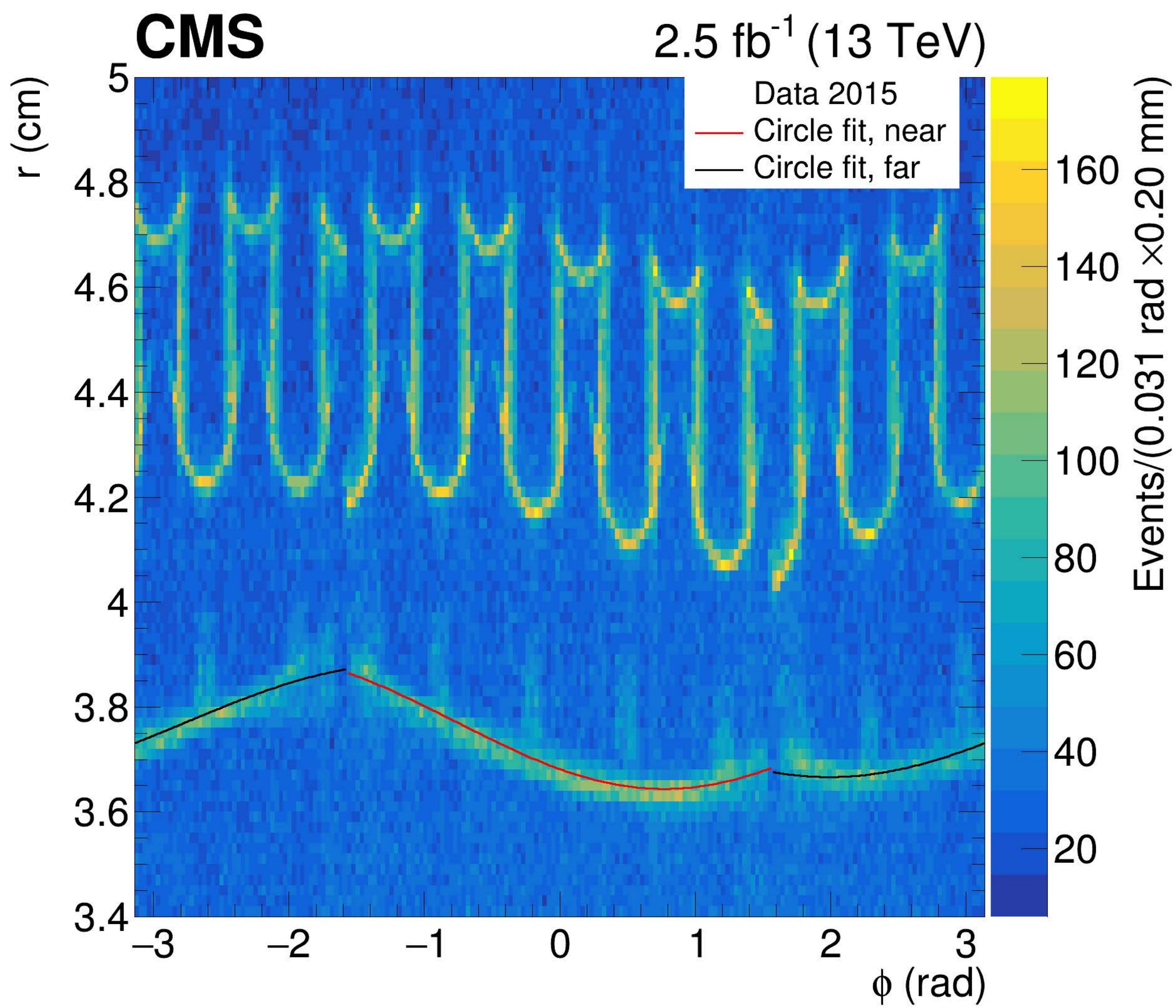
# radiography from secondary vertices

- inner shield
- inner pixel layer
- pixel sensors of inner layer
- cooling pipes (sensors are mounted on pipes)

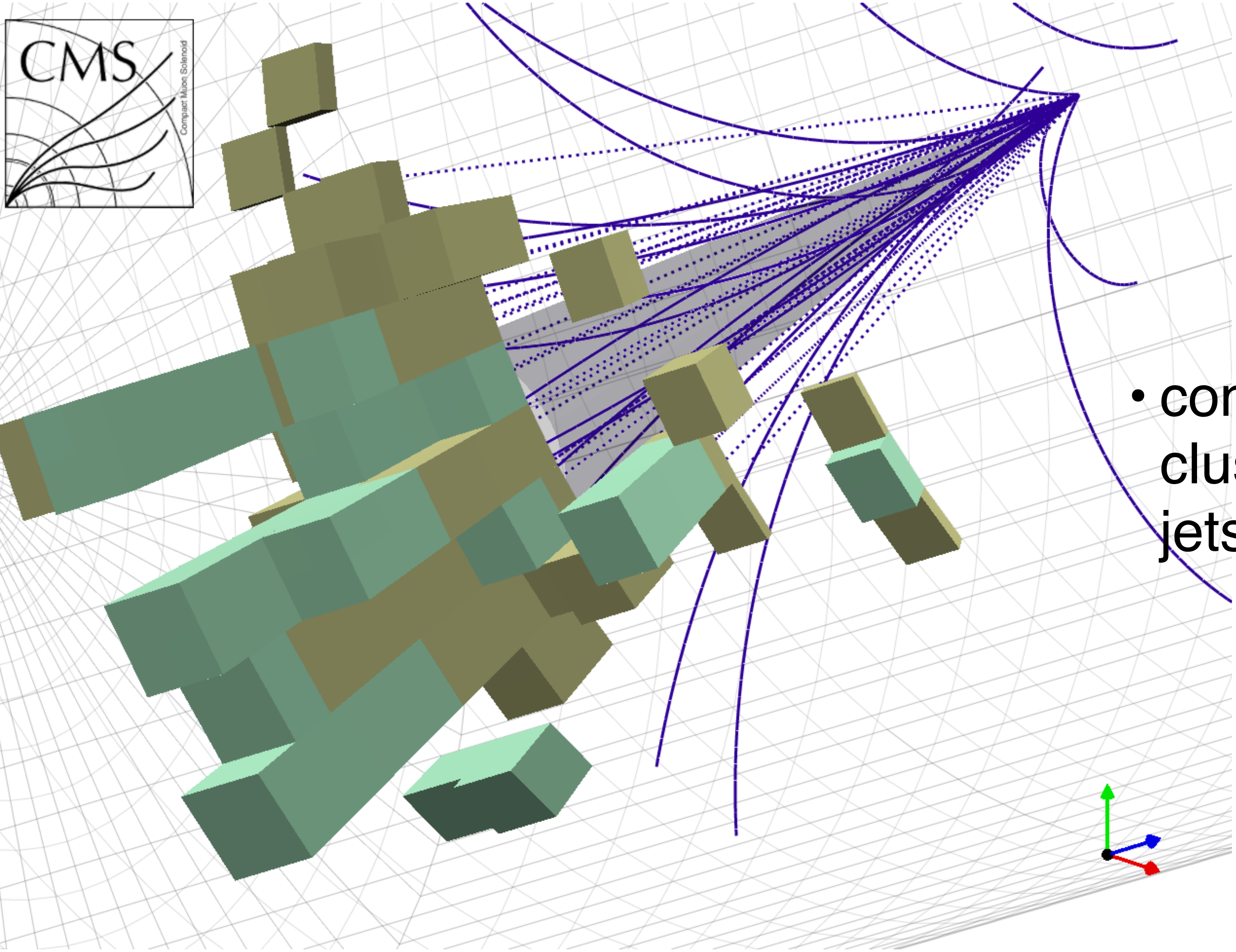


# radiography from secondary vertices

- same as previous slide in r-phi plane



# jets




- combine calorimeter clusters and tracks into jets ???

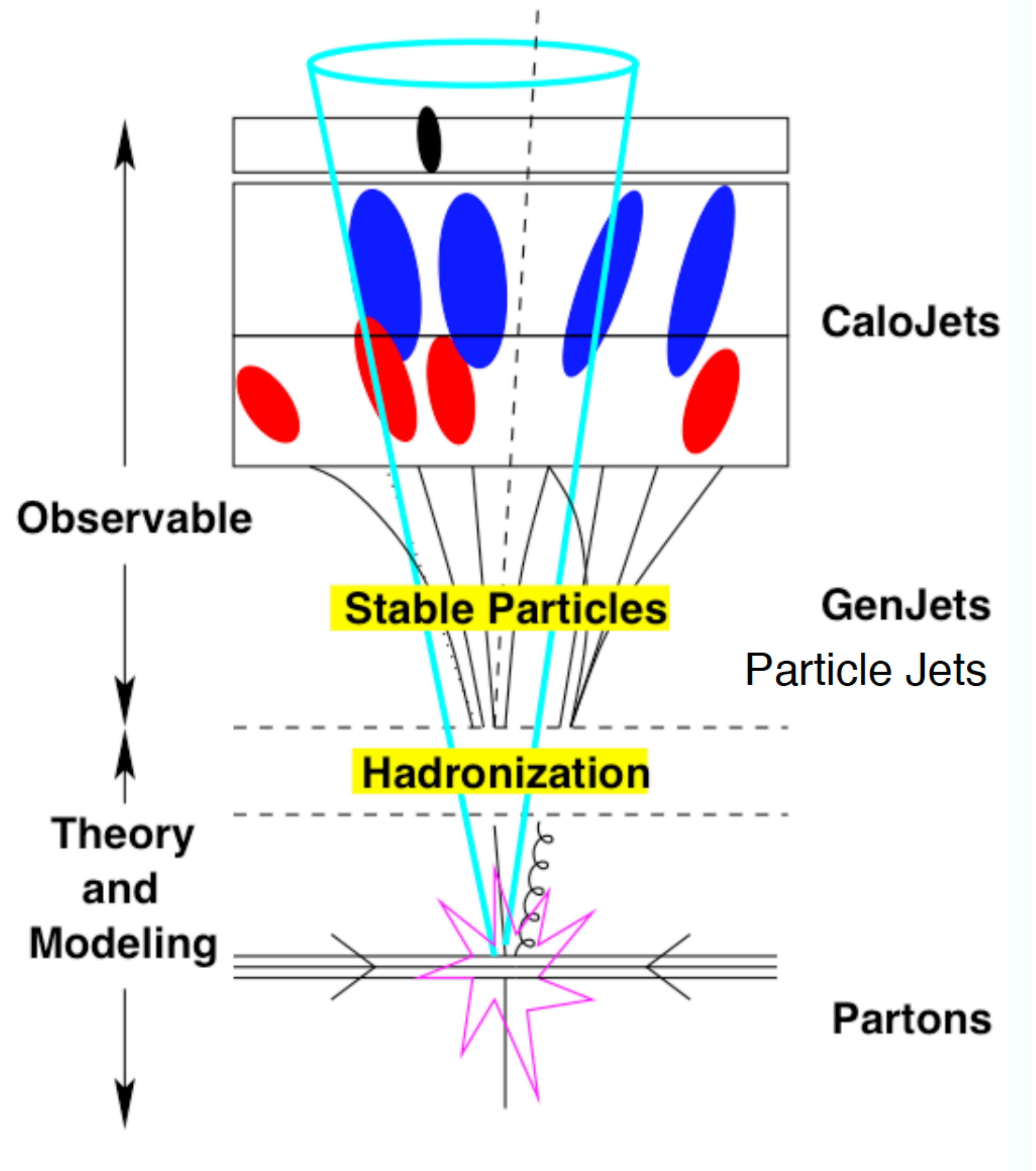
# jets

Jets are collimated bunches of stable hadrons, originating from partons (quarks and gluons) after fragmentation/ parton showering and hadronization

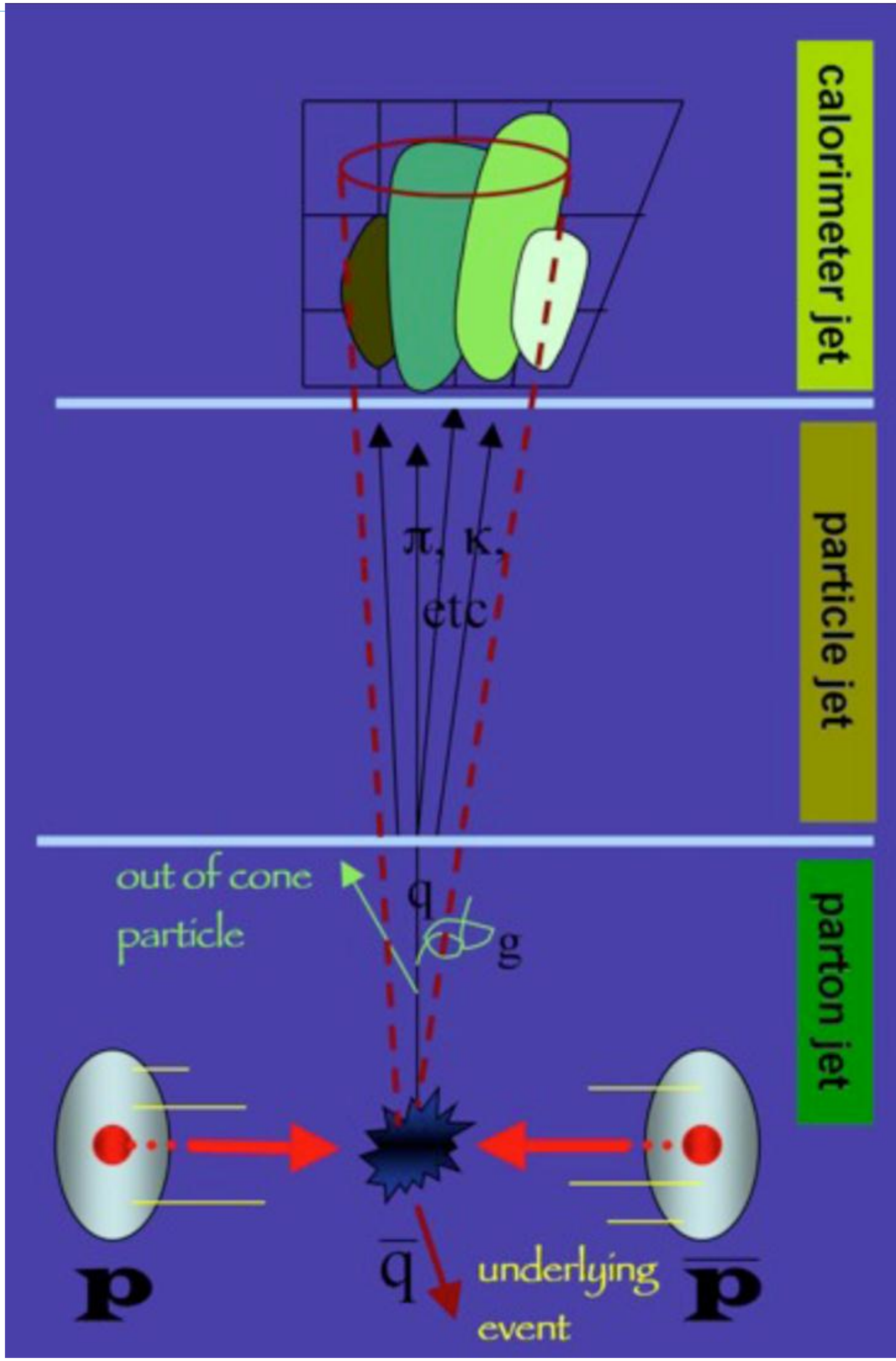
Jet reconstruction is the approximate attempt to reverse-engineer the quantum mechanical process of fragmentation and hadronization

Note! this is not a unique procedure → several different approaches

Jets are the observable objects to relate experimental observations to theory predictions formulated in terms of quarks and gluons



# types of jets




## Calorimeter jets:
(extracted from calo clusters)
- Understanding of detector response
- Knowledge of dead material
- Correct signal calibration
(- potentially include tracks)

From measured energy to particle energy
Compensate energy loss due to neutrinos, nuclear interactions ...

## Hadron/particle jets:
(might include electrons, muons, ...)

- Hadronization
- Fragmentation
- Parton shower
- Particle decays

From particle energy to original „parton energy“
Compensate hadronization, energy inside/outside jet cone

## Parton jets:
(quarks and gluons)

- Proton-proton interactions
- Initial and final state radiation
- Underlying event

**Needs calibration!**

# jet reconstruction

**jet clustering algorithms:**

- today sequential algorithms are mainly used
- define measure of distance between particles in an event

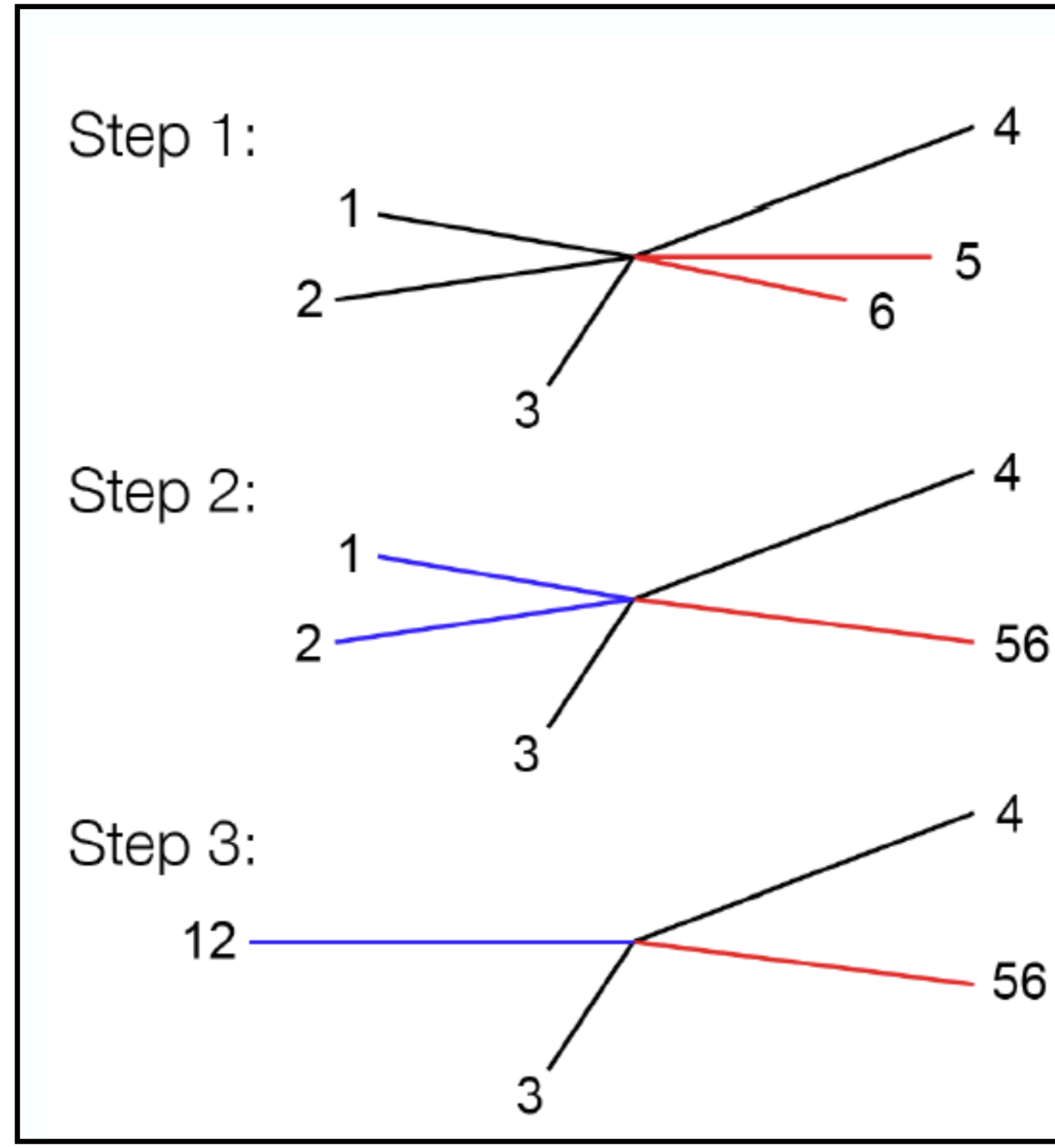
$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p})\frac{\Delta R_{ij}}{R}$$
$$d_i = k_{\mathrm{T},i}^{2p}$$

- where $k_{\mathrm{T},i}$ is the transverse momentum of particle $i$, and $R$ is defined as

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$$

- sequentially compute all distances $d_{ij}$ and $d_i$ find smallest
- declare as jet and remove from list, if $d_i$ is smallest
- the two smallest are combined (sum of fourvectors)
- the procedure is repeated until all particles are used

- the parameter $p$ governs the relative power of energy versus geometrical scales
- the case $p$=0 is a purely geometric algorithm




**1 = kT**
**0 = Cambridge / Aachen**
**-1 Anti-kT**

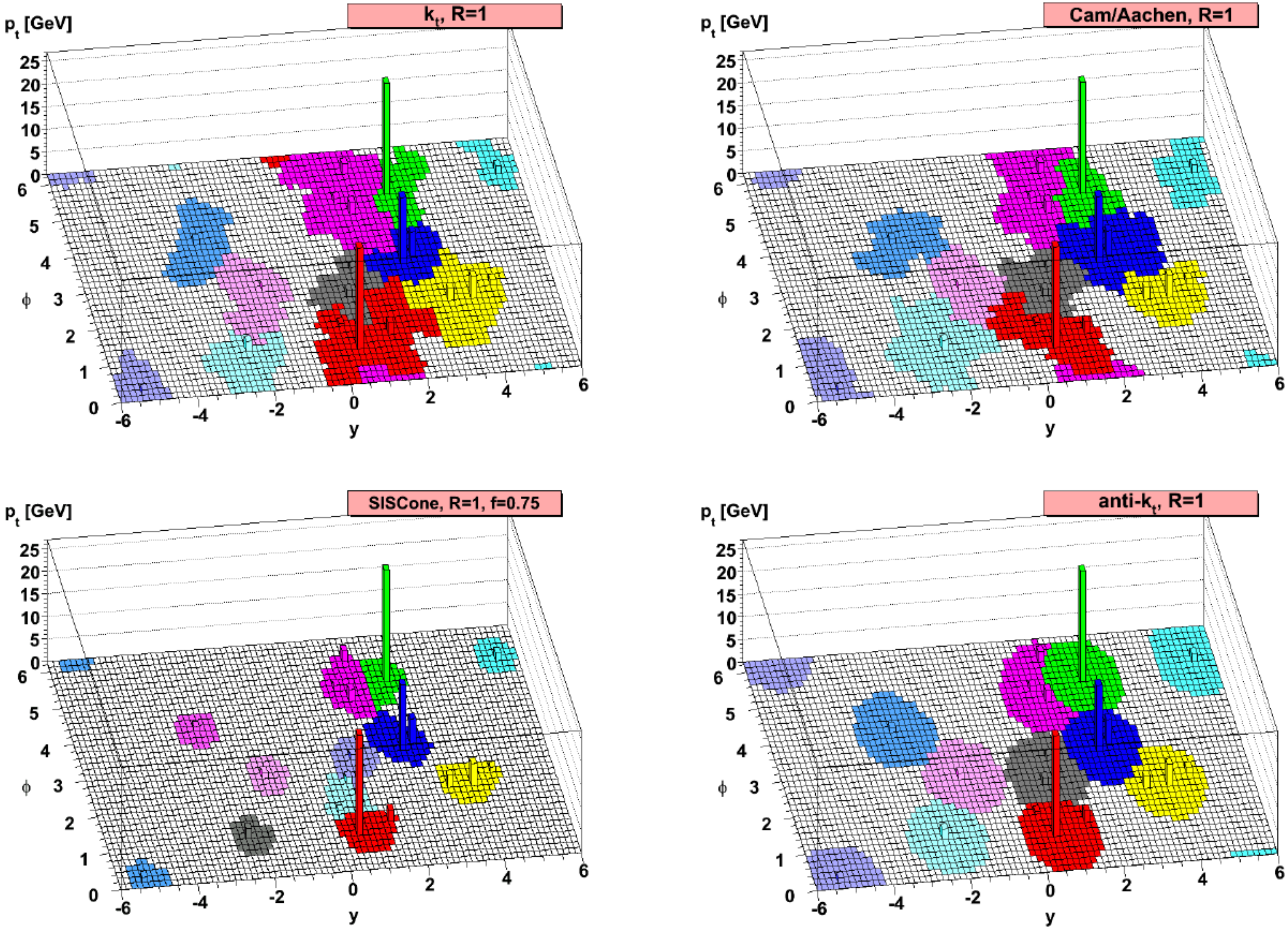
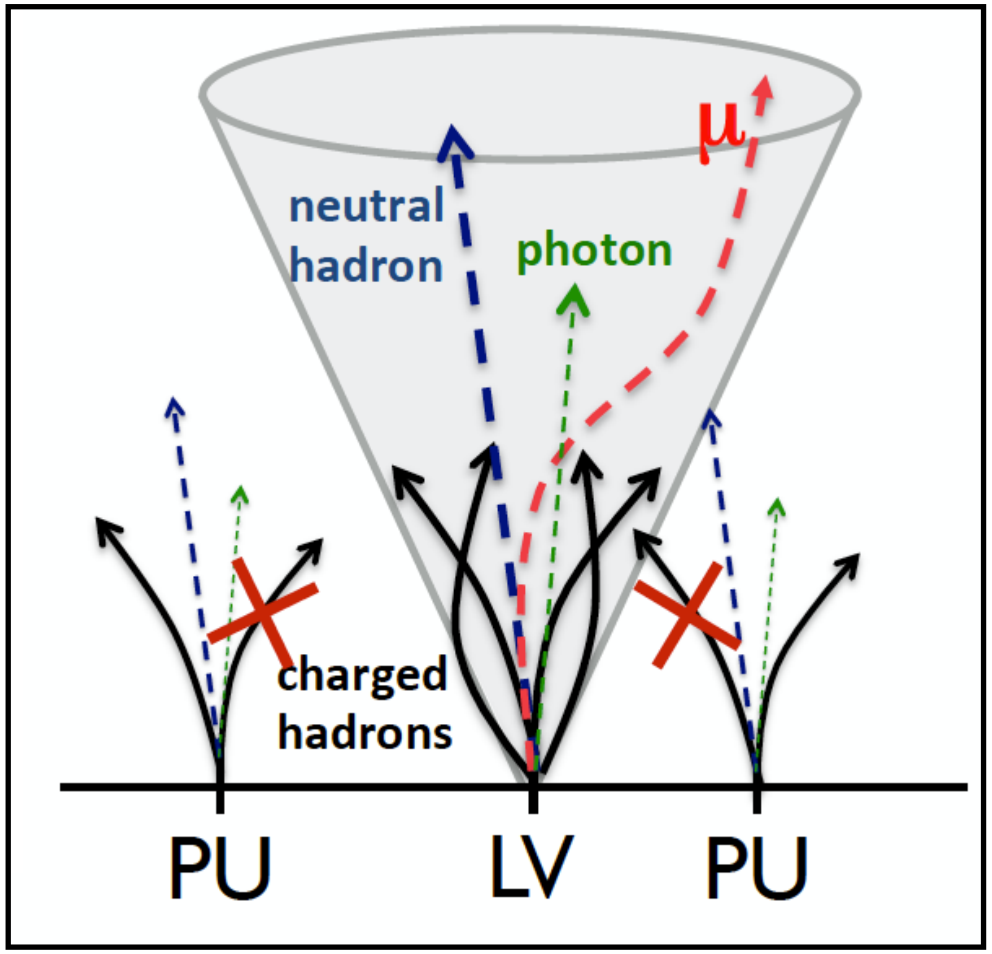



Figure 1: A sample parton-level event (generated with Herwig [8]), together with many random soft "ghosts", clustered with four different jets algorithms, illustrating the "active" catchment areas of the resulting hard jets. For $k_t$ and Cam/Aachen the detailed shapes are in part determined by the specific set of ghosts used, and change when the ghosts are modified.

# pileup

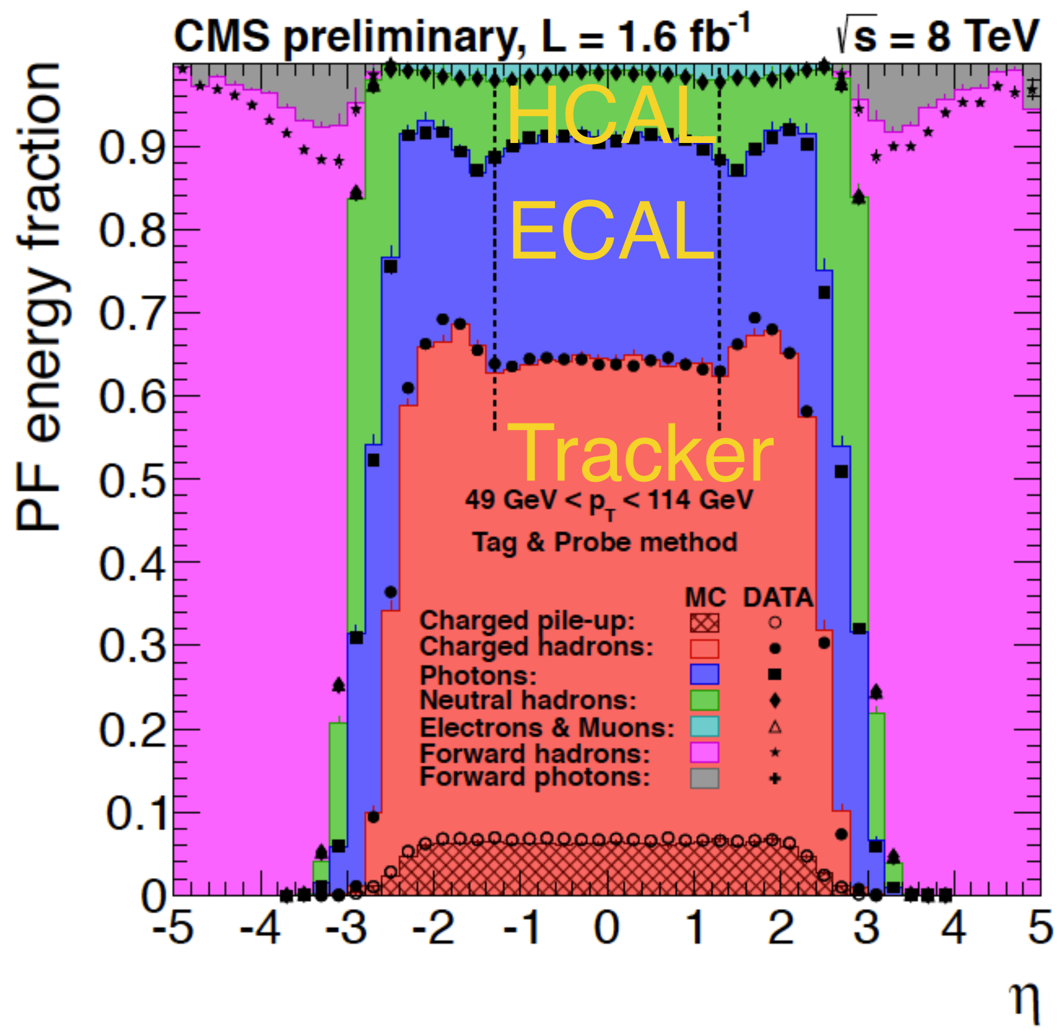


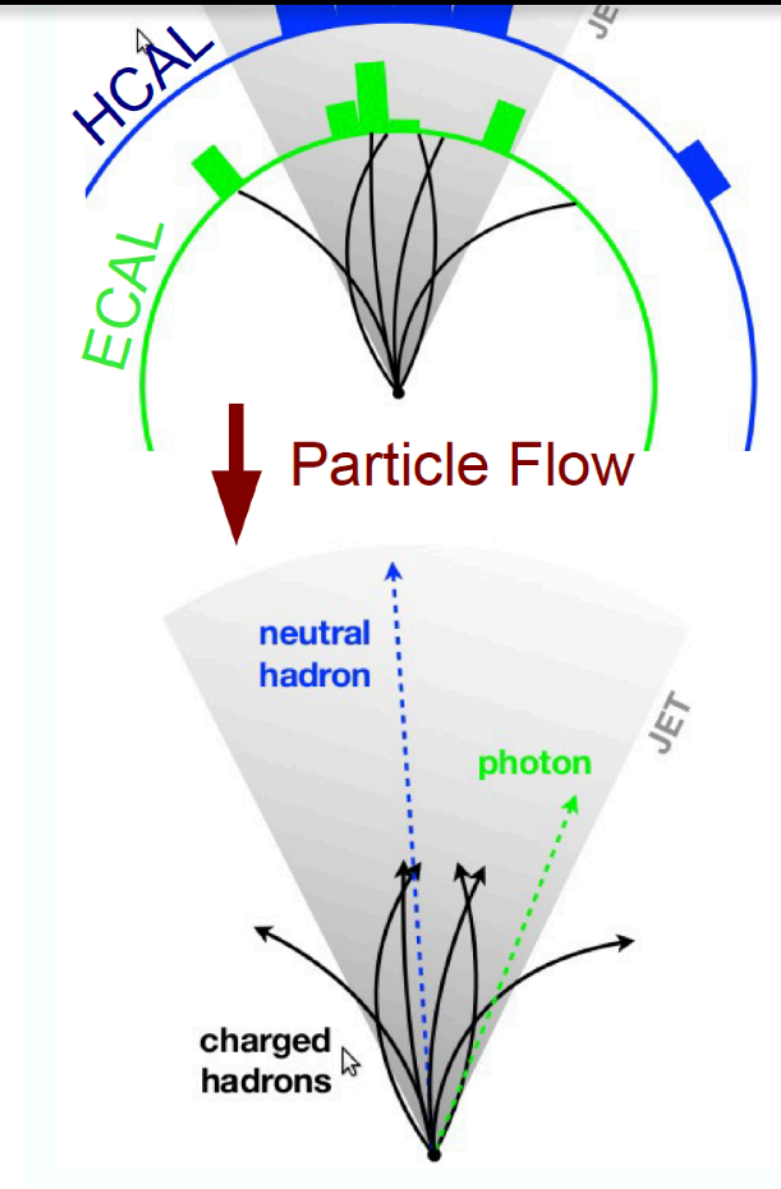

- pileup contaminates the jet with energy contributions that don't belong there
- various pileup-suppression algorithms exist and some are used
- CHS: charged hadron subtraction:
  - if a particle does not originate from the signal vertex, subtract it from the jet
- neutral contributions cannot be tracked to the vertex
  - energy offset depending on number of vertices (details later)

- PU removal also pays a role in the topic of jet substructure and JET energy scale corrections (see next week)
- it is not only  a standalone problem which can be fixed at one point in the jet reconstruction sequence
- it is woven into the various steps of jet energy reconstruction at different levels

# CMS: particle flow

- jet composition:



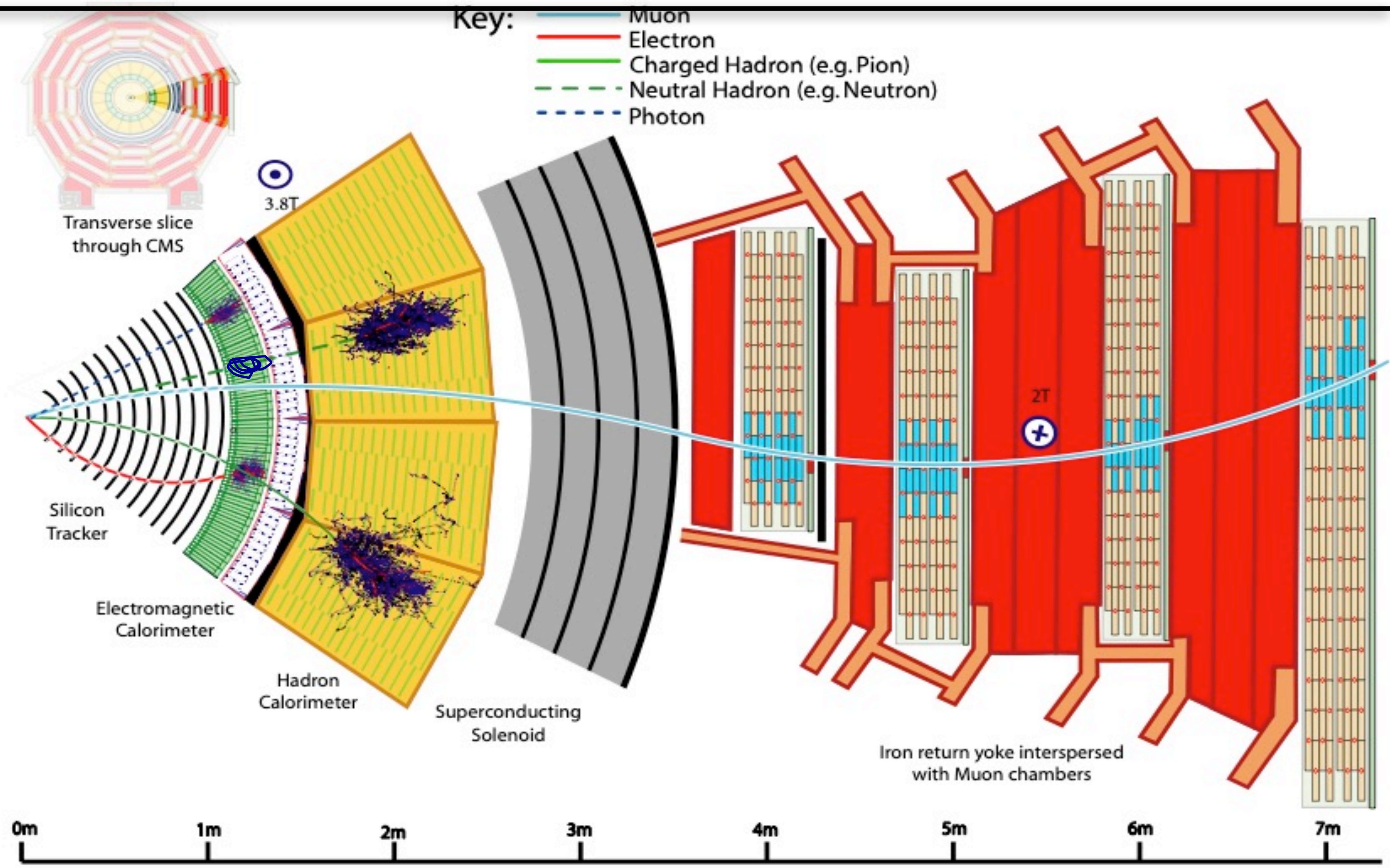

- **PF combines information from all subdetectors to optimise resolution:**

| Detector | $p_T$-resolution (range) | η/Φ-segmentation |
|---|---|---|
| Tracker | 0.6% (0.2 GeV) – 5% (500 GeV) | 0.002 x 0.003 (first pixel layer) |
| ECAL | 1% (20 GeV) – 0.4% (500 GeV) | 0.017 x 0.017 (barrel) |
| HCAL | 30% (30 GeV) – 5% (500 GeV) | 0.087 x 0.087 (barrel) |

# CMS: particle flow

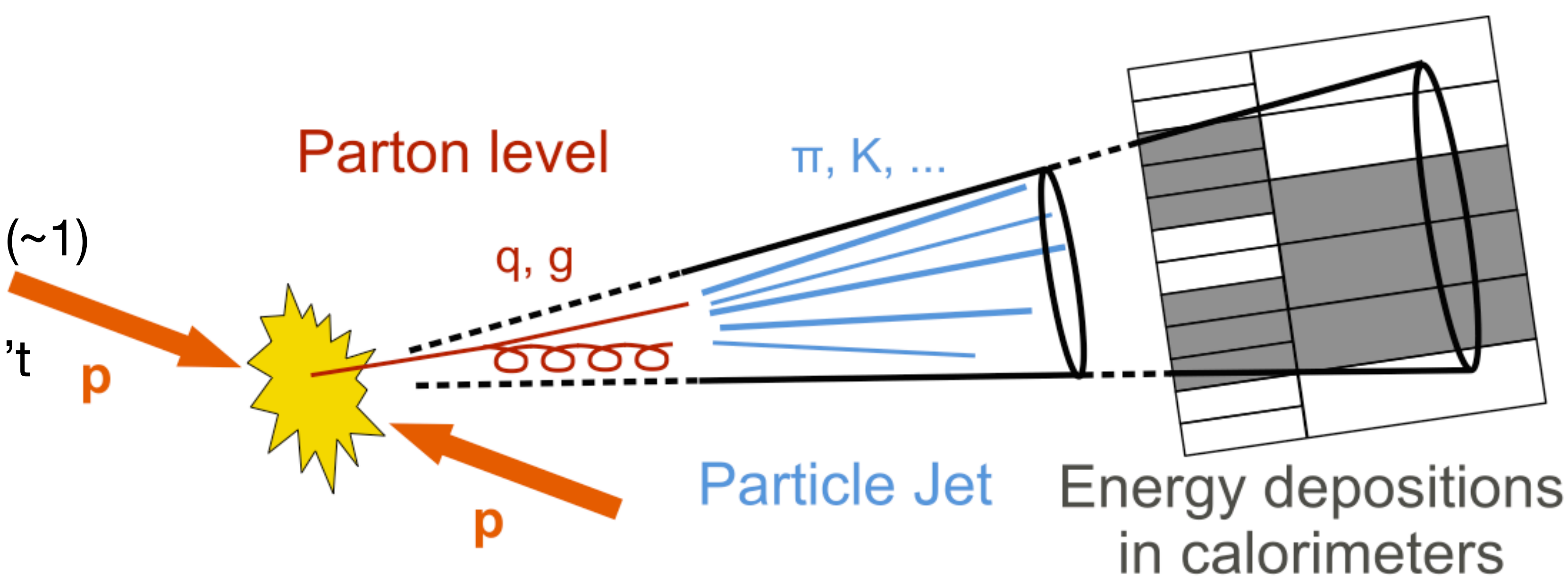
Making optimal use of all subdetectors for particle reconstruction and identification

To improve efficiencies,
fake rates,
momentum resolution,
angular resolution,
charge identification



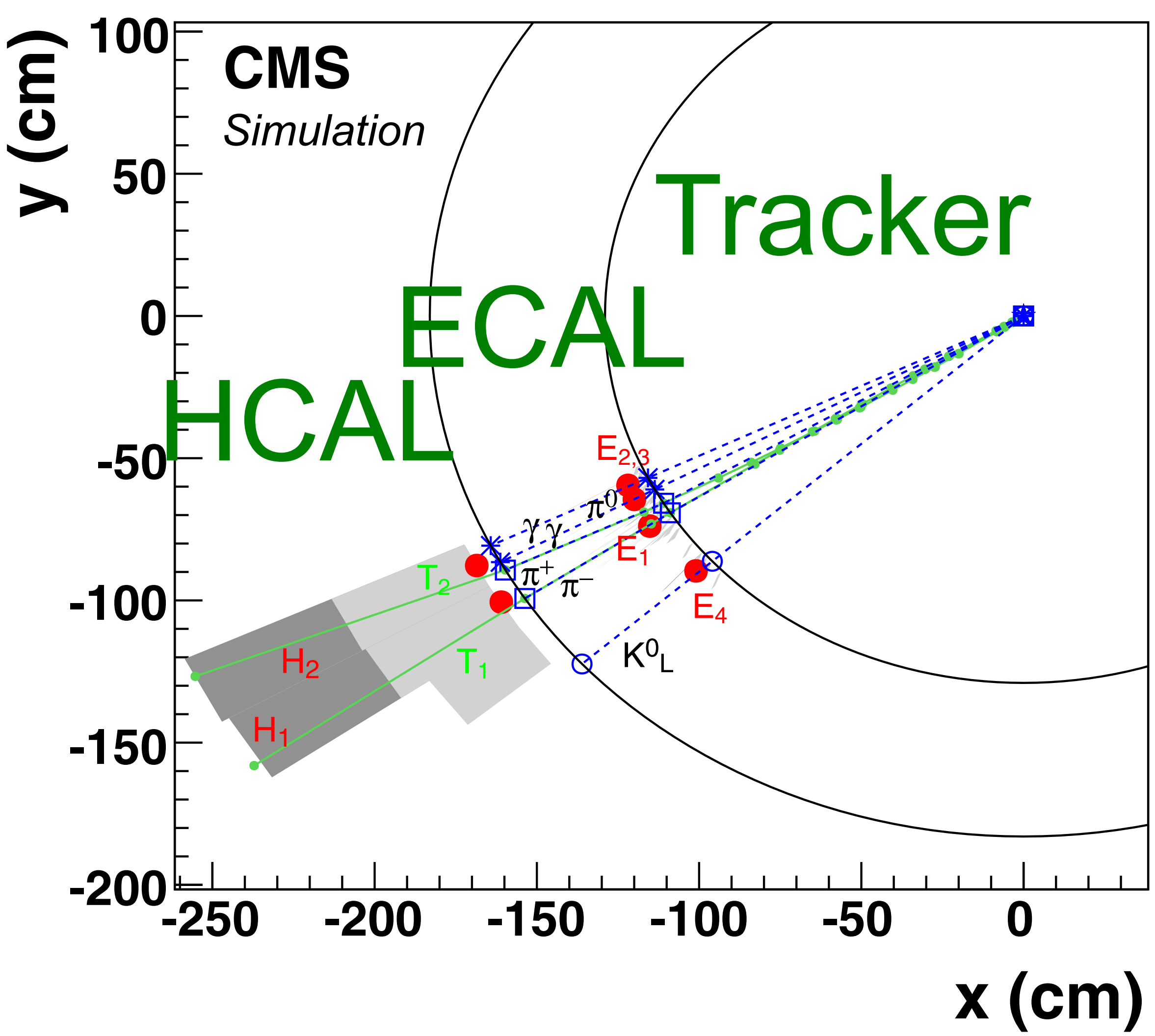

Reconstruction using only ECAL+HCAL suffers from

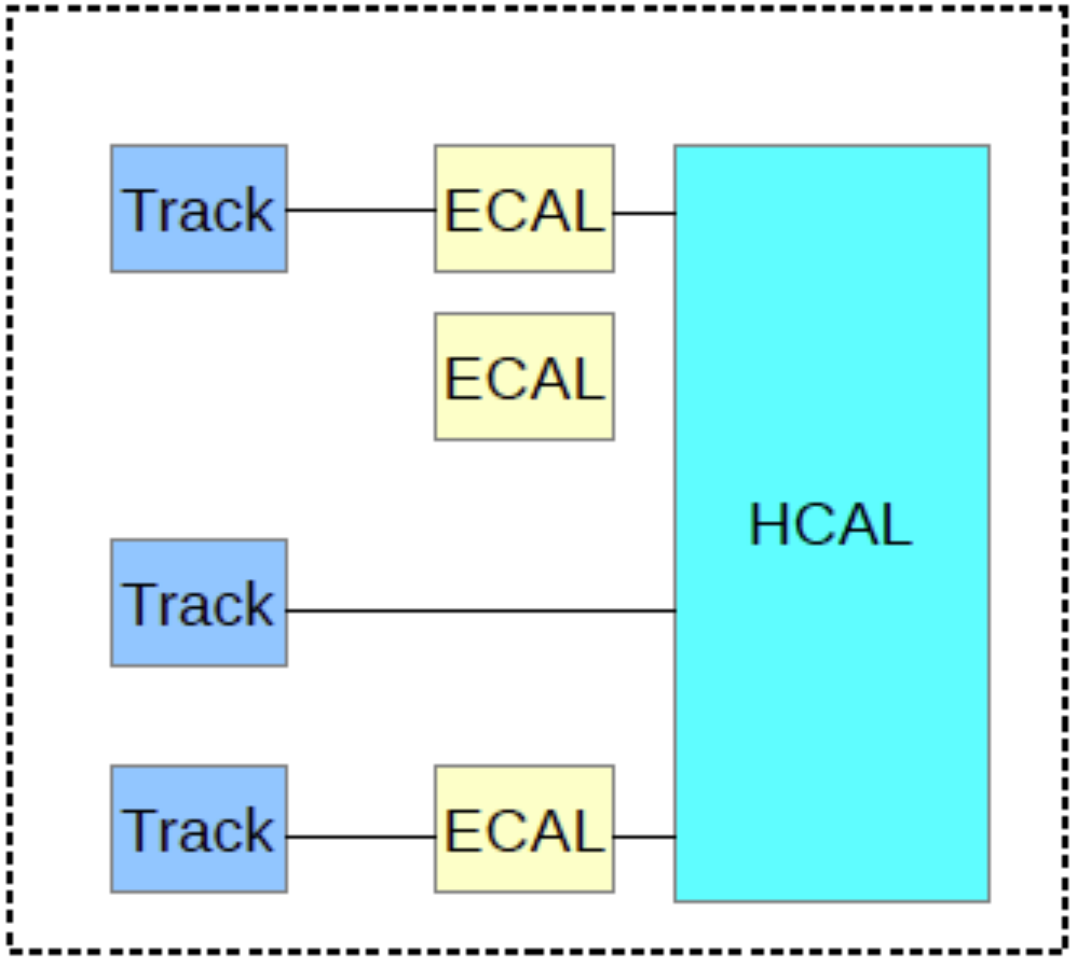
- Limited energy resolution in HCAL
- Limited angular resolution in HCAL
- Different energy response for photons (~1) and hadrons (~0.7) in ECAL
- Low momentum charged hadrons don't reach calorimeters (4T magnet)
- Muons are not measured (MIPs)



# CMS: particle flow

- ECAL-only → photon
- ECAL+HCAL or HCAL-only → neutral hadron
- track+ECAL → electron or charged hadron
- track+ECAL+HCAL → check ECAL+HCAL energy in hadron hypothesis
  1. compatible → charged hadrons
  2. track>Calo → remove worst tracks
  3. calo>Track → extra photons and neutral hadrons

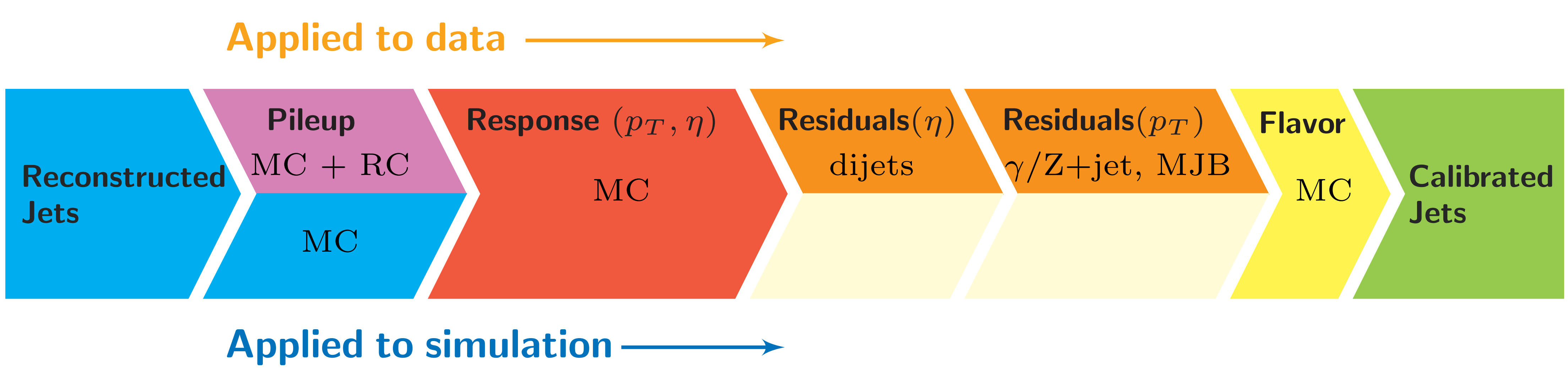


# jet energy calibration

- the initially reconstructed jet energy (sum of the constituents) is not the true jet energy, because of
  - pileup
  - calorimeter response (hadronic sampling calorimeter)
  - jet flavour (b-jets and c-jets contain neutrinos which are not measured)
  - mismodeling (simulation is never perfect)

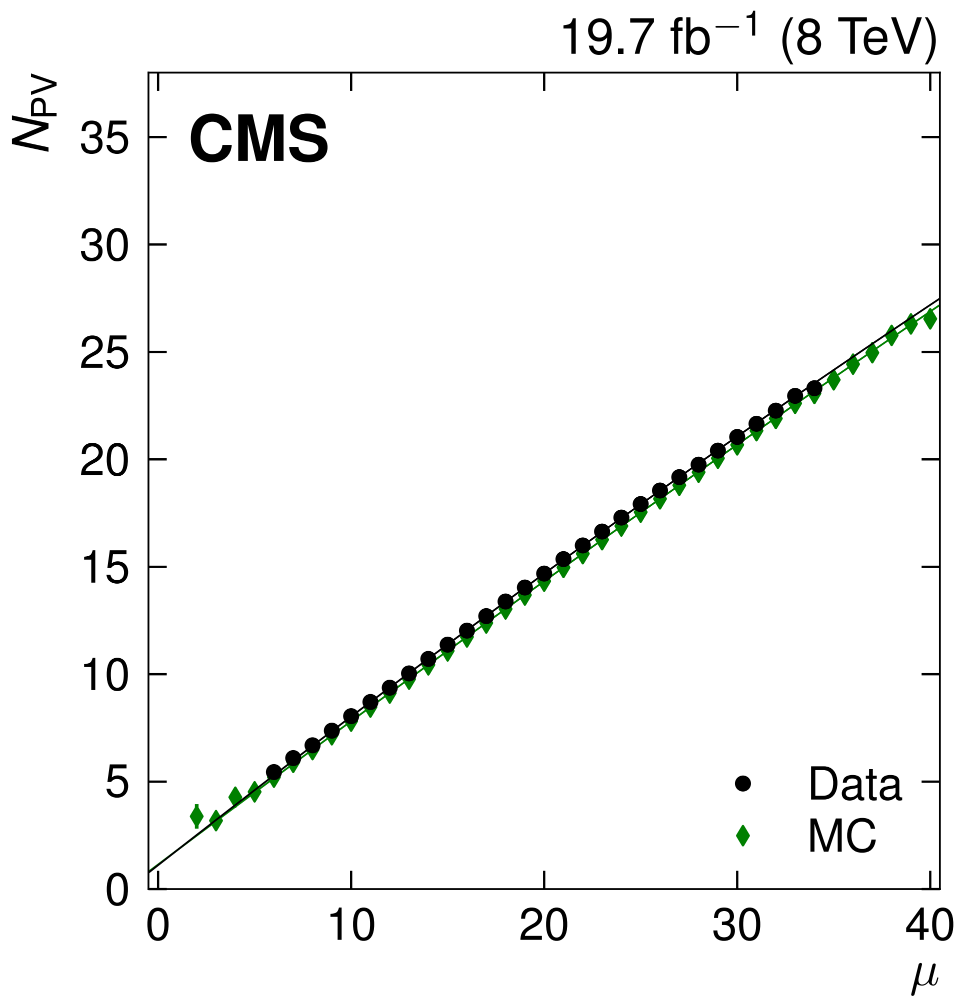
- a staged jet energy calibration is applied:



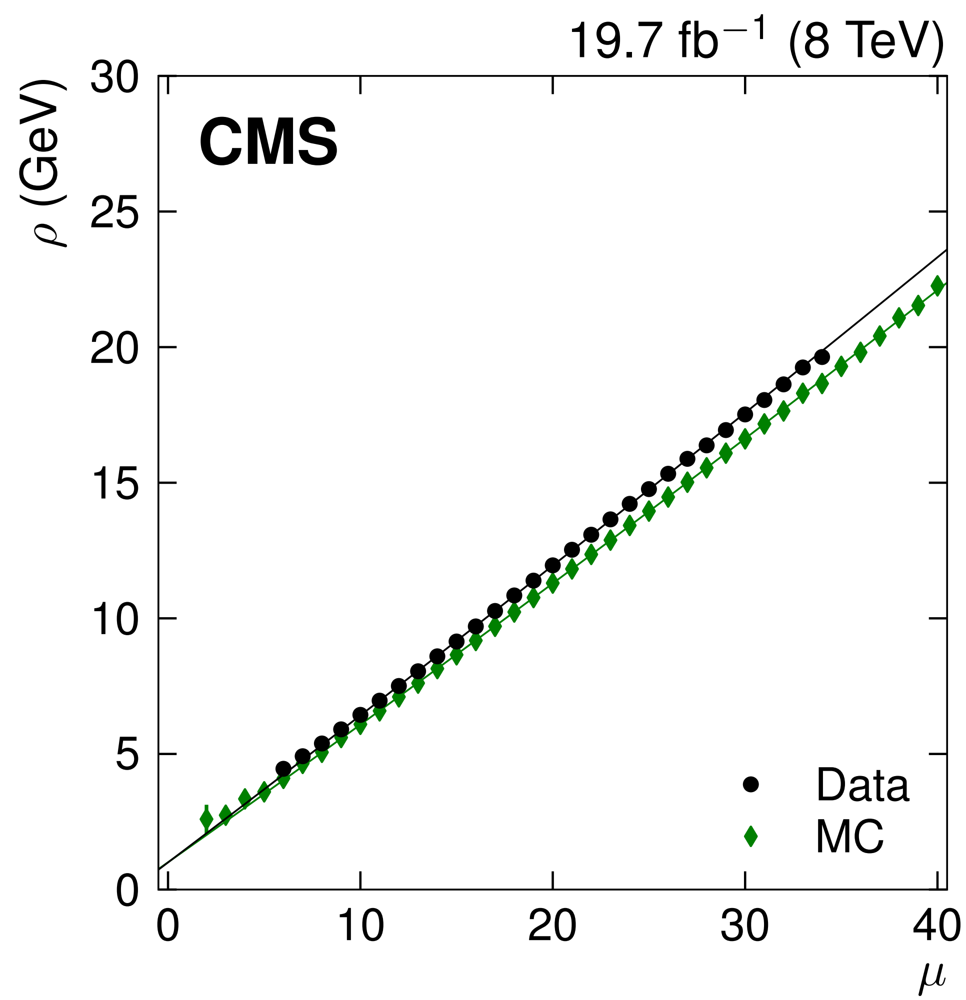
# jet energy calibration

- pileup correction:

reconstructed number of PV




true number of PV



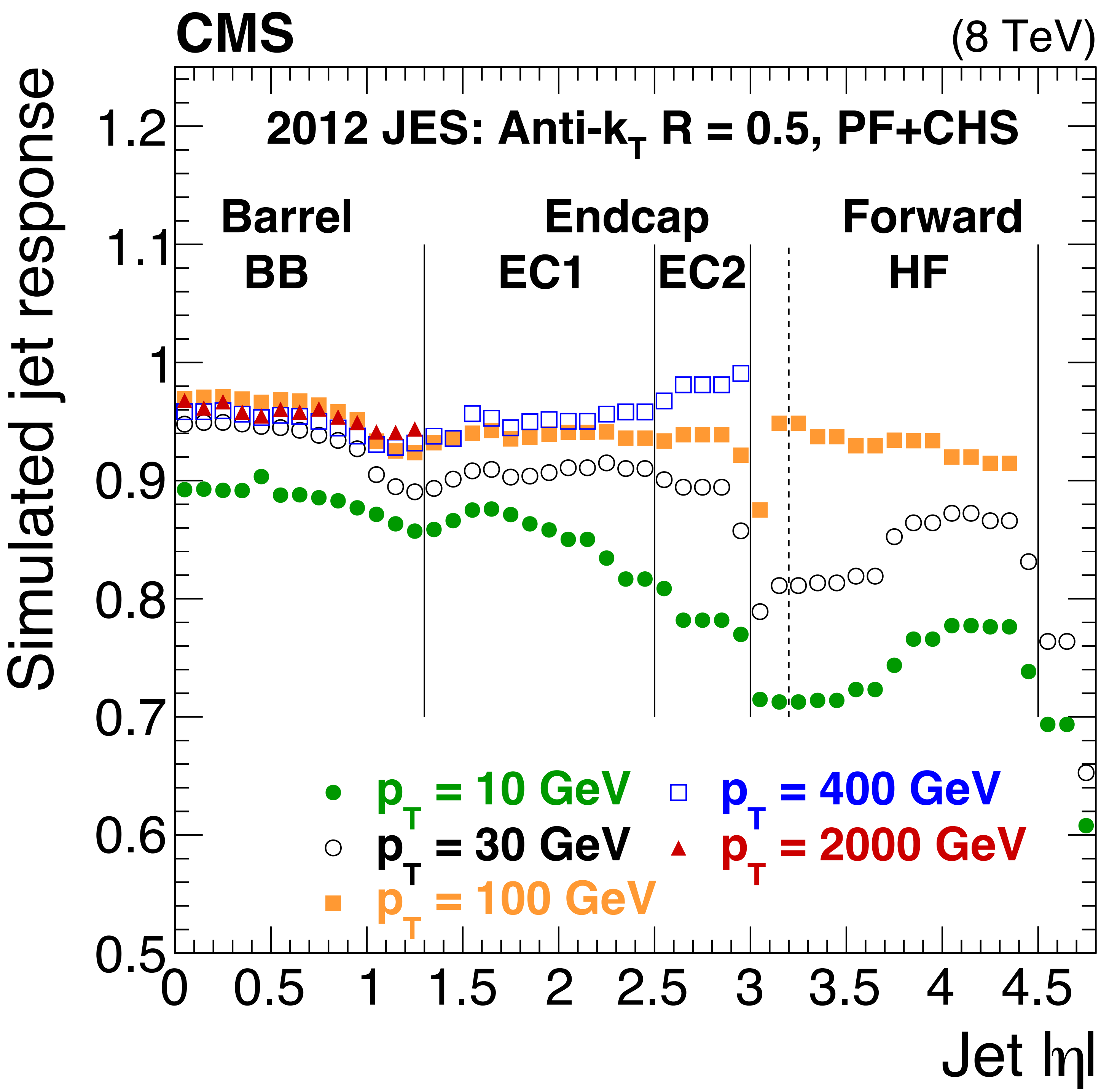

true number of PV

offset energy density

- after CHS, the remaining offset due to PU is subtracted as function of the PV multiplicity

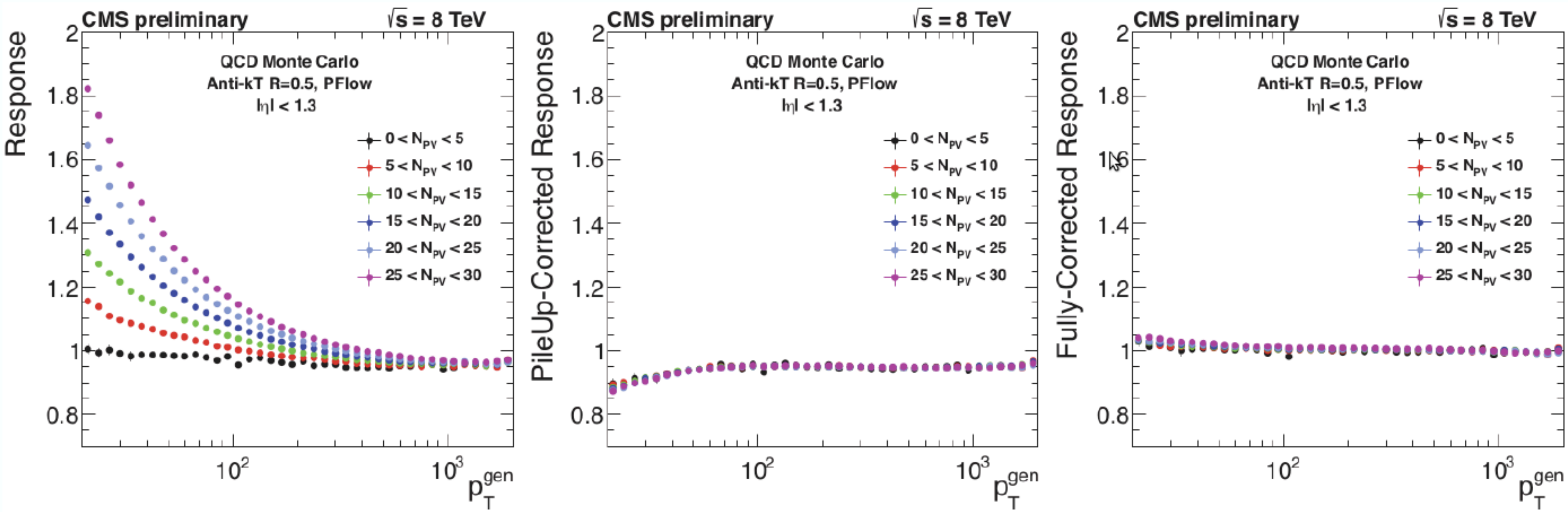
# jet energy calibration

- simulated jet response:




- as far as the simulation can be trusted, the response is corrected by "correction factors"

# jet energy calibration




**before corrections**

- high contributions from PU (especially at low $p_t$)

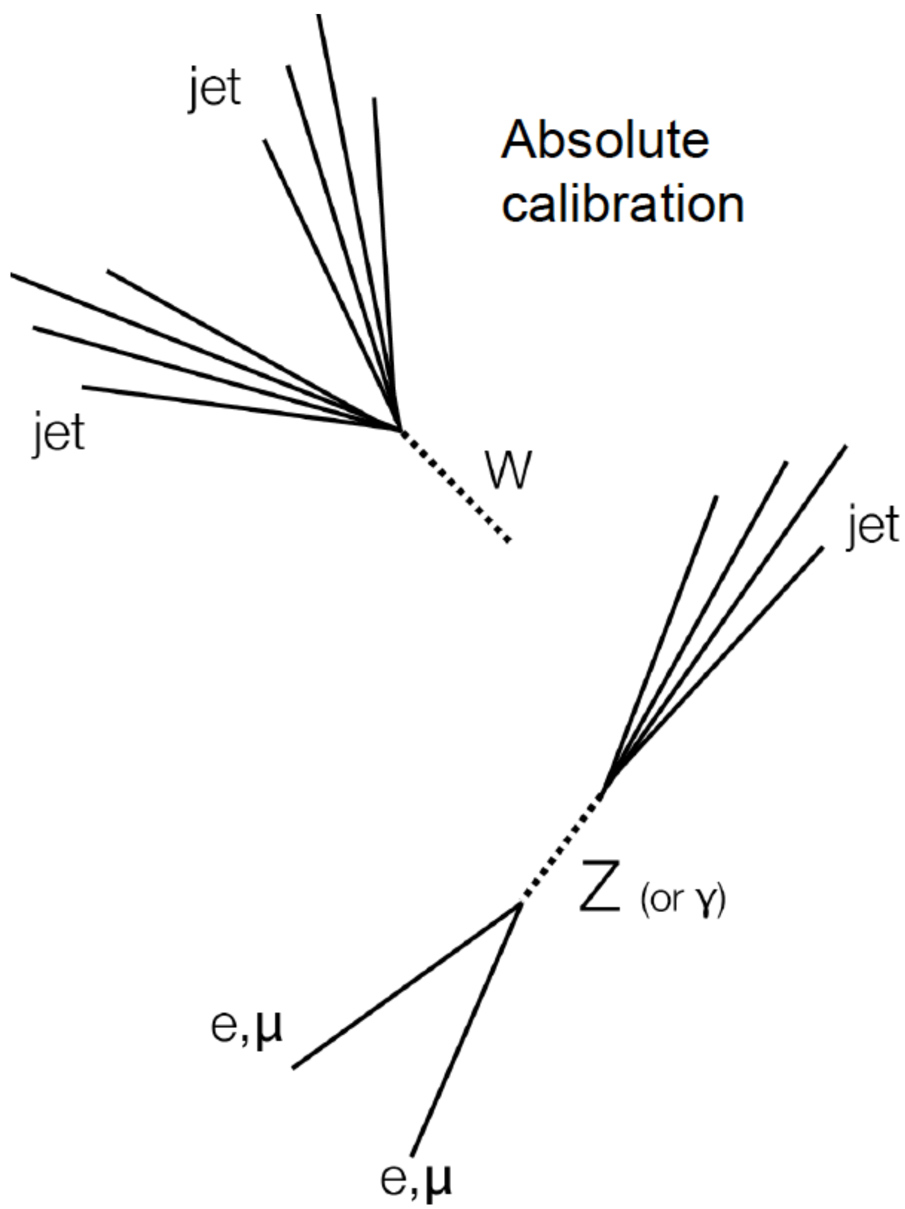
**after PU corrections**

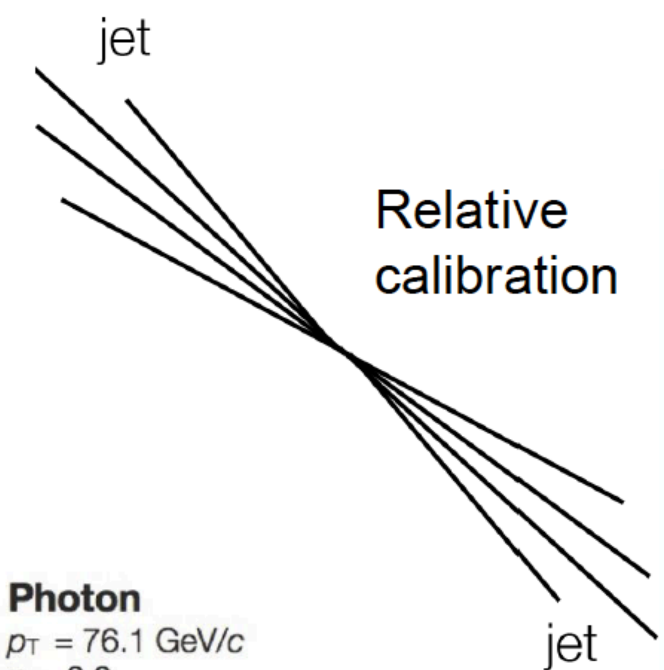
- PU dependency removed
- consistent for any number of PU interaction $N_{PV}$

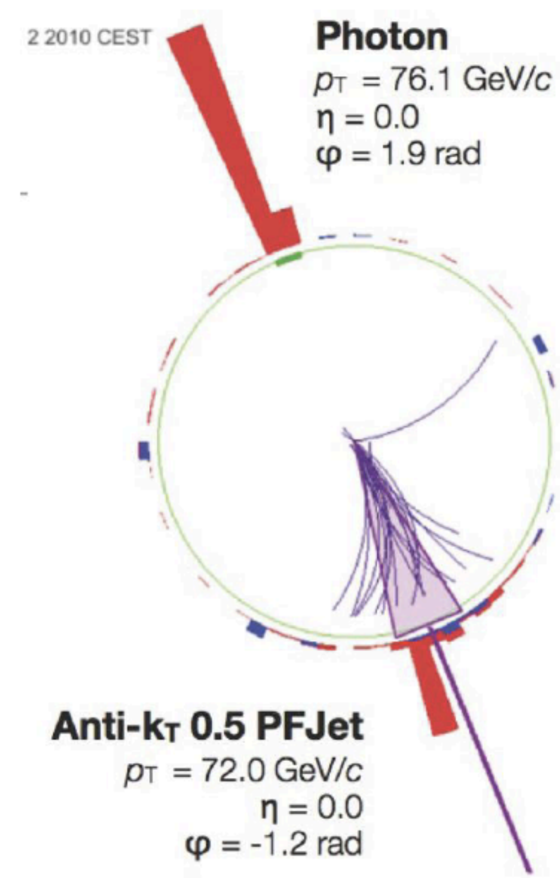
**after MC truth corrections**

# jet energy calibration

- **residual corrections:**
  - after the corrections based on simulation, the residual data/simulation disagreements are determined
  - e.g. from “balance” of objects

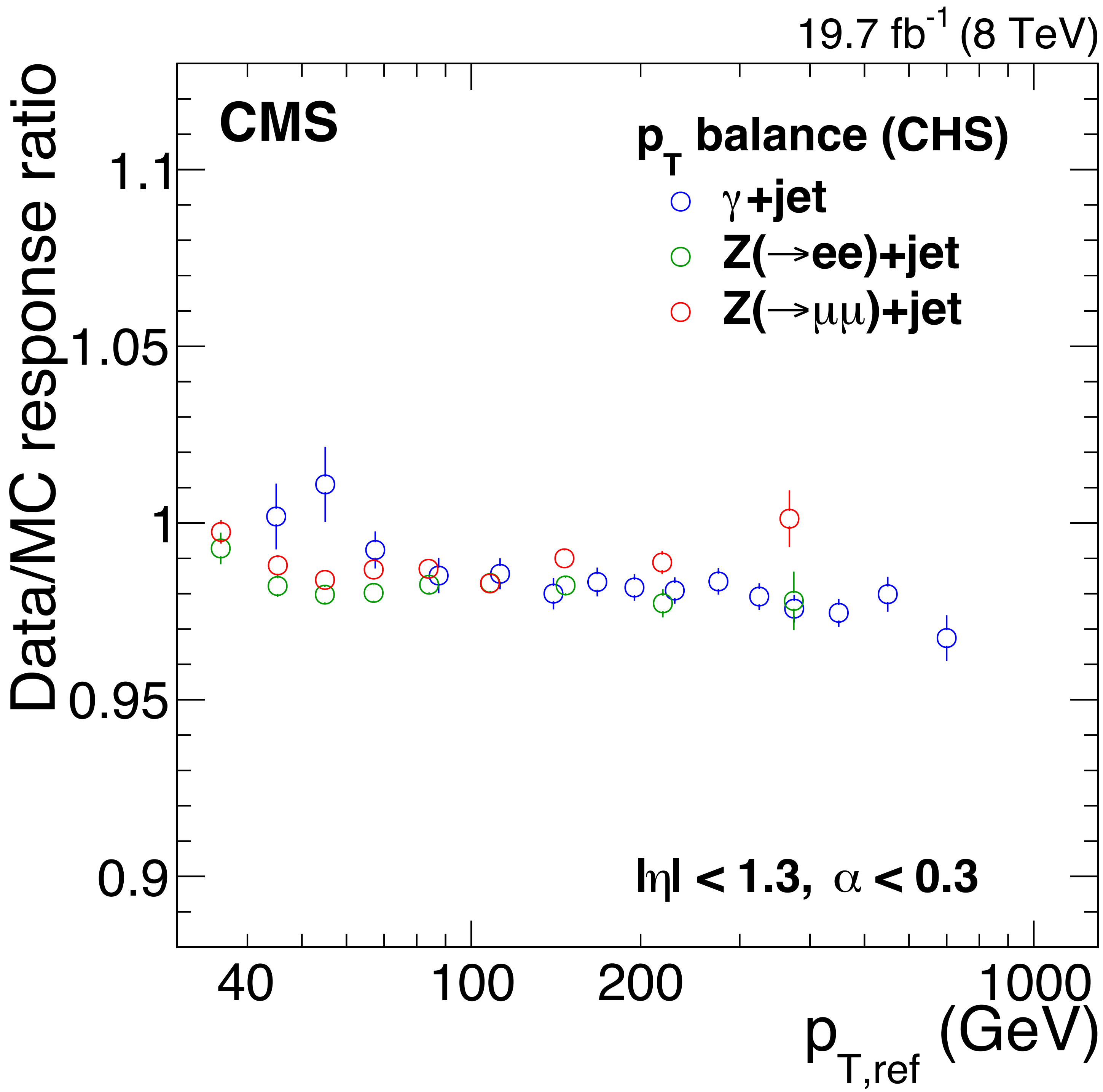


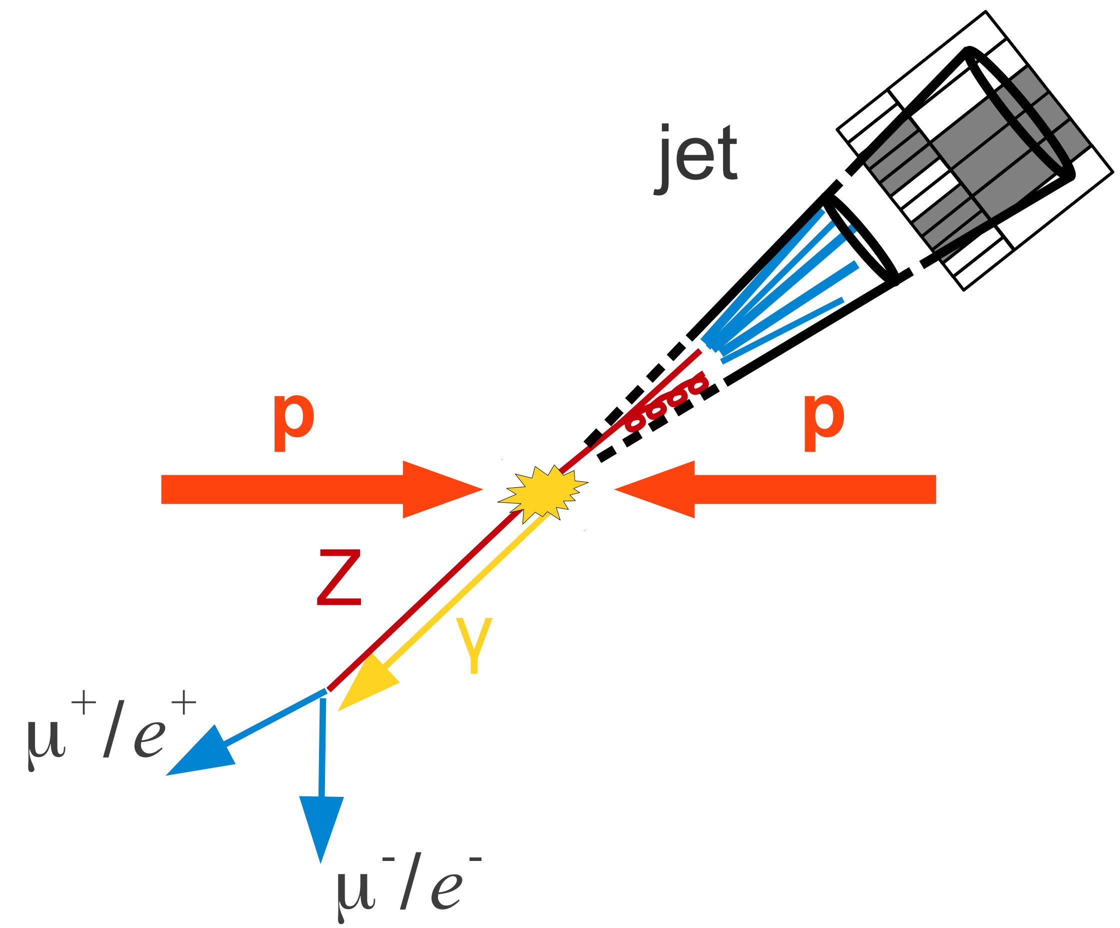



# jet energy calibration

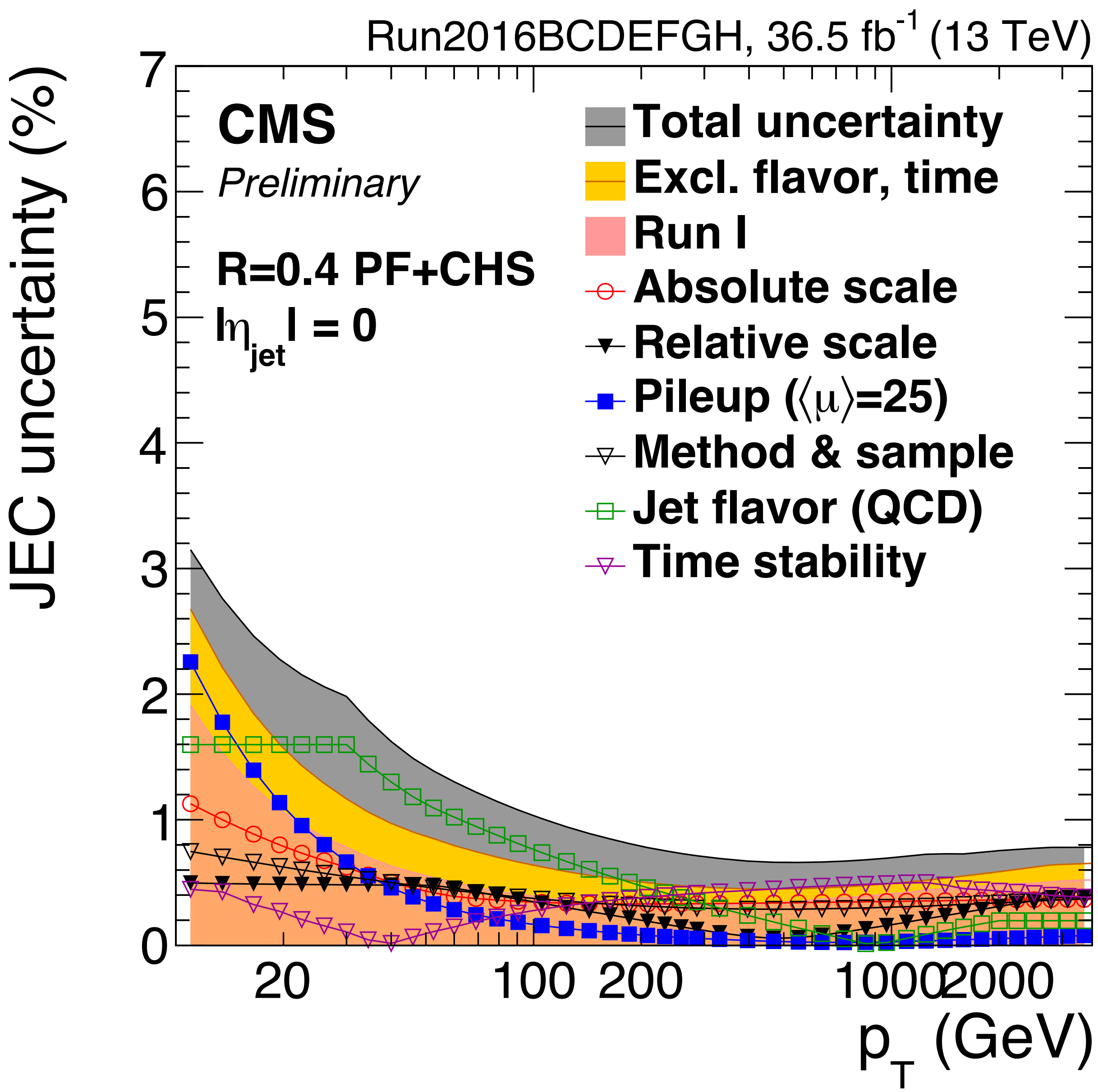
- residual data/simulation disagreements

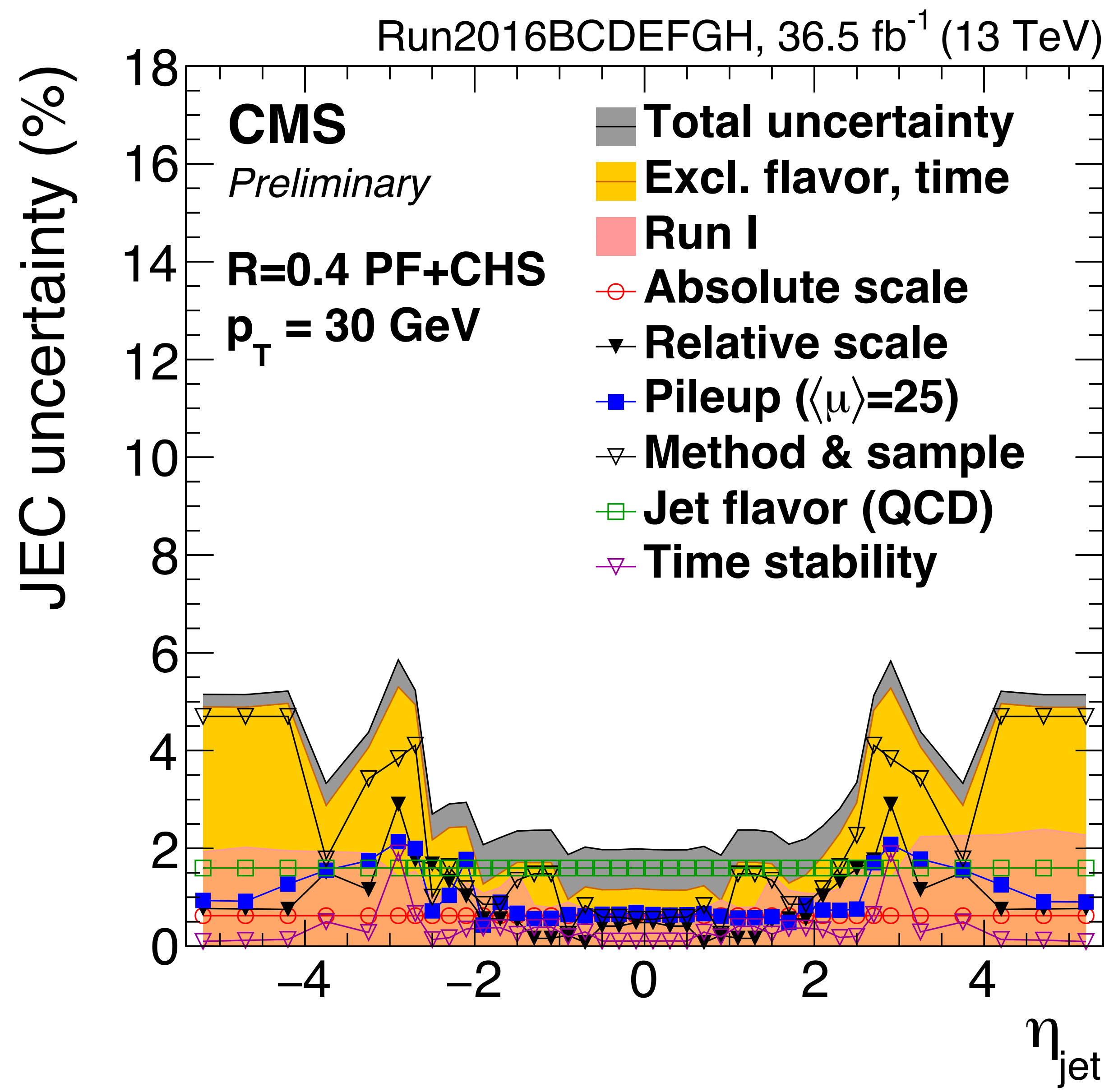


# jet energy calibration

- resulting uncertainties in the jet energy scale:

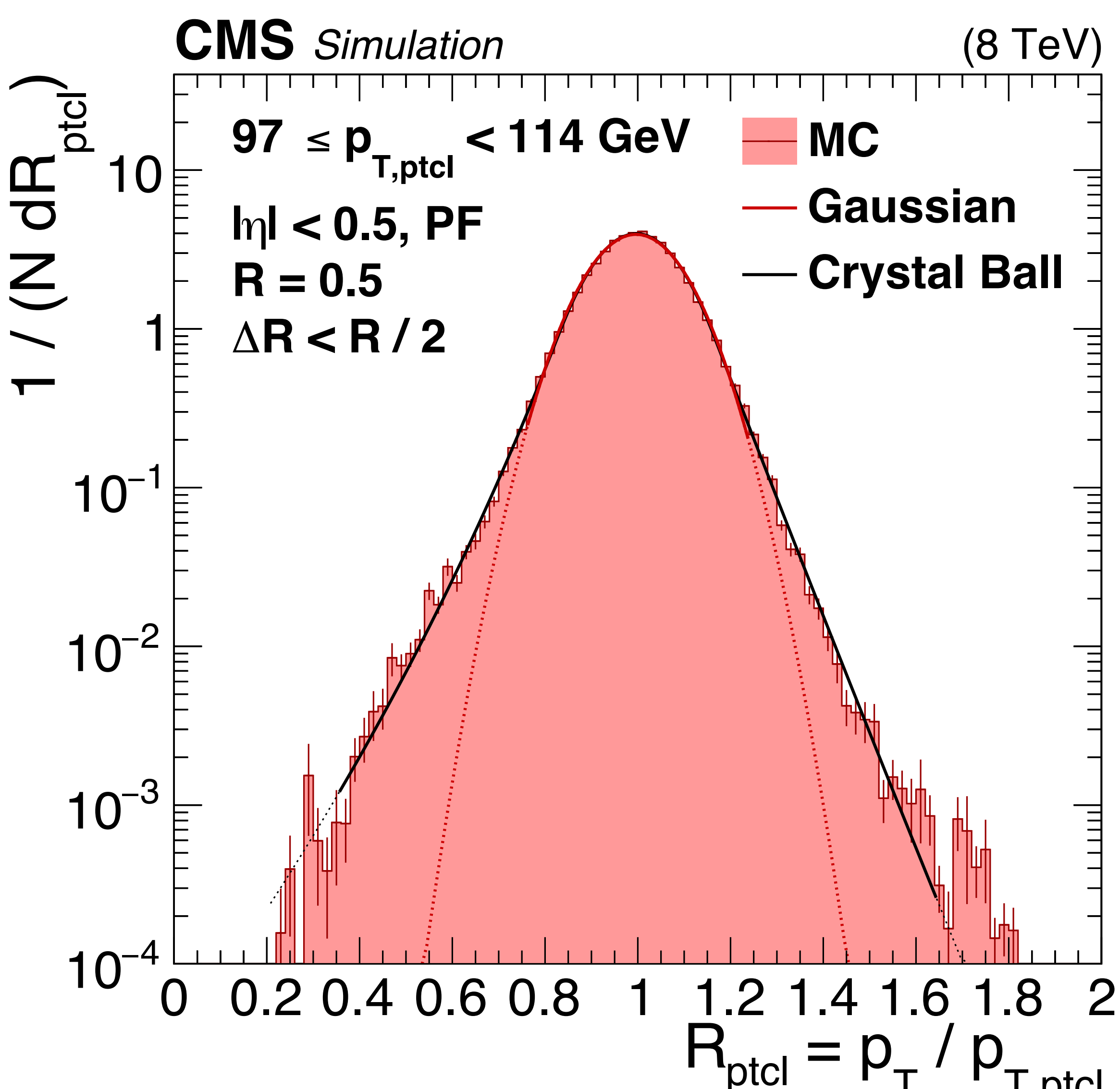


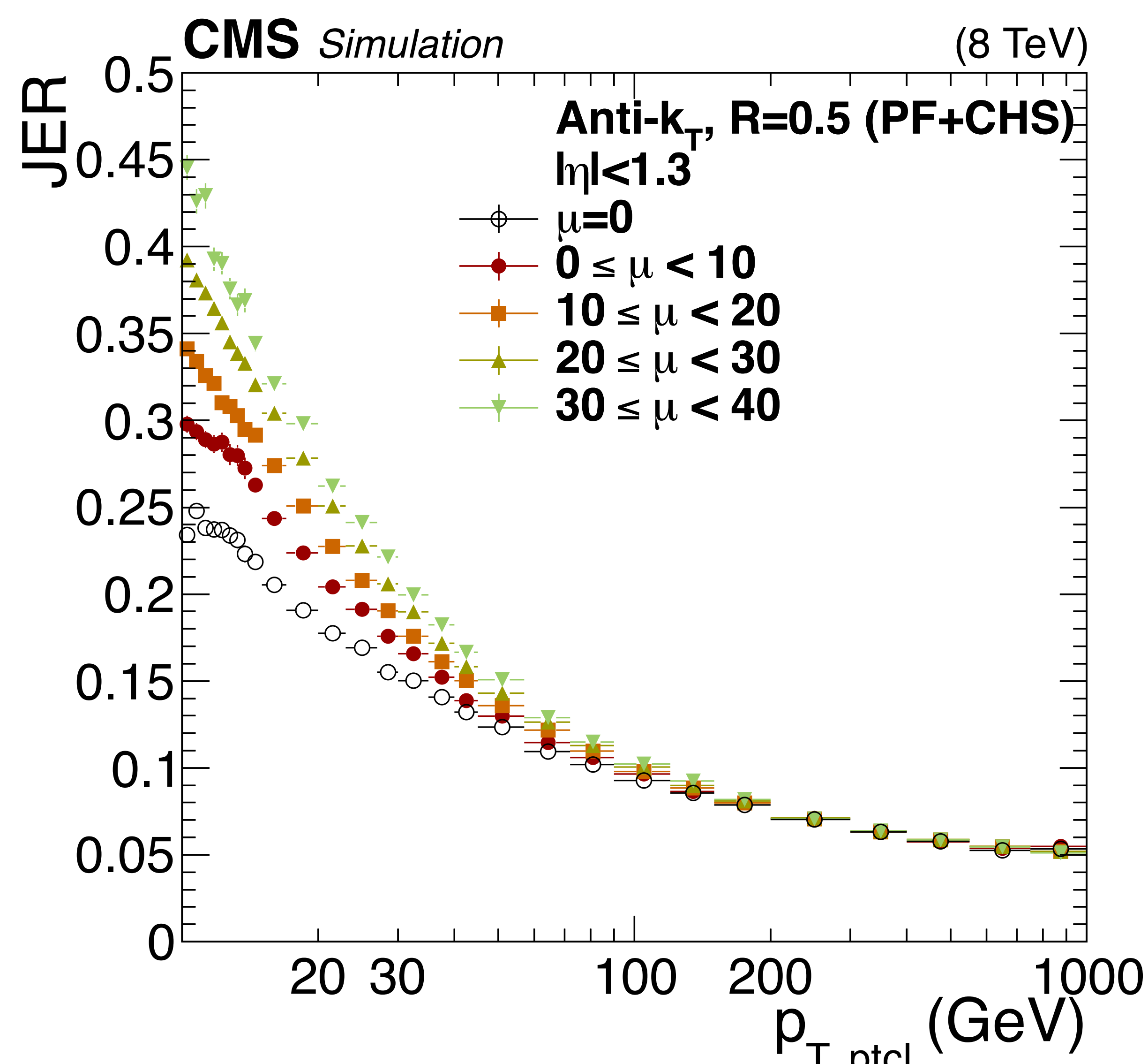



# jet energy resolution

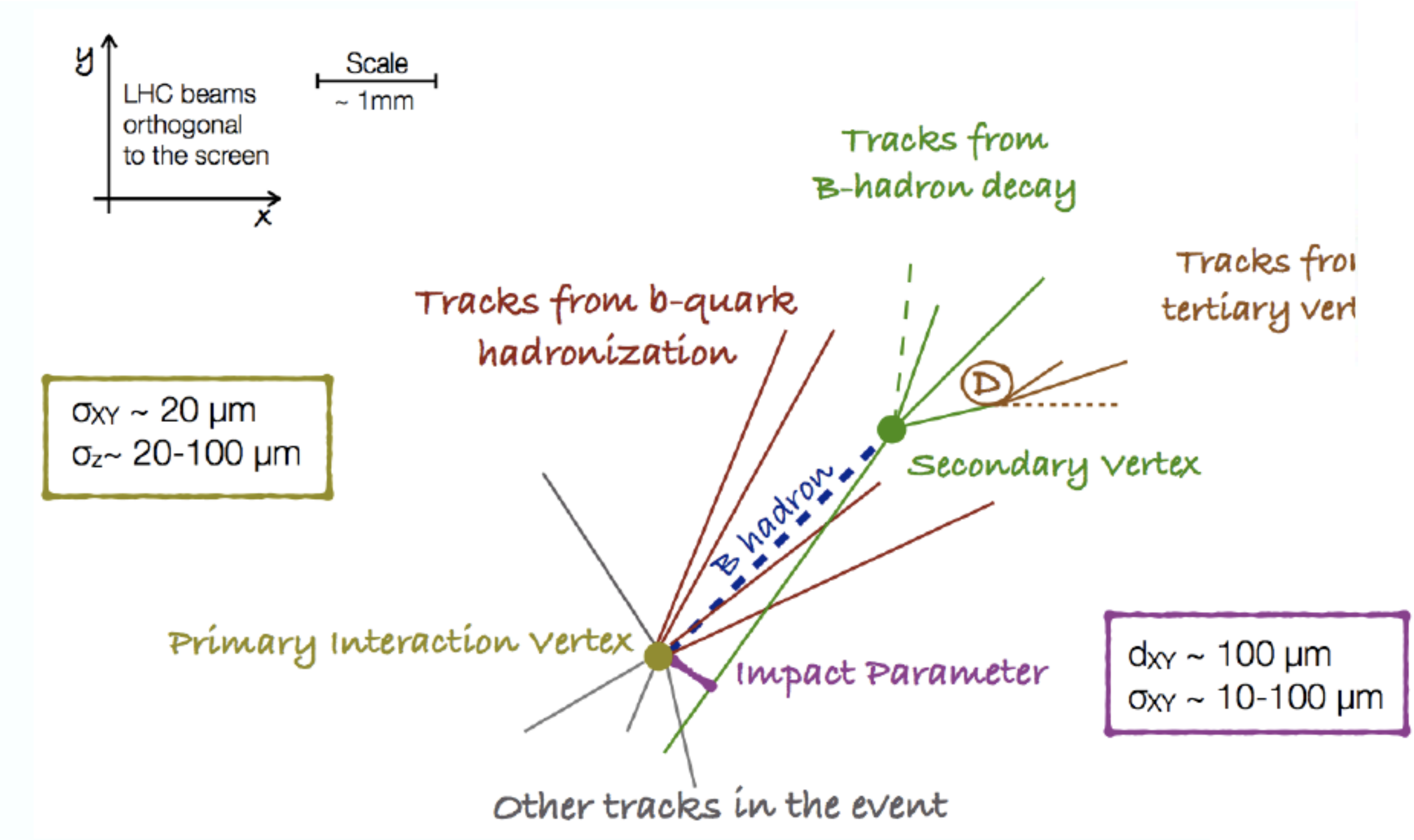
**careful:**

- scale uncertainty and resolution are two different things
- scale uncertainty is the uncertainty in the knowledge of the jet energy calibration
- each jet is miscalibrated in the same way
- resolution refers to the random (statistical) fluctuation jet-by-jet



# jet flavour identification




➜ **more details about particle ID methods next week**

# basics

- we end up with a list of few high-level “physics objects” in a “file”:
    - jets (flavour-tagged jets)
    - electrons
    - photons
    - muons
    - missing transverse momentum
- the physics objects have properties such as
    - momentum
    - energy
    - isolation from other particles
- from these high-level physics objects, we have to **reconstruct** what happened in the proton proton collision
    - this reconstruction is educated with **simulation**

# simulation

**several stages of event simulation**

- proton-proton collision:
  - event generation (theory)

- interaction with detector
  - GEANT 4

- response of the detector electronics (including trigger)
  - digitisation with detector specific software

---

- at this level, the simulation output looks exactly like **real data**
  - with the difference that we know what happened in the pp collision

- all the remaining steps (track reconstruction, jet reconstruction, …) are done in the exact same way in data and simulation

- at the end we compare simulation with data to draw conclusions

# physics analysis

- in the exercises we will now start a physics analysis project, that will continue through the end of the semester
- we will also use the time slot of the GEANT course on 4th and 11th February fort this
- we will study an actual dataset from the CMS experiment
- we need some basic understanding of top quarks for this (next slides)

# the top quark

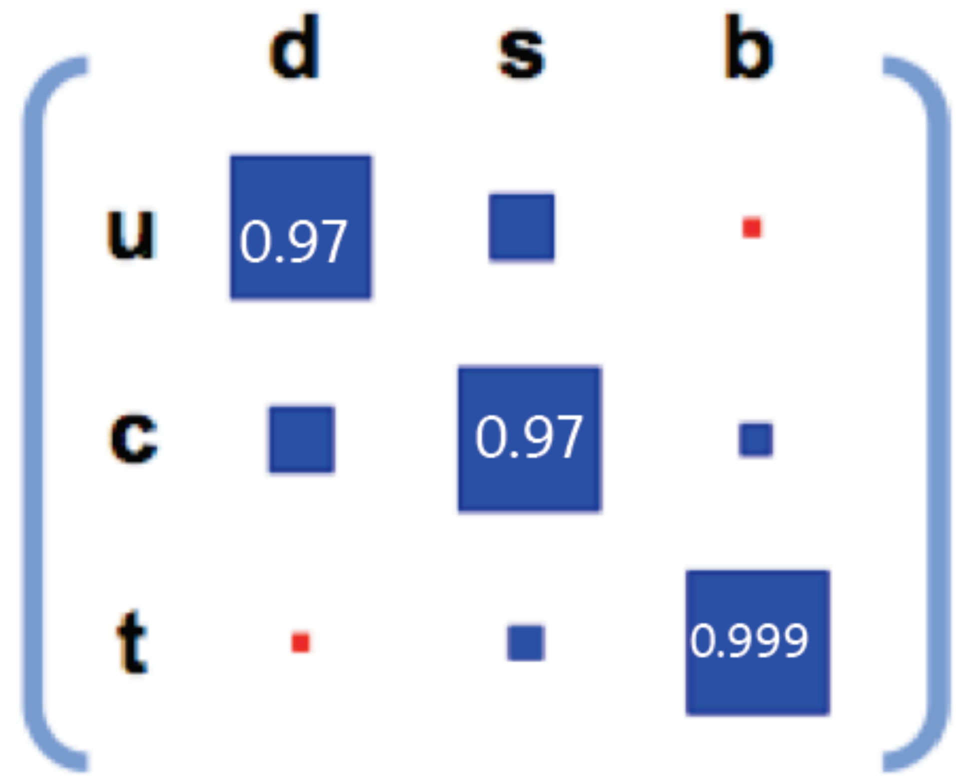
**Why is the Top quark special?**

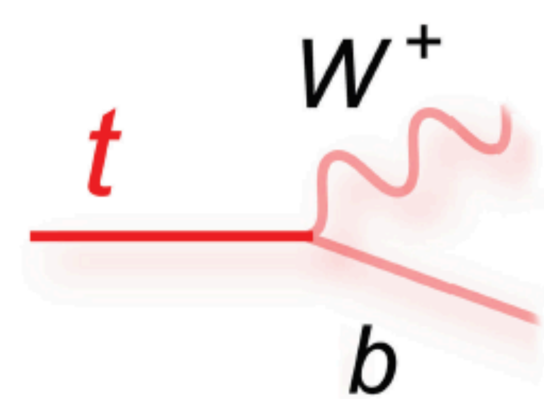
- It is the heaviest SM particle!
- Coupling to Higgs field (Yukawa couplings) are ~1
- Its live time is shorter than the characteristic hadronization time scale
  - Tops decay before fragmentation
  - Top quark decays carry information about spin correlations
- It decays exclusively in *W+b*

**Many searches for physics beyond the SM are connected to top physics**:

- Searches for fourth generation quarks
- SUSY searches (important background, but also final states with tops)
- *Z'* → ttbar ...

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

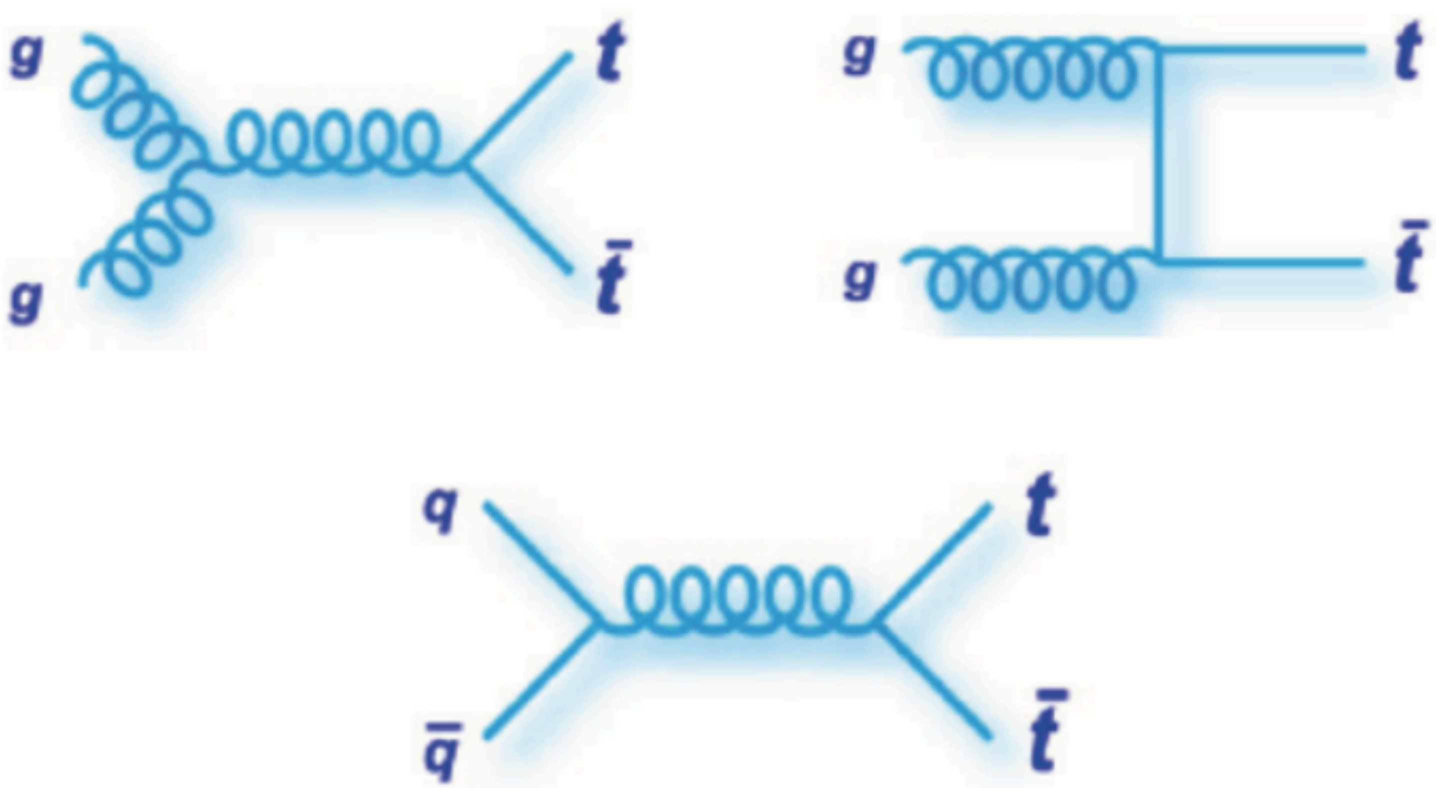


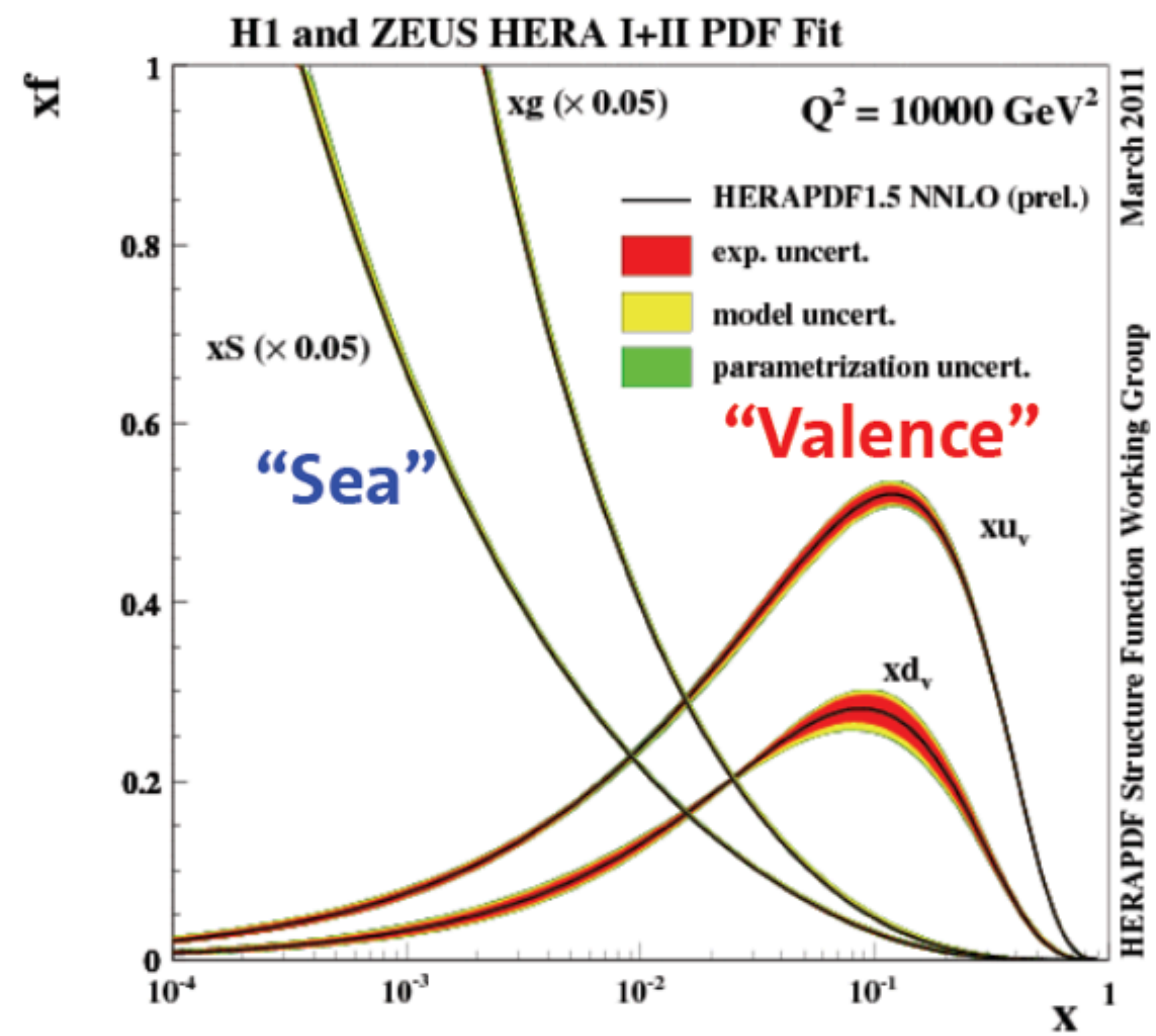



# production

- LHC is a Top factory
- gg and qqbar are complementary at Tevatron and LHC

| | LHC (7 TeV) | Tevatron (1.96 TeV) |
|---|---|---|
| gg | ~80% | ~15% |
| qqbar | ~20% | ~85% |



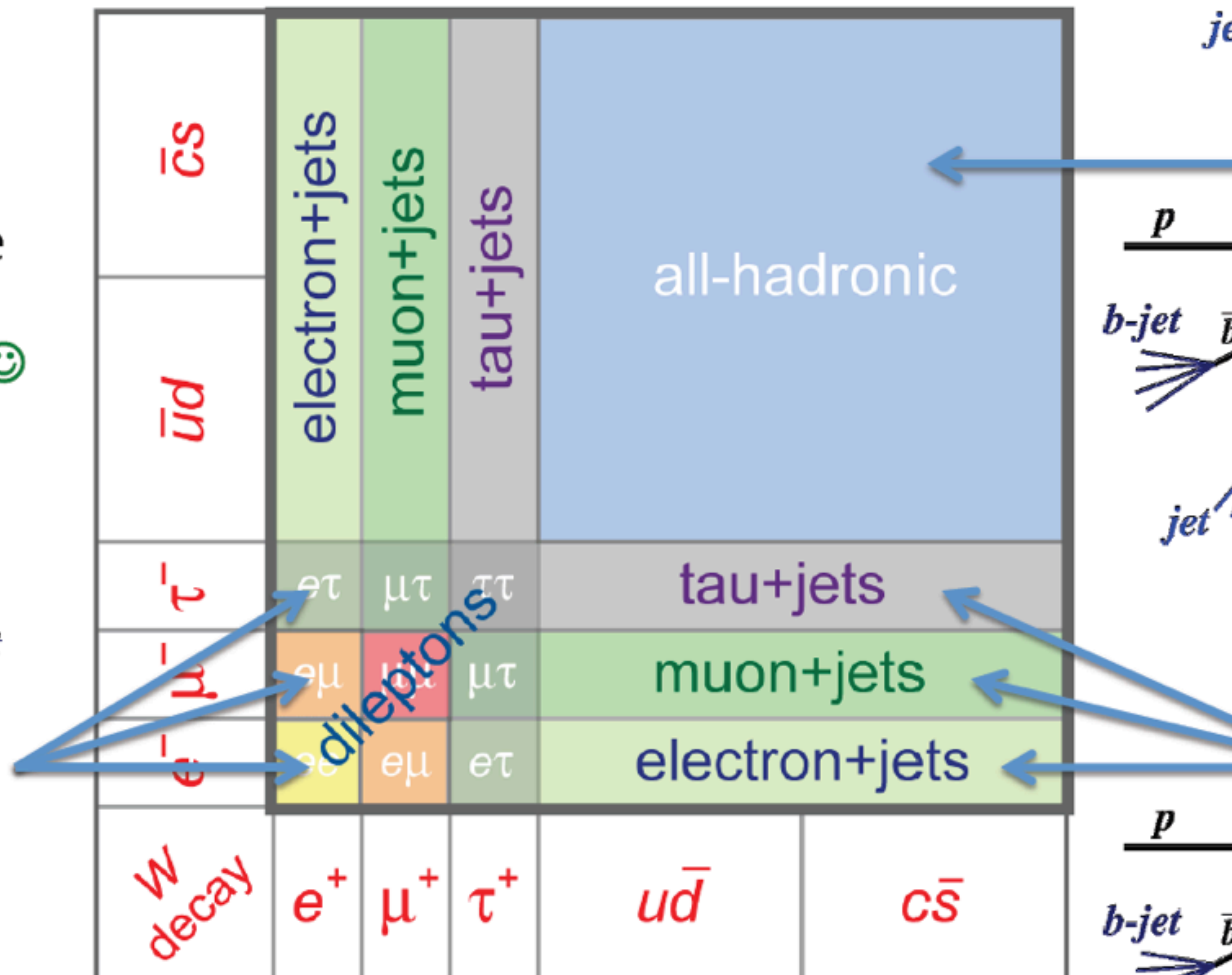



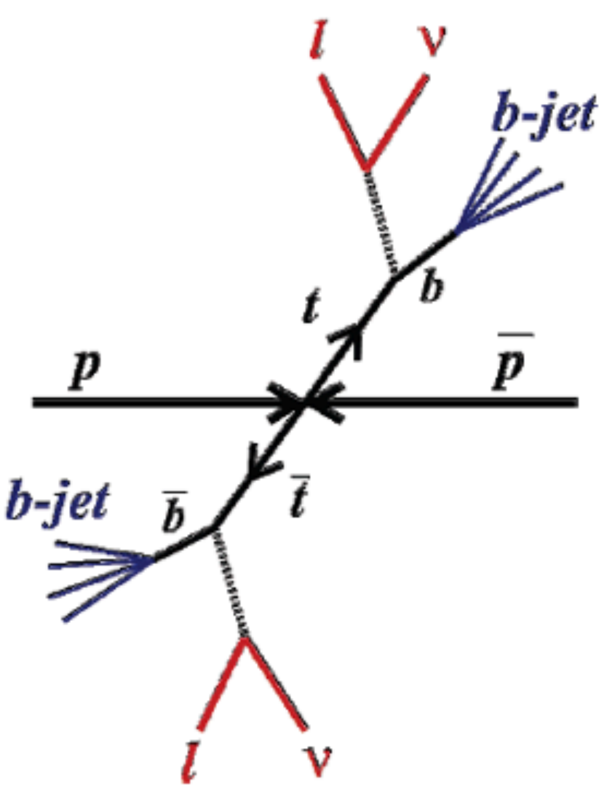

| $\sigma_{tt}$[pb] | Tevatron | LHC (7 TeV) | LHC (8 TeV) |
|---|---|---|---|
| NLO | $6.68^{+0.36+0.23}_{-0.75-0.22}$ | $158.1^{+19.5+6.8}_{-21.2-6.2}$ | $226.2^{+27.8+9.2}_{-29.7-8.3}$ |
| NNLO | $7.00^{+0.21+0.29}_{-0.31-0.25}$ | $160.9^{+11.1+7.2}_{-11.5-6.7}$ | $229.8^{+16.5+9.7}_{-16.7-9.0}$ |
| **NNLL** | $7.15^{+0.21+0.30}_{-0.20-0.25}$ | $162.4^{+6.7+7.3}_{-6.9-6.8}$ | $231.8^{+9.6+9.8}_{-9.9-9.1}$ |

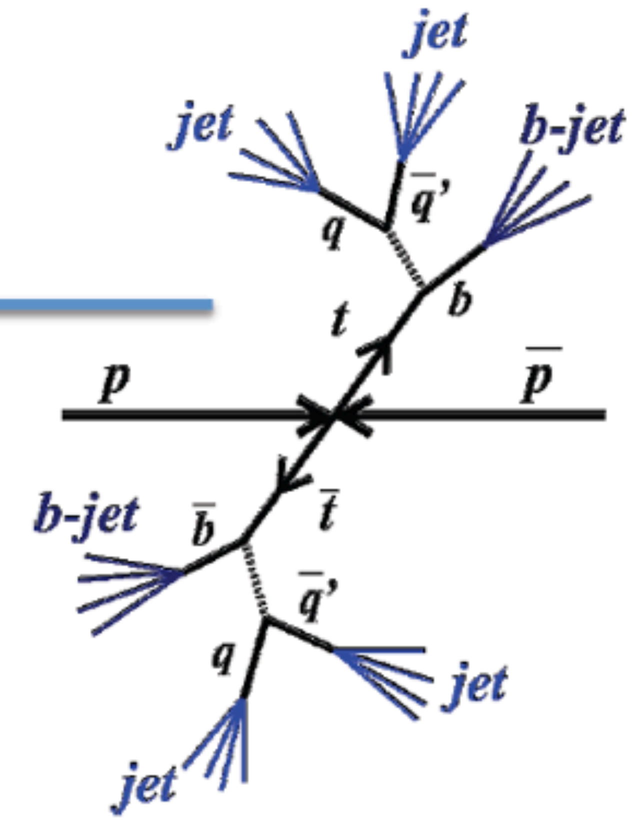
# decay

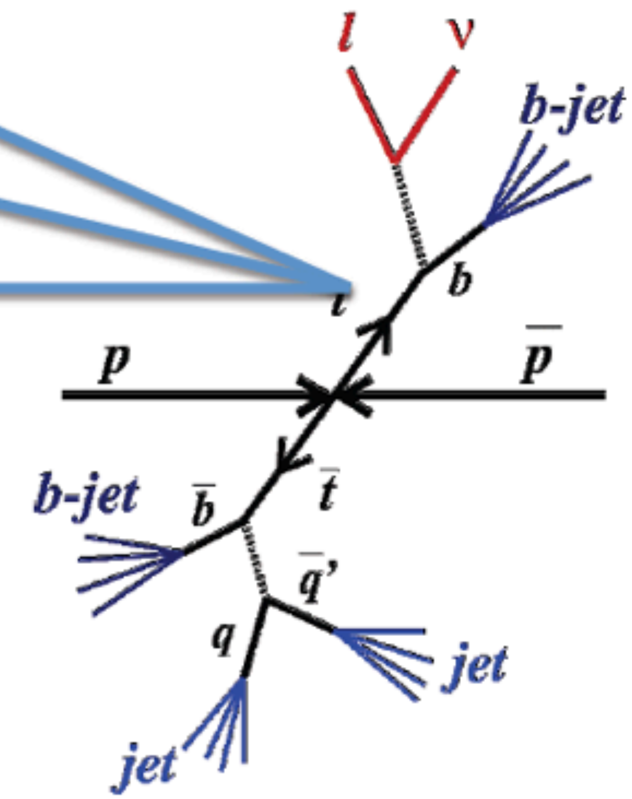
**Decay channels with leptons:**

- Low branching ratio ☹
- Clean signature
- Smaller combinatorics ☺
- Smaller backgrounds ☺



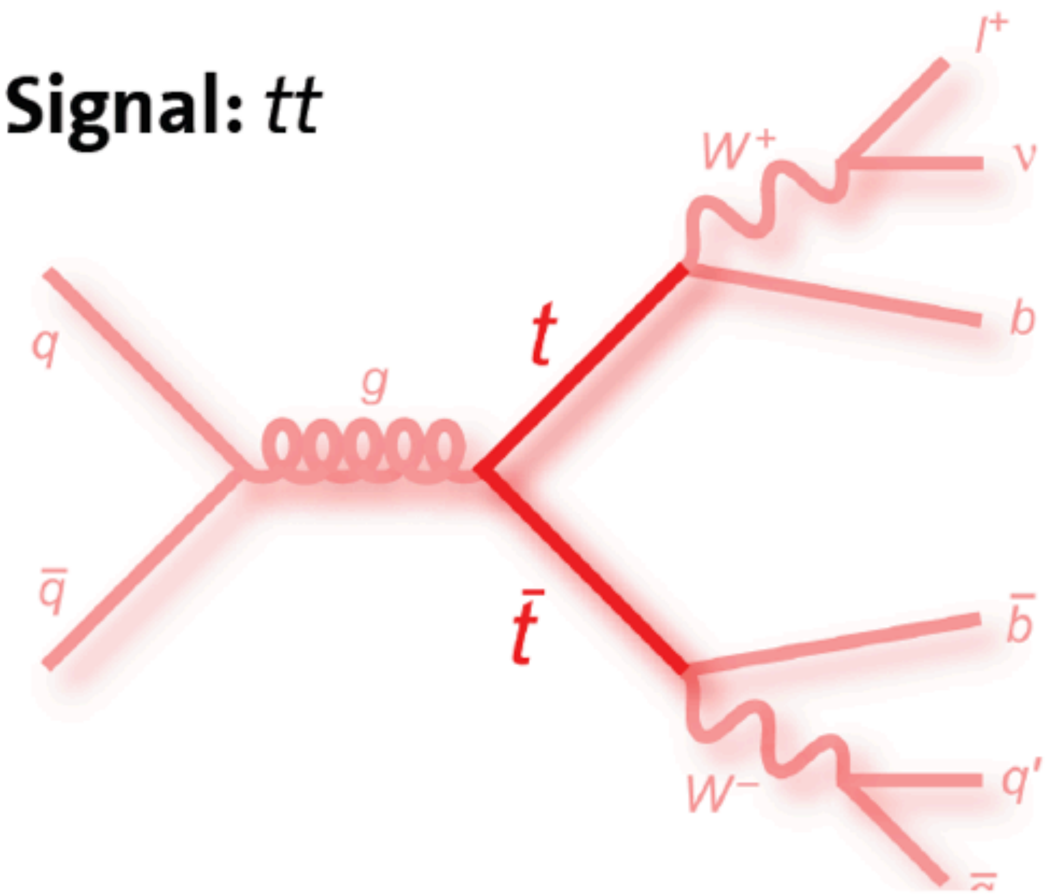



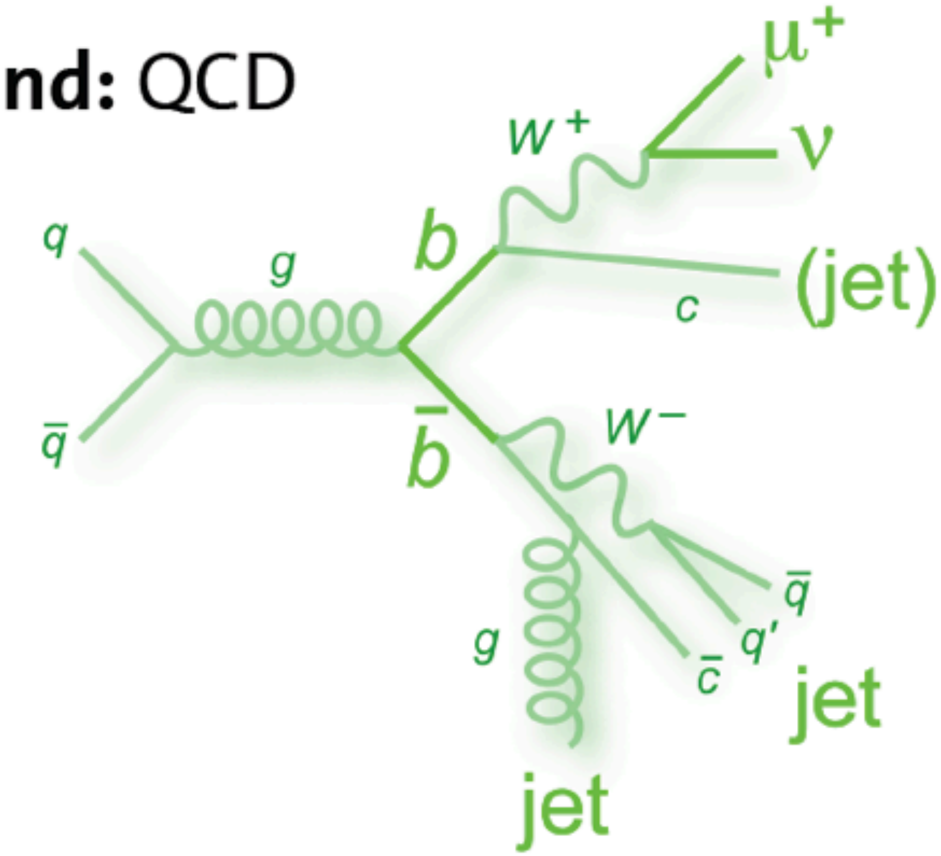



# signal and background

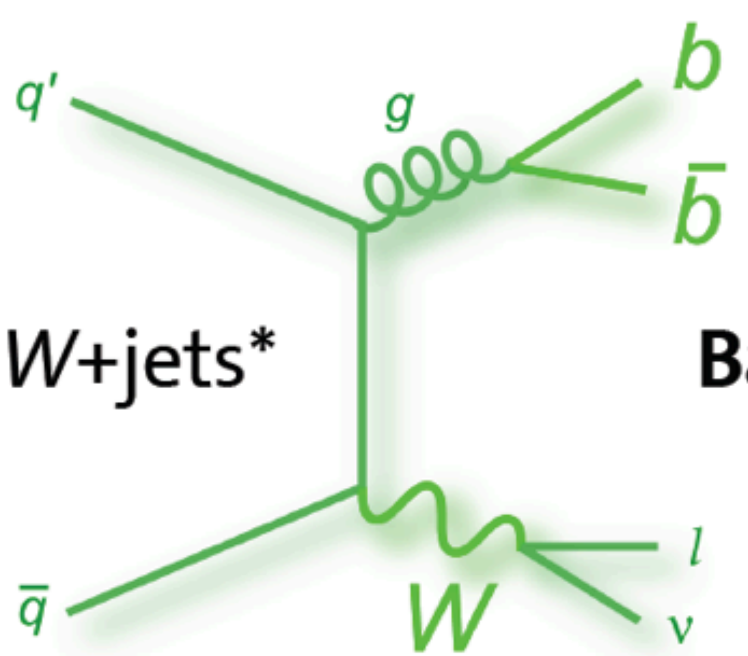
**Signal:** *tt*



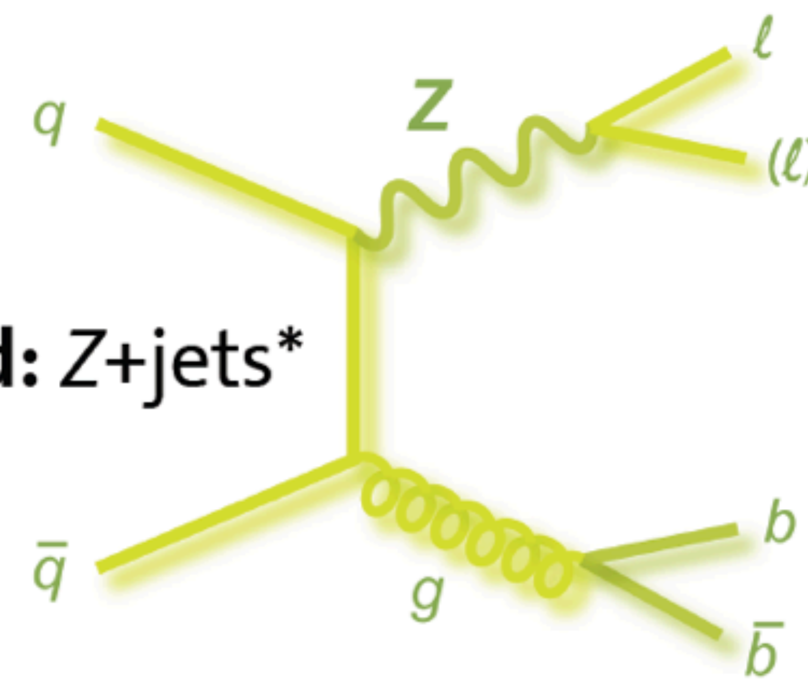

**Background:** QCD




**Background:** *W*+jets*




**Background:** *Z*+jets*




* not necessarily b-jets

# ingredients of an LHC physics analysis

- find out “how” signal and background processes look like
  - study distributions of measured quantities
  - this is only possible with simulation

- reject background events by “selecting” signal-like events

- determine exactly how many “signal” events are lost by this selection
  - == signal selection efficiency
- determine exactly how many “background” events survive the selection
  - == background contamination

- determine other efficiencies for loss of events
  - trigger efficiency
  - b-tagging efficiency
  - electron/muon reconstruction efficiency …

$$\boxed{N_{\mathrm{signal}} = \frac{N_{\mathrm{observed}} - N_{\mathrm{background}}}{\mathrm{efficiency}}}$$

**this week's exercise follows**

**some files and a 6-page documentation of the analysis is uploaded to moodle**

# plan for the exercise

- introduction to the structure of the code will be given now by A. Pozdnyakov
- look around and get yourself familiar with files and code
- read the 6-page documentation
- start working on the first set of questions
- this exercise will continue until the end of the semester, so you have sufficient time

## 1 Warmup

The trigger for this tutorial selects events which contain one or more muons as discussed in the documentation and explanation.

- Find out how often there is more than one isolated, reconstructed muon in data (histogram of the muon multiplicity)! Where could these additional muons come from?
- Calculate the invariant mass of two muons of opposite charge (manually and/or using the ROOT functionality of adding two fourvectors)! Only use isolated muons.
- Display the invariant mass distribution of two muons in a histogram (hint: try different axis ranges)!
- Compare your results to MC simulation (display simulation and data in the same histogram). Make sure you select triggered events only for the simulated samples!