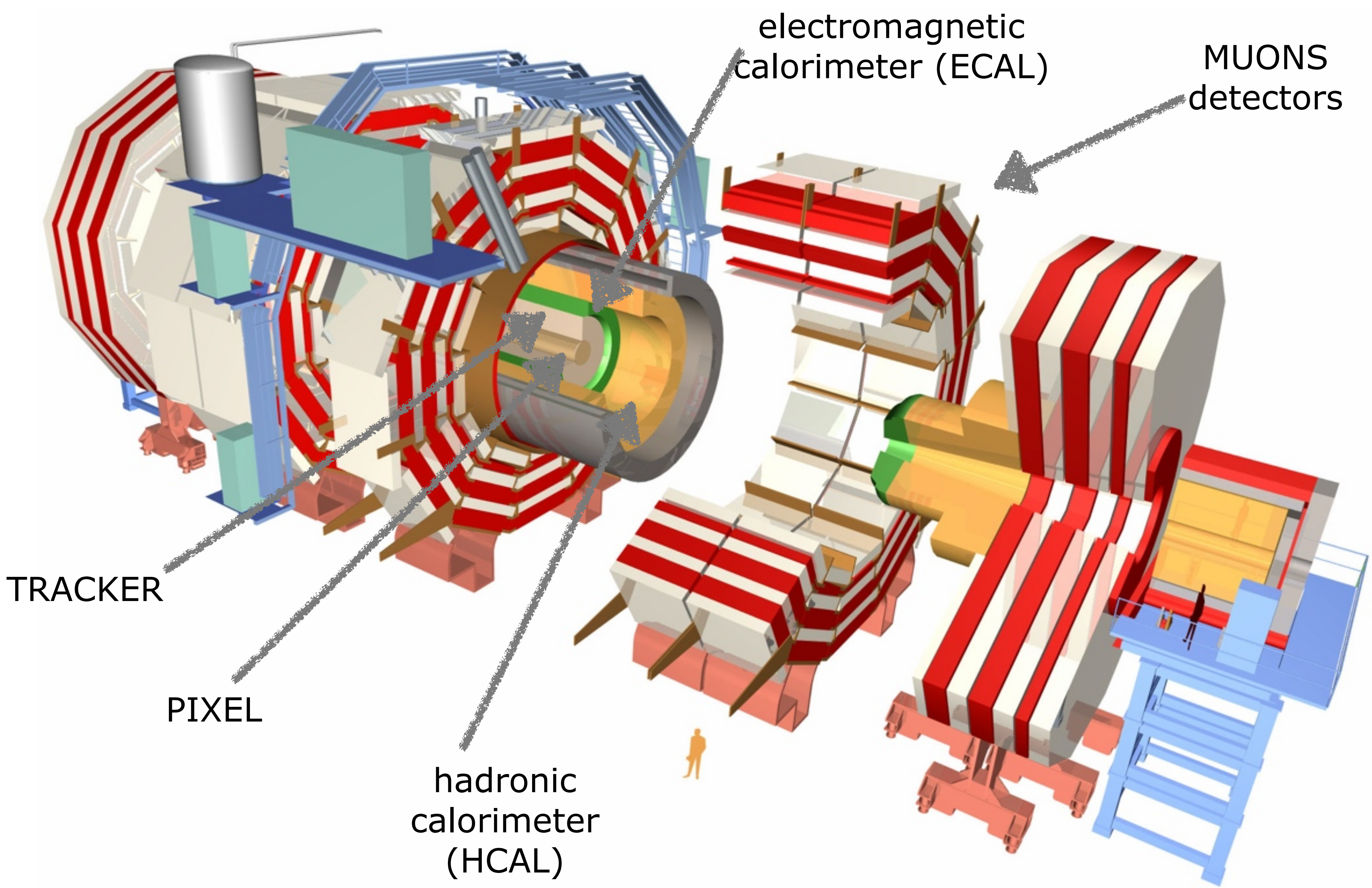
introduction

- my interpretation of “data analysis techniques” is here “doing a data analysis”
- follow the steps from the beginning (data taking) to the end (the result)
  - the luminosity
  - the trigger, from the point of view of the analysis
  - the reconstruction and detector response
  - the simulation
  - differential cross-section measurement: a di-jet correction
  - searches: the H > WW > lvlv
  - multivariate techniques

thanks to the following people, for interesting discussions, for liberally “borrowing” slides, or both: D. Benedetti, C. Bernet, T. Camporesi, G. Cowan, K. Cranmer, K. Ellis, S. Gennai , A. Ghezzi, A. Hoecker, R. Van Kooten, M. Nguyen, M. Paganoni, M. Pelliccioni, E. Rizvi, R. Rossin ...

physics objects reconstruction

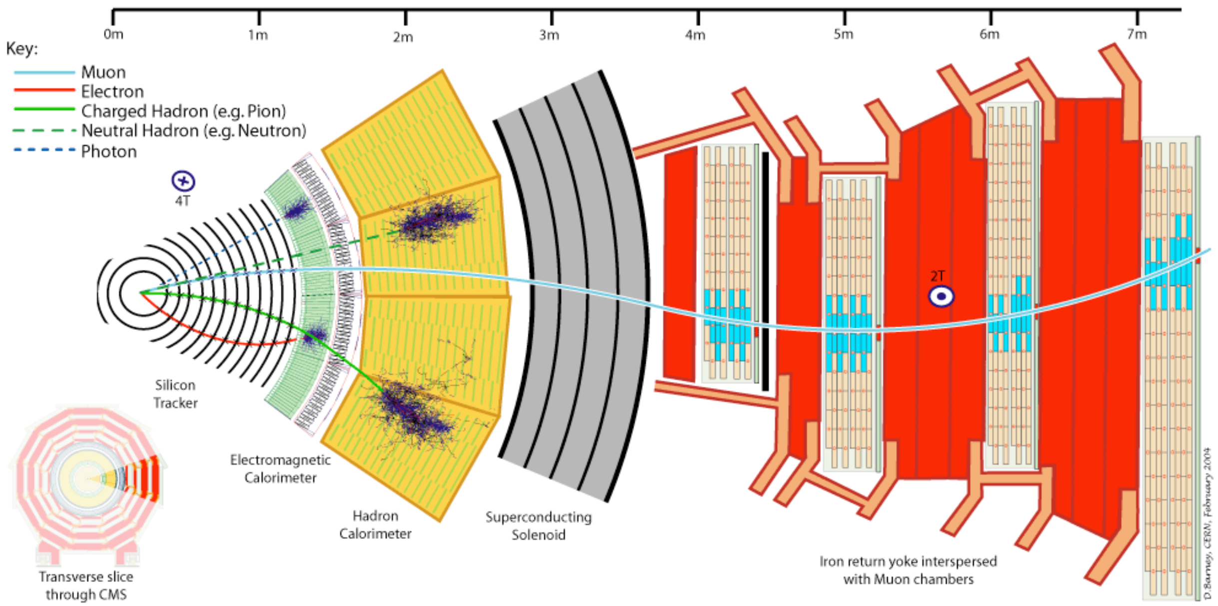
the cms detector



physics objects reconstruction

- obtain **physics objects from the detector response**
  - hits in the tracker and muon detectors
  - energy deposits in the calorimeters
- two ways are available in CMS
  - **single objects** reconstruction: build final objects (e.g. muons, electrons, jets) from the detector response
  - **particle-flow** reconstruction: build a coherent list of stable particles and produce the analysis objects on top of them

the particle flow



hits in the tracker → tracker tracks

cells in the calorimeter → calorimetric clusters

hits in muon detectors → muon tracks

link the single objects with geometrical requirements on the extrapolated trajectories and create **blocks**

the particle flow

blocks

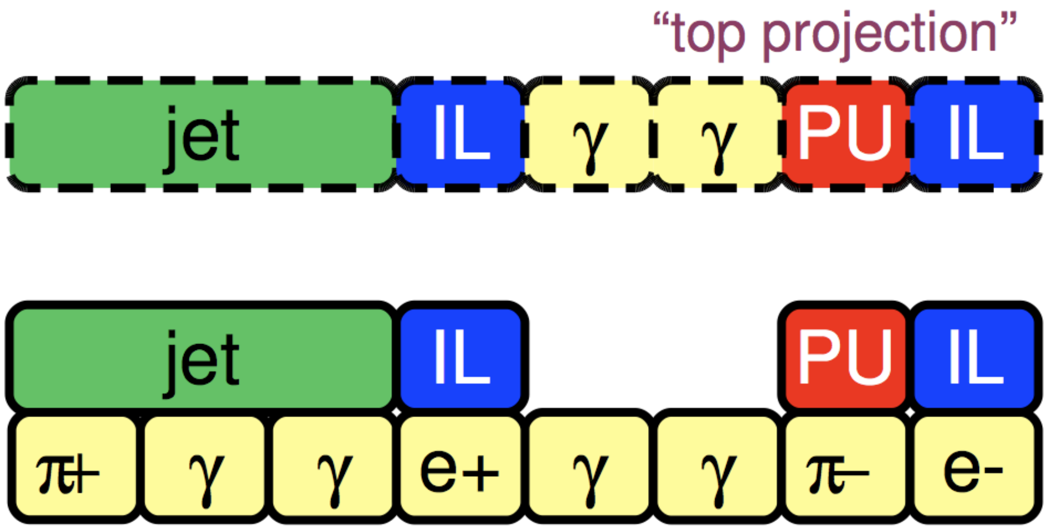
identify muons, promote the block

identify electrons, promote the block

match the remaining tracks to clusters, define charged hadrons and neutrals from calorimetric excess

the remaining calorimetric deposits define neutrals

the charged energy contribution is measured well from the tracker

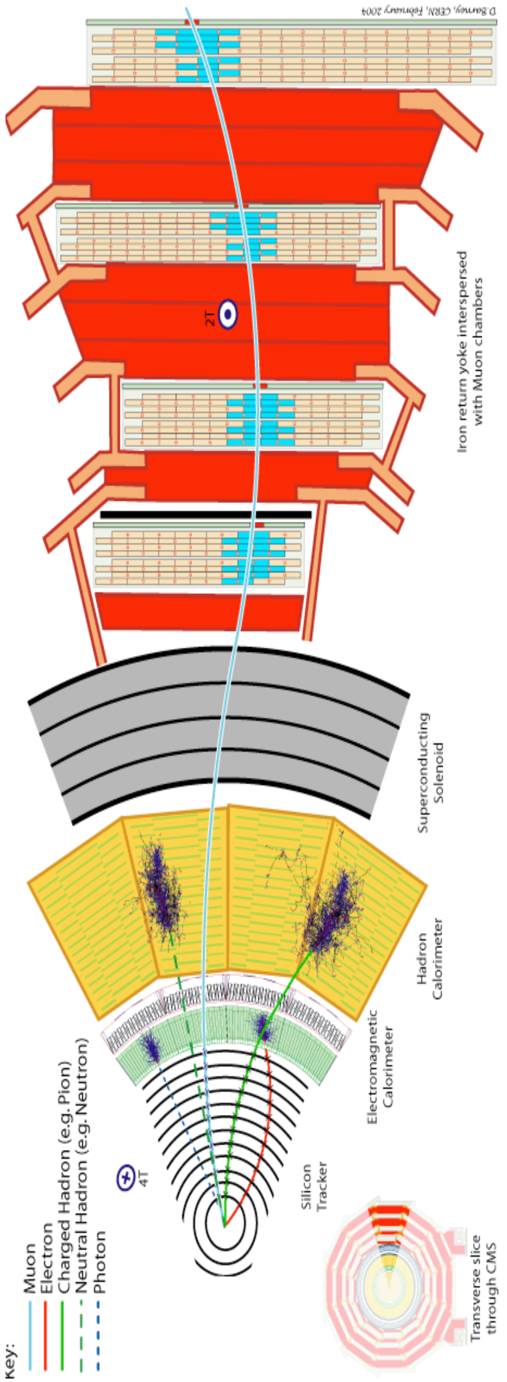


$$+ \; \overrightarrow{\mathrm{MET}} = -\sum_{i=0}^{N_{particles}} \vec{E}_T^{\,i}$$

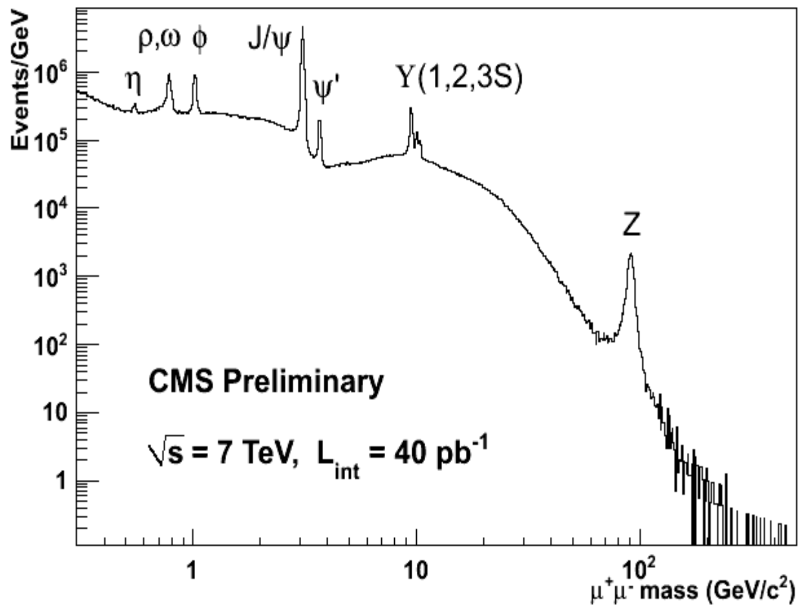
The list of PFCandidates

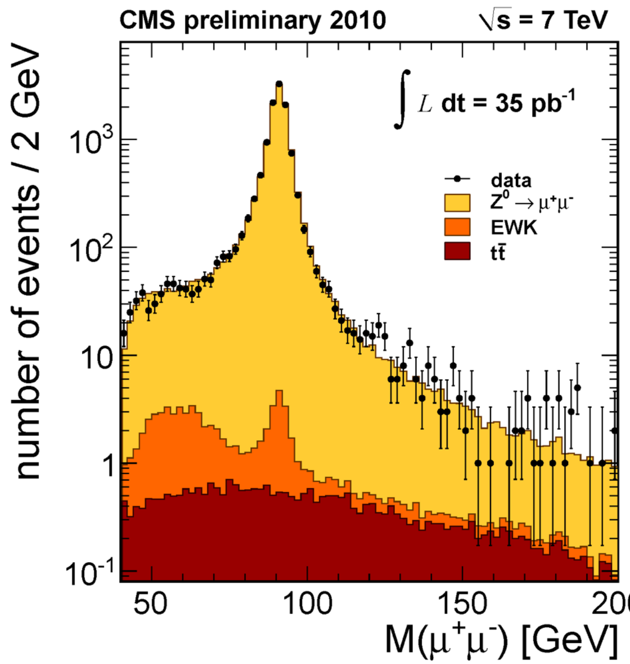
the list of particles obtained (candidates) is used for high level objects classification and reconstruction, to be used in the analysis

muon reconstruction

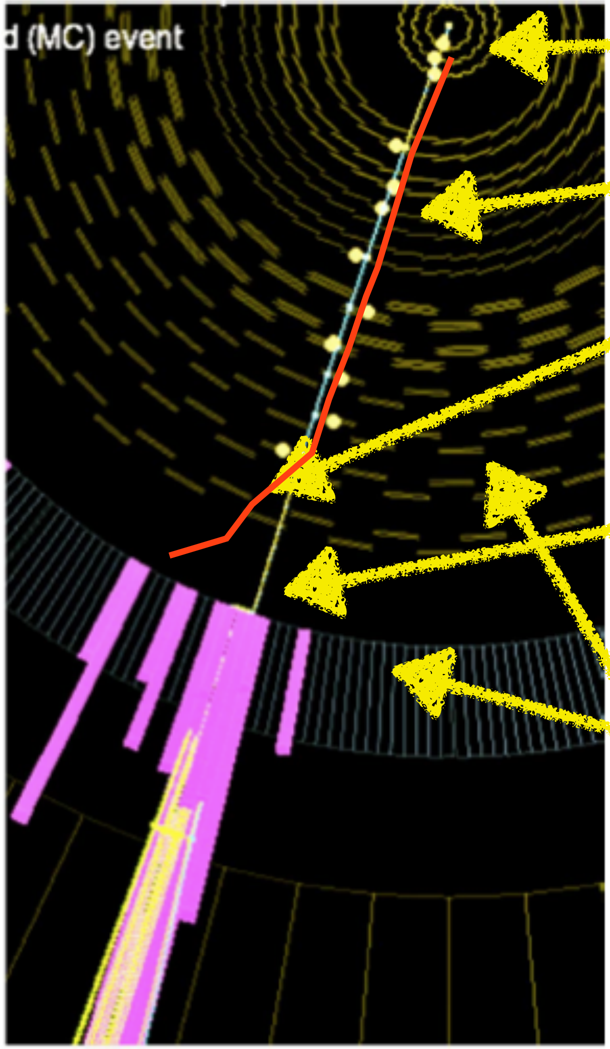


- high **purity** = fit with hits in both tracker and muon
- high **efficiency** = fit in the tracker + confirmation in the muon detector
- **momentum determination** from both tracker and muons information: best resolution from the tracker for p_T < 200 GeV, from the muons above (effect of multiple scattering)
- above 1 TeV, the **bremsstrahlung** is significant





electron reconstruction



Seeding

electron tracking: GSF

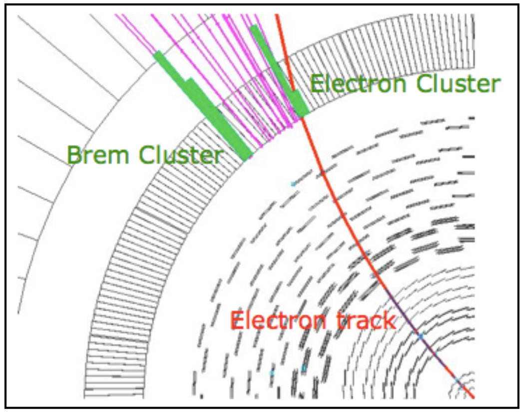
cleaning of GSF track duplicates

Identification of all the electron energy deposits in the ECAL

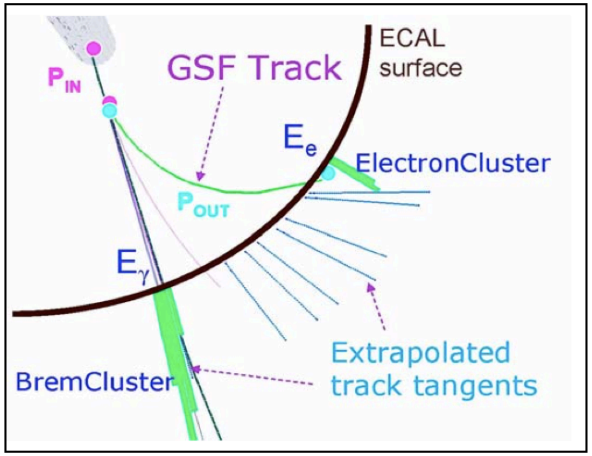
electron 4-mom determination.

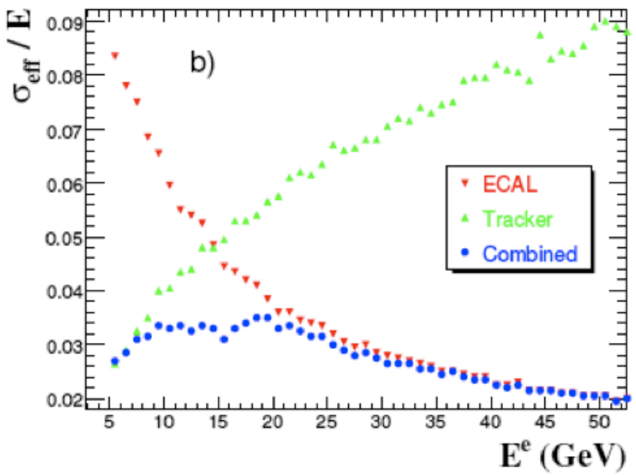
Identification against charged-particles interacting in the ECAL

from ECAL clusters or tracks:



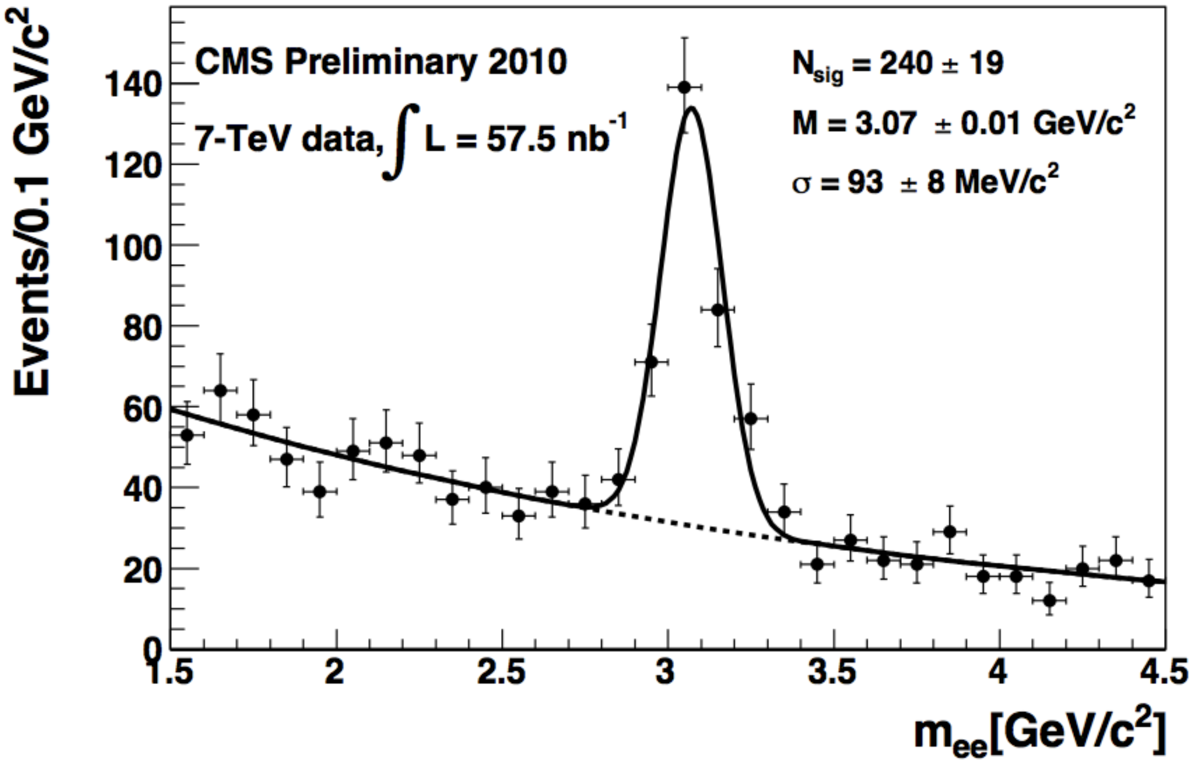
from ECAL footprint or tracks extrapolation:





use ECAL at high p_T, tracker at low p_T

electron reconstruction



search for the decay:

$$J/\Psi \to e^+e^-$$

contamination sources:

- real electrons, either from photon conversions or from semi-leptonic b-hadron decays,
- mis-identified charged hadrons.

- at most one hit missing in the pixel detector (reduce conversions)
- electrons originate from the same vertex (reduce the b-decay background)
- quality cuts to reject charged hadrons contamination
- opposite charge

photon reconstruction

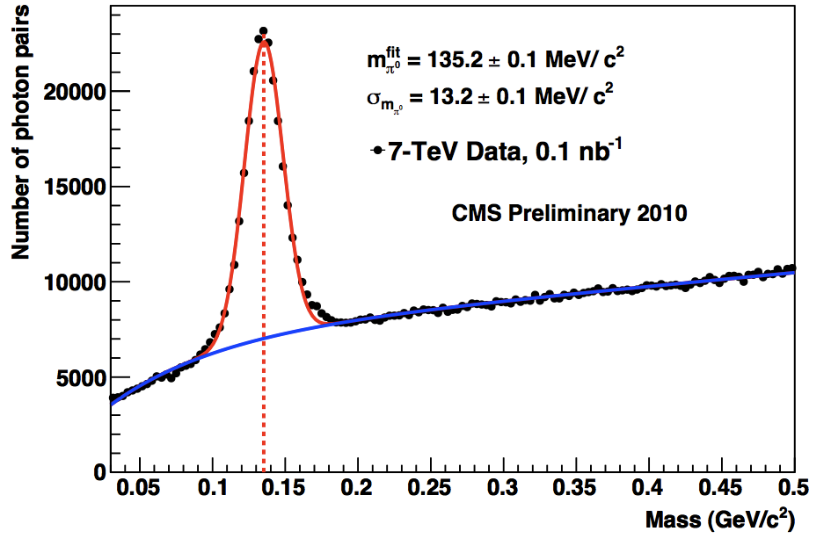
- **ECAL clusters** not associated to a track, nor a deposit in the hadronic calorimeter

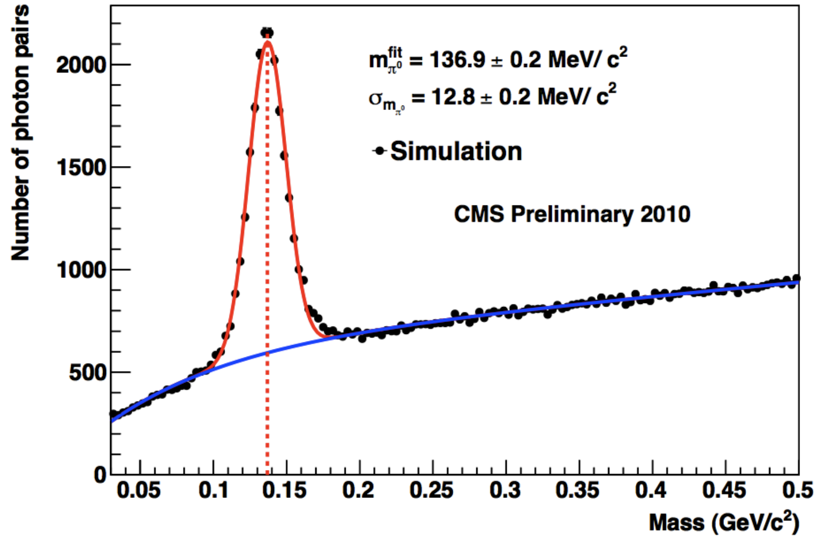
- ECAL detector response is **calibrated**, to account for the effect of the noise cut on the single crystals readout

- **check the photons energy scale** calibration with 2010 data, by looking at the π^0 peak position

- pair all photons with at least 400 MeV energy

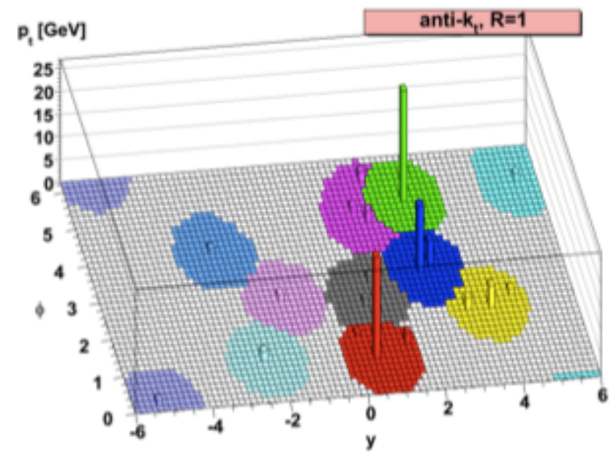
- determine the peak position with a combined fit of signal + background

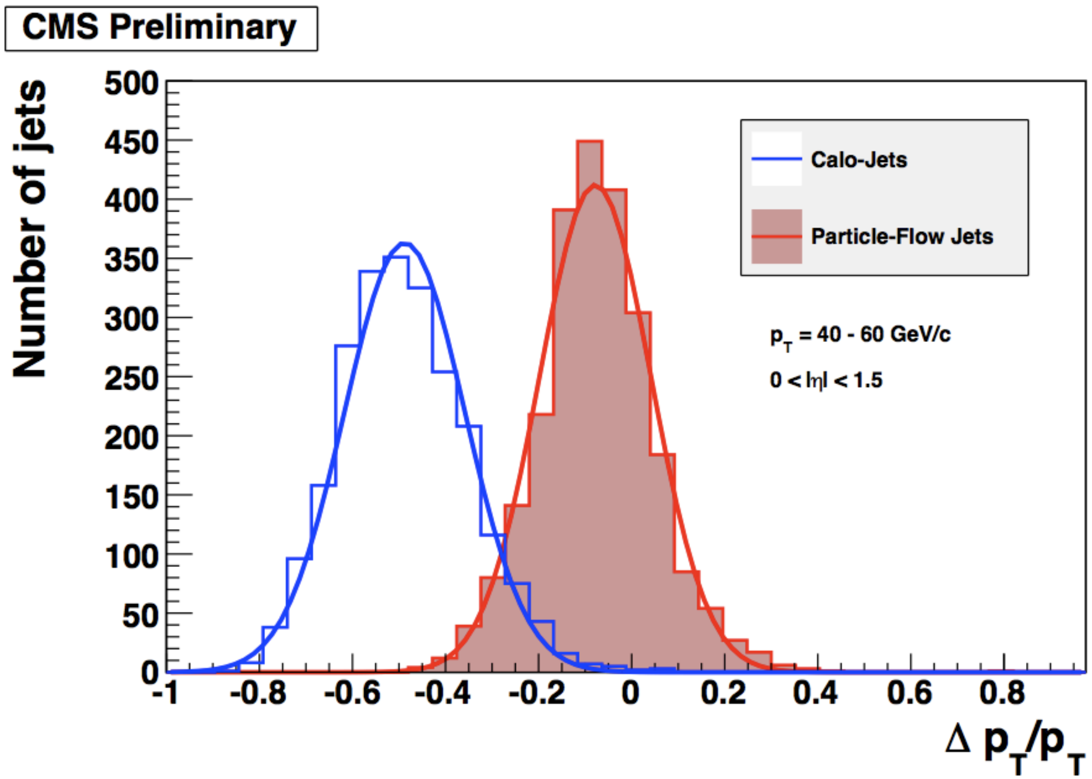


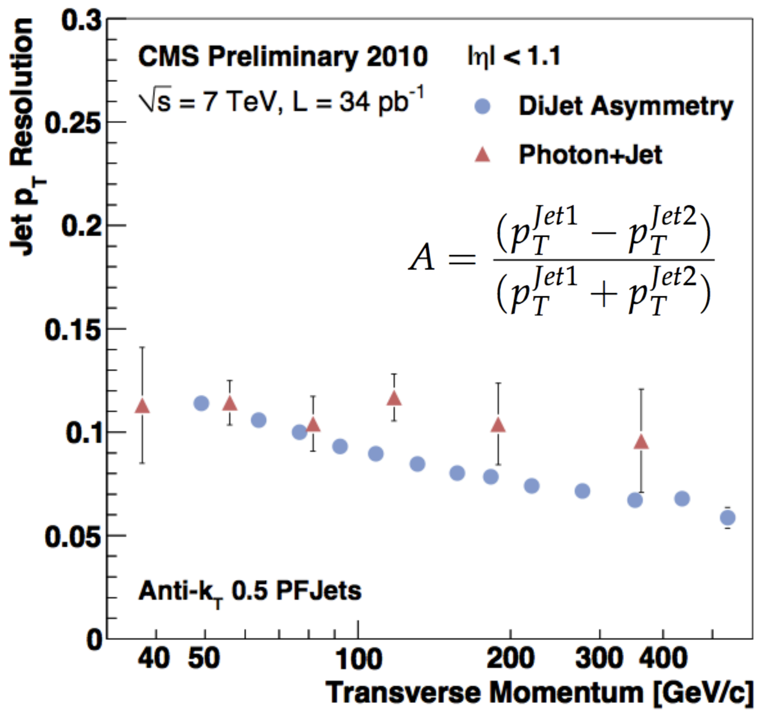


jet reconstruction

- jets are reconstructed with the AKT5 algorithm
- for the single object reconstruction: with calorimetric deposits
- for the particle-flow: with particle flow candidates



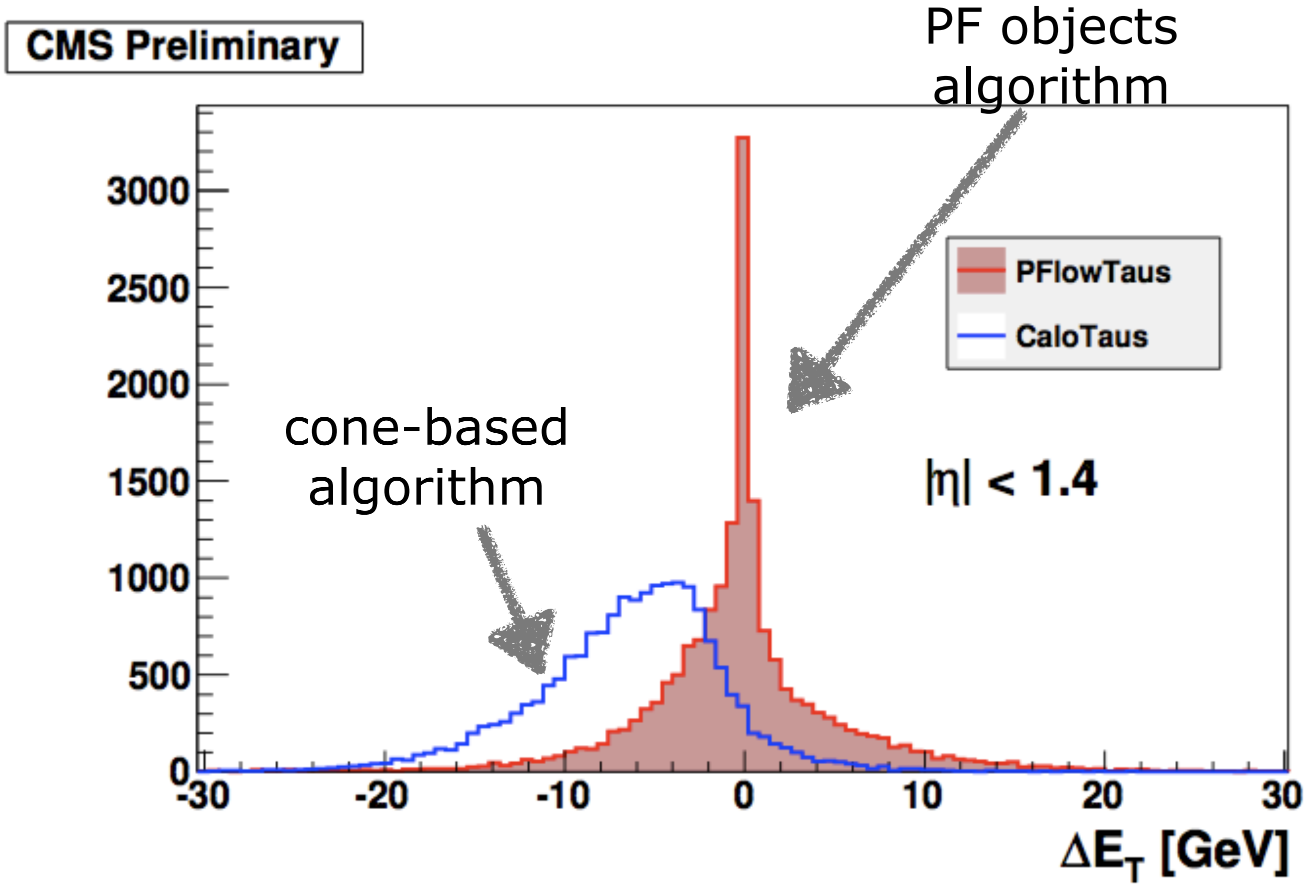




M. Cacciari, G. Salam, **The anti-kt jet clustering algorithm**

tau reconstruction

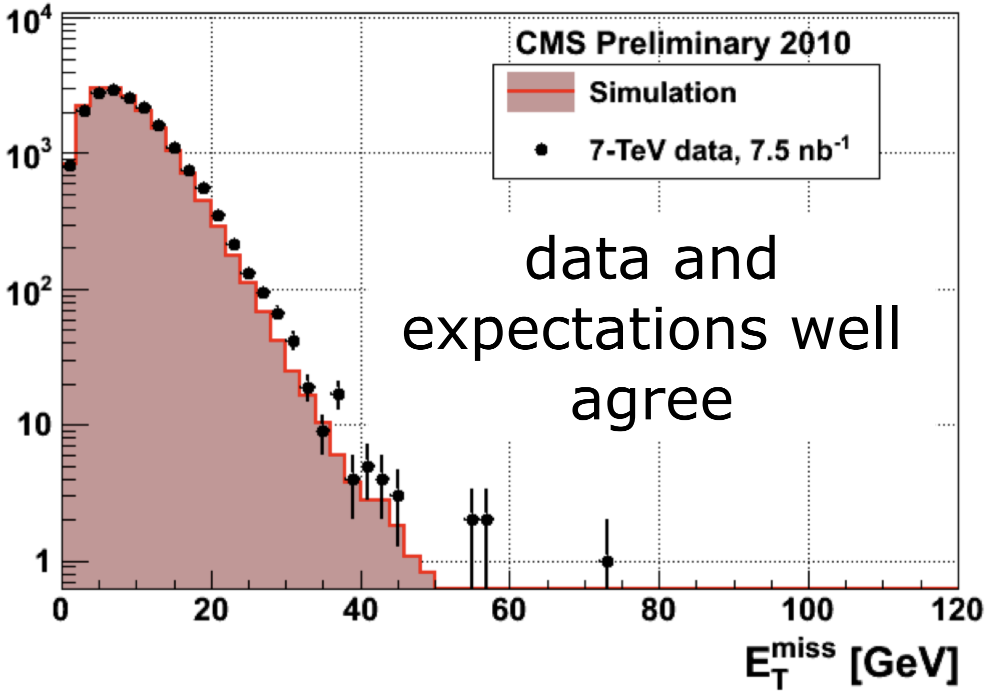
- reconstructed as narrow jets in the standard case, as the sum of the particles compatible with the tau decay in a narrow cone in the particle flow case

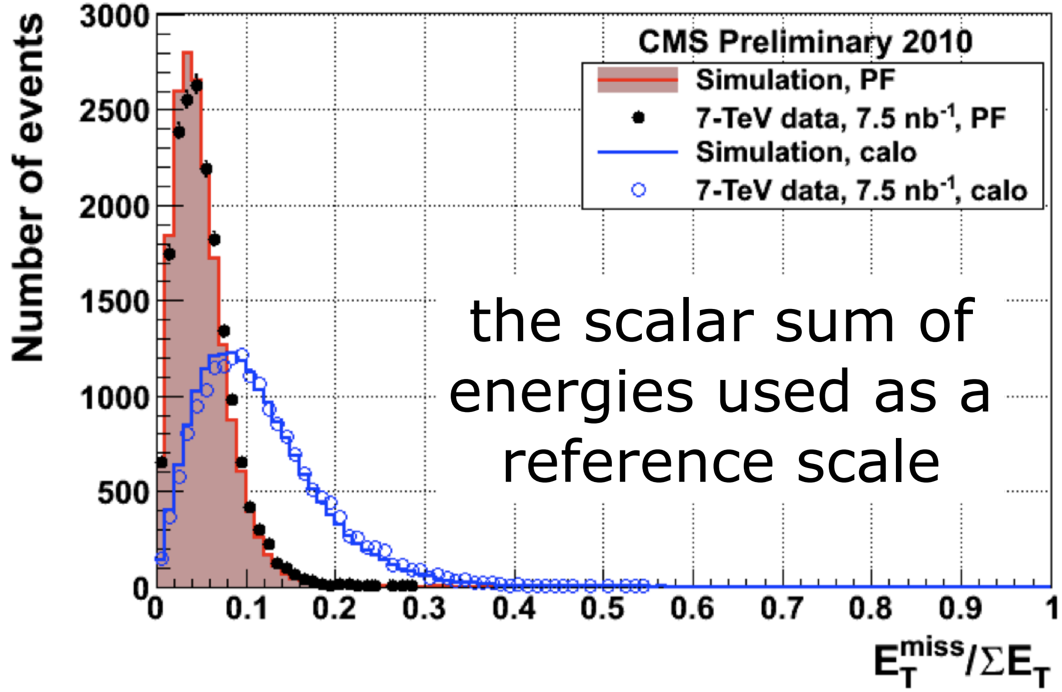


reconstructed taus E_T compared to the expected one, test performed on a simulated $Z > \tau\tau$ sample

missing energy reconstruction

- derived from (minus) the sum of “all the rest”
- sensitive to uncertainties in all the other physics objects
- noise effects, mis-calibrations, etc. generate fake missing energy in events without missing energy
- perform a test on a di-jet sample

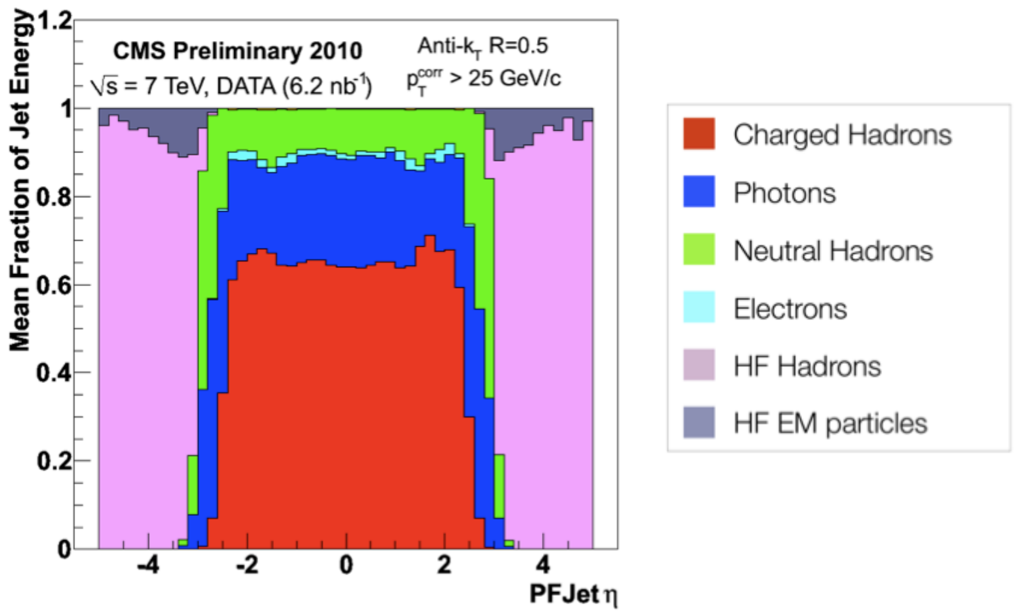




reconstruction: in summary

- the reconstruction obtains from the detector measurements the physics objects in the final state
- in a coherent way, to close the kinematics (as much as possible)
- making use of the most precise sub-detector
- reconstruction and identification are not (always) disentangled, for example electrons need to be separated from jets
- data-driven techniques necessary to assess the performances

jet composition: only for neutral hadrons one cannot profit of tracker measurements



detector response

- the detector response is not perfect
- the output of the reconstruction needs to be **calibrated for the detector response**
- use **known physics processes** to get the calibrations and the relative uncertainty
- for example
  - **resonances** for leptons (energy scale, tag&probe)
  - **cosmic** rays (alignments)
  - transverse momentum **balances**
  - ....

ECAL calibration

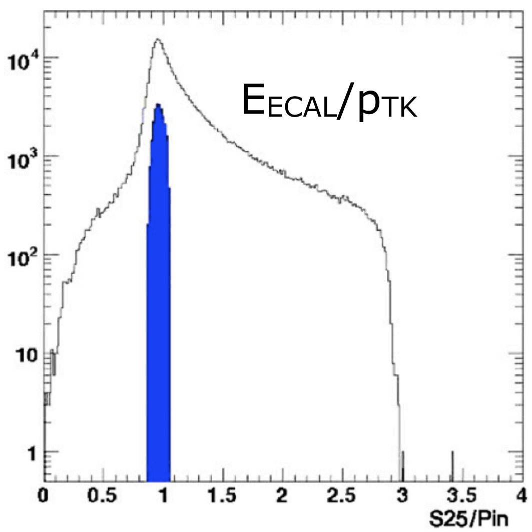
- each ECAL channel needs a calibration factor to equalize the response of all detector elements
- for electrons, the energy is measured in the tracker and in the ECAL
- find the calibration coefficients by minimizing a X^2 of:

energy in single elements

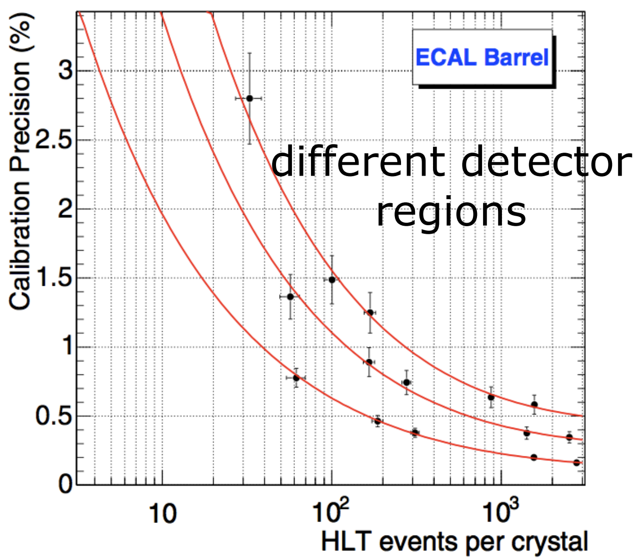
$$|\mathcal{E} \times \bar{c} - \mathcal{P}|$$

electrons momenta

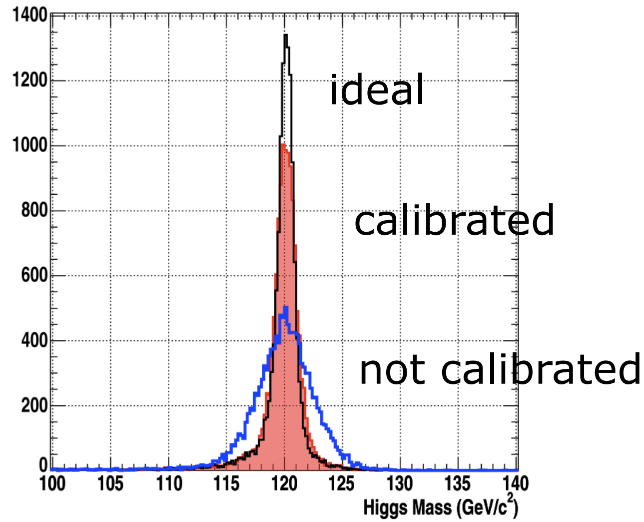
unknown coefficients



select good isolated electrons



statistical trend with luminosity



effect on $H \rightarrow \gamma\gamma$ invariant mass

jet energy corrections

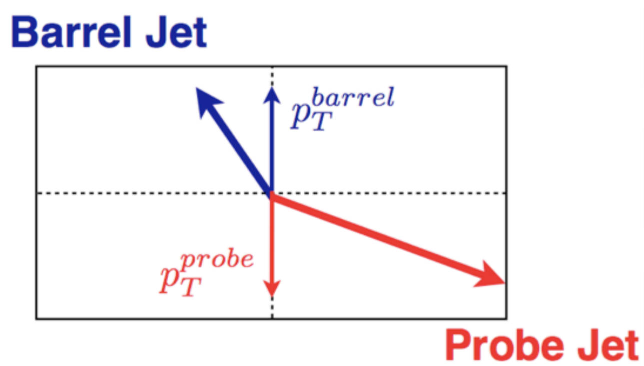
- the jet energy scale needs to be calibrated, as a function of various variables

Mandatory			Optional			
L1: Offset	L2: η Relative	L3: p_T Absolute	L4: "EMF"	L5: Flavor	L6: UE	L7: Parton

detector noise effects, pile-up

tag&probe like: di-jets events assumed to be balanced, get a relative correction

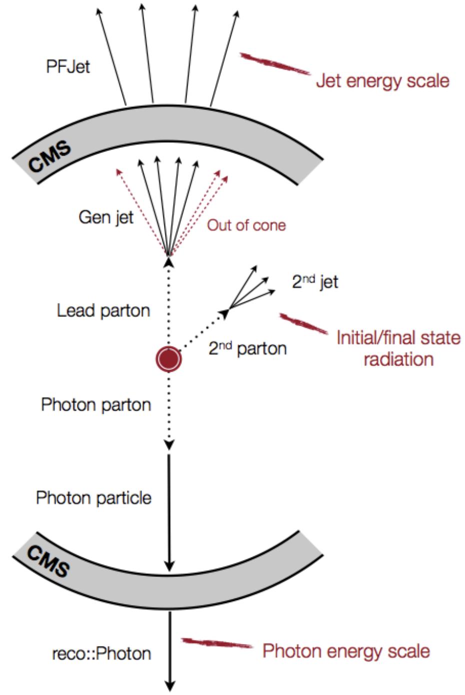
ϒ+jet balance in the transverse plane



$$p_T^{dijet} = \frac{p_T^{probe} + p_T^{barrel}}{2}$$

$$B = \frac{p_T^{probe} - p_T^{barrel}}{p_T^{dijet}}$$

$$r = \frac{2 + \langle B \rangle}{2 - \langle B \rangle}$$



the simulation

the simulation

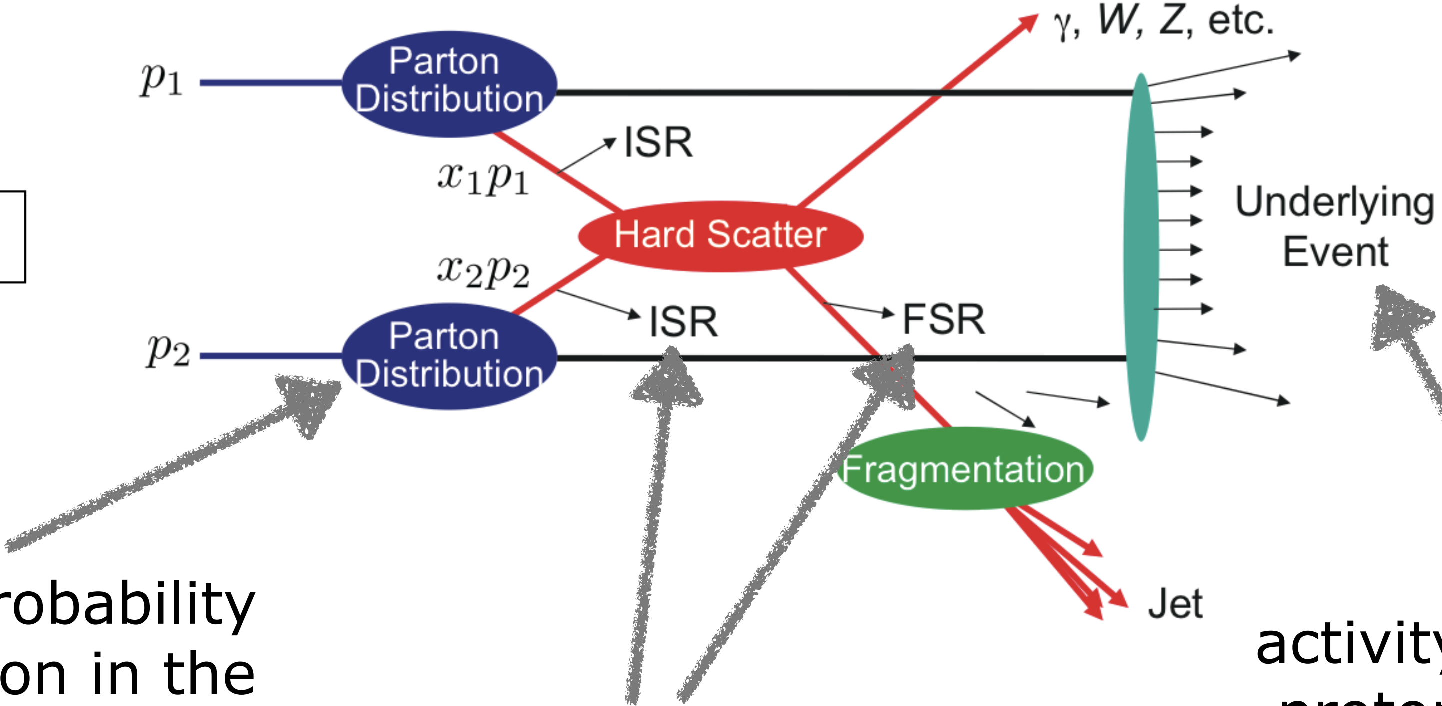
$$\sigma = \frac{N_{obs} - N_{bkg}}{\varepsilon \cdot \int \mathcal{L} dt}$$

$$\varepsilon = \varepsilon_{tr} \cdot \varepsilon_{reco} \cdot \varepsilon_{ID} \cdot \varepsilon_{sel}$$

- calculate what fraction of events from a given decay **falls within the detector acceptance and the selections** of the analysis
- need a **forecast of how the event develops** in space, after the interaction
- the **simulations** are necessary both for known physics objects (Z, W production) and, of course, to build searches for new physics
- the **uncertainty** in the input parameters is source of systematics

the simulation

- calculate **inclusive cross-sections**
- calculate **differential cross sections** as a function of variables of interest in the analysis
- provide **simulated events**, that mimic Physics, and have on average the behaviour foreseen by the theoretical model



LHC

detector

parton probability distribution in the proton

radiation in the process

activity due to the proton remnants

the physics event generation

add the underlying event

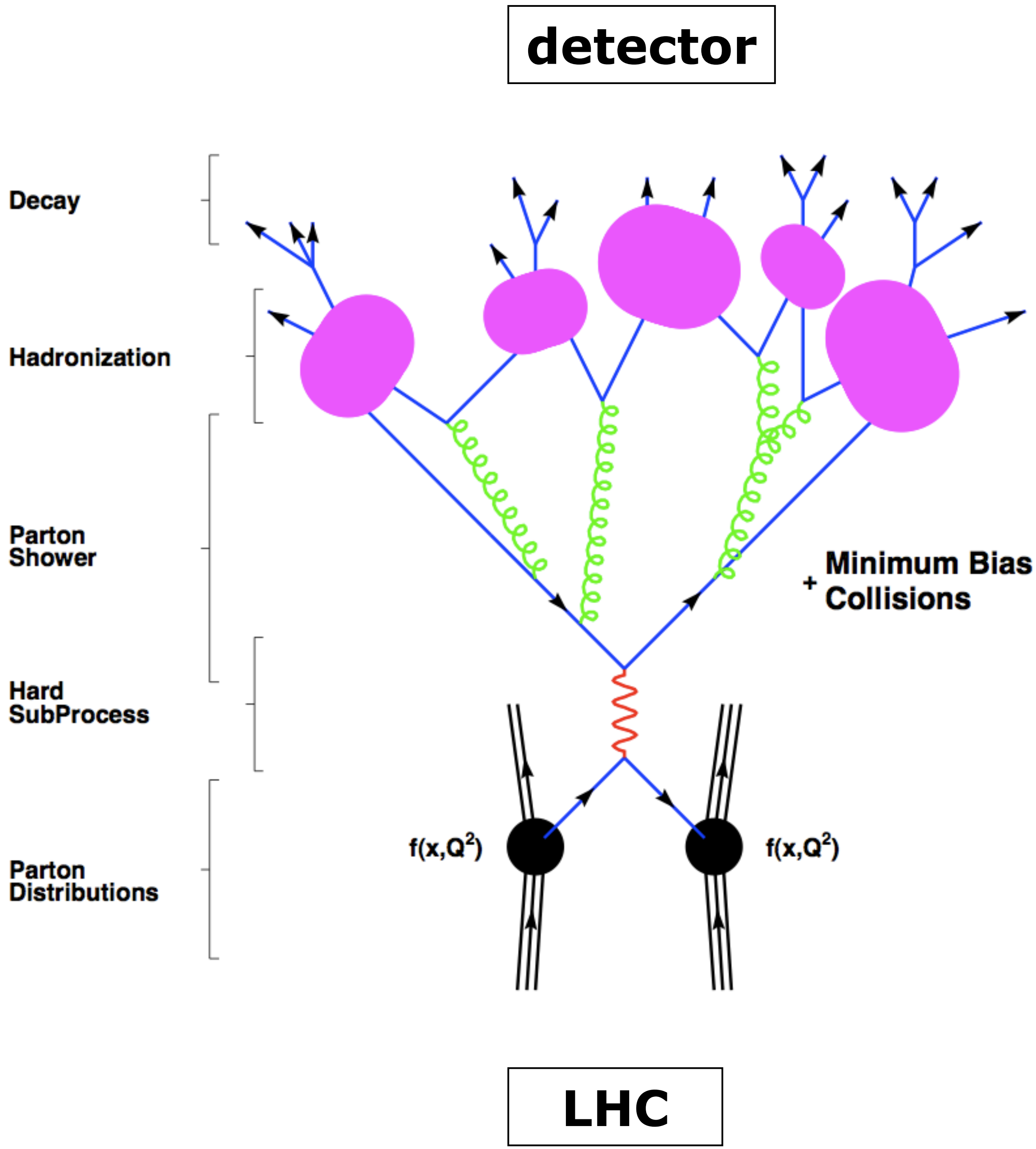
let hadronic decay

hadronize partons

add the parton showers

add initial and final state radiations

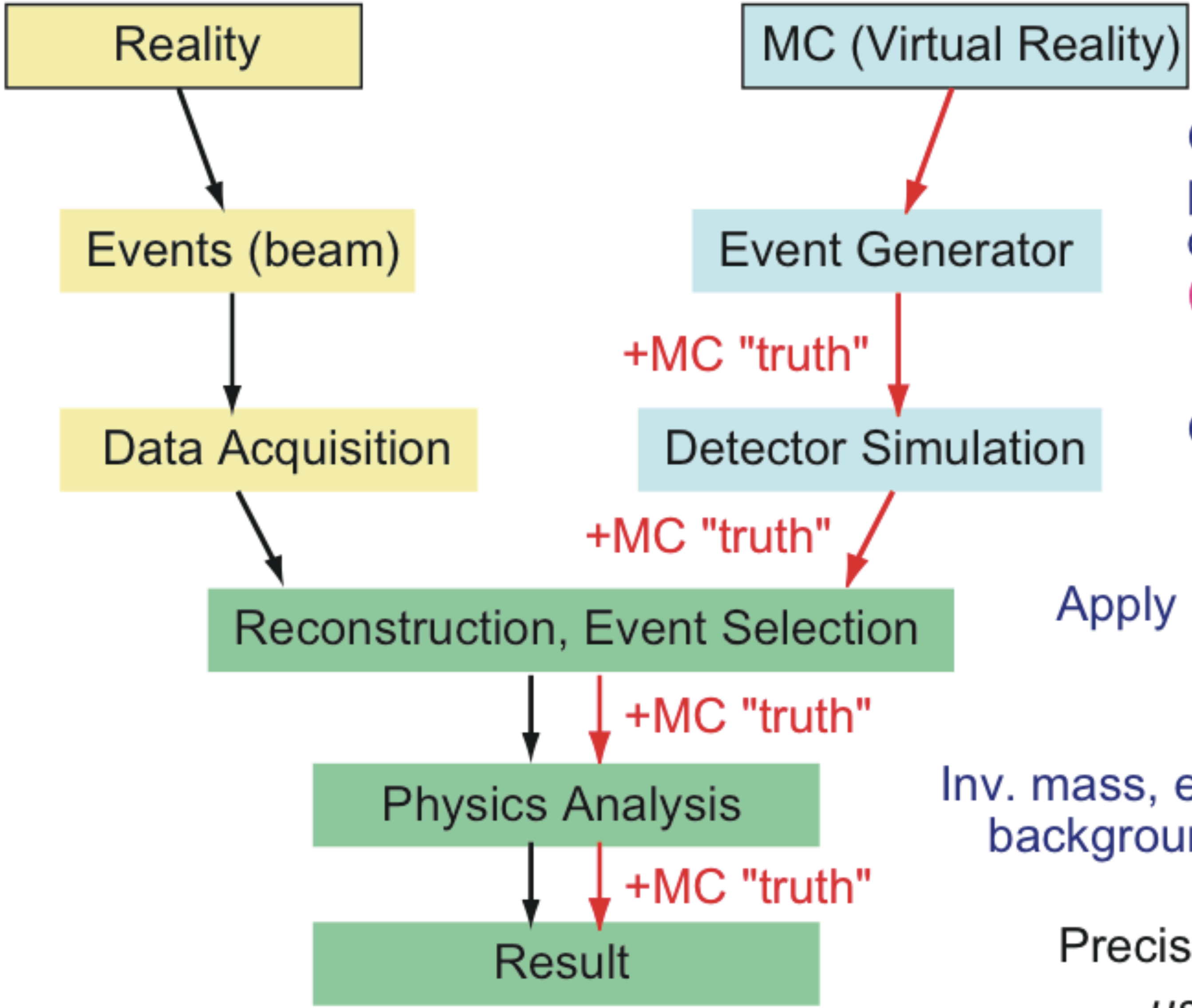
generate hard process



the simulation of the detector

- each experiment creates a **simulation of the detector**
- the GEANT program uses generator output (4-vectors) and simulates the **interaction of particles within the detector volume** (need a good description of the geometry):
  - particle ionization in trackers
  - energy deposition in calorimeters
  - intermediate particle decays/radiation
- the GEANT code is merged with (experiment specific) **detector simulation**
- final output: the response of the electronics readout
- MC events are in the same format as real raw data → **MC events are in the same format as real raw data**

the samples processing



Generates 4-vectors for the particles, resonances, ang. dist., decays, etc. (PYTHIA, HERWIG, ALPGEN, Sherpa...)

Generates detector relevant quantities (GEANT 4)

Apply boundary conditions Acceptance

Inv. mass, efficiency, purity backgrounds, *any* dist.

Precision $\approx 1/\sqrt{N}$

usually:

$$N^{\mathrm{MC}}_{\mathrm{signal}} \gg N^{\mathrm{Data}}_{\mathrm{signal}}$$

$$N^{\mathrm{MC}}_{\mathrm{backg}} \ll N^{\mathrm{Data}}_{\mathrm{backg}} \text{ (QCD!)}$$

levels of simulation

Three typical levels of MC simulation:

- Full

  Particle → Energy Deposit → Detector Response → Electronics → Analog Signal → Digitization

  Time consuming, smaller samples

- "Fast" or parameterized

  Intelligently smeared 4-vectors, effiiciencies, noise (from data and full MC)

  And/or calorimeter shower libraries

  Larger samples

- Toy

  Only throw from the handful of prob. dist. functions that you care about (with correlations)

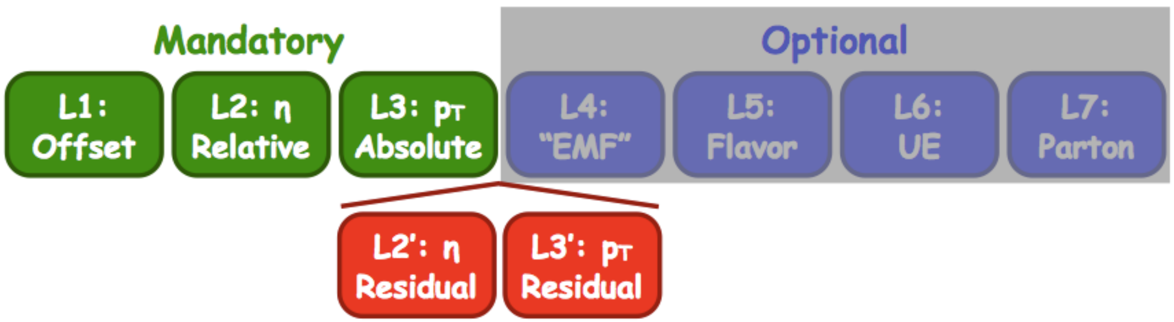
  "Roll your own", usually write (easy in root!) and run yourself

  Crazy-large samples, quickly

  To determine probability of fluctuations, checks for systematic effects, etc..

beware of your simulation

- the simulation is a multi-dimensional **parametrization** of the knowledge of the detector and standard model predictions
    - is the theoretical simulation correct for the analysis?
        - additional jets production is crucial for analyses that apply a jet veto
        - spin correlations in the Higgs decay need to be treated correctly
    - is the behaviour of the simulation in **agreement** with data, in the phase space of interest for the analysis?

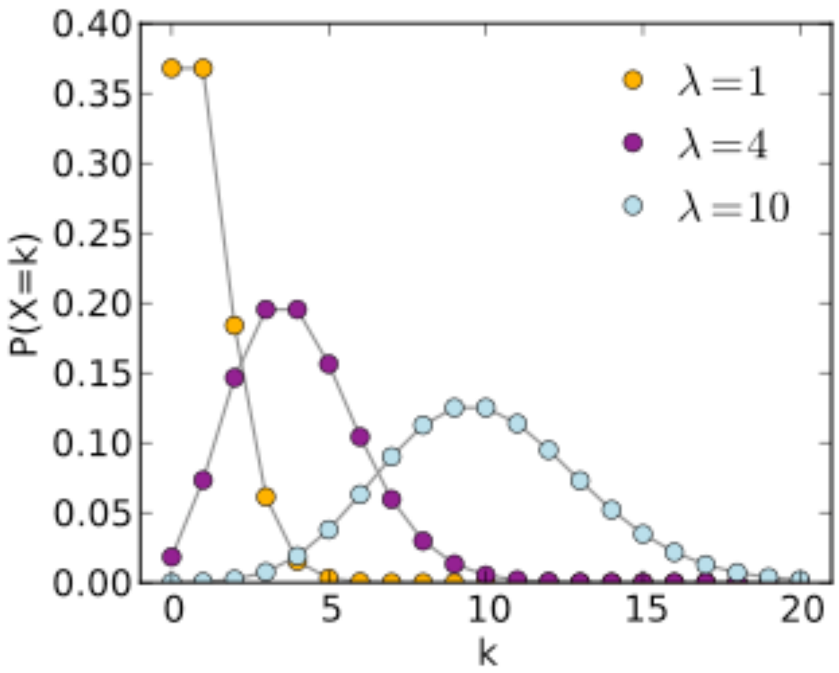


when there's agreement, **use it**: the jet energy scale at CMS is calculated as a correction factor to the one obtained from simulation

the pile-up

the pile-up

- At LHC, the interaction rate is higher than the bunch crossing rate
- Within a bunch crossing in LHC, more interactions happen
- An event of interesting physics will be **recorded together with other events overlapped**, that are proton-proton interactions with low physics interest
- they are equivalent to a non-interesting event (**minimum bias**)



- given an average number of interactions, the number of PU events per bunch-crossing is expected to have roughly a poissonian distribution

measure the pile-up

- multiply the luminosity (per bunch) by the minimum bias cross-section (71.3 mb) gets the expected rate per bunch:

$$\mathrm{Rate}_{\mathrm{pileup}_{\mathrm{xing,ls}}} = \mathscr{L}_{\mathrm{xing,ls}} \cdot \sigma_{\mathrm{minimum\ bias}}$$

- divide by the revolution frequency of a bunch to get the number of PU events:

$$\mathscr{N}_{\mathrm{pileup}_{\mathrm{xing,ls}}} = \frac{\mathscr{L}_{\mathrm{xing,ls}} \cdot \sigma_{\mathrm{minimum\ bias}}}{\mathrm{cirulation\ rate}}$$

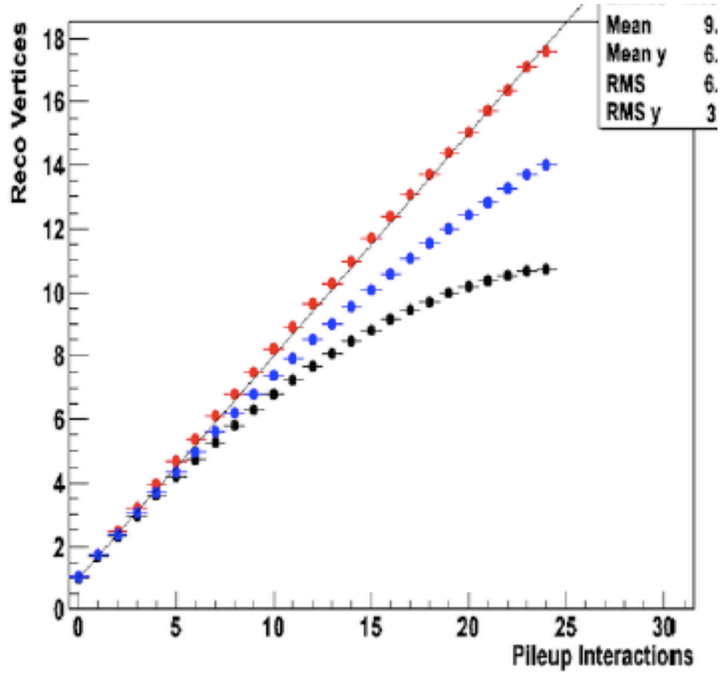
- calculate average distributions over longer periods, weighting by the luminosities

effects of pile-up

- fill in the detector with deposits:
  - **jet reconstruction** algorithms incorporate pile-up deposits
  - **lepton isolation** cones are filled in with pile-up deposits
  - **new jets** might appear in the event
  - more hits in the **tracker** appear
  - the **trigger** is affected
  - **MET** resolution worsens
  - ....

how to deal with it

- apply strict r**equirements on the vertexing of tracks** - need a precise vertex reconstruction algorithm
- measure the **pile-up density** event by event, and use it to subtract from the jets energy a pile-up term (FastJet)
- do the same with isolation cones



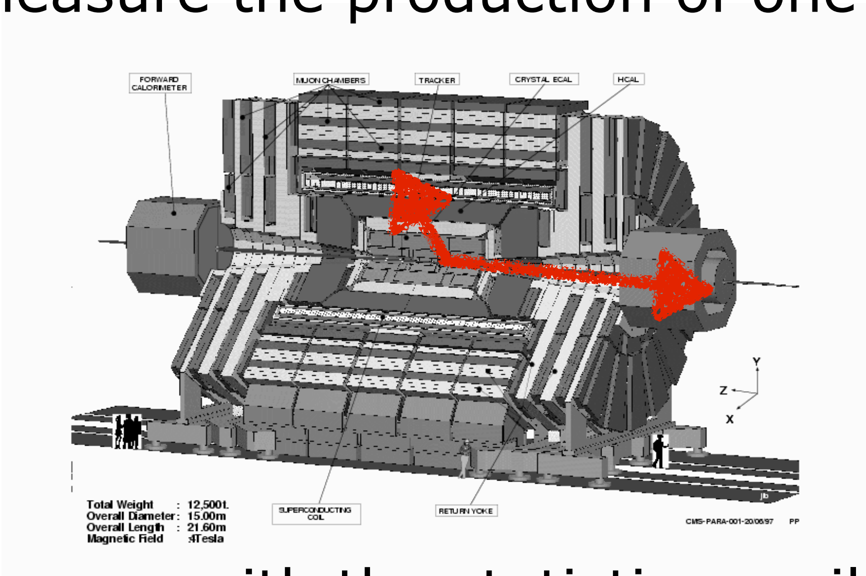
- subtract in the isolation cone the contribution of tracks that do not aim at the same vertex of the lepton
- reconstruct the MET only with particles that aim at a given vertex

M. Cacciari, G. Salam and G. Soyez, **FastJet** http://www.lpthe.jussieu.fr/~salam/fastjet/

dijet cross-section

dijet cross-section

- measure the production of one central + one forward jet in CMS



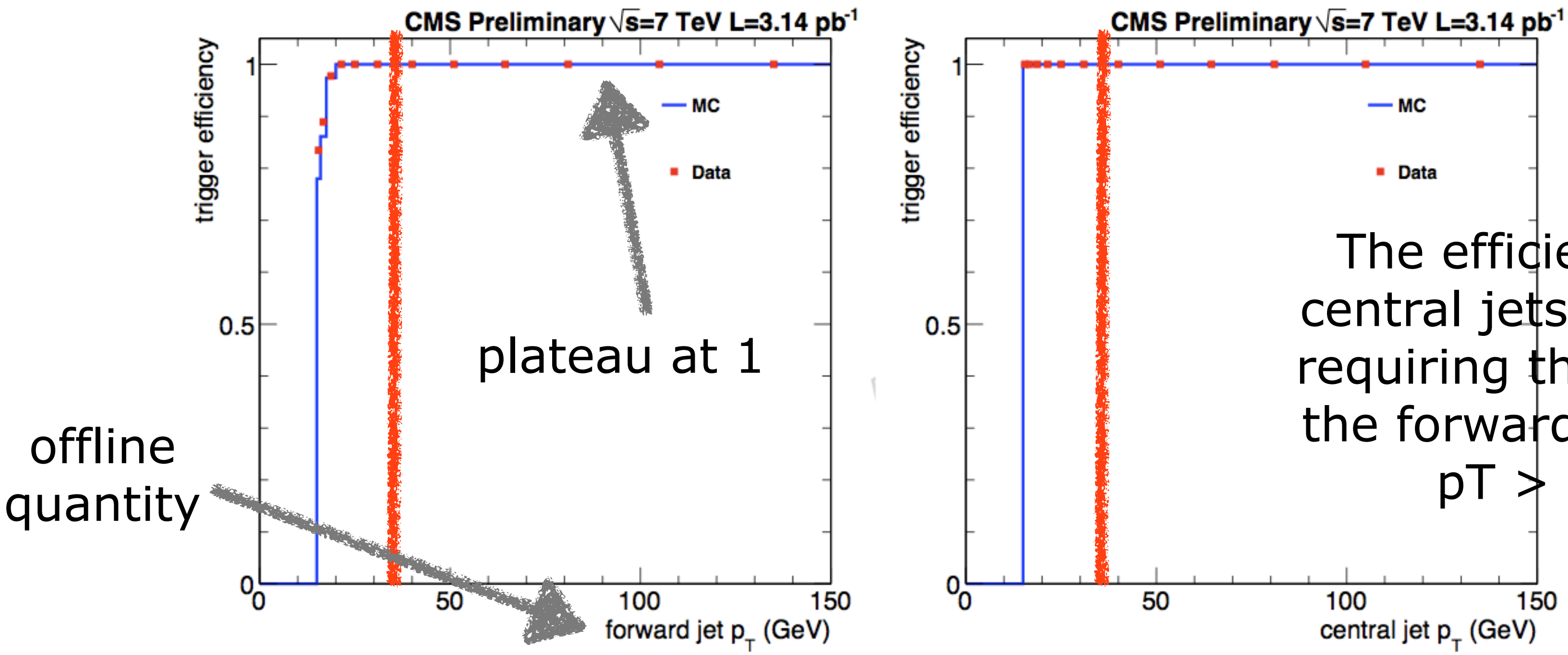
$$\frac{d^4\sigma}{dp_T^c\, dp_T^f\, d\eta^c\, d\eta^f}$$

- to cope with the statistics available, the measurement is done versus pT only, integrated over the central and forward regions, averaged over eta

$$\begin{cases} \dfrac{d^2\sigma}{dp_T^f d\eta^f} = \dfrac{1}{\Delta\eta^f} \cdot \left.\dfrac{d^4\sigma}{dp_T^c\, dp_T^f\, d\eta^c\, d\eta^f}\right|_{p_T^c>35\mathrm{GeV}\ \wedge\ |\eta^c|<2.8} \\[2em] \dfrac{d^2\sigma}{dp_T^c d\eta^c} = \dfrac{1}{\Delta\eta^c} \cdot \left.\dfrac{d^4\sigma}{dp_T^c\, dp_T^f\, d\eta^c\, d\eta^f}\right|_{p_T^f>35\mathrm{GeV}\ \wedge\ 3.2<|\eta^f|<4.7} \end{cases}$$

the analysis definition

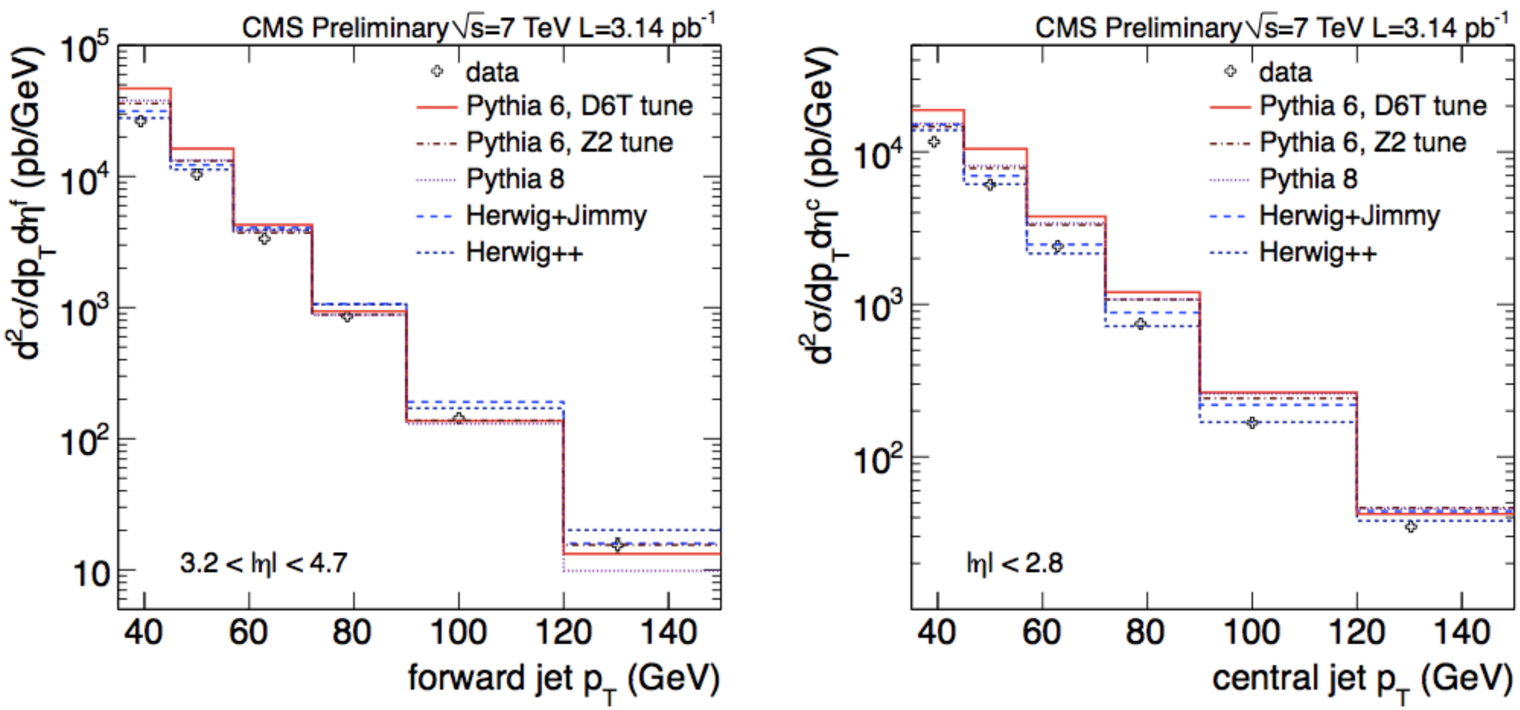
- simple topology: at least one central and one forward jet with p_T > 35 GeV in the event
- the one with the highest p_T is used in each region
- first question: do we trigger these events? di-jet trigger with a raw calorimeter energy threshold of $(E_{T,1} + E_{T,2})/2 > 15$ GeV
- measure the trigger efficiency with the bootstrapping method (wrt the minimum-bias one)

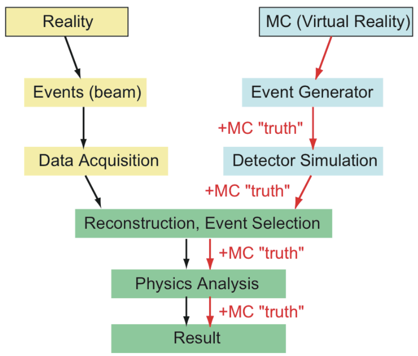


The efficiency for the central jets is calculated requiring that the jets in the forward region have pT > 35 GeV

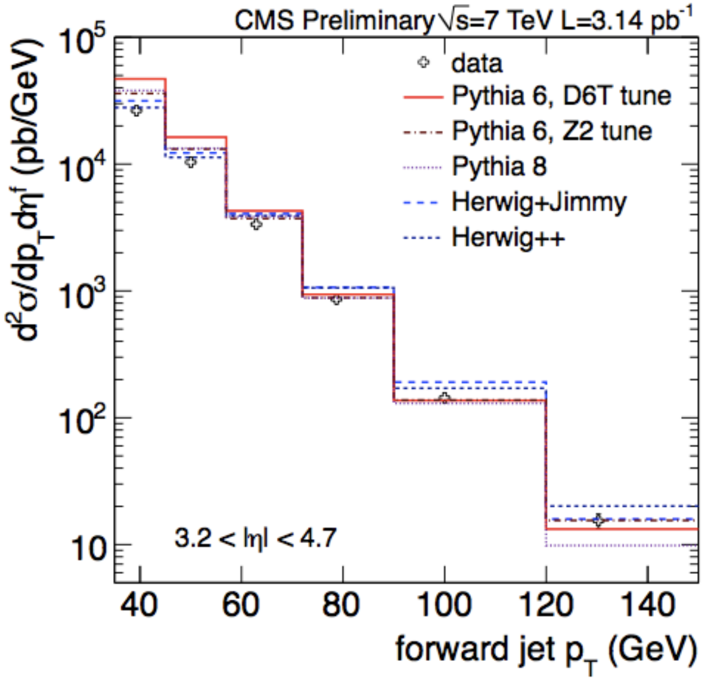
the observed cross sections

- count the number of events in bins of p_T, for the forward and central regions separately (at reconstruction level)
- comparison to some montecarlo predictions is possible, since the simulated events propagated though the whole chain
- what can be done for simulations that do not reach the end of the chain?

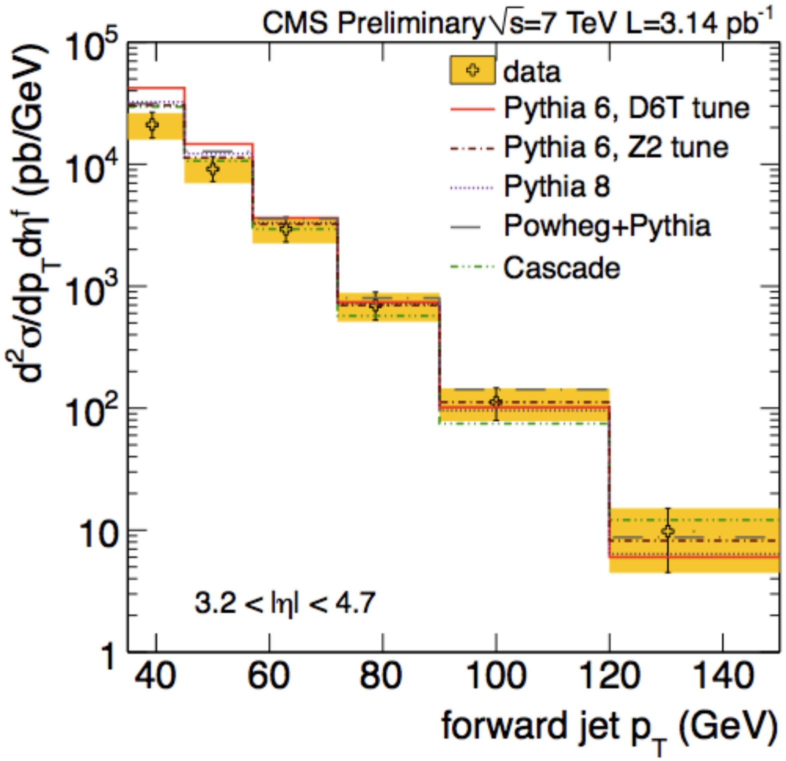




the unfolding



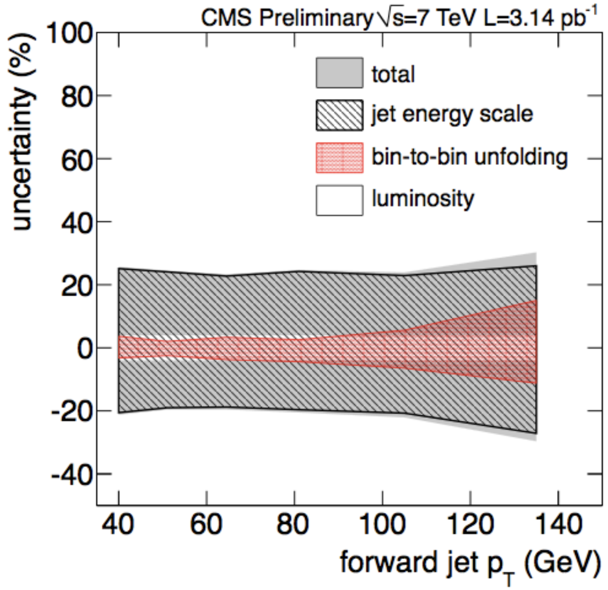
- the rather large jet energy resolution (10%) can give rise to migration effects among the bins
- the interaction with the detector can change the shape of the cross-section
- need to unfold the distribution to the hadron level



- pick one simulation
- reweight the simulated events on quantities at hadron level, to let it match the data
- use the ratio between hadron and detector level to get correction factors to be applied to the data

systematics

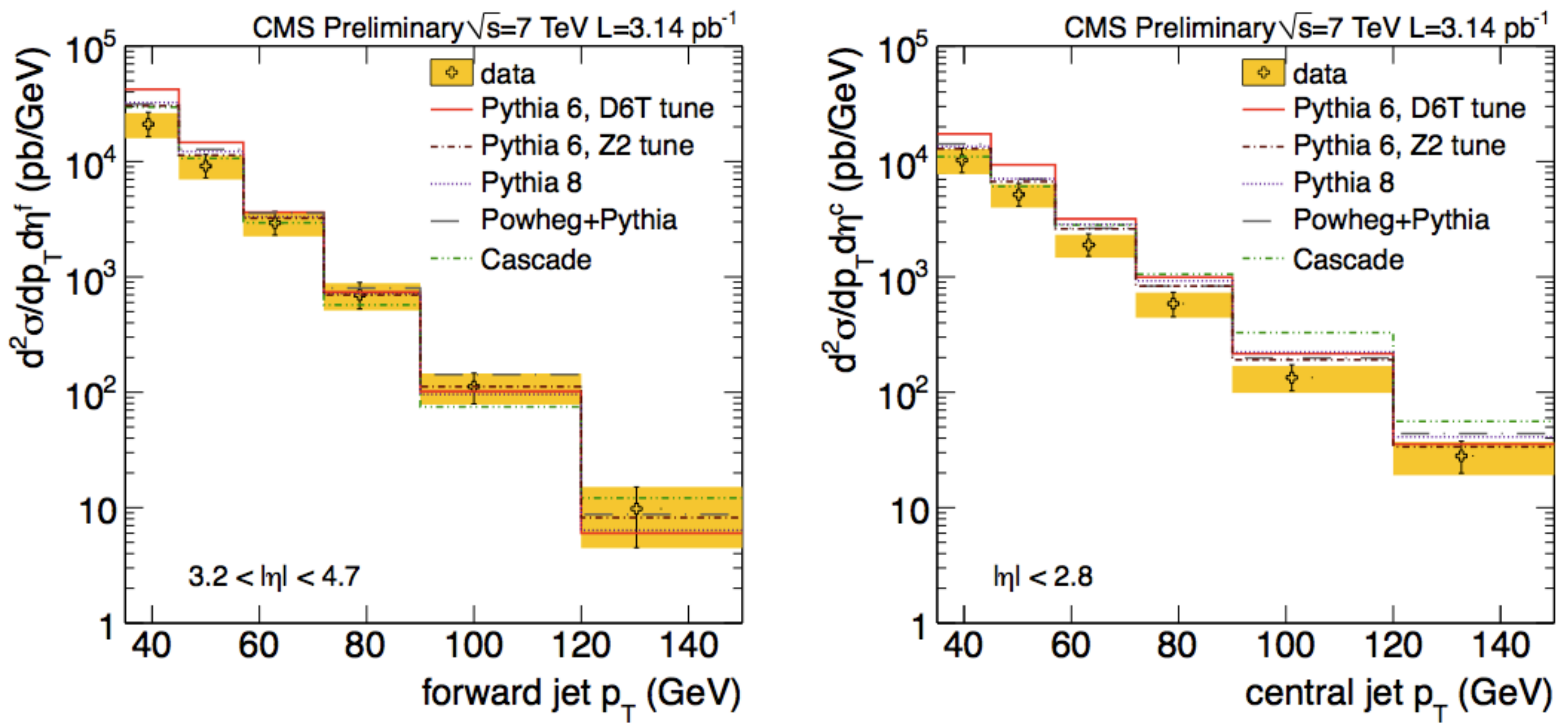
source	uncertainty [%]
JES	25
jet energy resolution	<5
PU	$\simeq 5$
luminosity	4
corrections	<10
HLT	≈ 0
total	30



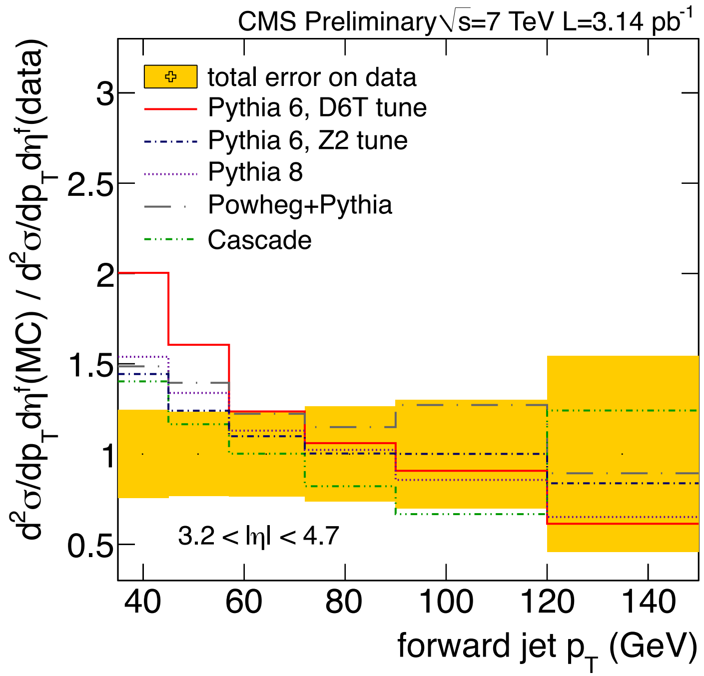
- propagating the initial uncertainties through the analysis
  - **jet energy scale (JES)**: coherently vary all the jets p_T of $\pm\sigma$ in he analysis and compare the results
  - **jet energy resolution**: assume a better (worse) resolution and propagate the effect
- comparing the effects of different initial choices (PU, corrections)
  - **PU**: perform the analysis with all the events, or the ones wit a single vertex
  - **unfolding**: calculate the factors with several simulations and combine the results

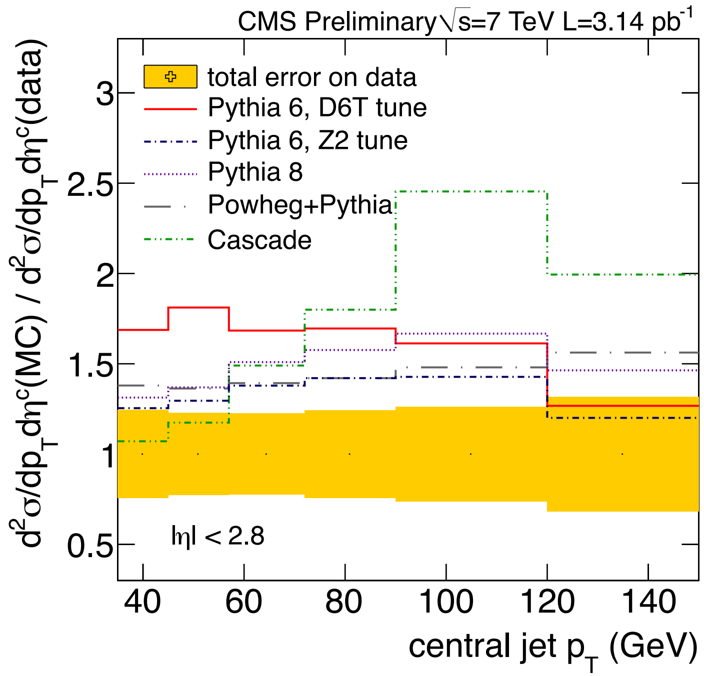
the cross-section

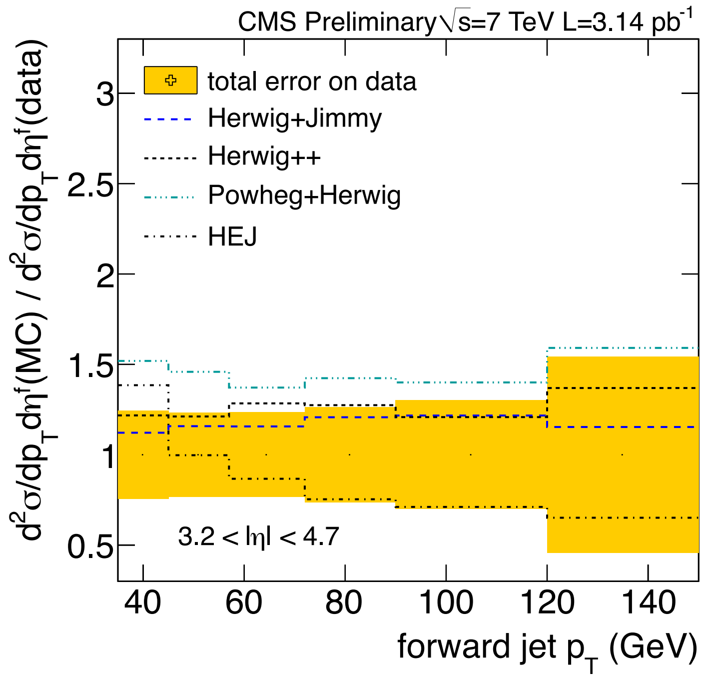
- such final state can give informations on multi-parton interaction and multi-jet production
- study different types of parton radiation dynamics (DGLAP, BKKL, CCFM)
- compare the results to simulations that implement the different behaviours

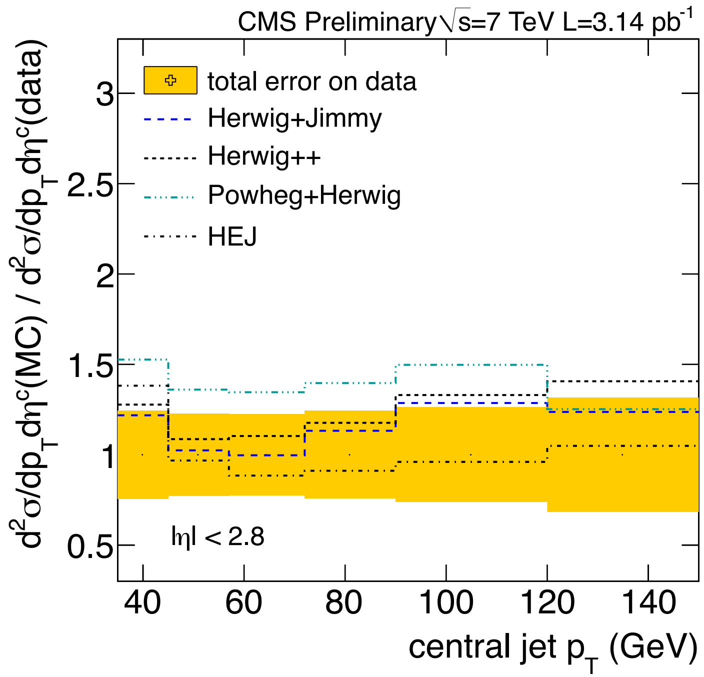


the result


CMS Preliminary √s=7 TeV L=3.14 pb-1
total error on data
Pythia 6, D6T tune
Pythia 6, Z2 tune
Pythia 8
Powheg+Pythia
Cascade
3.2 < |η| < 4.7
d2σ/dpTdηf(MC) / d2σ/dpTdηf(data)
forward jet pT (GeV)


CMS Preliminary √s=7 TeV L=3.14 pb-1
total error on data
Pythia 6, D6T tune
Pythia 6, Z2 tune
Pythia 8
Powheg+Pythia
Cascade
|η| < 2.8
d2σ/dpTdηc(MC) / d2σ/dpTdηc(data)
central jet pT (GeV)


CMS Preliminary √s=7 TeV L=3.14 pb-1
total error on data
Herwig+Jimmy
Herwig++
Powheg+Herwig
HEJ
3.2 < |η| < 4.7
d2σ/dpTdηf(MC) / d2σ/dpTdηf(data)
forward jet pT (GeV)


CMS Preliminary √s=7 TeV L=3.14 pb-1
total error on data
Herwig+Jimmy
Herwig++
Powheg+Herwig
HEJ
|η| < 2.8
d2σ/dpTdηc(MC) / d2σ/dpTdηc(data)
central jet pT (GeV)

H > WW > lvlv

one plot for the Higgs boson