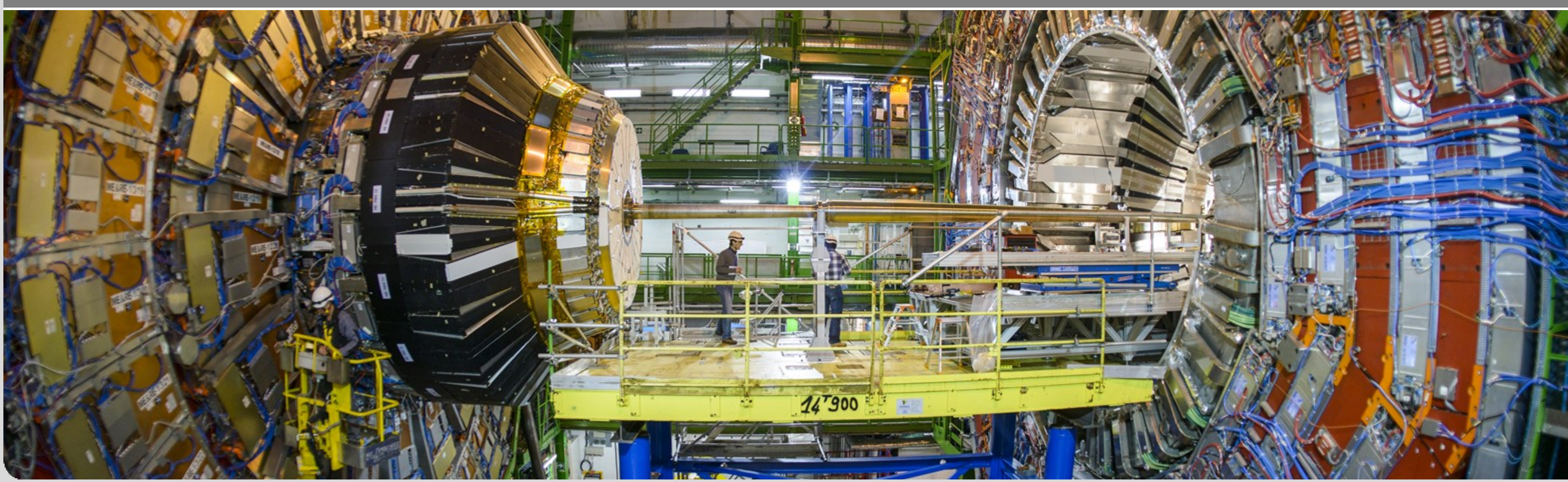
Karlsruhe Institute of Technology

Data Analysis

Stefan Wayand

09. June 2016

INSTITUTE OF EXPERIMENTAL PARTICLE PHYSICS (IEKP) – PHYSICS FACULTY



Schedule for today

- Which objects can be identified by a particle detector ?
- What tasks are covered by the Analysis?

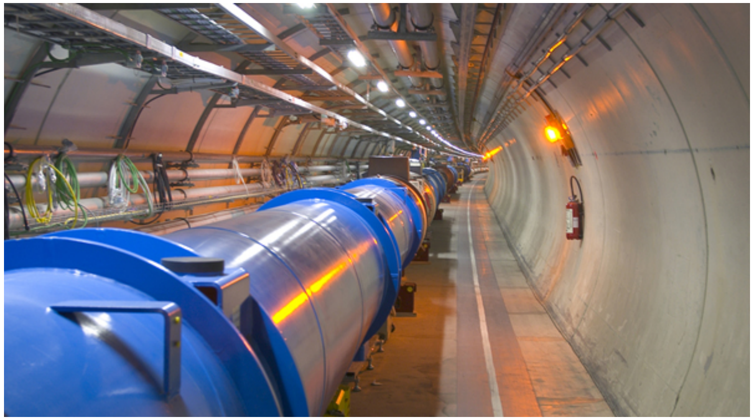
3 Modeling the background

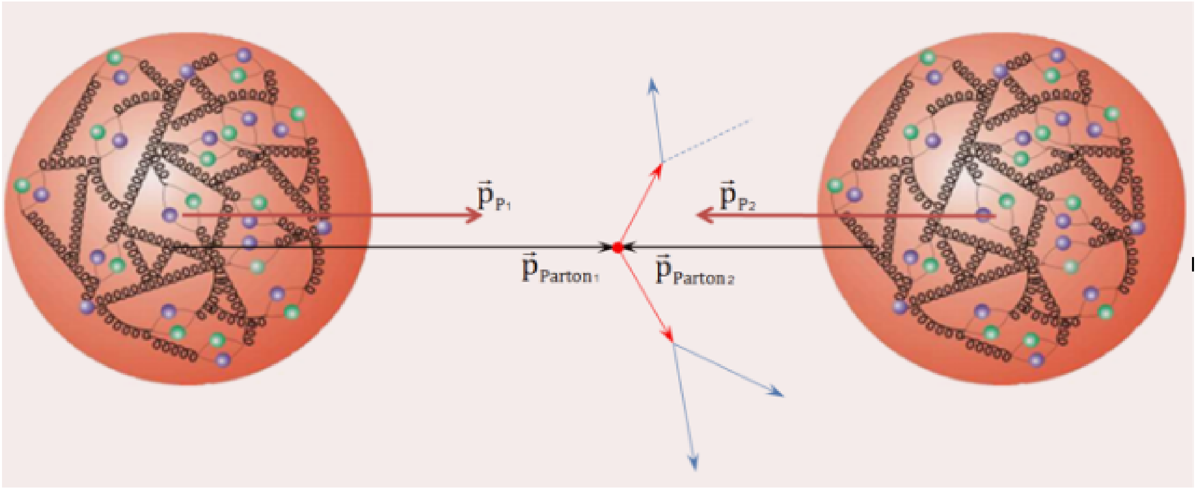
2 Techniques used to understand the reco objects

1 Basics about object reconstruction

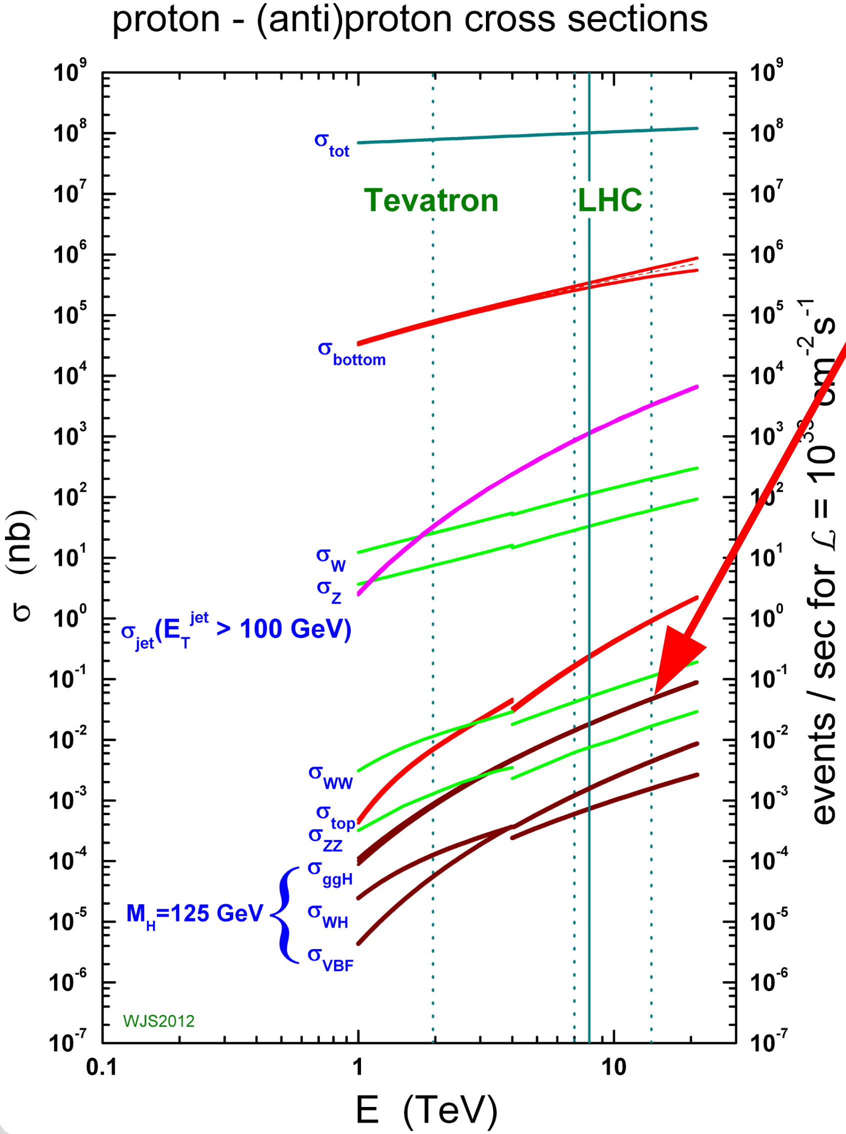
Recap


$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$
$+ i \bar{\psi} \not{D} \psi + h.c.$
$+ \psi_i y_{ij} \psi_j \phi + h.c.$
$+ |D_\mu \phi|^2 - V(\phi)$
IBM




$\vec{p}_{P_1}$
$\vec{p}_{P_2}$
$\vec{p}_{Parton_1}$
$\vec{p}_{Parton_2}$

Master plan



Need to understood a large variety of particle physics processes to find the Higgs

- Understand the reconstructed objects
- Search in well defined final states (H → bb/ττ/WW/ZZ/γγ). Choose your triggers
- Define the search region (optimize signal to background ratio): cuts / shapes / MVA
- Model the background processes and estimate the signal yields

---

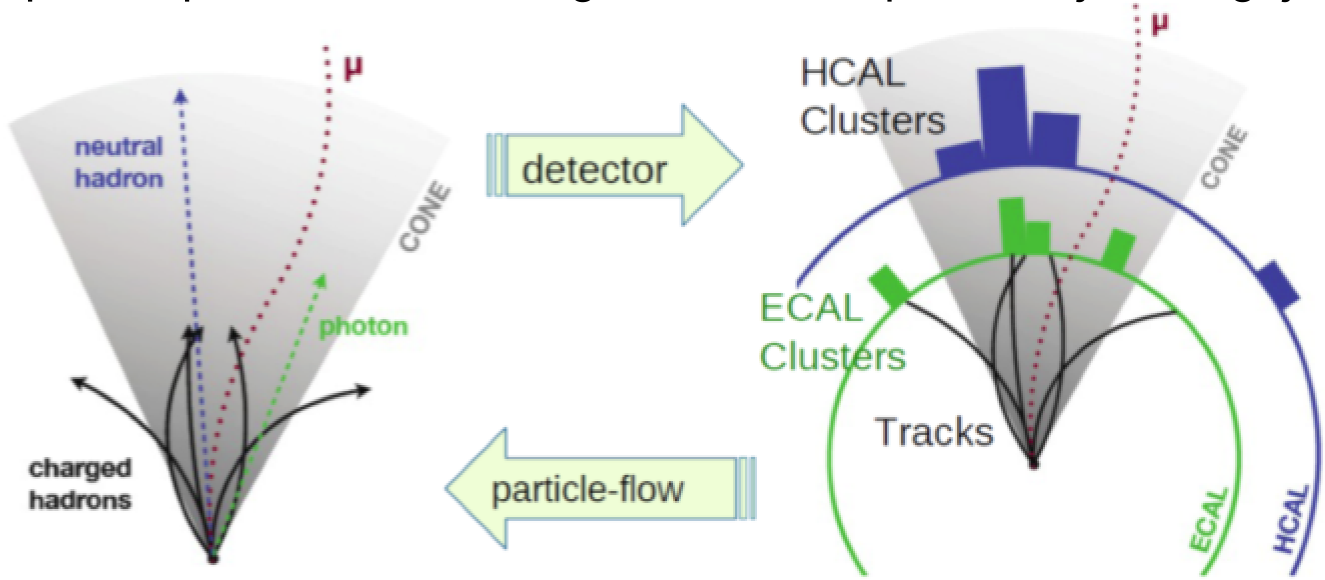
Feed into your statistical model to quantize the result

High-level reconstruction: Particle Flow

- Attempts to reconstruct and identify all particles in the event
  → need matching between calorimeter (fine granularity ECAL) and tracker
- Optimally combines information from all sub-detectors to give best four-momentum measurement of each particle type:

Charged hadrons, **neutral hadrons**, **electrons**, **photons** and **muons**

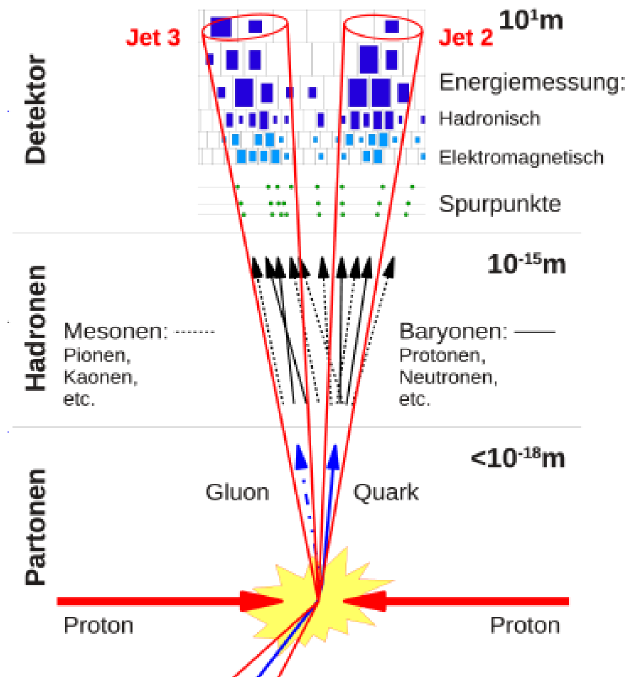
- Also improves performance for higher-level composite objects e.g. jets, MET



Reconstruction of Objects

1. **combine sub-detectors** to classify all stable objects, i.e. find electrons, muons, photons, hadrons. (In CMS provided by the "particle flow" algorithm)

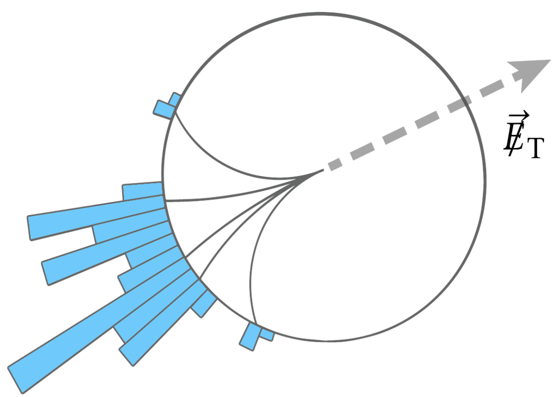
2. **cluster objects** into "jets" (relation between measured final state objects & hard partons) two types of algorithms:
   1. **"cone"**: geometrically assign objects to the leading object
   2. **sequentially combine** closest pairs of objects – different measures of "distance" exist (kT, anti-kT) with some variation of resolution parameter, which determines "jet size"



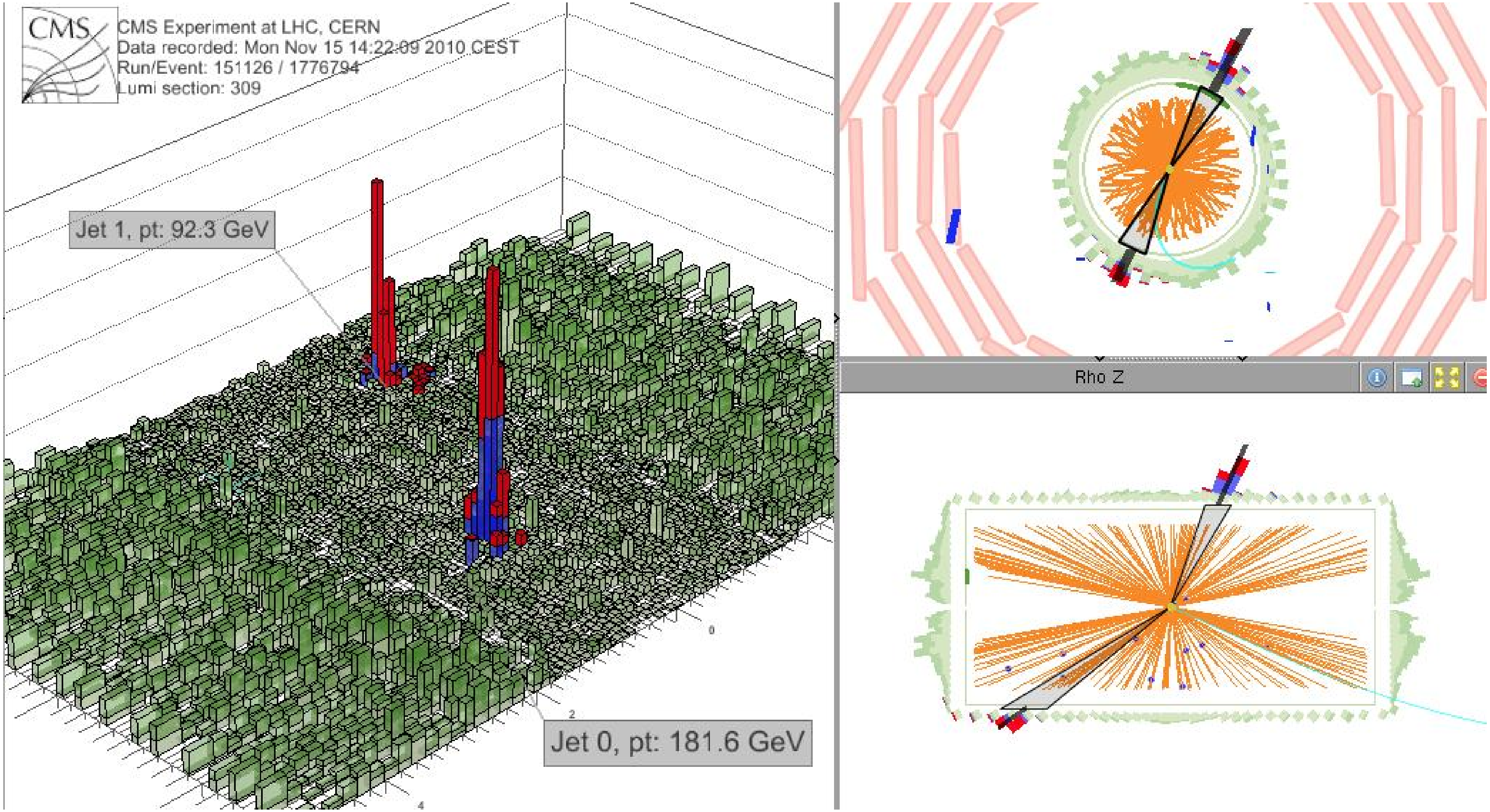
3. determine **missing transverse momentum (energy) called MET:**

$$p_{T\,\mathrm{miss}} = - \sum_{\mathrm{all\ partilces}} \vec{p}_{T\,i}$$

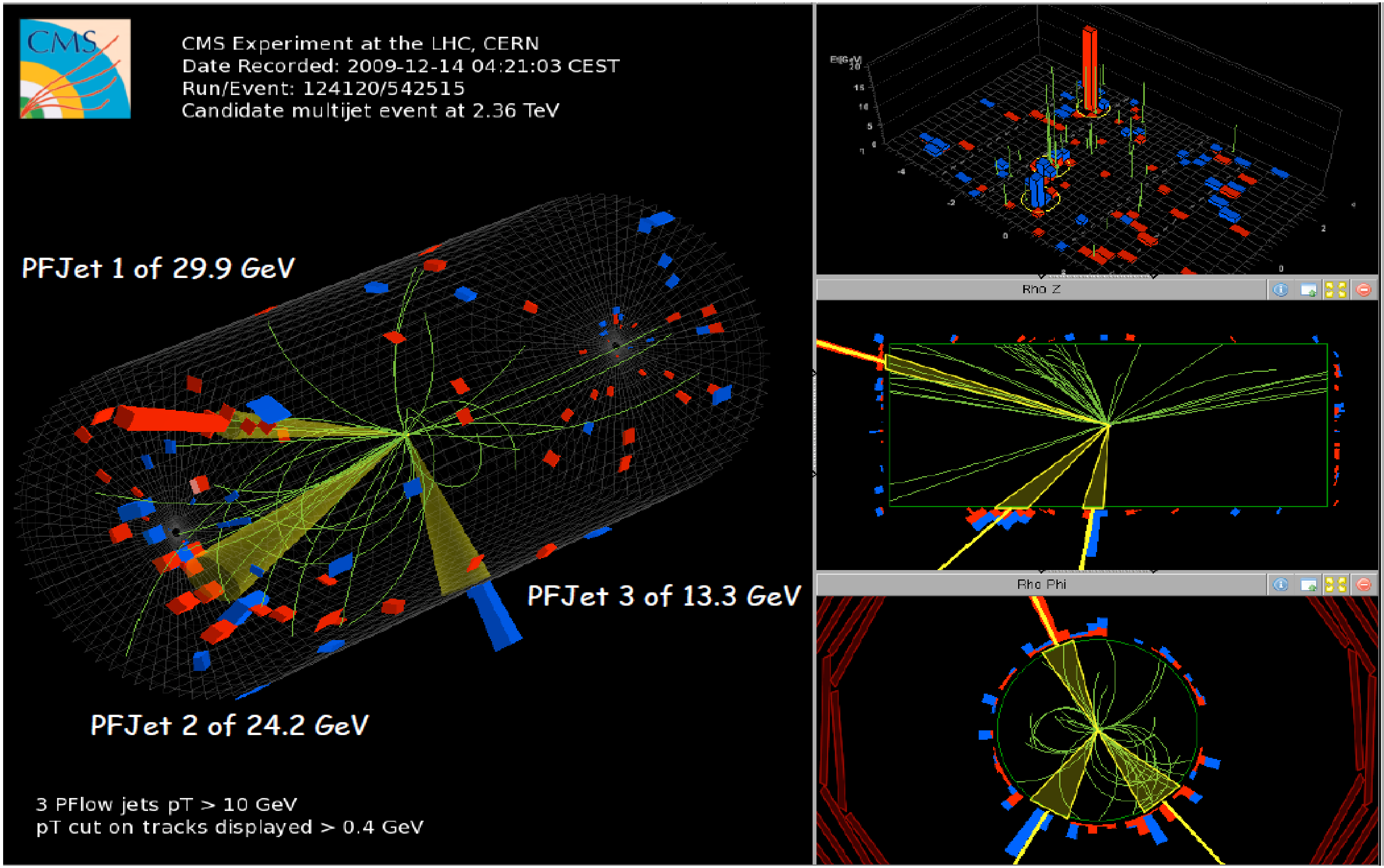
carried away by undetectable particles. In SM neutrinos, "new physics" provides more of them (e.g. dark matter)



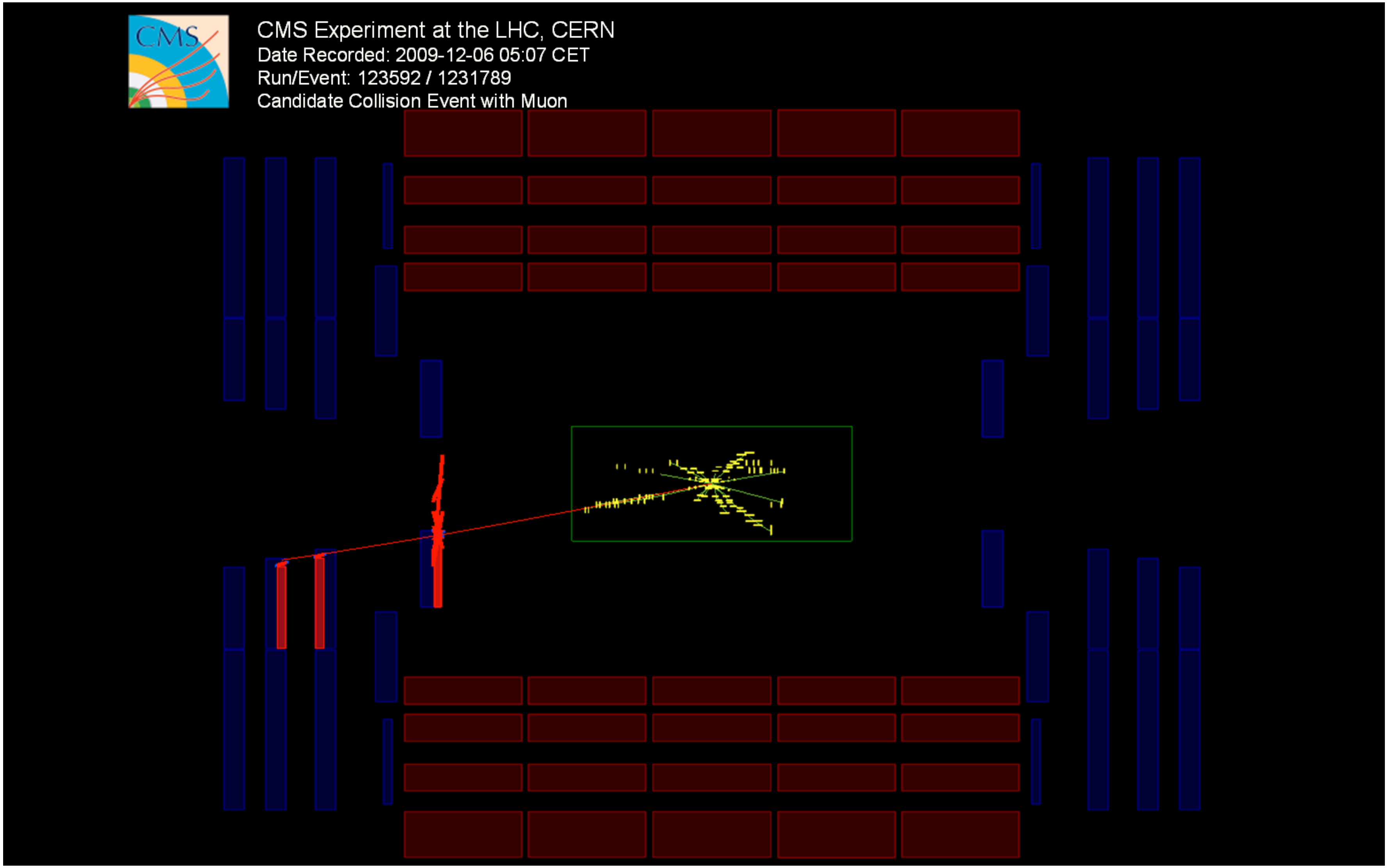
Two-Jet event in the CMS Detector



Three-Jet event in the CMS Detector

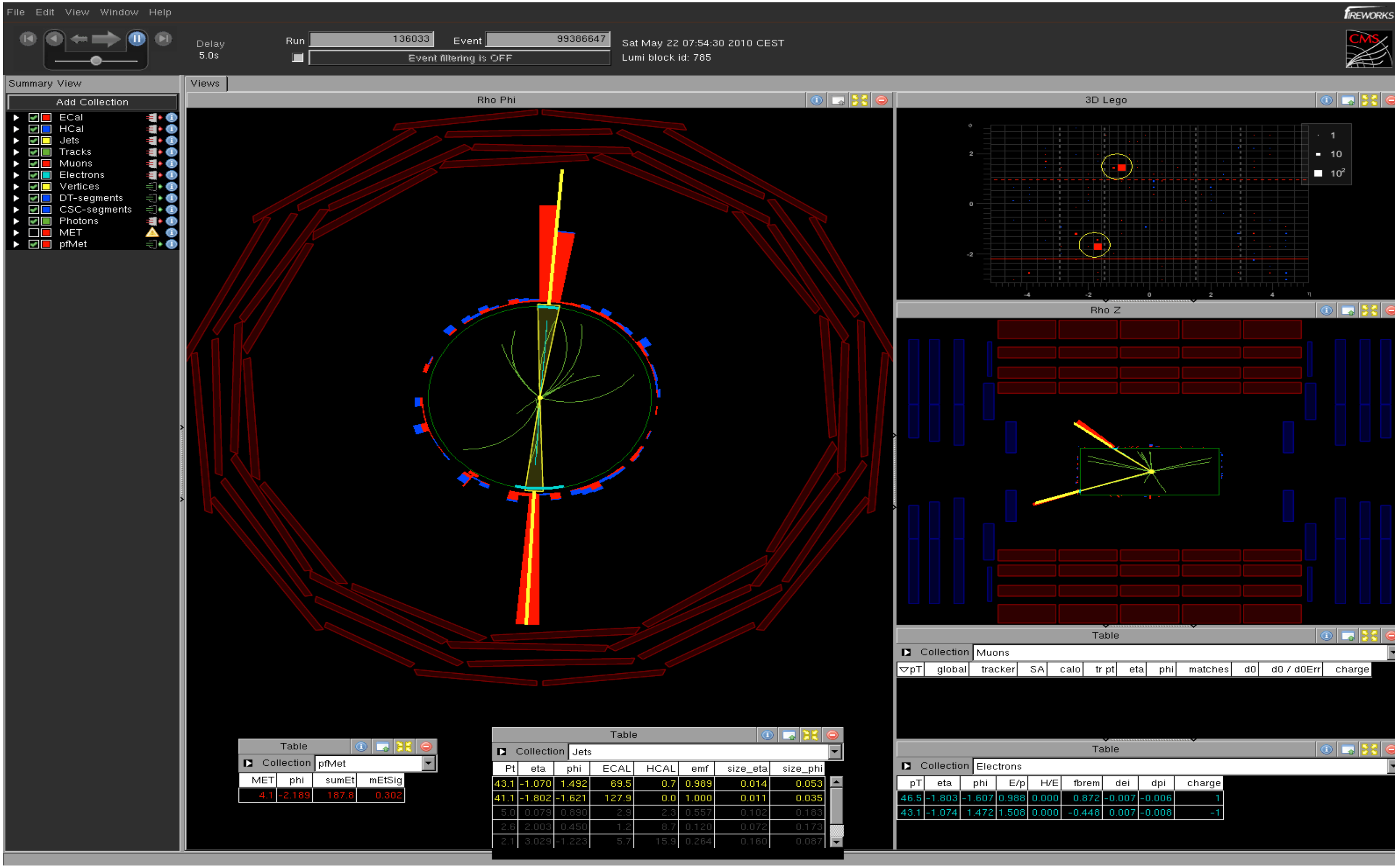


Event with an end-cap muon



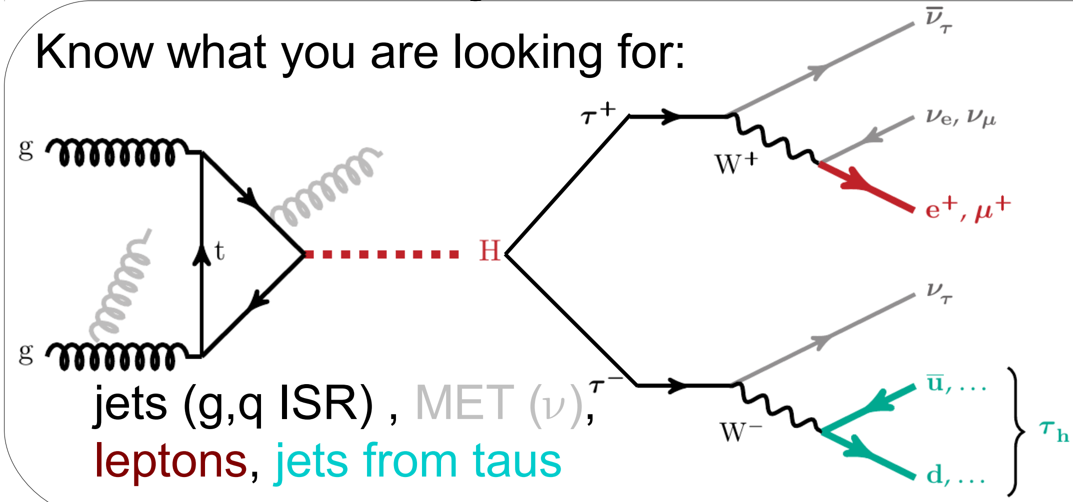
Two electrons in the CMS Detector

Karlsruhe Institute of Technology



Start the Analysis

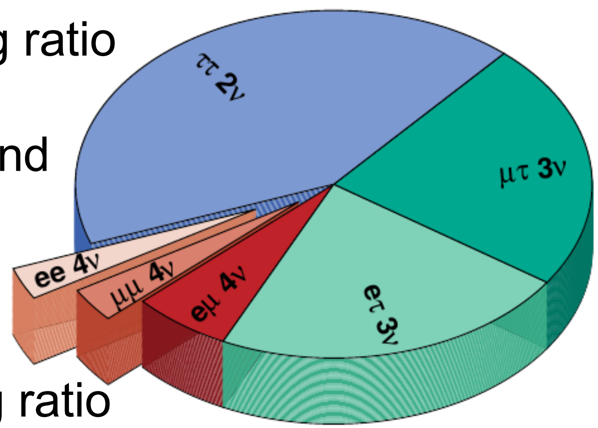
Know what you are looking for:



jets (g,q ISR) , MET (ν),
leptons, jets from taus

Decay Mode	Resonance	Branching Ratio / %
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$		17.4
$\tau^- \to \pi^- \nu_\tau$	$\pi(140)$	11.6
$\tau^- \to \pi^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \to \pi^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	10.8
$\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$		4.8
Other hadronic modes		1.7
All hadronic modes		64.8

- largest branching ratio
- hard to trigger
- largest background



- good branching ratio
- moderate trigger thresholds

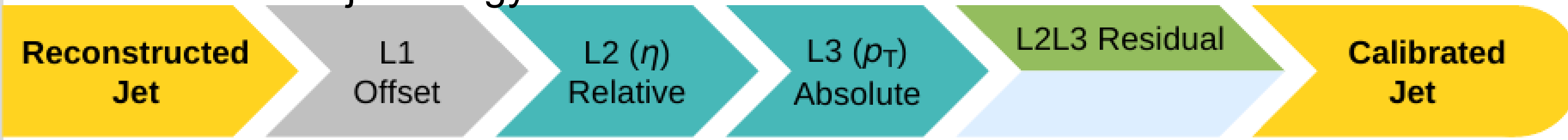
- smallest branching ratio
- smallest background (eμ)

In the final analysis all final states were considered (except ee/μμ)

Institute of Experimental Particle Physics (IEKP)

Object calibration

Calibration of the jet energy in CMS ...



... is a multi-step procedure, driven by data and MC

Level 1: offset correction for pile-up and electronic noise
Level 2: relative (η) corrections
Level 3: absolute p_T correction
MC and special balanced events

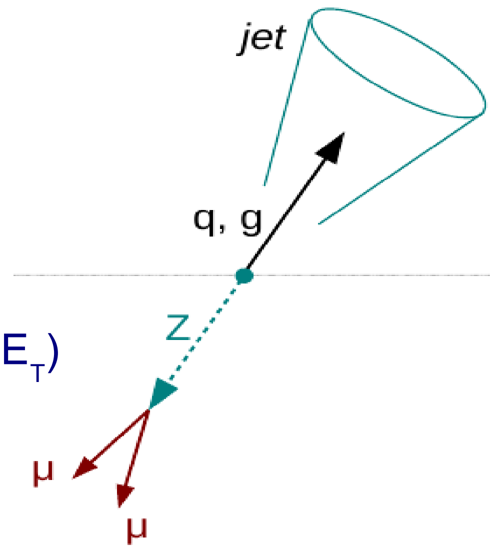
residual corrections from events with selected topology:

Level 2 residual η

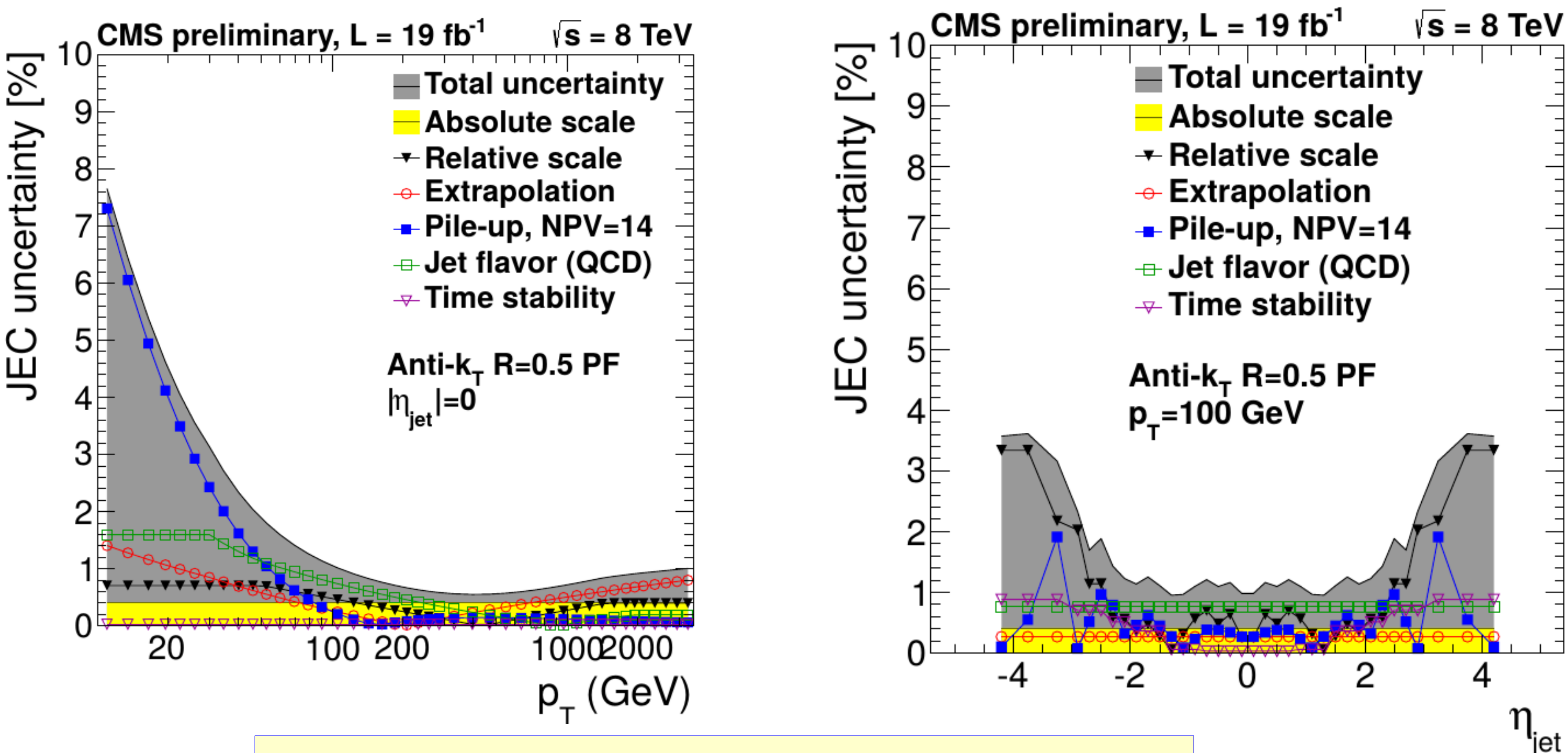
from measured di-jet events, assuming the two jets have the same E_T)

Level 3 residual p_T

from measured Z+jet & photon+jet, jet blanced by Z/γ



Object calibration (Jets)

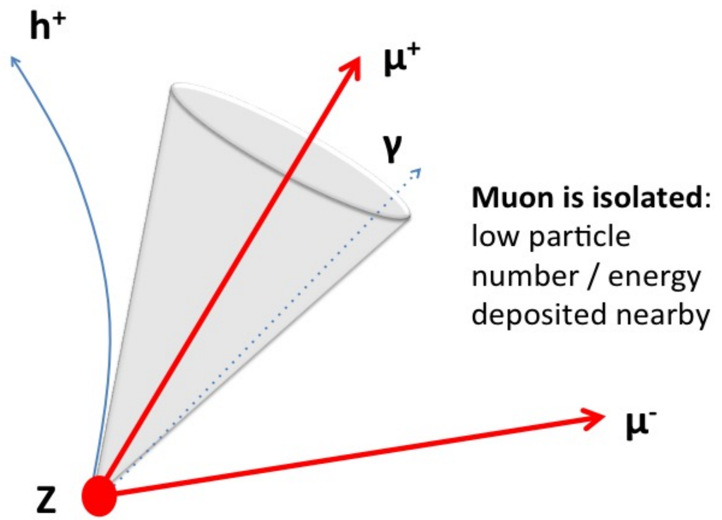


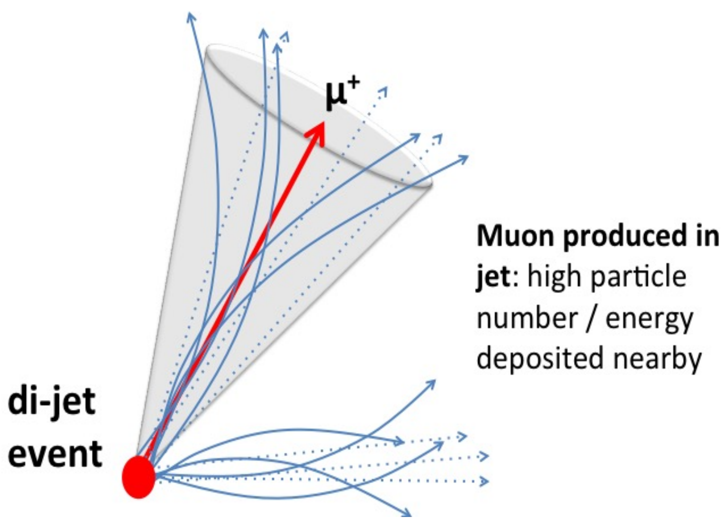
Precision of Jet energy calibration reaches 1 % !

Result is also propagated into MET which helps to improve MET resolution

Object identification and object isolation

- **Identification:** The true particle type can be ambiguous
  - "Is it an electron or a pion?" → can apply object criteria to increase purity of a particle type, e.g. small hadronic energy / EM energy → more likely to be an electron
- **Isolation:** powerful handle to reduce background from jets
  - We are often interested in leptons produced from decays of top quarks, W bosons, Z bosons, Higgs etc
  - These electroweak processes are 'clean' compared to QCD → less activity in the region around lepton direction





Determination of efficiencies

1. take efficiencies from simulation   not always believable !

   check classification in simulated data vs. truth, i.e. determine

   ε_{MC} = fraction of correctly selected objects

   (probability to select background determined in the same way)

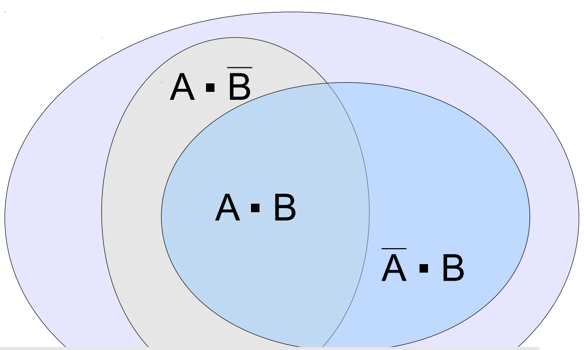
2. **design data-driven methods** using redundancy of at least two variables discriminating signal and background

   – tag & probe method:

   select very hard on one criterion, even with low efficiency, check result obtained by second criterion

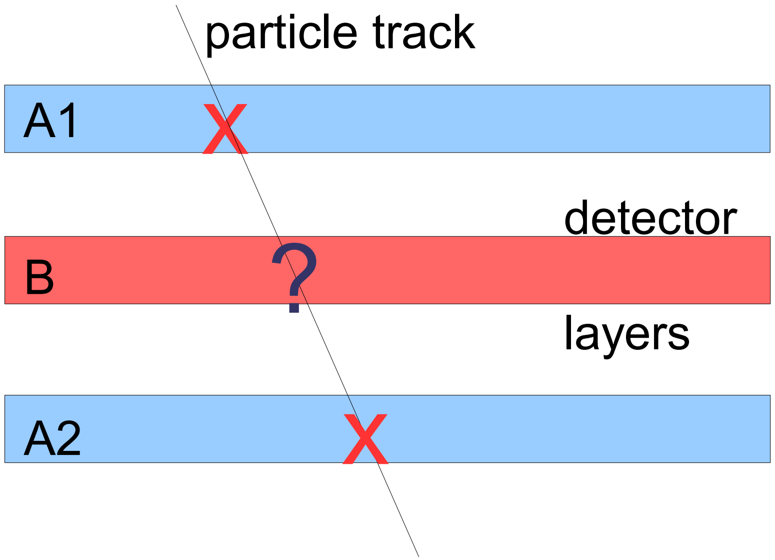
Illustration: two **<u>independent</u>** criteria A, B

$$\epsilon_B = \frac{n(A\cdot B)}{n(A\cdot B)+n(A\cdot\bar{B})}$$



Important: selecting on A must not affect B, i.e. A and B must be uncorrelated !

Tag and Probe: Example 1



Hits in layers A1 and A2 define valid particle track (tag)

probe hit in layer B

Coincidence of Layers A1 and A2 guarantees high purity of the tag (protects against random noise)

allows determination of efficiency of layer B

$$\Rightarrow \quad \epsilon_B = \frac{n_B}{n_{A1 \cdot A2}}$$

Institute of Experimental Particle Physics (IEKP)

Trigger efficiencies

Determination of trigger efficiencies depends on
existence of independent selection methods

Important to ensure redundancy when building trigger systems !

Trigger information must be stored for later use in efficiency determination !

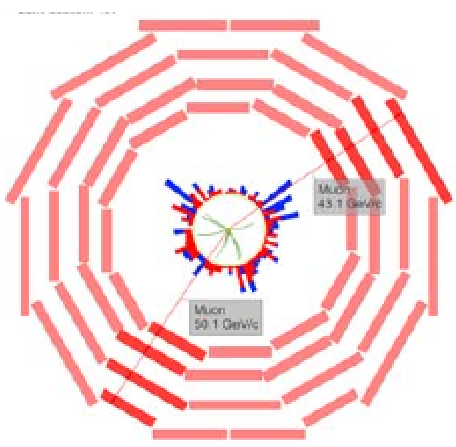
typical methods:

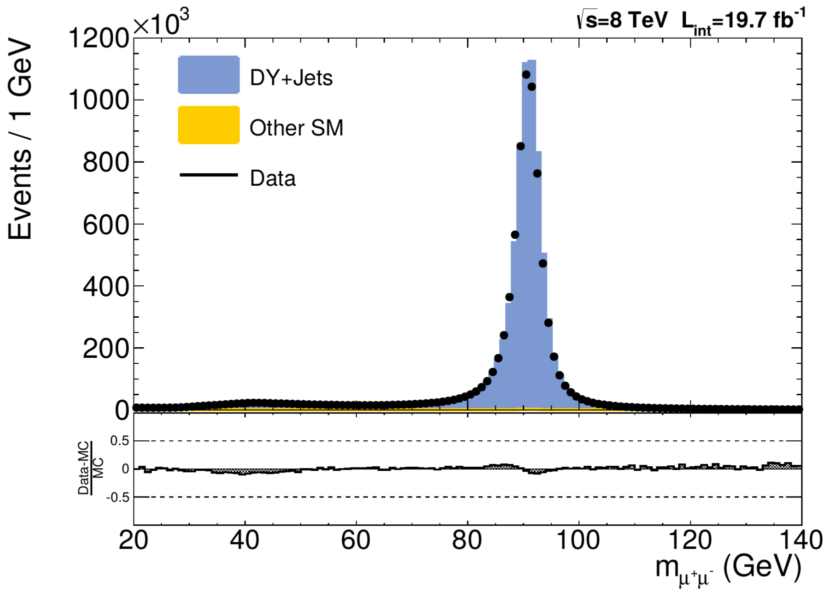
- use trigger from independent sub-systems
- trigger at lower threshold (typically pre-scaled to run at acceptable rates)
  → probe higher-threshold triggers
- trigger on pairs of objects at low threshold,
  → probe higher threshold on each member of the pair
  !!! potential bias, because higher-threshold trigger depends on same input signals as the tag !!!
- trigger only one object of a pair and use an off-line criterion to identify 2nd member of the pair and probe trigger decision on it

Tag and Probe: Example 2

criterion A: a tight muon/electron and
one other track with tight selection on Z mass **("tag")** thus
selecting Z → μμ (or ee) (which is possible with very high purity)
→ 2nd track also is a muon/electron with very high probability

criterion B: 2nd track selected by trigger (or analysis) **("probe")**
allows measurement of trigger efficiency
(or selection efficiency) of second muon





Statistical error on efficiency

determination of efficiencies is a clear application of **binomial statistics**:

*number of successes **k** in **n** trials at probability **p** per trial*

Binomial Distribution

$$P(k;p,n) = \binom{n}{k} p^k (1-p)^{n-k}, k = 1, \ldots, n \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Expectation value $\mathrm{E}[k] = np$ **Variance** $\mathrm{V}[k] = np(1-p)$

Error on efficiency: insert measured efficiency $\epsilon = k/n$ in formula for variance (instead of true (but unknown) selection efficiency p !)

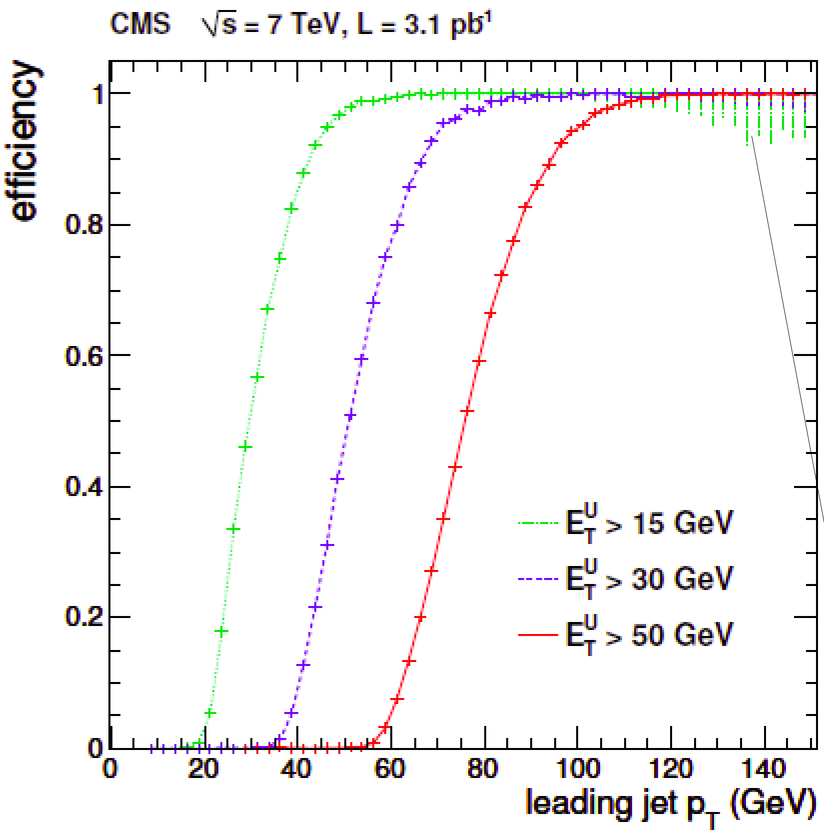
$$\rightarrow \quad \sigma_\epsilon = \frac{\sqrt{\epsilon(1-\epsilon)}}{\sqrt{n}}$$

if this is not justified due to very small statistics, a more sophisticated method of "interval estimation" is needed to specify a confidence range on the measured efficiency:

→ Clopper-Pearson method

Typical “turn-on” curves of trigger efficiencies

(calorimeter jet trigger on transverse energy of jets, CMS experiment)



Remarks:

- efficiency at 100% only far beyond “nominal” threshold
- trigger efficiencies vary with time (depend on “on-line” calibration constants)
- to be safe and independent of trigger efficiencies, analyses should use cuts on reconstructed objects that are tighter than trigger requirements

2nd remark: errors determined as 68% confidence interval by application of Clopper-Person method per bin; this explains the (counter-intuitive) large uncertainties on the >15 GeV trigger at high pT: there were just no events observed where trigger was inefficient.

LESSON: sophisticated methods are not always plausible !

More complicated observables

Calculate **derived quantities** from objects,

– transverse momentum or energy, *at hadron colliders where rest system of an interaction is boosted along z direction*

$$\vec{p_T} = \sum_i \vec{p_T}_i$$

$$E_T^2 = \sum_i {m_i}^2 + \vec{p_T}_i^2$$

– missing transverse momentum, from all particles in an event, *assuming total transverse momentum of zero in each event, measures effects of invisible particles (neutrinos in the SM, but there are others in extended theories)*

$$\vec{p}_{T\,\mathrm{miss}} = -\sum_{\mathrm{all\ partilces}} \vec{p}_{T\,i}$$

– “transverse mass” $M_T^2 = E_T^2 - p_T^2$

More complicated observables

Calculate **derived quantities** from objects,

– transverse momentum or energy, *at hadron colliders where rest system of an interaction is boosted along z direction*

$$\vec{p_T} = \sum_i \vec{p_T}_i$$

$$E_T^2 = \sum_i {m_i}^2 + \vec{p_T}_i^2$$

– missing transverse momentum, from all particles in an event, *assuming total transverse momentum of zero in each event, measures effects of invisible particles (neutrinos in the SM, but there are others in extended theories)*

$$\vec{p}_{T\,\mathrm{miss}} = - \sum_{\mathrm{all\ \ partilces}} \vec{p}_{T\,i}$$

– "transverse mass" $M_\mathrm{T} = \sqrt{2 \cdot E_\mathrm{T}^{miss} \cdot p_T^{Wlep} (1 - \cos \Delta\phi)}$

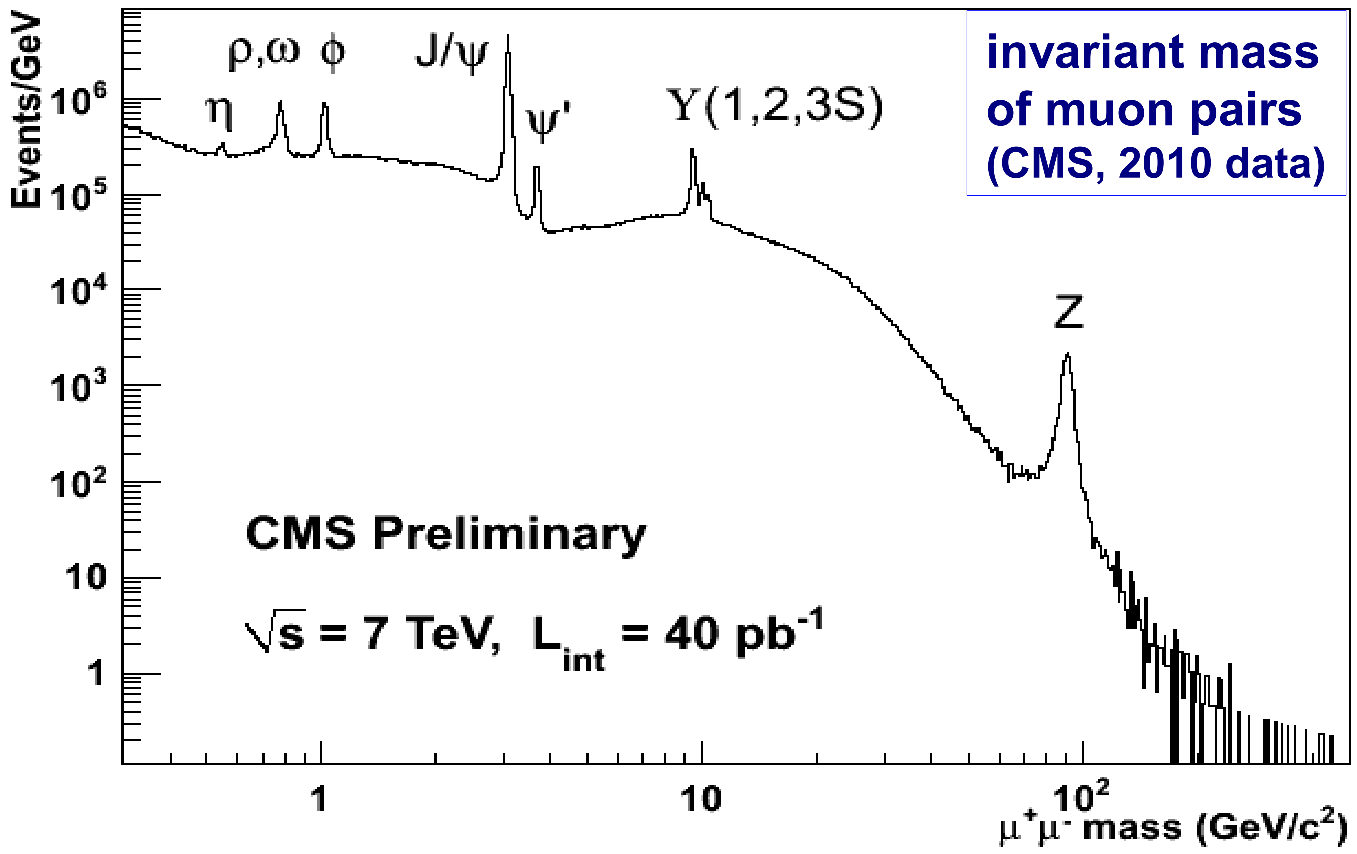
– event shape variables (for QCD analyses) to classify jet topologies

– all kinds of "classifiers" using MVA techniques *for object or event classification*

Invariant mass

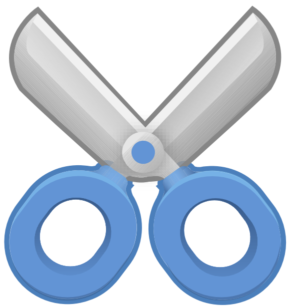
60 years of particle physics in only one year:

Example of a very simple selection:
just the invariant mass of muon pairs in events with one muon trigger



Event Selection

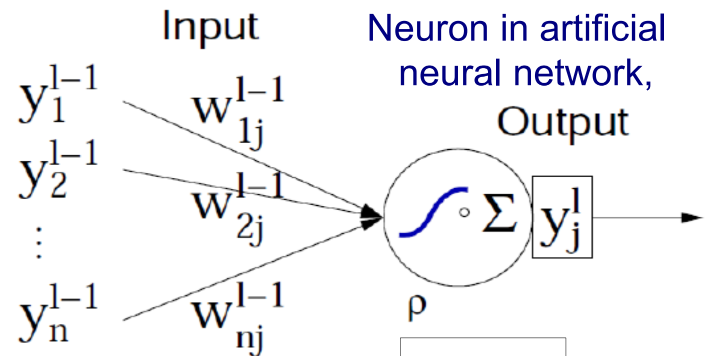
CUTS



Multivariate Analysis (MVA)

e.g. decorrelated likelihood, artificial neural networks, boosted decision trees

Neuron in artificial neural network,



lepton p_T,η

lepton isolation

MET

muon identification

tau identification

electron identification

Number of objects (e,μ,τ, jets)

invariant mass off di-tau system

Time, amount of work, complexity, better separation

Need to understand the efficiencies on signal and background, the uncertainties and possible correlations

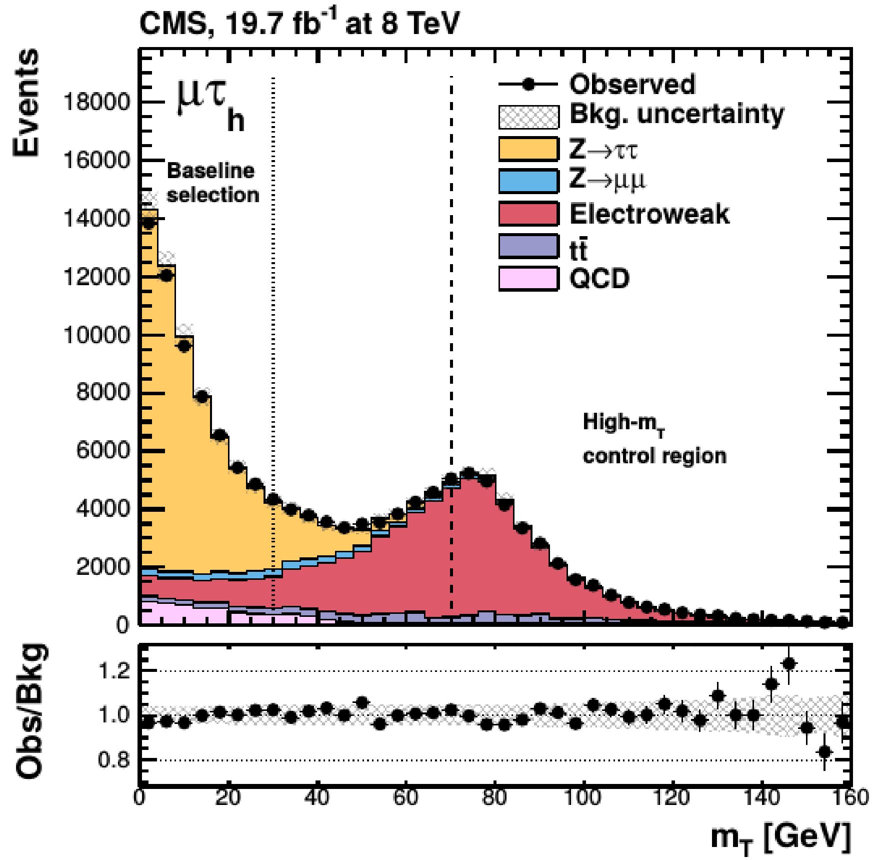
Modeling of Background: part I

– shape take from MC

– extrapolation from "side band"
assuming "simple" background shape or by taking background shape from simulation

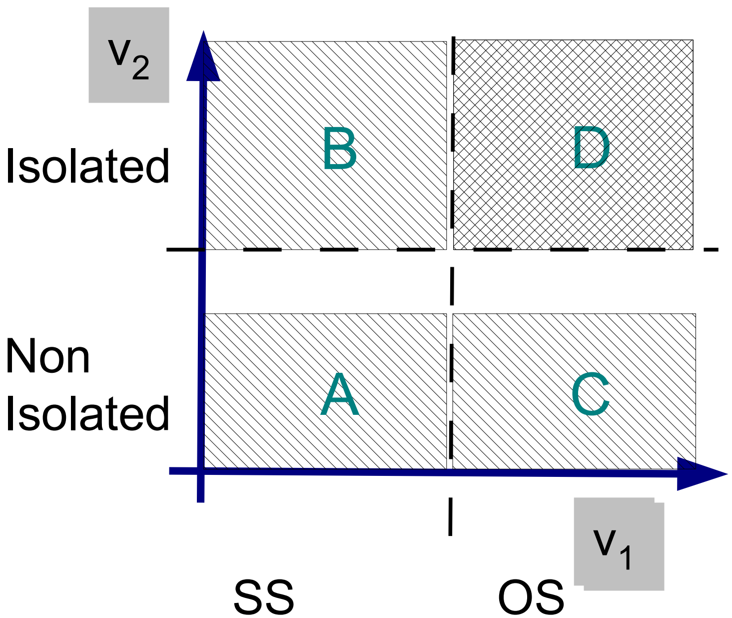
- event counting in background regions, extrapolation under signal assuming (simple) model

- fit of signal + background model to the observed data



Modeling of Background: part II

– ABCD – Method ...



Assumptions:

- two independent variables v1 and v2 for background
- signal only in region D

$$\rightarrow \quad n_D^{bkg} = n_C \frac{n_B}{n_A}$$

... a **data driven estimate** of *background under a signal*

Example: Take the ratio of same-sign (A) and opposite-sign (B) non isolated (invert isolation criteria) leptons to predict the amount of QCD fakes.

– more advanced methods exist to **exploit two uncorrelated variables** to predict the background shape under a signal, see e.g. "sPlot method" in ROOT.

Institute of Experimental Particle Physics (IEKP)

Modeling of Background: part III

Hybrid events: data + Monte Carlo: $Z \rightarrow \tau\tau$ background in the $H \rightarrow \tau\tau$ search

- $H \rightarrow \mu\mu$ has very low branching ratio, hence there is no $H \rightarrow \mu\mu$ under $H \rightarrow \mu\mu$
- $Z \rightarrow \mu\mu$ and $Z \rightarrow \tau\tau$ are very similar (lepton universality of weak decay)

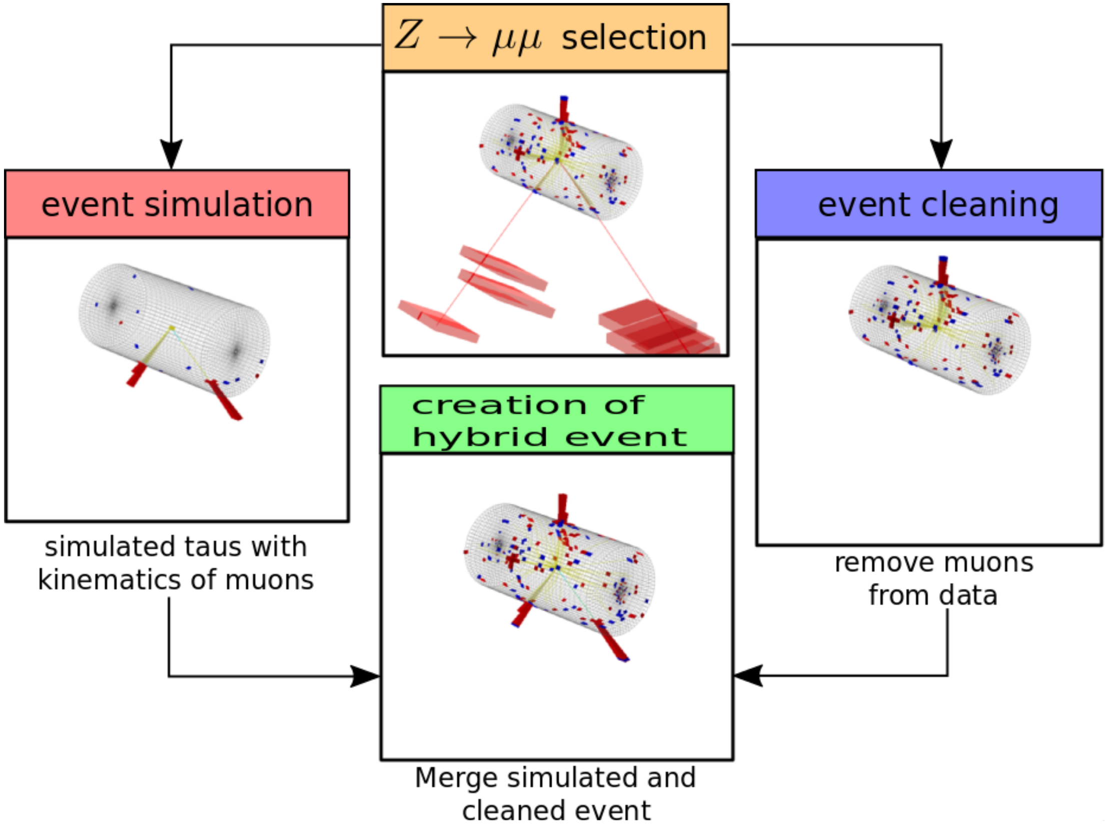
idea:

replace real μ in $Z \rightarrow \mu\mu$ events with simulated τ to model Z background under H signal

advantages:

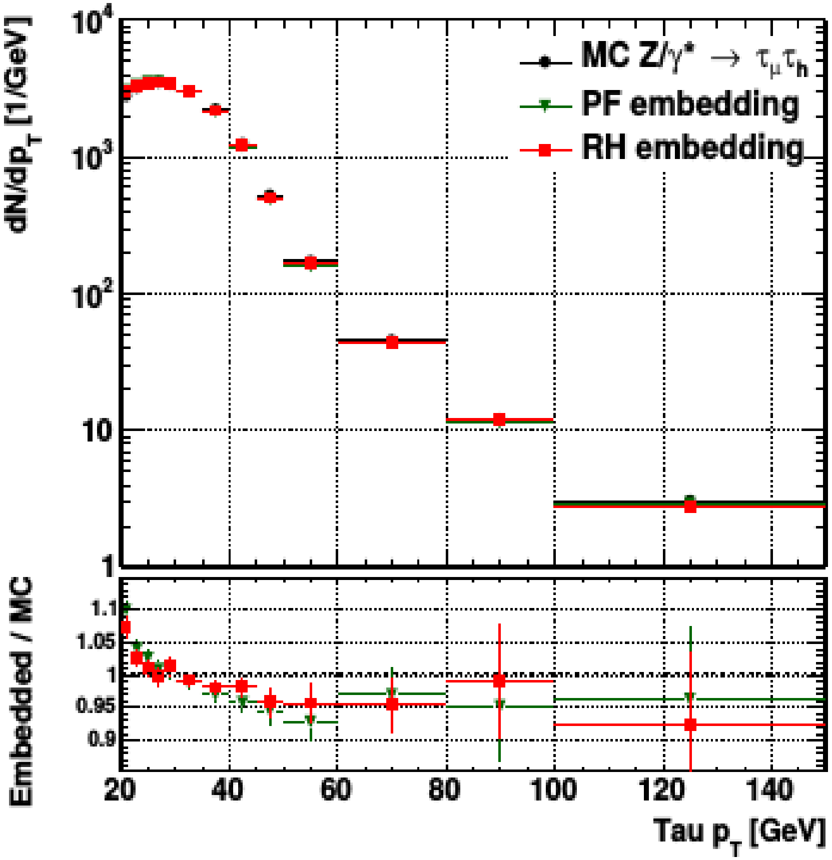
– non-leptonic part of event is from real data, esp. important in presence of pile-up

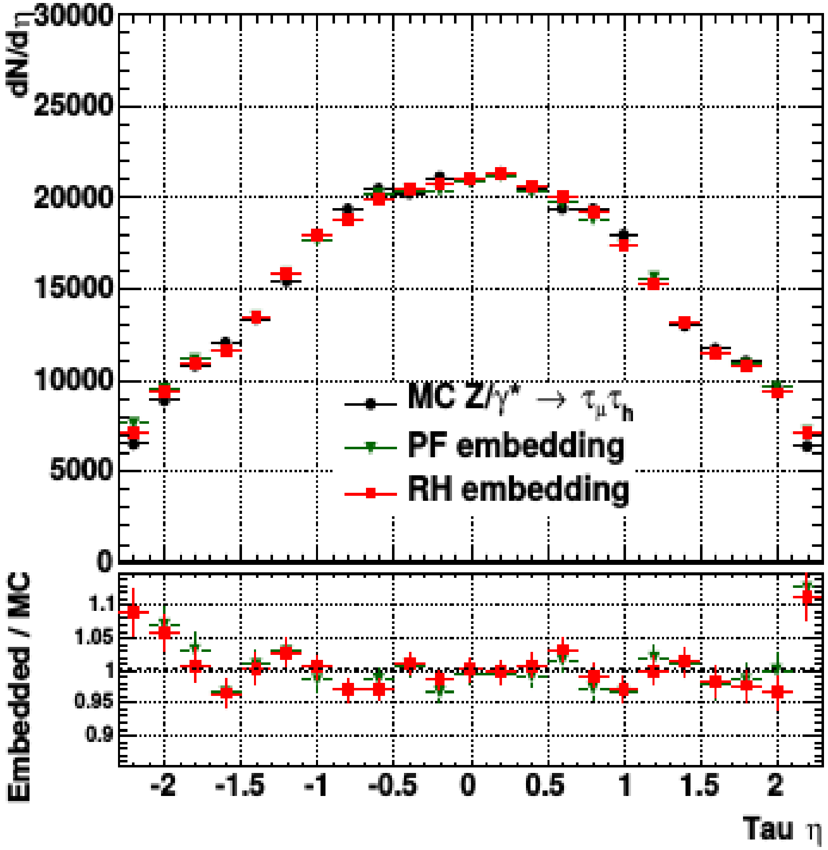
- leptonic part can be well and easily modeled
- important cross check of full simulation via MC



“Closure Test”

demonstrate that method works on simulated events





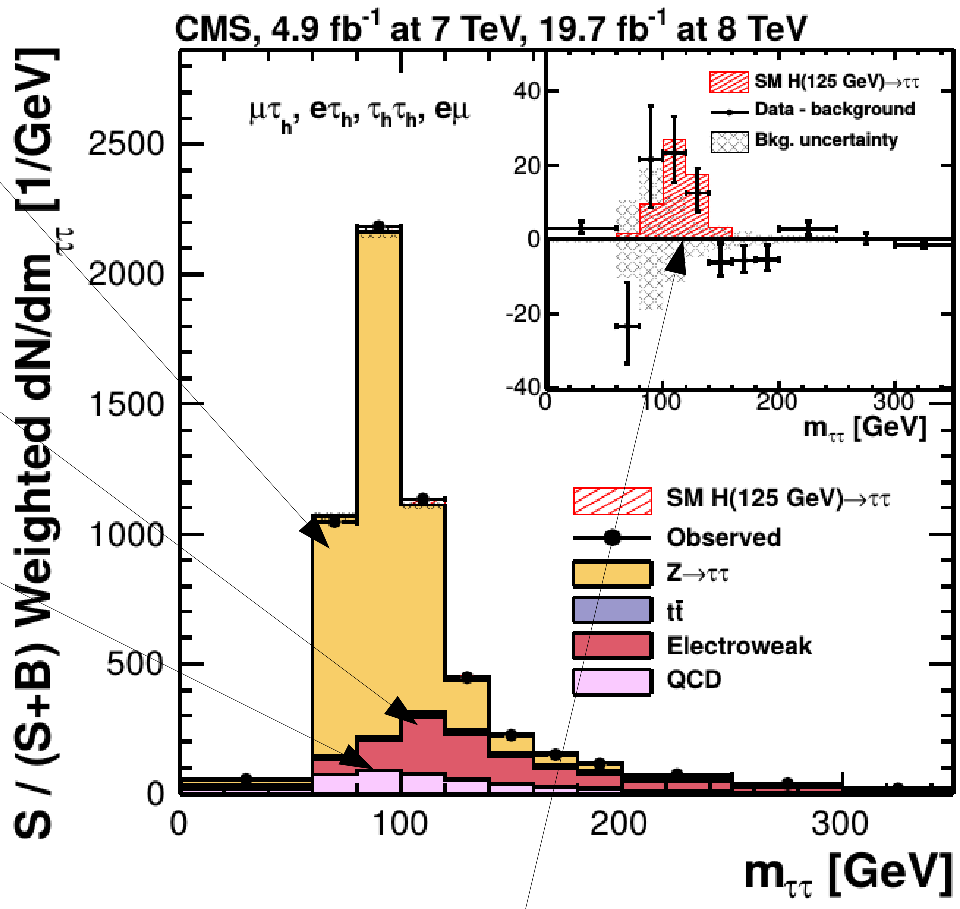
from PhD thesis Armin Burgmeier, Karlsruhe - DESY, June 2014

Summary and Outlook

Embedding /MC validated

Shape fit in "sideband" region

ABCD Method



Coming next:

statistical analysis of rare signals