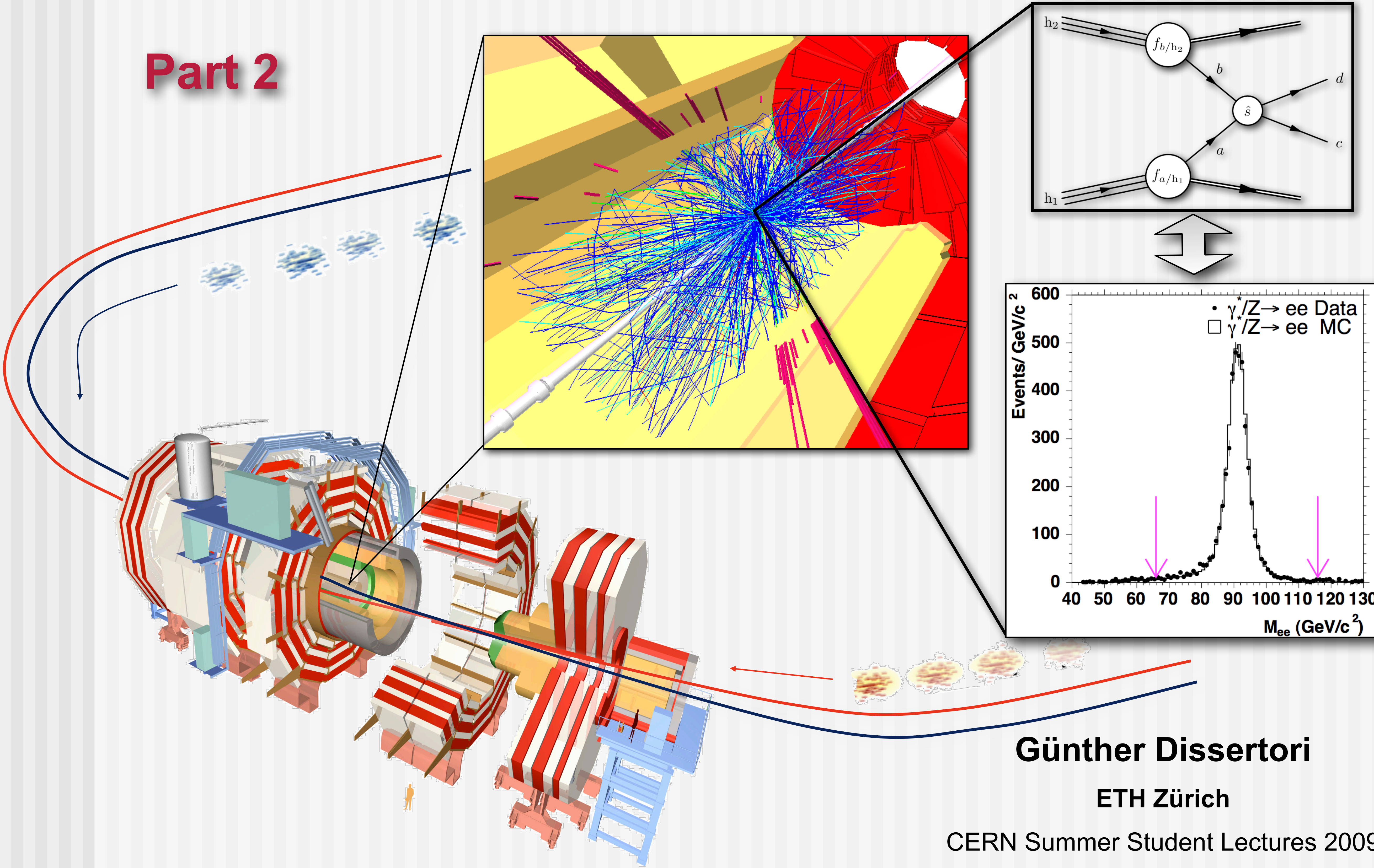
From Raw Data to Physics Results

Part 2



Günther Dissertori

ETH Zürich

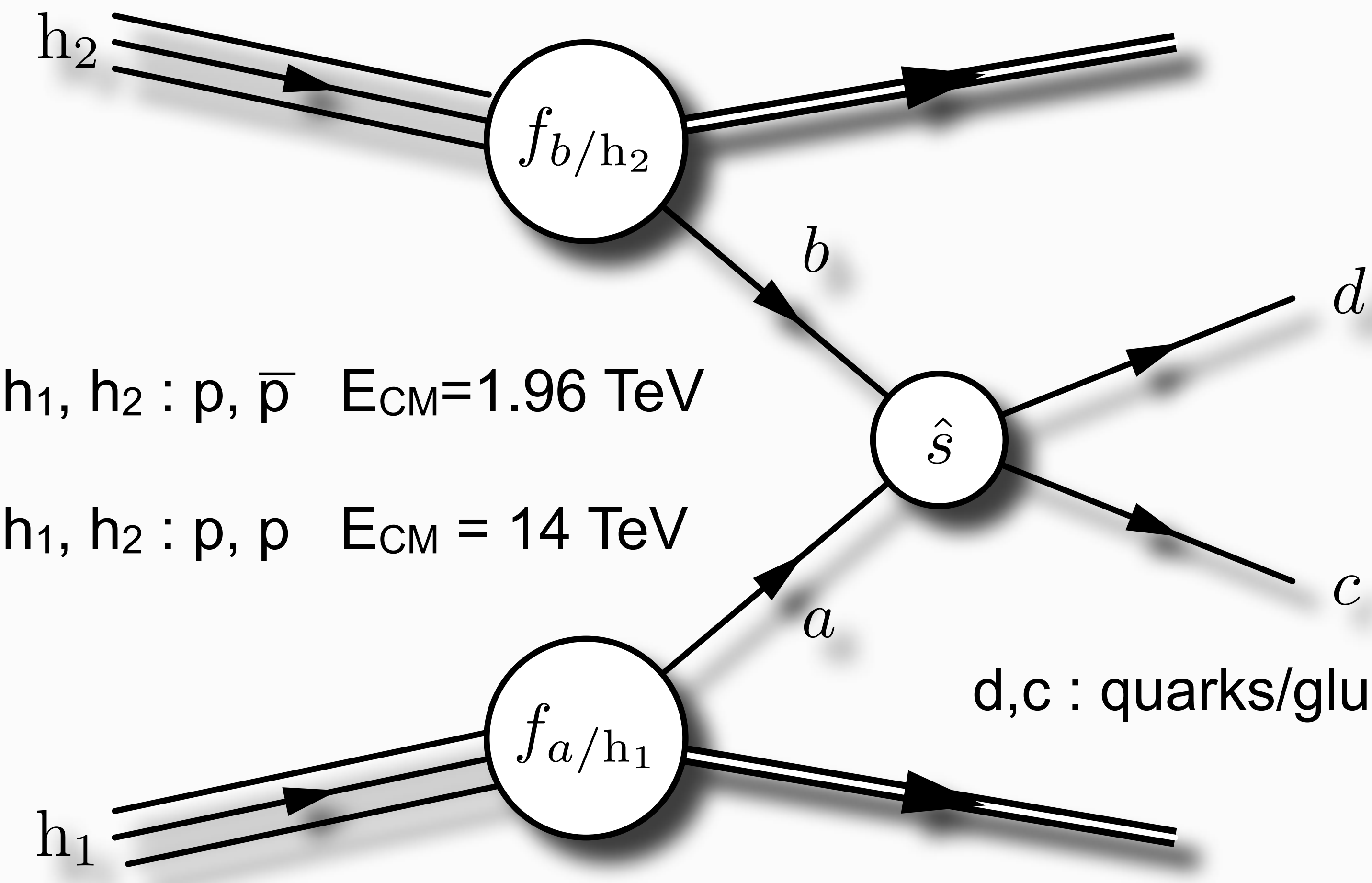
CERN Summer Student Lectures 2009

A "more complicated" example

"The greater the obstacle, the more glory in overcoming it." (Moliere)

JET production at hadron colliders

- at the Tevatron, or in the future at the LHC



h_1, h_2 : p, $\bar{p}$ E_{CM}=1.96 TeV

h_1, h_2 : p, p E_{CM} = 14 TeV

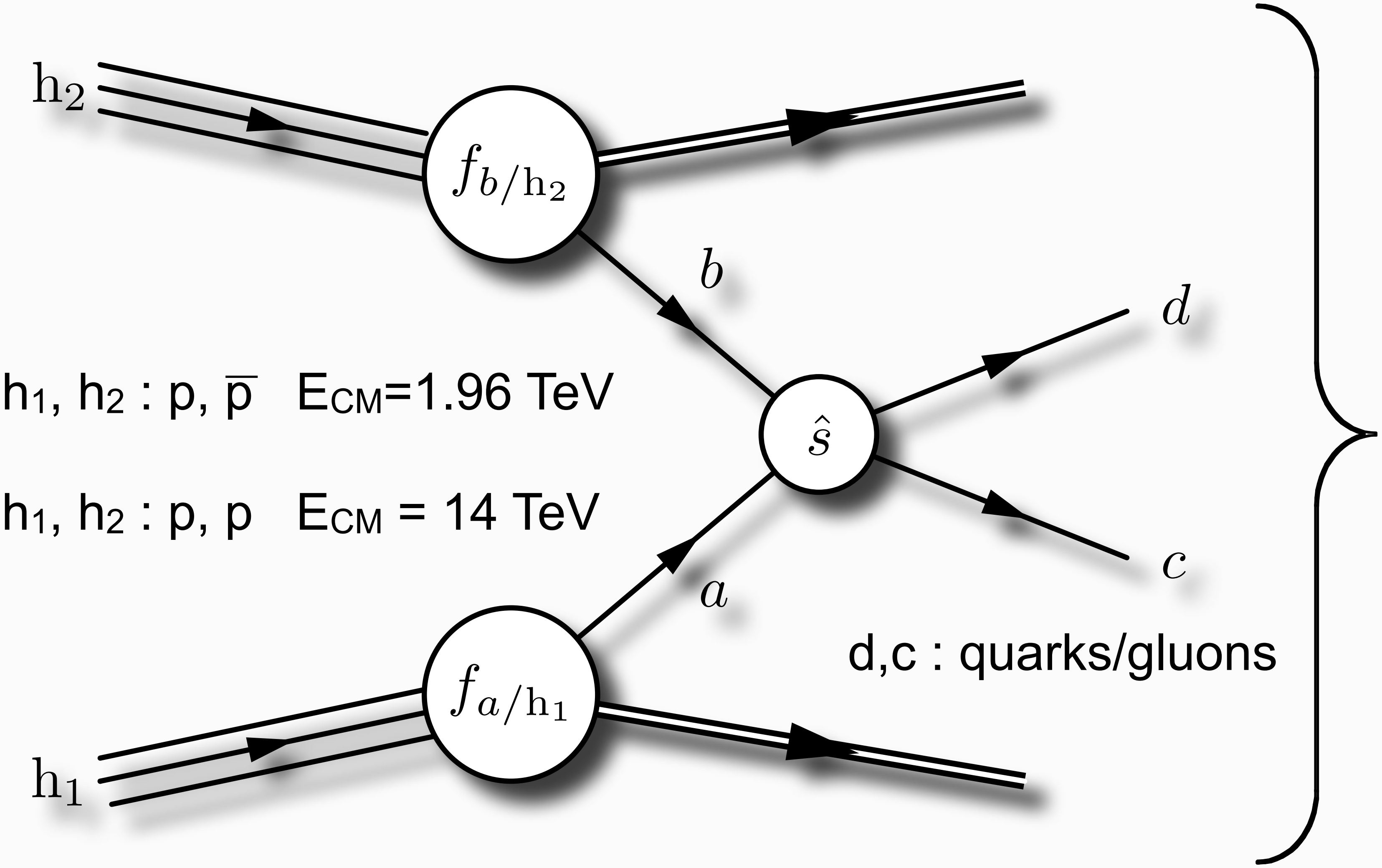
d,c : quarks/gluons

see also lecture by B. Heinemann

JET production at hadron colliders

see also lecture by B. Heinemann

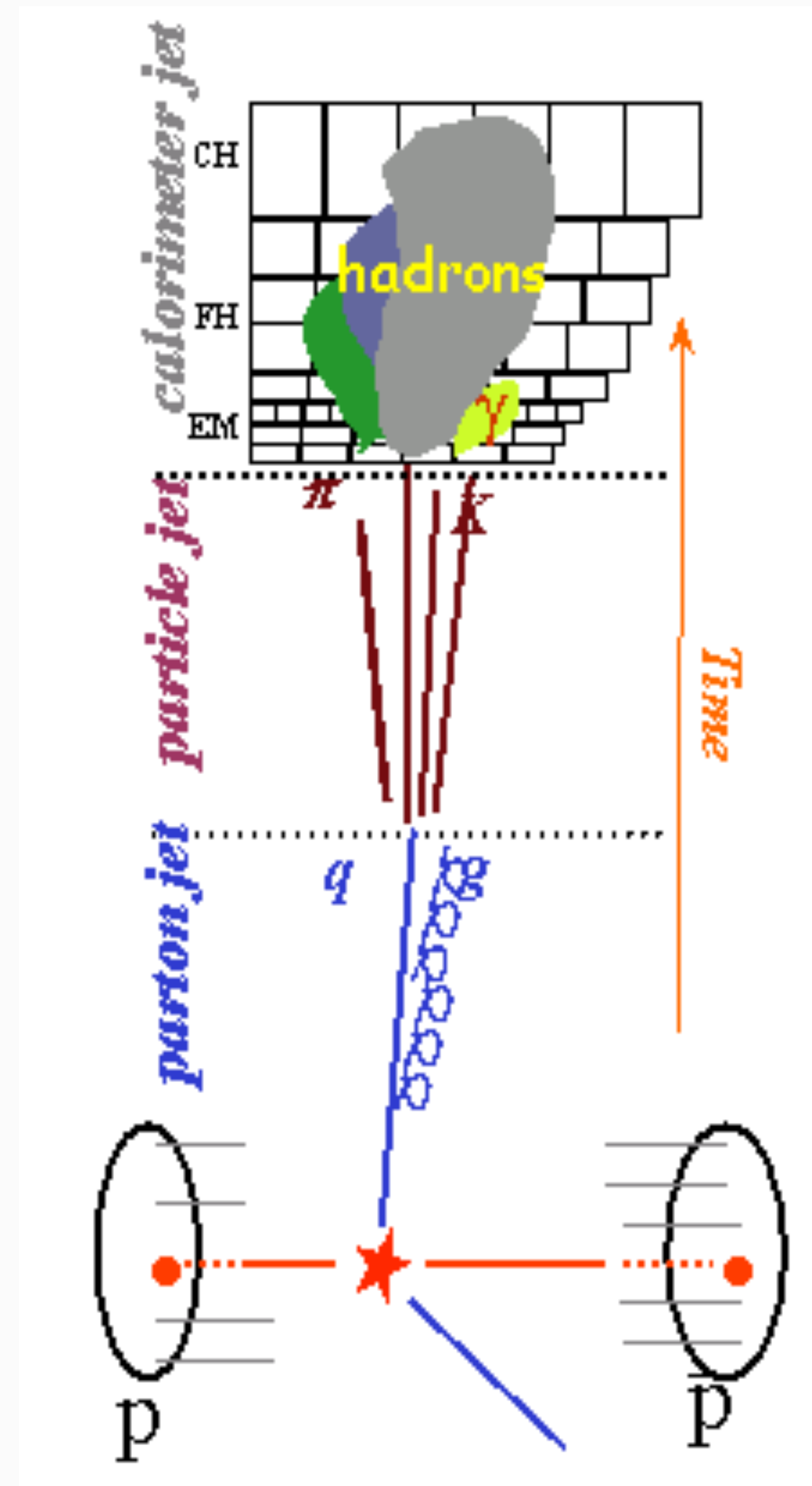
- at the Tevatron, or in the future at the LHC



h_1, h_2 : p, $\bar{p}$ E_{CM}=1.96 TeV

h_1, h_2 : p, p E_{CM} = 14 TeV

d,c : quarks/gluons

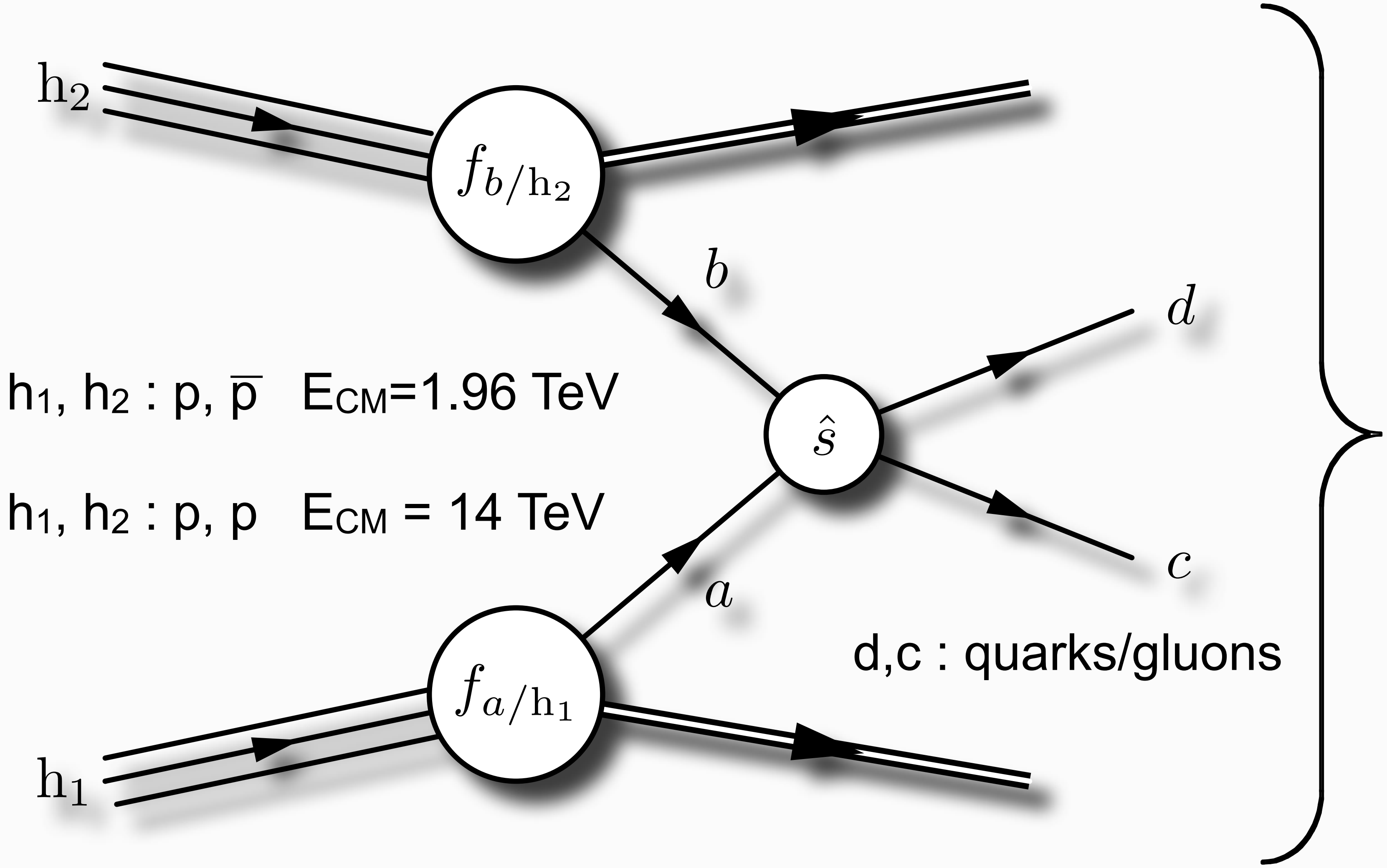


JET production at hadron colliders

ETH Institute for Particle Physics

see also lecture by B. Heinemann

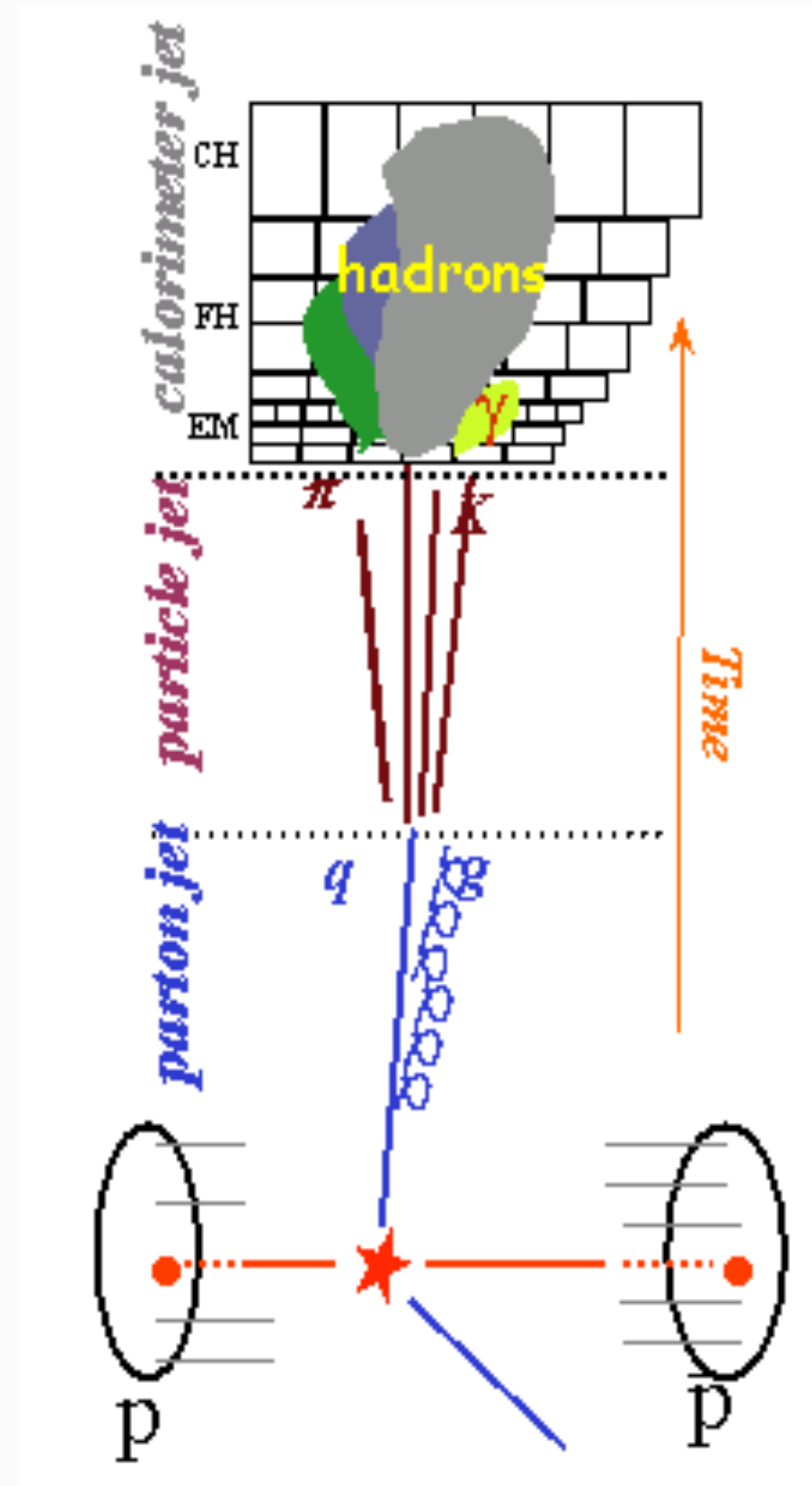
- at the Tevatron, or in the future at the LHC



h_1, h_2 : p, $\bar{p}$ E_{CM}=1.96 TeV

h_1, h_2 : p, p E_{CM} = 14 TeV

d,c : quarks/gluons

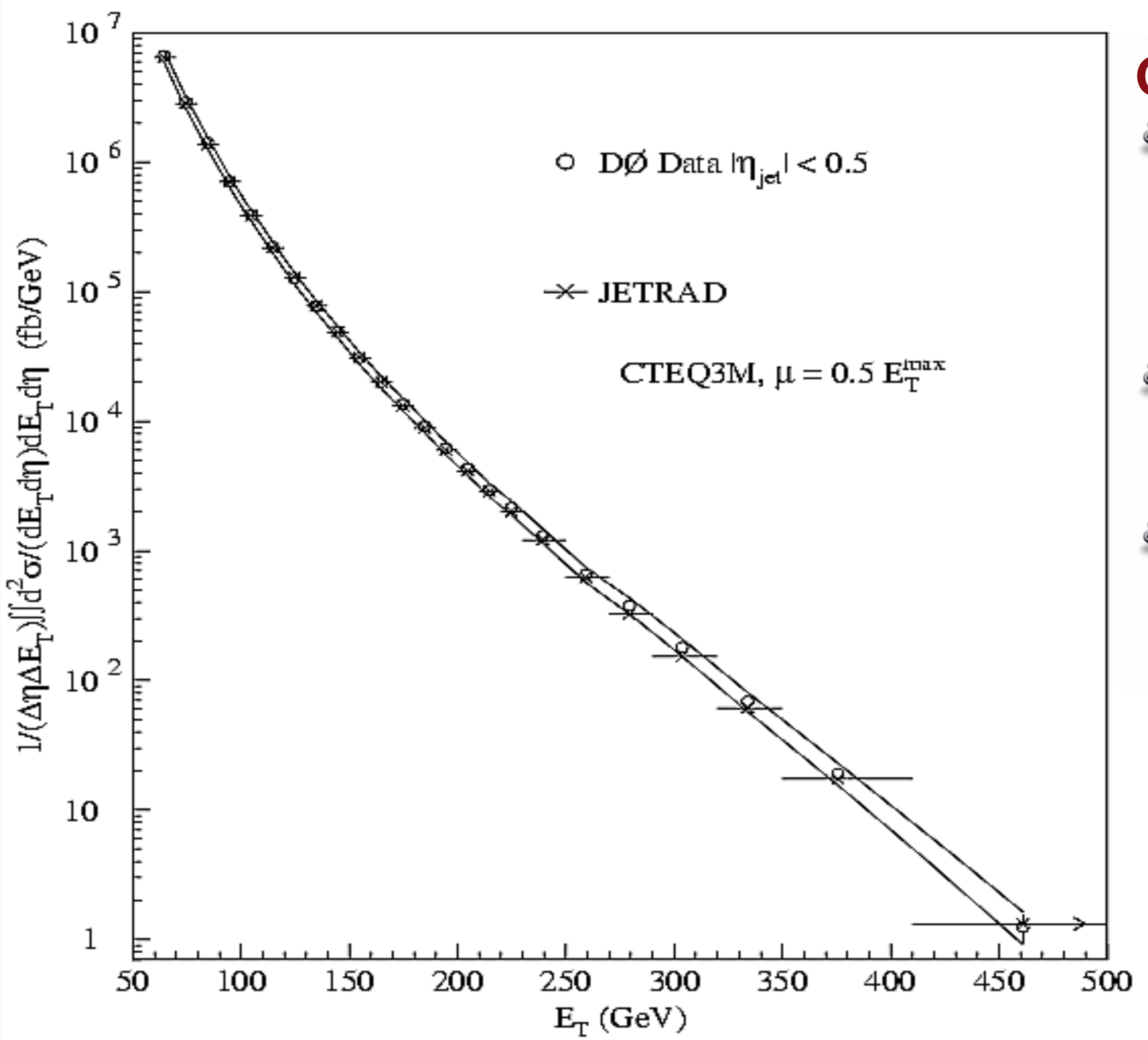


Goal

- measure probability that quarks/gluons are produced with a certain energy, at a certain angle
- Problem : do not observe quarks and gluons directly, only hadrons, which appear collimated into **jets**
- Reconstruct tracks and/or energy clusters in the calorimeter

What do we have to measure?

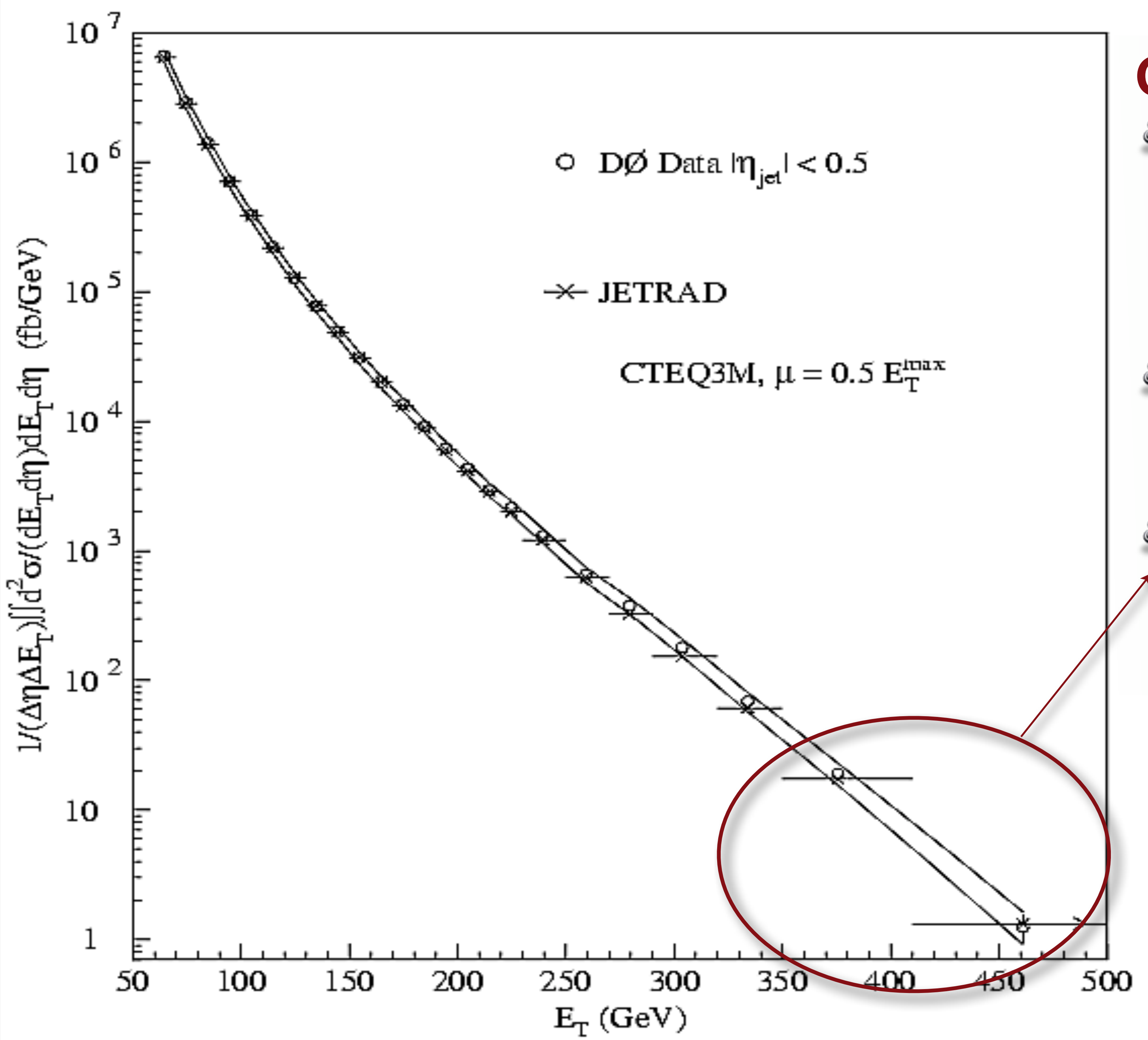
ETH Institute for Particle Physics



Goal

- measure cross section (probability) that jets are produced with a certain transverse energy E_T, within a certain rapidity range
- Test of perturbative QCD, over many orders of magnitude!
- Look at very high energy tail, new physics could show up there in form of excess (eg. sub-structure of quarks?)

What do we have to measure?

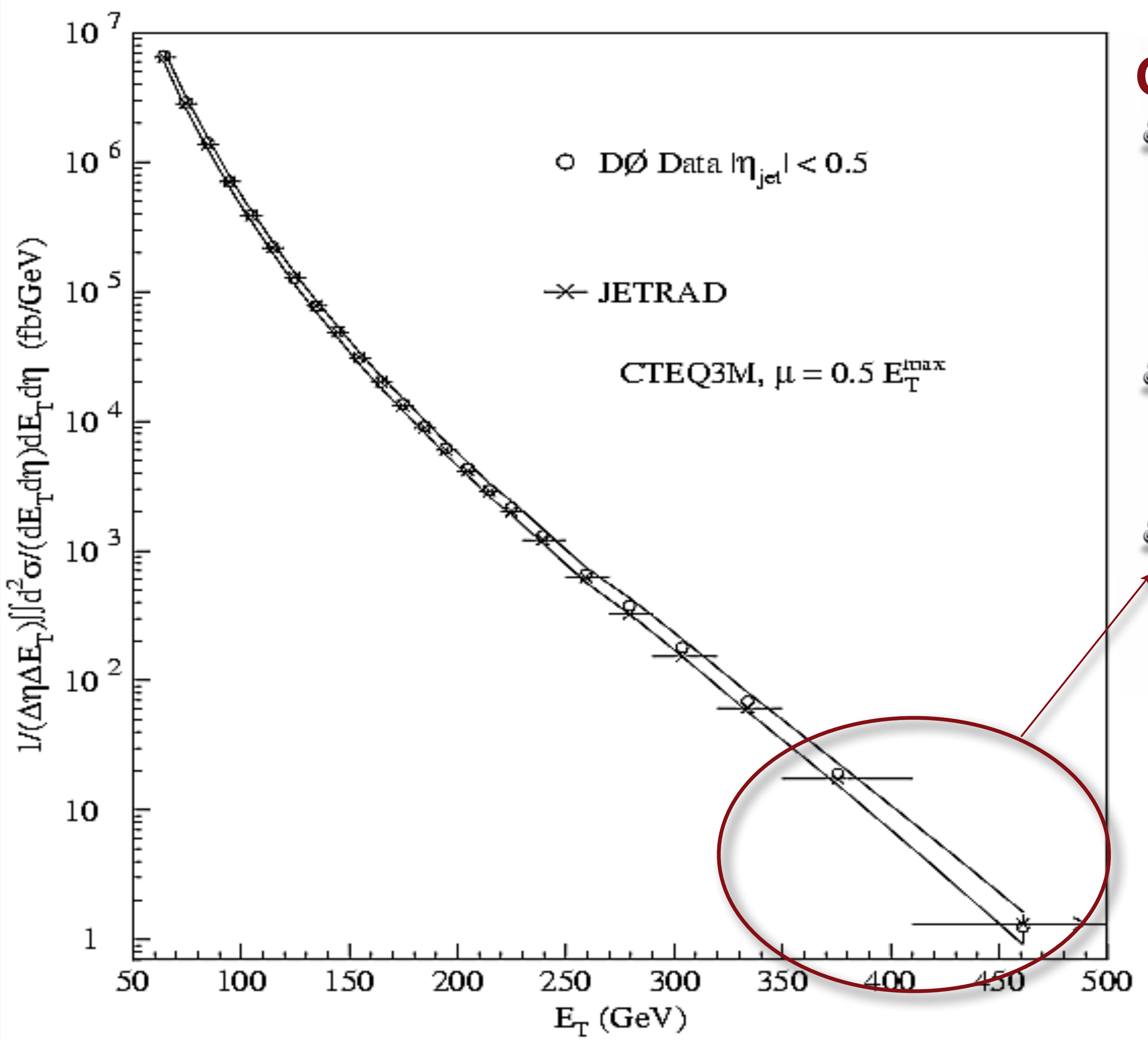


Goal

- measure cross section (probability) that jets are produced with a certain transverse energy E_T, within a certain rapidity range
- Test of perturbative QCD, over many orders of magnitude!
- Look at very high energy tail, new physics could show up there in form of excess (eg. sub-structure of quarks?)

What do we have to measure?

ETH Institute for Particle Physics



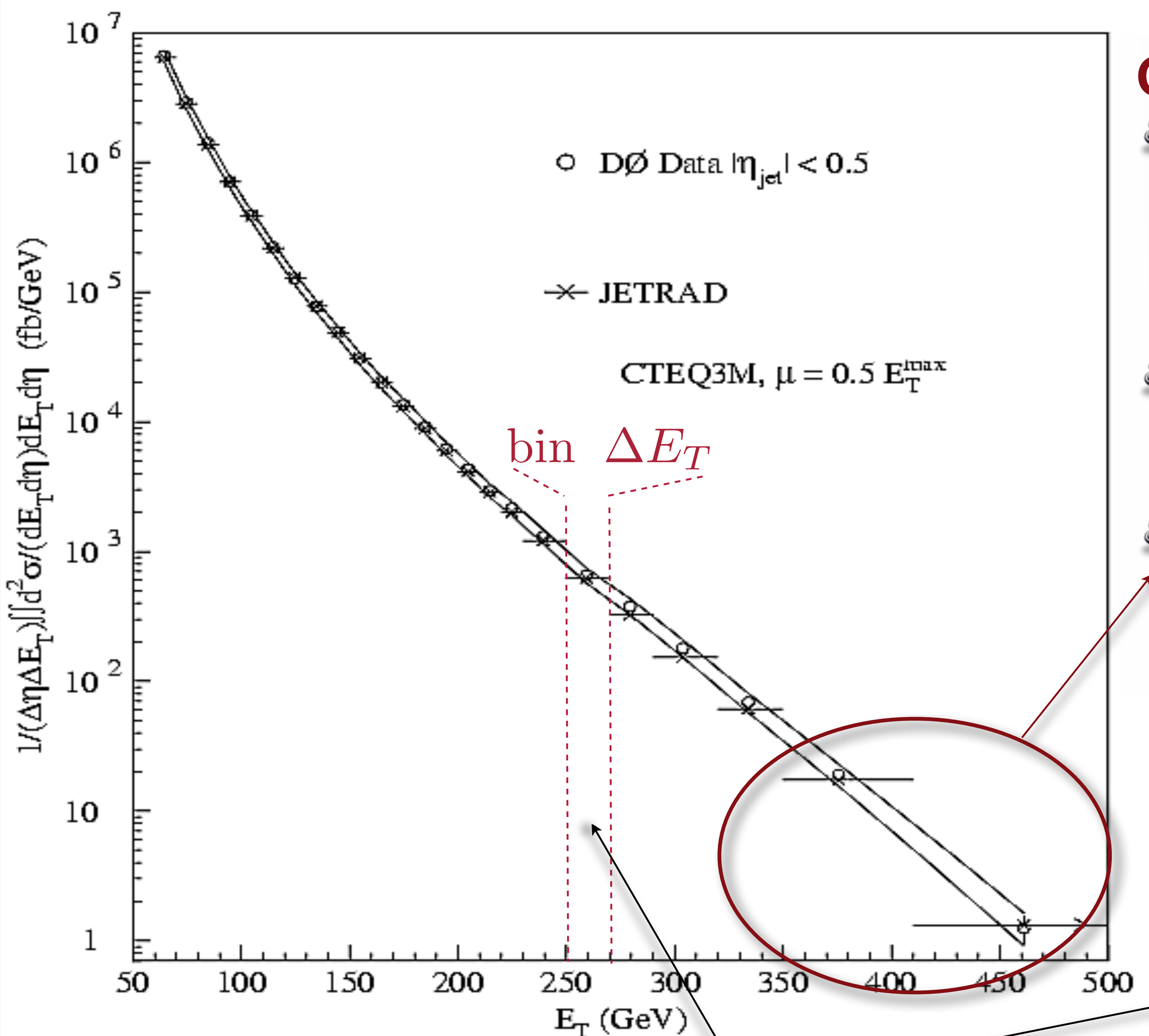
Goal

- measure cross section (probability) that jets are produced with a certain transverse energy E_T, within a certain rapidity range
- Test of perturbative QCD, over many orders of magnitude!
- Look at very high energy tail, new physics could show up there in form of excess (eg. sub-structure of quarks?)

can be calculated in pert. QCD

$$\left\langle \frac{d^2\sigma}{dE_T d\eta} \right\rangle = \frac{N}{\Delta E_T \Delta\eta\, \epsilon\, \mathcal{L}}$$

What do we have to measure?



Goal

- measure cross section (probability) that jets are produced with a certain transverse energy E_T, within a certain rapidity range
- Test of perturbative QCD, over many orders of magnitude!
- Look at very high energy tail, new physics could show up there in form of excess (eg. sub-structure of quarks?)

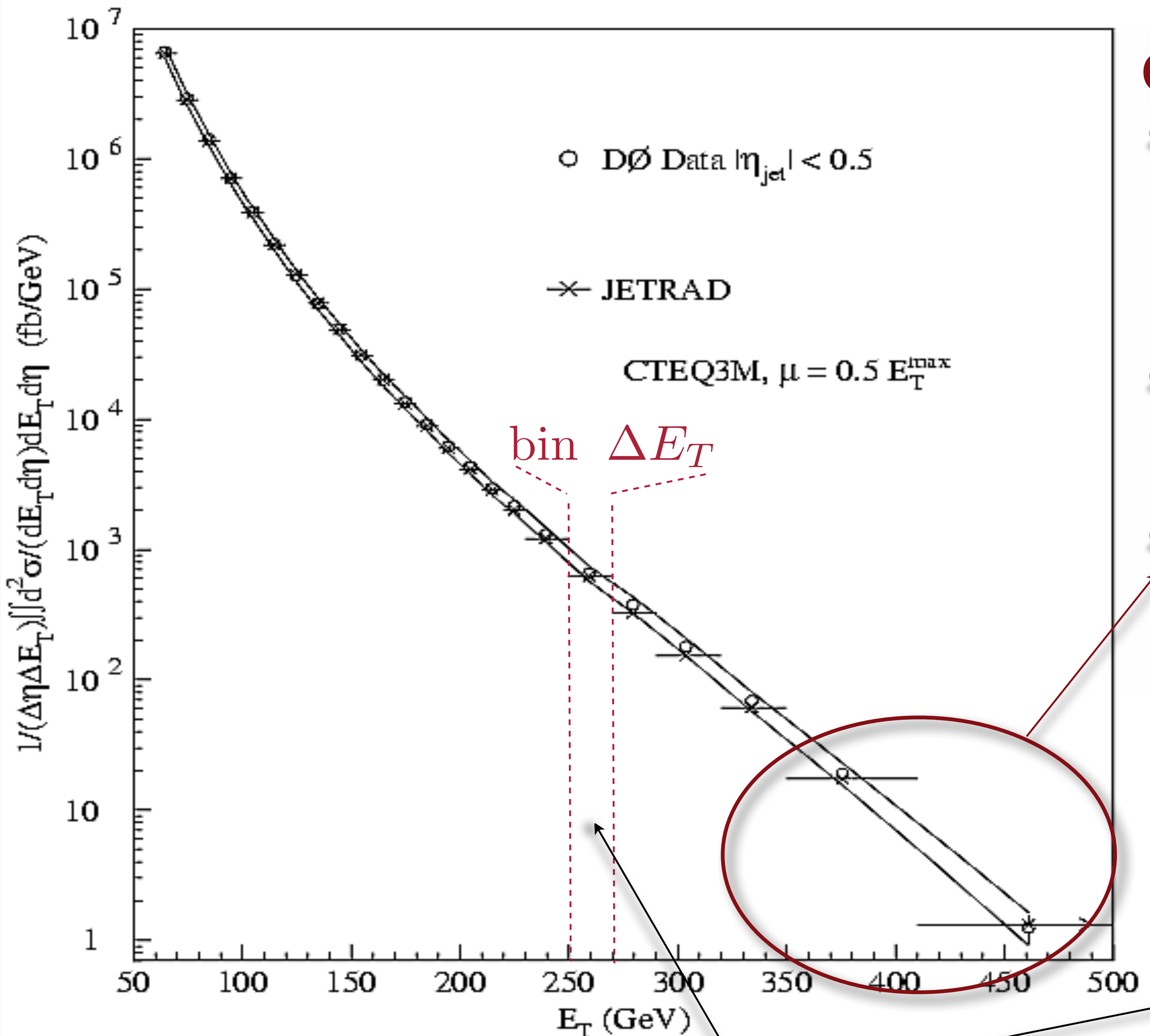
can be calculated in pert. QCD

$$\left\langle \frac{d^2\sigma}{dE_T d\eta} \right\rangle = \frac{N}{\Delta E_T \Delta\eta\, \epsilon\, \mathcal{L}}$$

- count number of events, N, in this bin
- for a certain range in rapidity (angle) $\Delta\eta$

What do we have to measure?

ETH Institute for Particle Physics



Goal

- measure cross section (probability) that jets are produced with a certain transverse energy E_T, within a certain rapidity range
- Test of perturbative QCD, over many orders of magnitude!
- Look at very high energy tail, new physics could show up there in form of excess (eg. sub-structure of quarks?)

can be calculated in pert. QCD

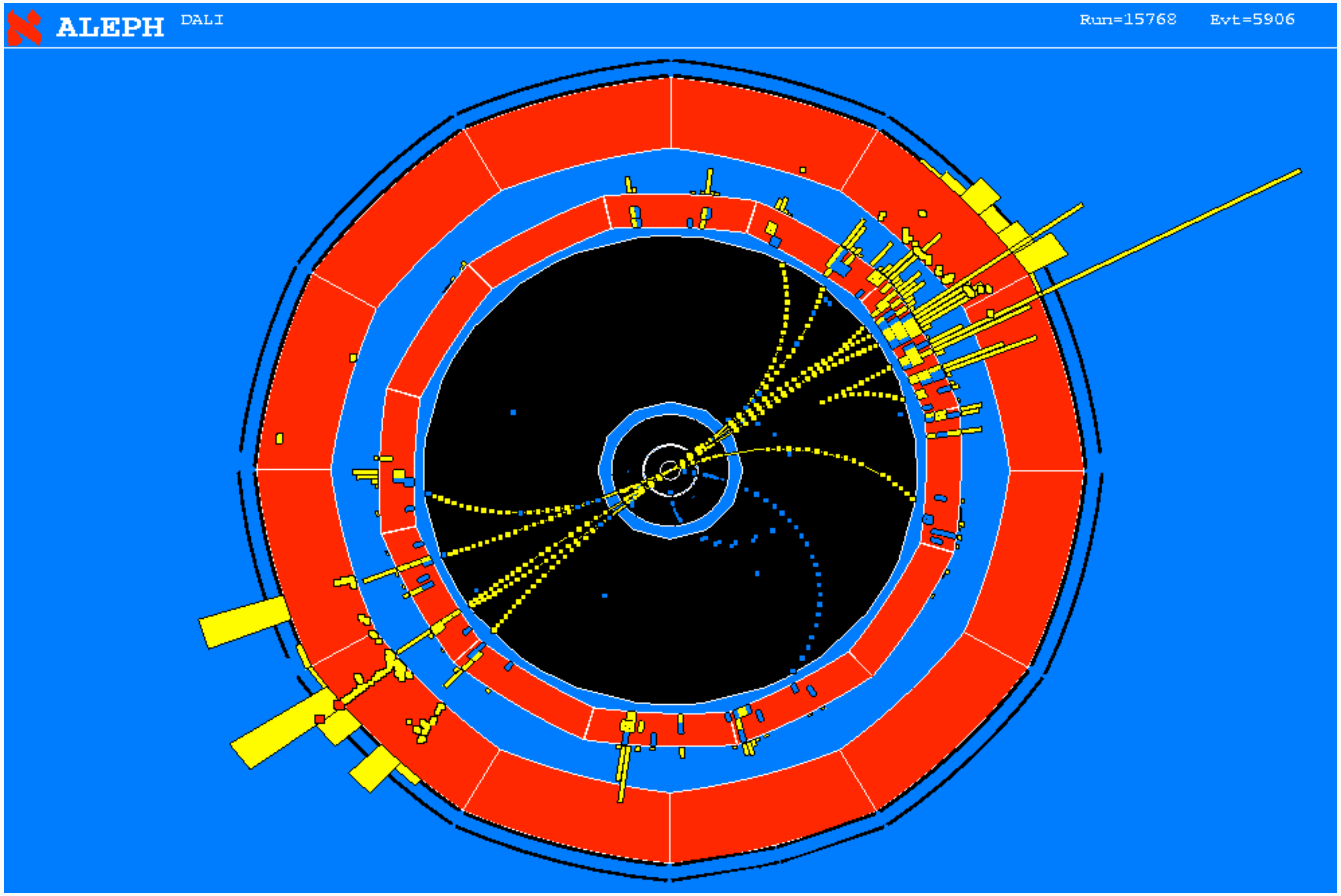
$$\left\langle \frac{d^2\sigma}{dE_T d\eta} \right\rangle = \frac{N}{\Delta E_T \Delta\eta\, \epsilon\, \mathcal{L}}$$

- count number of events, N, in this bin
- for a certain range in rapidity (angle) $\Delta\eta$

efficiency to reconstruct jets

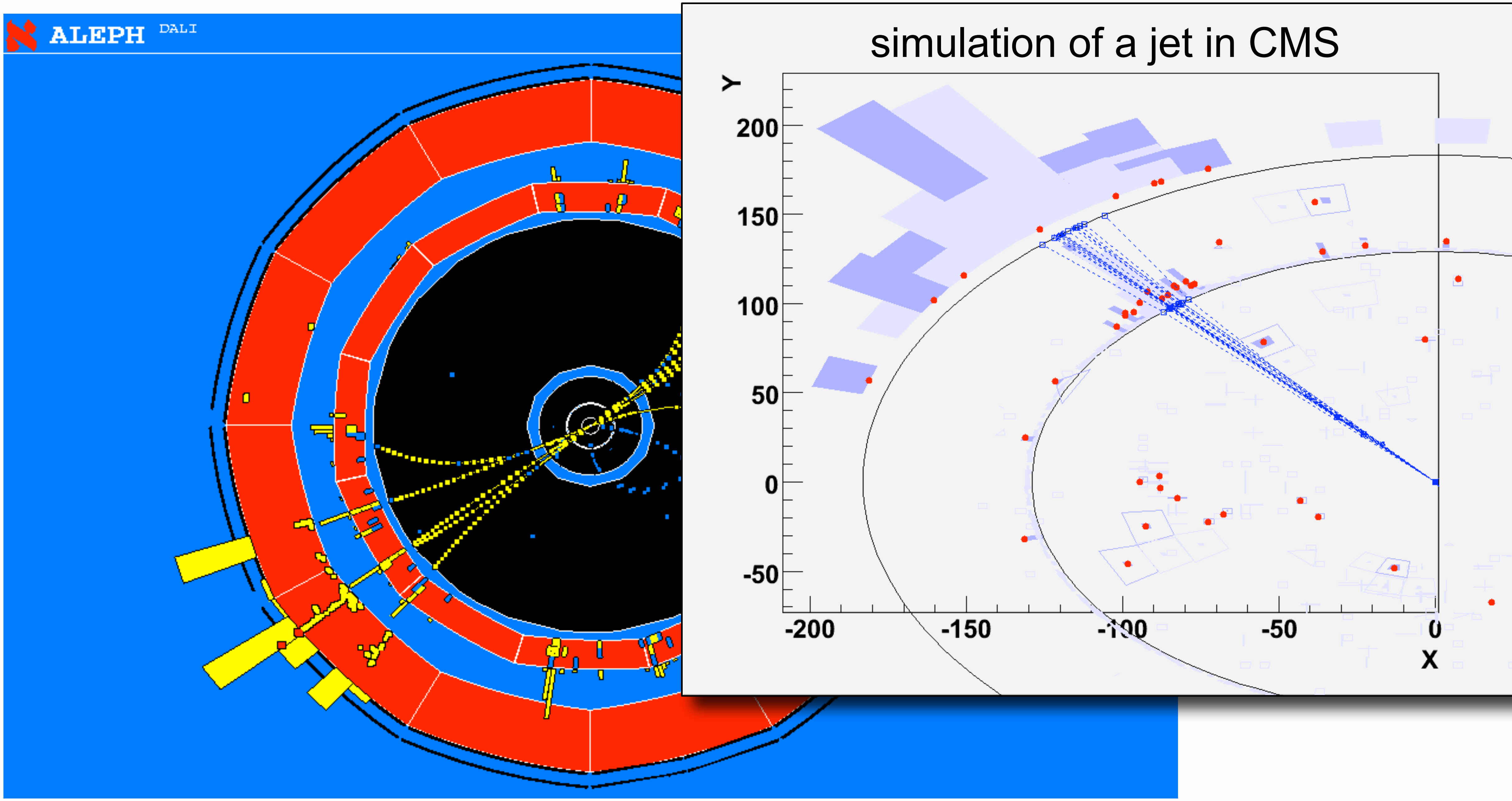
integrated accelerator luminosity

What is a jet?



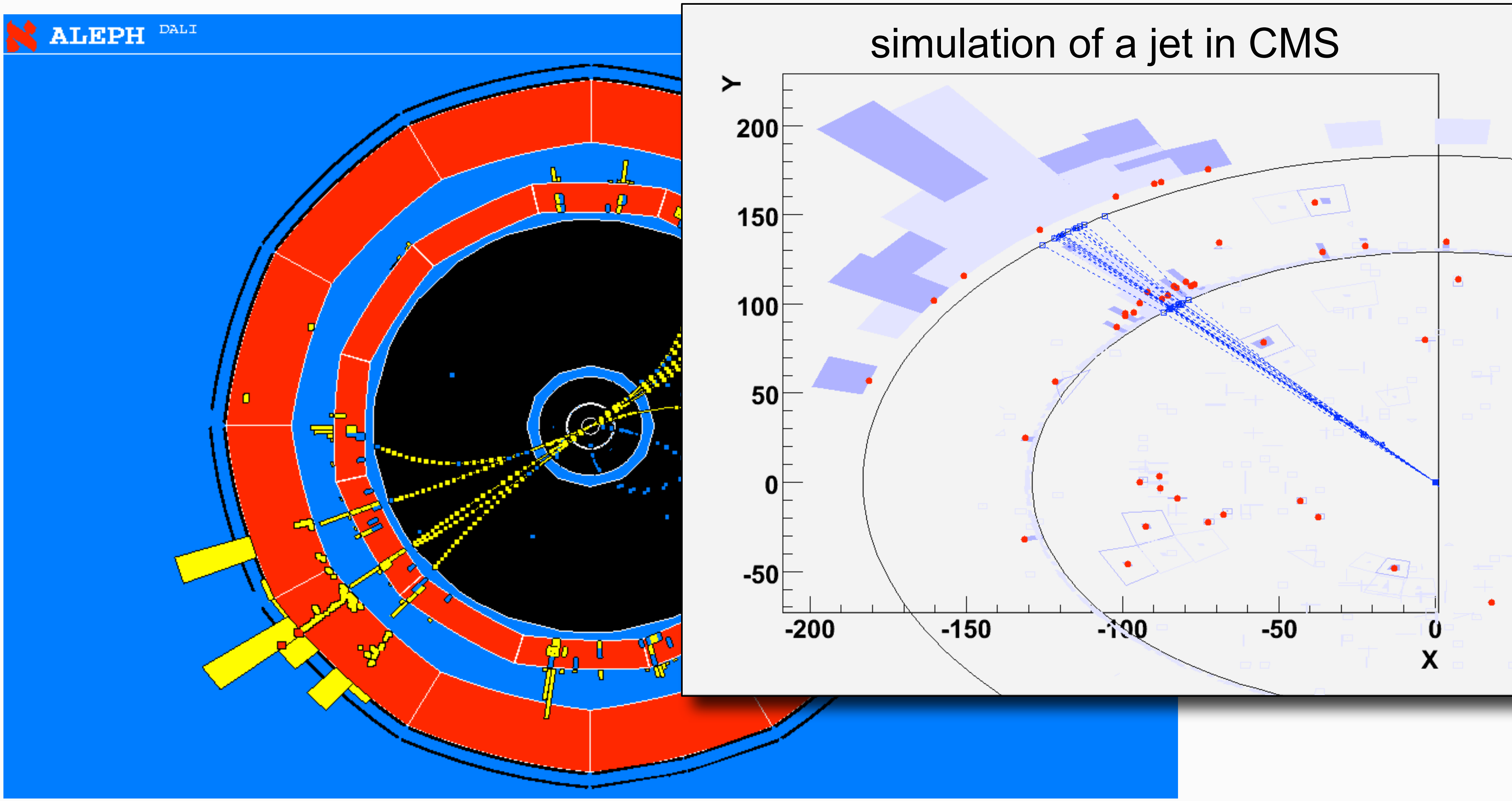
"cluster/spray of particles (tracks, calorimeter deposits) or flow of energy in a restricted angular region"

What is a jet?



“cluster/spray of particles (tracks, calorimeter deposits) or flow of energy in a restricted angular region”

What is a jet?

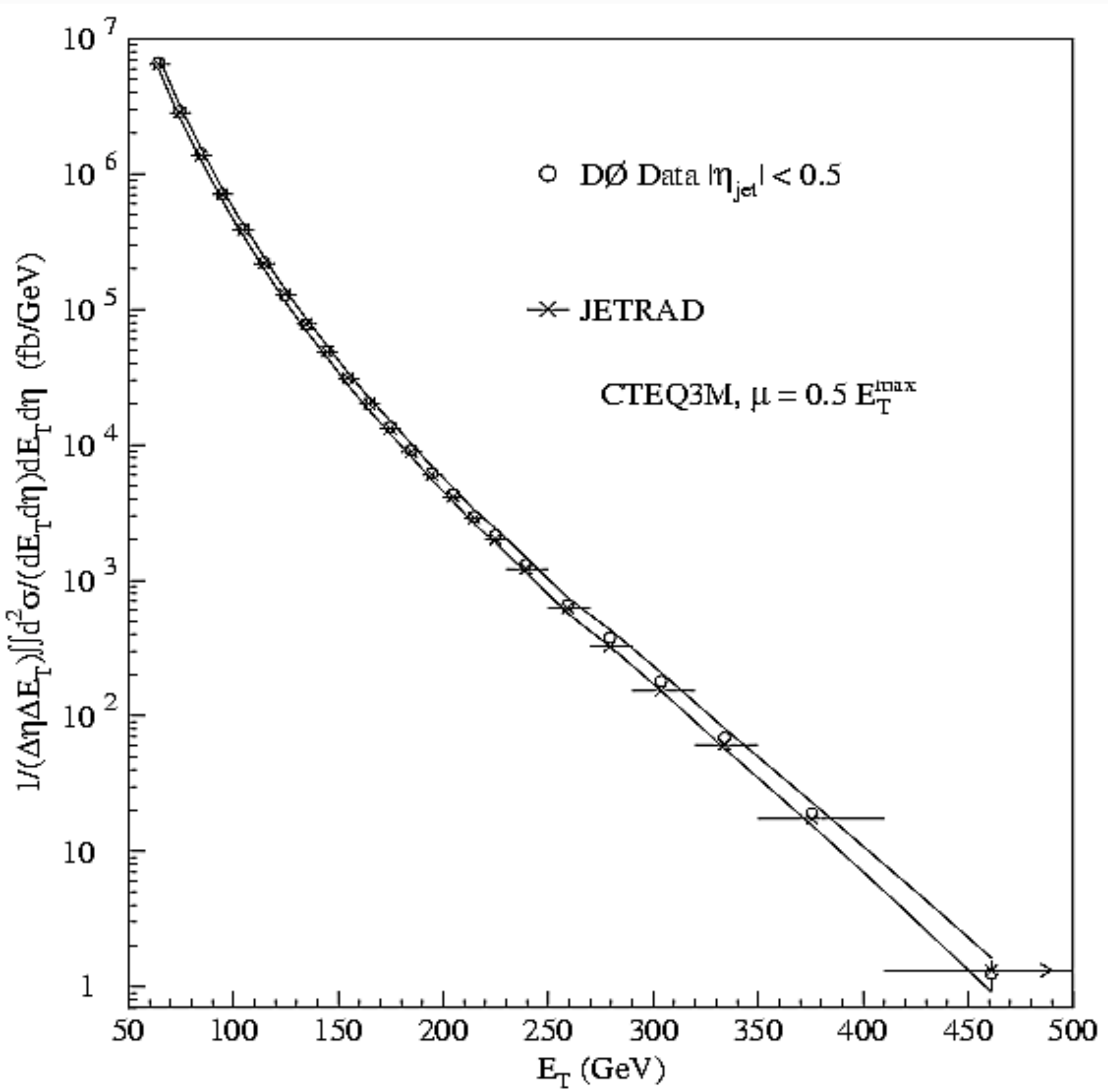


“cluster/spray of particles (tracks, calorimeter deposits) or flow of energy in a restricted angular region”

- clear : need some algorithmic definition. See later..

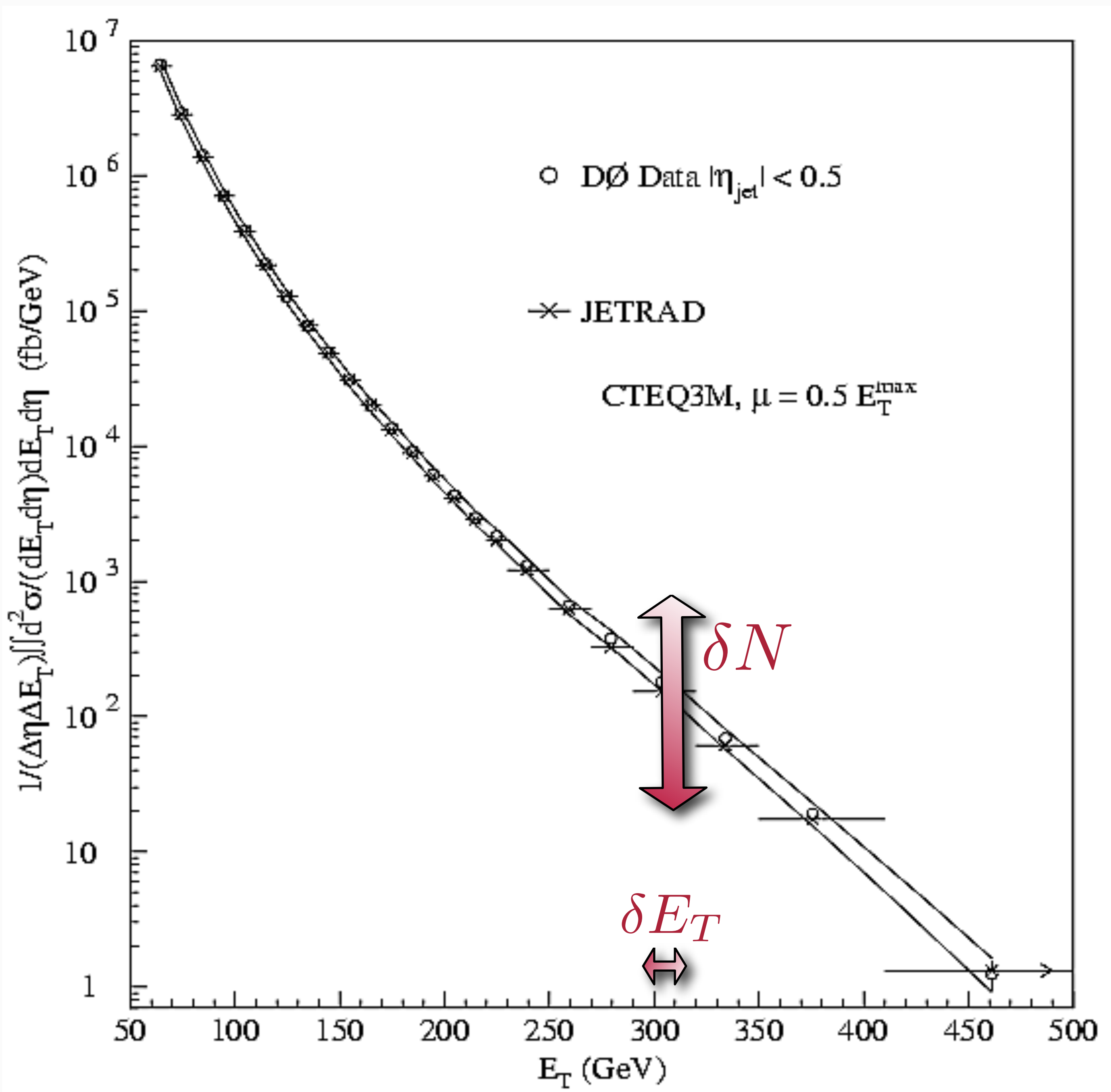
Problem 1 : Energy scale

- Question : how well do we know the energy calibration?
- Critical because of very steeply falling spectrum!



Problem 1 : Energy scale

- Question : how well do we know the energy calibration?
- Critical because of very steeply falling spectrum!



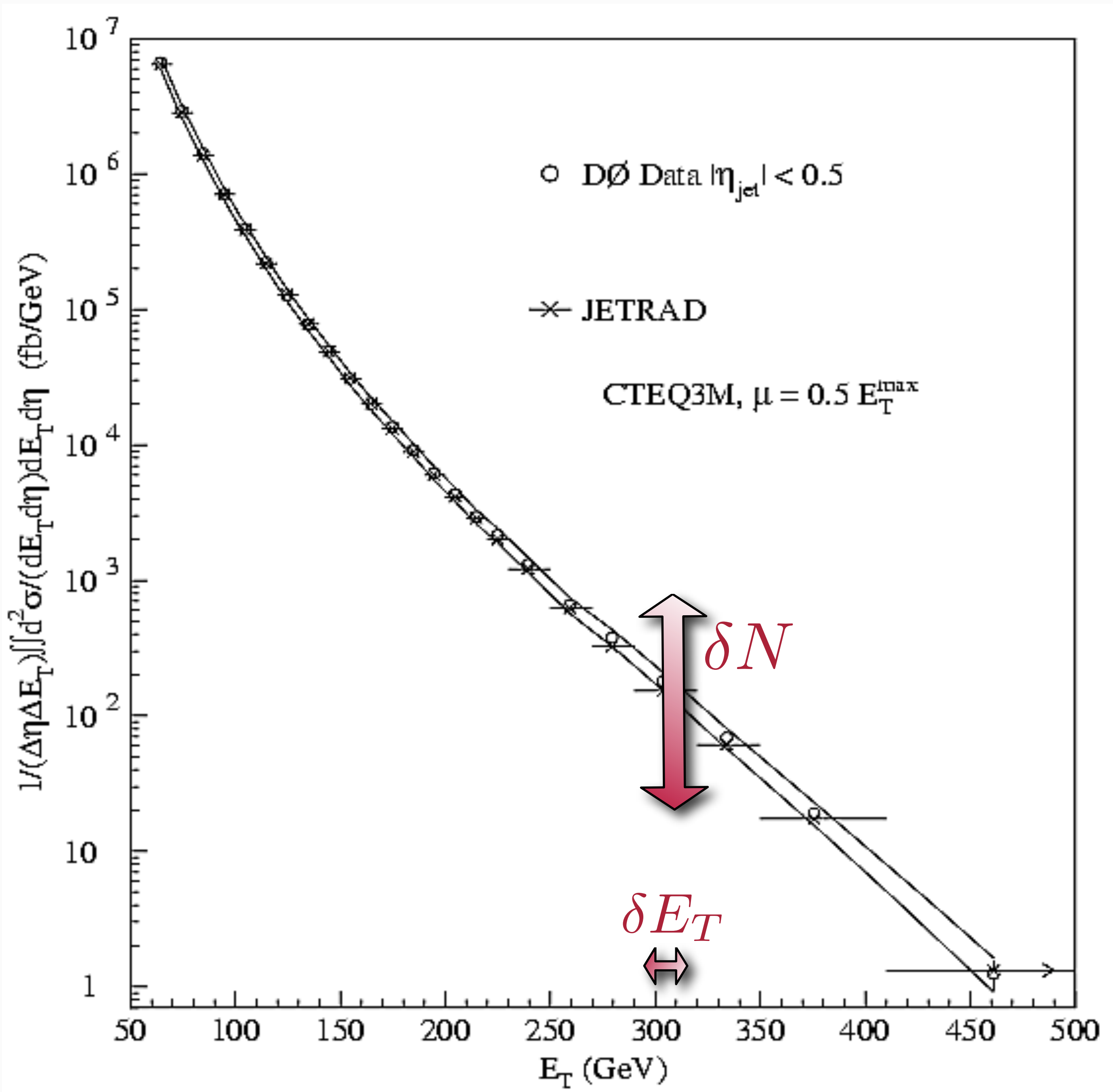
$$\frac{d^2\sigma}{dE_T d\eta} \approx \text{const} \cdot E_T^{\ -6}$$

relative uncertainties

$$\frac{\delta N}{N} \approx 6 \cdot \frac{\delta E_T}{E_T}$$

Problem 1 : Energy scale

- Question : how well do we know the energy calibration?
- Critical because of very steeply falling spectrum!



$$\frac{d^2\sigma}{dE_T d\eta} \approx \text{const} \cdot E_T{}^{-6}$$

relative uncertainties

$$\frac{\delta N}{N} \approx 6 \cdot \frac{\delta E_T}{E_T}$$

so beware:

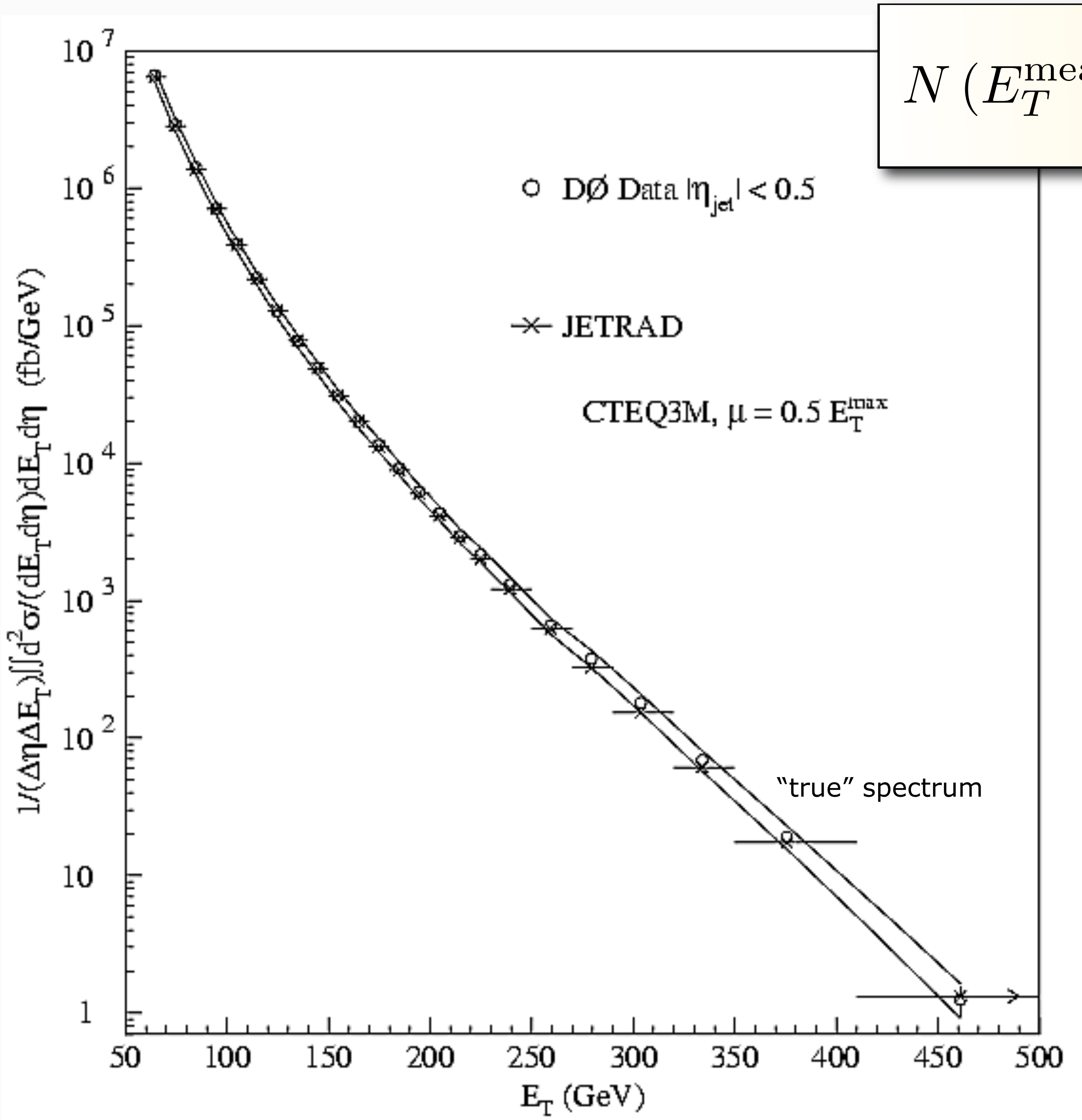
eg. an uncertainty of **5%** on absolute energy scale (calibration)

➔ an uncertainty of **30%** (!) on the measured cross section

Problem 2 : Energy resolution

- The energy resolution can distorts the spectrum
- Again : Critical because of very steeply falling spectrum!

$$N\left(E_T^{\mathrm{meas}}\right) = \int_0^\infty N\left(E_T^{\mathrm{true}}\right) \cdot \mathrm{Resol}\left(E_T^{\mathrm{meas}}, E_T^{\mathrm{true}}\right) \mathrm{d}E_T^{\mathrm{true}}$$



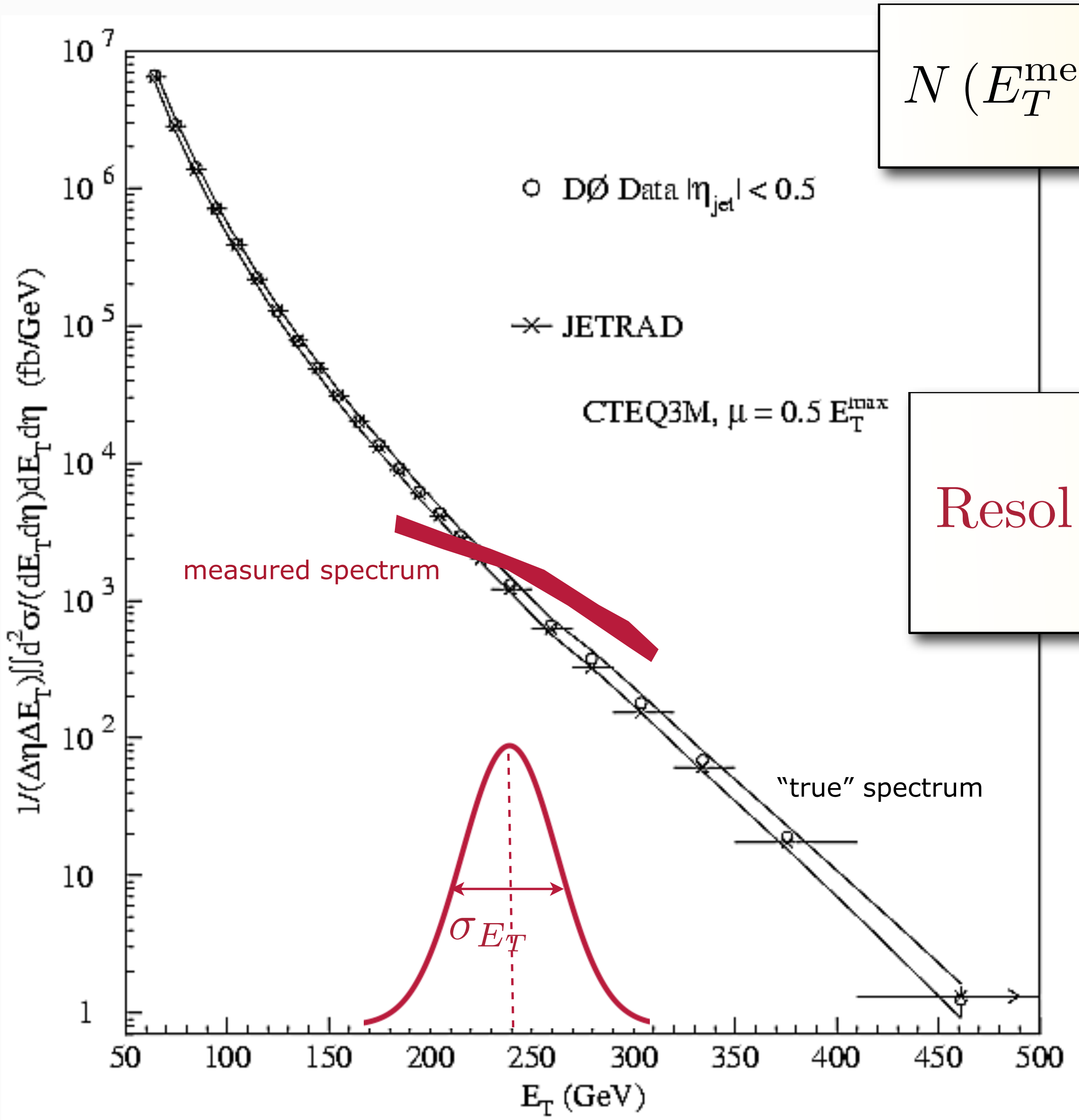
Problem 2 : Energy resolution

- The energy resolution can distorts the spectrum
- Again : Critical because of very steeply falling spectrum!

$$N\left(E_T^{\mathrm{meas}}\right) = \int_0^{\infty} N\left(E_T^{\mathrm{true}}\right) \cdot \mathrm{Resol}\left(E_T^{\mathrm{meas}}, E_T^{\mathrm{true}}\right) \mathrm{d}E_T^{\mathrm{true}}$$

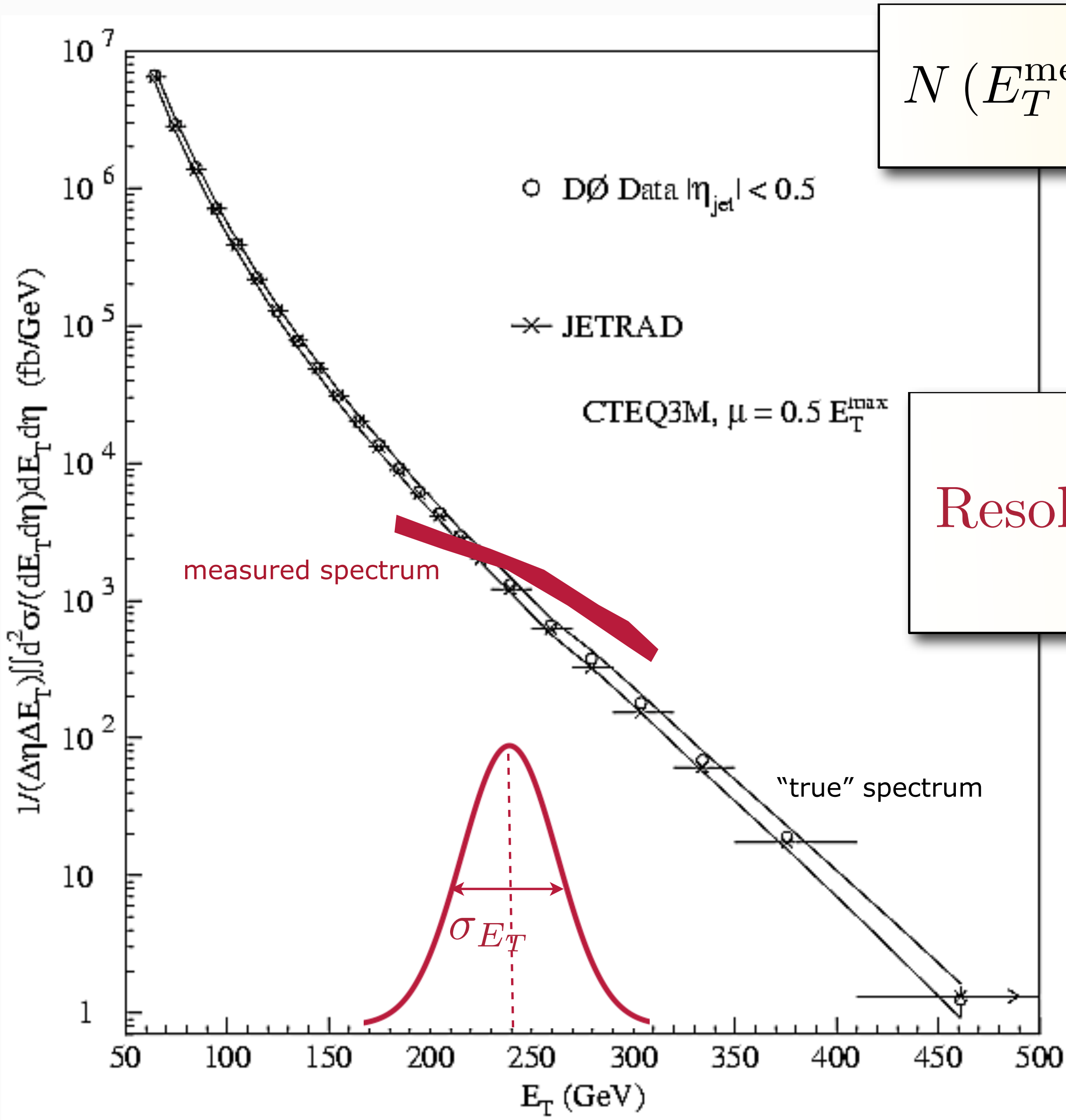
eg. Gaussian resolution function

$$\mathrm{Resol}\left(E_T^{\mathrm{meas}}, E_T^{\mathrm{true}}\right) \propto \exp\left[-\frac{\left(E_T^{\mathrm{meas}} - E_T^{\mathrm{true}}\right)^2}{\sigma_{E_T}^2}\right]$$



Problem 2 : Energy resolution

- The energy resolution can distorts the spectrum
- Again : Critical because of very steeply falling spectrum!



$$N\left(E_T^{\text{meas}}\right) = \int_0^\infty N\left(E_T^{\text{true}}\right) \cdot \text{Resol}\left(E_T^{\text{meas}}, E_T^{\text{true}}\right) \mathrm{d}E_T^{\text{true}}$$

eg. Gaussian resolution function

$$\text{Resol}\left(E_T^{\text{meas}}, E_T^{\text{true}}\right) \propto \exp\left[-\frac{\left(E_T^{\text{meas}} - E_T^{\text{true}}\right)^2}{\sigma_{E_T}^2}\right]$$

so beware:

A bad energy resolution can distort the true spectrum

➔ have to determine the energy resolution

➔ have to "unfold" the measured spectrum

➔ problem is minimized if bin width ~ σ_{E_T}

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate "high-level" physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured
  - apply event selection (cuts)

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured
  - apply event selection (cuts)
  - estimate efficiencies and backgrounds, eg. from MC simulation

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate "high-level" physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured
  - apply event selection (cuts)
  - estimate efficiencies and backgrounds, eg. from MC simulation
  - if distributions are measured : take care of absolute calibrations and effects because of detector resolution/smearing
    - correct for these effects

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured
  - apply event selection (cuts)
  - estimate efficiencies and backgrounds, eg. from MC simulation
  - if distributions are measured : take care of absolute calibrations and effects because of detector resolution/smearing
    - correct for these effects
  - determine statistical and systematic uncertainties

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate "high-level" physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured
  - apply event selection (cuts)
  - estimate efficiencies and backgrounds, eg. from MC simulation
  - if distributions are measured : take care of absolute calibrations and effects because of detector resolution/smearing
    - correct for these effects
  - determine statistical and systematic uncertainties
  - compare with theory, found a deviation, something new?

Where do we stand now?

- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured
  - apply event selection (cuts)
  - estimate efficiencies and backgrounds, eg. from MC simulation
  - if distributions are measured : take care of absolute calibrations and effects because of detector resolution/smearing
    - correct for these effects
  - determine statistical and systematic uncertainties
  - compare with theory, found a deviation, something new?
    - if yes, book the ticket to Stockholm

Where do we stand now?

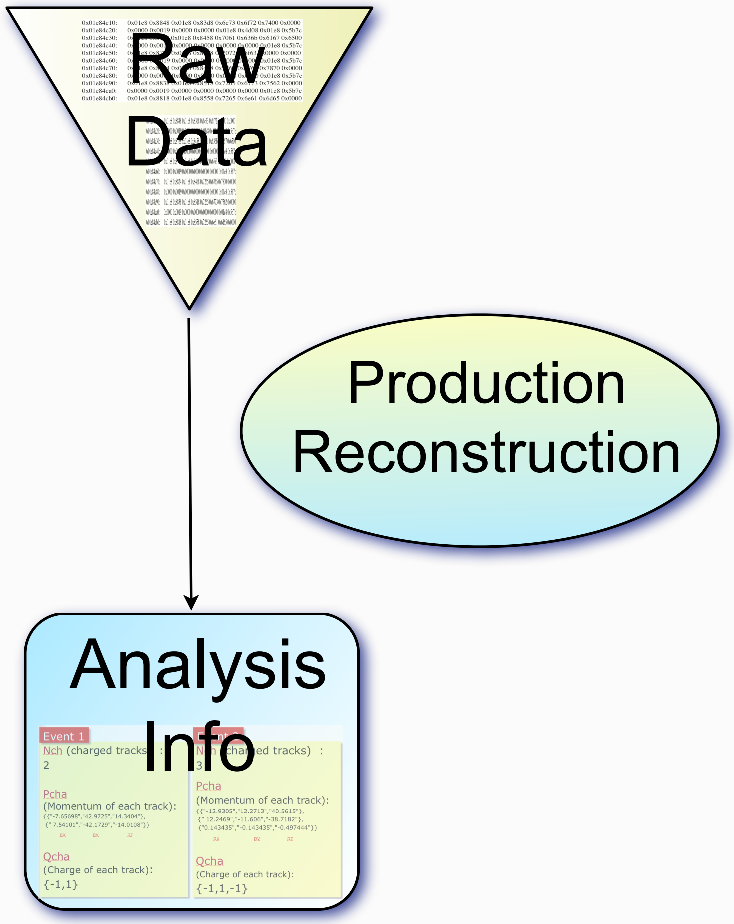
- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate "high-level" physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  - store all these quantities/objects event per event
- The data analysis
  - define the theoretically computed observable(s) to be measured
  - apply event selection (cuts)
  - estimate efficiencies and backgrounds, eg. from MC simulation
  - if distributions are measured : take care of absolute calibrations and effects because of detector resolution/smearing
    - correct for these effects
  - determine statistical and systematic uncertainties
  - compare with theory, found a deviation, something new?
    - if yes, book the ticket to Stockholm
  - determine parameters, eg. by fitting the prediction to the data

How is all this done in practice?

"The only place you'll find SUCCESS before WORK is in the dictionary" (May B. Smith)

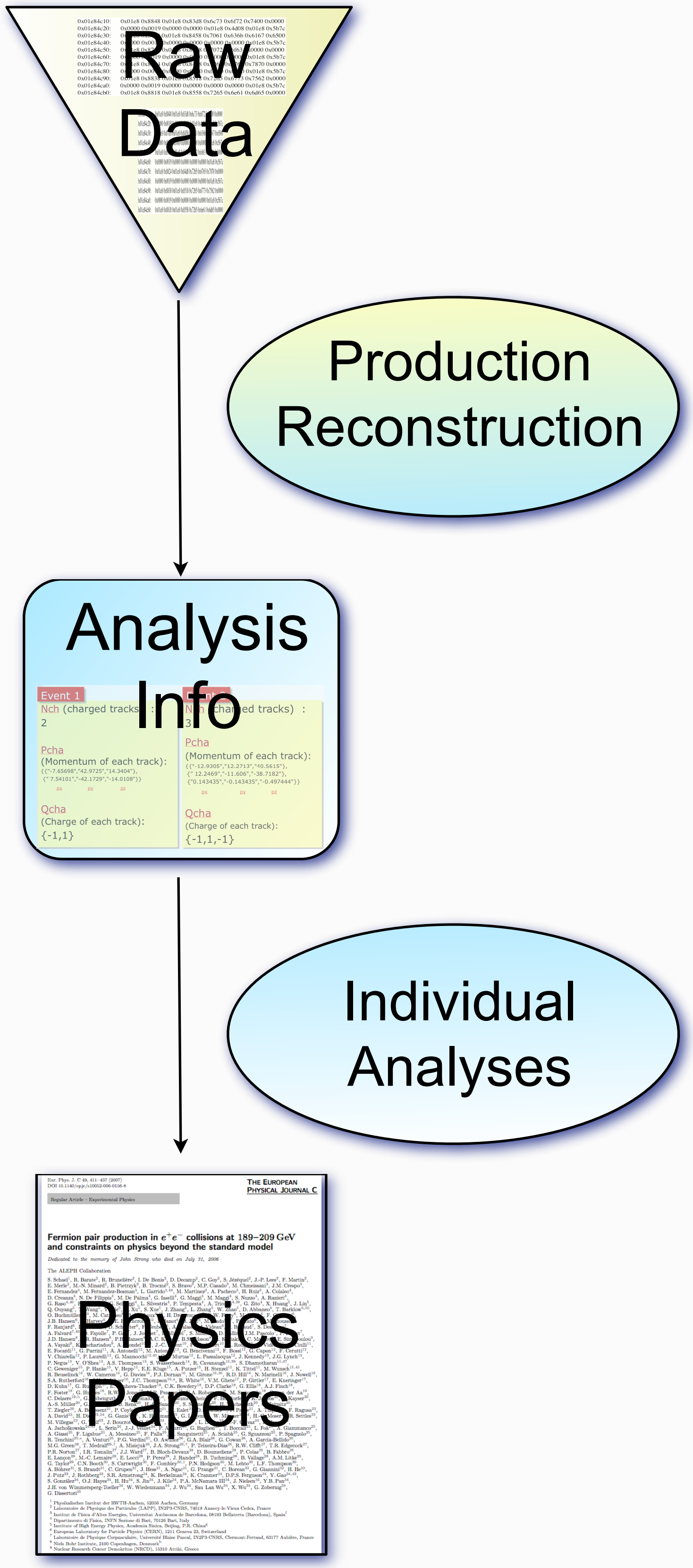
The process in practice

- The reconstruction step is usually done in common
  - “Tracks”, “particle ID”, “calorimeter towers” etc are general concepts, not analysis-specific. Common algorithms make it easier to understand how well they work
  - “very coordinated” data access



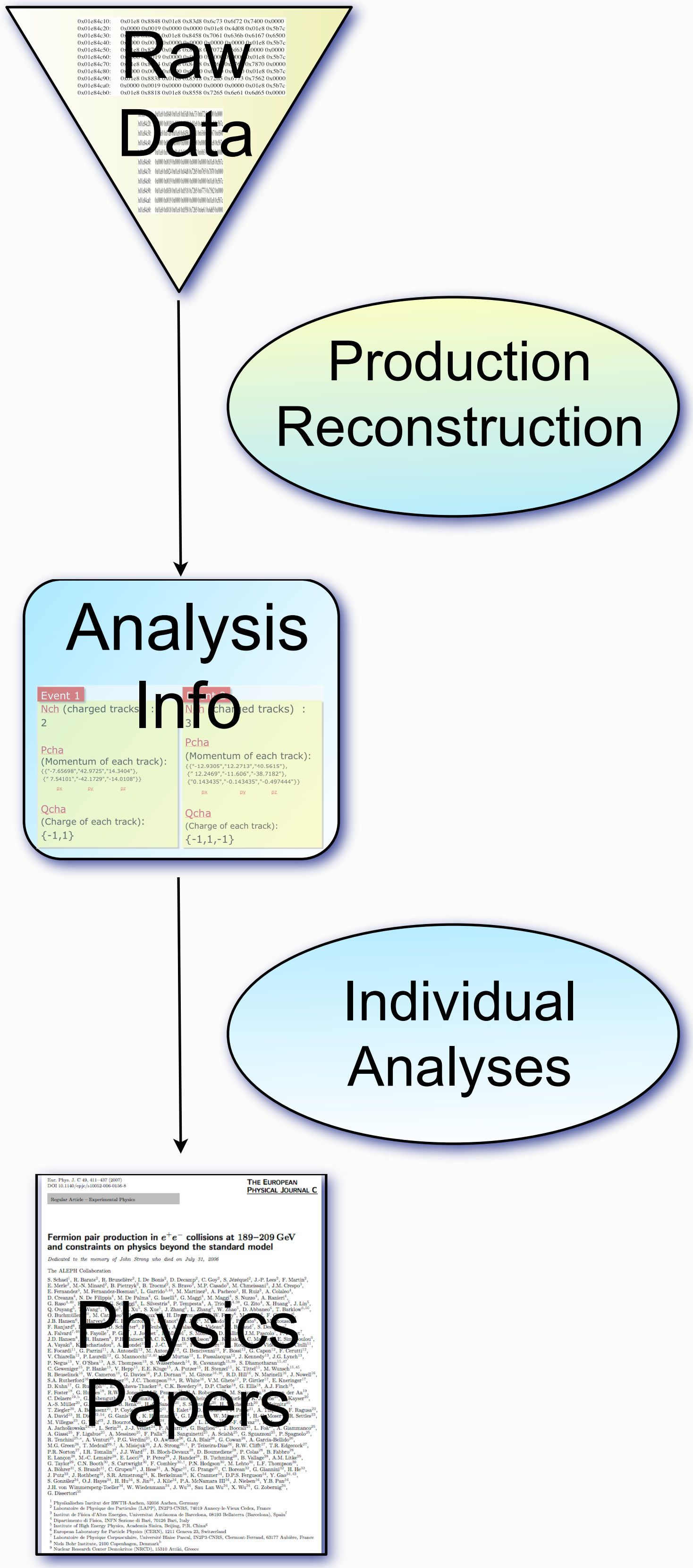
The process in practice

- The reconstruction step is usually done in common
  - "Tracks", "particle ID", "calorimeter towers" etc are general concepts, not analysis-specific. Common algorithms make it easier to understand how well they work
  - "very coordinated" data access
- Analysis is a very individual thing
  - Many different measurements being done at once
  - Small groups working on topics they are interested in
  - Many different time scales for these efforts
  - "chaotic" data access



The process in practice

- The reconstruction step is usually done in common
  - “Tracks”, “particle ID”, “calorimeter towers” etc are general concepts, not analysis-specific. Common algorithms make it easier to understand how well they work
  - “very coordinated” data access
- Analysis is a very individual thing
  - Many different measurements being done at once
  - Small groups working on topics they are interested in
  - Many different time scales for these efforts
  - “chaotic” data access
- Collaborations build offline computing systems to handle all this



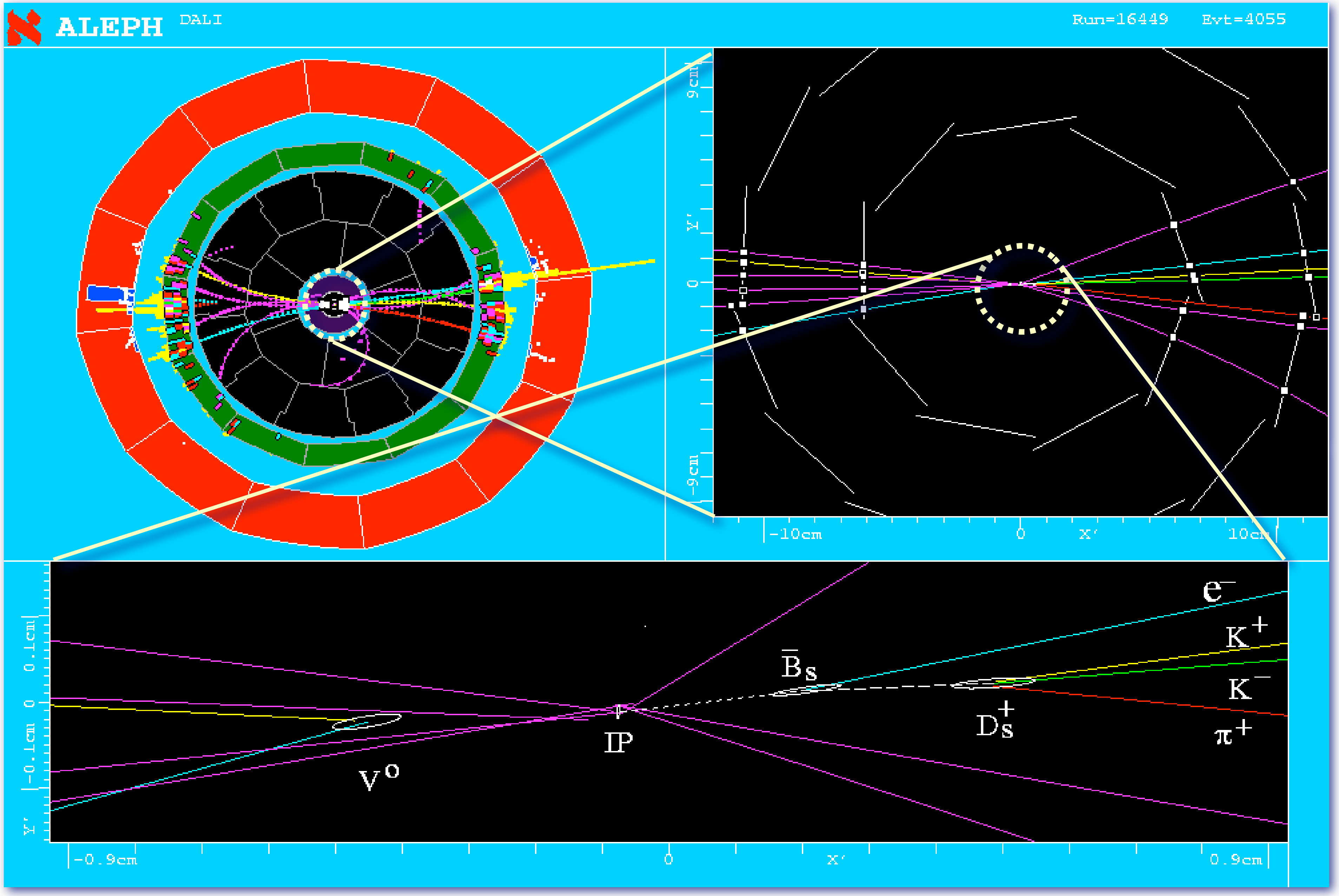
Track finding

Why does tracking need to be done well?

- Determine how many charged particles were created in an event
- Measure their momentum
  - direction, magnitude
  - combine these to look for decays of particles with known masses
  - only final stable particles are visible

Why does tracking need to be done well?

- Determine how many charged particles were created in an event
- Measure their momentum
  - direction, magnitude
  - combine these to look for decays of particles with known masses
  - only final stable particles are visible
- Measure spatial trajectories
  - combine to look for separated vertices, indicating particles with long lifetimes



Track Fitting

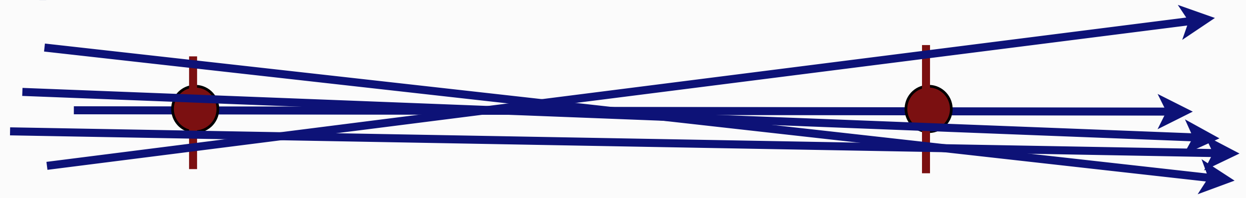
- 1D straight line fit as simple case
- Two perfect measurements
  - away from interaction point
  - no measurement uncertainty
  - just draw a straight line through them and extrapolate

- Imperfect measurements give less precise results

Track Fitting

- 1D straight line fit as simple case
- Two perfect measurements
  - away from interaction point
  - no measurement uncertainty
  - just draw a straight line through them and extrapolate

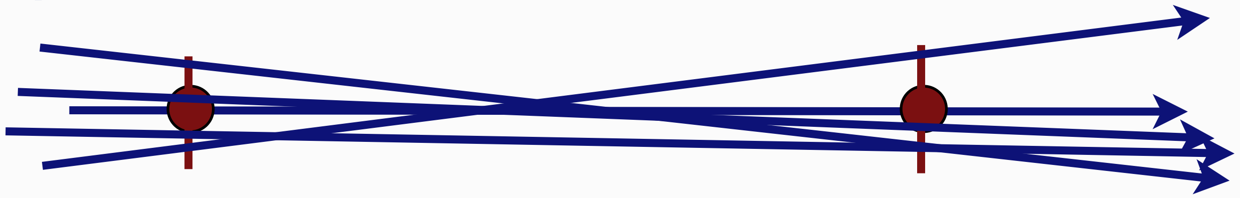
- Imperfect measurements give less precise results
  - the farther you extrapolate, the less you know



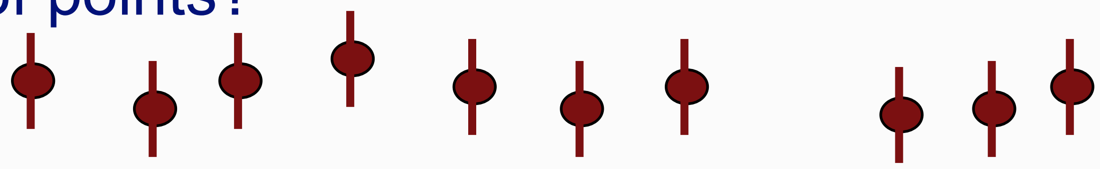
Track Fitting

- 1D straight line fit as simple case
- Two perfect measurements
  - away from interaction point
  - no measurement uncertainty
  - just draw a straight line through them and extrapolate

- Imperfect measurements give less precise results
  - the farther you extrapolate, the less you know



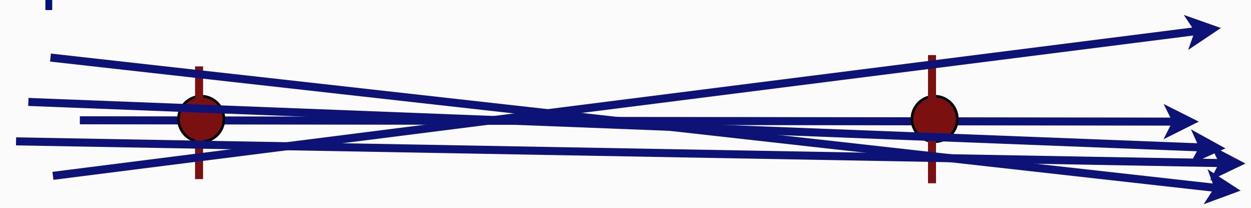
- Smaller errors and more points help to constrain the possibilities.
  But how to find the best point from a large set of points?



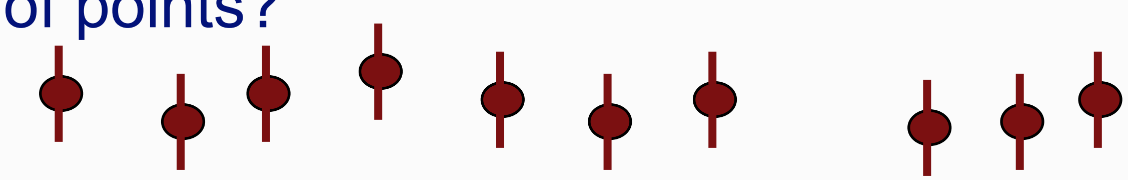
Track Fitting

- 1D straight line fit as simple case
- Two perfect measurements
  - away from interaction point
  - no measurement uncertainty
  - just draw a straight line through them and extrapolate

- Imperfect measurements give less precise results
  - the farther you extrapolate, the less you know



- Smaller errors and more points help to constrain the possibilities.
  But how to find the best point from a large set of points?

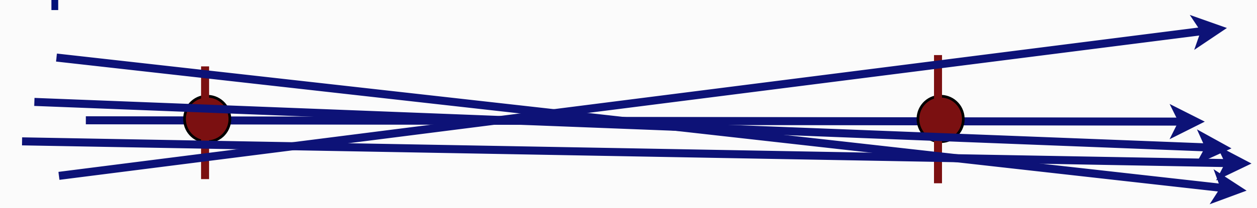


- Quantitatively
  - parameterize a track:
  - In case of straight line $y(x) = \theta\, x + d$ or, eg., helix in case of magnetic field present

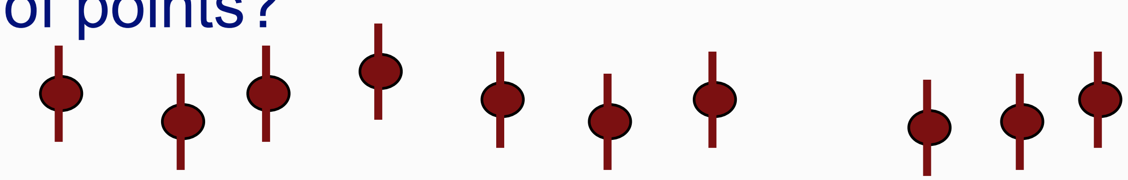
Track Fitting

- 1D straight line fit as simple case
- Two perfect measurements
  - away from interaction point
  - no measurement uncertainty
  - just draw a straight line through them and extrapolate

- Imperfect measurements give less precise results
  - the farther you extrapolate, the less you know



- Smaller errors and more points help to constrain the possibilities. But how to find the best point from a large set of points?



- Quantitatively
  - parameterize a track:
  - In case of straight line $\boxed{y(x) = \theta\, x + d}$ or, eg., helix in case of magnetic field present

predicted track position at i^{th} hit

position of i^{th} hit

  - Find track parameters by Least-Squares-Minimization

$$\boxed{\chi^2 = \sum_{i=1}^{n_{\text{hits}}} \frac{(y_i - y(x_i))^2}{\sigma_i^2}}$$

uncertainty of i^{th} measurement

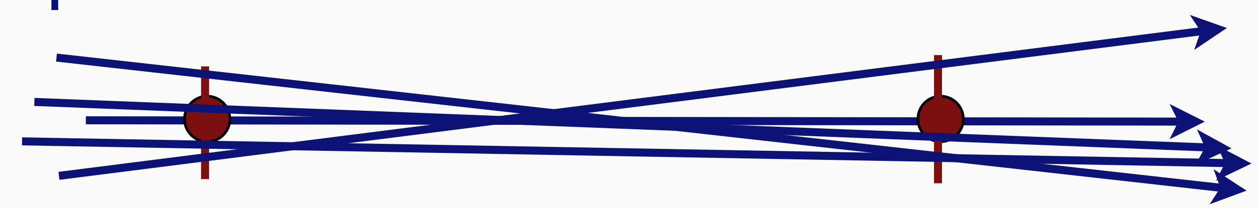
see also lecture by G. Cowan

Track Fitting

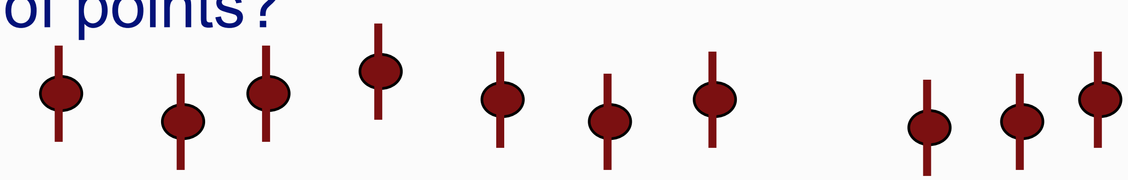
ETH Institute for Particle Physics

- 1D straight line fit as simple case
- Two perfect measurements
  - away from interaction point
  - no measurement uncertainty
  - just draw a straight line through them and extrapolate

- Imperfect measurements give less precise results
  - the farther you extrapolate, the less you know



- Smaller errors and more points help to constrain the possibilities. But how to find the best point from a large set of points?



- Quantitatively
  - parameterize a track:
  - In case of straight line $\boxed{y(x) = \theta\, x + d}$ or, eg., helix in case of magnetic field present
  - Find track parameters by Least-Squares-Minimization
  - Obtain also uncertainties on track parameters

$\delta\theta \qquad \delta d$

position of i^{th} hit

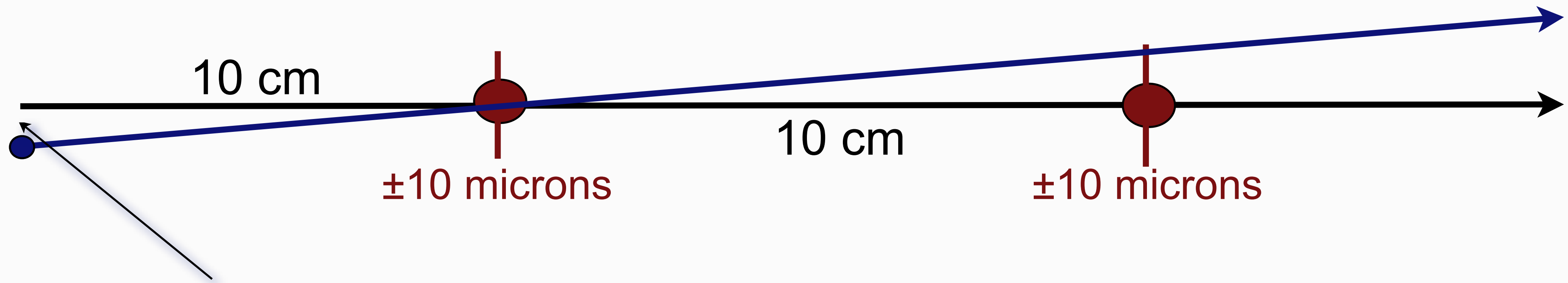
predicted track position at i^{th} hit

$$\boxed{\chi^2 = \sum_{i=1}^{n_{\text{hits}}} \frac{(y_i - y(x_i))^2}{\sigma_i^2}}$$

uncertainty of i^{th} measurement

see also lecture by G. Cowan

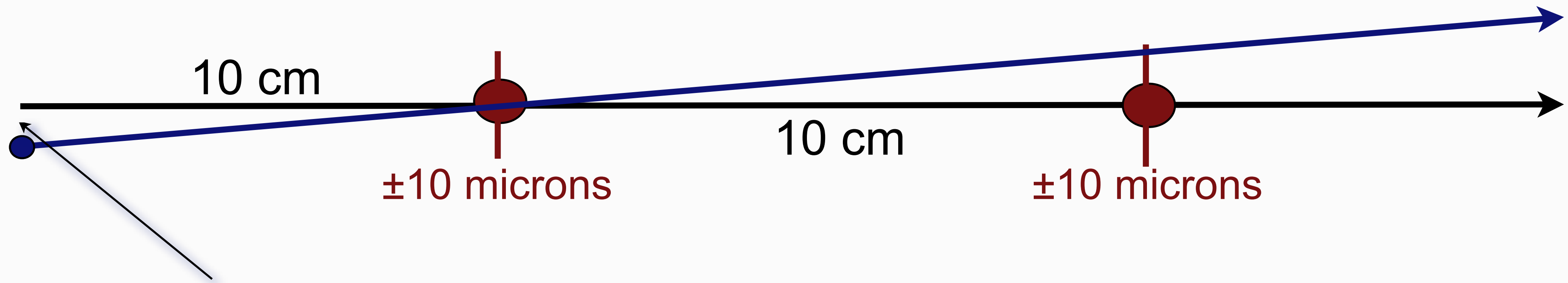
“Typical” size of errors



- Error δd on position is about ±10 microns
- Error $\delta\Theta$ on angle is about ±0.1 milliradians (±0.002 degrees)
- Satisfyingly small errors
  - allows separation of tracks that come from different particle decays (which can be separated at the order of mm)

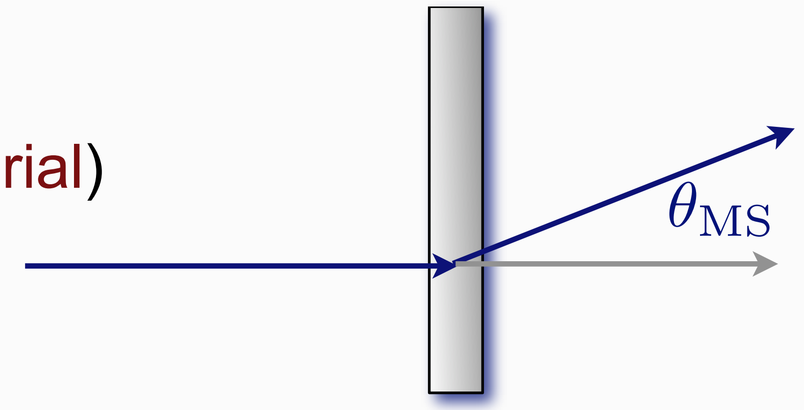
- However
  - we “see” particles by interaction with a detector (=material)

“Typical” size of errors

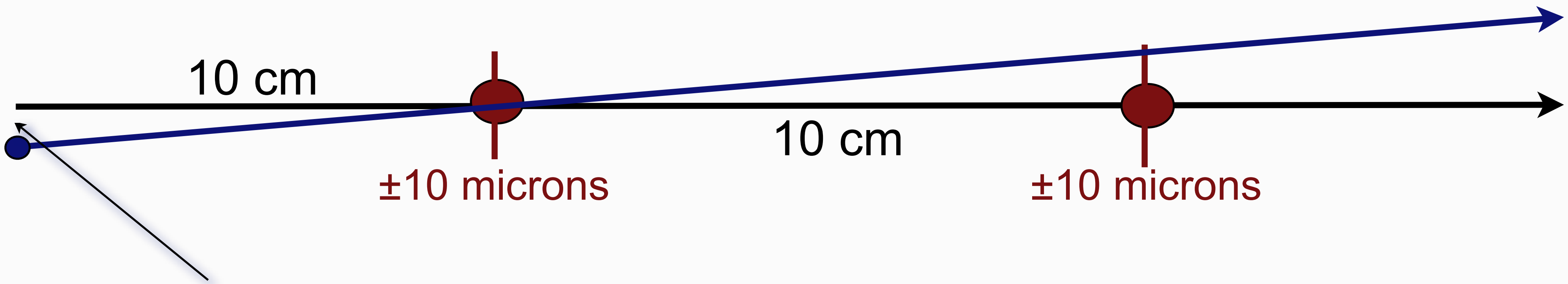


- Error δd on position is about ±10 microns
- Error $\delta\Theta$ on angle is about ±0.1 milliradians (±0.002 degrees)
- Satisfyingly small errors
  - allows separation of tracks that come from different particle decays (which can be separated at the order of mm)

- However
  - we “see” particles by interaction with a detector (=material)
  - interaction leads to : energy loss, change in direction

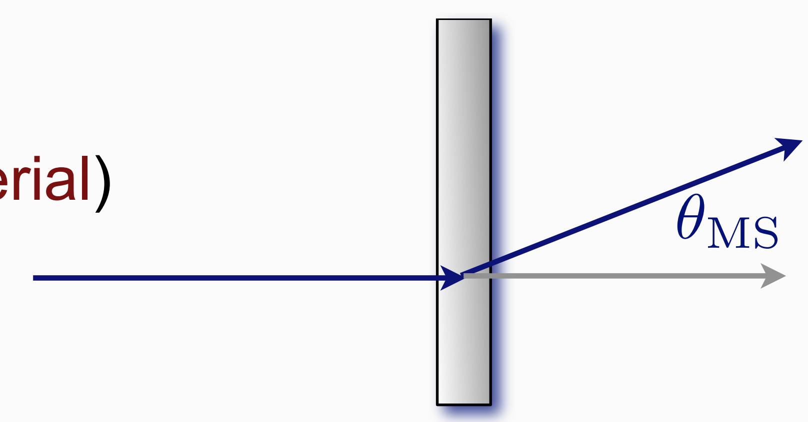


“Typical” size of errors



- Error δd on position is about ±10 microns
- Error δΘ on angle is about ±0.1 milliradians (±0.002 degrees)
- Satisfyingly small errors
  - allows separation of tracks that come from different particle decays (which can be separated at the order of mm)

- However
  - we “see” particles by interaction with a detector (=material)
  - interaction leads to : energy loss, change in direction
  - This is **Multiple Scattering**
    - Charged particles passing through matter “scatter” by a random angle



$$\sqrt{\langle\theta^2_{\mathrm{MS}}\rangle} = \frac{15\,\mathrm{MeV}/c}{\beta p}\sqrt{\frac{\mathrm{thickness}}{X_{\mathrm{rad}}}}$$

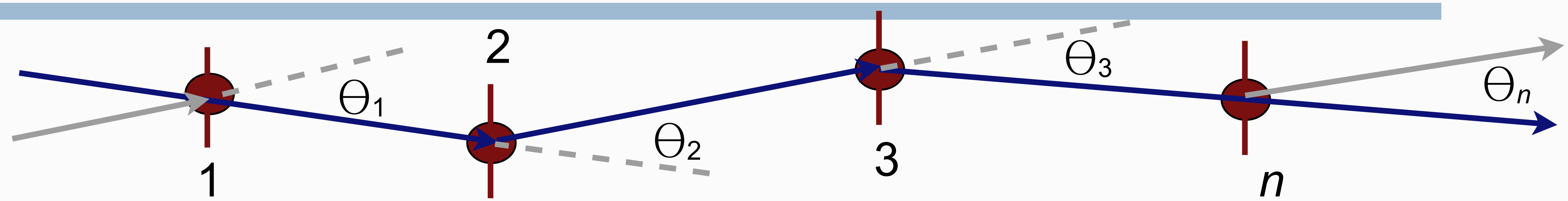
examples:

300 micron Si : RMS = 0.9 mrad / βp

1 mm Be : RMS = 0.8 mrad / βp

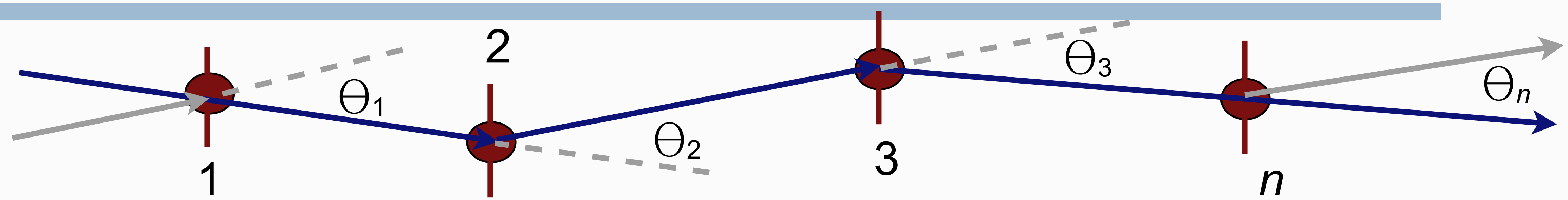
➔ leads to additional position errors

Kalman filter



- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)
  - For large *n*, computing time grows like O(n^3), quickly un-practicable

Kalman filter



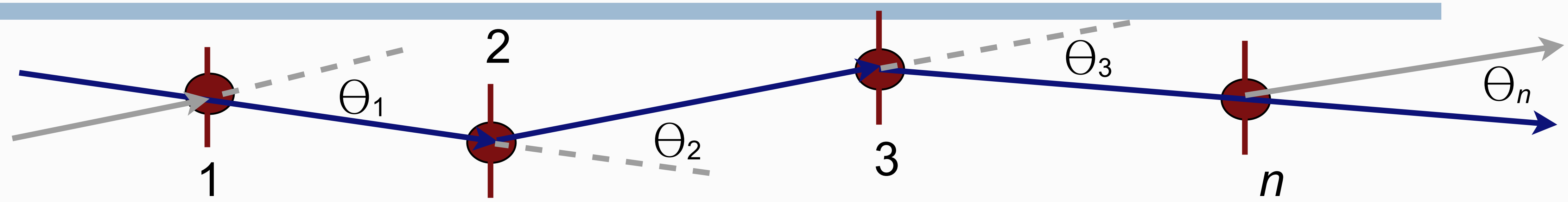
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$

- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

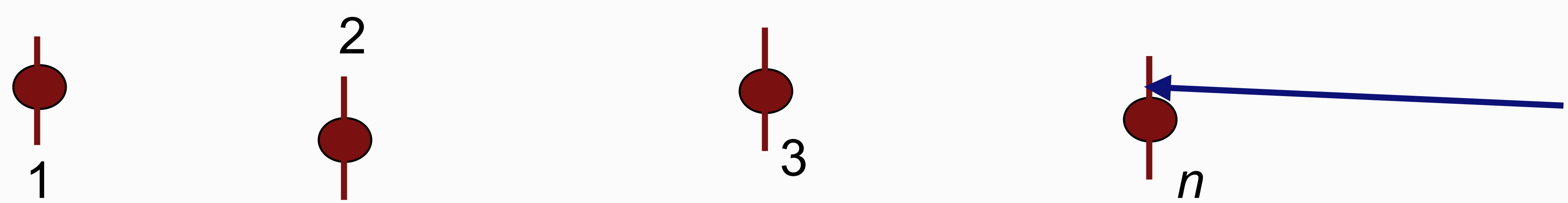
$$\chi^2 = \chi^2_{\mathrm{old}} + \sum_i \frac{\theta_i^2}{\sigma^2_{\mathrm{MS}}}$$

  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin

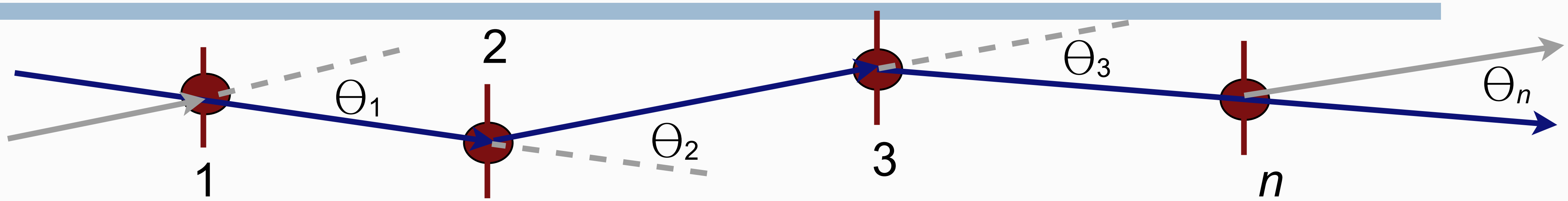
Kalman filter



- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns) $\chi^2 = \chi^2_{\text{old}} + \sum_i \frac{\theta_i^2}{\sigma^2_{\text{MS}}}$
  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin
- Instead, approximate, work inward *N* times



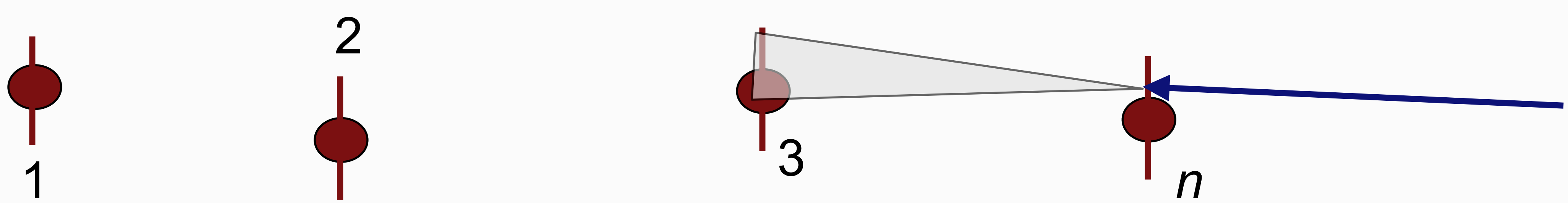
Kalman filter



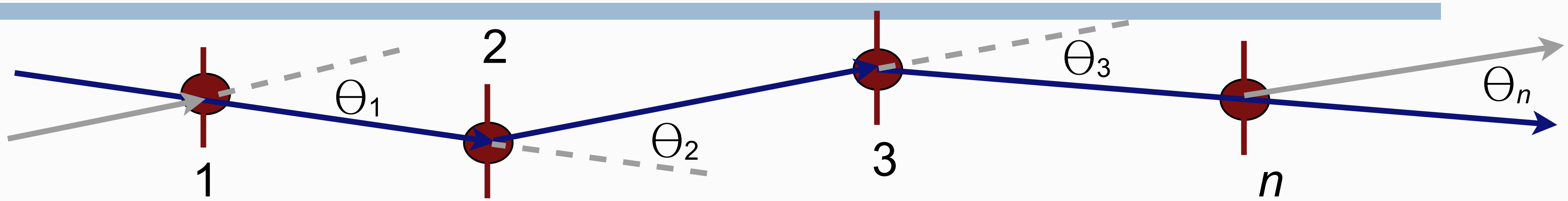
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)
  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin

$$\chi^2 = \chi^2_{\mathrm{old}} + \sum_i \frac{\theta_i^2}{\sigma^2_{\mathrm{MS}}}$$

- Instead, approximate, work inward *N* times



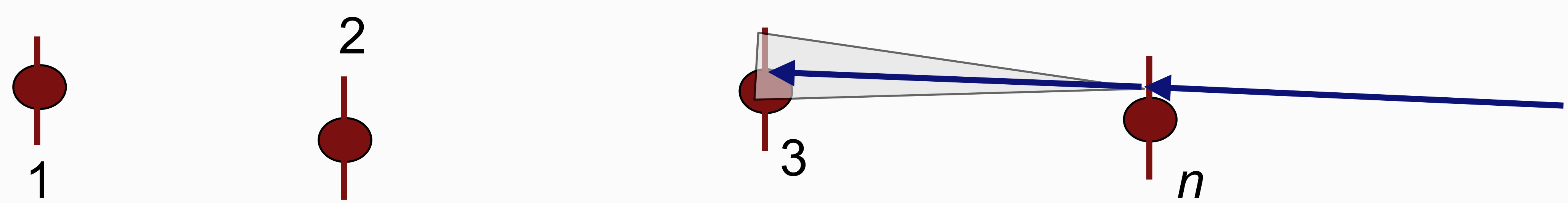
Kalman filter



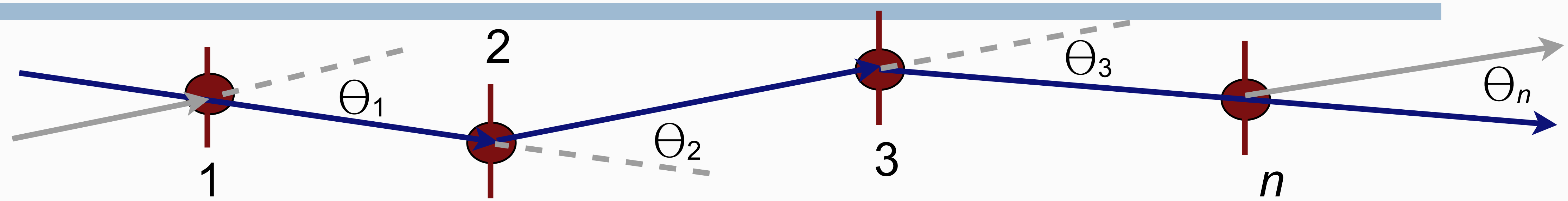
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

$$\chi^2 = \chi^2_{\rm old} + \sum_i \frac{\theta_i^2}{\sigma^2_{\rm MS}}$$

  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin
- Instead, approximate, work inward *N* times



Kalman filter

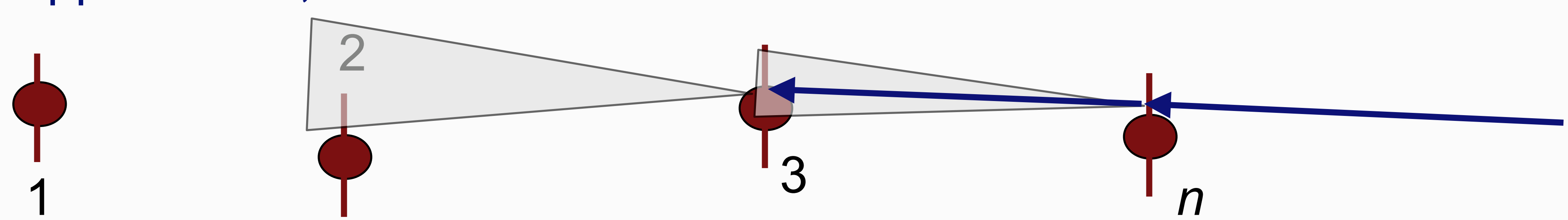


- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$

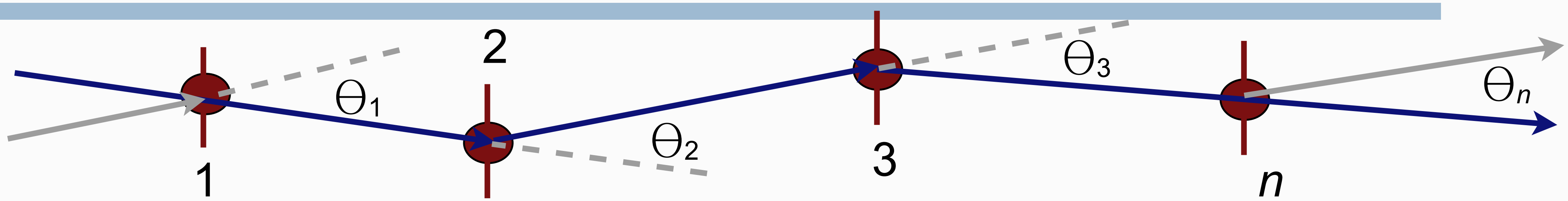
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)
  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin

$$\chi^2 = \chi^2_{\mathrm{old}} + \sum_i \frac{\theta_i^2}{\sigma^2_{\mathrm{MS}}}$$

- Instead, approximate, work inward *N* times



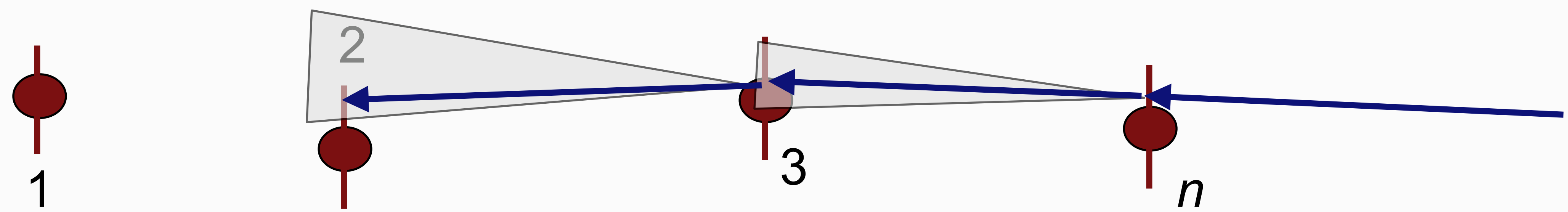
Kalman filter



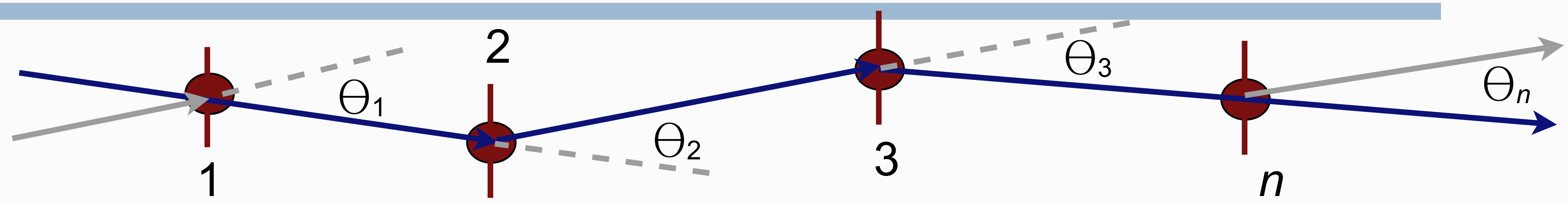
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

$$\chi^2 = \chi^2_{\rm old} + \sum_i \frac{\theta_i^2}{\sigma^2_{\rm MS}}$$

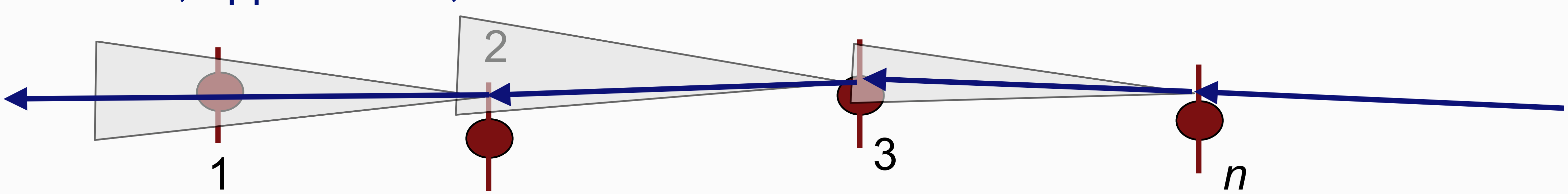
  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin
- Instead, approximate, work inward *N* times



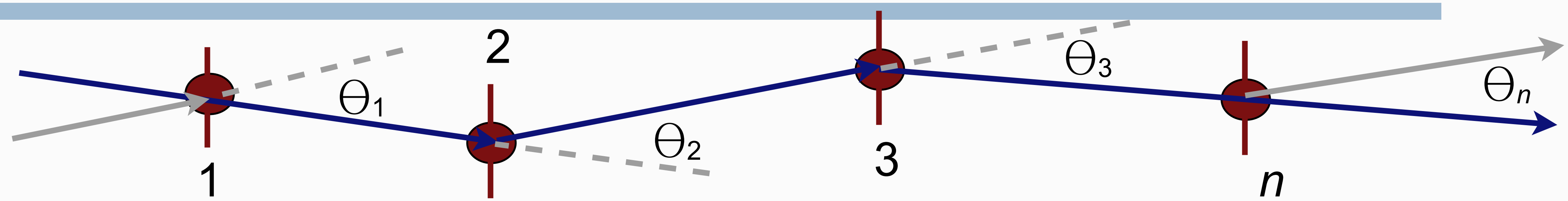
Kalman filter



- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)
  $$\chi^2 = \chi^2_{\text{old}} + \sum_i \frac{\theta_i^2}{\sigma^2_{\text{MS}}}$$
  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin
- Instead, approximate, work inward *N* times



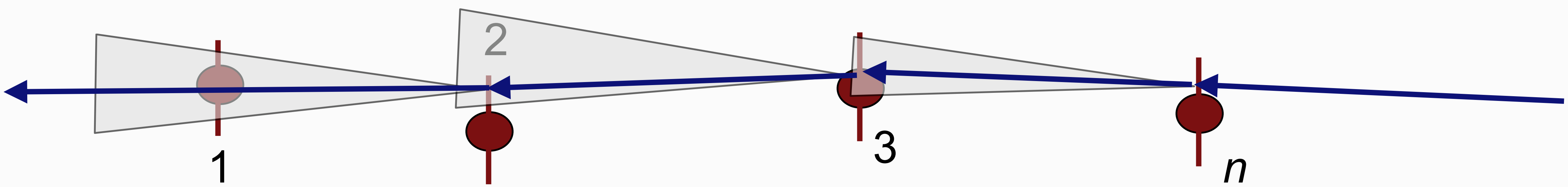
Kalman filter



- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

$$\chi^2 = \chi^2_{\mathrm{old}} + \sum_i \frac{\theta_i^2}{\sigma^2_{\mathrm{MS}}}$$

  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin
- Instead, approximate, work inward *N* times

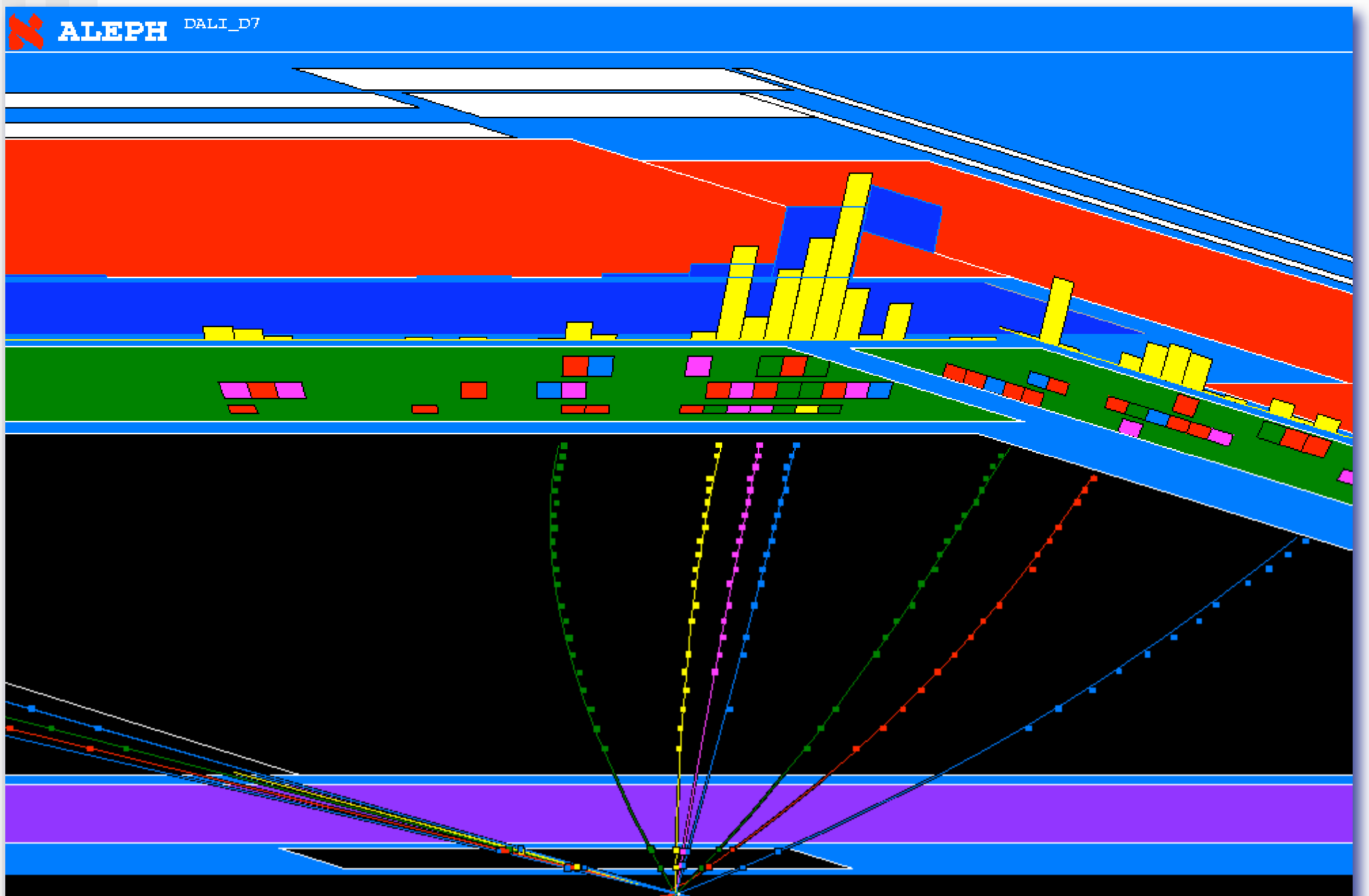


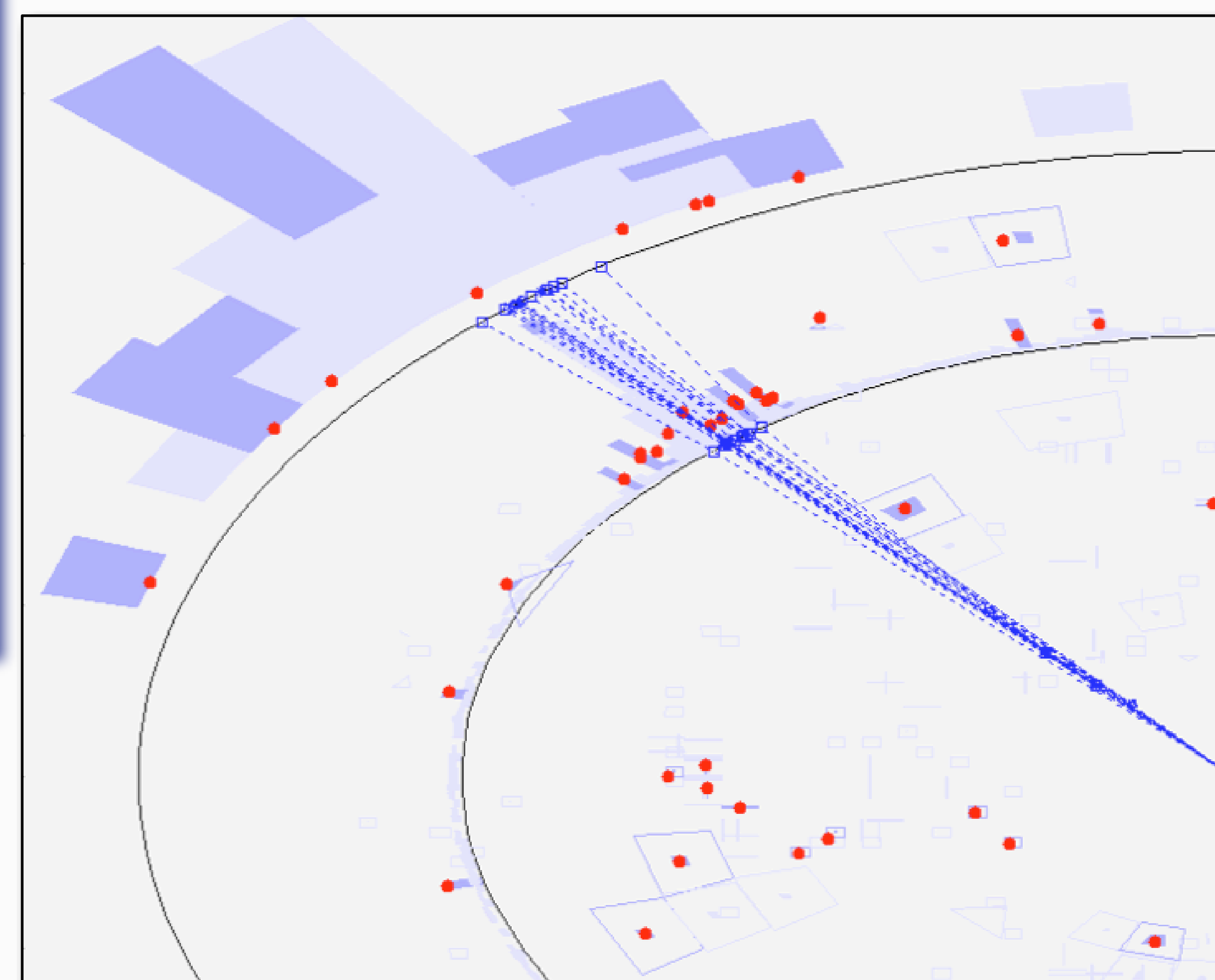
  - leads to O(*n*) computations!
  - in each step, make extrapolation to next layer, using information from current track parameters, expected scattering error, and measurement at next layer
  - Needs a starting estimate (seed) and may need some iterations, smoothing
  - This method is based on theory of the **Kalman Filter**

Calorimeter energy reconstruction

Goals

- Reconstruct energy deposited by charged and neutral particles
- Determine position of deposit, direction of incident particles
- Be insensitive to noise and “un-wanted” (un-correlated) energy





- and obtain the best possible resolution!

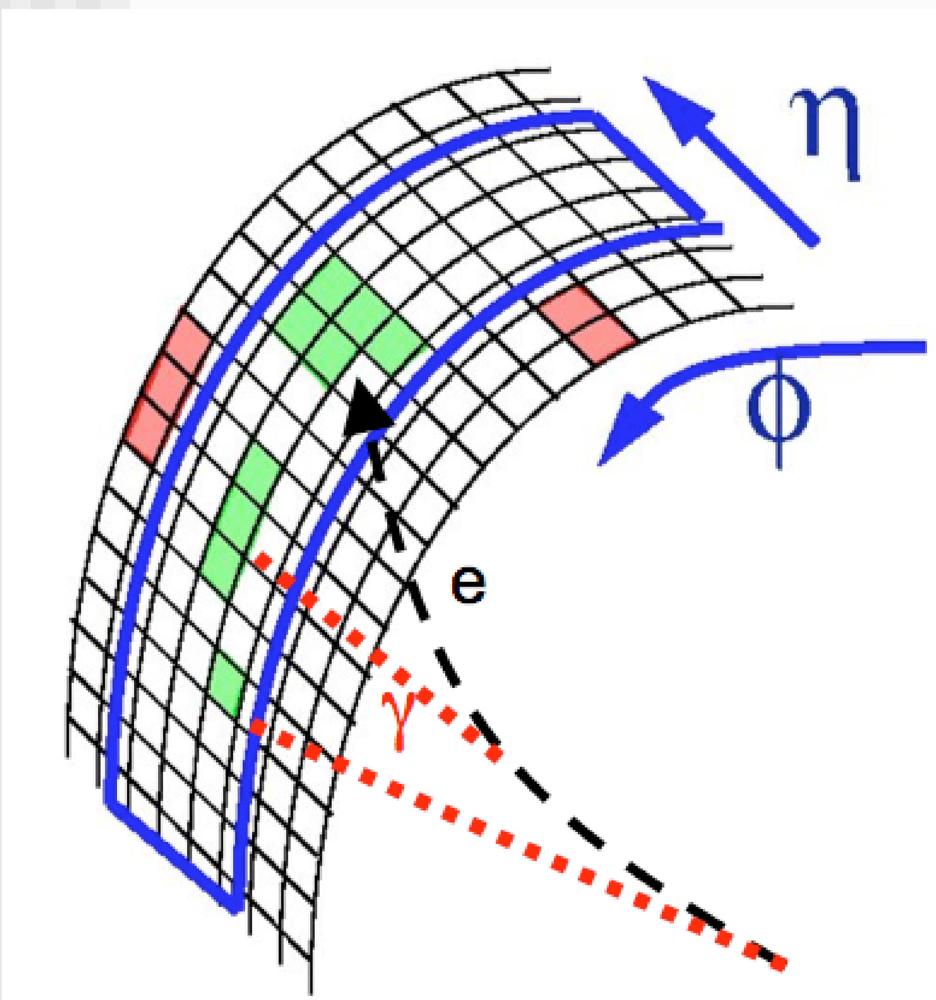
Clusters of energy

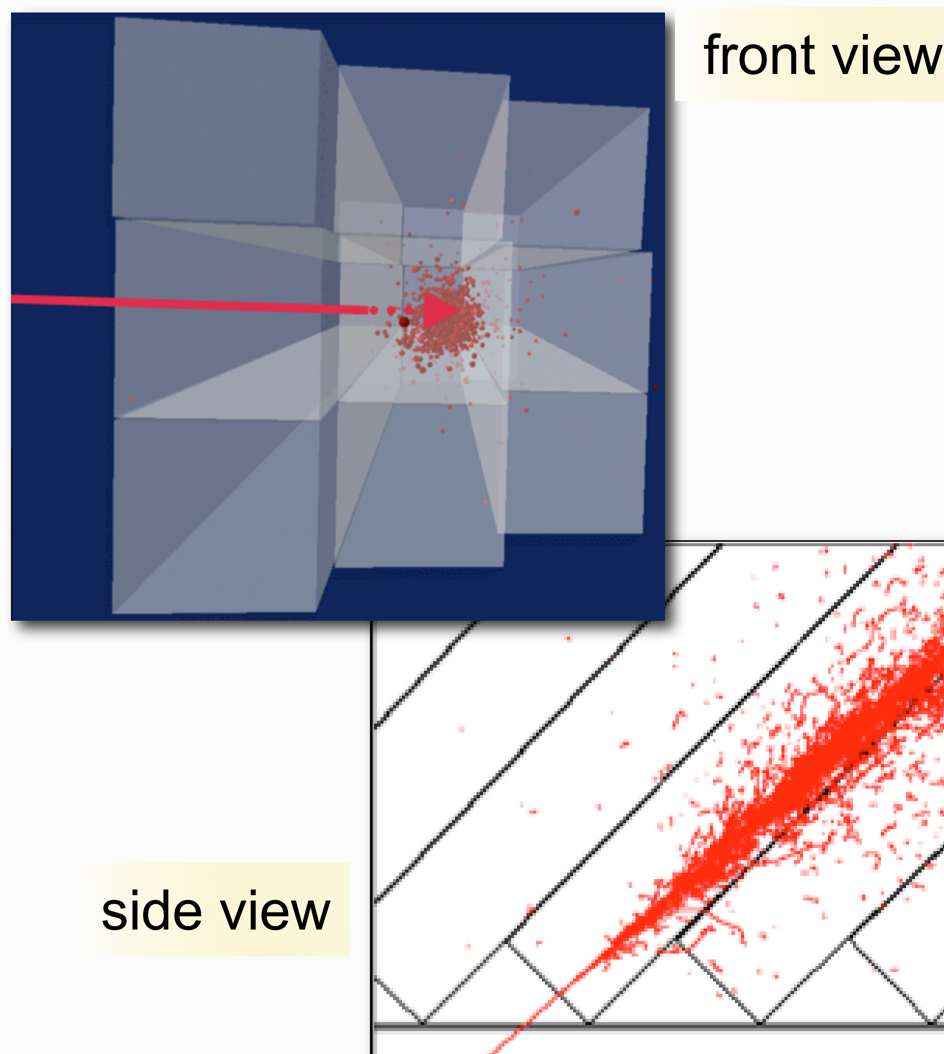
- Calorimeters are segmented in cells
- Typically a shower extends over several cells
  - Useful to reconstruct precisely the impact point from the “center-of-gravity” of the deposits in the various cells

Clusters of energy

ETH Institute for Particle Physics

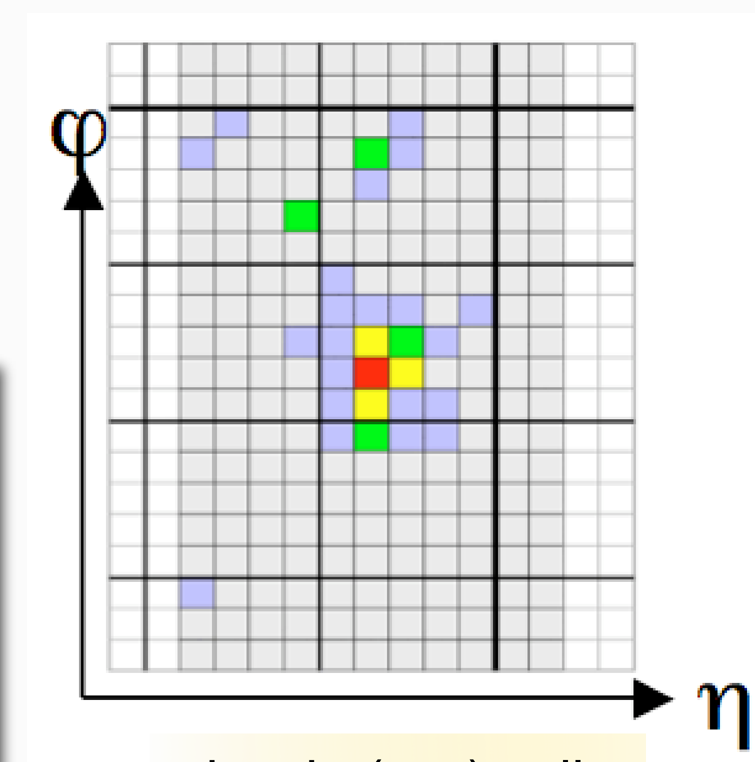
- Calorimeters are segmented in cells
- Typically a shower extends over several cells
  - Useful to reconstruct precisely the impact point from the “center-of-gravity” of the deposits in the various cells
- Example CMS Crystal Calorimeter:
  - electron energy in central crystal ~ 80 %, in 5x5 matrix around it ~ 96 %





front view

side view

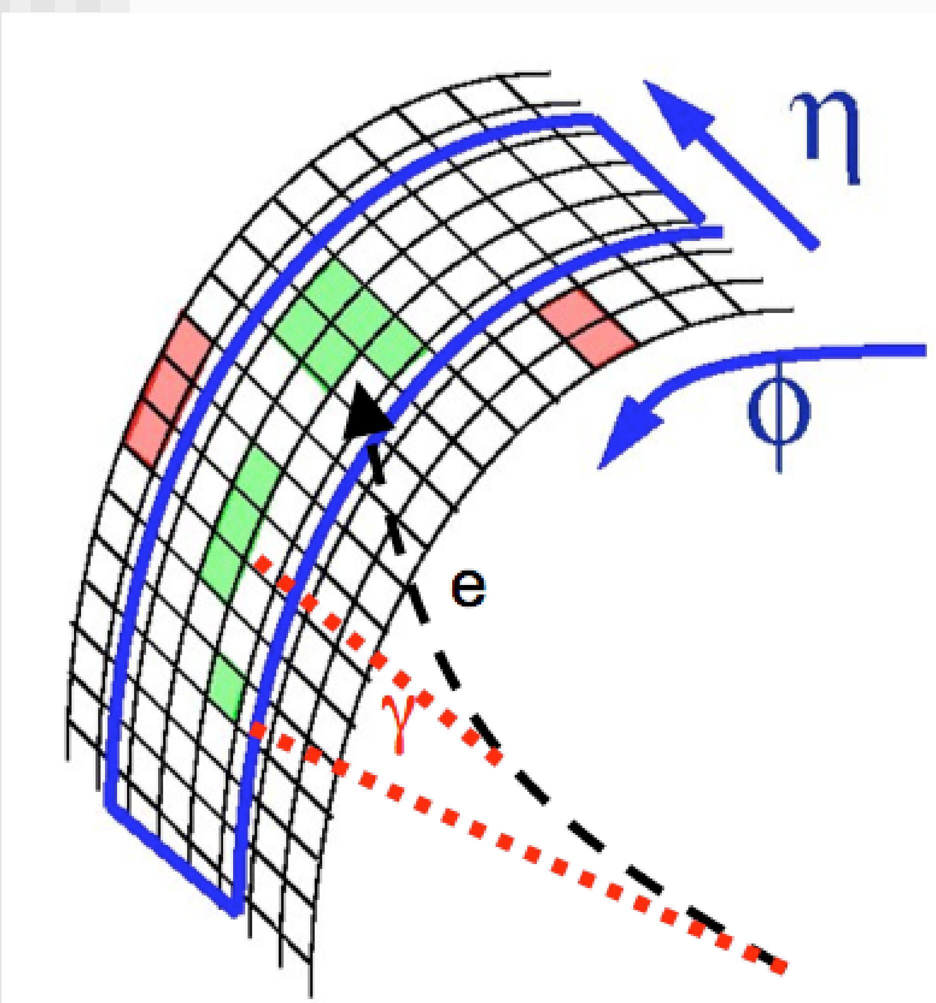


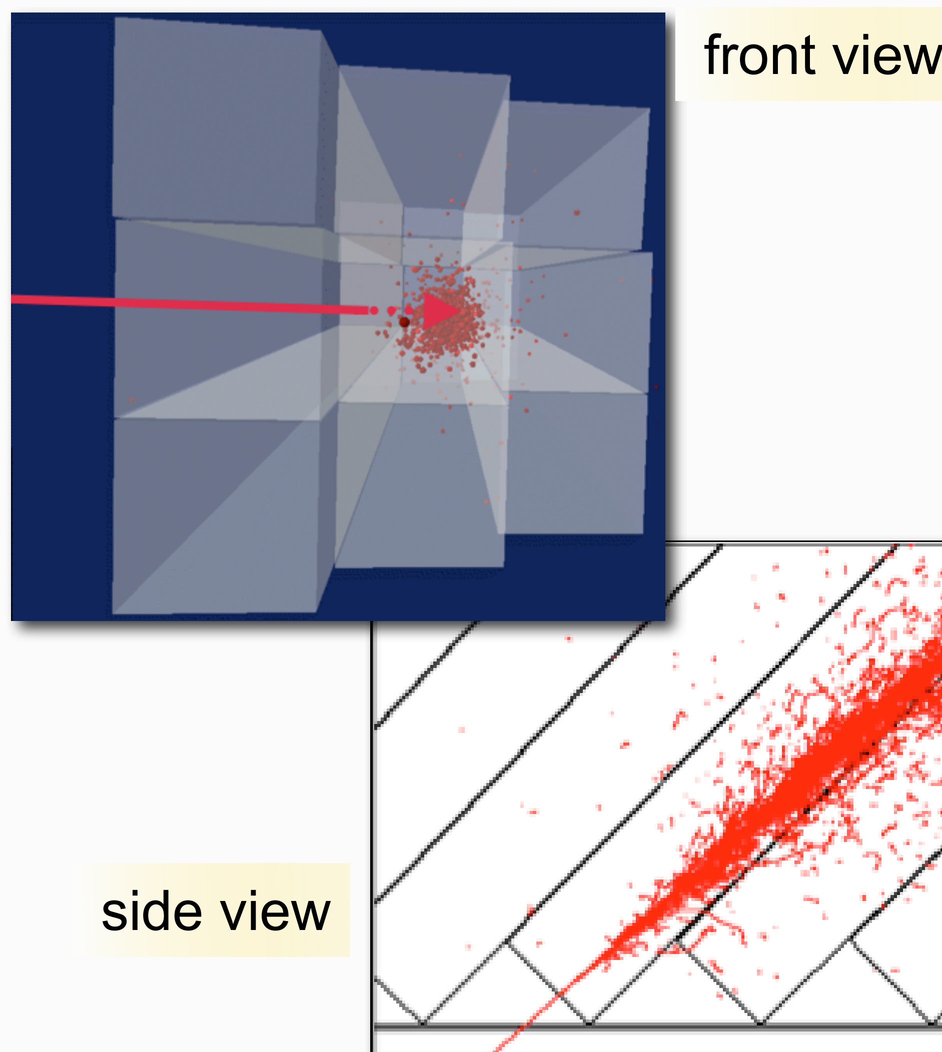
view in (φ,η) cells

Clusters of energy

ETH Institute for Particle Physics

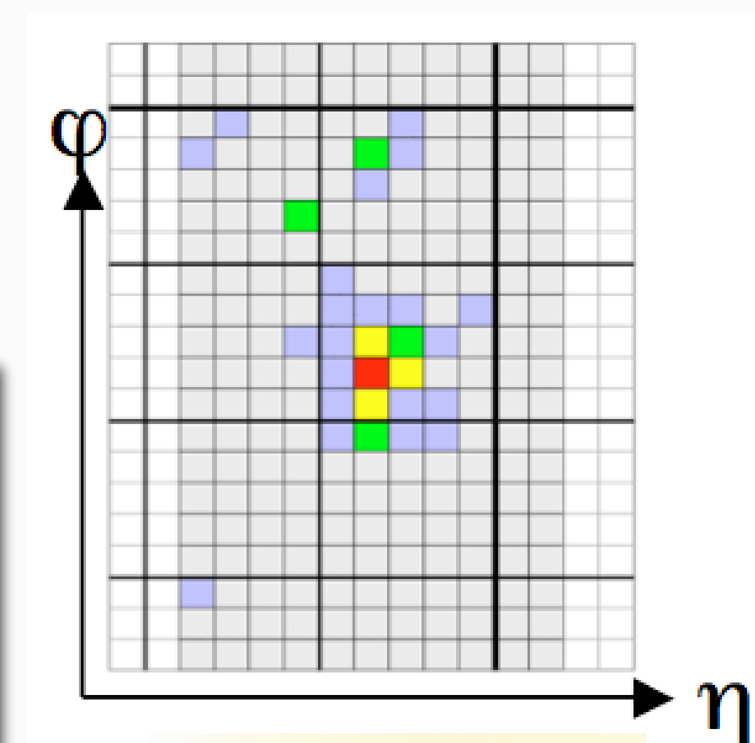
- Calorimeters are segmented in cells
- Typically a shower extends over several cells
  - Useful to reconstruct precisely the impact point from the “center-of-gravity” of the deposits in the various cells
- Example CMS Crystal Calorimeter:
  - electron energy in central crystal ~ 80 %, in 5x5 matrix around it ~ 96 %
- So task is : identify these clusters and reconstruct the energy they contain





front view

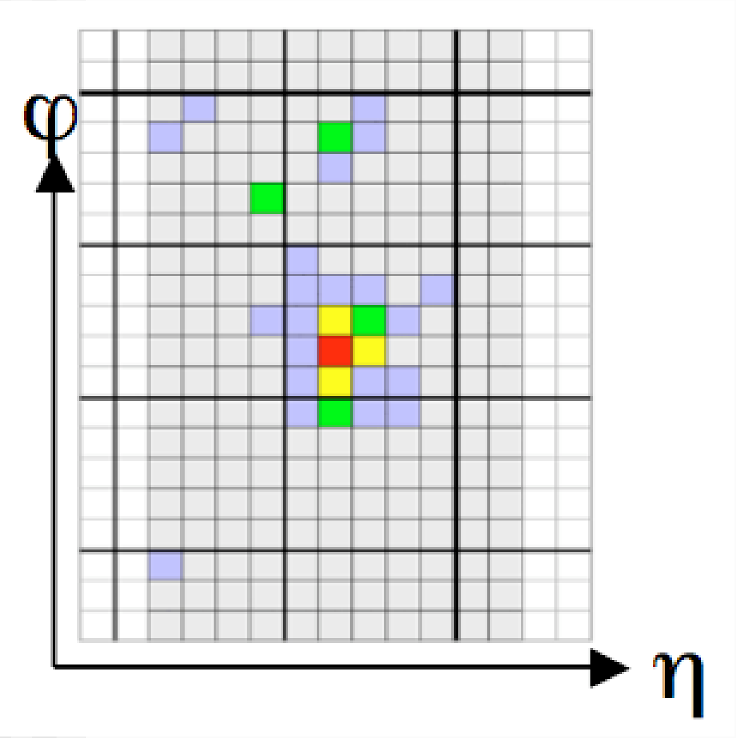
side view



view in (φ,η) cells

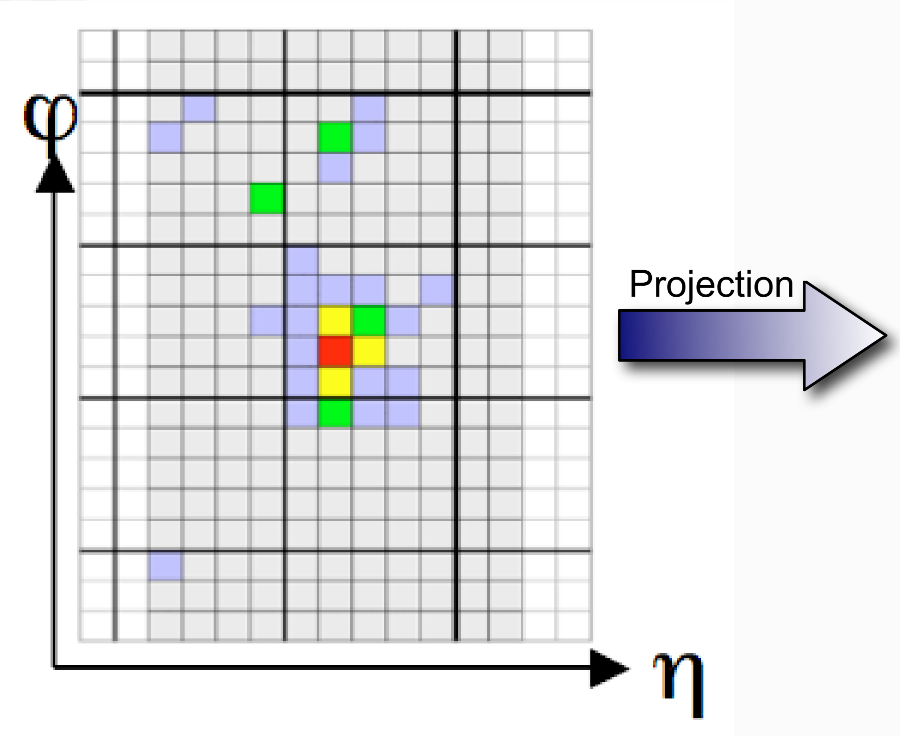
Cluster Finding

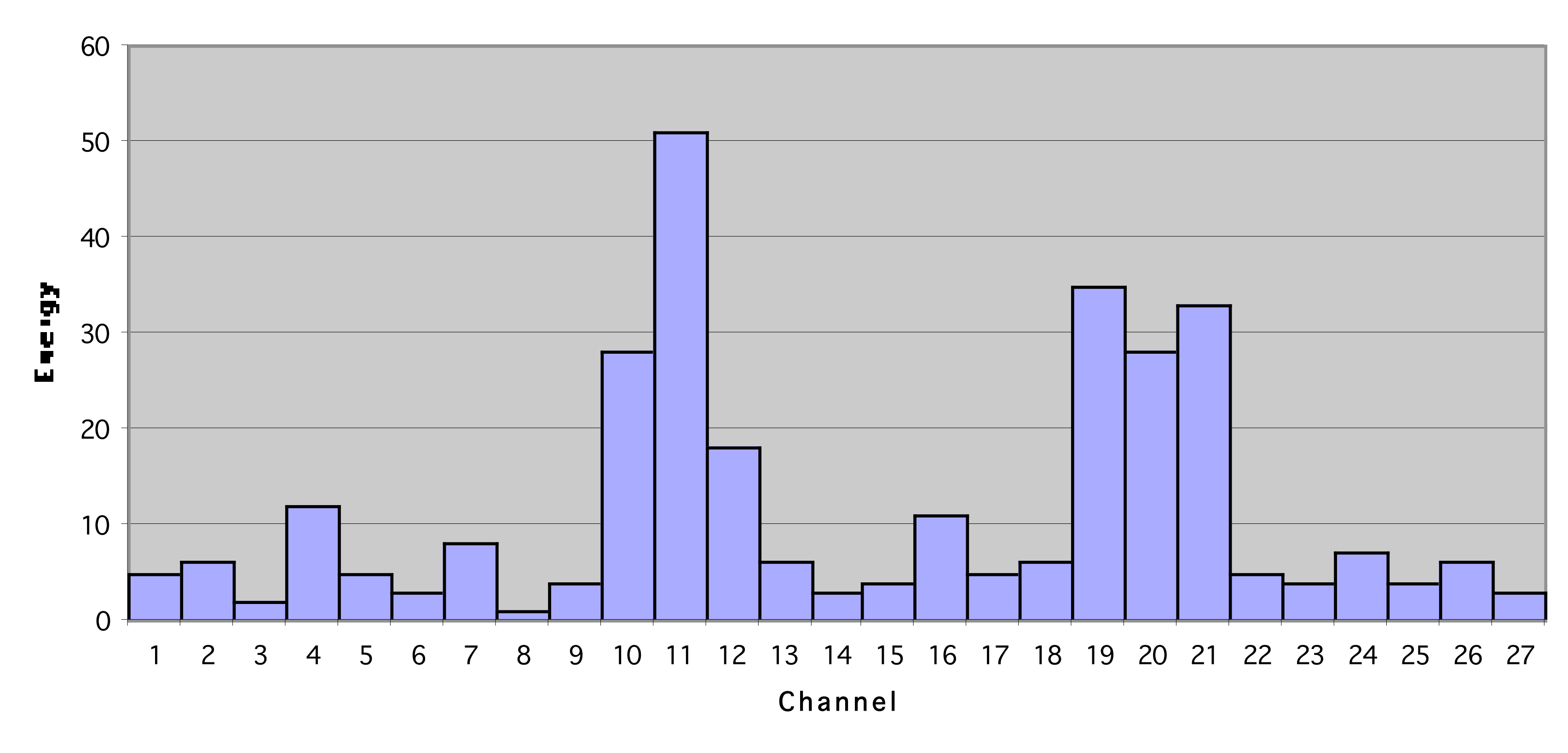
- Clusters of energy in a calorimeter are due to the original particles
  - Clustering algorithm groups individual channel energies
  - Don’t want to miss any; don’t want to pick up fakes



Cluster Finding

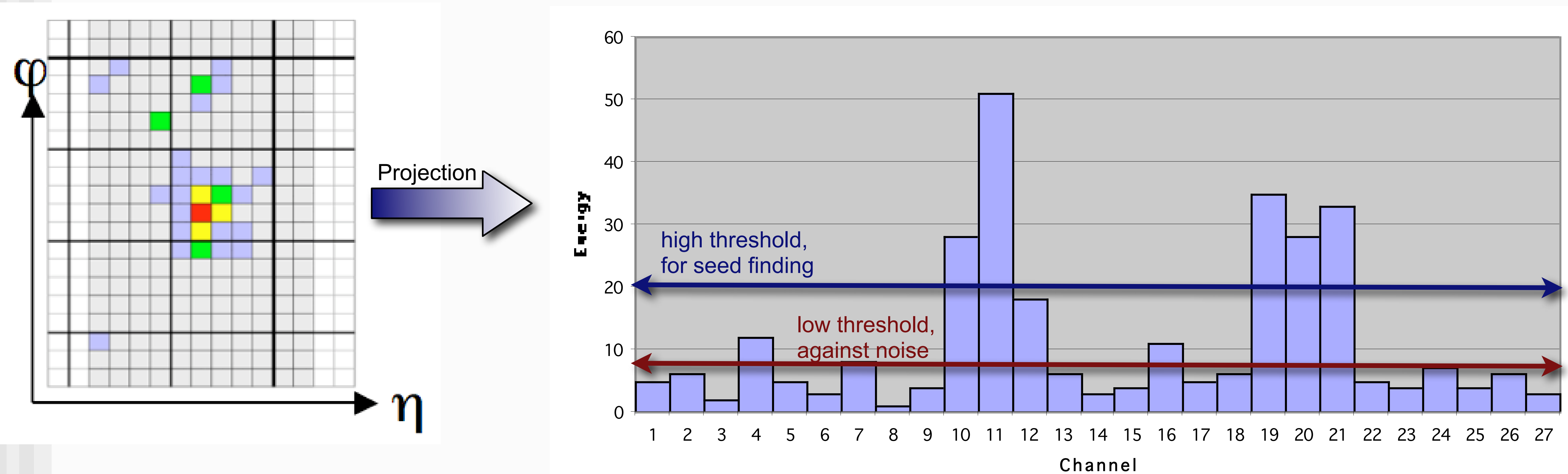
- Clusters of energy in a calorimeter are due to the original particles
  - Clustering algorithm groups individual channel energies
  - Don't want to miss any; don't want to pick up fakes





Cluster Finding

- Clusters of energy in a calorimeter are due to the original particles
  - Clustering algorithm groups individual channel energies
  - Don't want to miss any; don't want to pick up fakes



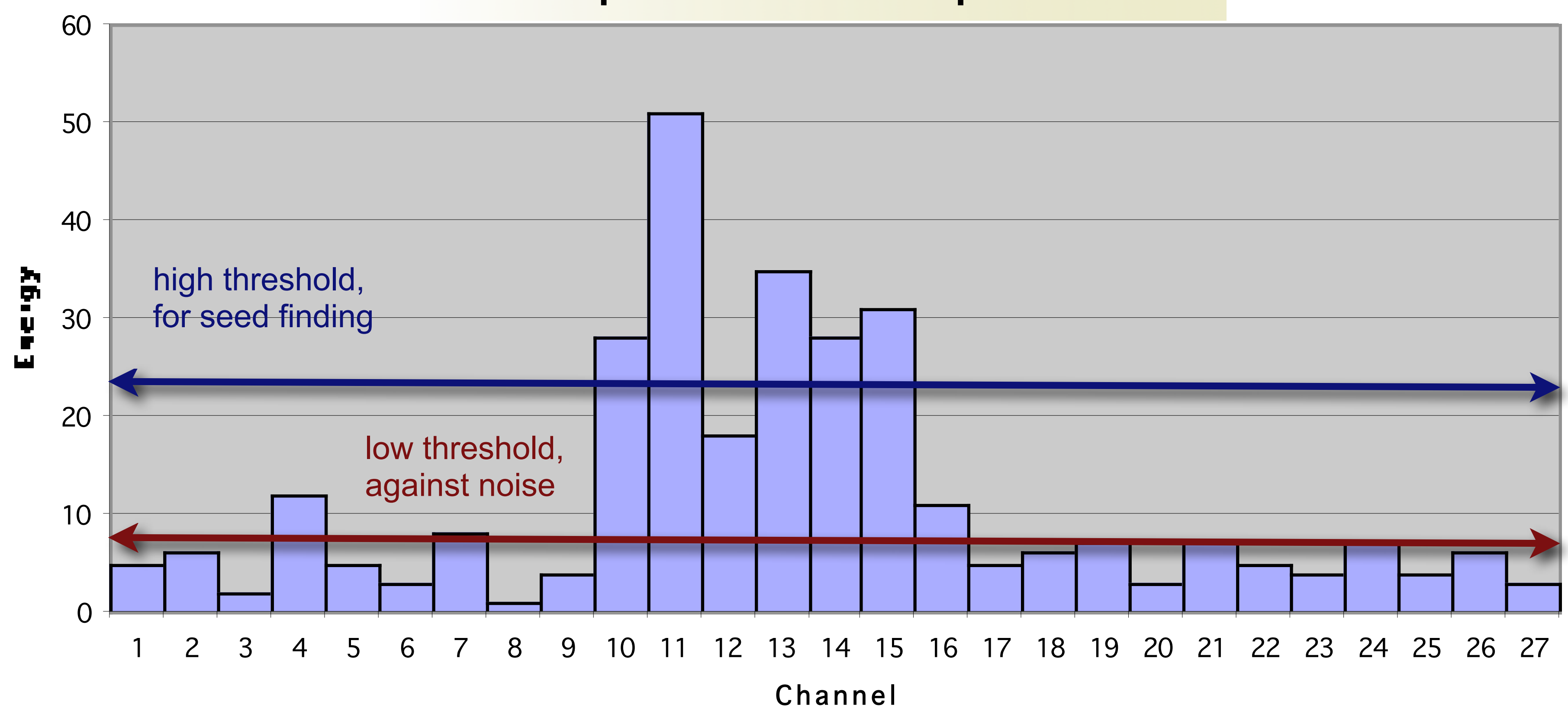
Simple example of an algorithm

- Scan for seed crystals = local energy maximum above a defined seed threshold
- Starting from the seed position, adjacent crystals are examined, scanning first in φ and then in η
- Along each scan line, crystals are **added to the cluster if**
  1. The crystal's energy is above the noise level (lower threshold)
  2. The crystal has not been assigned to another cluster already
  3. The previous crystal added (in the same direction) has higher energy

Difficulties

- Careful tuning of thresholds needed
  - needs usually learning phase
  - adapt to noise conditions
  - too low : pick up too much unwanted energy
  - too high : loose too much of “real” energy. Corrections/Calibrations will be larger
- Sometimes several clustering stages, in order separate or combine nearby clusters

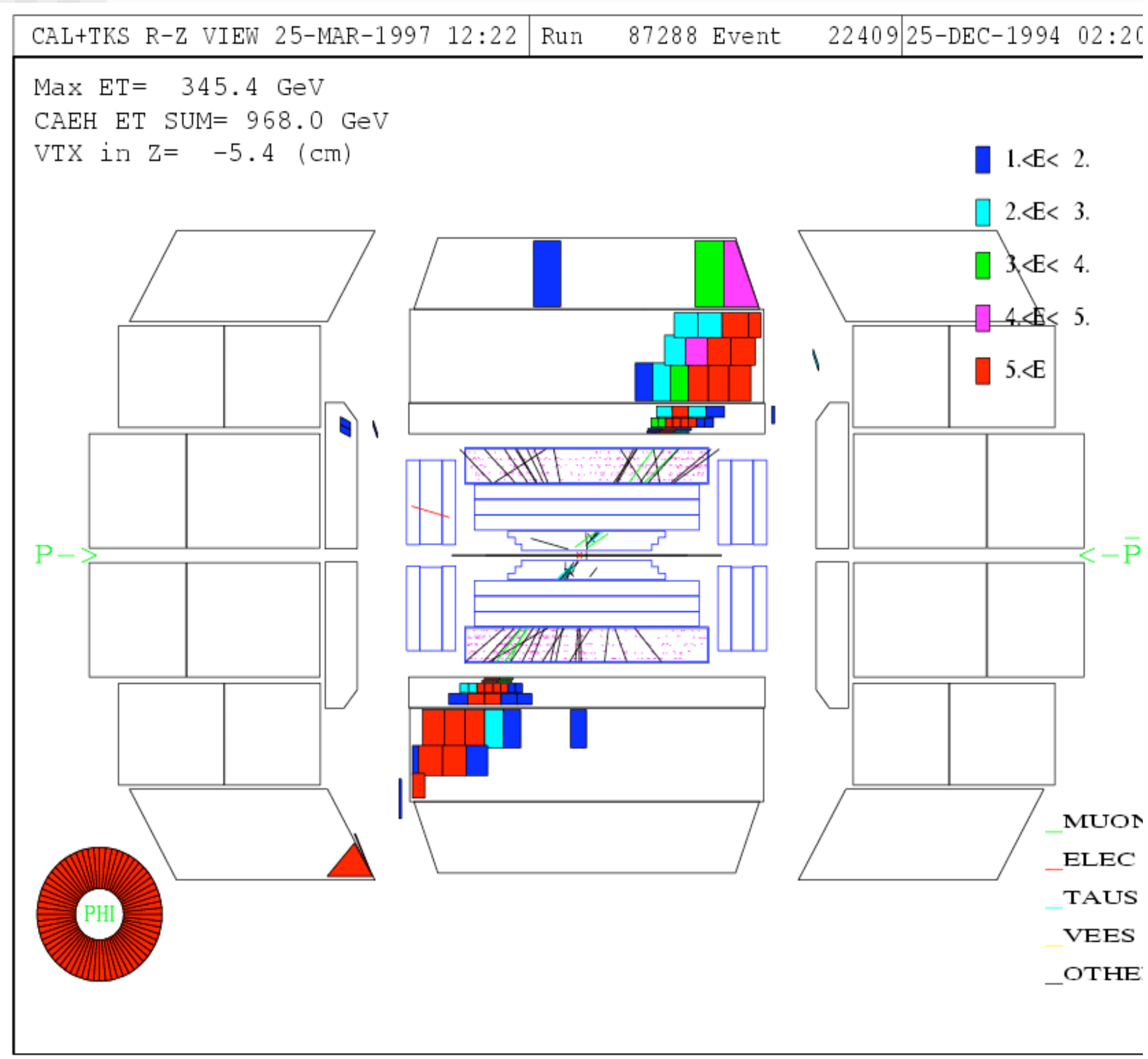
example : one lump or two?

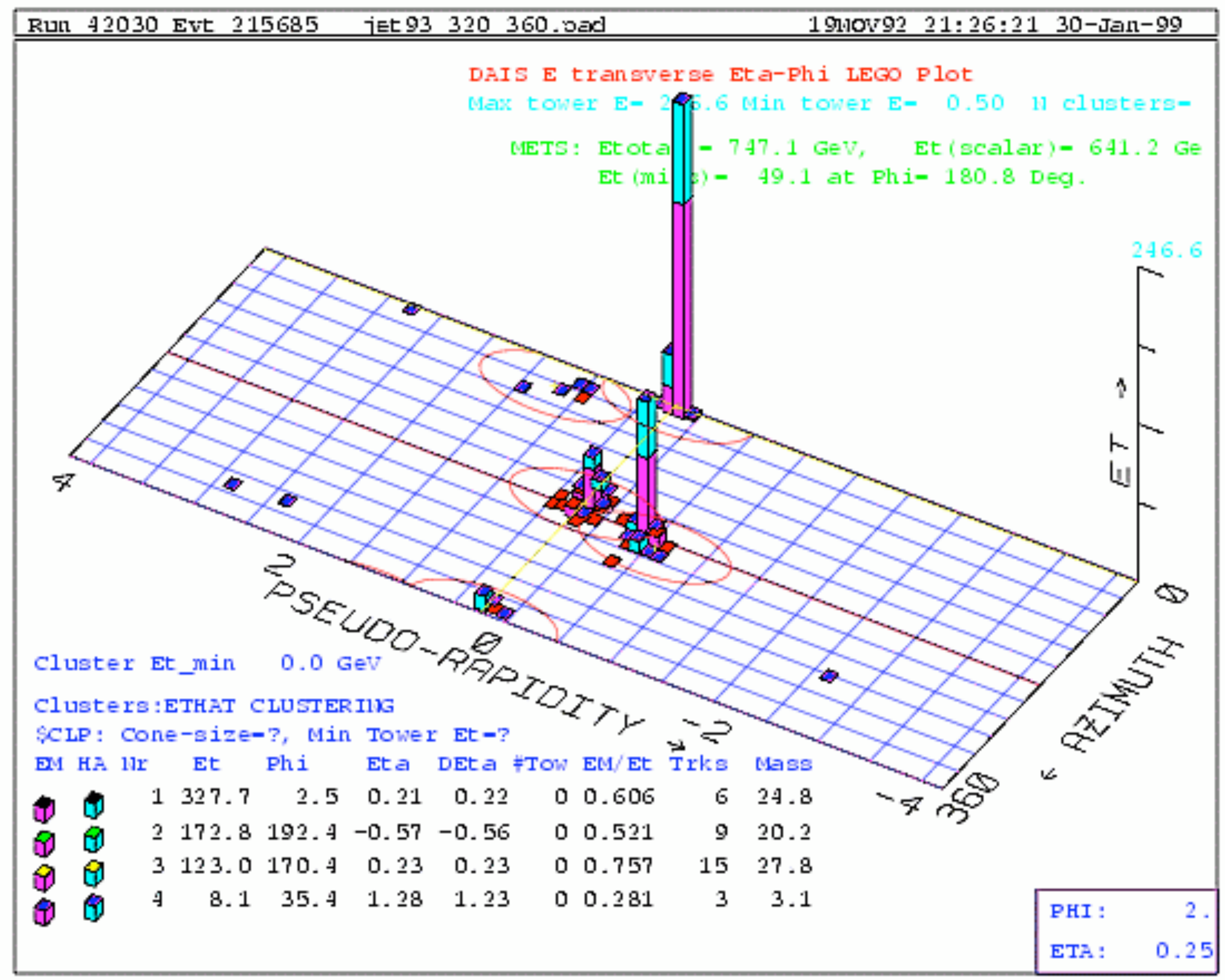


Jet Algorithms

Jets in Hadron Collider Detectors

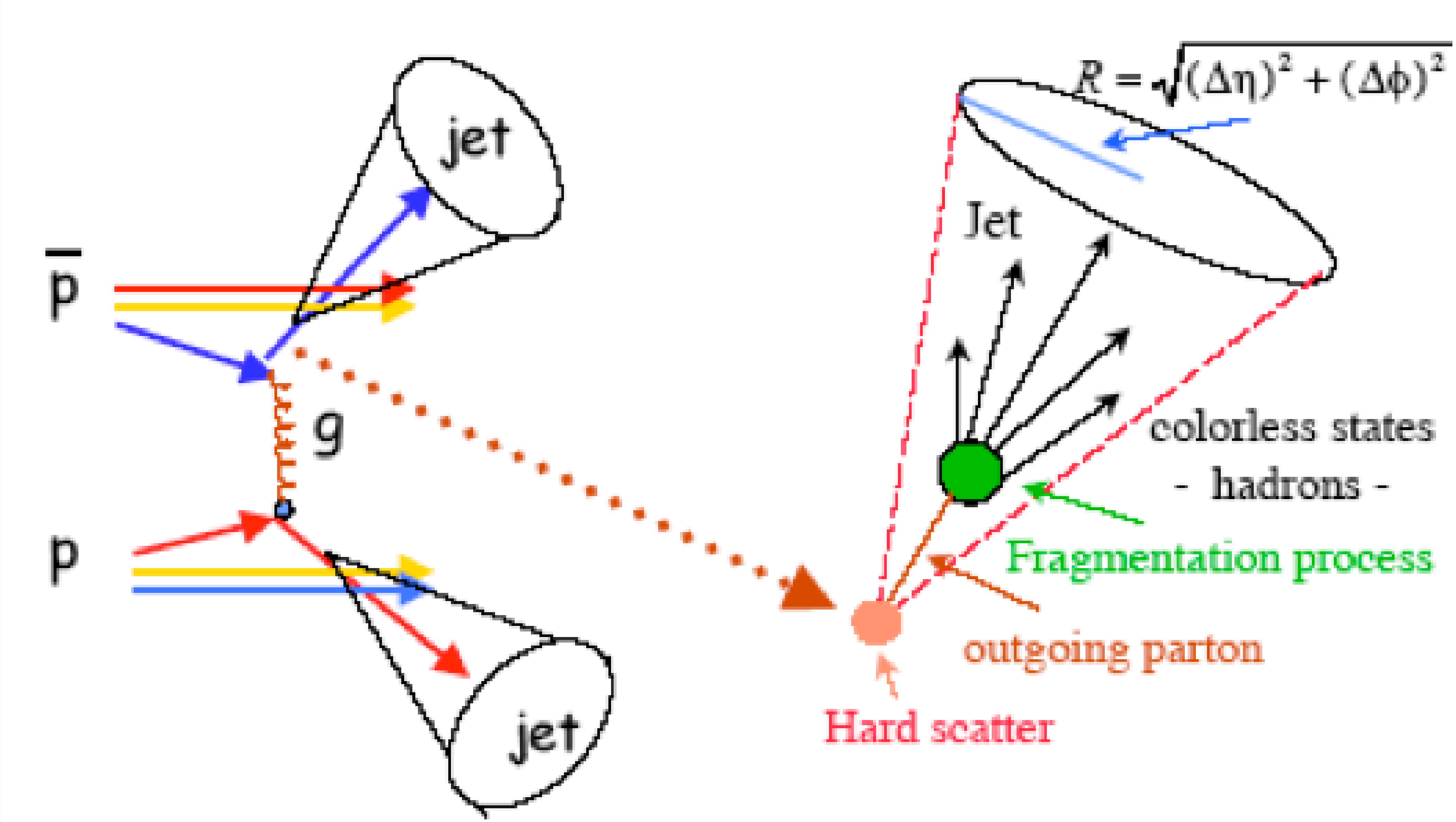
Jets in DØ CDF





- Introducing a cone prescription seems "natural"...
- But how to make it more quantitative?
  - don't want people "guessing" at whether there are 2,3, ... jets

Cones...



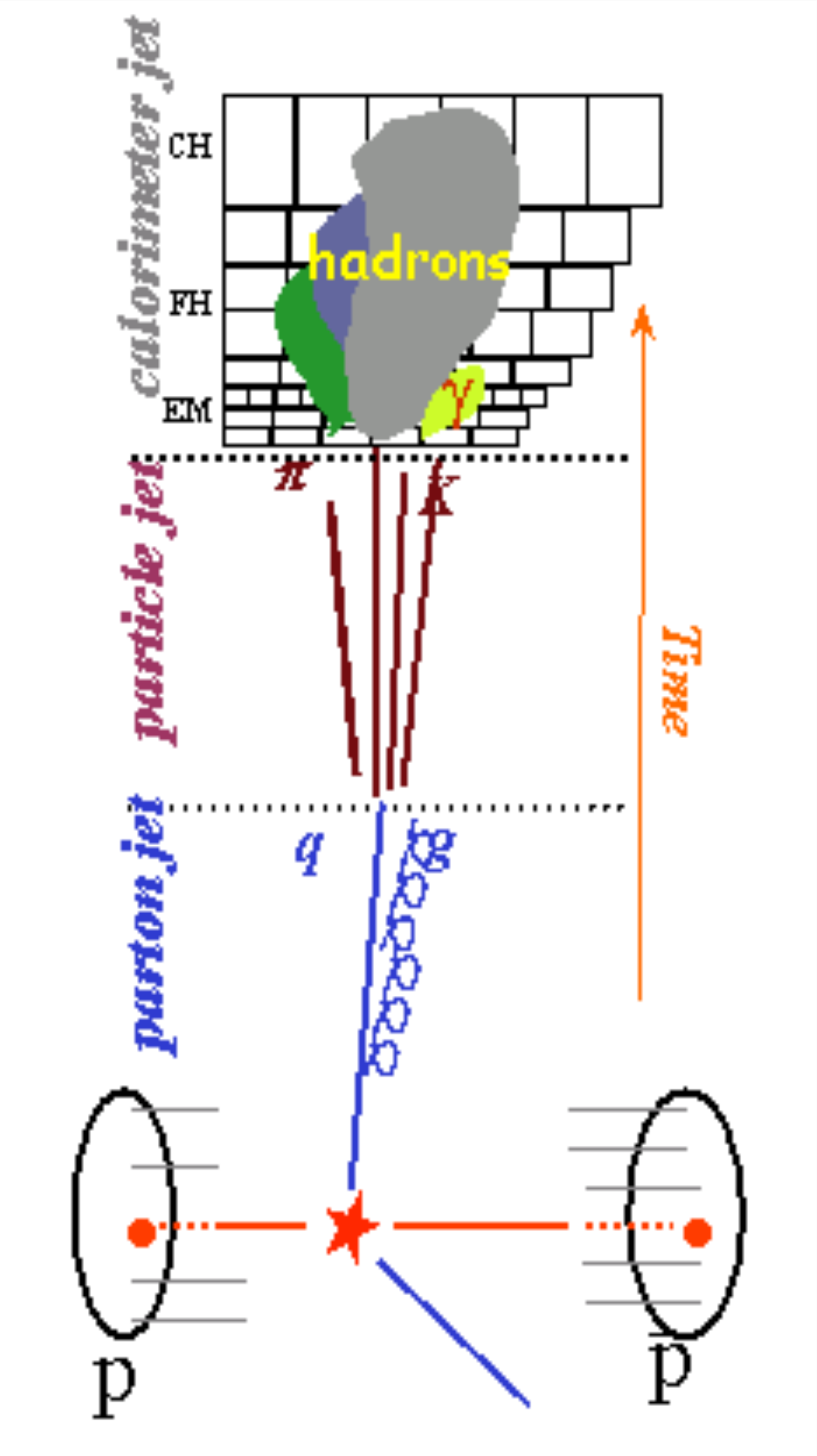
The natural (?) definition of a jet in a hadron collider environment

from J. Huston, CTEQ summer school 2004

Requirements

ETH Institute for Particle Physics

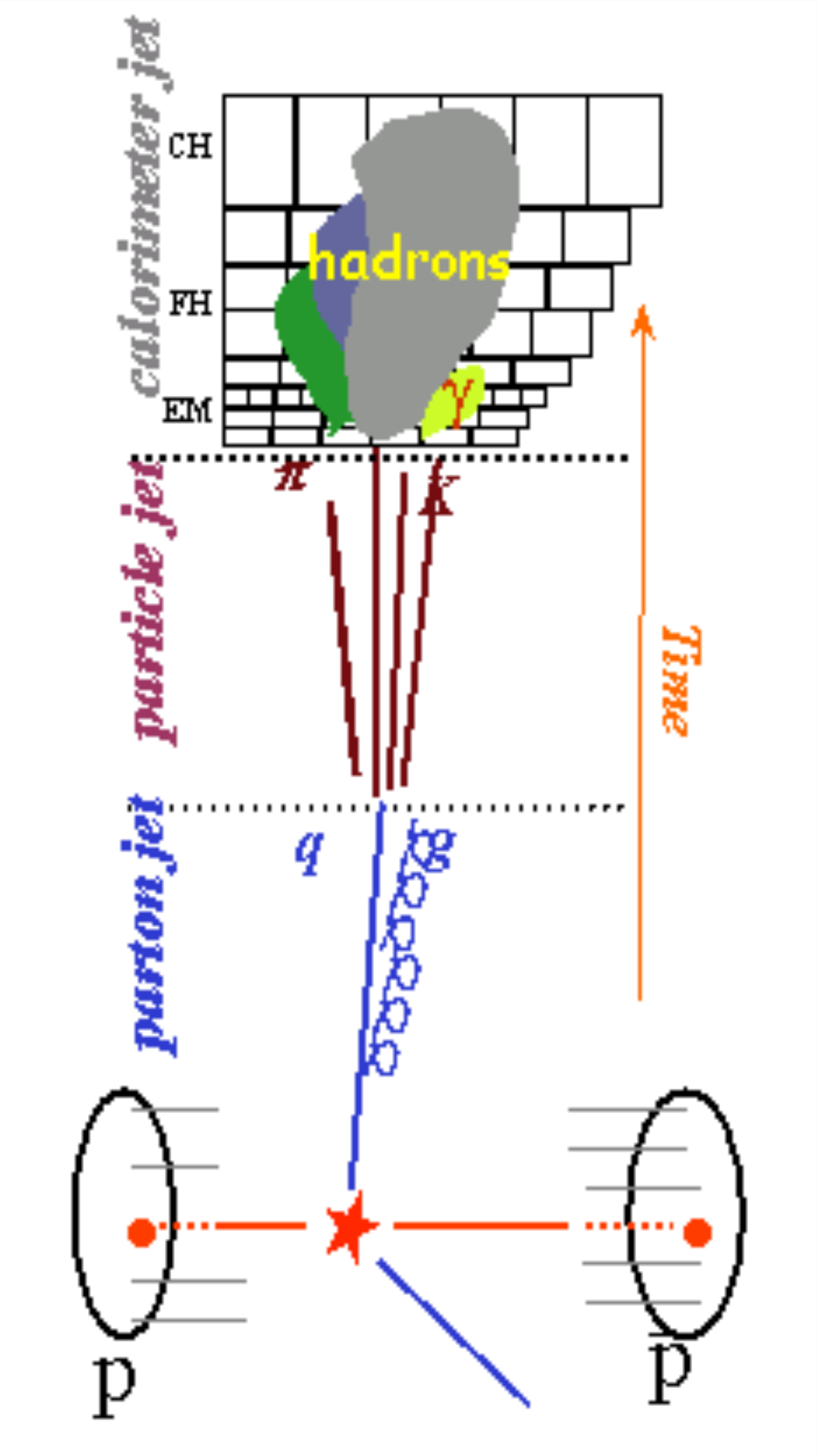
- Applicable at all levels
  - partons, stable particles
    - for theoretical calculations
  - measured objects (calorimeter objects, tracks, etc)
  - and always find the same jet



Requirements

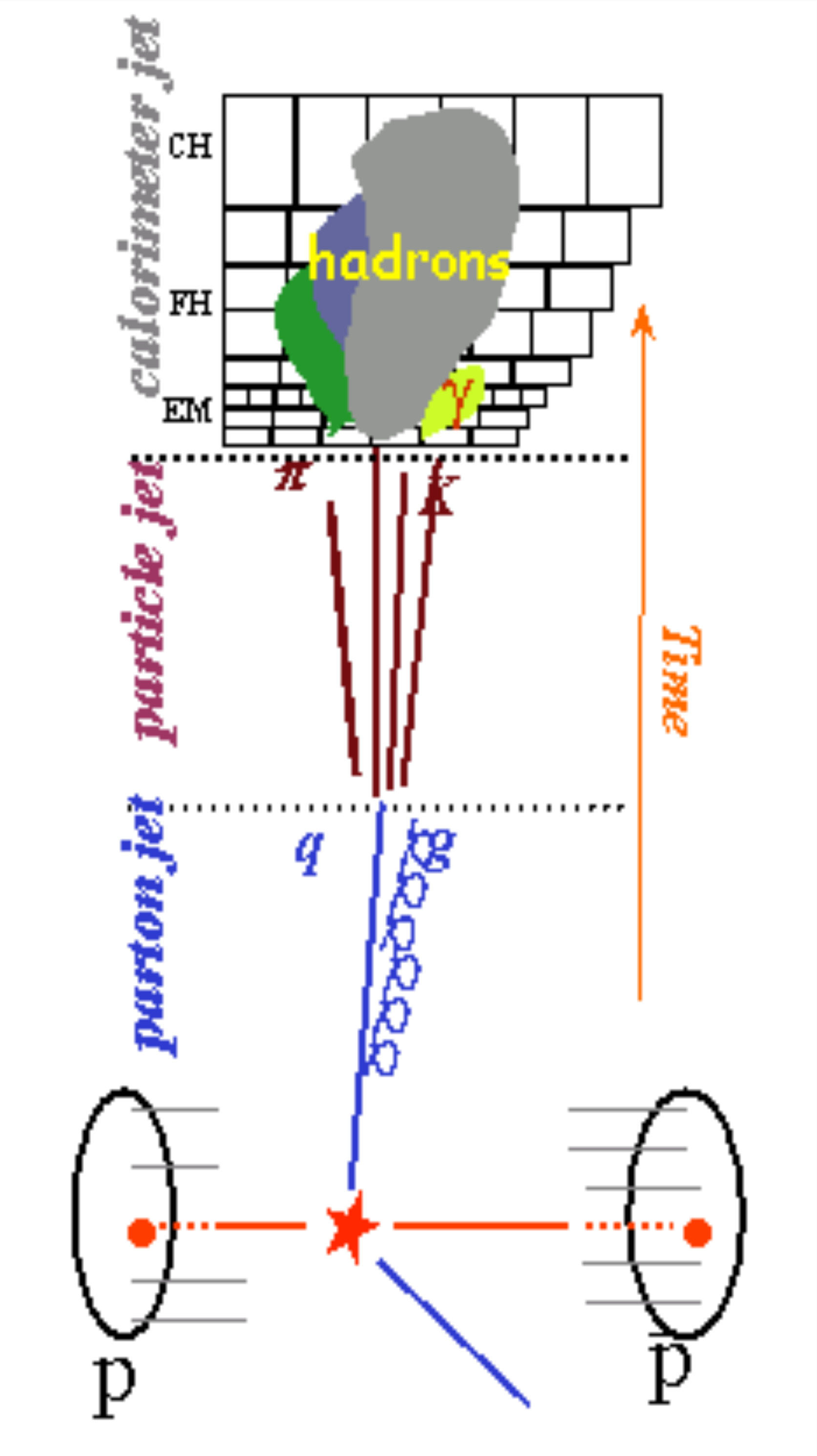
ETH Institute for Particle Physics

- Applicable at all levels
  - partons, stable particles
    - for theoretical calculations
  - measured objects (calorimeter objects, tracks, etc)
  - and always find the same jet
- Independent of the very details of the detector
  - example : granularity of the calorimeter, energy response,...



Requirements

- Applicable at all levels
  - partons, stable particles
    - for theoretical calculations
  - measured objects (calorimeter objects, tracks, etc)
  - and always find the same jet
- Independent of the very details of the detector
  - example : granularity of the calorimeter, energy response,...
- Easy to implement !



Requirements

- Applicable at all levels
  - partons, stable particles
    - for theoretical calculations
  - measured objects (calorimeter objects, tracks, etc)
  - and always find the same jet

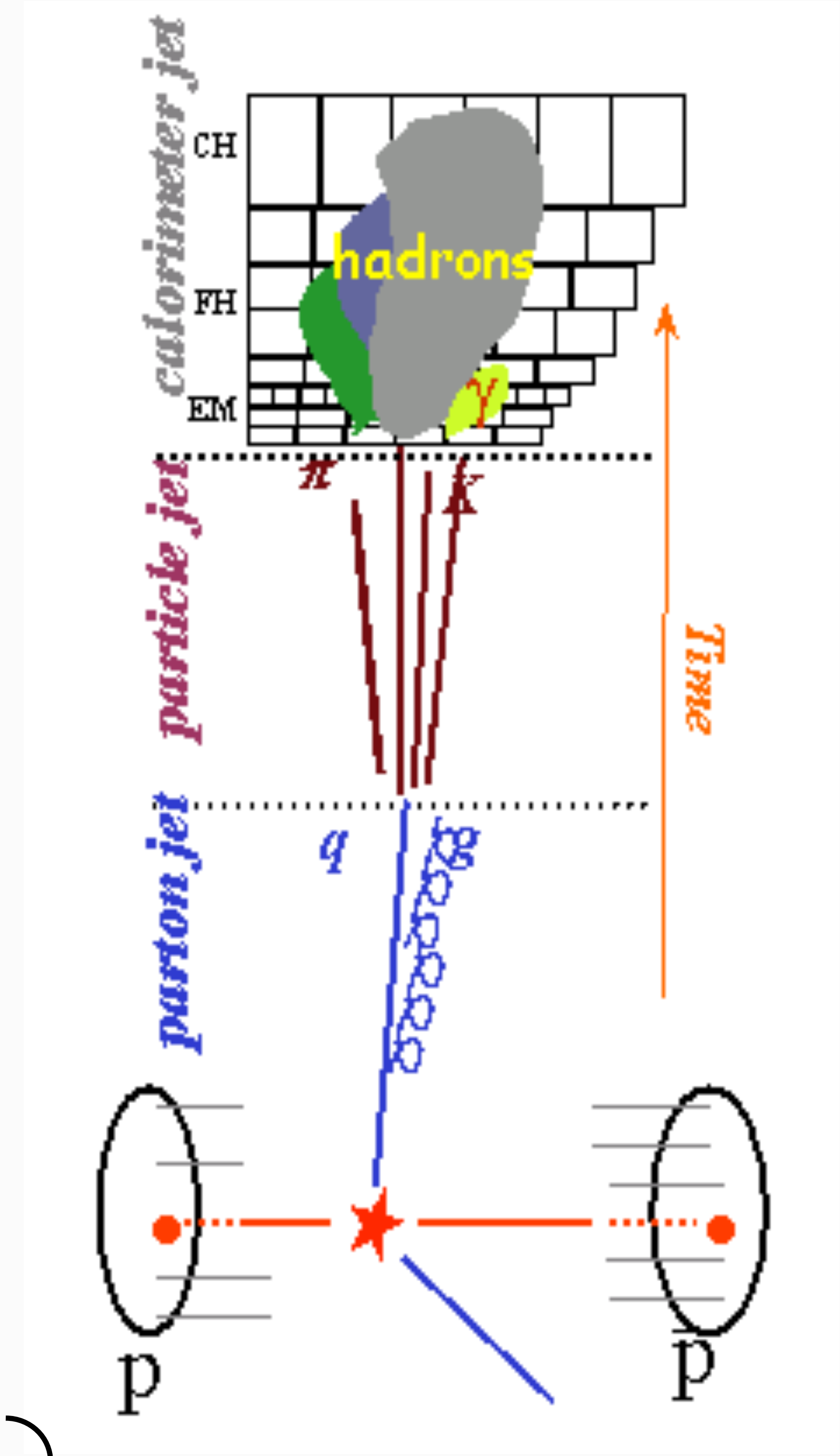
- Independent of the very details of the detector
  - example : granularity of the calorimeter, energy response,...

- Easy to implement !

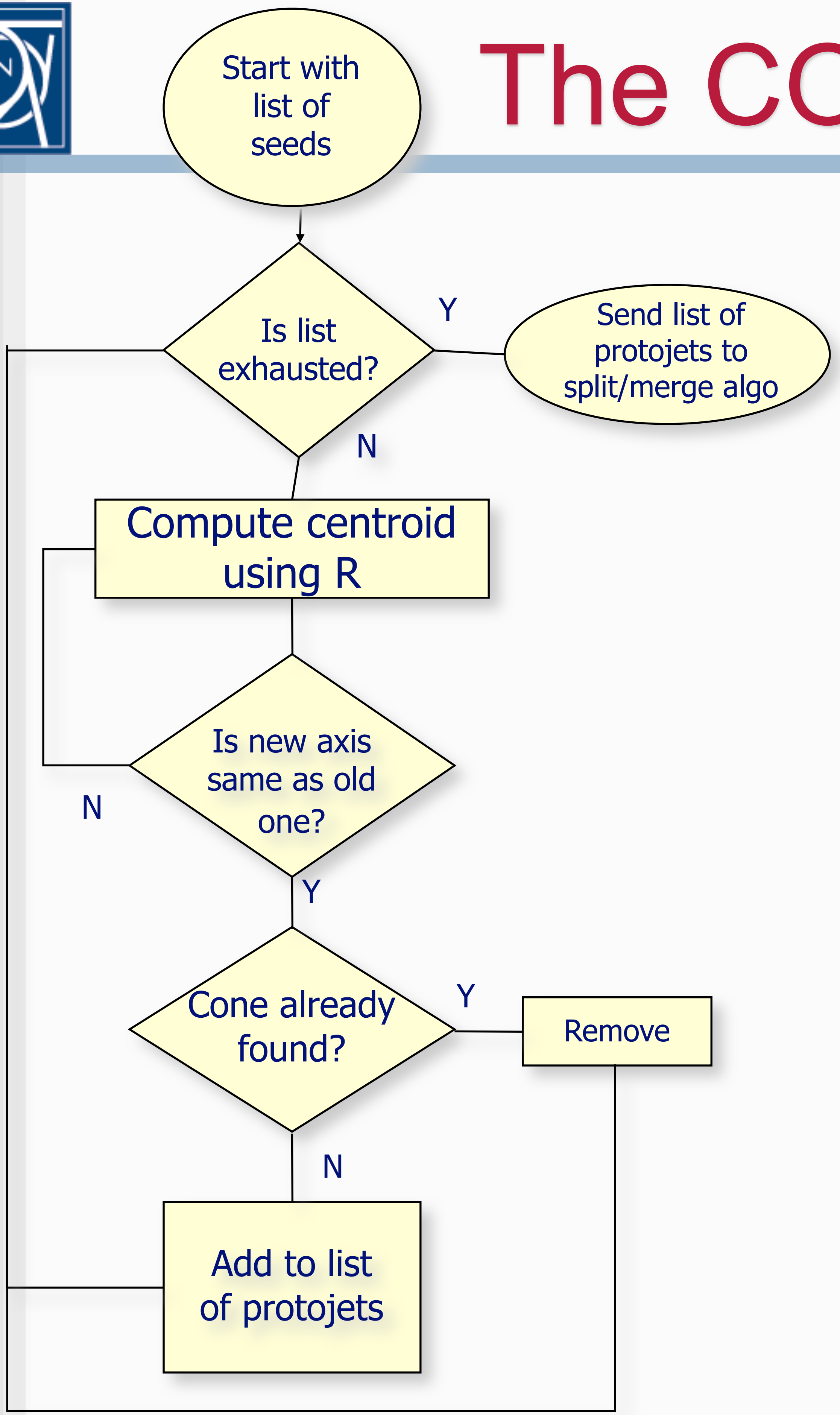
- Close correspondence between

$$P_{\text{parton}} \longleftrightarrow P_{\text{jet}}$$

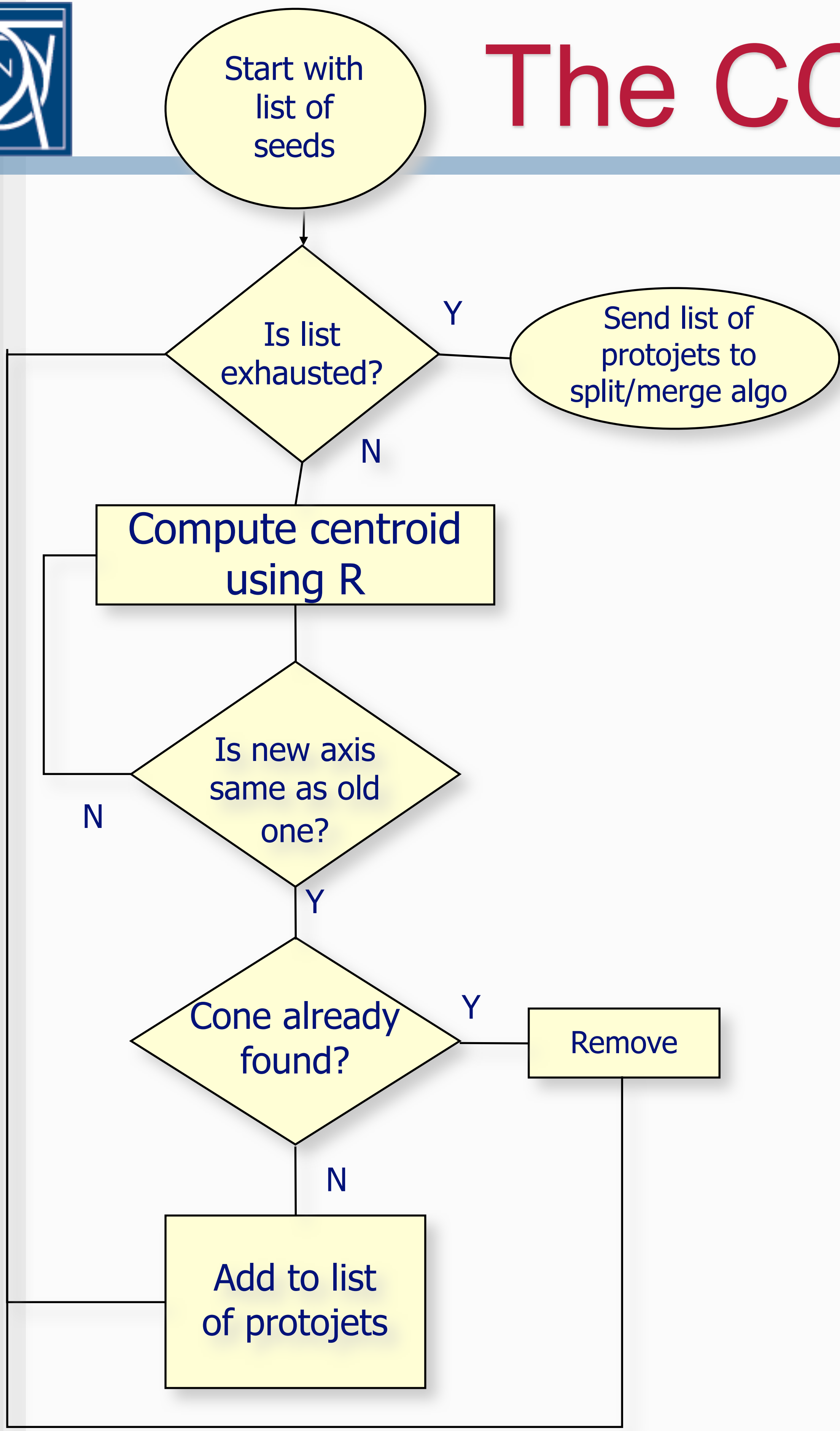
(Energy Momentum angle)



The CONE algorithm

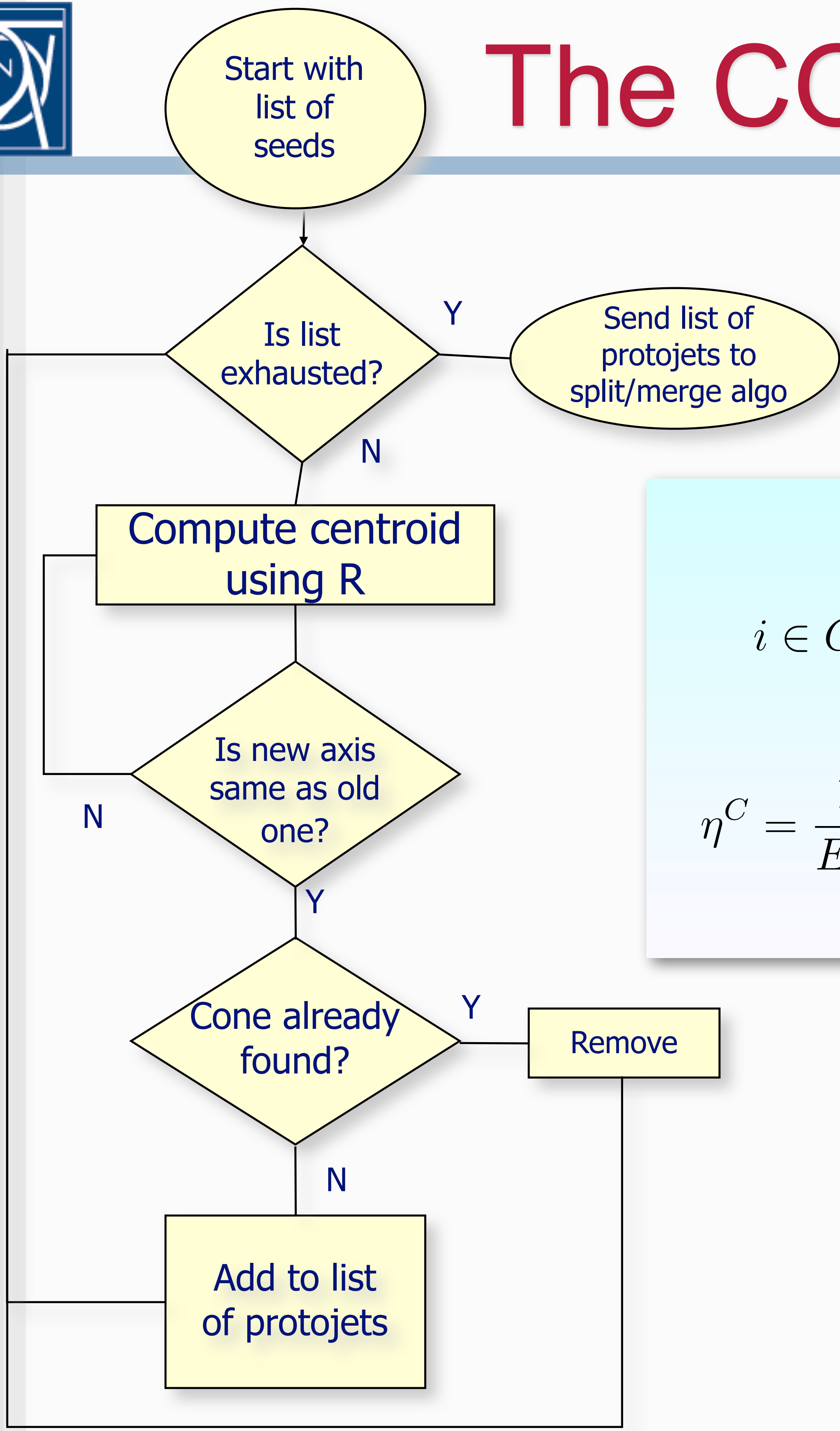


The CONE algorithm



Seeds: for example, energy deposits with transverse energy ($E_T = E \sin\theta$) > 2 GeV in a tower of the calorimeter

The CONE algorithm



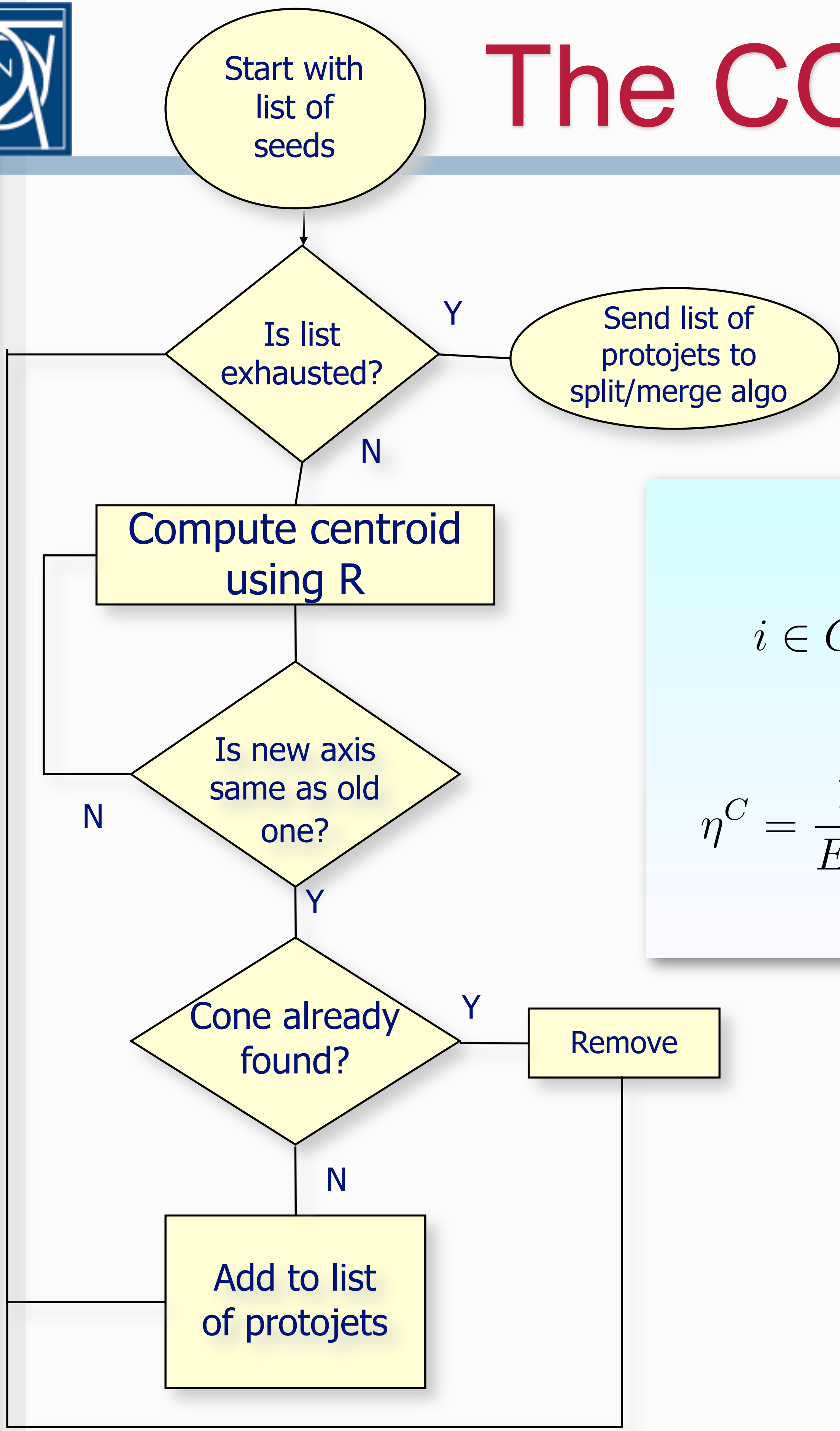
Seeds: for example, energy deposits with transverse energy (E_T = E sinθ) > 2 GeV in a tower of the calorimeter

Centroid (one possible def) :

$$i \in C \quad : \quad \sqrt{(\eta^i - \eta^C)^2 + (\Phi^i - \Phi^C)^2} \leq R \quad \text{cone radius}$$

$$\eta^C = \frac{1}{E_T^C} \sum_{i \in C} E_T^i \eta^i \quad ; \quad \Phi^C = \frac{1}{E_T^C} \sum_{i \in C} E_T^i \Phi^i \quad ; \quad E_T^C = \sum_{i \in C} E_T^i$$

The CONE algorithm

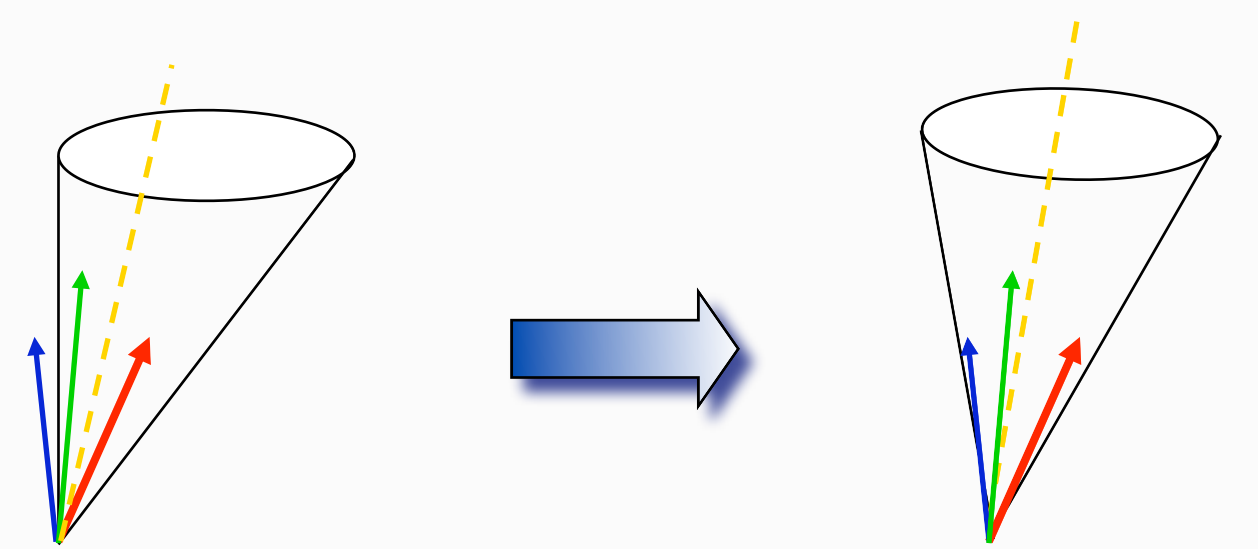


Seeds: for example, energy deposits with transverse energy (E_T = E sinθ) > 2 GeV in a tower of the calorimeter

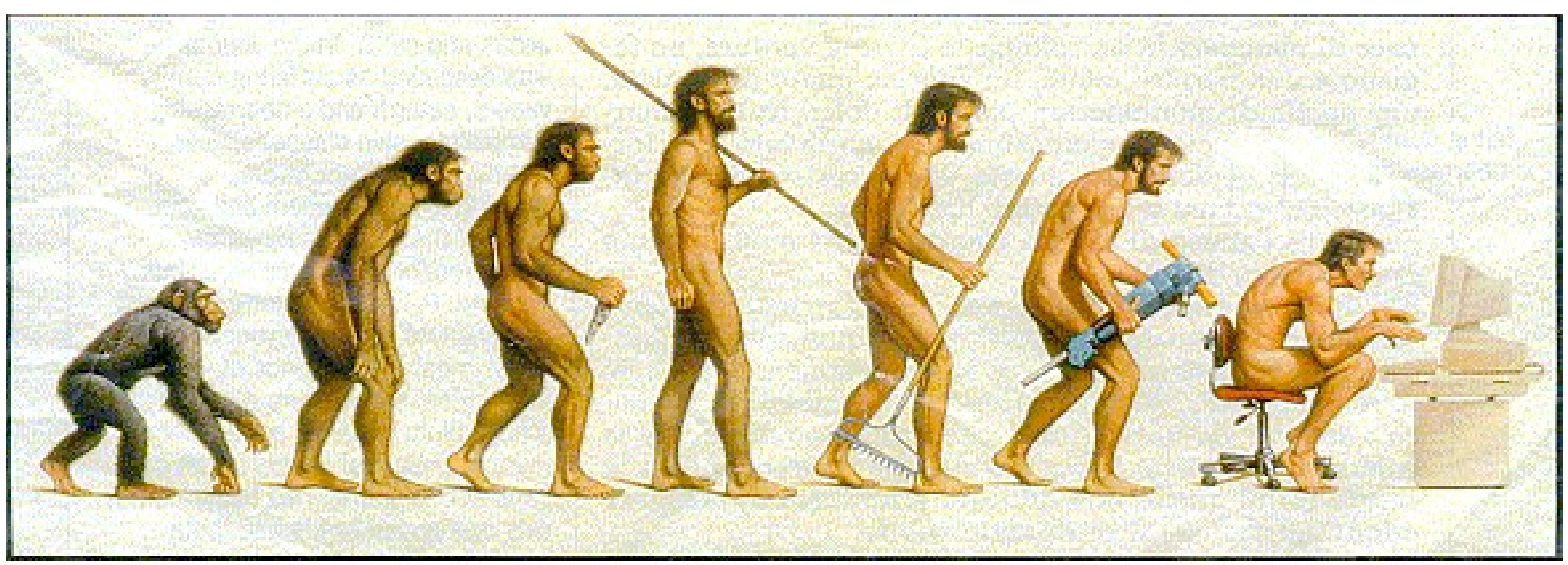
Centroid (one possible def) :

$$i \in C \quad : \quad \sqrt{(\eta^i - \eta^C)^2 + (\Phi^i - \Phi^C)^2} \leq R \quad \text{cone radius}$$

$$\eta^C = \frac{1}{E_T^C} \sum_{i \in C} E_T^i \eta^i \quad ; \quad \Phi^C = \frac{1}{E_T^C} \sum_{i \in C} E_T^i \Phi^i \quad ; \quad E_T^C = \sum_{i \in C} E_T^i$$



The computing behind all this



Somewhere, something went terribly wrong

Some numbers

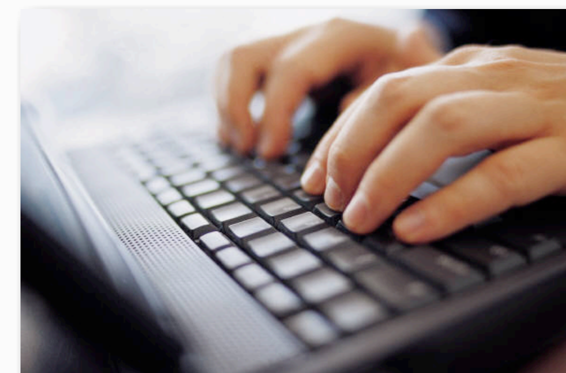
- Examples from CMS, estimates
  - Rate of events streaming out from High-Level Trigger farm ~150 Hz
  - each event has a size of the order of 200 kByte

Some numbers

- Examples from CMS, estimates
  - Rate of events streaming out from High-Level Trigger farm ~150 Hz
  - each event has a size of the order of 200 kByte
- CMS will record ~100k top-quark events per day
  - among about 10^7 events in total per day
  - will have roughly 150 “physics” days per year
  - thus about 10^9 evts/year, a few Pbyte

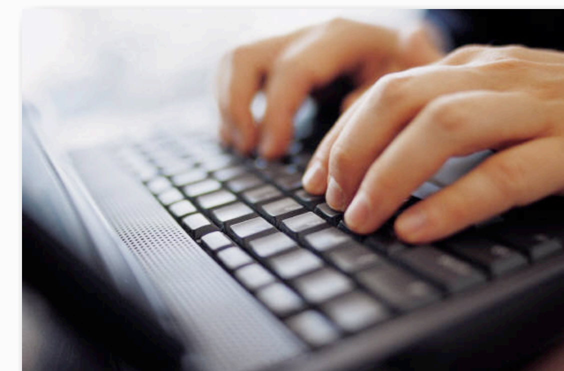
Some numbers

- Examples from CMS, estimates
  - Rate of events streaming out from High-Level Trigger farm ~150 Hz
  - each event has a size of the order of 200 kByte
- CMS will record ~100k top-quark events per day
  - among about 10^7 events in total per day
  - will have roughly 150 “physics” days per year
  - thus about 10^9 evts/year, a few Pbyte
- “prompt” processing
  - Expect to do first reprocessing step within one day
  - Reco time per event on std. CPU: < 5 sec (on lxplus)
  - Note : will have to reprocess several times
    - new/better algorithms, updated calibrations, etc.



Some numbers

- Examples from CMS, estimates
  - Rate of events streaming out from High-Level Trigger farm ~150 Hz
  - each event has a size of the order of 200 kByte
- CMS will record ~100k top-quark events per day
  - among about 10^7 events in total per day
  - will have roughly 150 “physics” days per year
  - thus about 10^9 evts/year, a few Pbyte

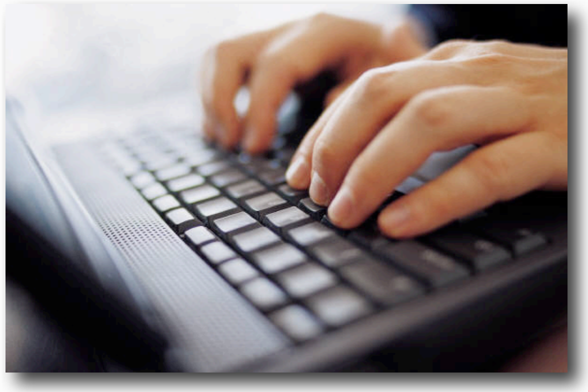


- “prompt” processing
  - Expect to do first reprocessing step within one day
  - Reco time per event on std. CPU: < 5 sec (on lxplus)
  - Note : will have to reprocess several times
    - new/better algorithms, updated calibrations, etc.
- Expect to simulate several 100s to 1000s of millions of events
  - will be mostly done at computing centres outside CERN
  - Simulation time per event now ~ 100 secs (eg. for QCD or top evts)

Some numbers

ETH Institute for Particle Physics

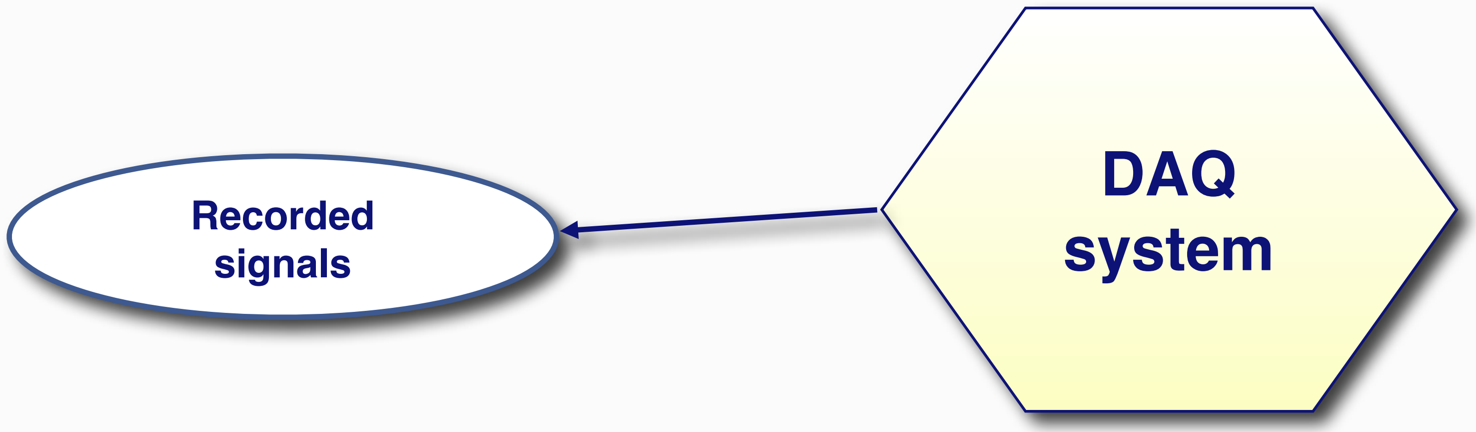
- Examples from CMS, estimates
  - Rate of events streaming out from High-Level Trigger farm ~150 Hz
  - each event has a size of the order of 200 kByte
- CMS will record ~100k top-quark events per day
  - among about 10^7 events in total per day
  - will have roughly 150 “physics” days per year
  - thus about 10^9 evts/year, a few Pbyte



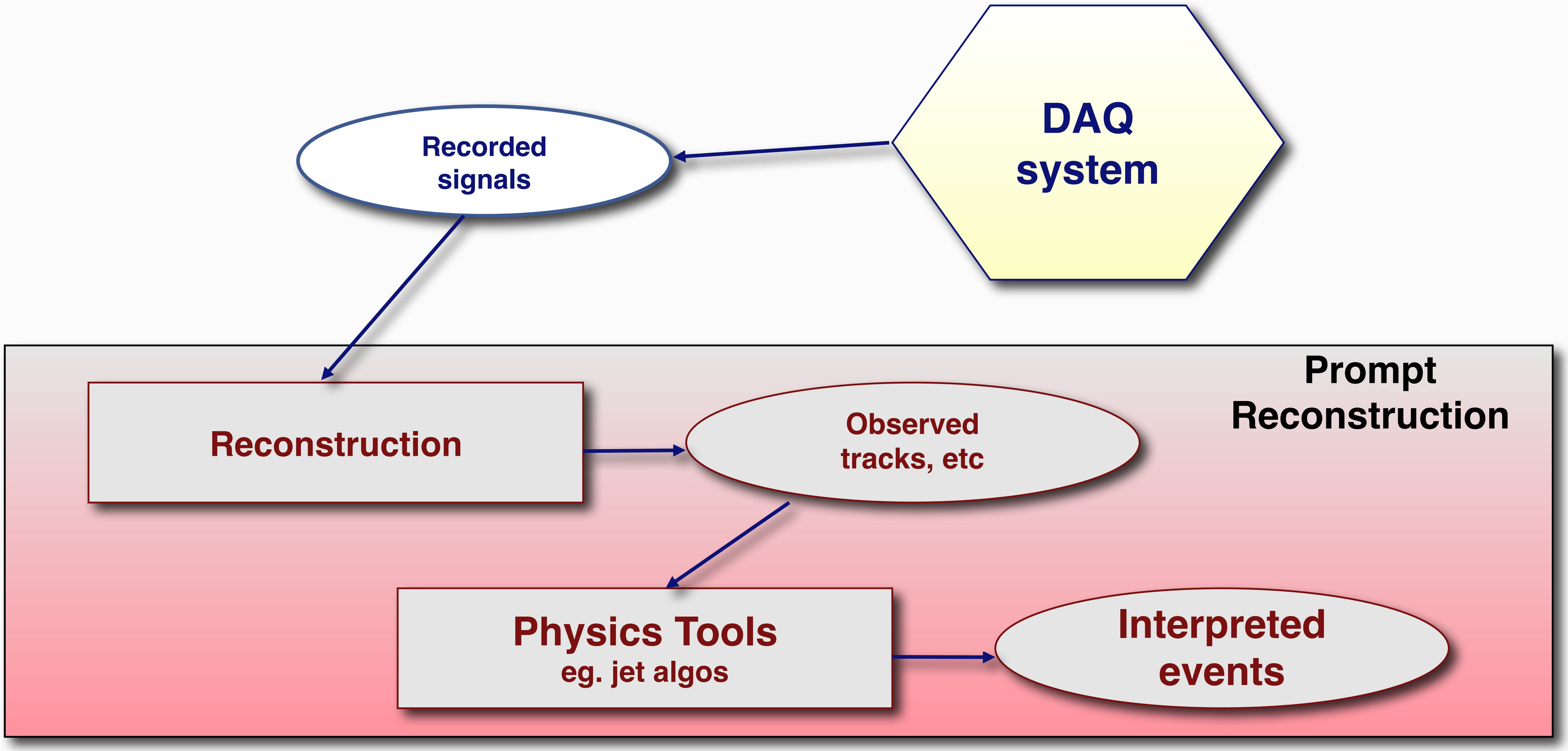
- “prompt” processing
  - Expect to do first reprocessing step within one day
  - Reco time per event on std. CPU: < 5 sec (on lxplus)
  - Note : will have to reprocess several times
    - new/better algorithms, updated calibrations, etc.
- Expect to simulate several 100s to 1000s of millions of events
  - will be mostly done at computing centres outside CERN
  - Simulation time per event now ~ 100 secs (eg. for QCD or top evts)

- Now : ~2 million lines of code (reconstruction and simulation)

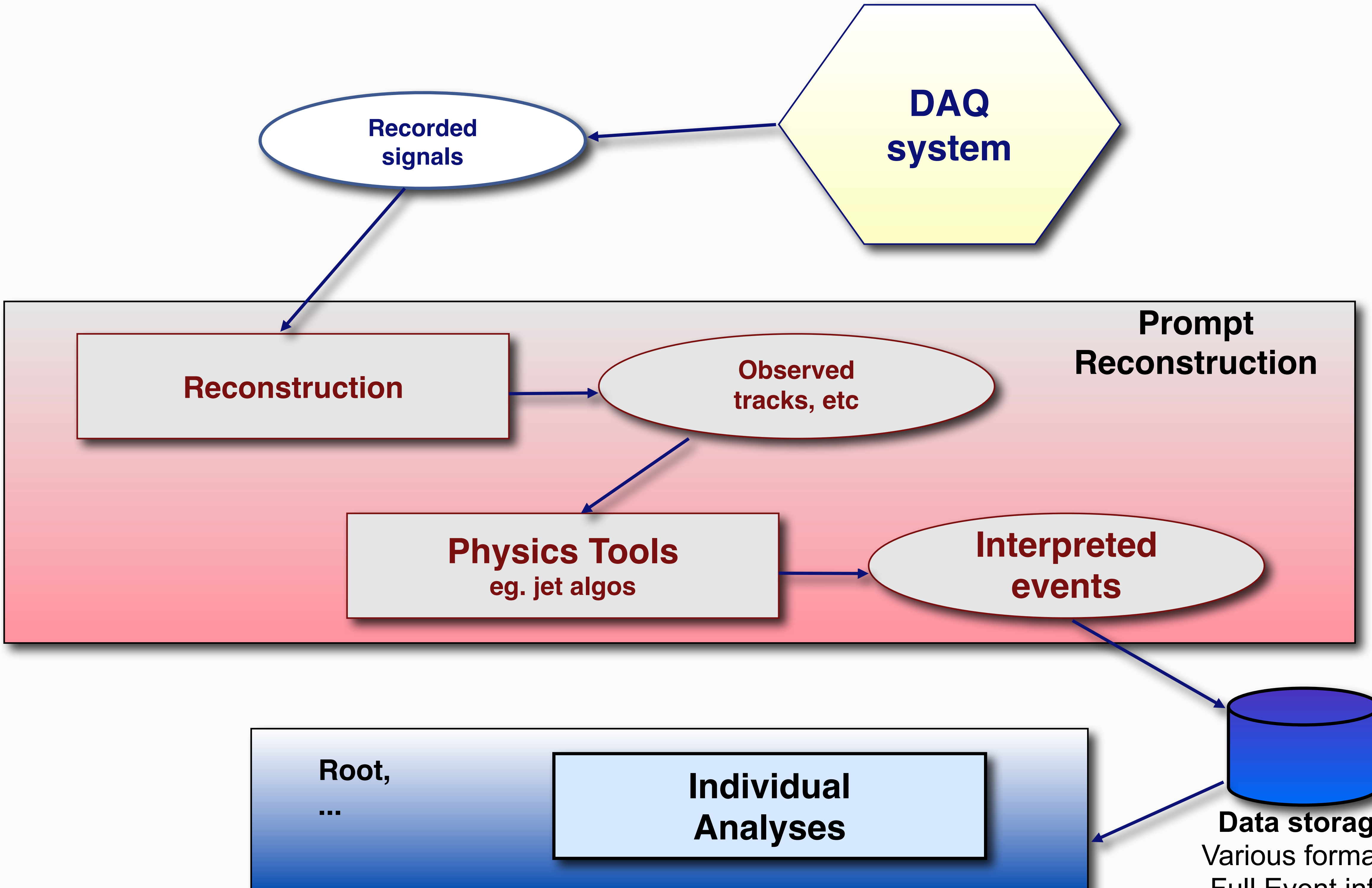
RECO flow



RECO flow

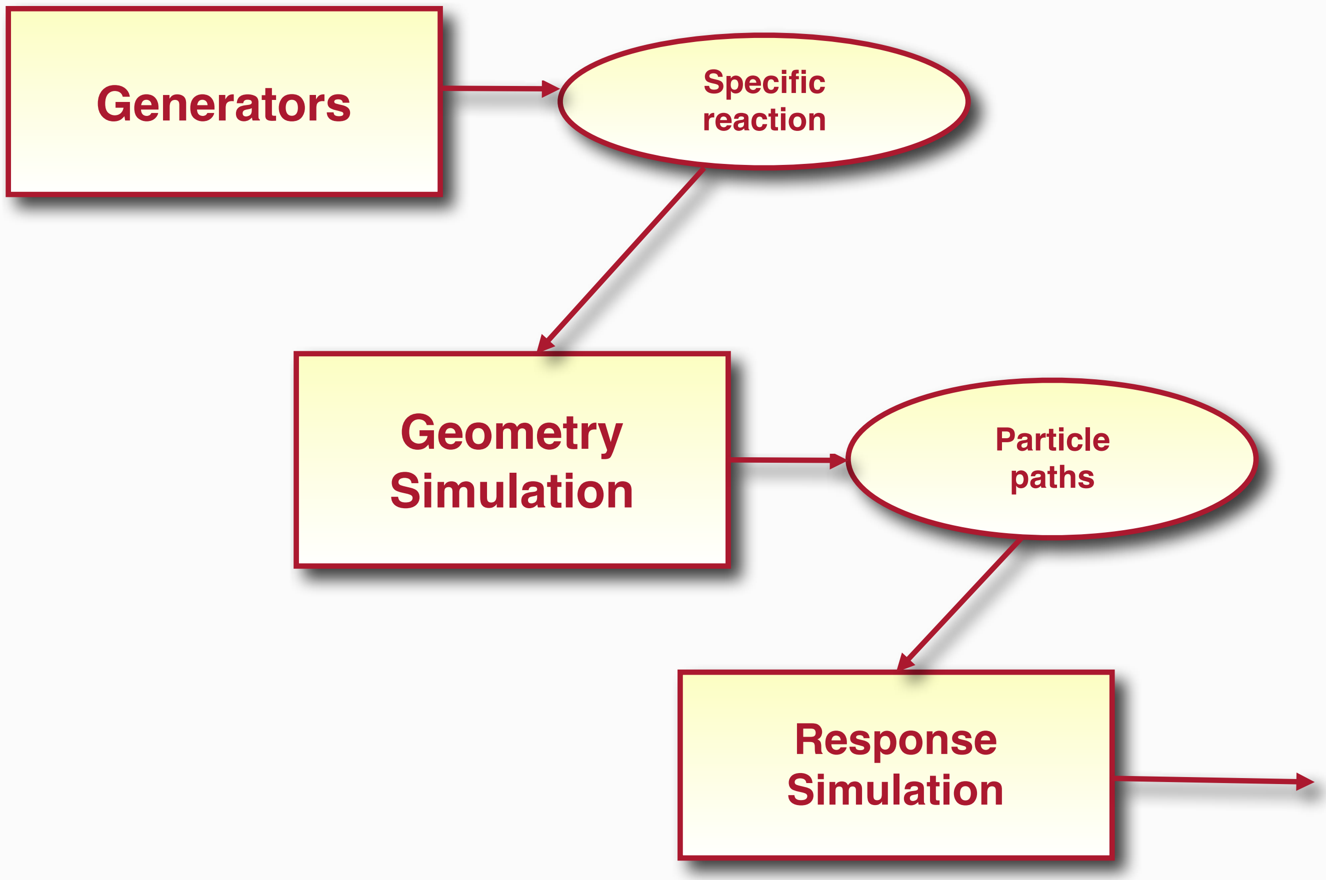

DAQ system
Recorded signals
Prompt Reconstruction
Reconstruction
Observed tracks, etc
Physics Tools
eg. jet algos
Interpreted events

RECO flow

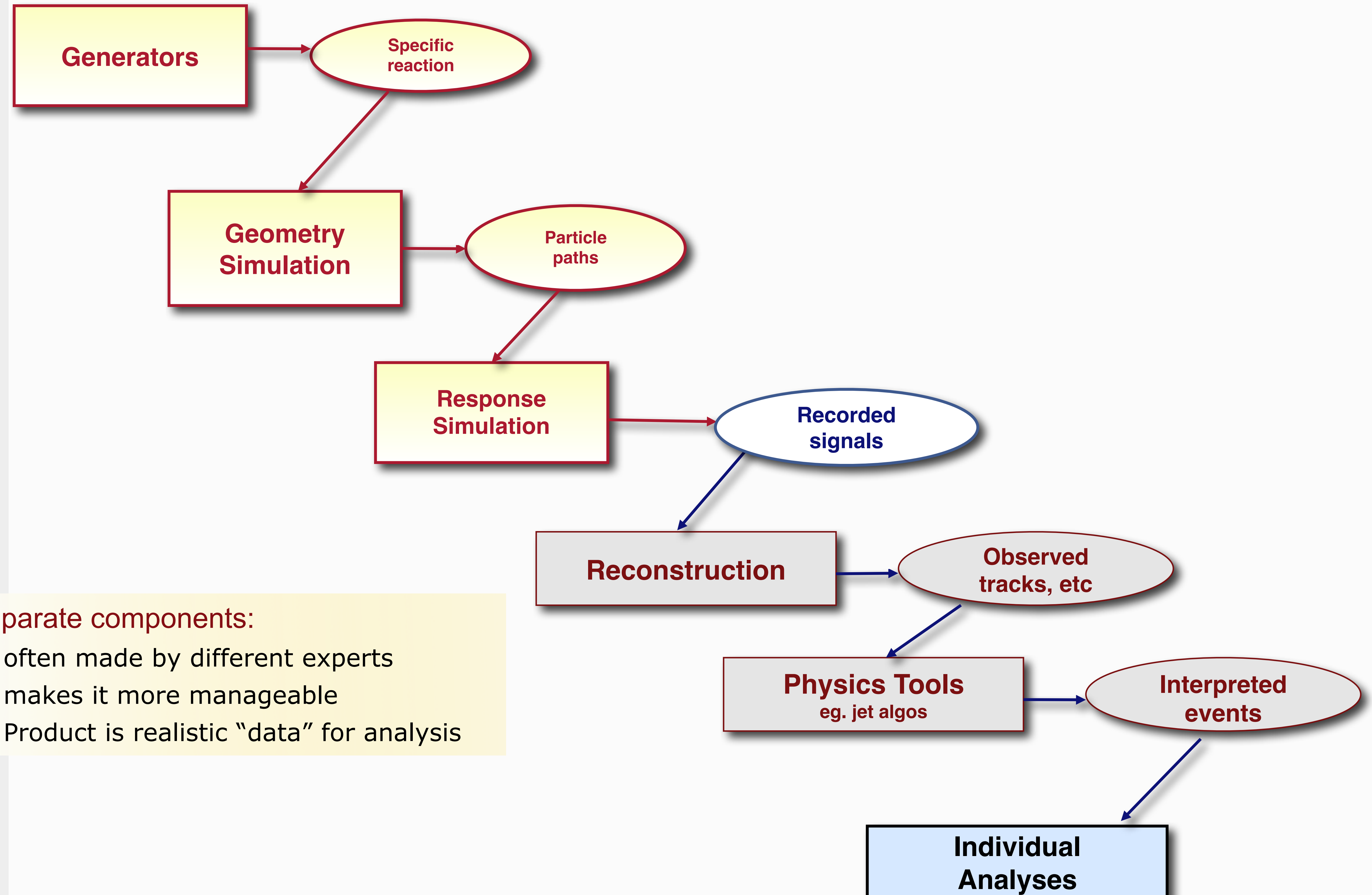


Data storage
Various formats:
Full Event info,
only RECO info,
reduced/selected RECO info

Flow of simulated data



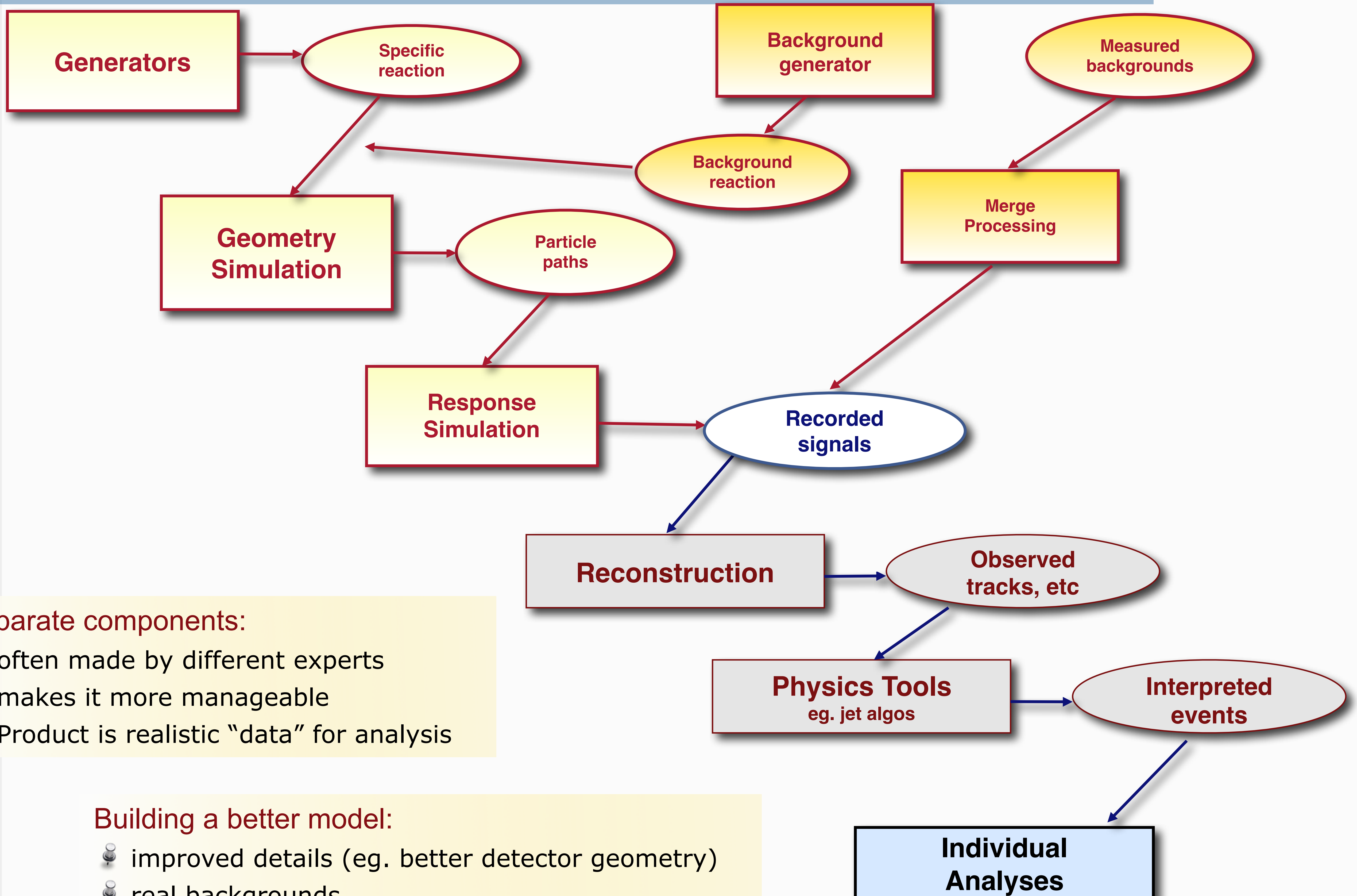
Flow of simulated data



Separate components:

- often made by different experts
- makes it more manageable
- Product is realistic “data” for analysis

Flow of simulated data



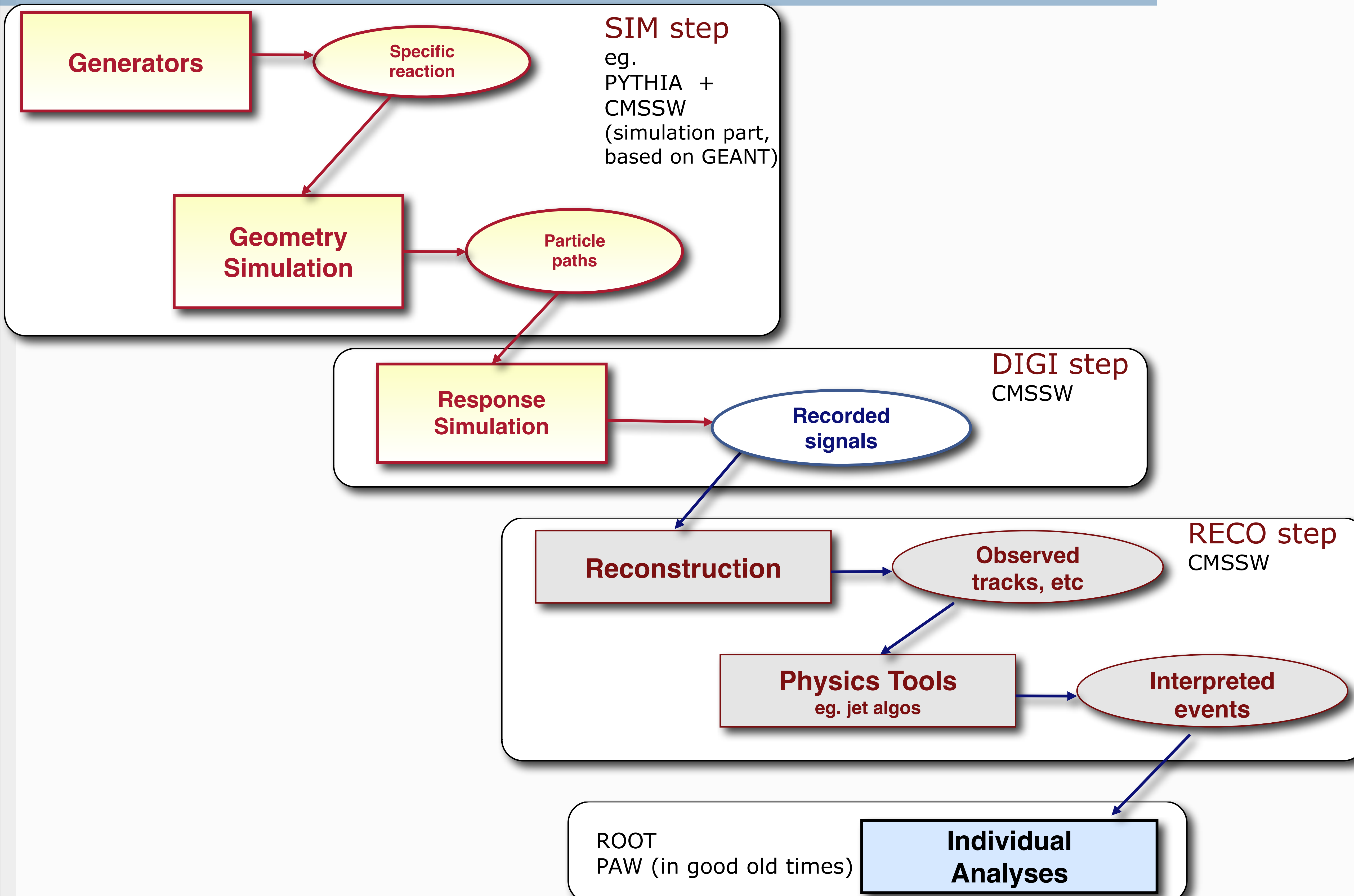
Separate components:

- often made by different experts
- makes it more manageable
- Product is realistic "data" for analysis

Building a better model:

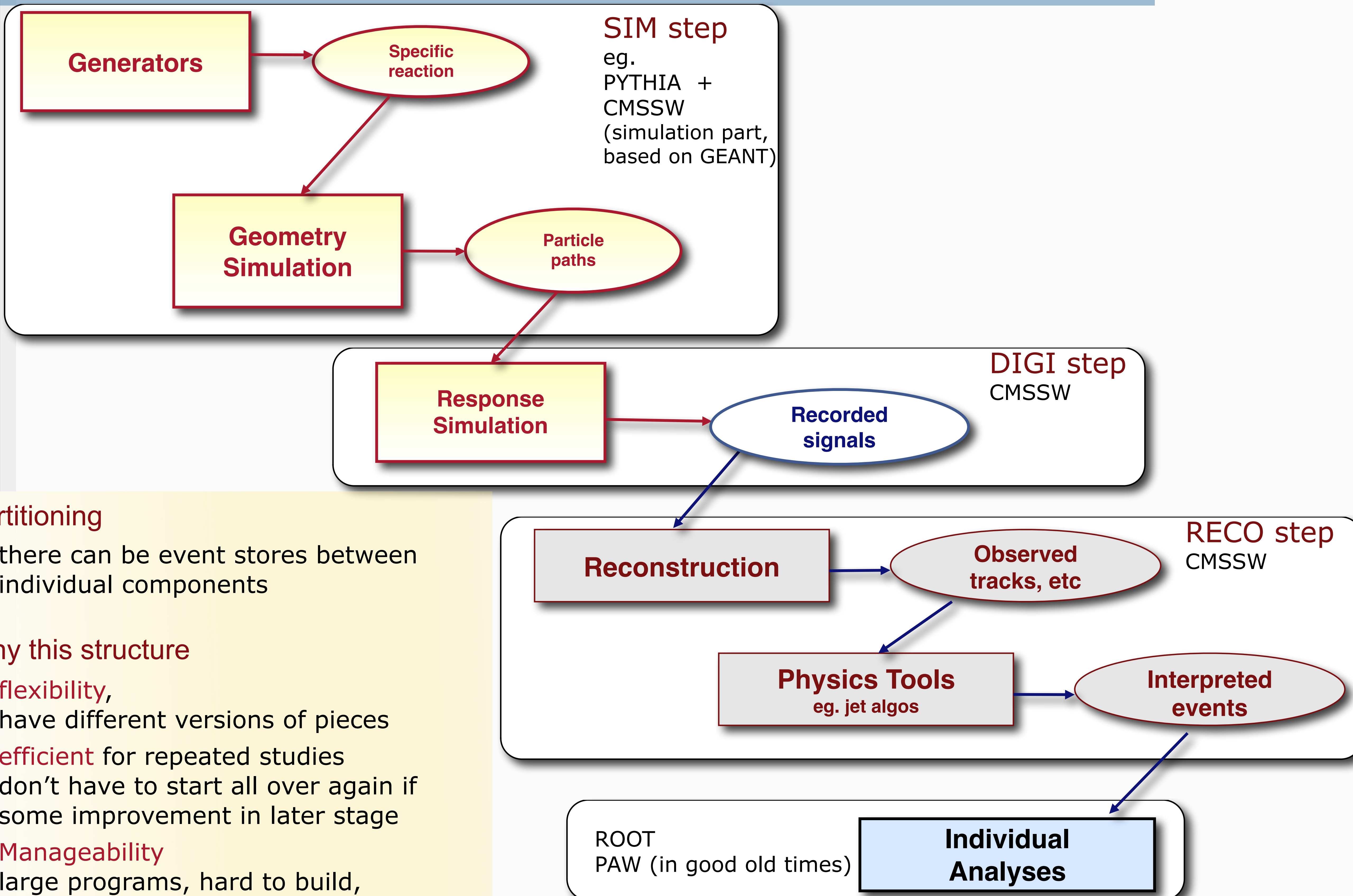
- improved details (eg. better detector geometry)
- real backgrounds

Partitioning production systems



Partitioning production systems

ETH Institute for Particle Physics



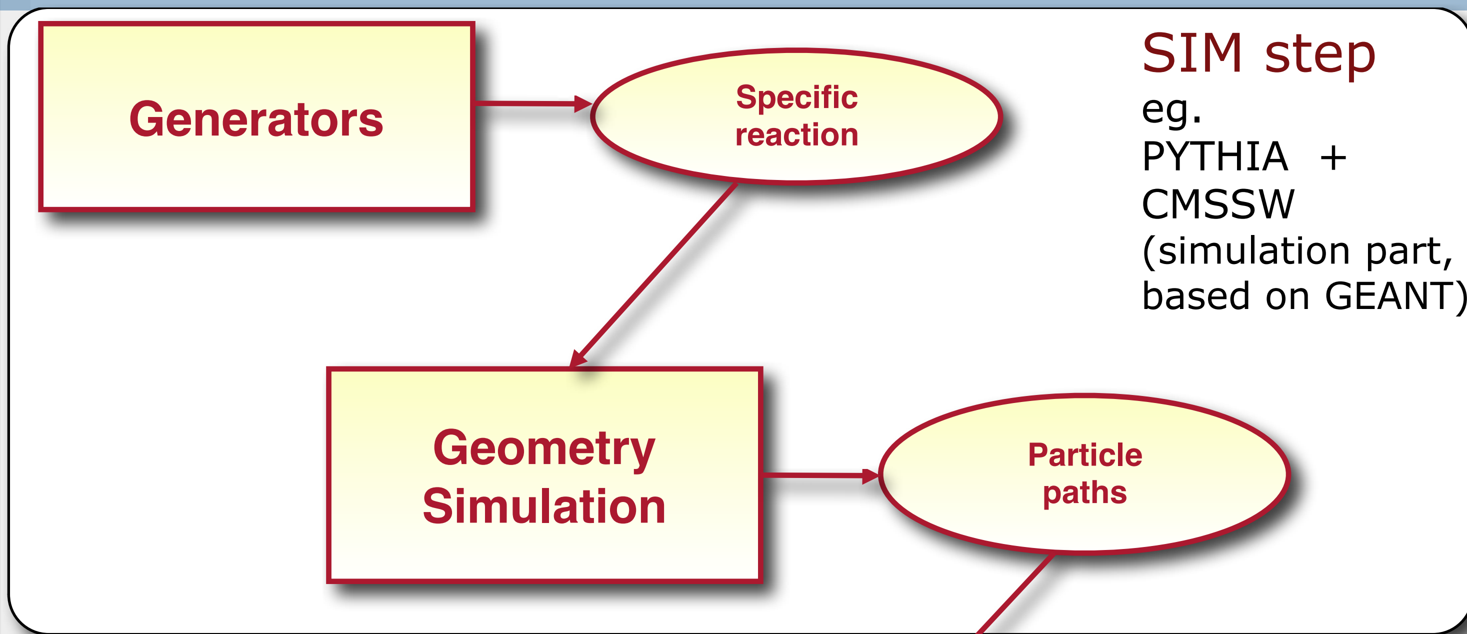
Partitioning

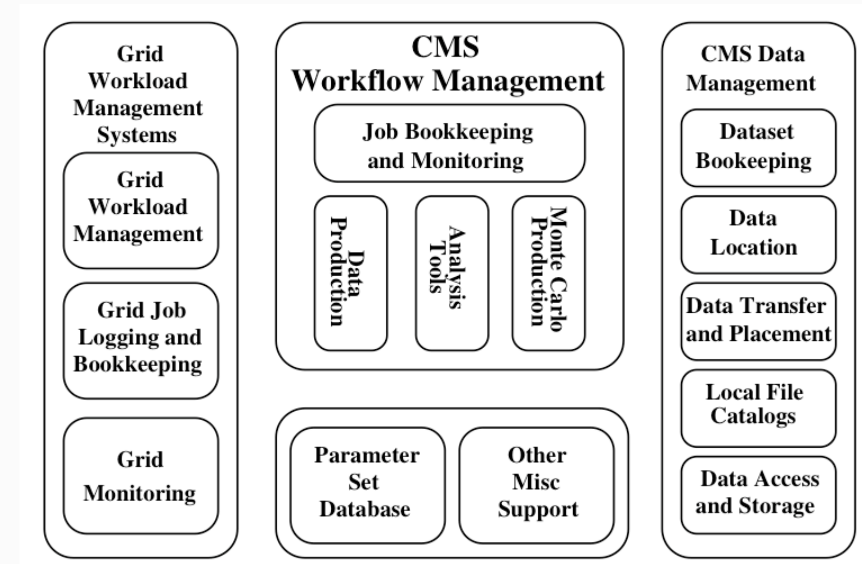
- there can be event stores between individual components

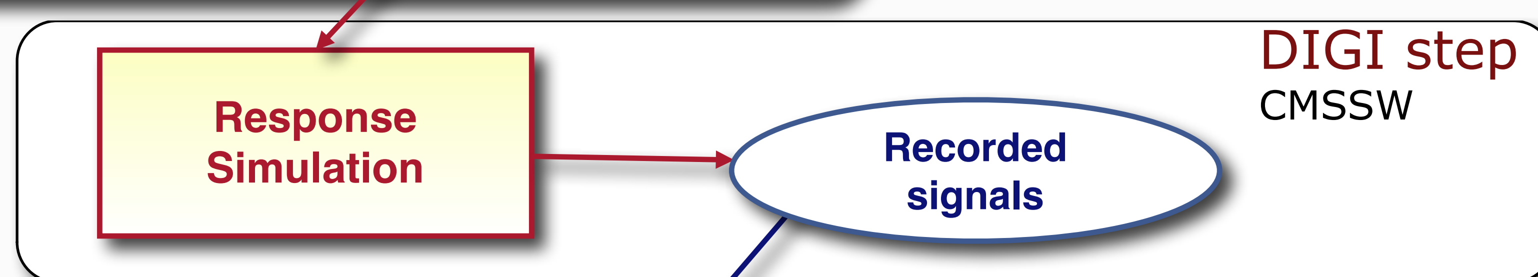
Why this structure

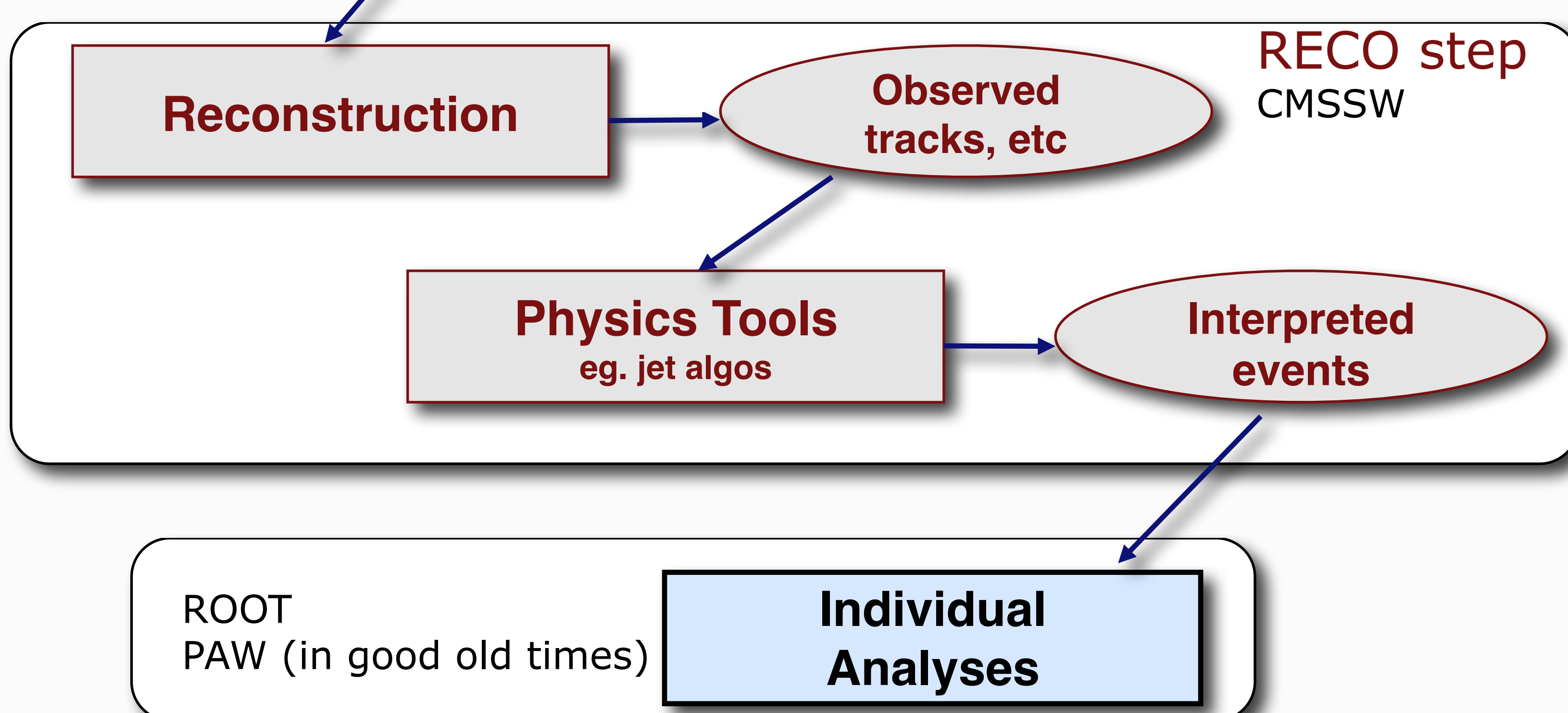
- flexibility,
  have different versions of pieces
- efficient for repeated studies
  don't have to start all over again if some improvement in later stage
- Manageability
  large programs, hard to build, understand, debug, maintain, ...

Partitioning production systems









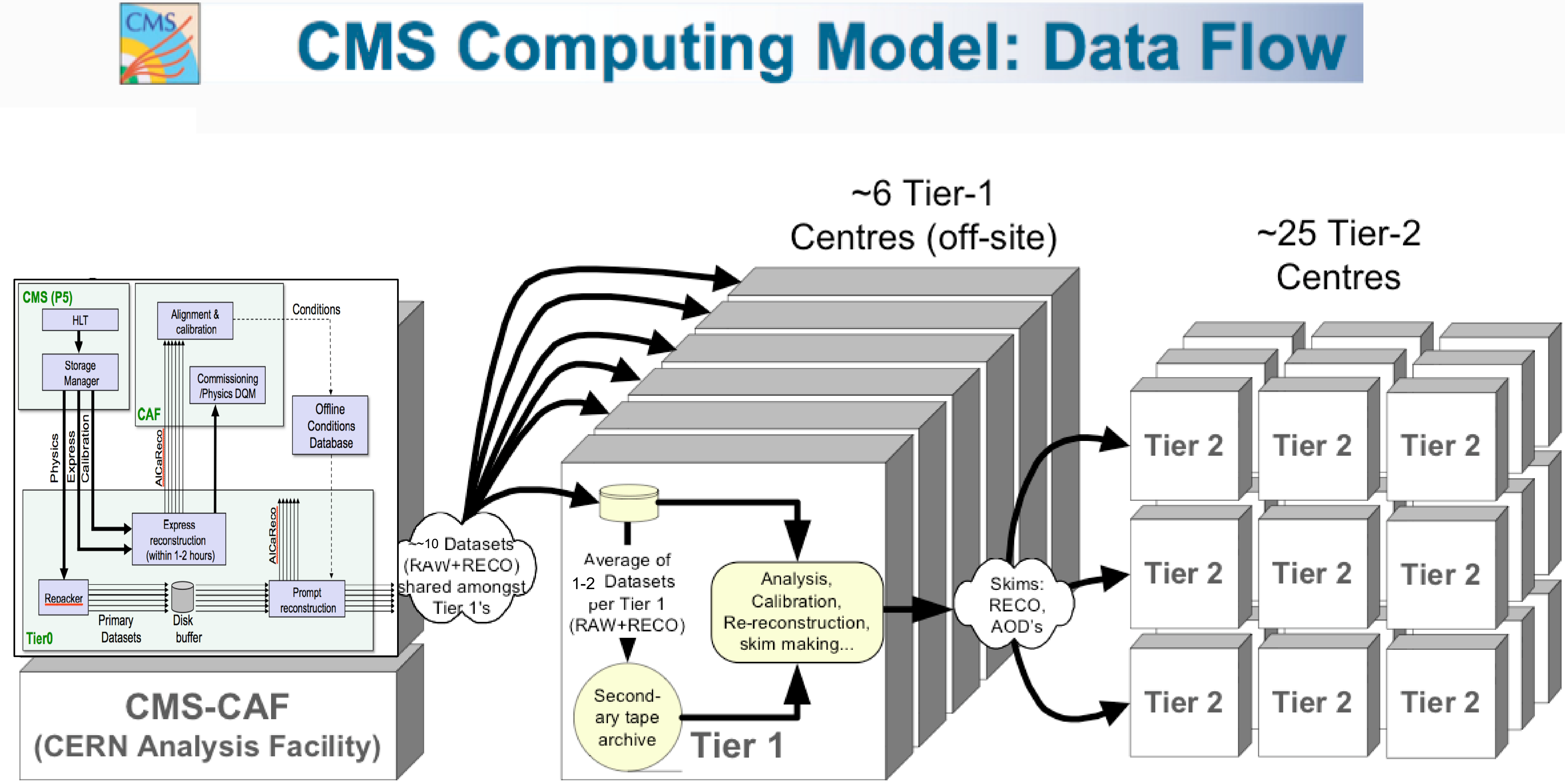
Partitioning

- there can be event stores between individual components

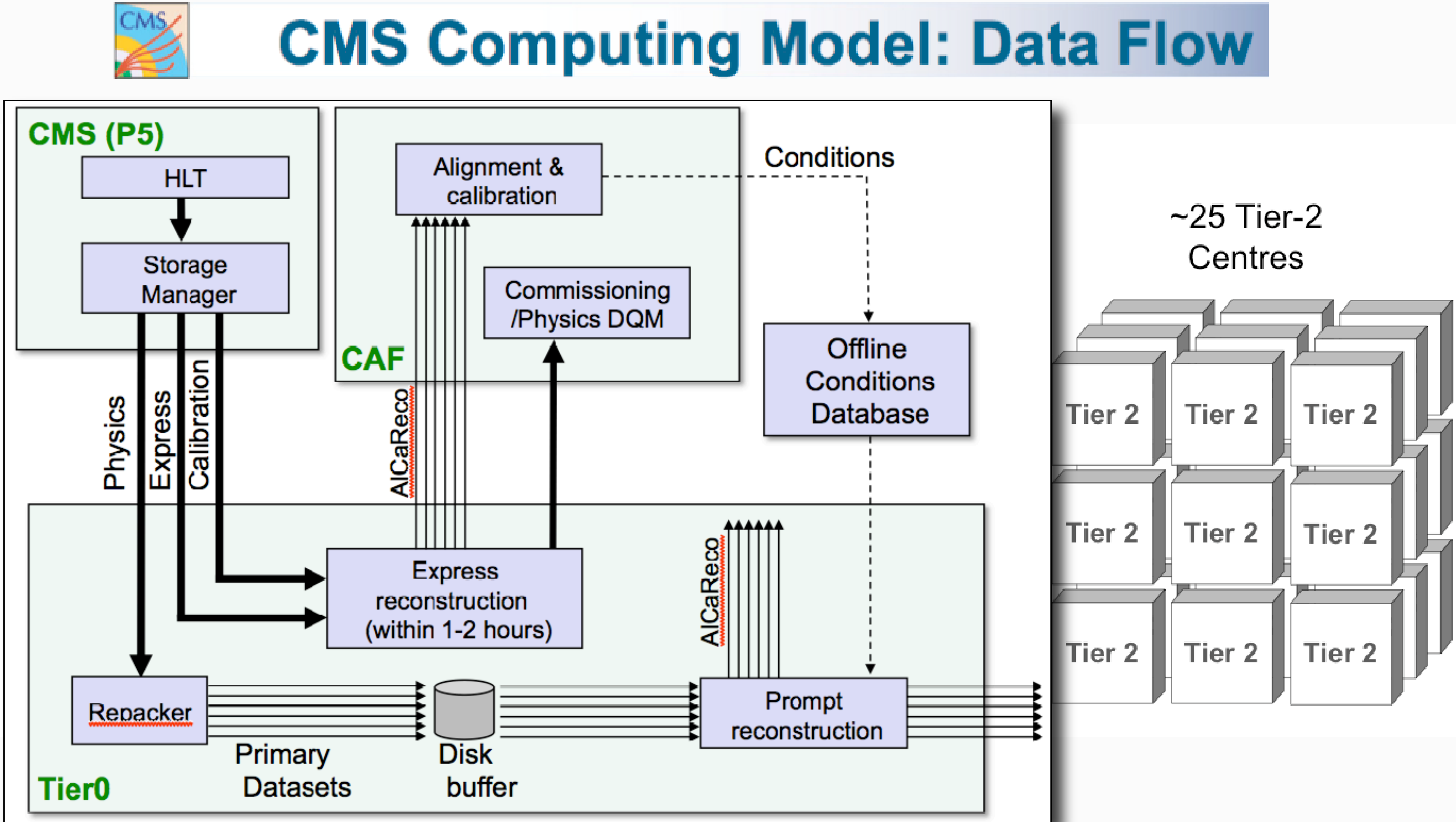
Why this structure

- flexibility,
  have different versions of pieces
- efficient for repeated studies
  don't have to start all over again if some improvement in later stage
- Manageability
  large programs, hard to build, understand, debug, maintain, ...

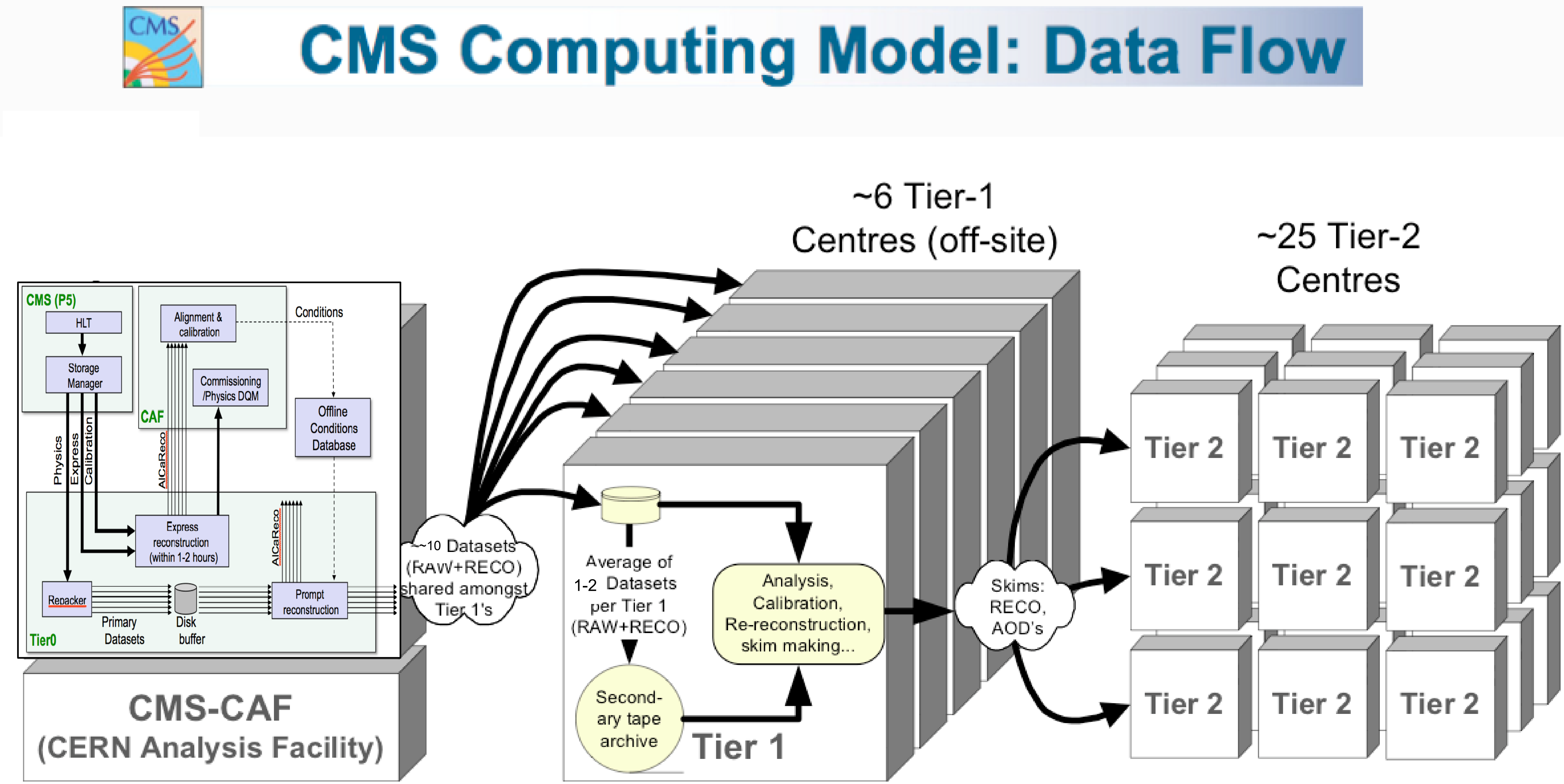
GRID distribution : example



GRID distribution : example



GRID distribution : example



Summary

"Doing something ordinary is a waste of time" (Madonna)

What wasn't covered

- Details on track fitting, Kalman filters
- Secondary Vertex finding
- Alignment
- Particle Identification
- Calibration techniques, "in-situ" methods
- Particle/Energy flow
- Trigger menus, their studies
- more details on parameter fitting,
  eg. lifetime and mass measurements
- how to estimate systematic errors
- Databases, persistent data storage
- Programming languages in use (F77, C, C++, JAVA, ...)
- .....

Conclusions

- Reconstruction and Analysis
  is how we get from raw data to physics papers

Conclusions

- Reconstruction and Analysis
  is how we get from raw data to physics papers

- On your way

Conclusions

ETH Institute for Particle Physics

- Reconstruction and Analysis
  is how we get from raw data to physics papers
- On your way
  - first you have too much information → reduce

Conclusions

ETH Institute for Particle Physics

- Reconstruction and Analysis
  is how we get from raw data to physics papers
- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses

Conclusions

ETH Institute for Particle Physics

- Reconstruction and Analysis
  is how we get from raw data to physics papers
- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses
- What makes it hard, but also exciting

Conclusions

- Reconstruction and Analysis
  is how we get from raw data to physics papers
- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses
- What makes it hard, but also exciting
  - many many cross checks

Conclusions

- Reconstruction and Analysis
  is how we get from raw data to physics papers

- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses

- What makes it hard, but also exciting
  - many many cross checks
  - more cross checks

Conclusions

- Reconstruction and Analysis
  is how we get from raw data to physics papers

- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses

- What makes it hard, but also exciting
  - many many cross checks
  - more cross checks
  - sometimes some “art” involved

Conclusions

ETH Institute for Particle Physics

- Reconstruction and Analysis
  is how we get from raw data to physics papers

- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses

- What makes it hard, but also exciting
  - many many cross checks
  - more cross checks
  - sometimes some “art” involved
  - tuning, evolutionary improvement

Conclusions

- Reconstruction and Analysis
  is how we get from raw data to physics papers
- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses
- What makes it hard, but also exciting
  - many many cross checks
  - more cross checks
  - sometimes some “art” involved
  - tuning, evolutionary improvement
- Even to me it is often a miracle that we can generate wonderful results from these complicated instruments!

Conclusions

ETH Institute for Particle Physics

- Reconstruction and Analysis
  is how we get from raw data to physics papers

- On your way
  - first you have too much information → reduce
  - sometimes too little information or little prior knowledge
    - make hypotheses

- What makes it hard, but also exciting
  - many many cross checks
  - more cross checks
  - sometimes some “art” involved
  - tuning, evolutionary improvement

- Even to me it is often a miracle that we can generate wonderful results from these complicated instruments!