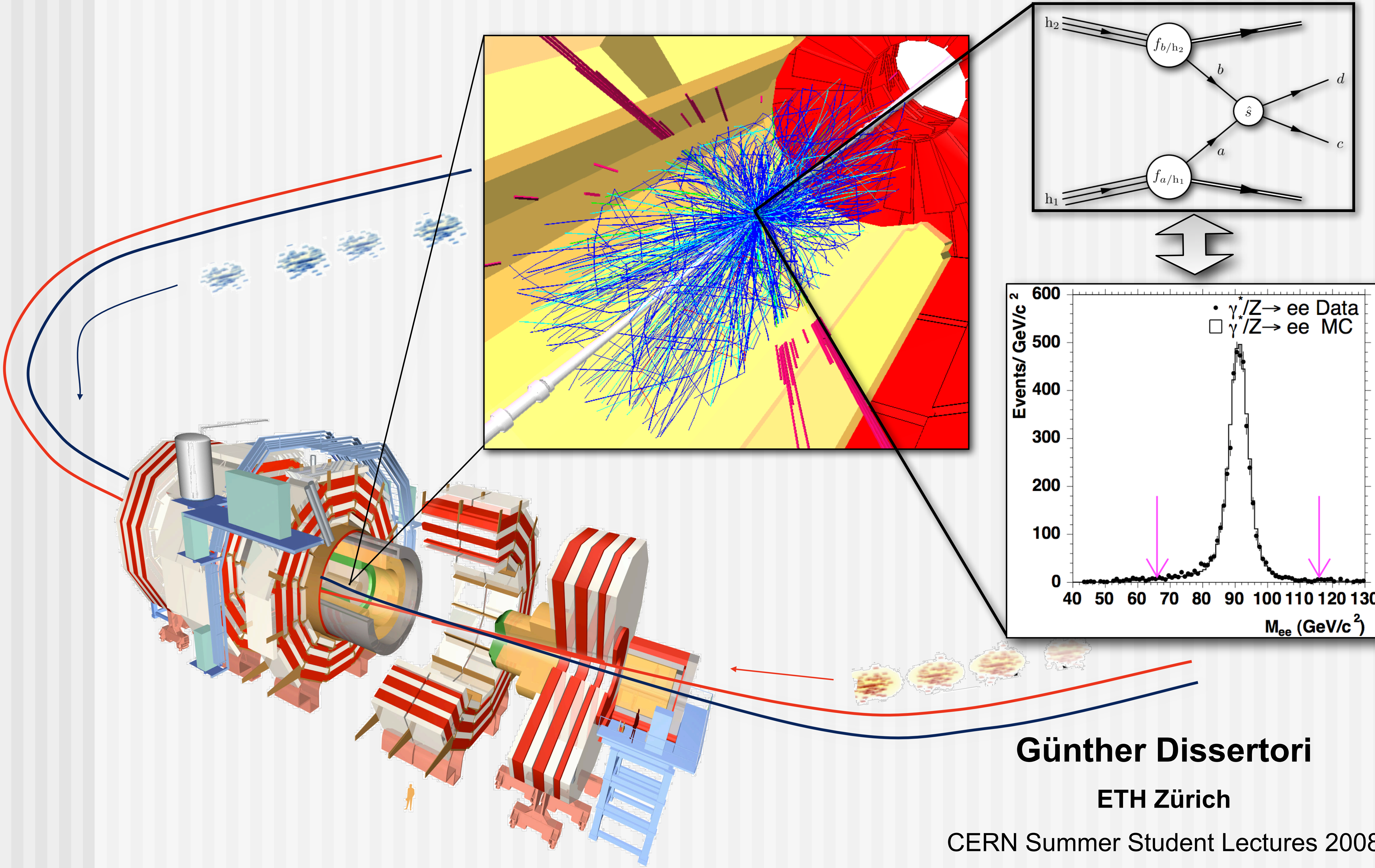
From Raw Data to Physics Results



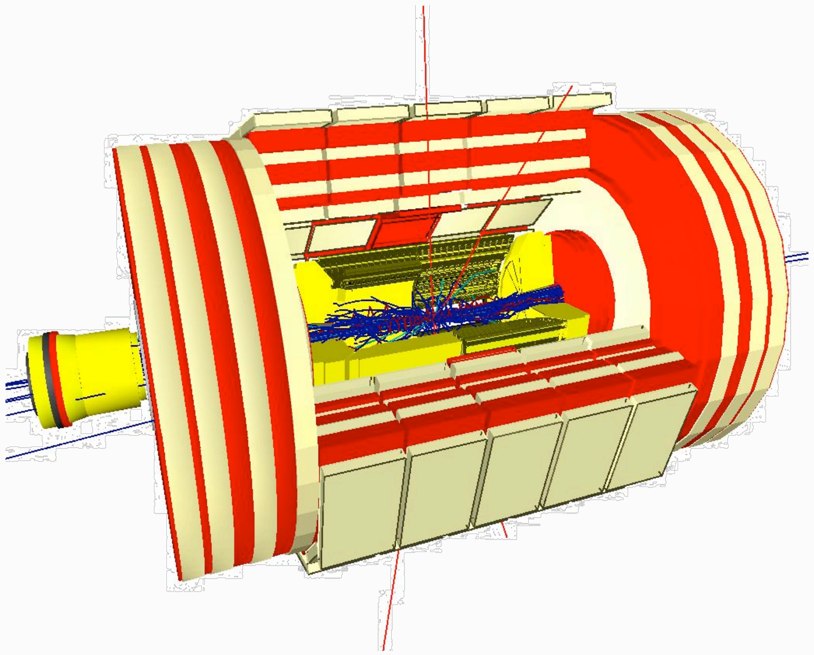
Günther Dissertori

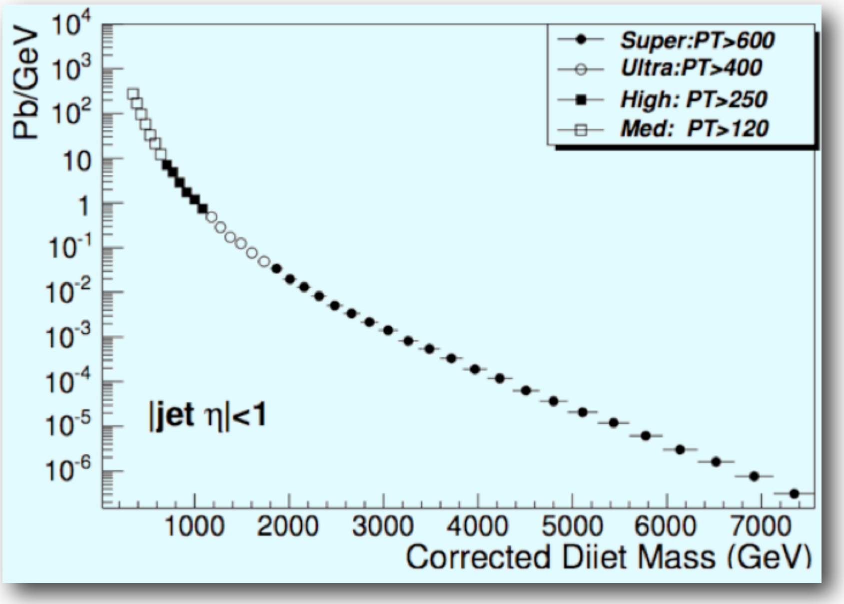
ETH Zürich

CERN Summer Student Lectures 2008

Outline

- “Executive Summary”
  - The whole story in a nutshell
- Some more details
  - Introduction
  - A simple example : Z decays
  - A more complicated example : Jets
- How is it done in practice?
  - Track and Calorimeter energy reconstruction
  - High-level algorithms : Jets
  - The computing part
- Summary



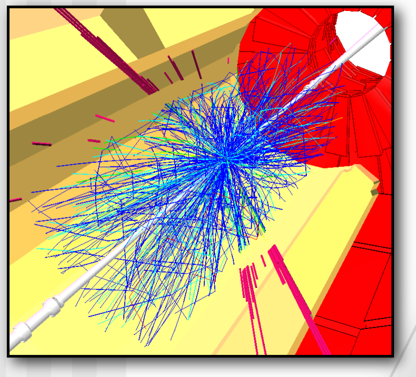


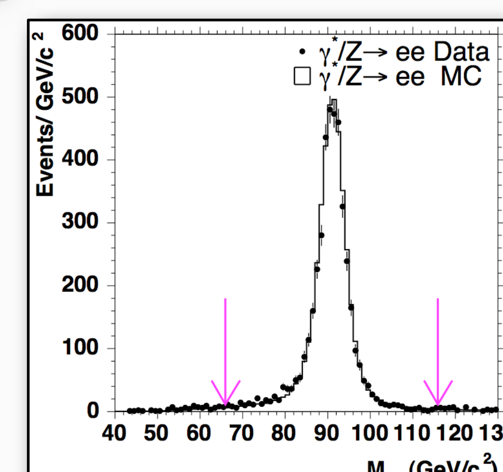
- Disclaimer : Several slides based on past CSS lectures by B. Jacobsen
- thanks also to J. Weng, T. Punz, A. Valassi

The whole lecture in a nutshell

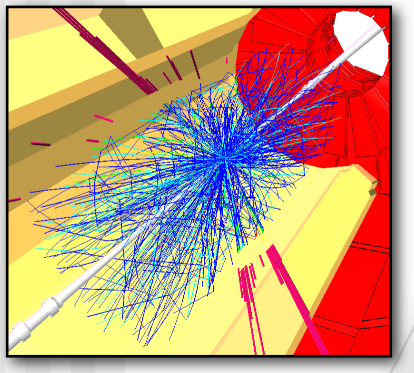
“But you should not leave immediately after this....”

Data Analysis Chain

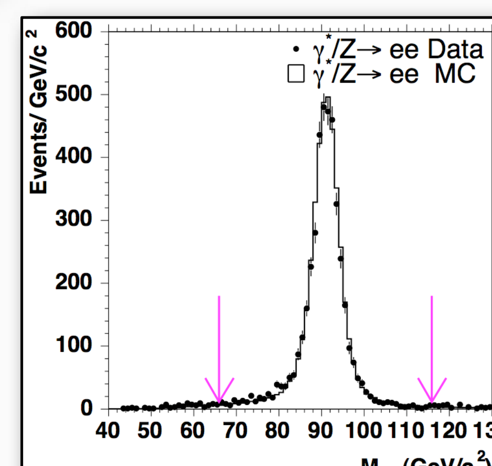




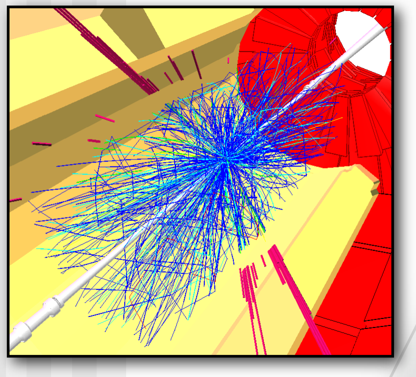
Data Analysis Chain



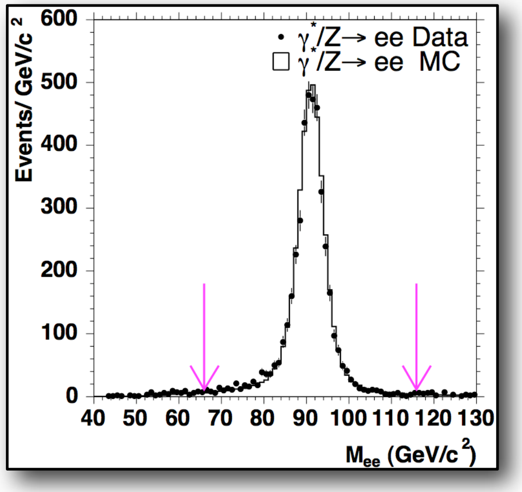
- Have to collect data from many channels on many sub-detectors (millions)



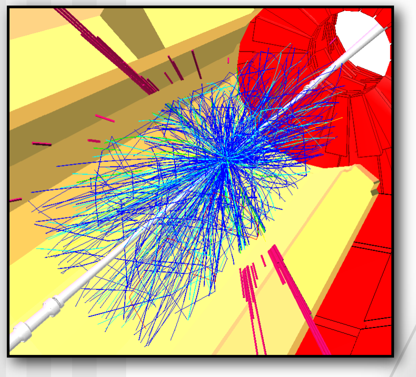
Data Analysis Chain



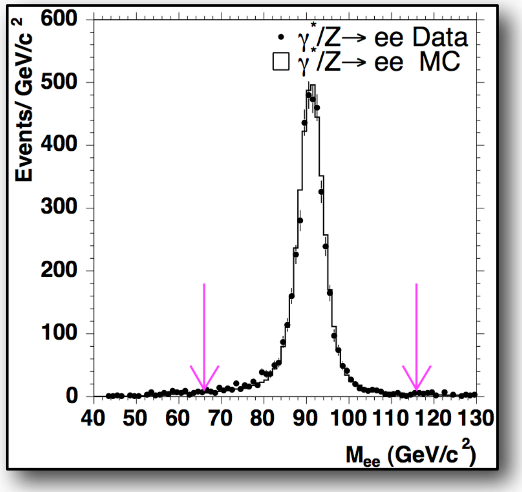
- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)



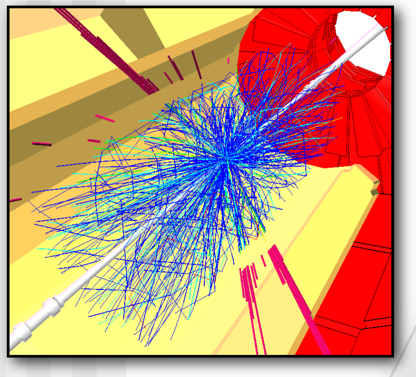
Data Analysis Chain



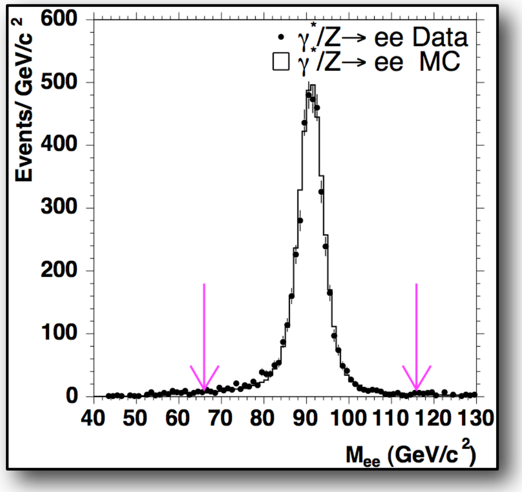
- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)
- Build the event (put info together)



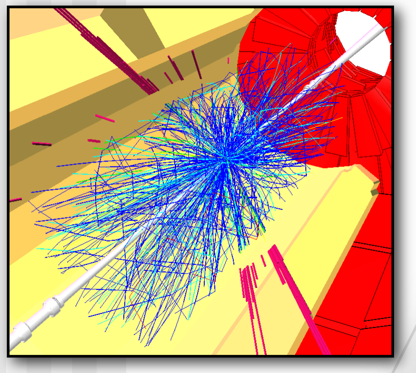
Data Analysis Chain



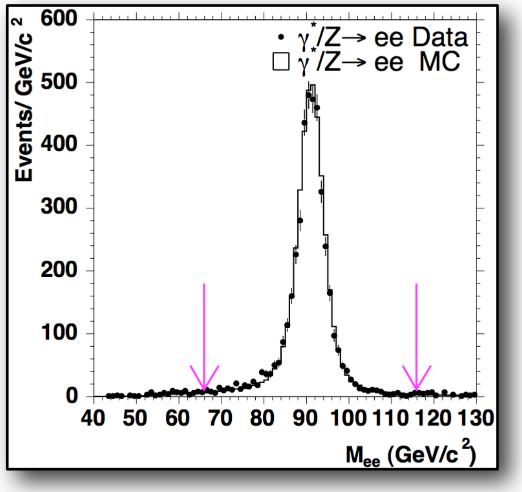
- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)
- Build the event (put info together)
- Store the data



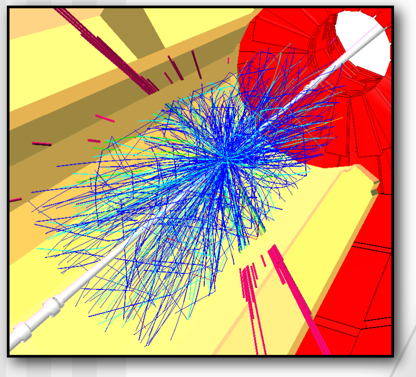
Data Analysis Chain



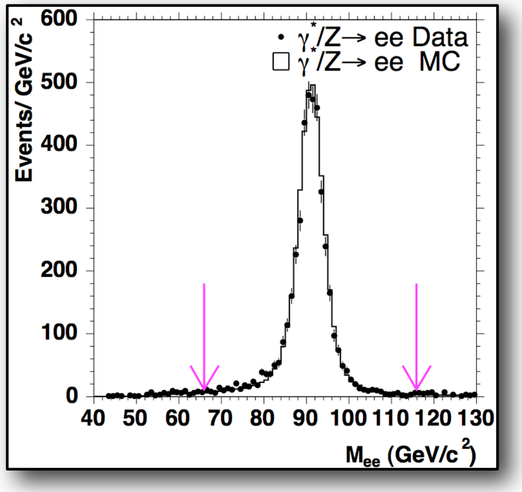
- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)
- Build the event (put info together)
- Store the data
- Analyze them
  - reconstruction, user analysis algorithms, data volume reduction



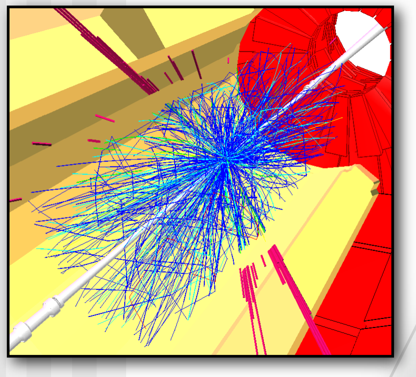
Data Analysis Chain



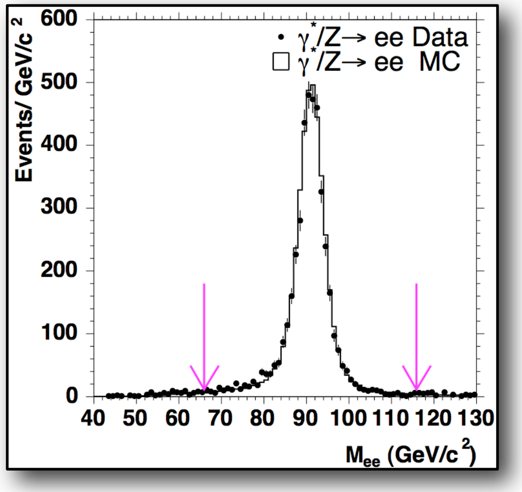
- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)
- Build the event (put info together)
- Store the data
- Analyze them
  - reconstruction, user analysis algorithms, data volume reduction
- do the same with a simulation
  - correct data for detector effects



Data Analysis Chain



- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)
- Build the event (put info together)
- Store the data
- Analyze them
  - reconstruction, user analysis algorithms, data volume reduction
- do the same with a simulation
  - correct data for detector effects
- Compare data and theory

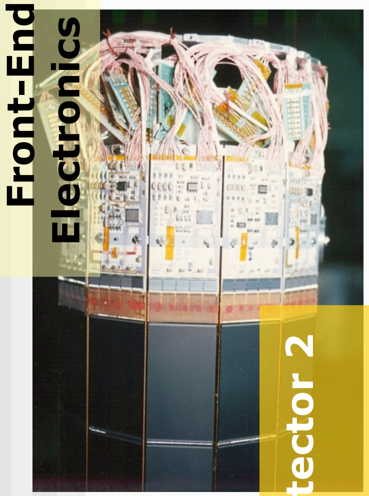


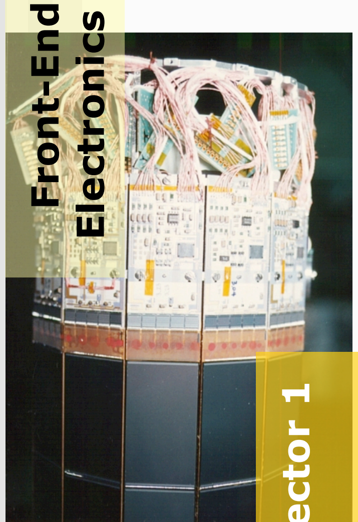
DAQ chain (see lectures by N. Neufeld)

DAQ chain (see lectures by N. Neufeld)

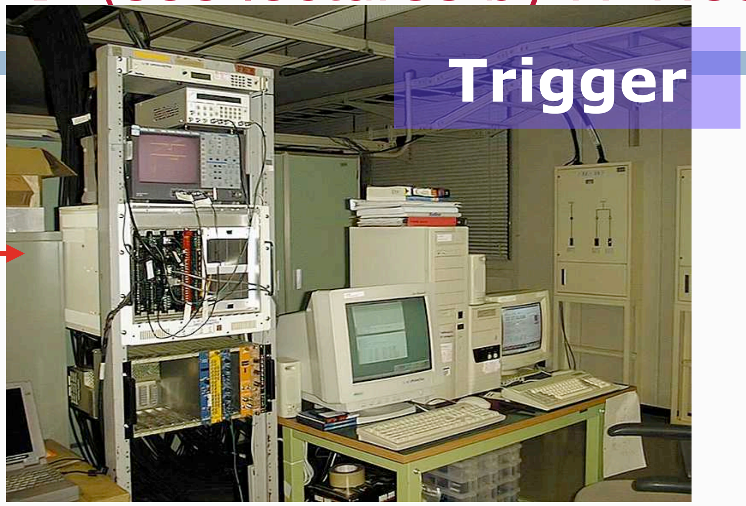
Detector Front-End

⋮



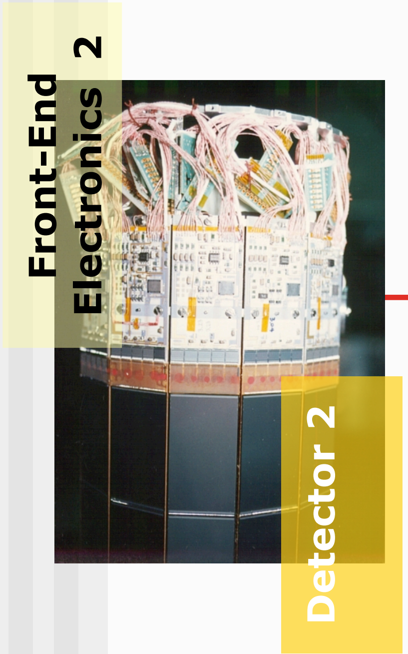


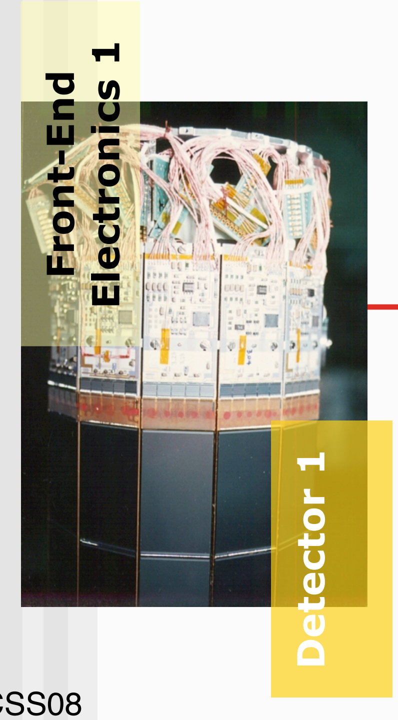
DAQ chain (see lectures by N. Neufeld)



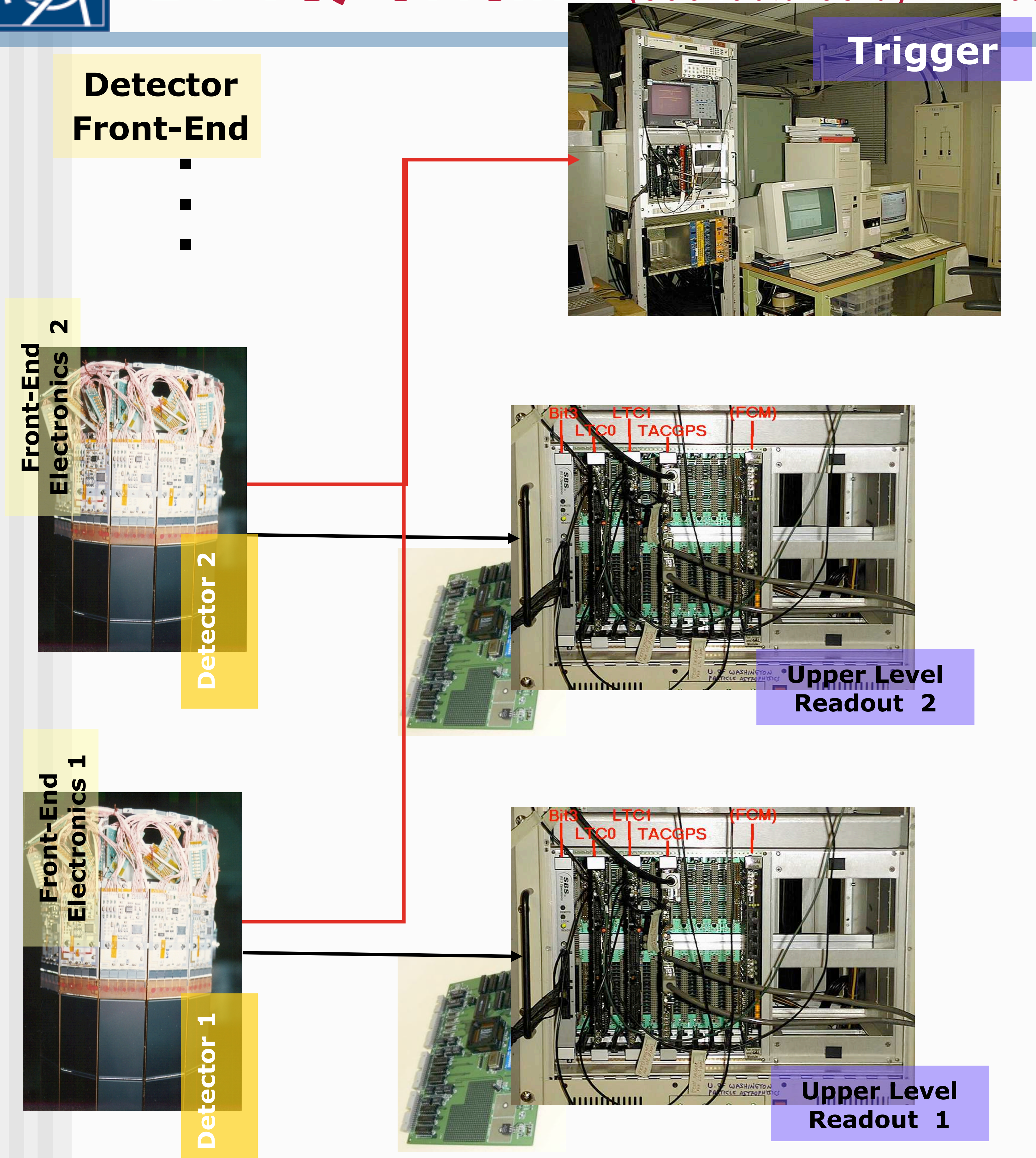
Detector Front-End

.
.
.

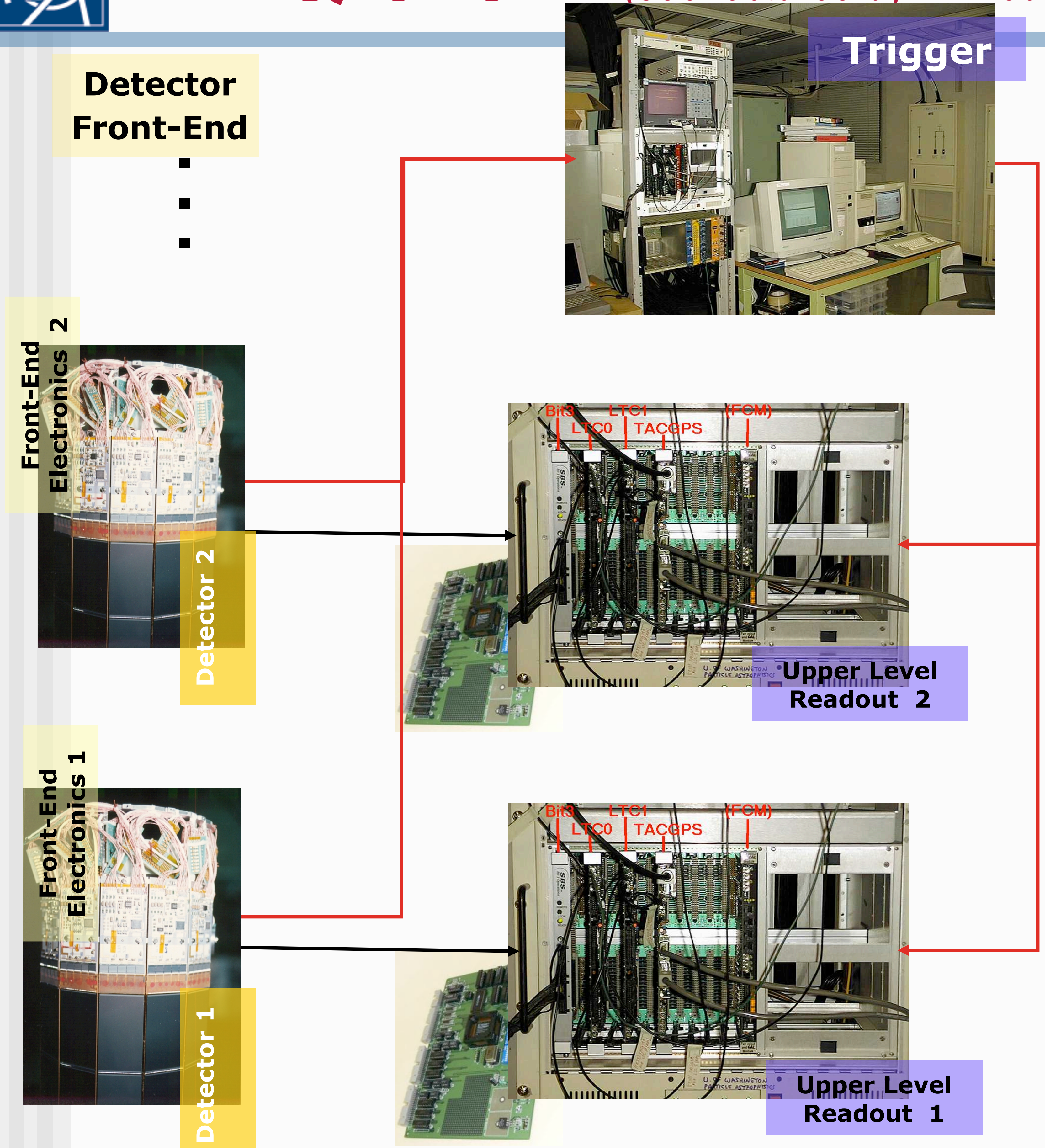




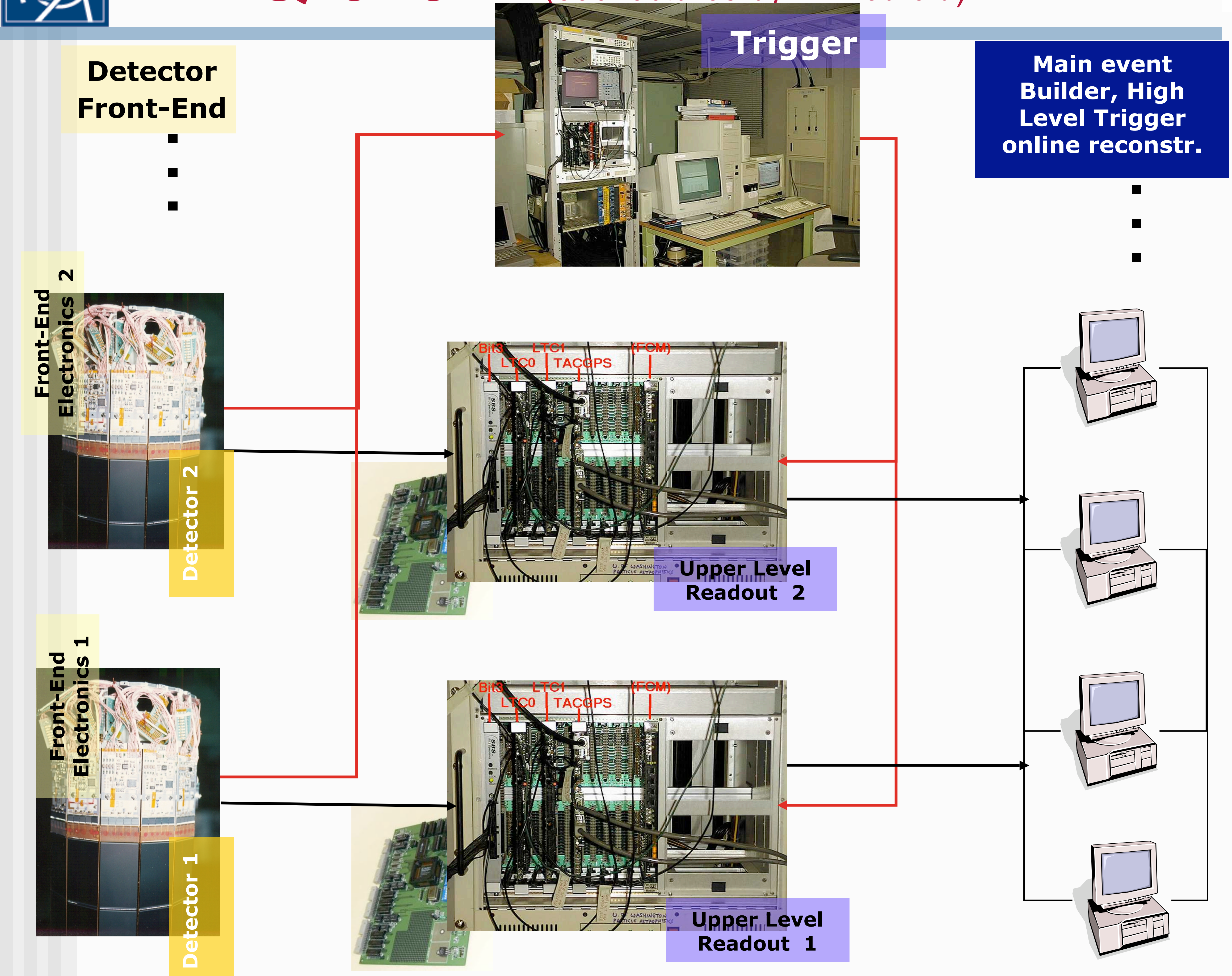
DAQ chain (see lectures by N. Neufeld)



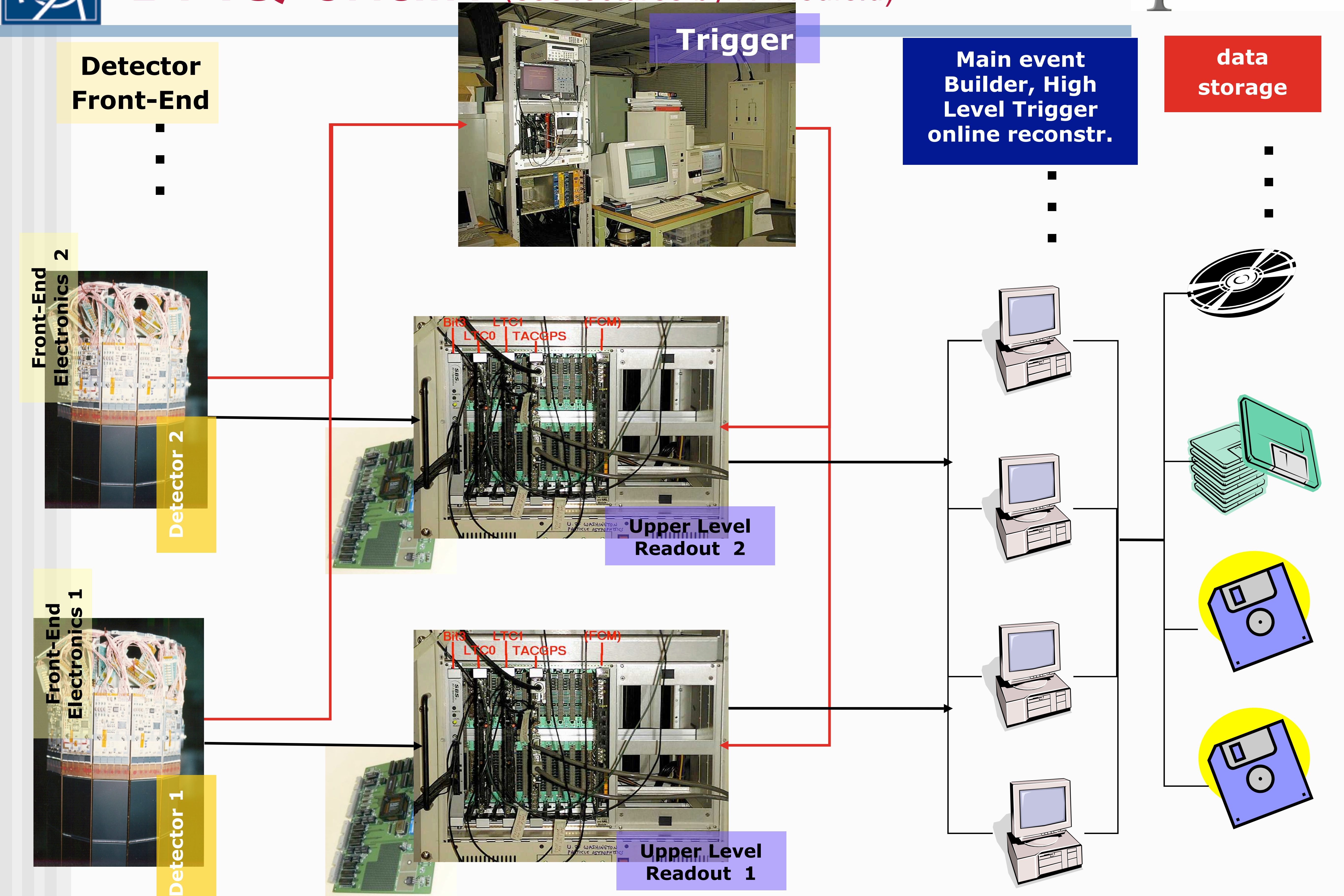
DAQ chain (see lectures by N. Neufeld)



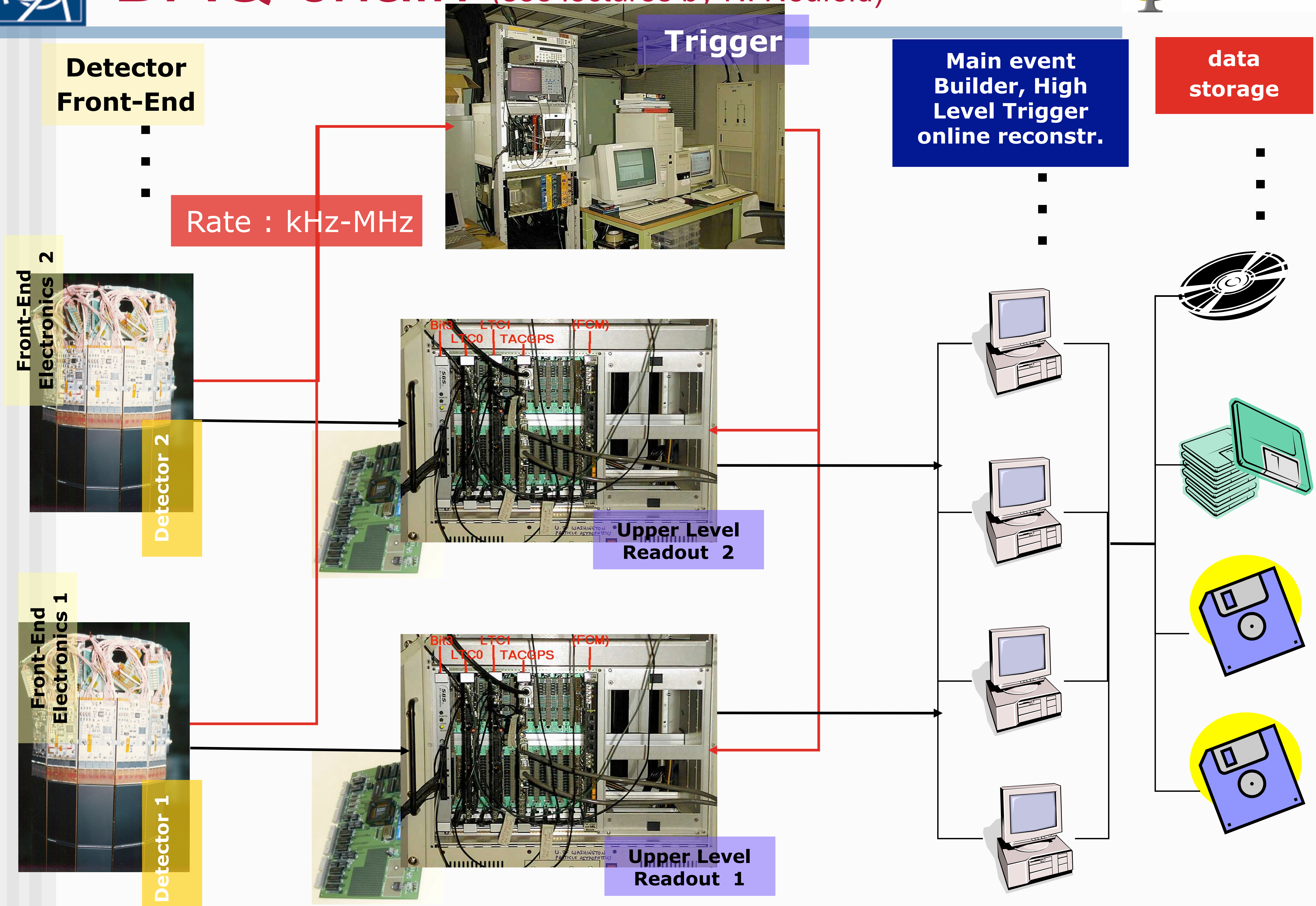
DAQ chain (see lectures by N. Neufeld)



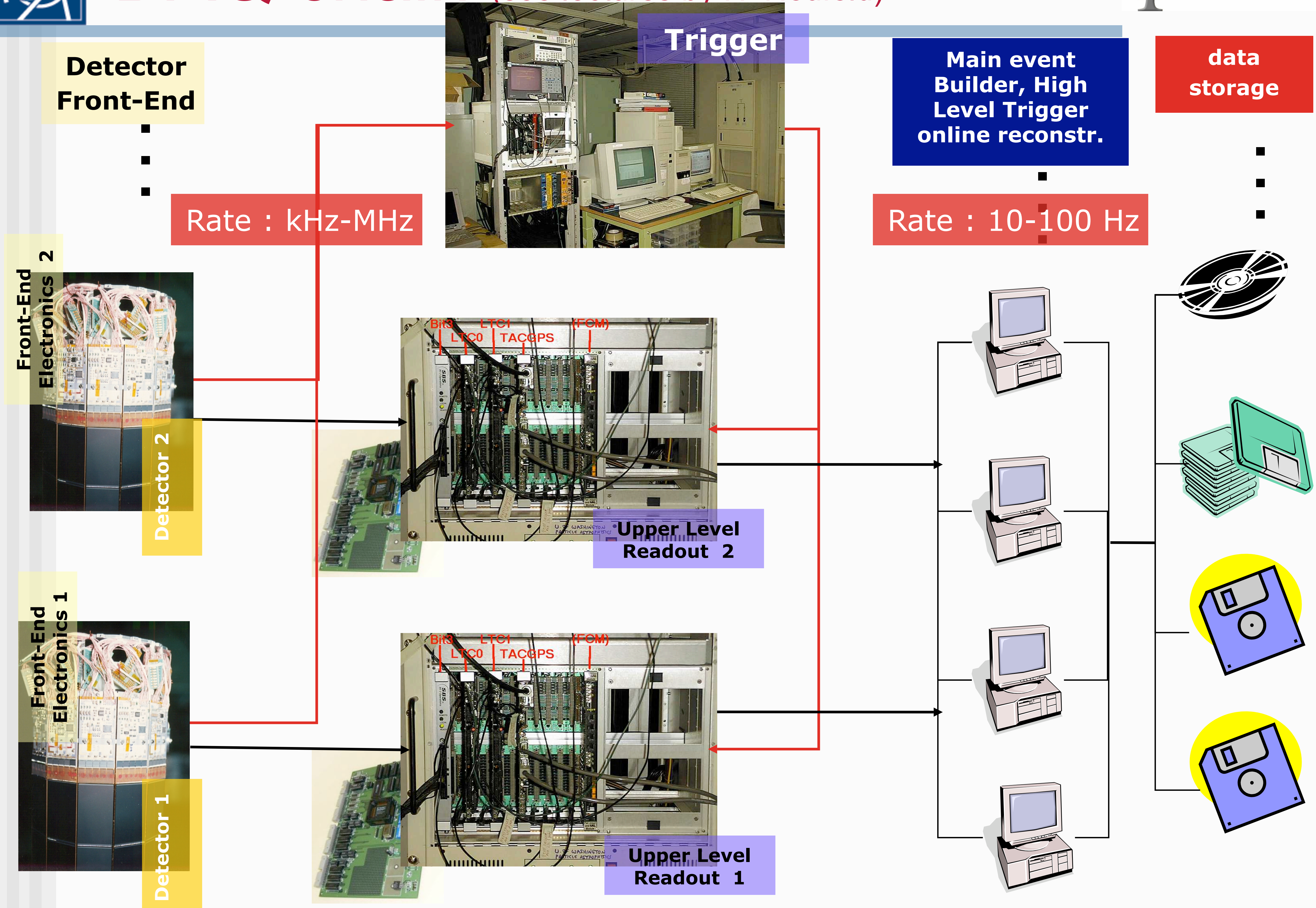
DAQ chain (see lectures by N. Neufeld)



DAQ chain (see lectures by N. Neufeld)



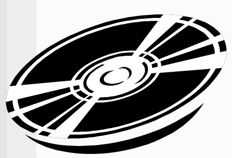
DAQ chain (see lectures by N. Neufeld)



Offline Analysis Chain

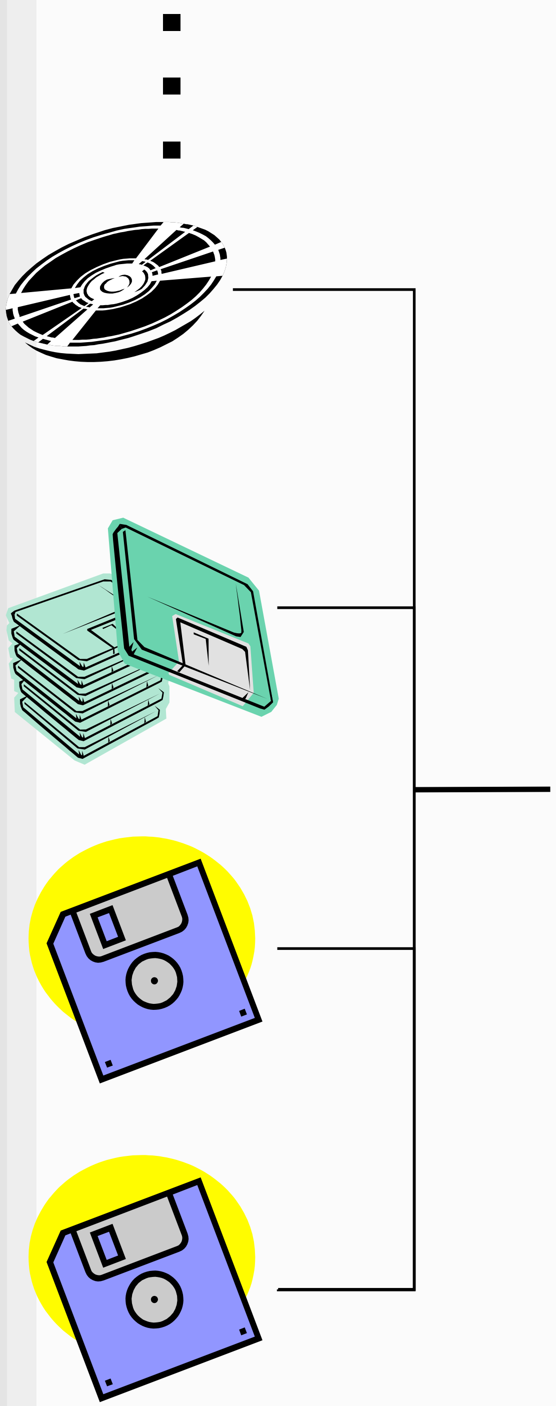
Offline Analysis Chain

data storage



Offline Analysis Chain

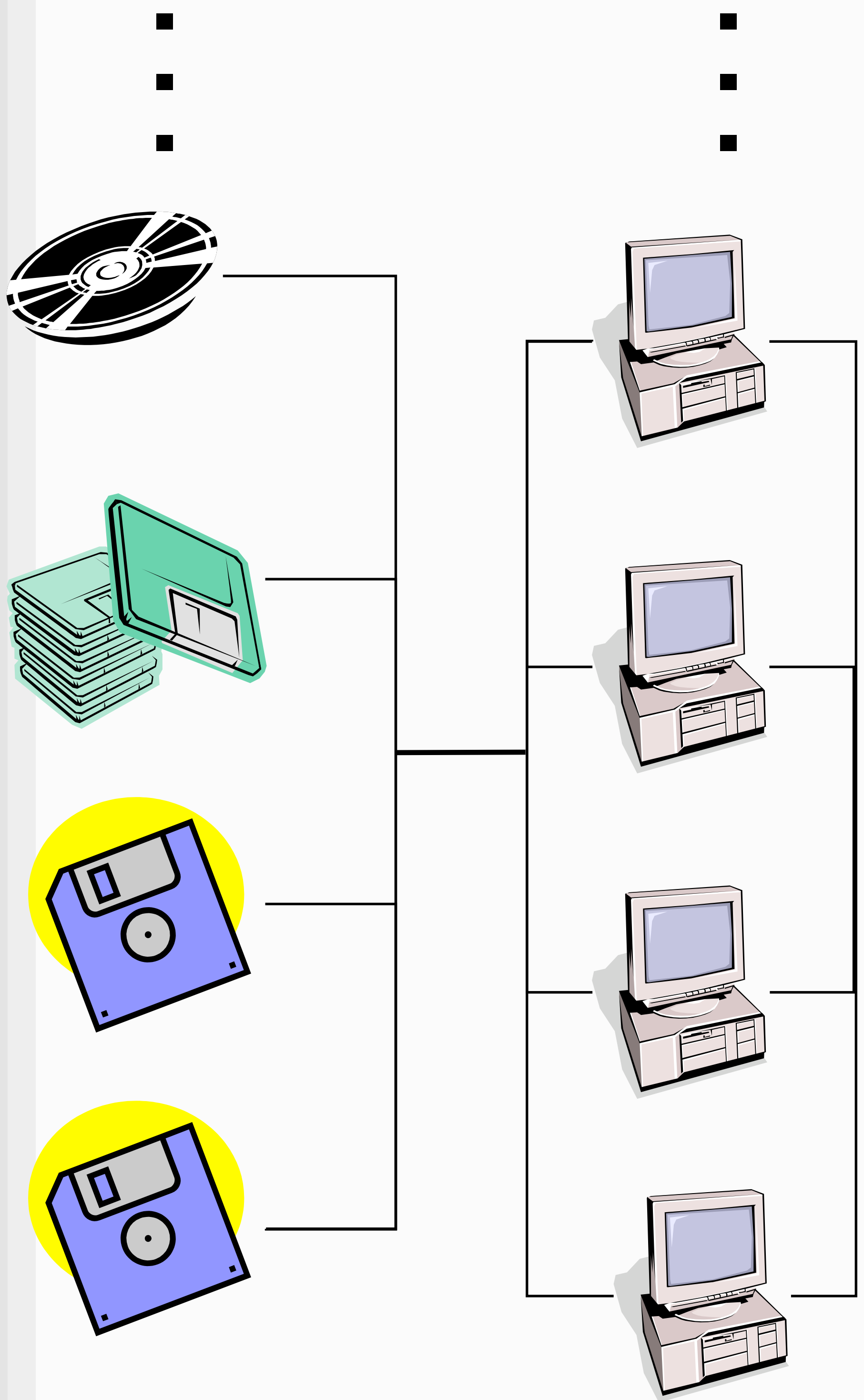
data storage



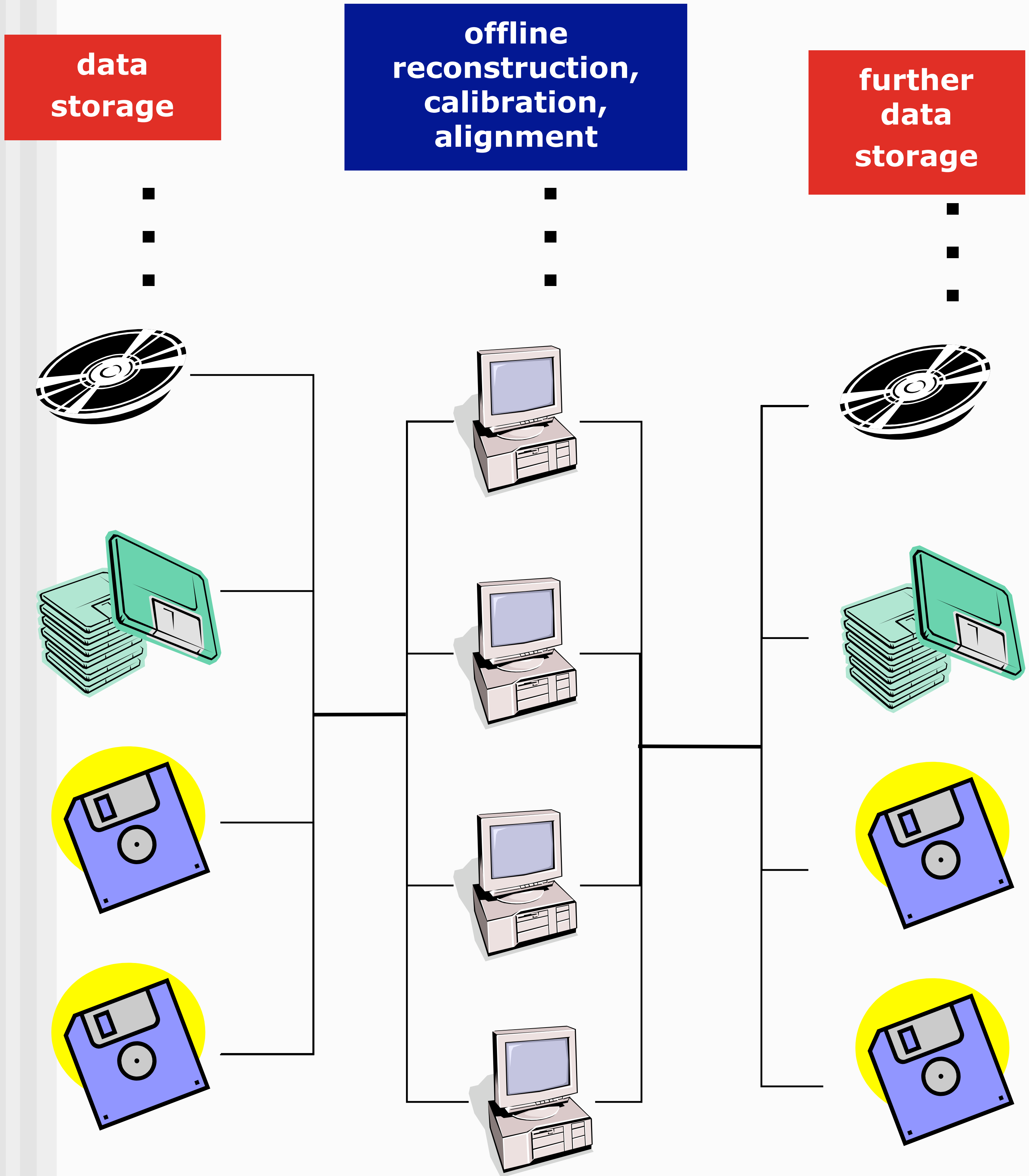
Offline Analysis Chain

data storage

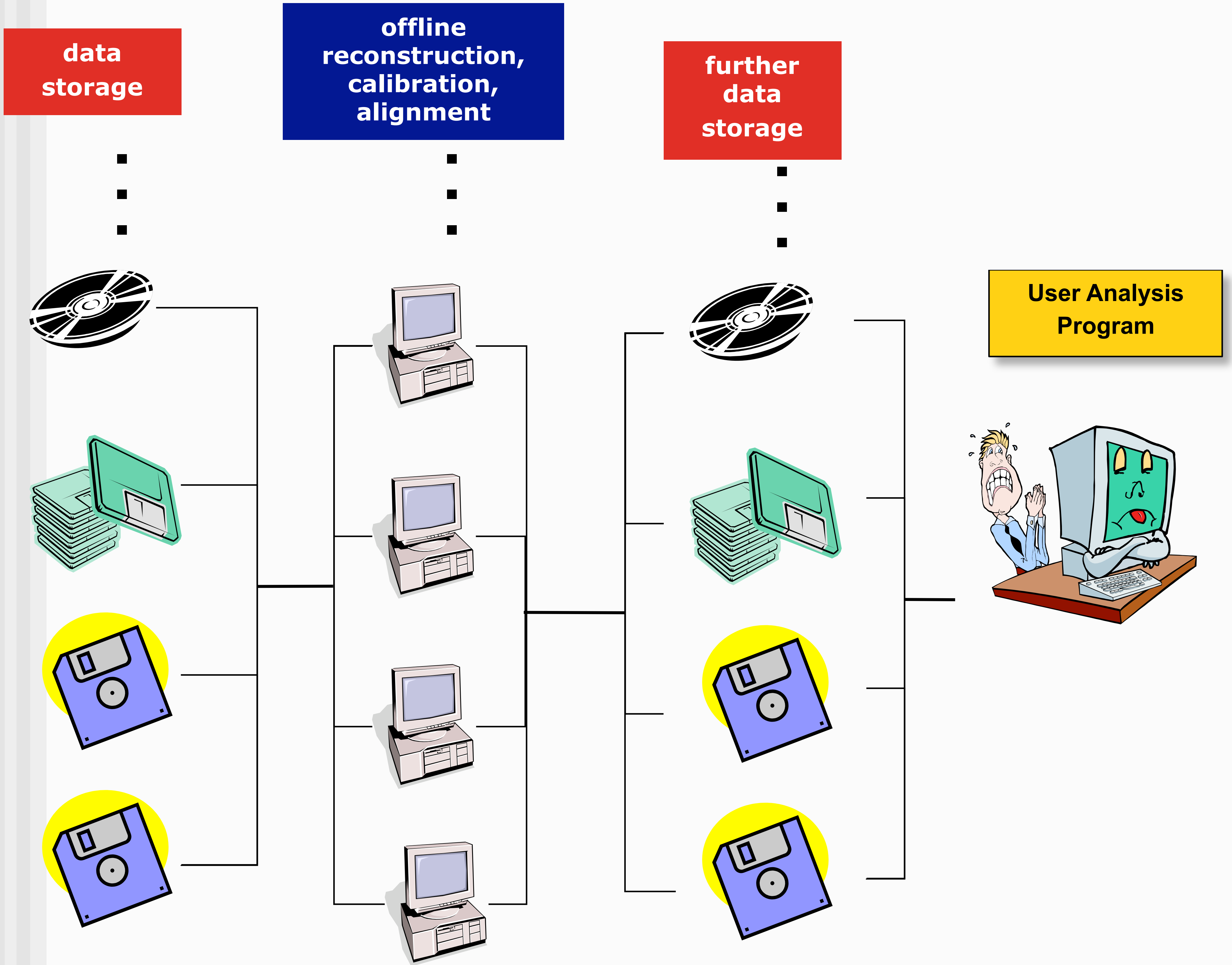
offline reconstruction, calibration, alignment



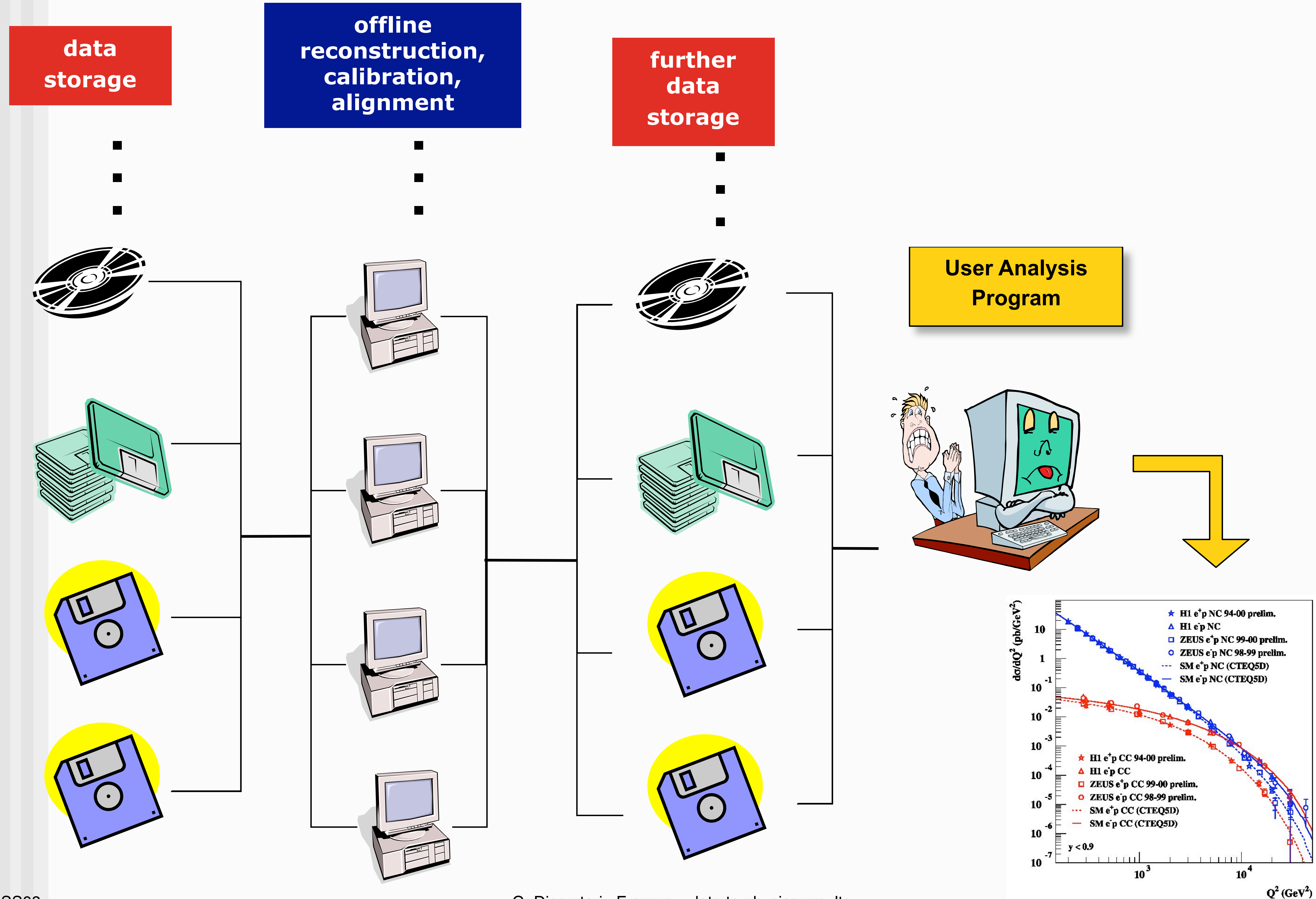
Offline Analysis Chain



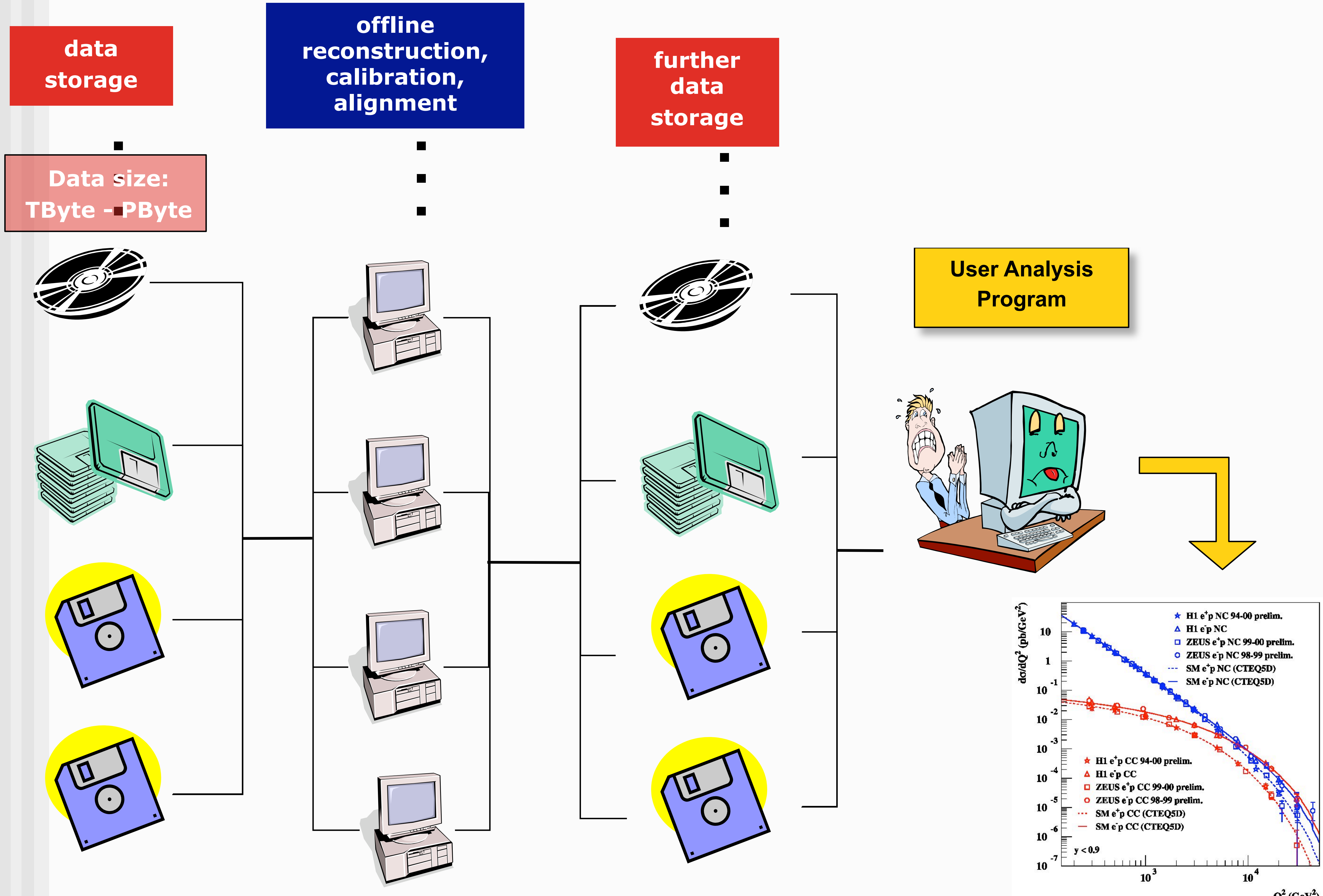
Offline Analysis Chain



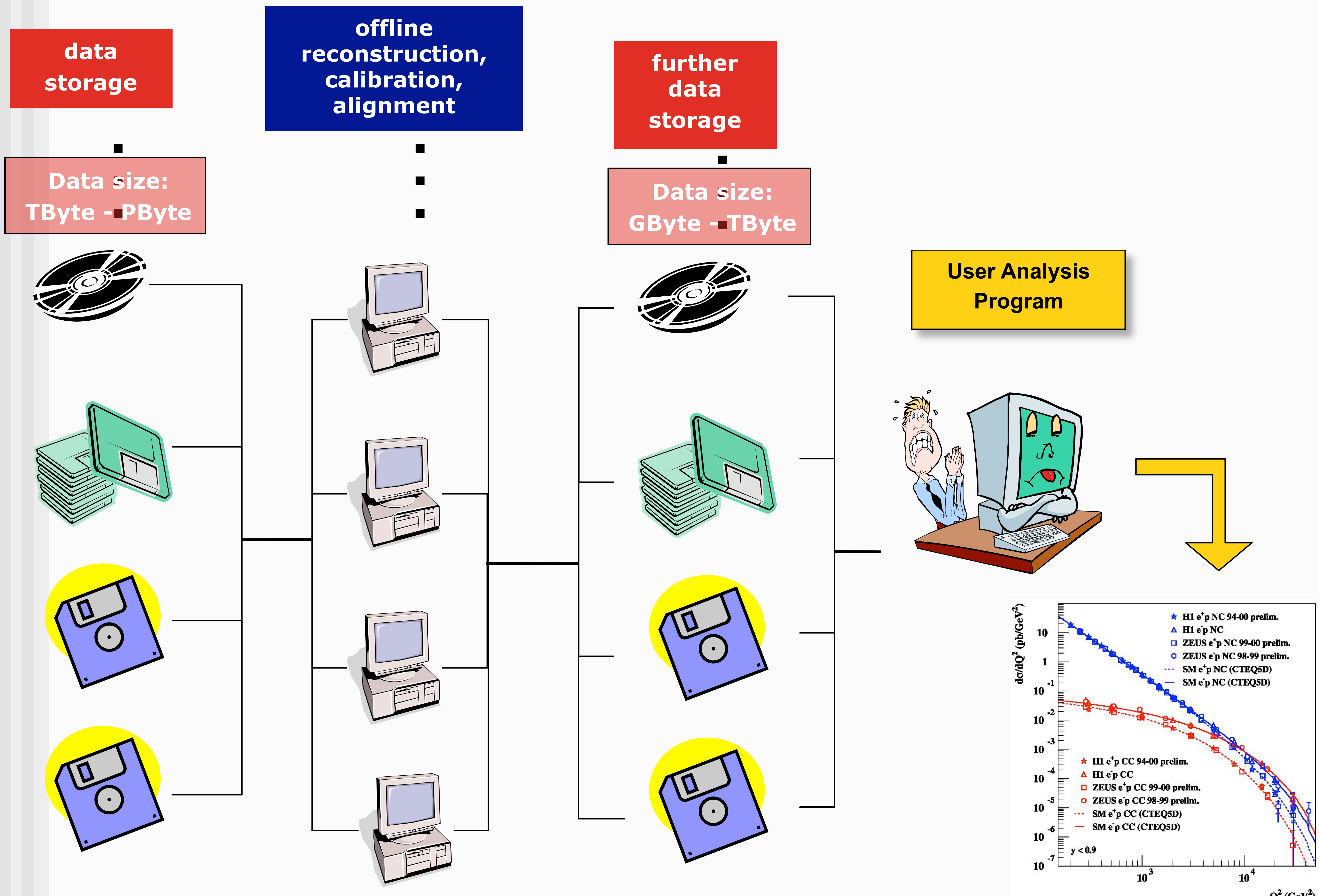
Offline Analysis Chain



Offline Analysis Chain



Offline Analysis Chain



Offline Analysis Chain

ETH Institute for Particle Physics

data storage

Data size: TByte - PByte

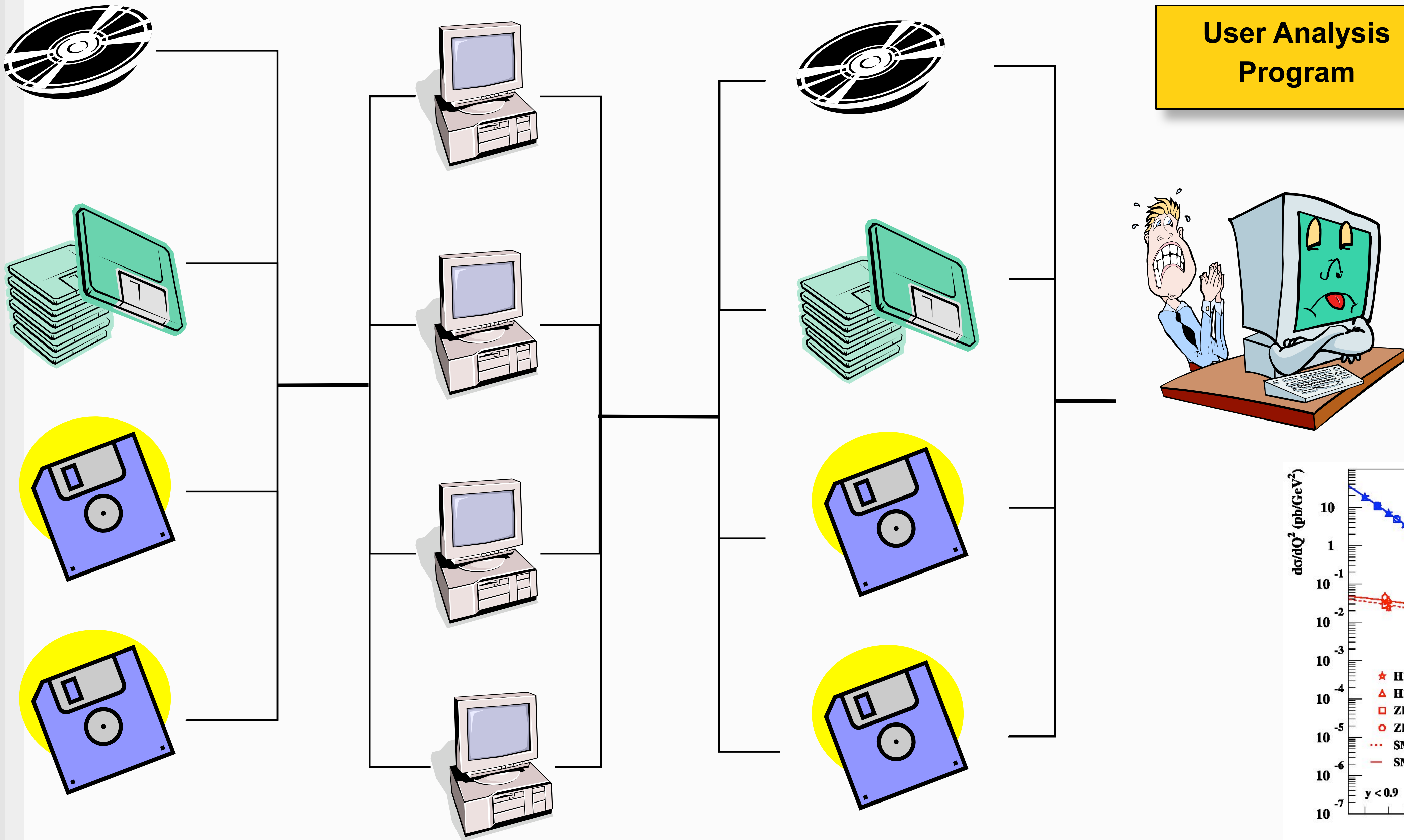
offline reconstruction, calibration, alignment

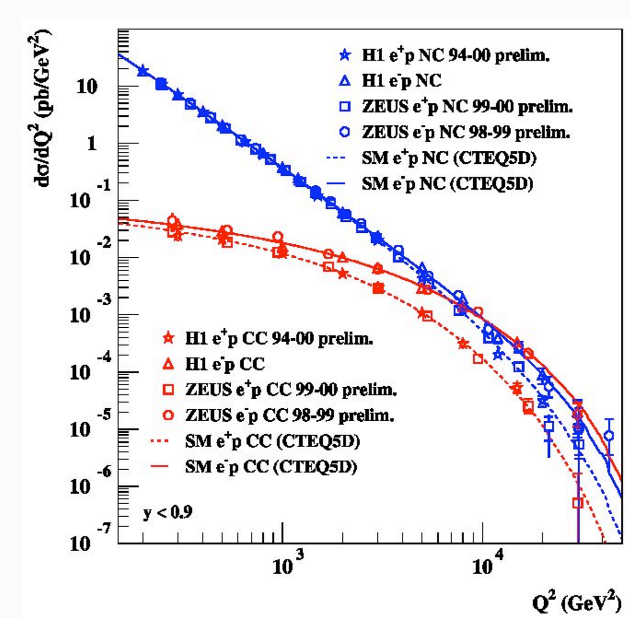
further data storage

Data size: GByte - TByte

User Analysis Program

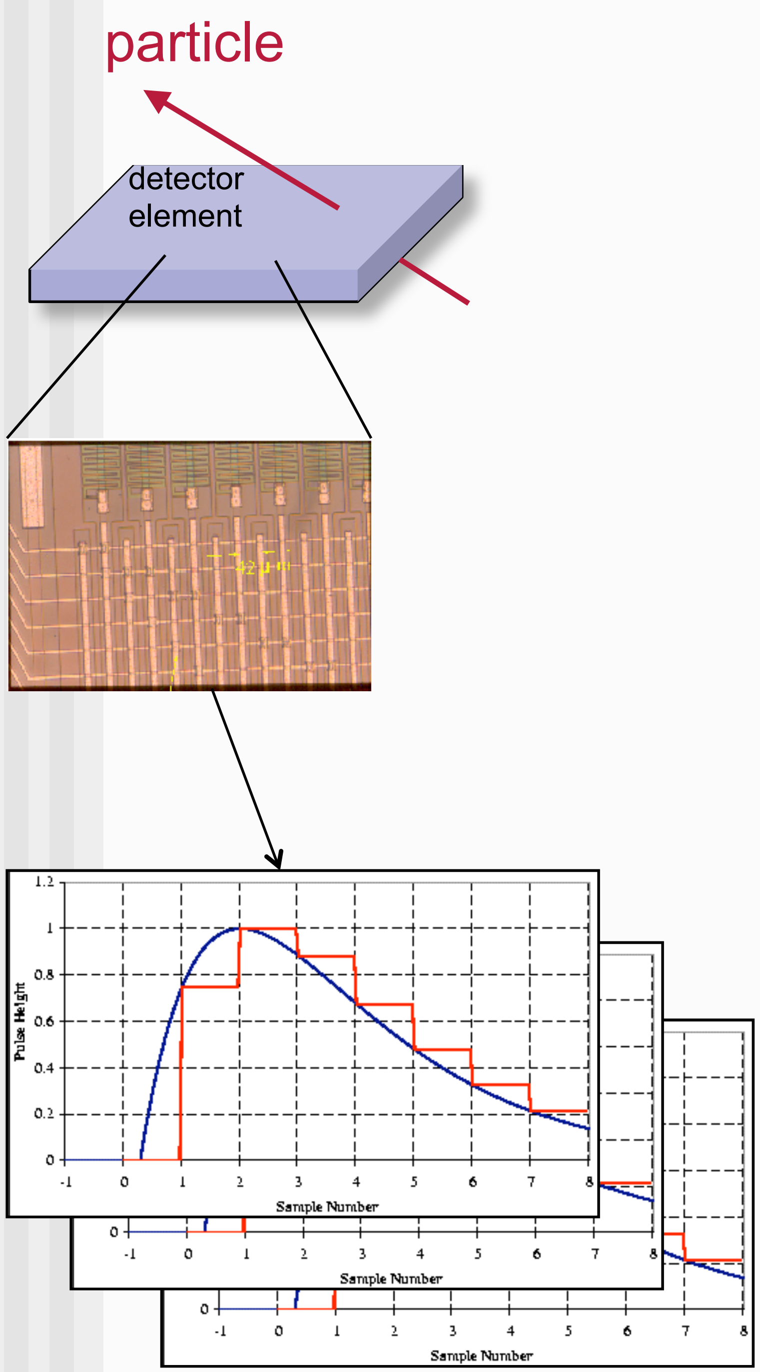
Data size: kByte





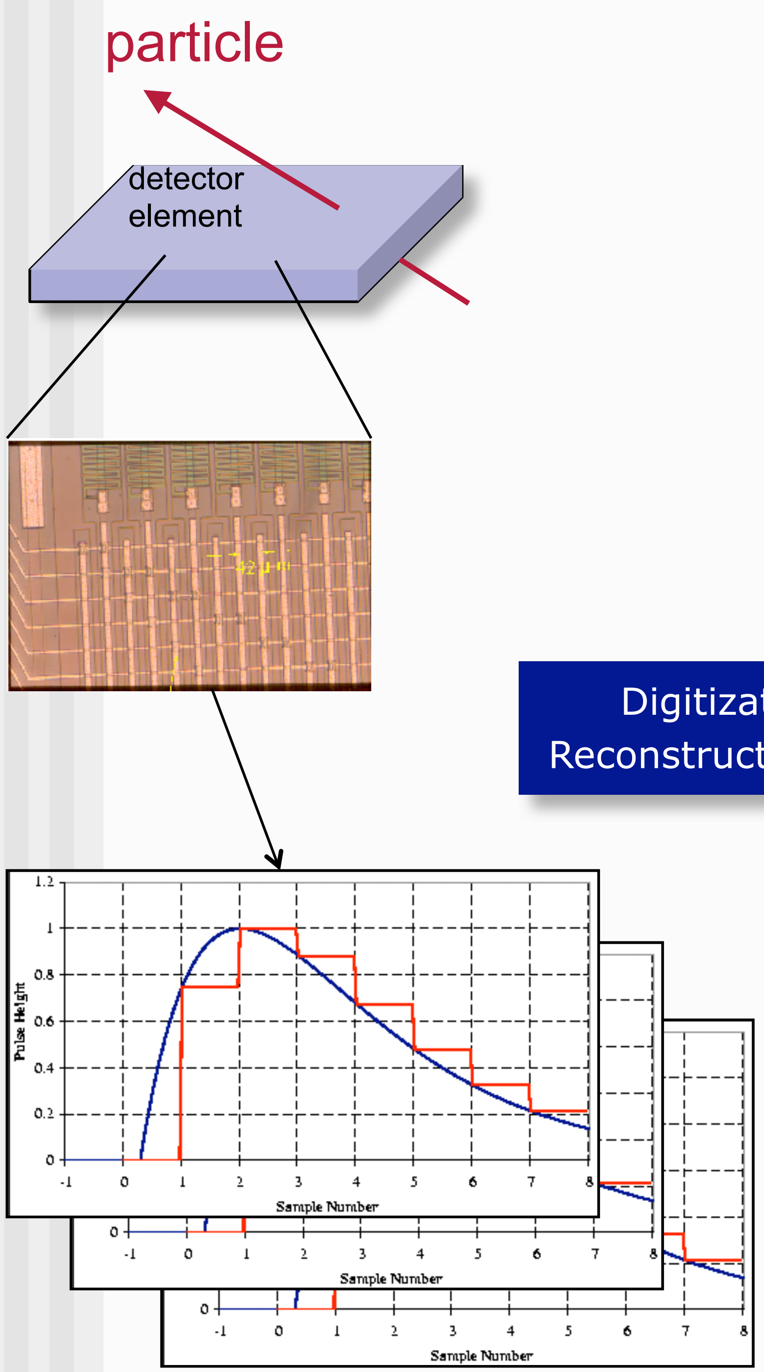
Data reduction/abstraction

Data reduction/abstraction



Analog signals

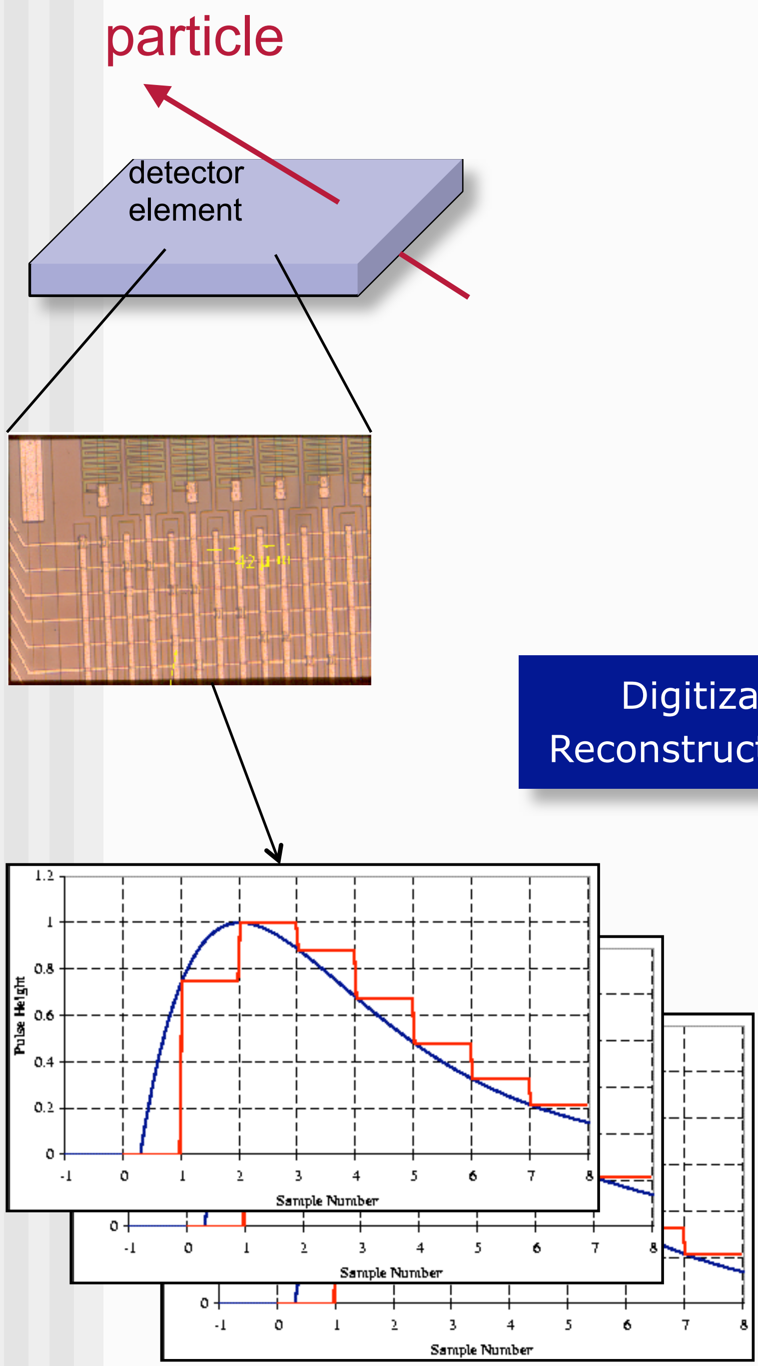
Data reduction/abstraction



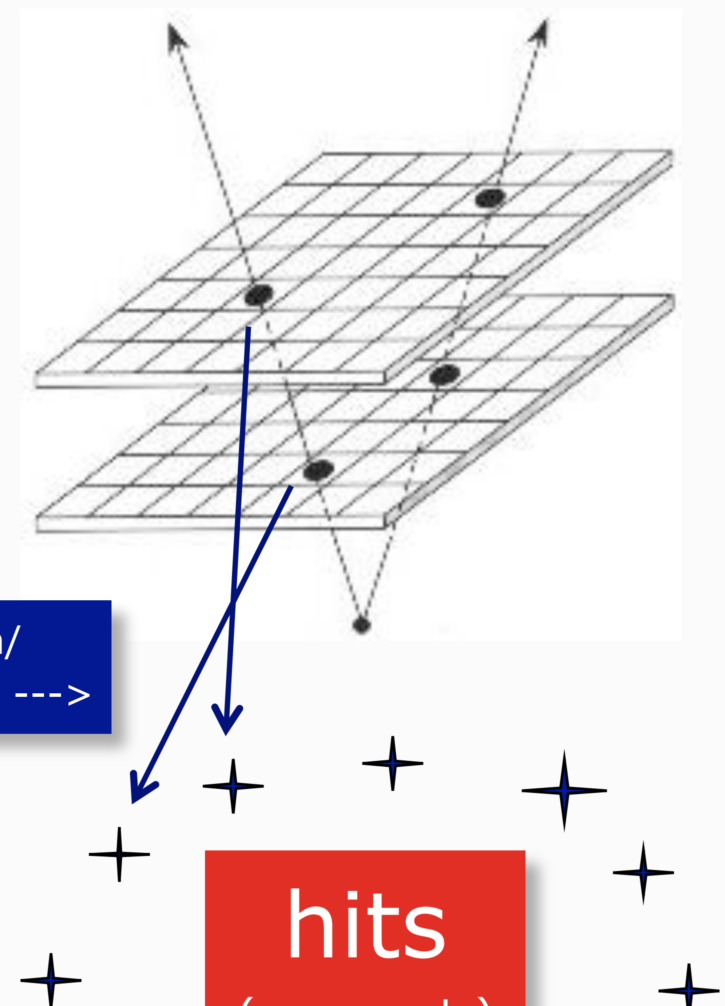
Digitization/
Reconstruction --->

Analog
signals

Data reduction/abstraction



Digitization/
Reconstruction --->



hits

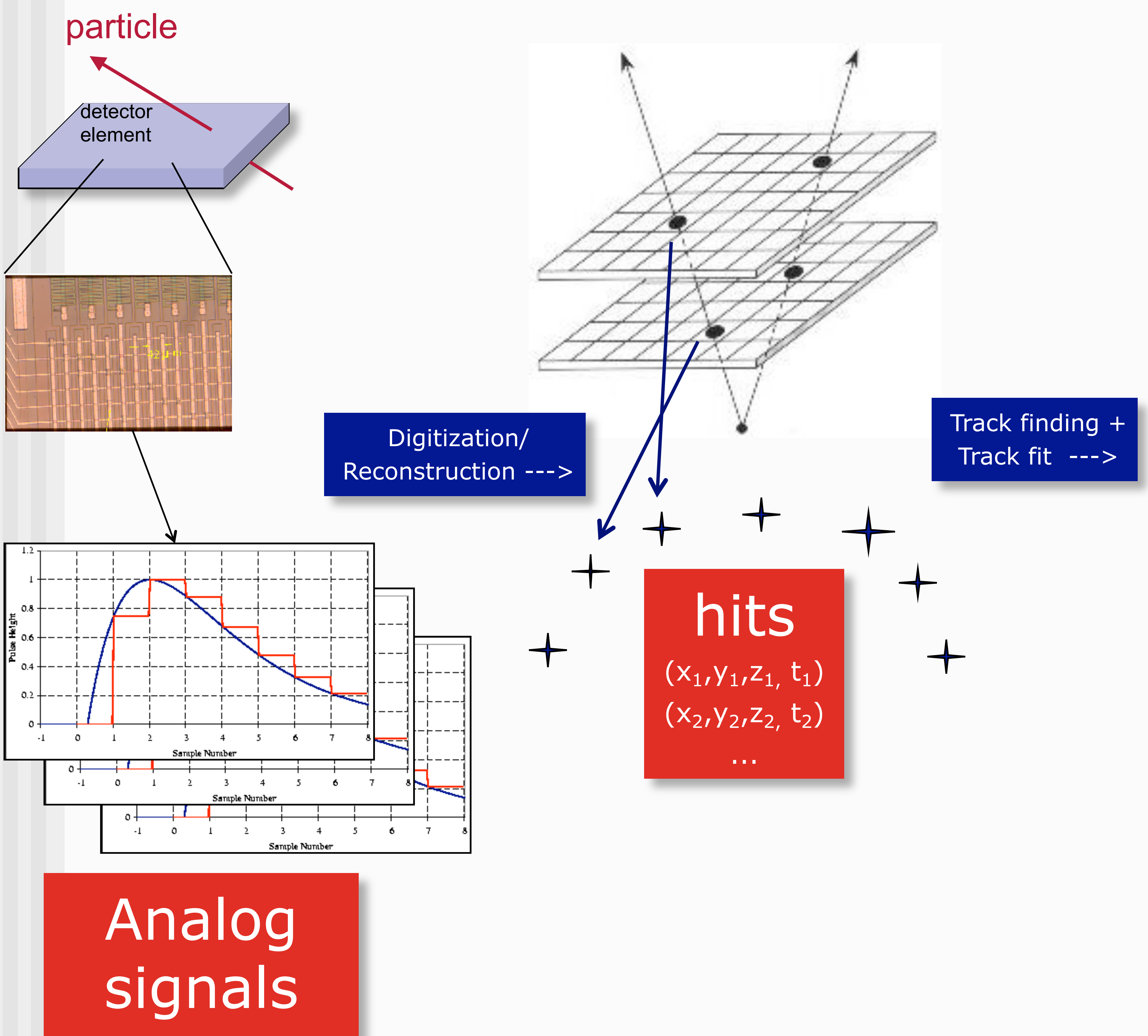
$(x_1,y_1,z_1,\ t_1)$

$(x_2,y_2,z_2,\ t_2)$

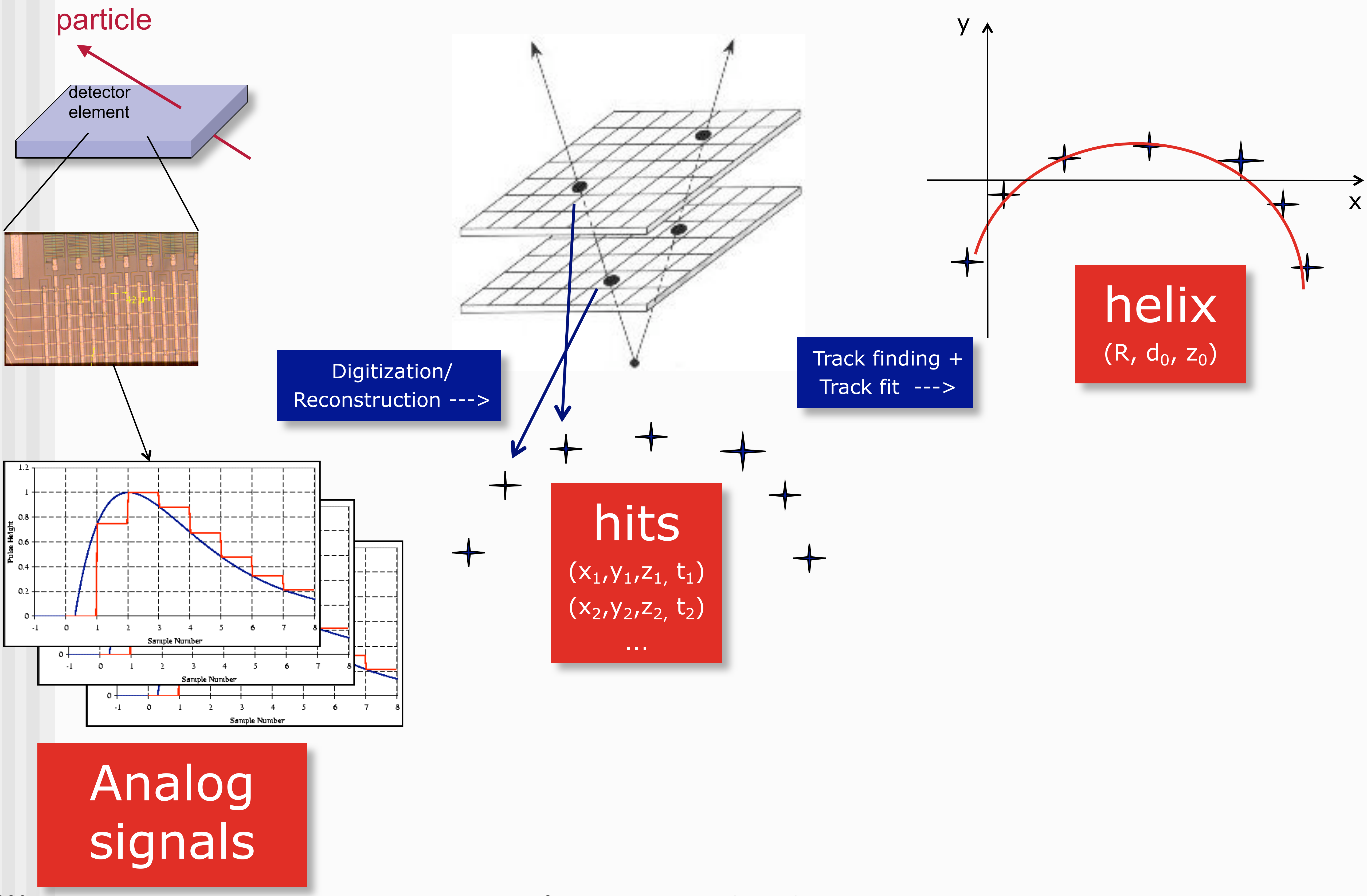
...

Analog signals

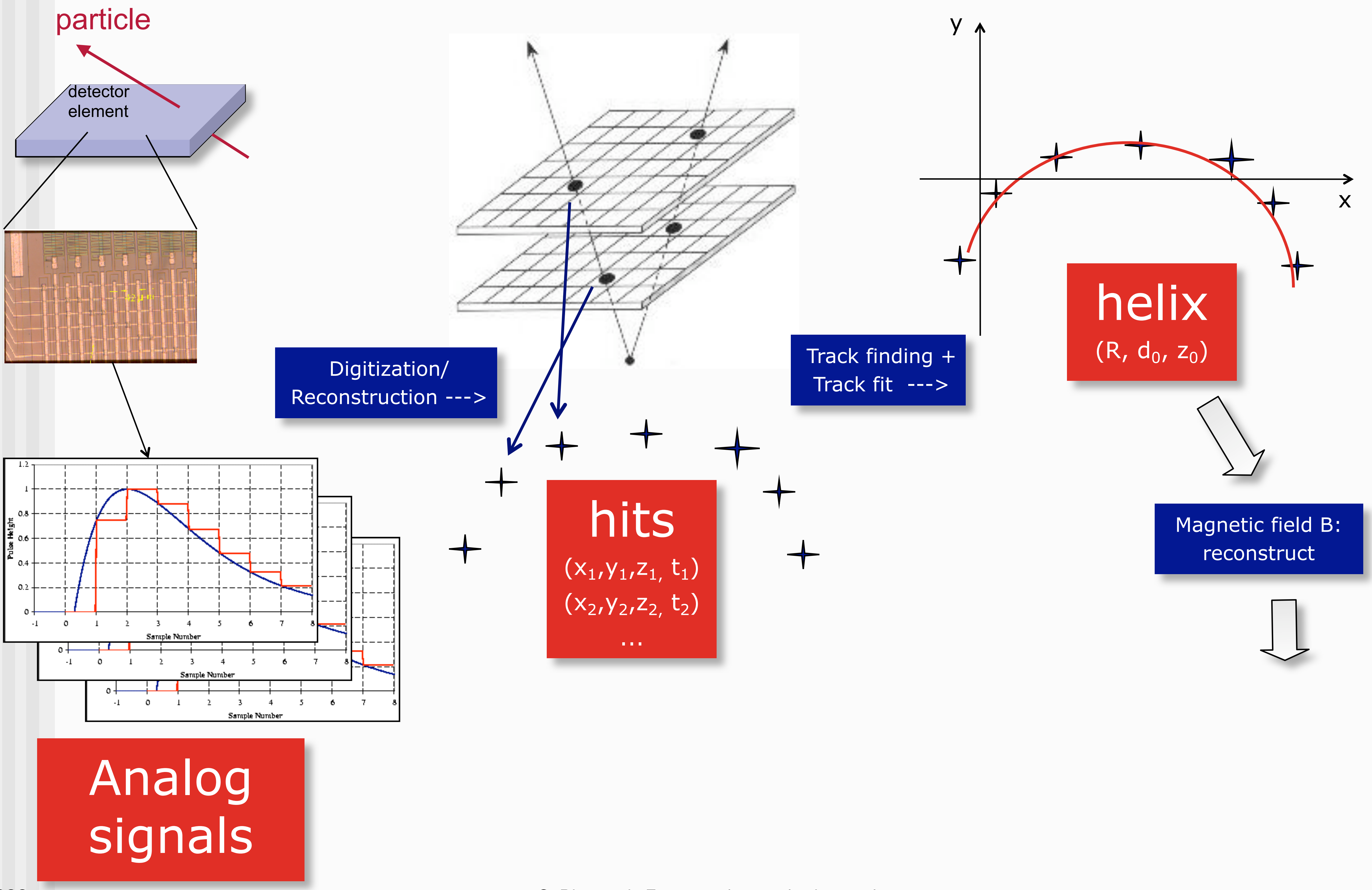
Data reduction/abstraction



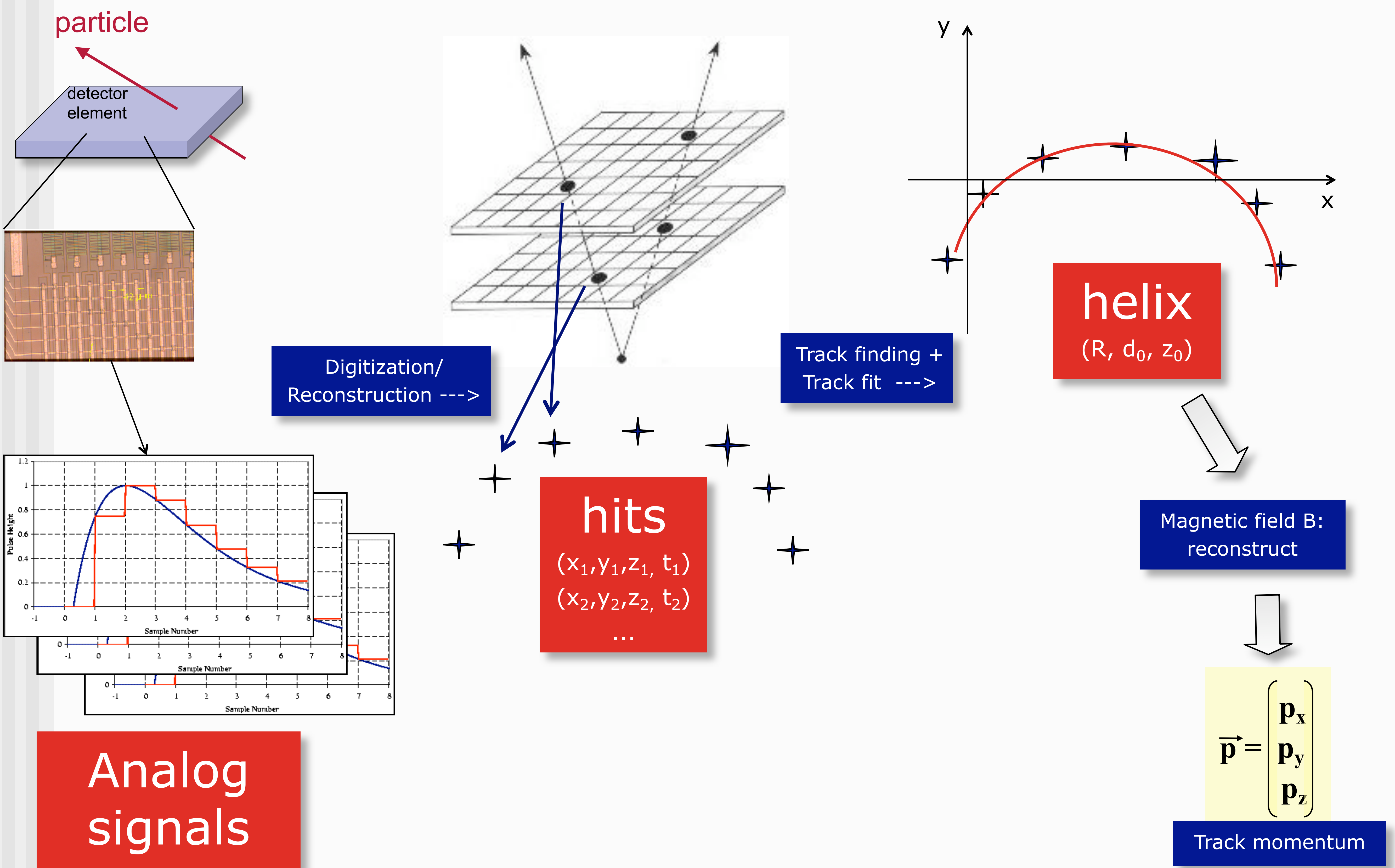
Data reduction/abstraction



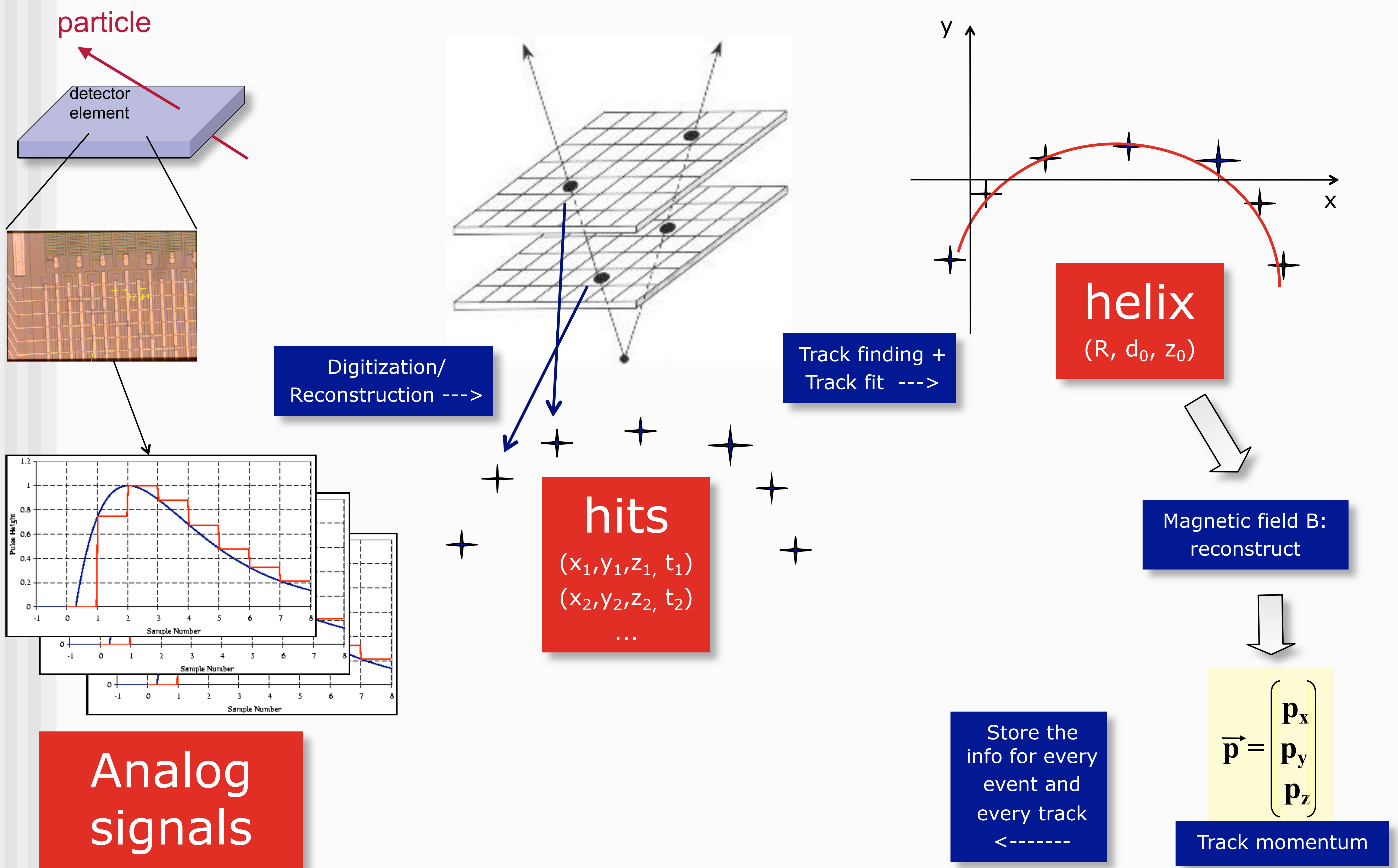
Data reduction/abstraction



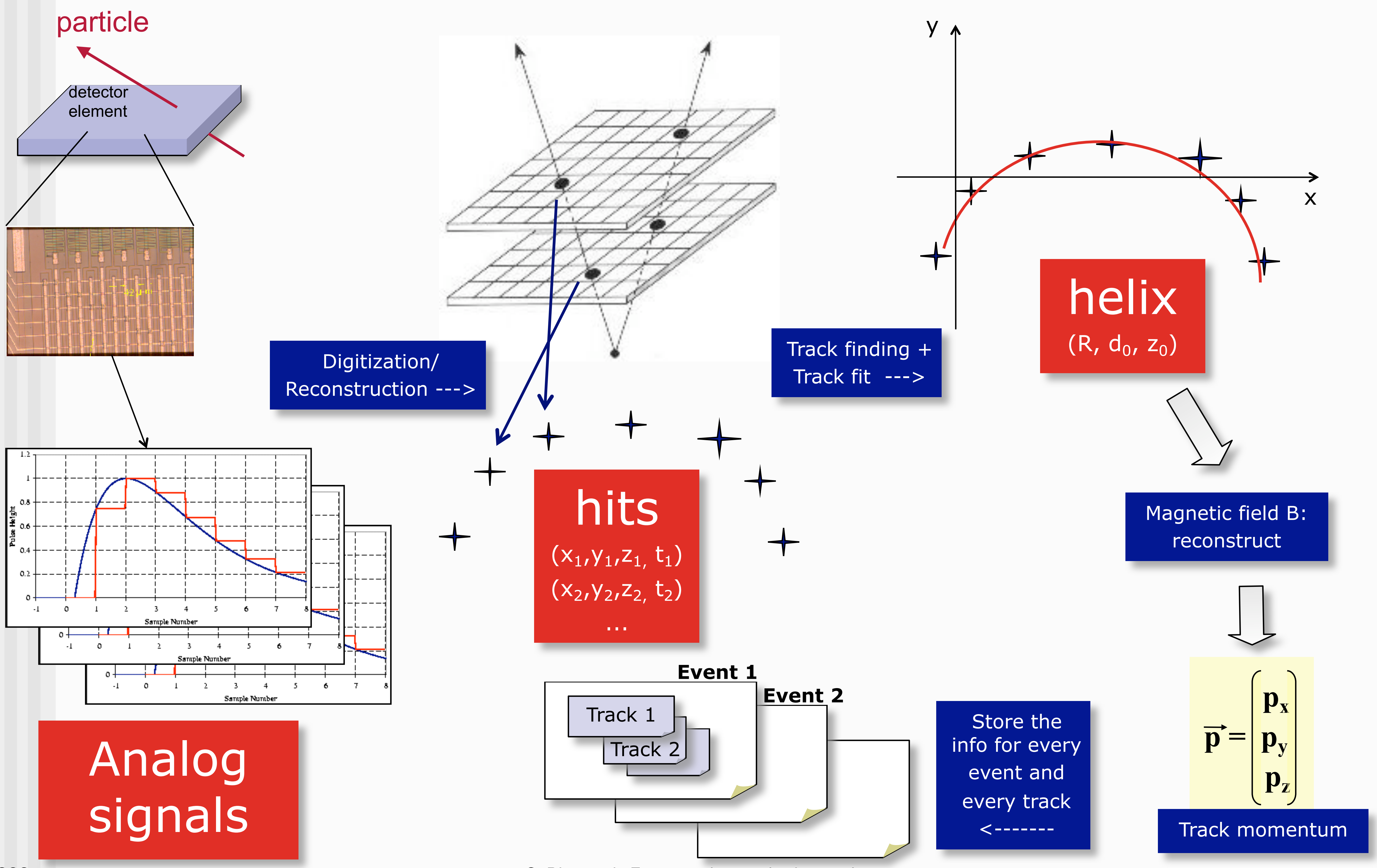
Data reduction/abstraction



Data reduction/abstraction



Data reduction/abstraction



High Level Data Storage

- Data are stored sequentially in files...

File A

High Level Data Storage

- Data are stored sequentially in files...

Event 1

Nch (charged tracks) :

2

Pcha

(Momentum of each track):

{{"-7.65698","42.9725","14.3404"},
{" 7.54101","-42.1729","-14.0108"}}

px py pz

Qcha

(Charge of each track):

{-1,1}

File A

High Level Data Storage

- Data are stored sequentially in files...

Event 1

Nch (charged tracks) :

2

Pcha

(Momentum of each track):

{{"-7.65698","42.9725","14.3404"},
{" 7.54101","-42.1729","-14.0108"}}

px py pz

Qcha

(Charge of each track):

{-1,1}

Event 2

Nch (charged tracks) :

3

Pcha

(Momentum of each track):

{{"-12.9305","12.2713","40.5615"},
{" 12.2469","-11.606","-38.7182"},
{"0.143435","-0.143435","-0.497444"}}

px py pz

Qcha

(Charge of each track):

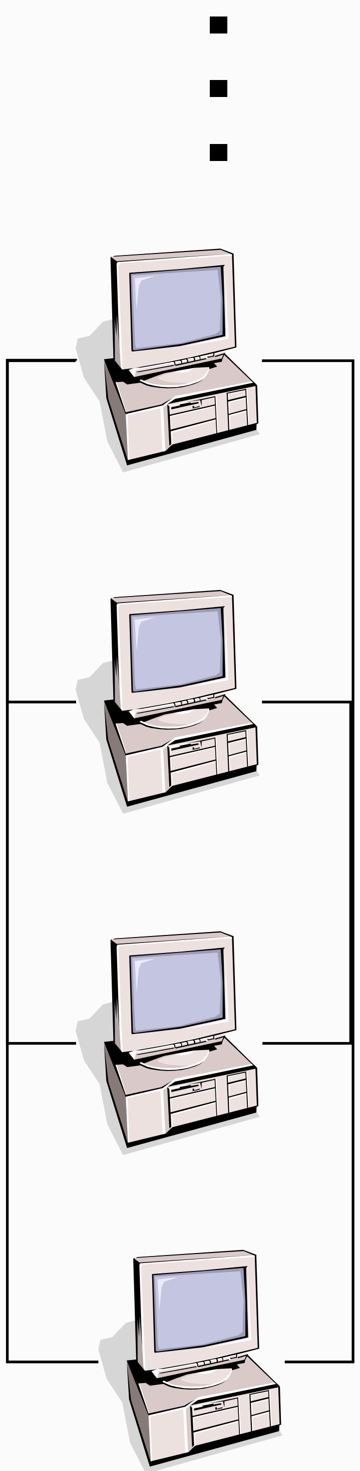
{-1,1,-1}

File A

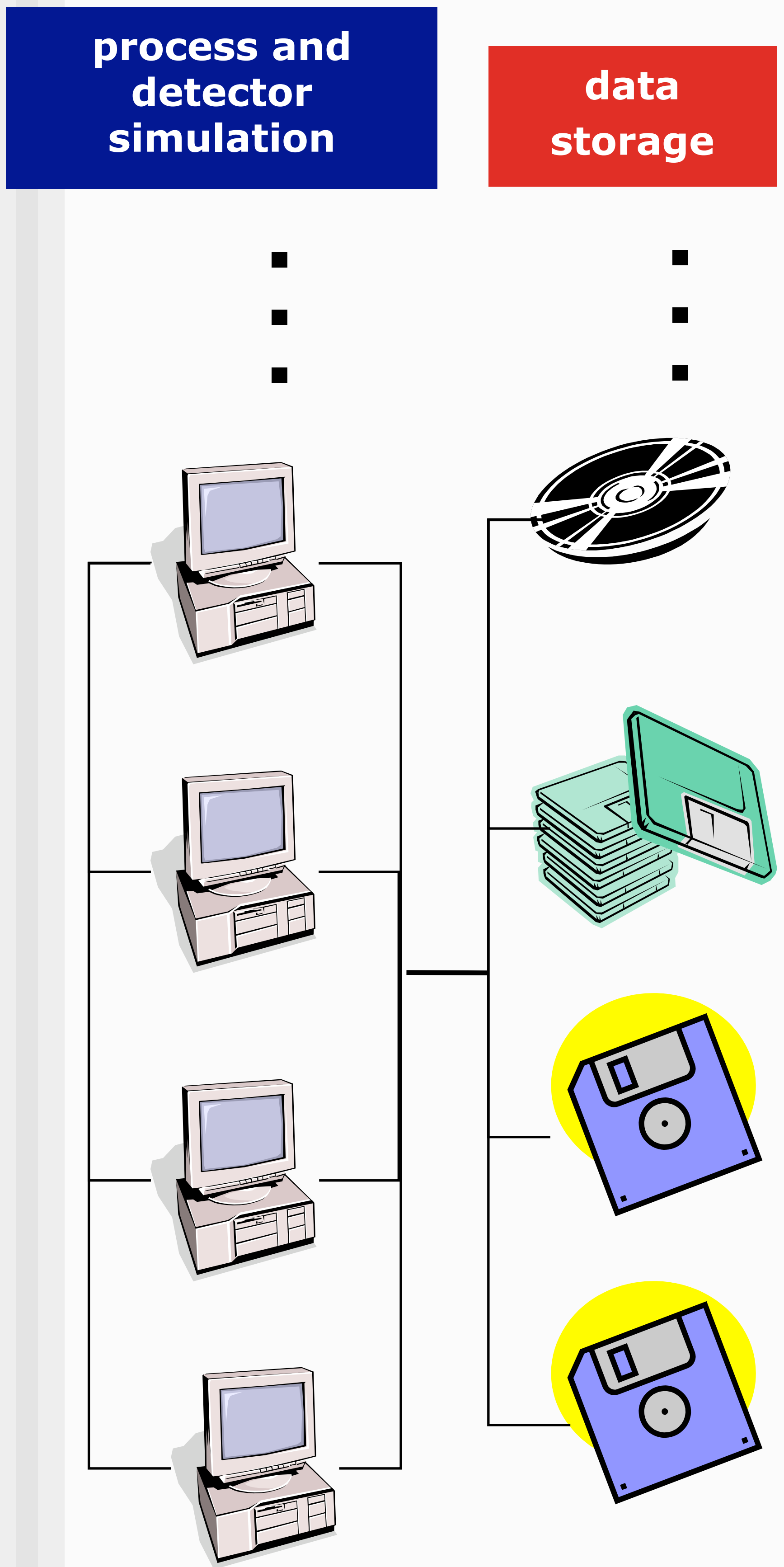
Simulation

Simulation

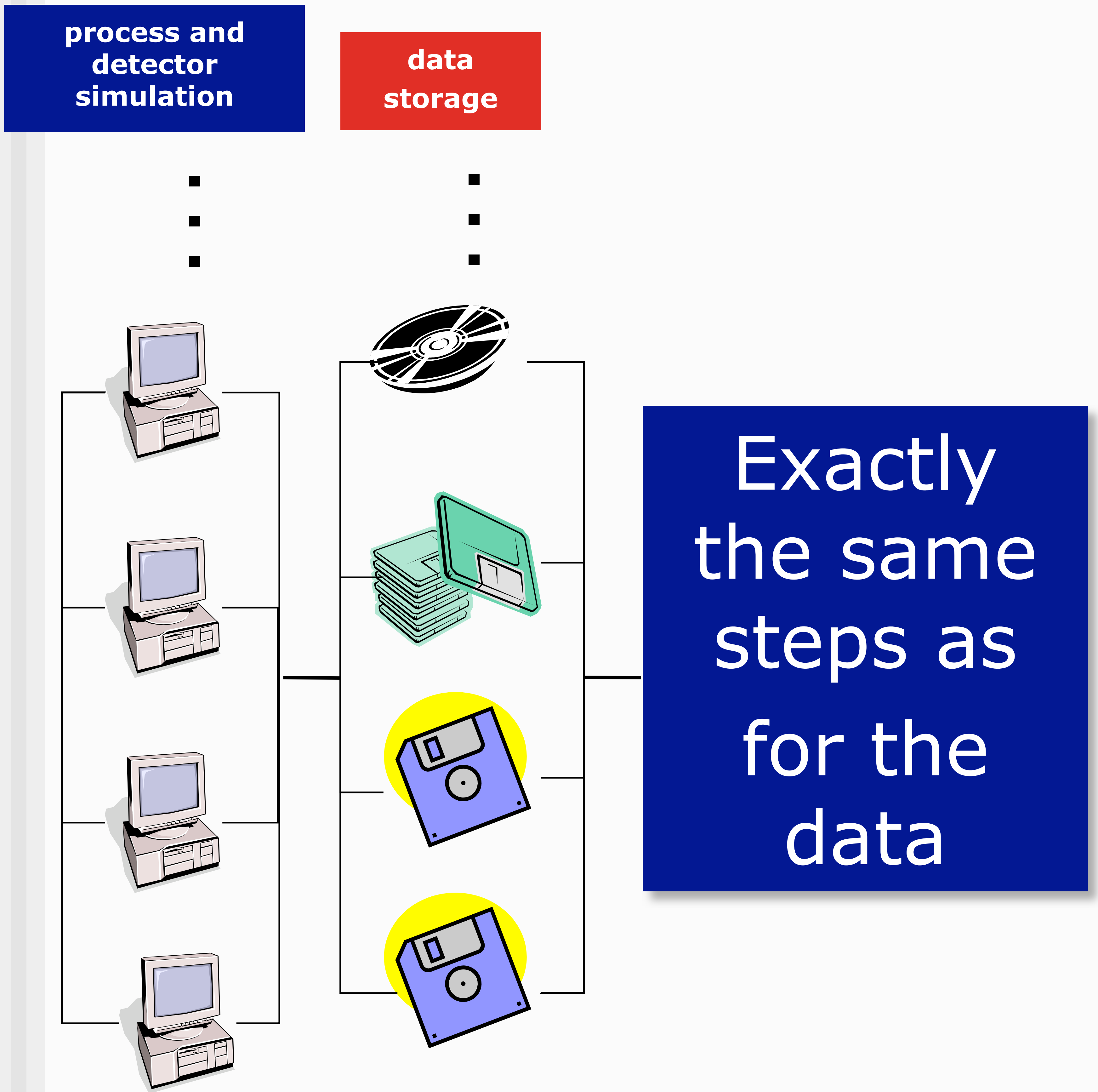
process and detector simulation



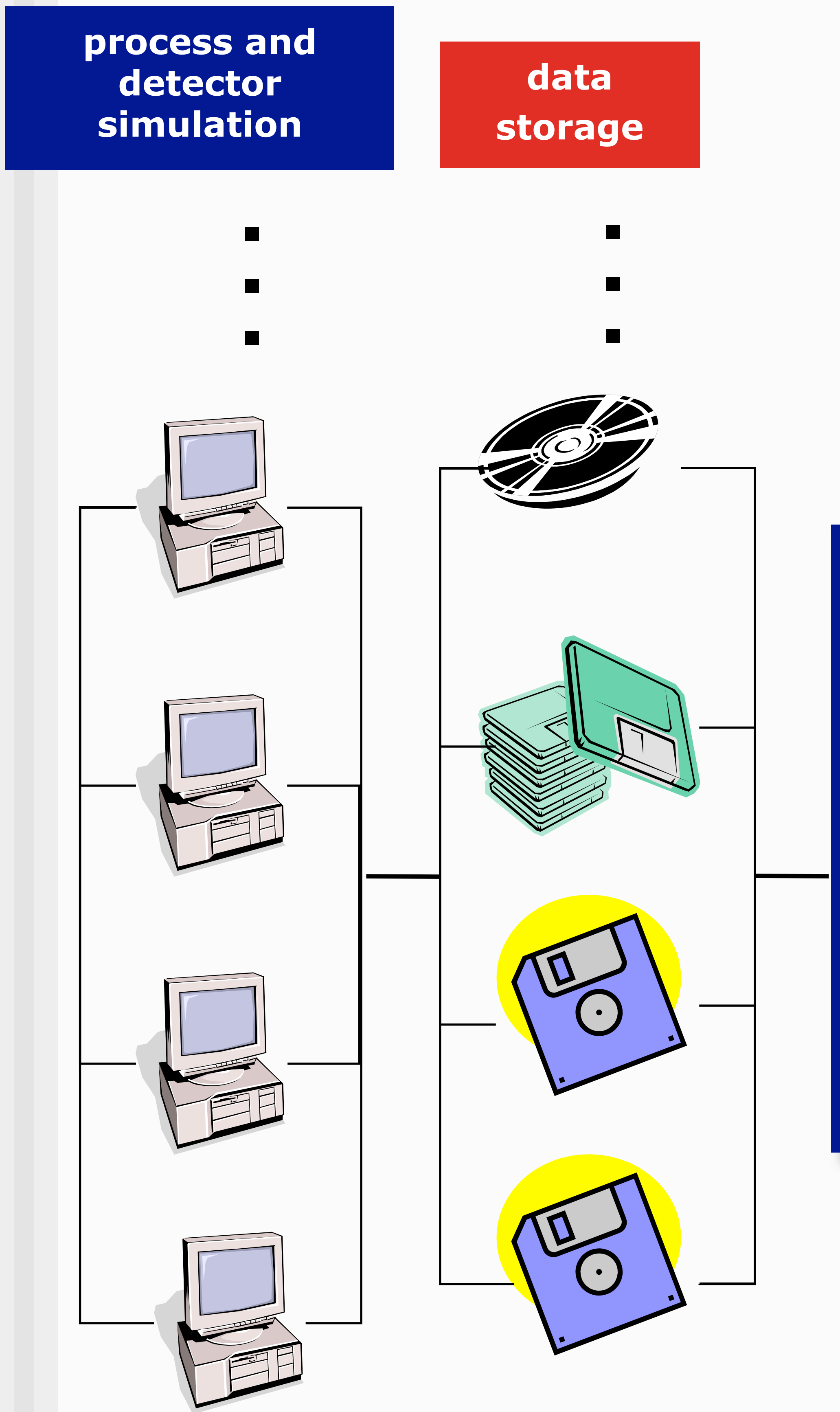
Simulation



Simulation


process and detector simulation
data storage
Exactly the same steps as for the data

Simulation

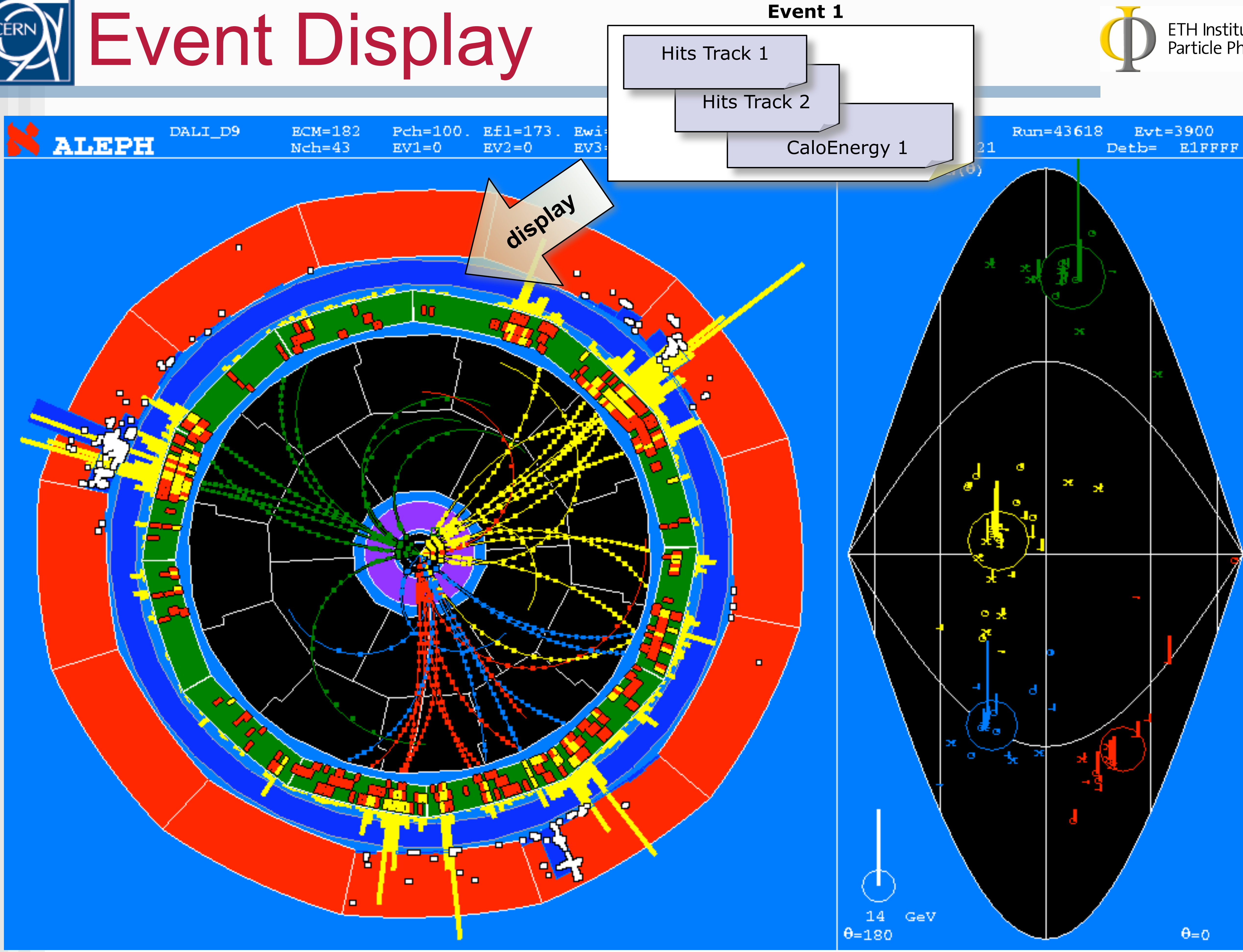


Exactly the same steps as for the data

Simulation of many (millions) of events

- simulate physics process
  e.g. $e^+e^- \rightarrow$ hadrons
  or $p\ p \rightarrow$ jets
- plus the detector response to the produced particles
- understand detector response and analysis parameters (lost particles, resolution, efficiencies, backgrounds)
- and compare to real data
- Note : simulations present from beginning to end of experiment, needed to make design choices

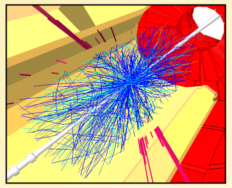
Event Display



And now let's go a little bit more into the details ...

Our Task

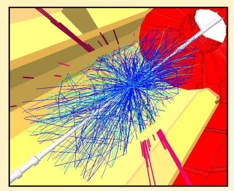
Reality



We use experiments to inquire about what "reality" (nature) does

Our Task

Reality



We use experiments to inquire about what "reality" (nature) does

Theory

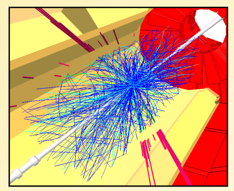
$$S = \mathrm{i} \int \mathrm{d}^4 x \, \mathcal{L}(x)$$

The goal is to understand in the most general; that's usually also the simplest.

- A. Eddington

Our Task

Reality



We use experiments to inquire about what "reality" (nature) does

We intend to fill this gap

Theory

$$S = \mathrm{i} \int \mathrm{d}^4x \, \mathcal{L}(x)$$

The goal is to understand in the most general; that's usually also the simplest.

- A. Eddington

Theory...

$$\mathcal{L} = -\tfrac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \tfrac{1}{4}B_{\mu\nu}B^{\mu\nu}$$

$\left\{\right.$ $W^{\pm}, Z, \gamma$ kinetic energies and self-interactions

$$+\overline{L}\gamma^{\mu}\left(i\partial_{\mu} - g\tfrac{1}{2}\tau\cdot\mathbf{W}_{\mu} - g'\tfrac{Y}{2}B_{\mu}\right)L$$
$$+\overline{R}\gamma^{\mu}\left(i\partial_{\mu} - g'\tfrac{Y}{2}B_{\mu}\right)R$$

$\left\{\right.$ lepton and quark kinetic energies and their interactions with $W^{\pm}, Z, \gamma$

$$+\left|\left(i\partial_{\mu} - g\tfrac{1}{2}\tau\cdot\mathbf{W}_{\mu} - g'\tfrac{Y}{2}B_{\mu}\right)\phi\right|^{2}$$
$$-V(\phi)$$

$\left\{\right.$ $W^{\pm}, Z, \gamma$ and Higgs masses and couplings

$$-(G_1\overline{L}\phi R + G_2\overline{L}\phi_c R + h.c.)$$

$\left\{\right.$ lepton and quark masses and coupling to Higgs

L ... left–handed fermion (l or q) doublet
R ... right–handed fermion singlet

$\mathcal{L}$ from QCD:

$$\mathcal{L} = \underbrace{\bar{q}\,(i\gamma^{\mu}\partial_{\mu} - m)\,q}_{\text{E}_{\text{kin}}(\text{q})} - \underbrace{g\,(\bar{q}\gamma^{\mu}T_a q)G^a_{\mu}}_{\text{Interaction q, g}} - \underbrace{\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{\text{E}_{\text{kin}}(\text{g}) \text{ includes self-interaction between gluons}}$$

eg.
the Standard Model

Theory...

$$\mathcal{L} = -\tfrac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \tfrac{1}{4}B_{\mu\nu}B^{\mu\nu}$$

$\Big\{$ $W^\pm, Z, \gamma$ kinetic energies and self-interactions

$$+\bar{L}\gamma^\mu\left(i\partial_\mu - g\tfrac{1}{2}\tau\cdot\mathbf{W}_\mu - g'\tfrac{Y}{2}B_\mu\right)L$$
$$+\bar{R}\gamma^\mu\left(i\partial_\mu - g'\tfrac{Y}{2}B_\mu\right)R$$

$\Big\{$ lepton and quark kinetic energies and their interactions with $W^\pm, Z, \gamma$

$$+\left|\left(i\partial_\mu - g\tfrac{1}{2}\tau\cdot\mathbf{W}_\mu - g'\tfrac{Y}{2}B_\mu\right)\phi\right|^2 - V(\phi)$$

$\Big\{$ $W^\pm, Z, \gamma$ and Higgs masses and couplings

$$-(G_1\bar{L}\phi R + G_2\bar{L}\phi_c R + h.c.)$$

$\Big\{$ lepton and quark masses and coupling to Higgs

L ... left–handed fermion (l or q) doublet
R ... right–handed fermion singlet

$\mathcal{L}$ from QCD:

$$\mathcal{L} = \bar{q}\underbrace{(i\gamma^\mu\partial_\mu - m)}_{E_{kin}(q)}q - g\underbrace{(\bar{q}\gamma^\mu T_a q)G^a_\mu}_{\text{Interaction } q,\, g} - \underbrace{\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{E_{kin}(g)}$$

$E_{kin}(g)$ includes self-interaction between gluons

eg.
the Standard Model

has parameters

coupling constants

masses

predicts:
cross sections,
branching ratios,
lifetimes, ...

Experiment...

ETH Institute for Particle Physics

```
0x01e84c10:   0x01e8 0x8848 0x01e8 0x83d8 0x6c73 0x6f72 0x7400 0x0000
0x01e84c20:   0x0000 0x0019 0x0000 0x0000 0x01e8 0x4d08 0x01e8 0x5b7c
0x01e84c30:   0x01e8 0x87e8 0x01e8 0x8458 0x7061 0x636b 0x6167 0x6500
0x01e84c40:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c50:   0x01e8 0x8788 0x01e8 0x8498 0x7072 0x6f63 0x0000 0x0000
0x01e84c60:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c70:   0x01e8 0x8824 0x01e8 0x84d8 0x7265 0x6765 0x7870 0x0000
0x01e84c80:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c90:   0x01e8 0x8838 0x01e8 0x8518 0x7265 0x6773 0x7562 0x0000
0x01e84ca0:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cb0:   0x01e8 0x8818 0x01e8 0x8558 0x7265 0x6e61 0x6d65 0x0000
0x01e84cc0:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cd0:   0x01e8 0x8798 0x01e8 0x8598 0x7265 0x7475 0x726e 0x0000
0x01e84ce0:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cf0:   0x01e8 0x87ec 0x01e8 0x85d8 0x7363 0x616e 0x0000 0x0000
0x01e84d00:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d10:   0x01e8 0x87e8 0x01e8 0x8618 0x7365 0x7400 0x0000 0x0000
0x01e84d20:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d30:   0x01e8 0x87a8 0x01e8 0x8658 0x7370 0x6c69 0x7400 0x0000
0x01e84d40:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d50:   0x01e8 0x8854 0x01e8 0x8698 0x7374 0x7269 0x6e67 0x0000
0x01e84d60:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d70:   0x01e8 0x875c 0x01e8 0x86d8 0x7375 0x6273 0x7400 0x0000
0x01e84d80:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d90:   0x01e8 0x87c0 0x01e8 0x8718 0x7377 0x6974 0x6368 0x0000
```

eg.

1/30th of an event in the BaBar detector

- get about 100 evts/sec

Experiment...

ETH Institute for Particle Physics

```
0x01e84c10:  0x01e8 0x8848 0x01e8 0x83d8 0x6c73 0x6f72 0x7400 0x0000
0x01e84c20:  0x0000 0x0019 0x0000 0x0000 0x01e8 0x4d08 0x01e8 0x5b7c
0x01e84c30:  0x01e8 0x87e8 0x01e8 0x8458 0x7061 0x636b 0x6167 0x6500
0x01e84c40:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c50:  0x01e8 0x8788 0x01e8 0x8498 0x7072 0x6f63 0x0000 0x0000
0x01e84c60:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c70:  0x01e8 0x8824 0x01e8 0x84d8 0x7265 0x6765 0x7870 0x0000
0x01e84c80:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c90:  0x01e8 0x8838 0x01e8 0x8518 0x7265 0x6773 0x7562 0x0000
0x01e84ca0:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cb0:  0x01e8 0x8818 0x01e8 0x8558 0x7265 0x6e61 0x6d65 0x0000
0x01e84cc0:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cd0:  0x01e8 0x8798 0x01e8 0x8598 0x7265 0x7475 0x726e 0x0000
0x01e84ce0:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cf0:  0x01e8 0x87ec 0x01e8 0x85d8 0x7363 0x616e 0x0000 0x0000
0x01e84d00:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d10:  0x01e8 0x87e8 0x01e8 0x8618 0x7365 0x7400 0x0000 0x0000
0x01e84d20:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d30:  0x01e8 0x87a8 0x01e8 0x8658 0x7370 0x6c69 0x7400 0x0000
0x01e84d40:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d50:  0x01e8 0x8854 0x01e8 0x8698 0x7374 0x7269 0x6e67 0x0000
0x01e84d60:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d70:  0x01e8 0x875c 0x01e8 0x86d8 0x7375 0x6273 0x7400 0x0000
0x01e84d80:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d90:  0x01e8 0x87c0 0x01e8 0x8718 0x7377 0x6974 0x6368 0x0000
```

eg.

1/30th of an event in the BaBar detector

- get about 100 evts/sec

"Address" :

- which detector element took the reading

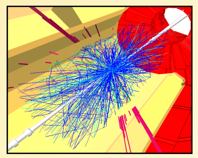
"Value(s)" :

- what the electronics wrote out

Making the connection

ETH Institute for Particle Physics

Reality



Raw Data

The imperfect measurement of a (set of) interactions in the detector

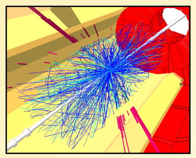
Theory

$$S = \mathrm{i} \int \mathrm{d}^4 x \, \mathcal{L}(x)$$

A small number of general equations, with some parameters (poorly or not known at all)

Making the connection

Reality



Raw Data

Events

The imperfect measurement of a (set of) interactions in the detector

A unique happening:
eg. Run 23458, event 1345
which contains a Z $\rightarrow \mu^+\mu^-$ decay

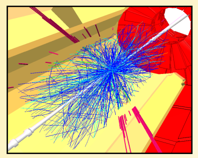
Theory

$$S = \mathrm{i} \int \mathrm{d}^4x \, \mathcal{L}(x)$$

A small number of general equations, with some parameters (poorly or not known at all)

Making the connection

Reality



Raw Data

The imperfect measurement of a (set of) interactions in the detector

Events

A unique happening:
eg. Run 23458, event 1345
which contains a Z $\rightarrow \mu^+\mu^-$ decay

Observables

cross sections (probabilities for interactions), branching ratios (BR), ratios of BRs, specific lifetimes, ...

Theory

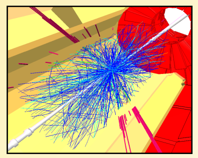
$$S = \mathrm{i} \int \mathrm{d}^4x \, \mathcal{L}(x)$$

A small number of general equations, with some parameters (poorly or not known at all)

Making the connection

ETH Institute for Particle Physics

Reality



Raw Data

The imperfect measurement of a (set of) interactions in the detector

Events

A unique happening:
eg. Run 23458, event 1345
which contains a $Z \rightarrow \mu^+\mu^-$ decay

Analysis : We “confront theory with experiment” by comparing the measured quantity (observable) with the prediction.

Observables

cross sections (probabilities for interactions), branching ratios (BR), ratios of BRs, specific lifetimes, ...

Theory

$$S = \mathrm{i} \int \mathrm{d}^4x\, \mathcal{L}(x)$$

A small number of general equations, with some parameters (poorly or not known at all)

ETH Institute for Particle Physics

A “simple” example

A simple example

- Measurement of e^+e^- annihilation into hadrons and muons:

sum over all quark flavours, which can be produced at a certain e^+e^- centre-of-mass energy E_{CM}, ,eg. d, u, s, c, b, t

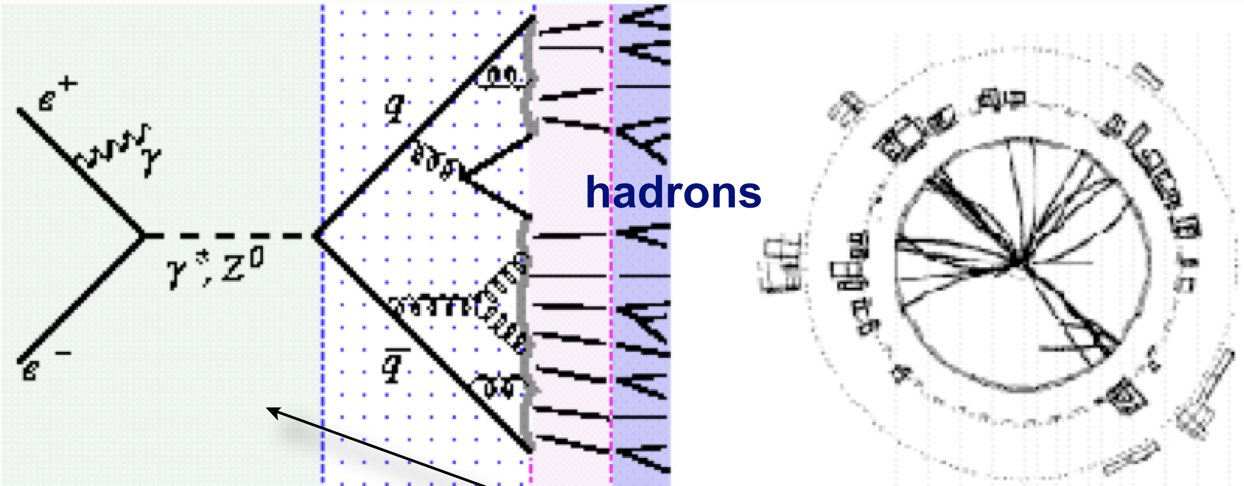
$$\boxed{R := \frac{\sigma(\mathrm{e}^+\mathrm{e}^- \to \mathrm{q}_f\bar{\mathrm{q}}_f)}{\sigma(\mathrm{e}^+\mathrm{e}^- \to \mu^+\mu^-)} = N_c \sum_f z_f^2}$$

Number of colours

electric charges of quarks, in units of electron charge

A simple example

- Measurement of e^+e^- annihilation into hadrons and muons:



Hadronic final state

- many charged tracks (>~ 10)
- sum of energy deposits in calorimeters not too far from centre-of-mass energy

sum over all quark flavours, which can be produced at a certain e^+e^- centre-of-mass energy E_{CM}, ,eg. d, u, s, c, b, t

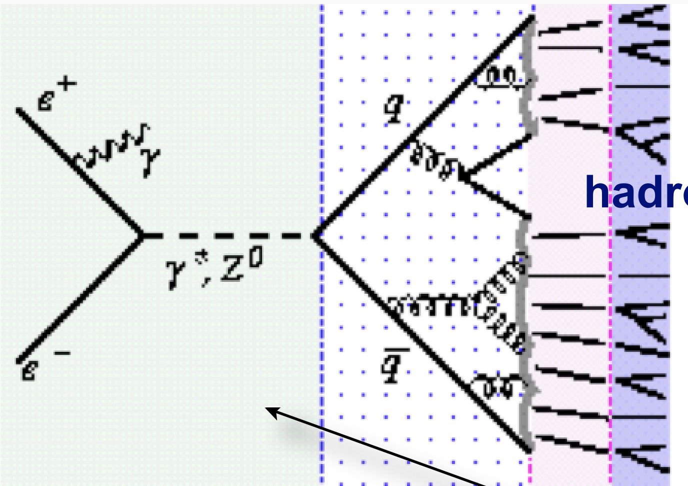
$$R := \frac{\sigma(e^+e^- \to q_f\bar{q}_f)}{\sigma(e^+e^- \to \mu^+\mu^-)} = N_c \sum_f z_f^2$$

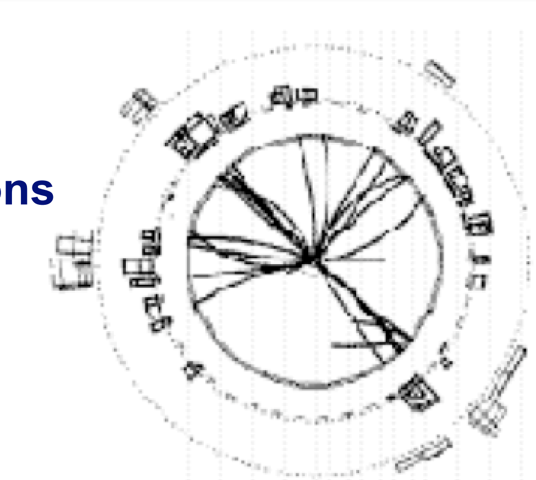
Number of colours

electric charges of quarks, in units of electron charge

A simple example

- Measurement of e^+e^- annihilation into hadrons and muons:





Hadronic final state

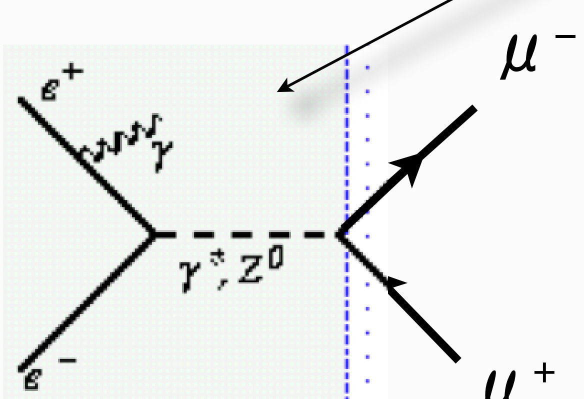
- many charged tracks (>~ 10)
- sum of energy deposits in calorimeters not too far from centre-of-mass energy

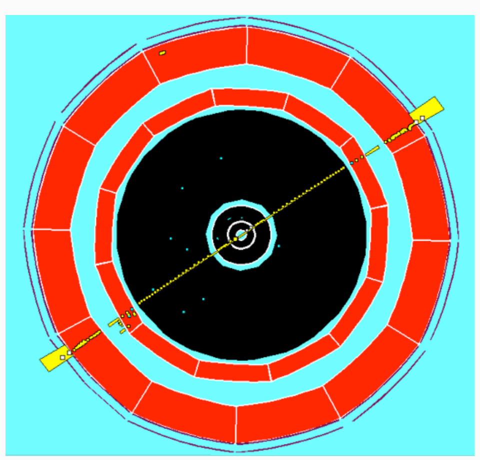
sum over all quark flavours, which can be produced at a certain e^+e^- centre-of-mass energy E_{CM}, ,eg. d, u, s, c, b, t

$$R := \frac{\sigma(e^+e^- \rightarrow q_f\bar{q}_f)}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} = N_c \sum_f z_f^2$$

Number of colours

electric charges of quarks, in units of electron charge





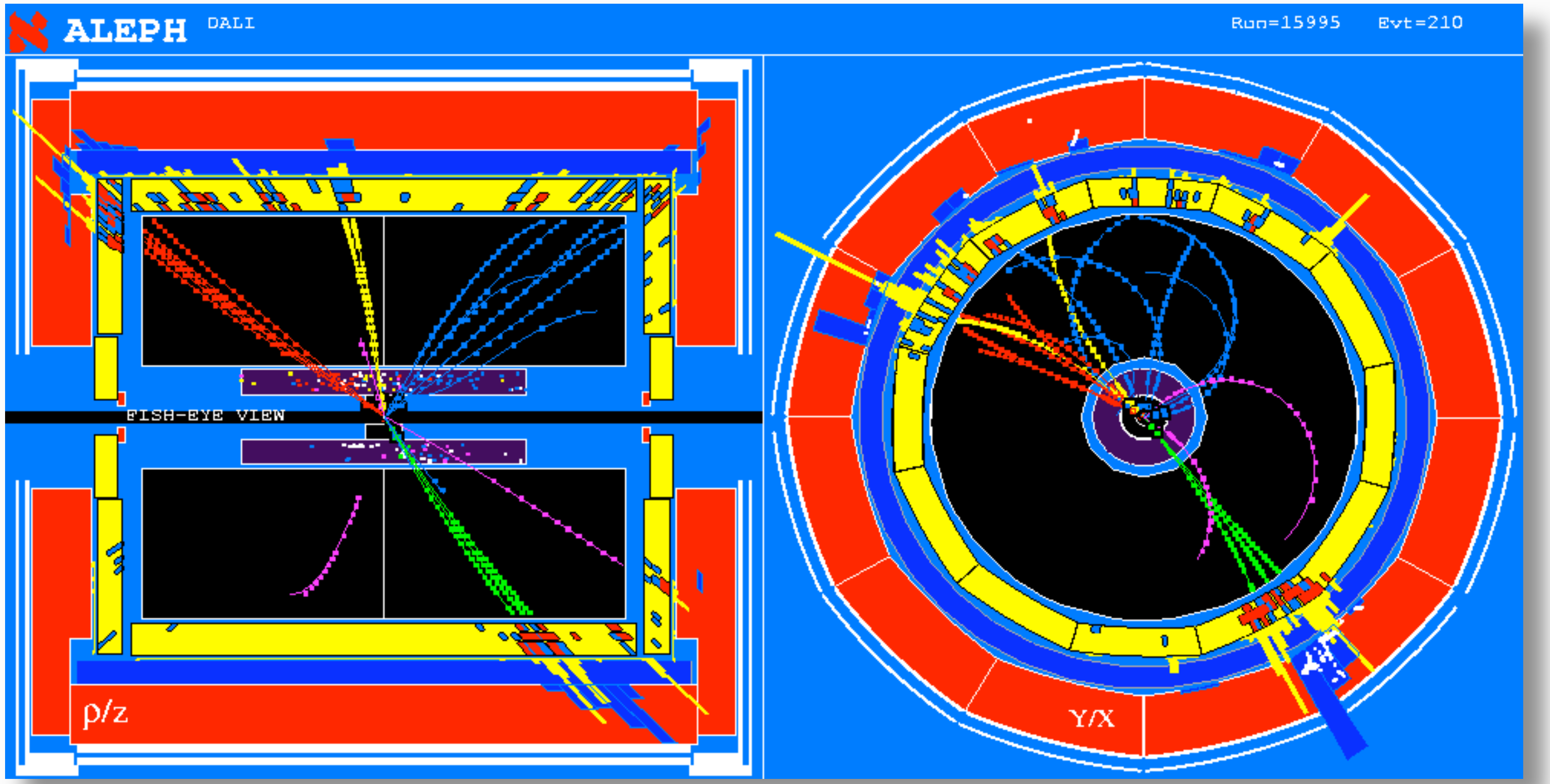
Muonic final state

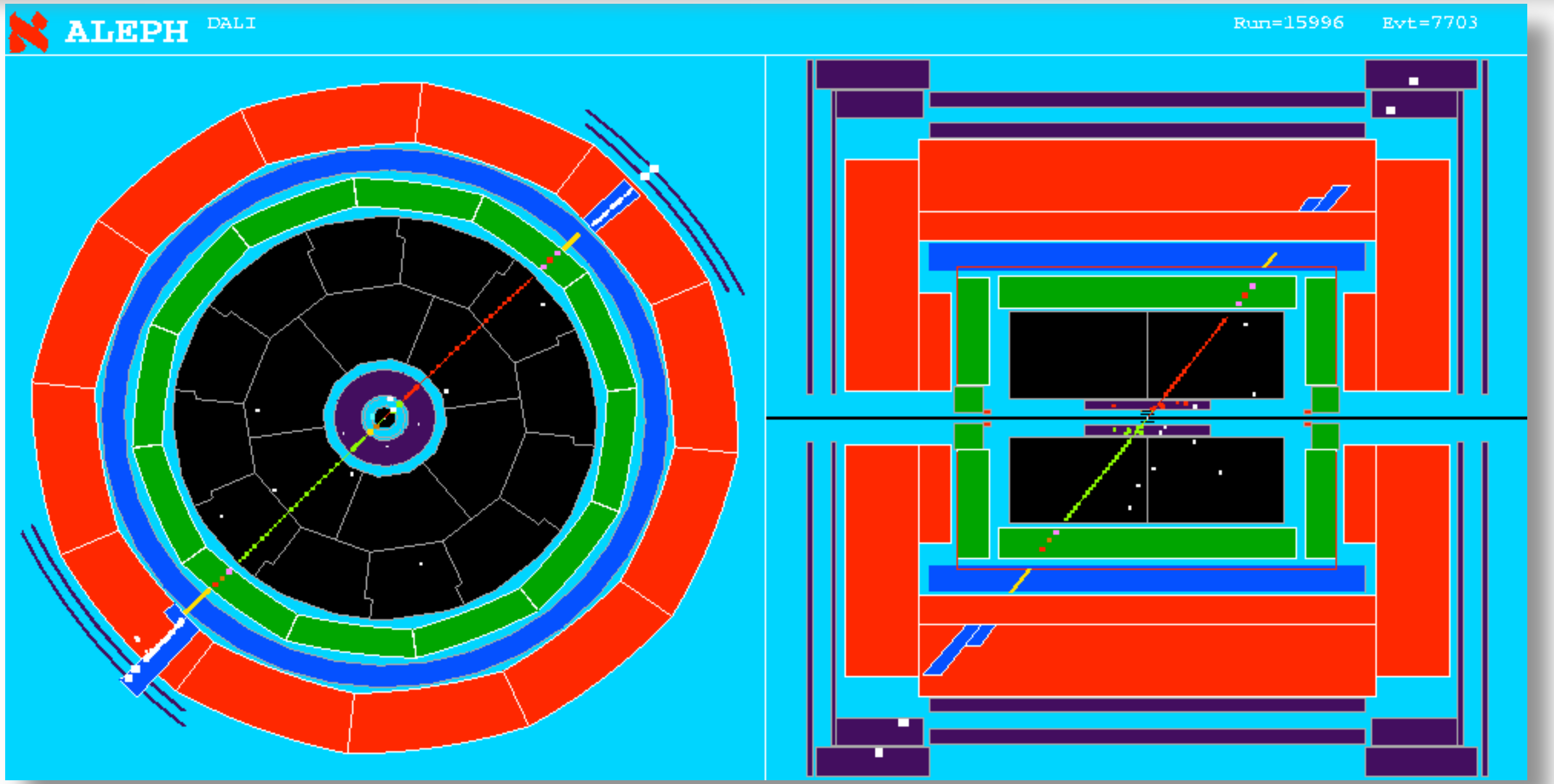
- two charged tracks, approx. back-to-back, with expected momentum (~ 1/2 E_{CM})
- right number of muon hits in outer layers (muons very penetrating, traverse whole detector)
- expected energy in calorimeter (electrons deposit all their energy, muons leave little)

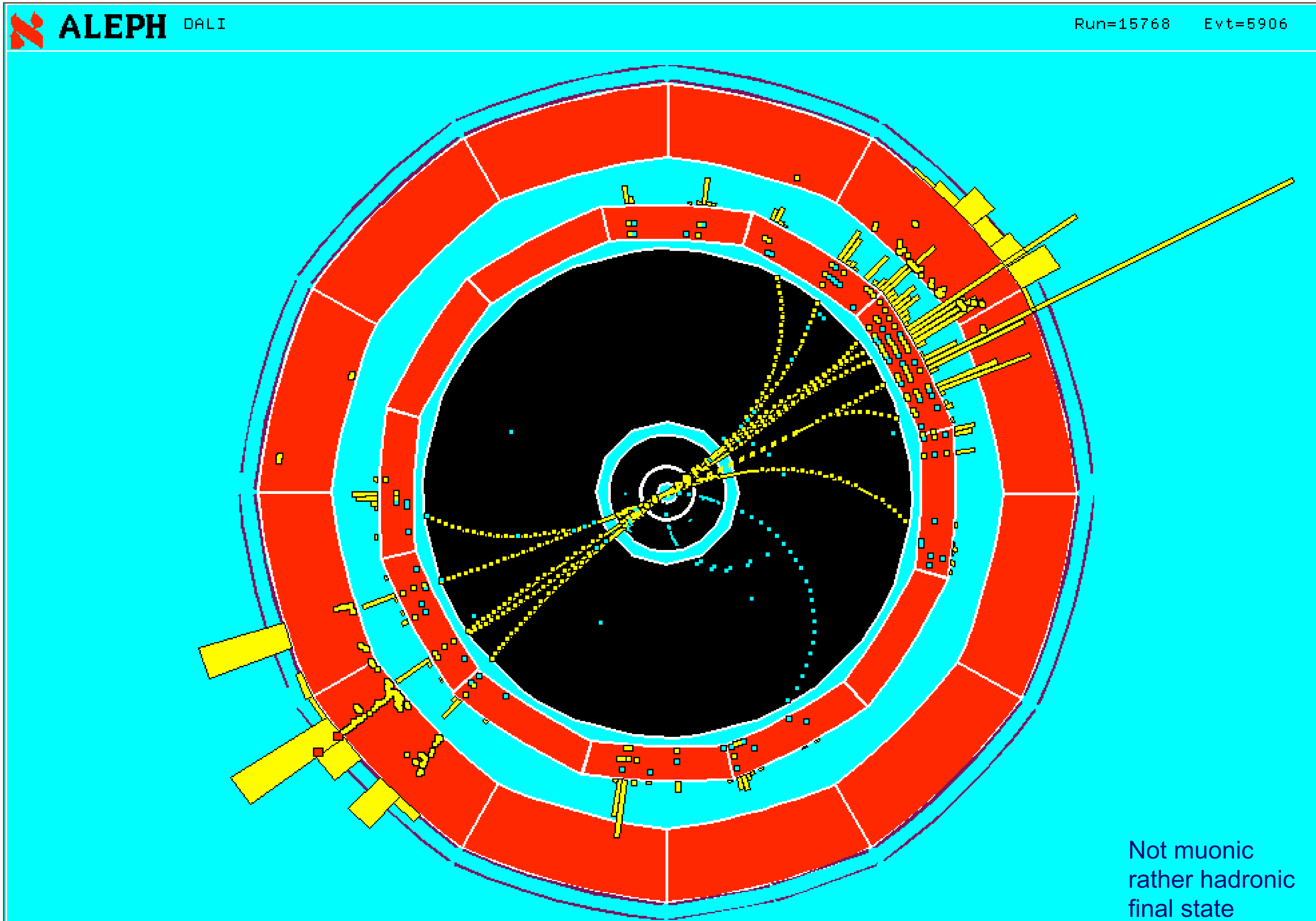
A “simple” counting experiment

ETH Institute for Particle Physics

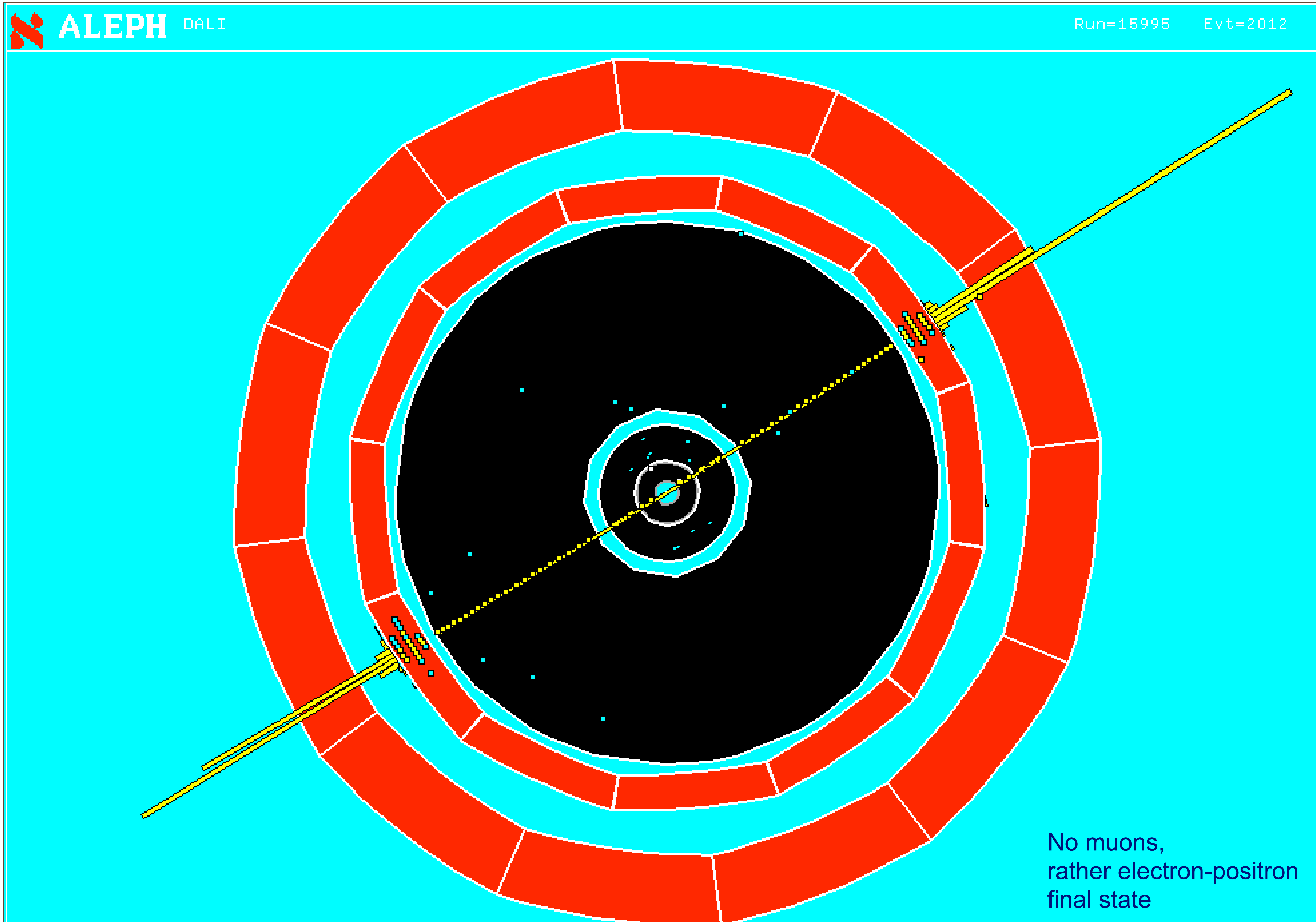
$$R = \frac{\#\text{ of} \quad \text{Hadron final states}}{\#\text{ of} \quad \text{Muon final states}}$$





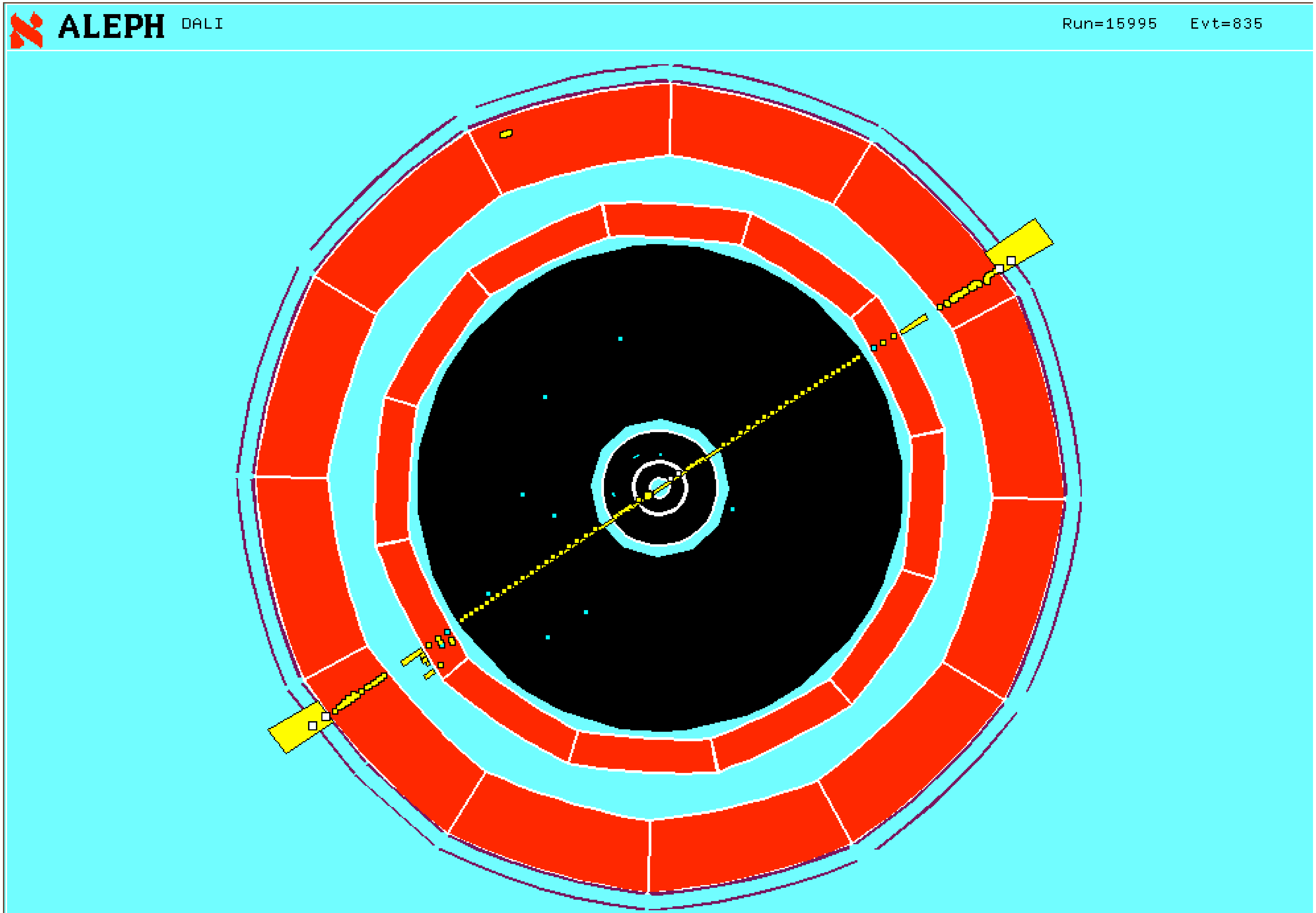


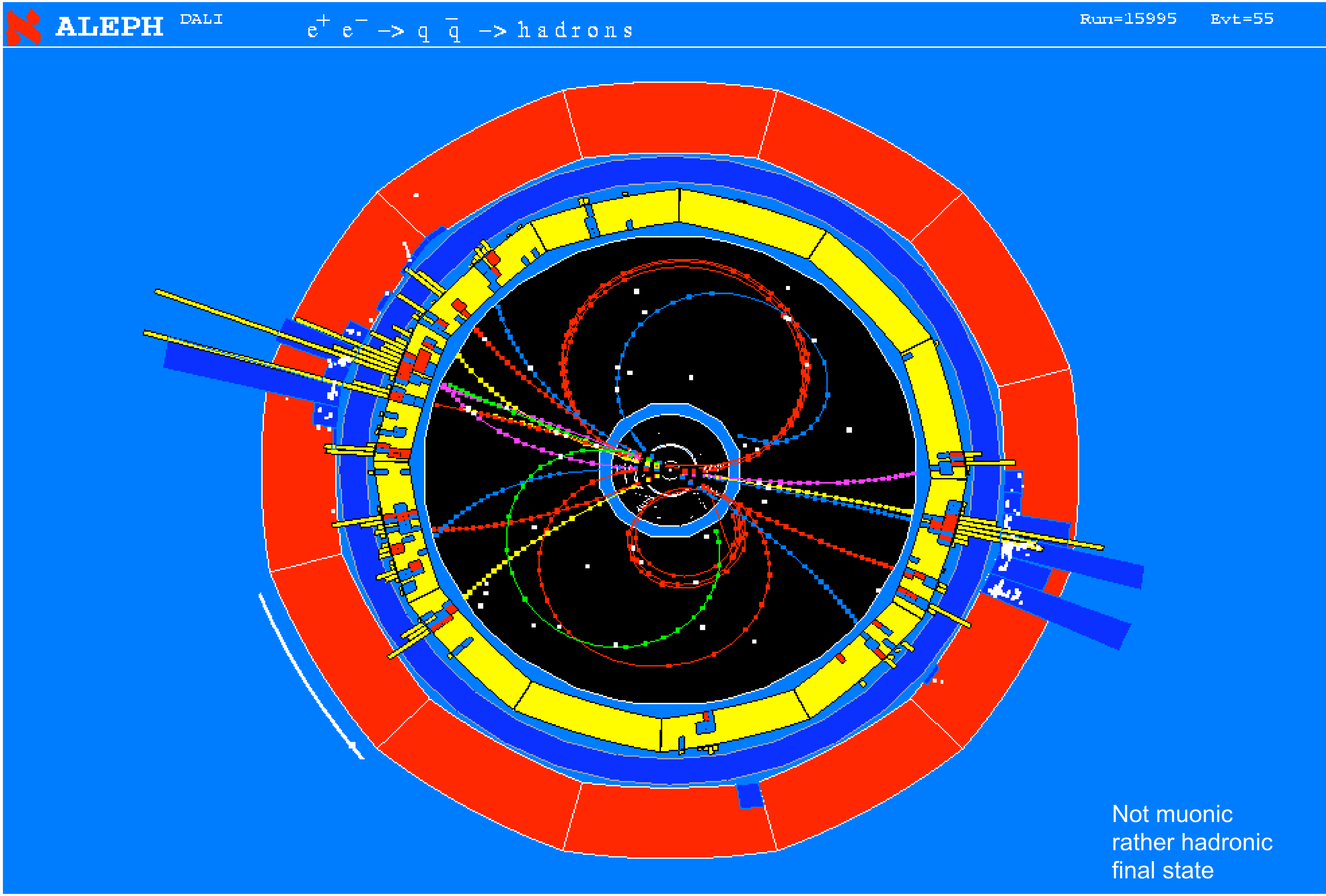
Not muonic
rather hadronic
final state



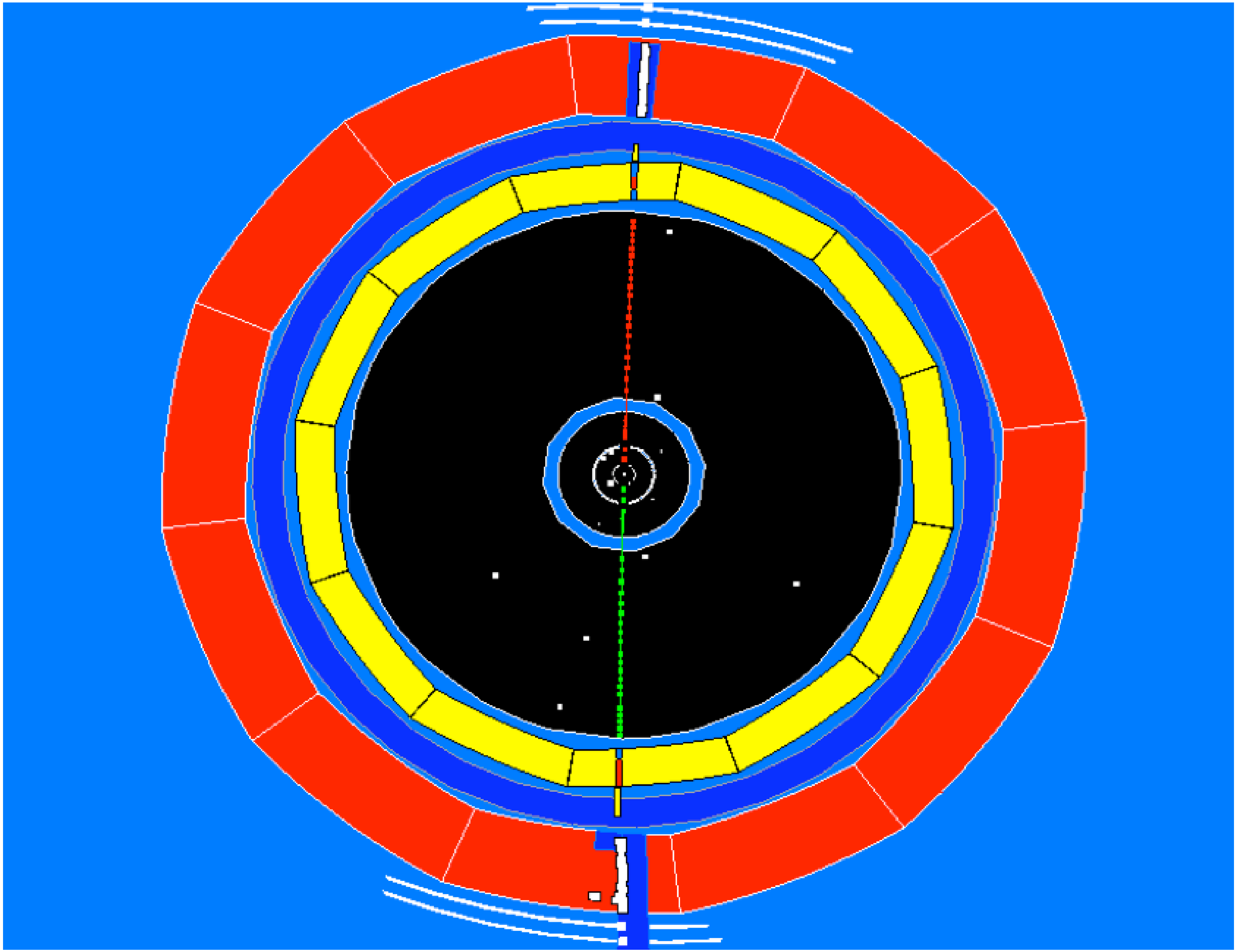
No muons,
rather electron-positron
final state

$Z \rightarrow \mu^+\mu^-$



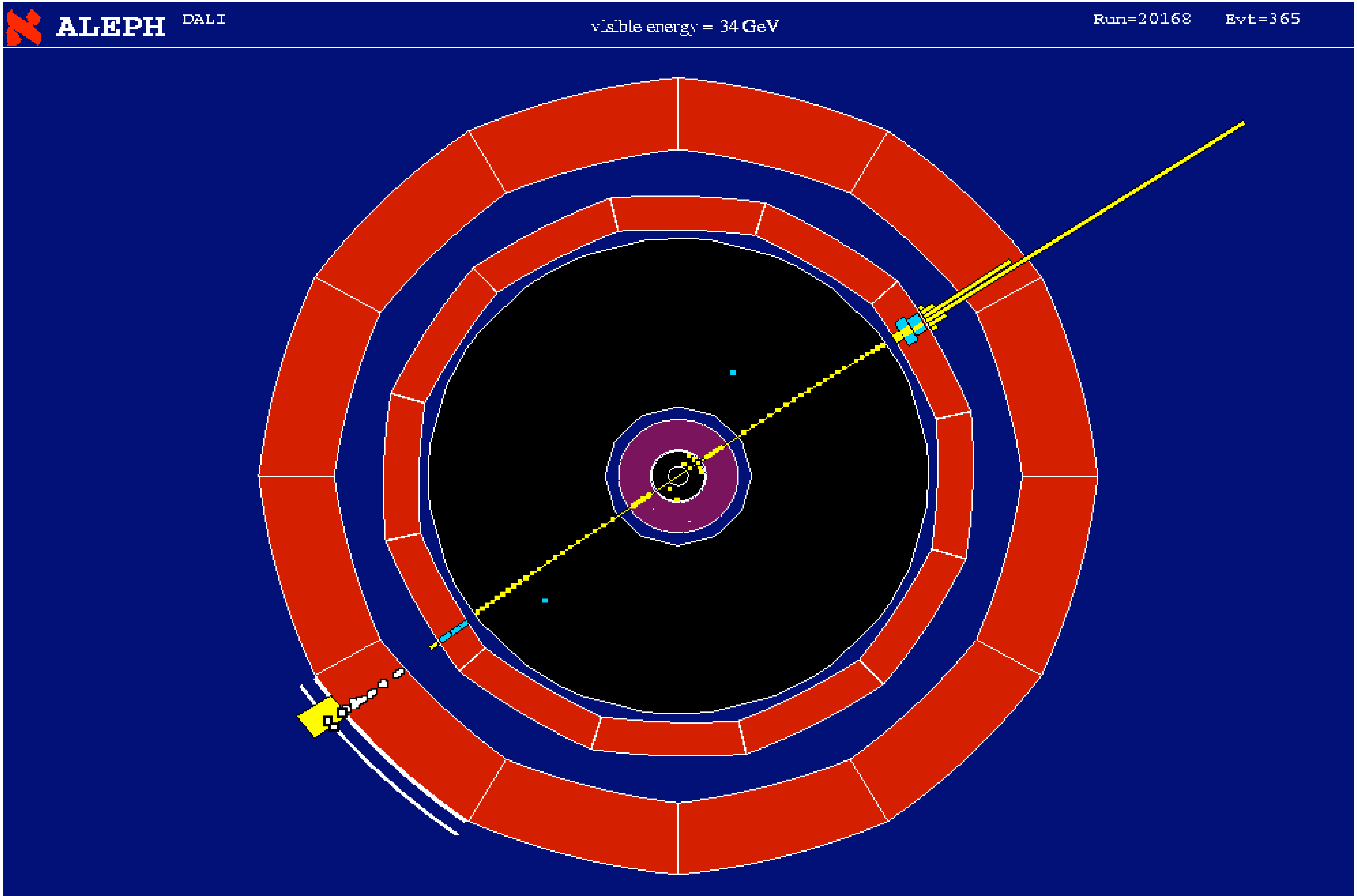

ALEPH
DALI
$e^+ e^- \to q \bar{q} \to$ hadrons
Run=15995
Evt=55
Not muonic
rather hadronic
final state

$Z \rightarrow \mu^+\mu^-$

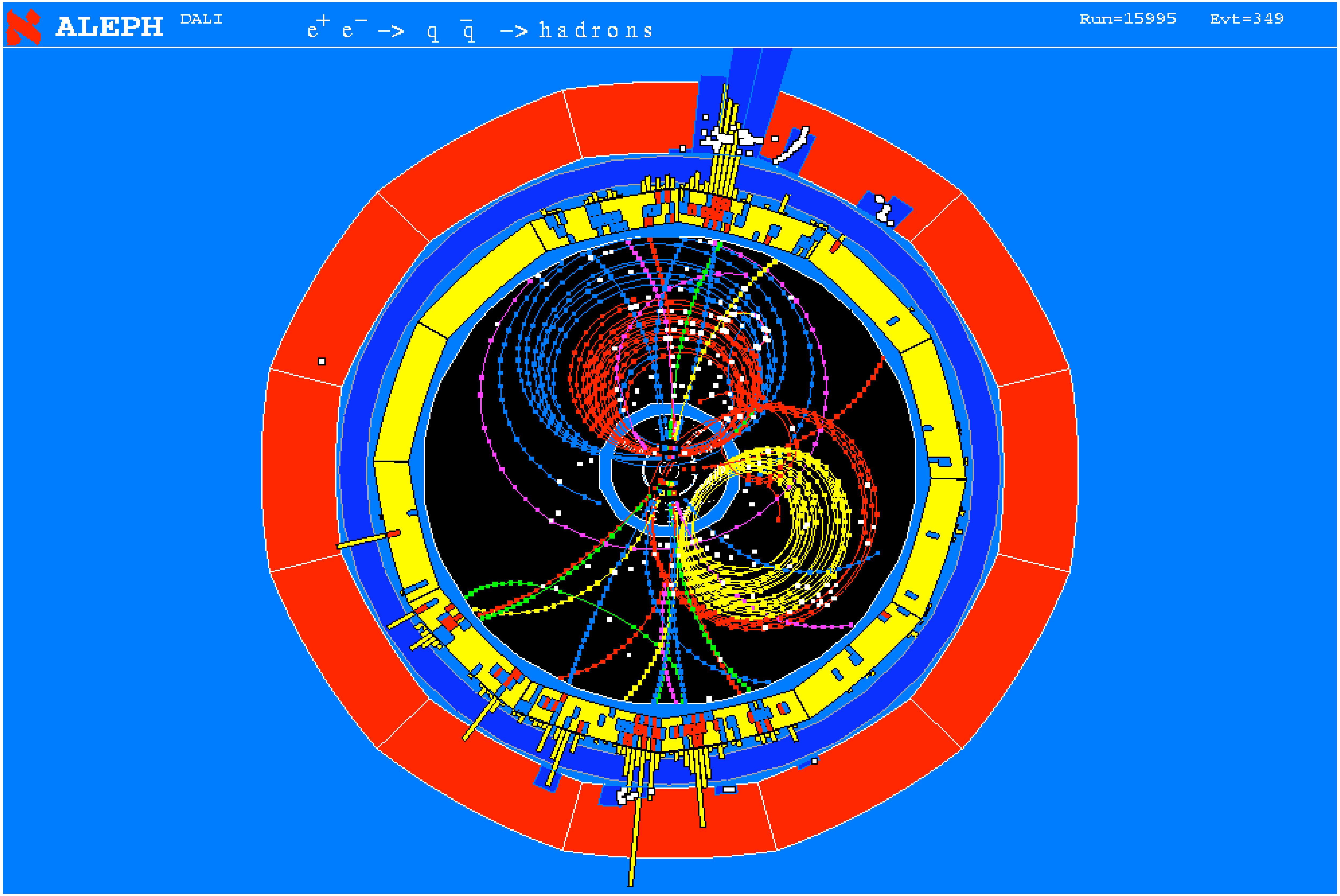


not $Z \rightarrow \mu^+\mu^-$

rather Z decay to $\tau^+\tau^-$,
one tau decayed to electron + 2 neutrinos
the other tau decayed to muon + 2 neutrinos



Not muonic, rather hadronic final state



And so on....

Result

For E_{CM} below the Z peak and above the Υ resonance we expect:

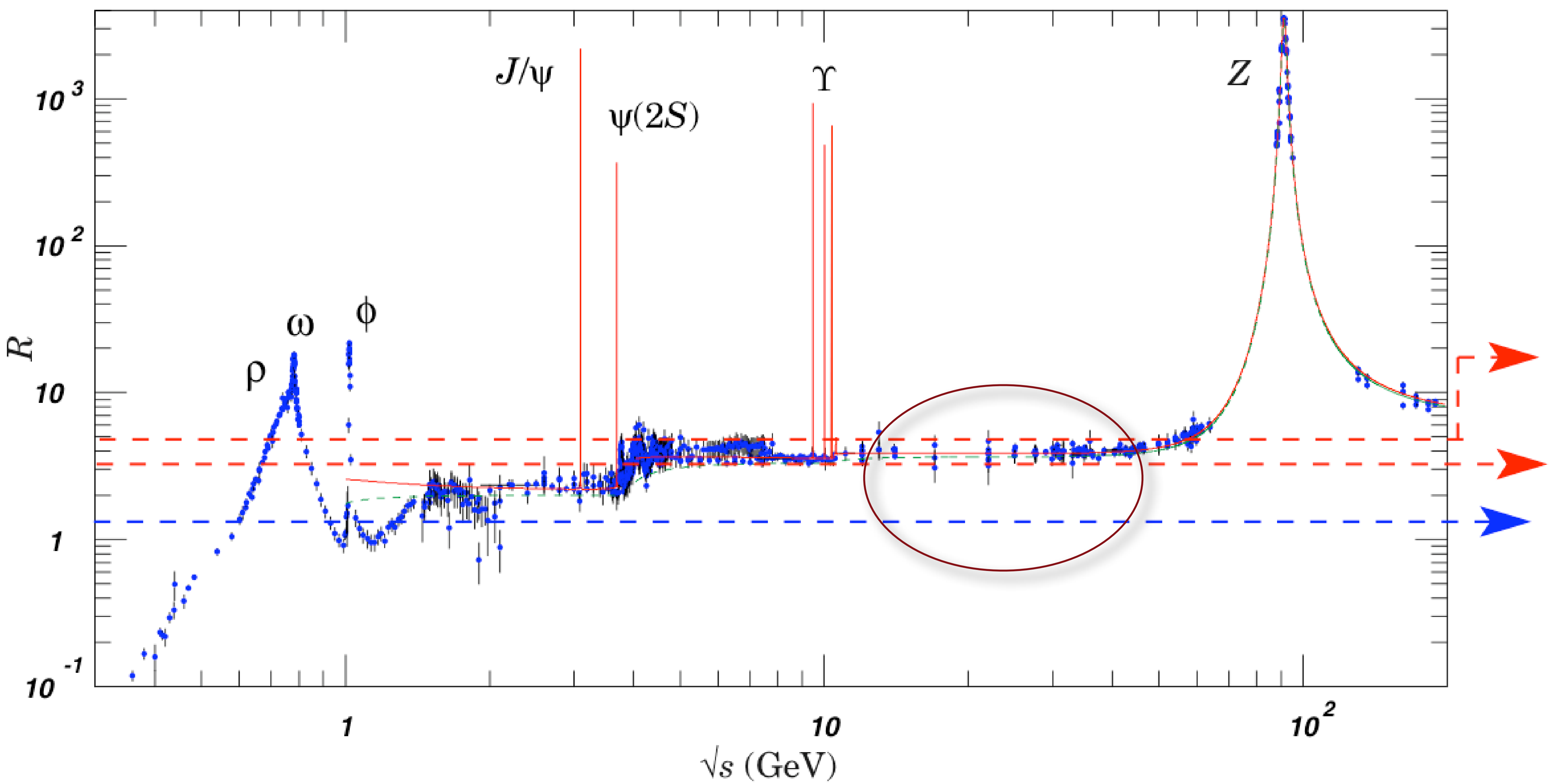
$$R = N_c \sum_f z_f^2 = N_c \cdot [\underbrace{(\tfrac{2}{3})^2}_{u} + \underbrace{(-\tfrac{1}{3})^2}_{d} + \underbrace{(-\tfrac{1}{3})^2}_{s} + \underbrace{(\tfrac{2}{3})^2}_{c} + \underbrace{(-\tfrac{1}{3})^2}_{b}] = N_c \cdot \frac{11}{9}$$

Result

For E_{CM} below the Z peak and above the Υ resonance we expect:

$$R = N_c \sum_f z_f^2 = N_c \cdot [(\tfrac{2}{3})^2 + (-\tfrac{1}{3})^2 + (-\tfrac{1}{3})^2 + (\tfrac{2}{3})^2 + (-\tfrac{1}{3})^2] = N_c \cdot \frac{11}{9}$$

u d s c b



6 quarks,
3 colours (45/9)

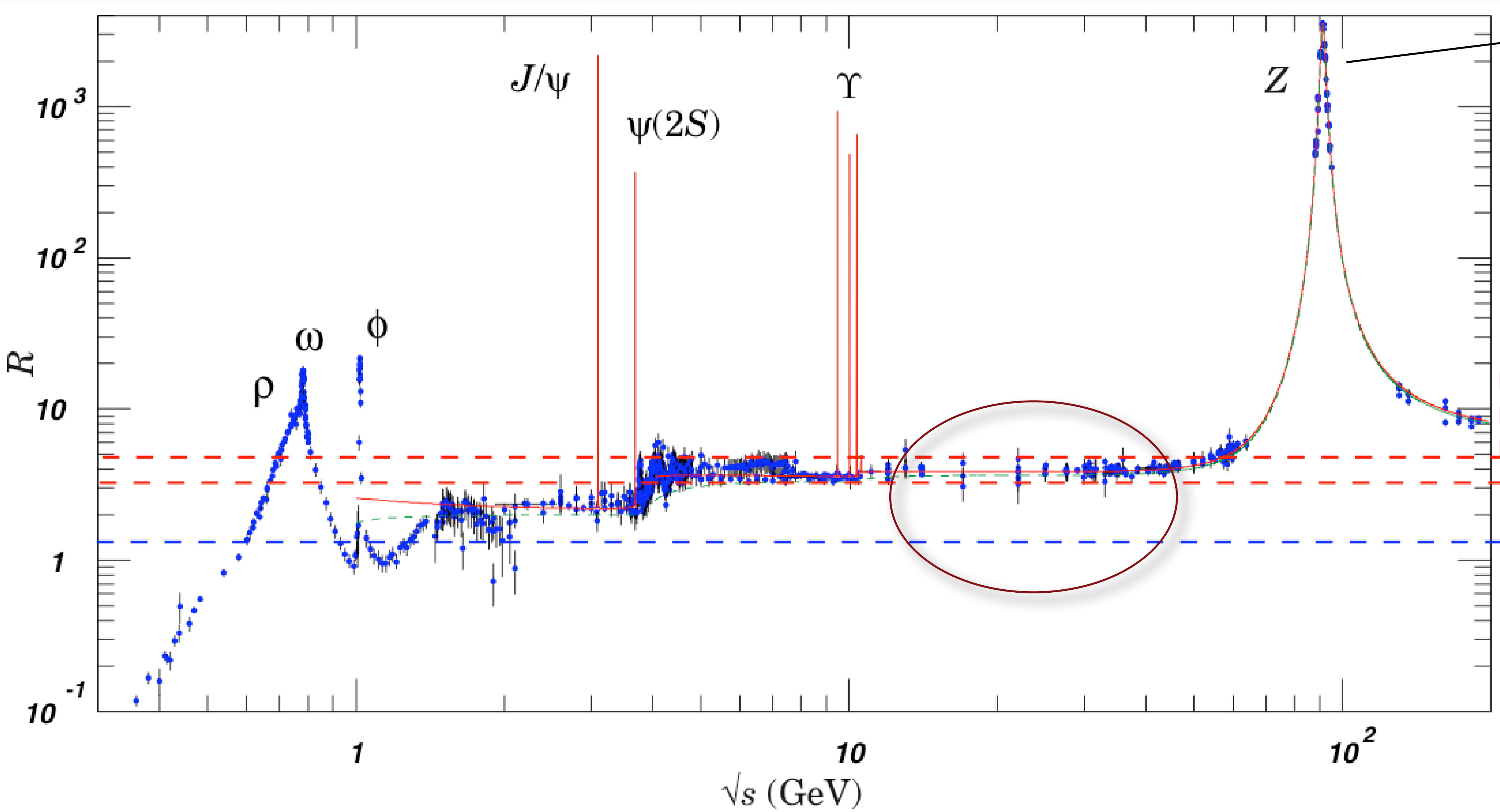
5 quarks,
3 colours (33/9)

5 quarks,
1 colour (11/9)

Result

For E_{CM} below the Z peak and above the Υ resonance we expect:

$$R = N_c \sum_f z_f^2 = N_c \cdot [\underbrace{(\tfrac{2}{3})^2}_{u} + \underbrace{(-\tfrac{1}{3})^2}_{d} + \underbrace{(-\tfrac{1}{3})^2}_{s} + \underbrace{(\tfrac{2}{3})^2}_{c} + \underbrace{(-\tfrac{1}{3})^2}_{b}] = N_c \cdot \frac{11}{9}$$



At Z peak :

have to include also couplings to Z

6 quarks, 3 colours (45/9)

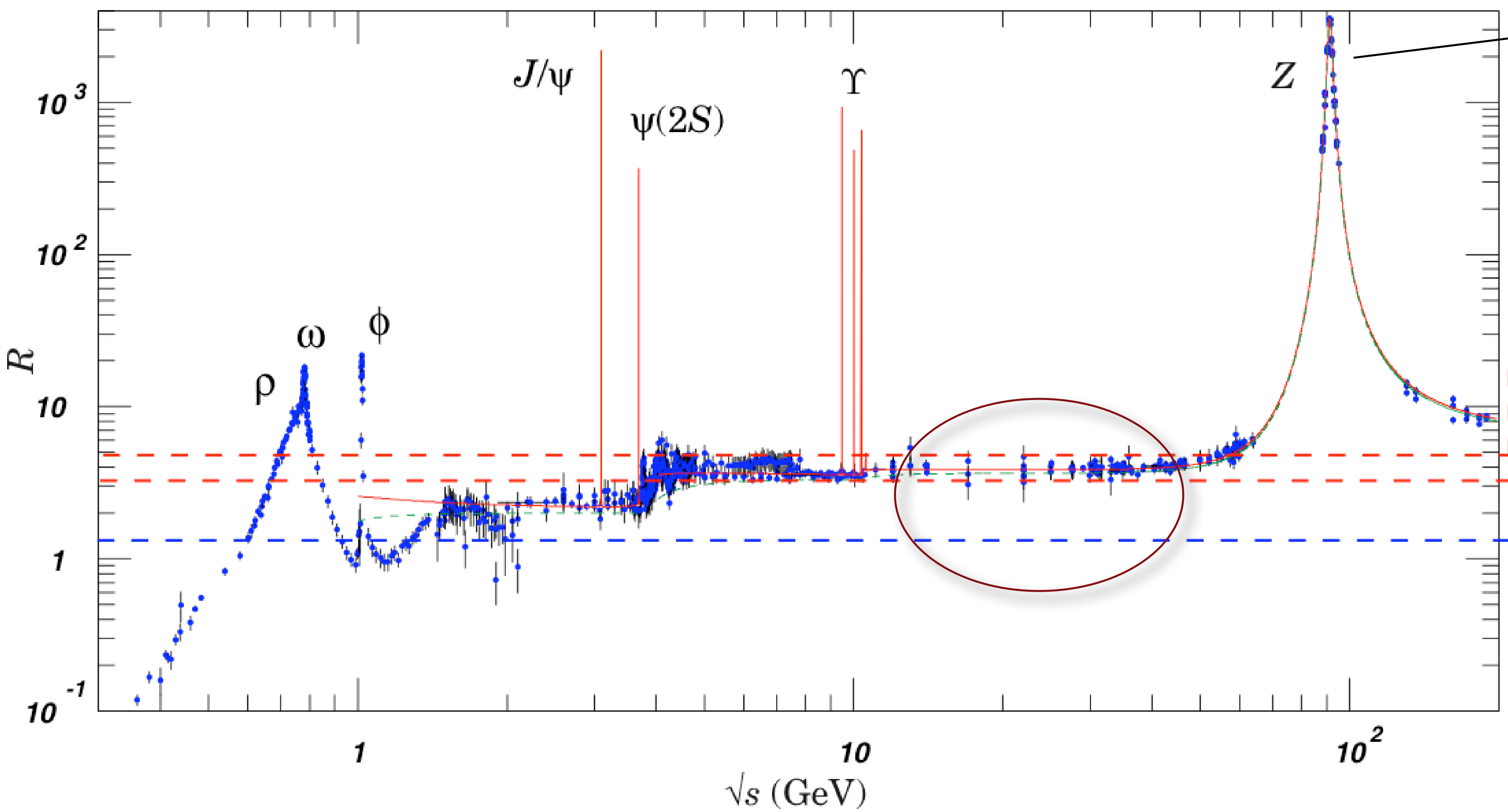
5 quarks, 3 colours (33/9)

5 quarks, 1 colour (11/9)

Result

For E_{CM} below the Z peak and above the Υ resonance we expect:

$$R = N_c \sum_f z_f^2 = N_c \cdot [\underbrace{(\tfrac{2}{3})^2}_{u} + \underbrace{(-\tfrac{1}{3})^2}_{d} + \underbrace{(-\tfrac{1}{3})^2}_{s} + \underbrace{(\tfrac{2}{3})^2}_{c} + \underbrace{(-\tfrac{1}{3})^2}_{b}] = N_c \cdot \frac{11}{9}$$



At Z peak :

have to include also couplings to Z

6 quarks,
3 colours (45/9)

5 quarks,
3 colours (33/9)

5 quarks,
1 colour (11/9)

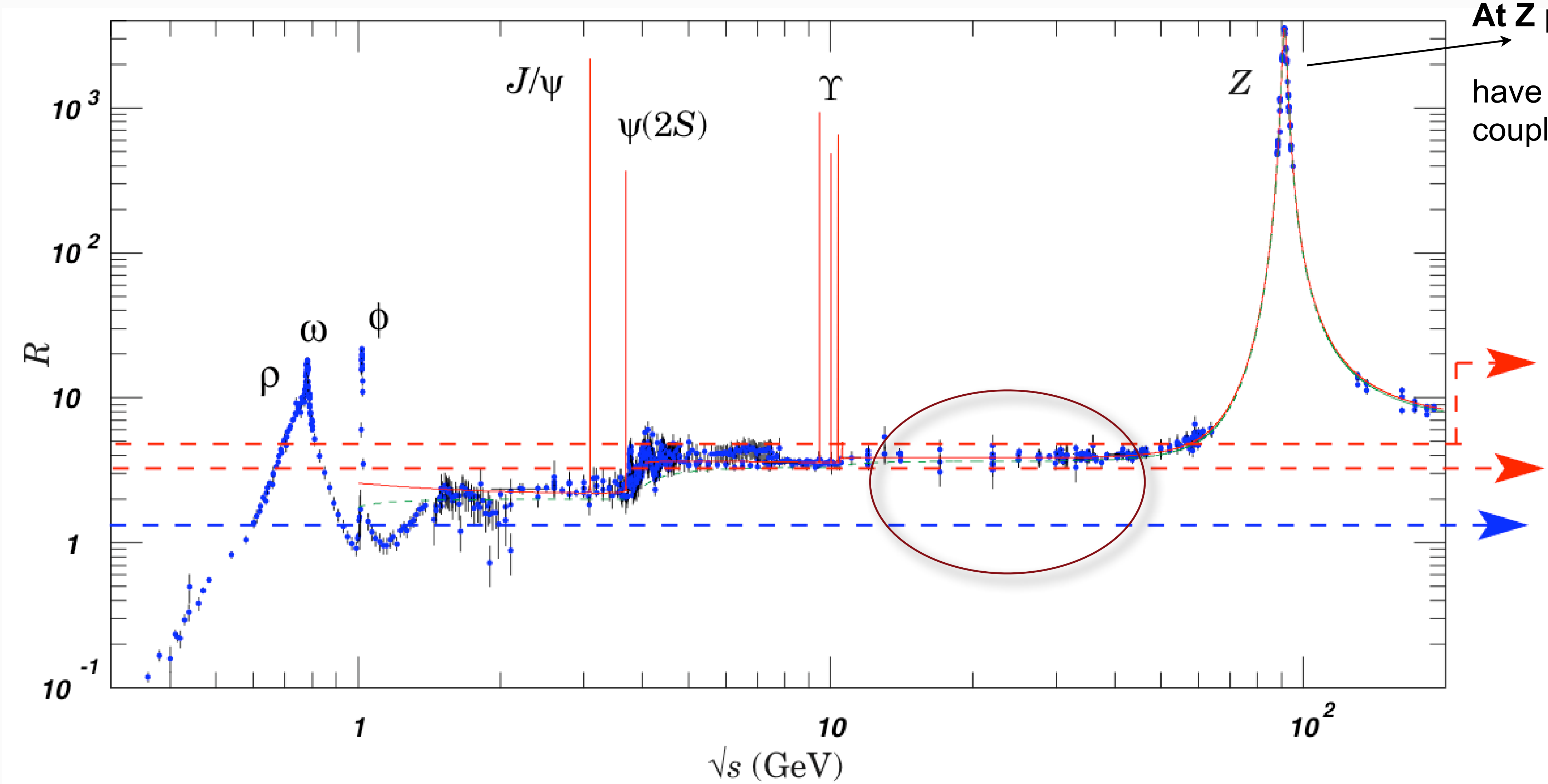
- Confirmation of : Number of colours = 3 !

Result

For E_{CM} below the Z peak and above the Υ resonance we expect:

$$R = N_c \sum_f z_f^2 = N_c \cdot [(\tfrac{2}{3})^2 + (-\tfrac{1}{3})^2 + (-\tfrac{1}{3})^2 + (\tfrac{2}{3})^2 + (-\tfrac{1}{3})^2] = N_c \cdot \frac{11}{9}$$

u d s c b



- Confirmation of : Number of colours = 3 !

Note : small remaining difference : because of QCD correction (gluon radiation) = 1 + α_s / π

Uncertainties

- Just having a "counting result" is not all, there's lot more to do!

Uncertainties

- Just having a "counting result" is not all, there's lot more to do!
- Statistical error
  - We saw 2 muon events, could easily have been 1 or 3
  - Those fluctuations go like the square-root of the number of events

$$BR\left(Z^0 \to \mu^+\mu^-\right) = \frac{N_{\mu\mu}}{N_{total}} \pm \frac{\sqrt{N_{\mu\mu}}}{N_{total}}$$

  - To reduce this uncertainty, you need to record lots (millions) of events in the detector, and process them

see also lecture by G. Cowan

Uncertainties

- Just having a "counting result" is not all, there's lot more to do!
- Statistical error
  - We saw 2 muon events, could easily have been 1 or 3
  - Those fluctuations go like the square-root of the number of events

$$BR\left(Z^0 \rightarrow \mu^+\mu^-\right) = \frac{N_{\mu\mu}}{N_{total}} \pm \frac{\sqrt{N_{\mu\mu}}}{N_{total}}$$

  - To reduce this uncertainty, you need to record lots (millions) of events in the detector, and process them
- Systematic error
  - What if you only see 50% of the $\mu^+\mu^-$ events?
    - because of event selection (cut), detector imperfections, poor understanding, etc.

"efficiency"

$$N_{\mu\mu_{seen}} = \varepsilon N_{\mu\mu}$$

Uncertainties

ETH Institute for Particle Physics

see also lecture by G. Cowan

- Just having a "counting result" is not all, there's lot more to do!
- Statistical error
  - We saw 2 muon events, could easily have been 1 or 3
  - Those fluctuations go like the square-root of the number of events

$$BR\left(Z^0 \rightarrow \mu^+\mu^-\right) = \frac{N_{\mu\mu}}{N_{total}} \pm \frac{\sqrt{N_{\mu\mu}}}{N_{total}}$$

  - To reduce this uncertainty, you need to record lots (millions) of events in the detector, and process them
- Systematic error
  - What if you only see 50% of the $\mu^+\mu^-$ events?
    - because of event selection (cut), detector imperfections, poor understanding, etc.

"efficiency"

$$N_{\mu\mu\,\mathrm{seen}} = \varepsilon N_{\mu\mu}$$

$$BR\left(Z^0 \rightarrow \mu^+\mu^-\right) = \frac{N_{\mathrm{seen}}/\varepsilon}{N_{total}}$$

$$\varepsilon = 0.50 \pm 0.05$$

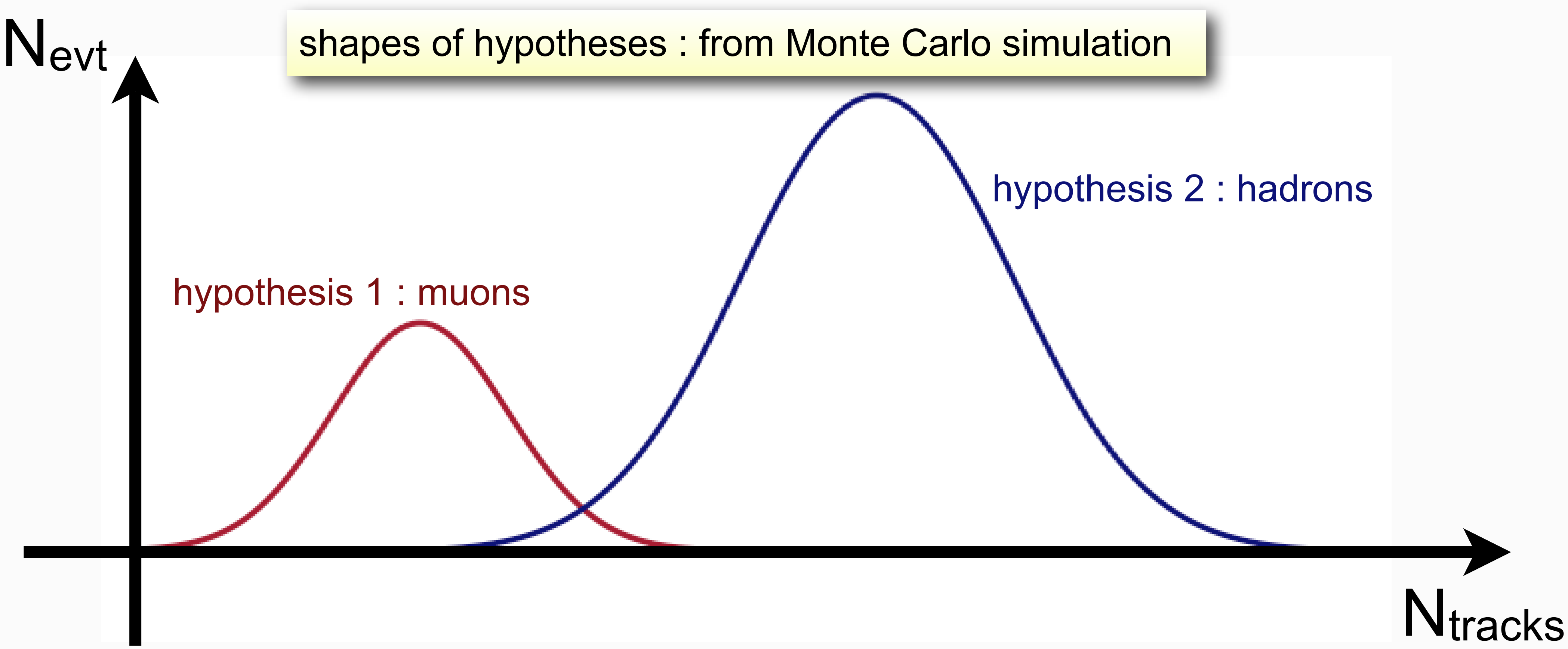
- from statistical error of detector simulation
- imperfect modeling of geometry in simulation
- model of muon interactions in simulation, etc

Event selection and all that jazz

- Event per event have to decide how to categorize it
  - eg. do we call it a muon event, or a hadronic event?
  - how do we estimate the efficiency?
  - Define an event selection, eg. “cut-based”
  - see statistics lectures, *hypothesis testing* etc...

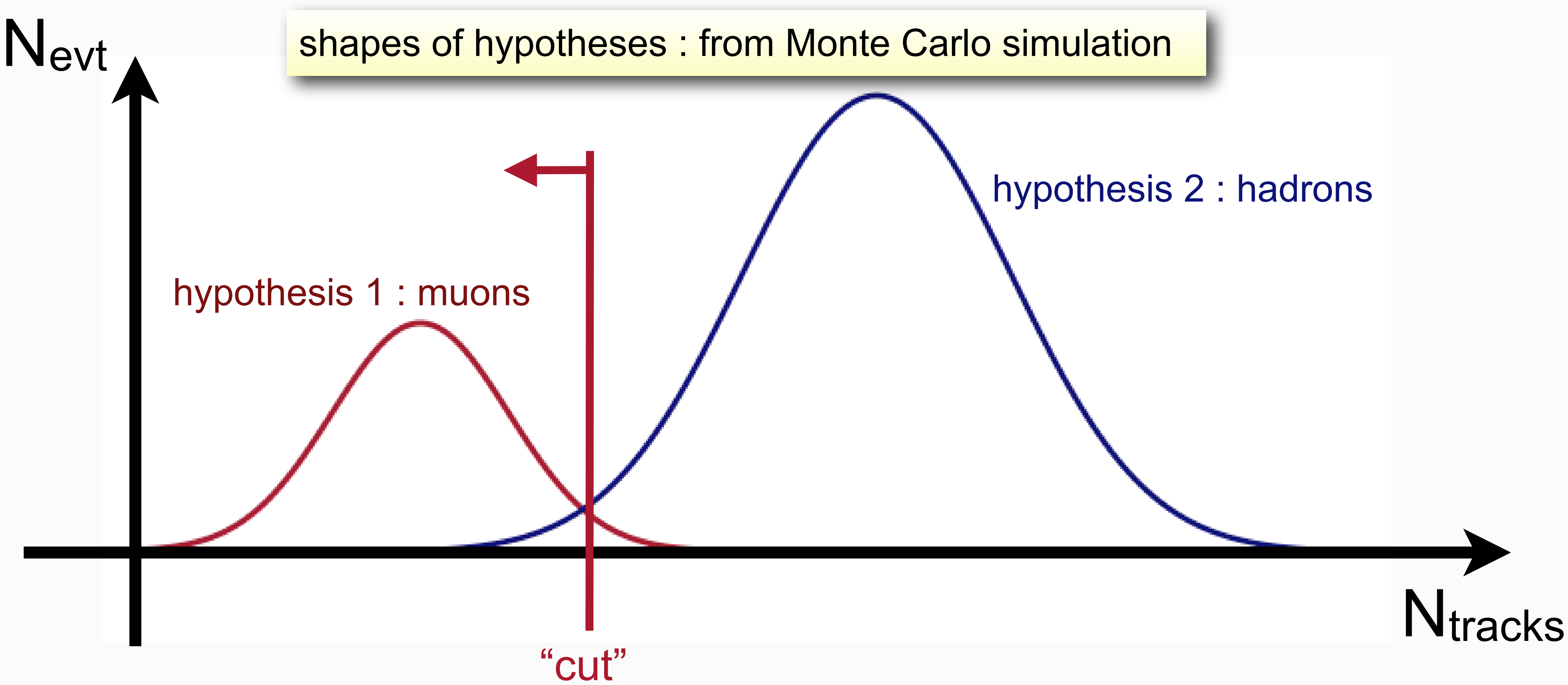
Event selection and all that jazz

- Event per event have to decide how to categorize it
  - eg. do we call it a muon event, or a hadronic event?
  - how do we estimate the efficiency?
  - Define an event selection, eg. “cut-based”
  - see statistics lectures, *hypothesis testing* etc...



Event selection and all that jazz

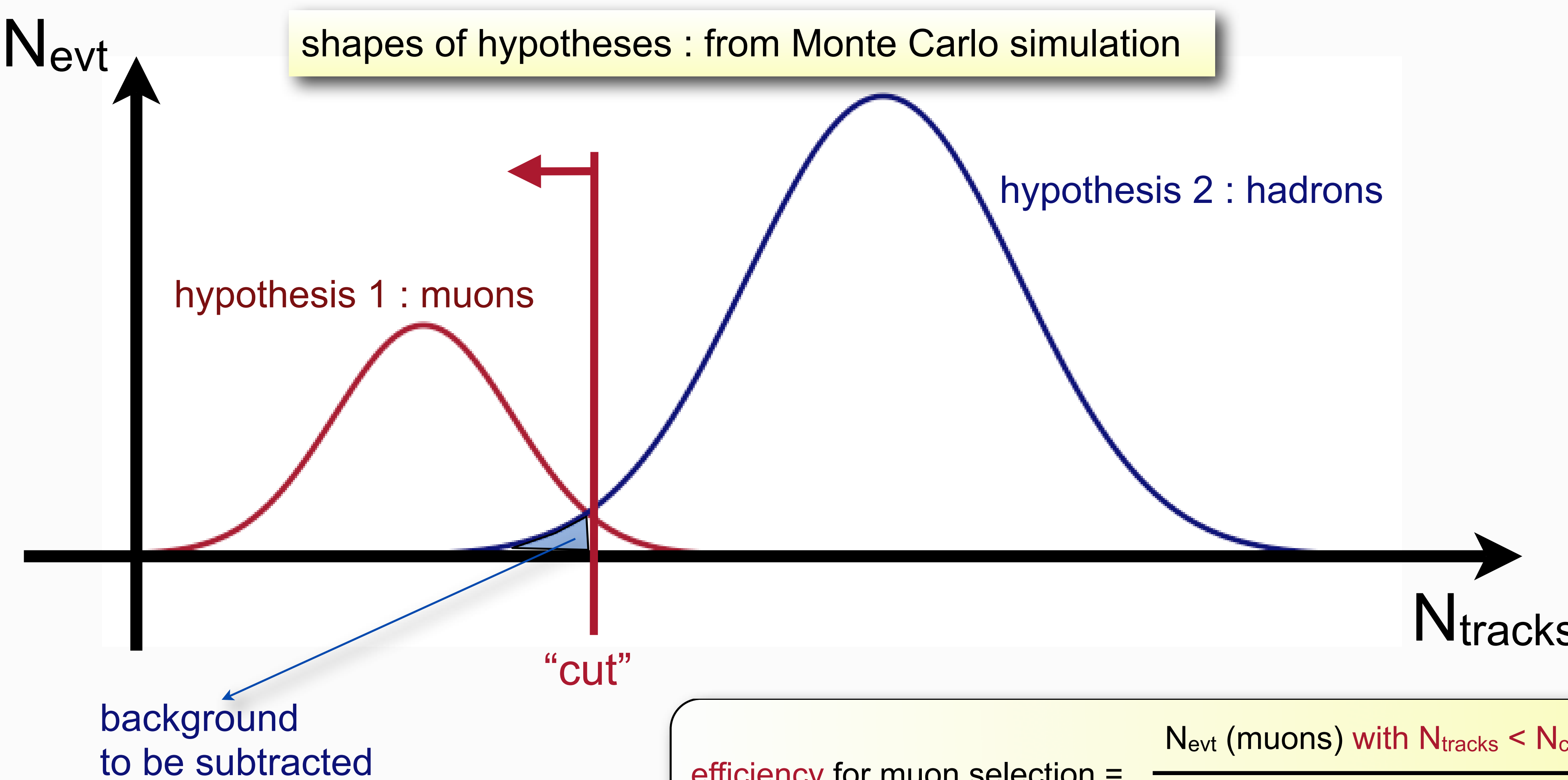
- Event per event have to decide how to categorize it
  - eg. do we call it a muon event, or a hadronic event?
  - how do we estimate the efficiency?
  - Define an event selection, eg. “cut-based”
  - see statistics lectures, *hypothesis testing* etc...



$$\text{efficiency for muon selection} = \frac{N_{evt}\text{ (muons) with } N_{tracks} < N_{cut}}{\text{all } N_{evt}\text{ (muons)}}$$

Event selection and all that jazz

- Event per event have to decide how to categorize it
  - eg. do we call it a muon event, or a hadronic event?
  - how do we estimate the efficiency?
  - Define an event selection, eg. “cut-based”
  - see statistics lectures, *hypothesis testing* etc...



$$\text{efficiency for muon selection} = \frac{N_{evt}\,(\text{muons})\ \text{with}\ N_{tracks} < N_{cut}}{\text{all}\ N_{evt}\,(\text{muons})}$$

see also lecture by G. Cowan

A “more complicated” example

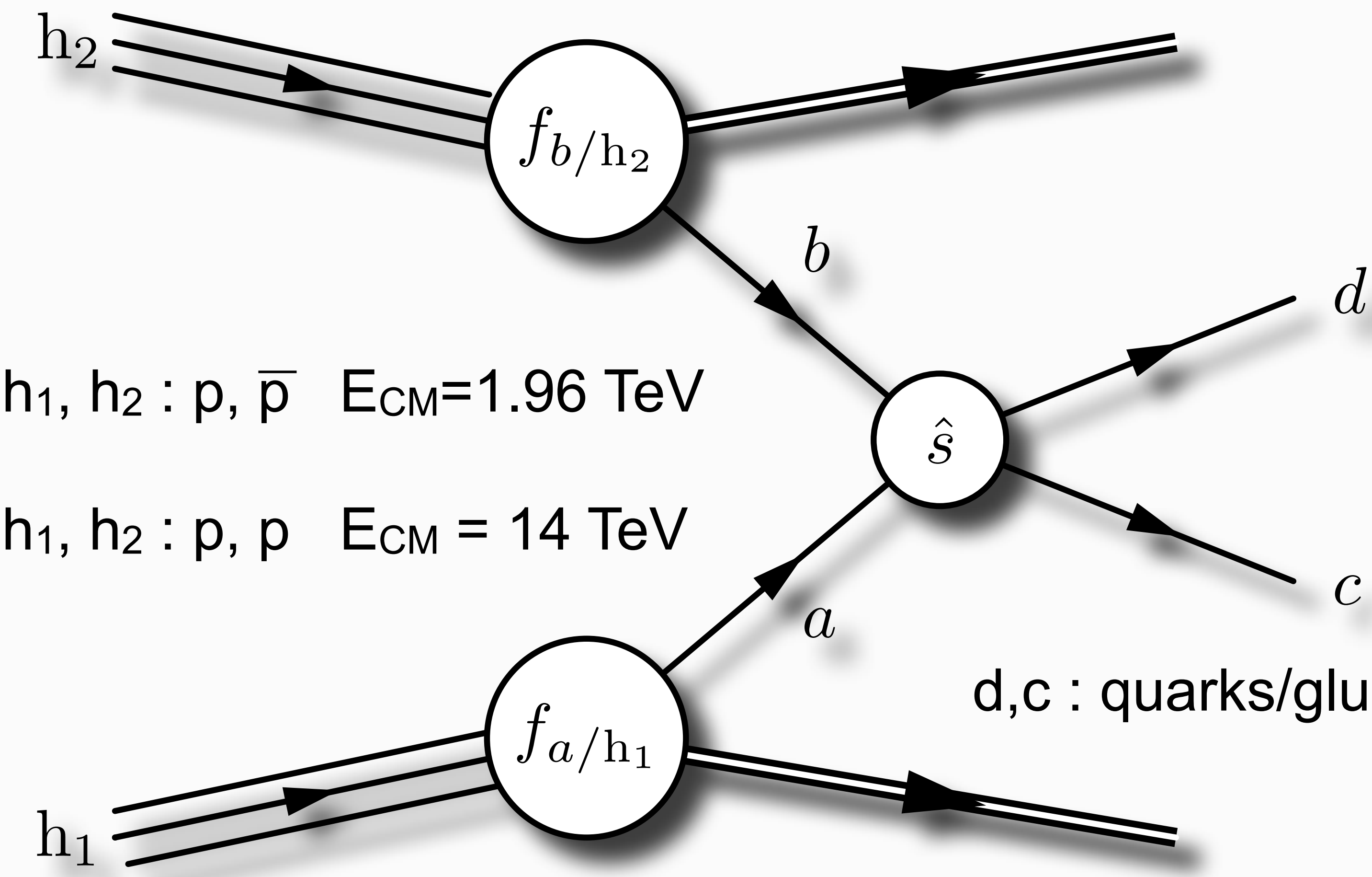
“The greater the obstacle, the more glory in overcoming it.” (Moliere)

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

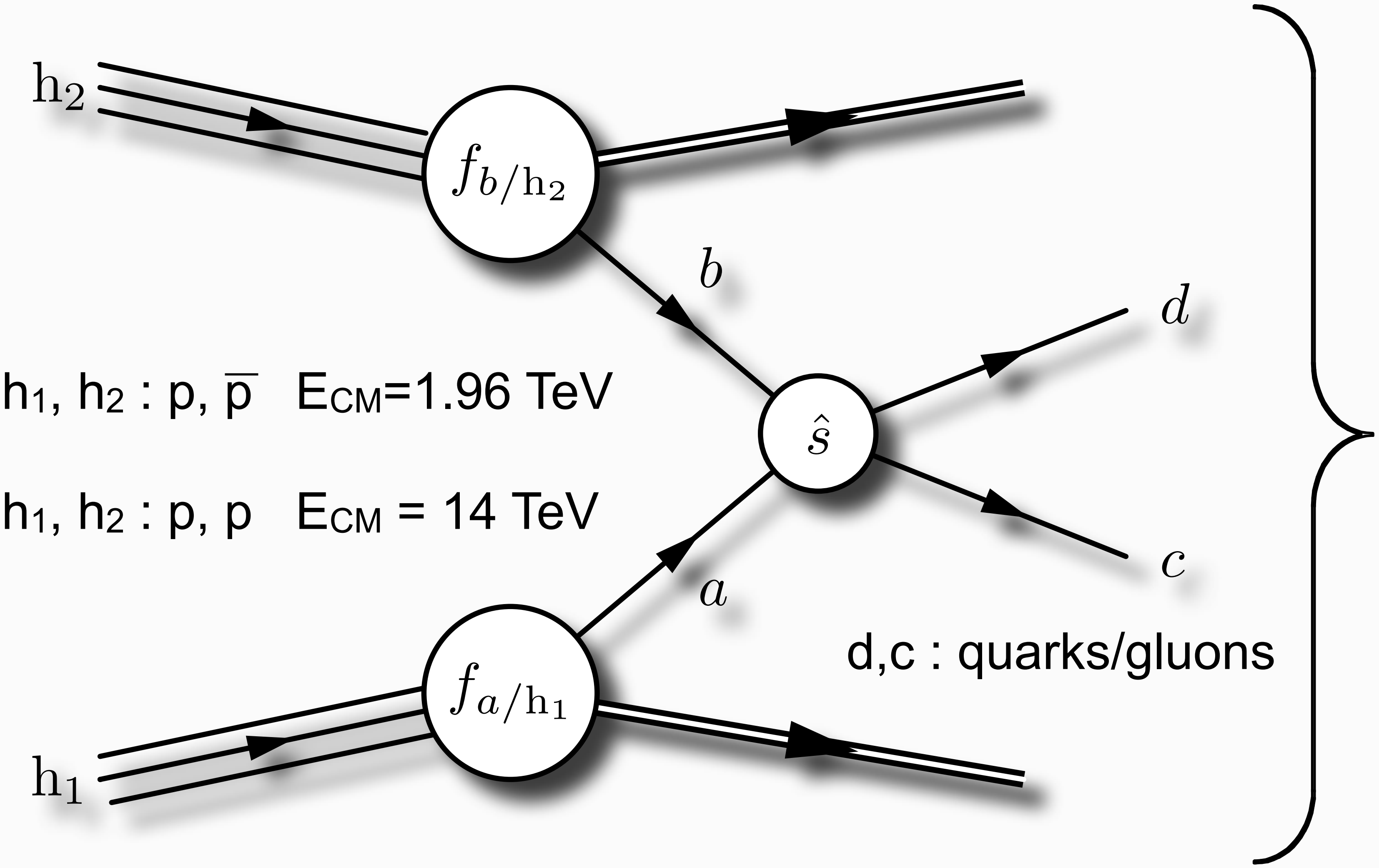
d,c : quarks/gluons

JET production at hadron colliders

ETH Institute for Particle Physics

see also lecture by B. Heinemann

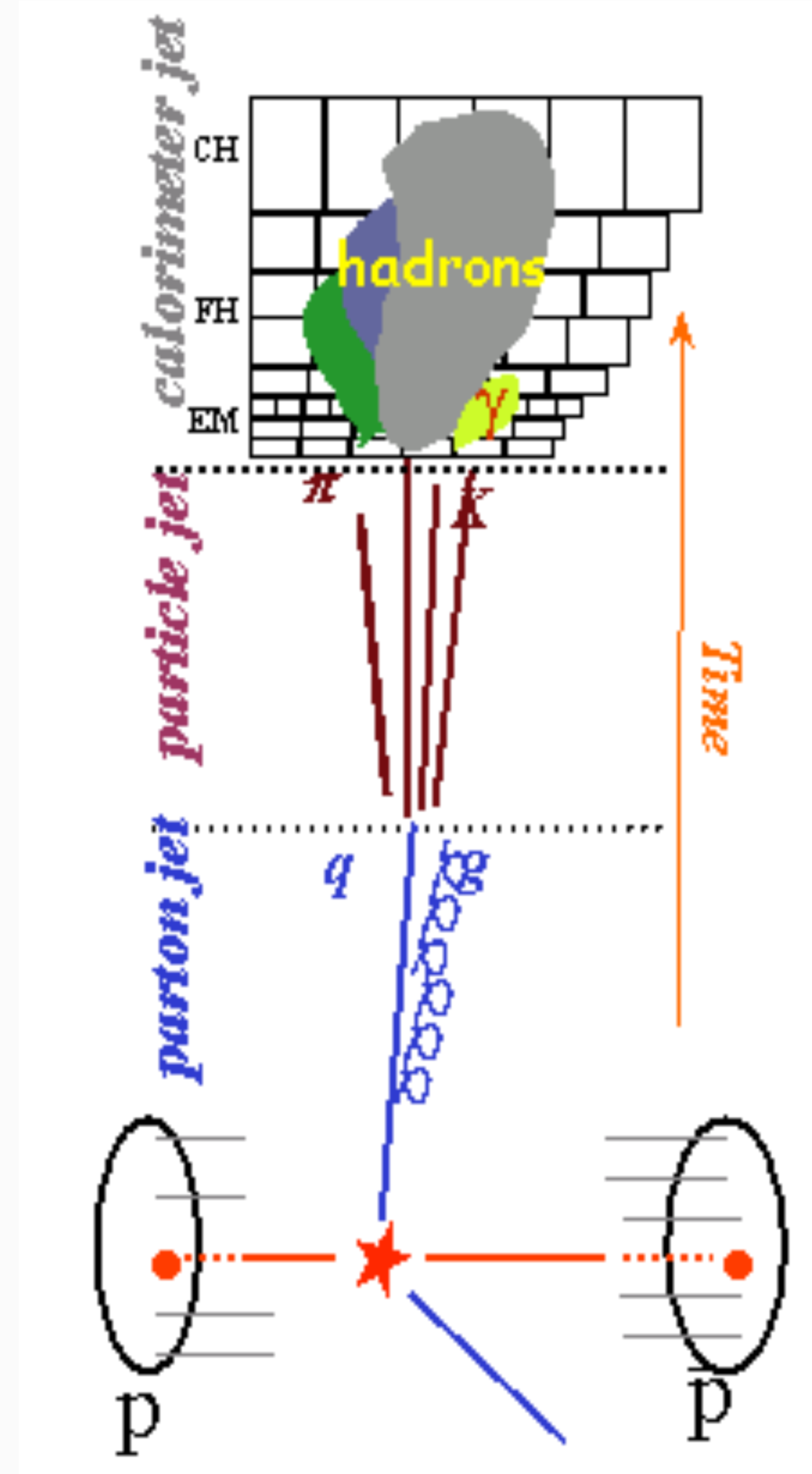
- at the Tevatron, or in the future at the LHC



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

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

d,c : quarks/gluons

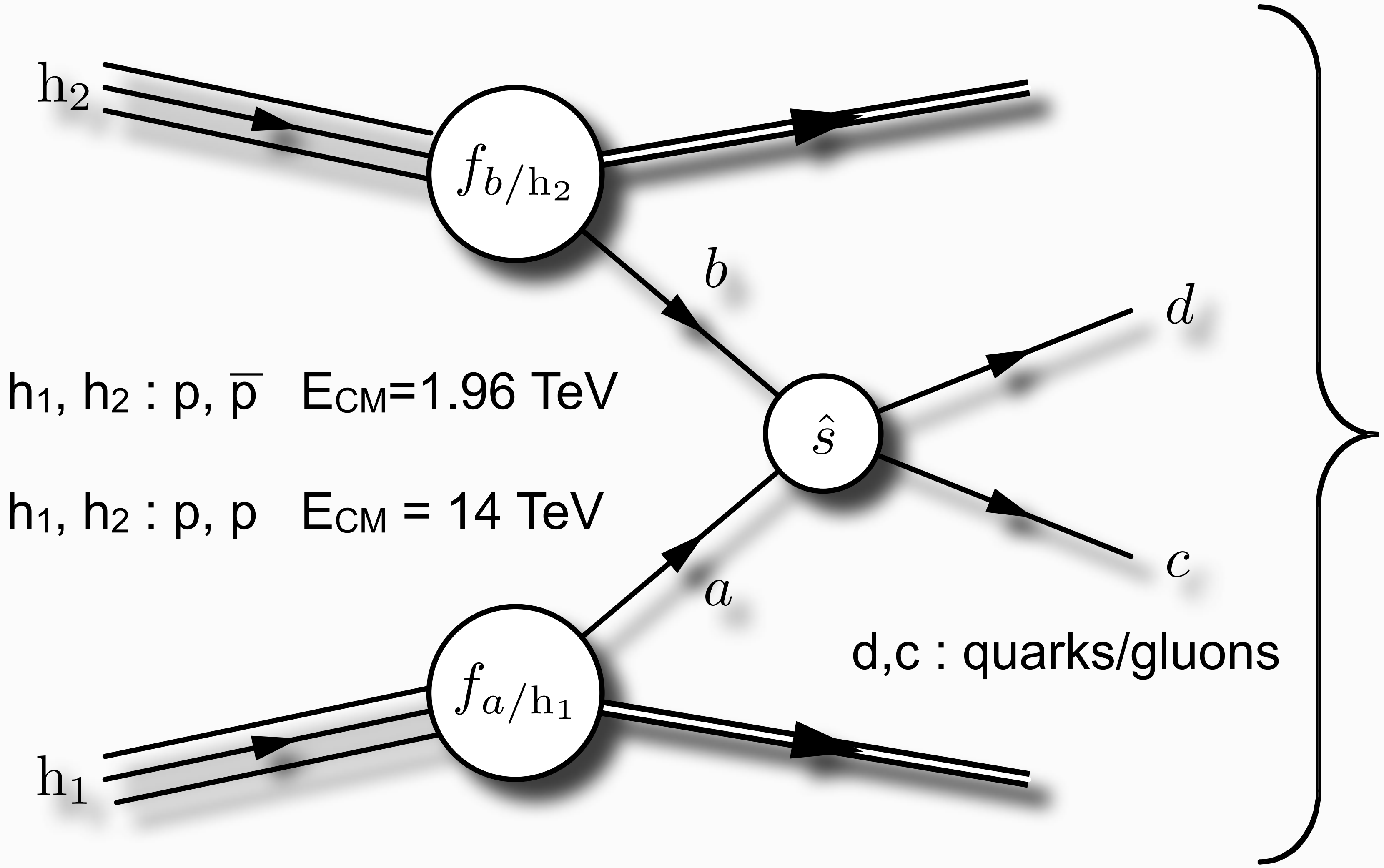


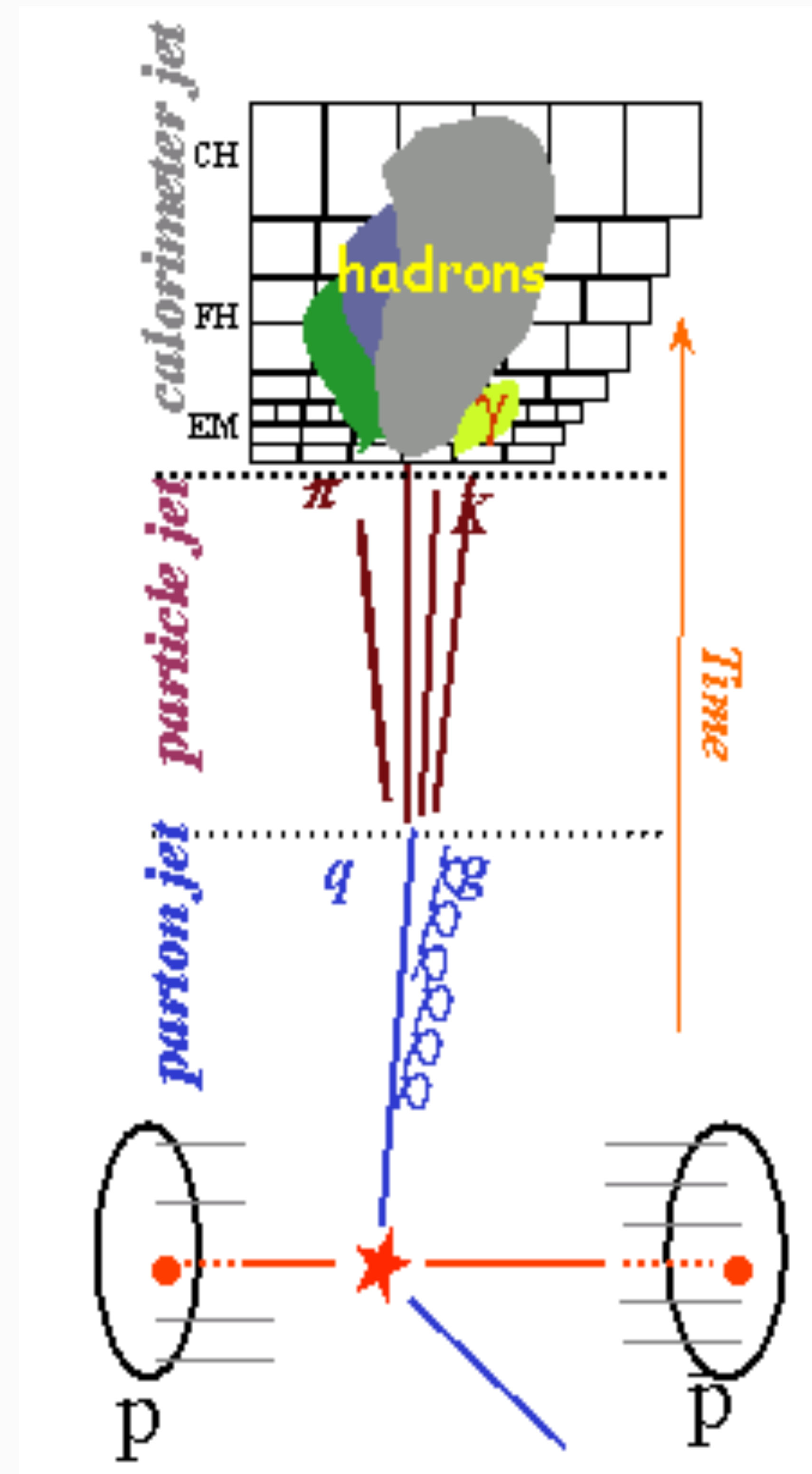
JET production at hadron colliders

ETH Institute for Particle Physics

see also lecture by B. Heinemann

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



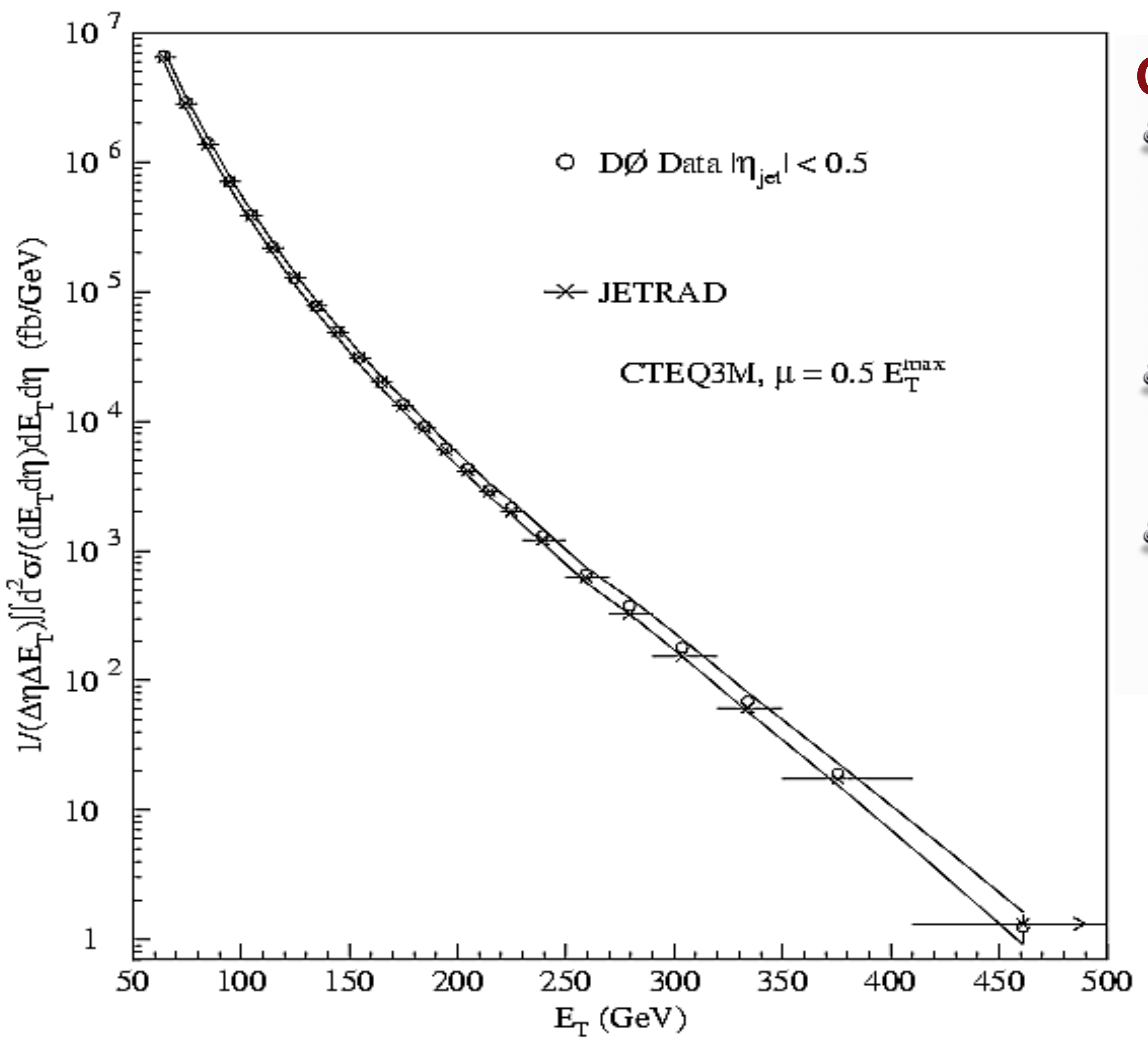


Goal

- 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

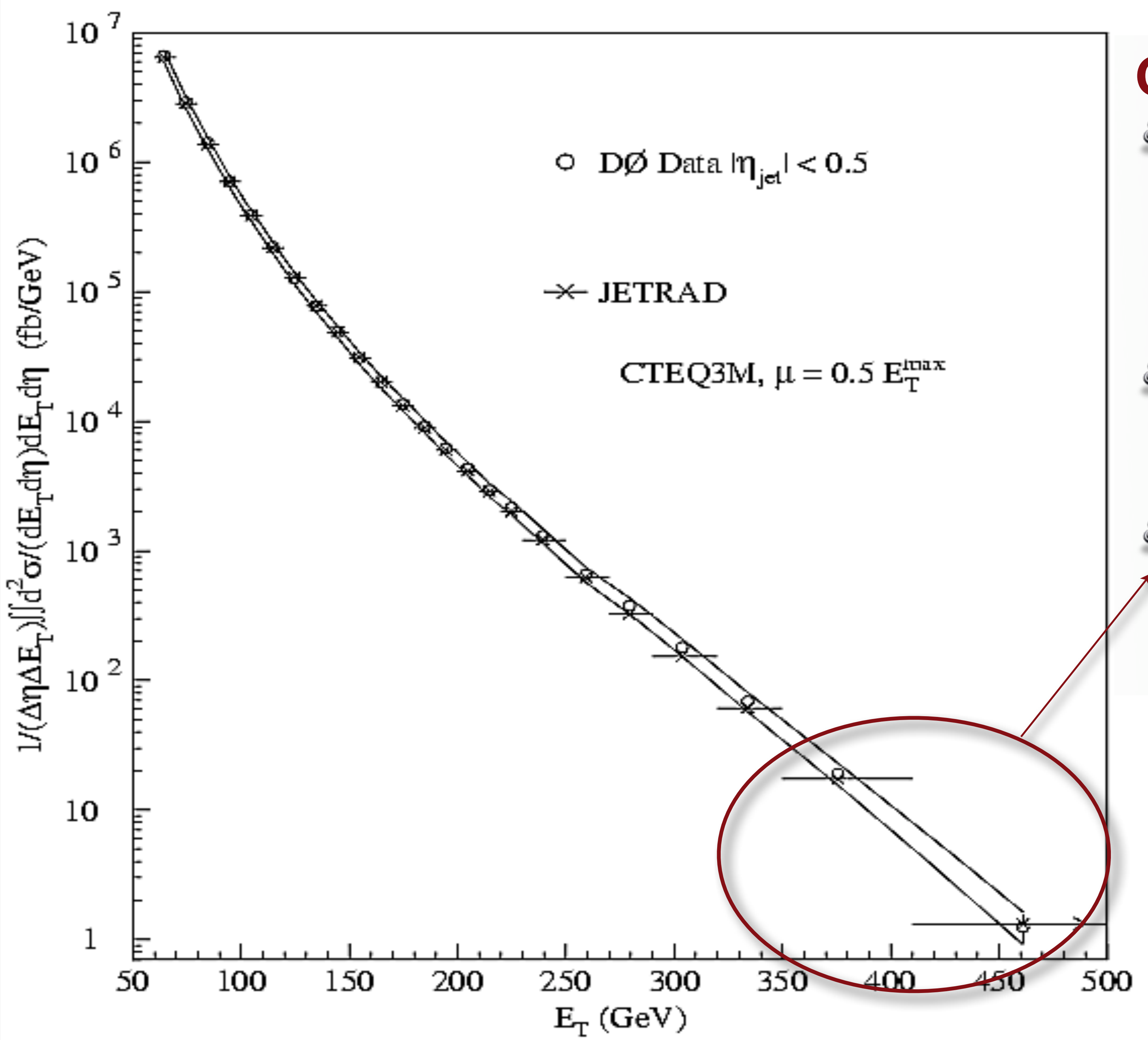


Goal

- 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?

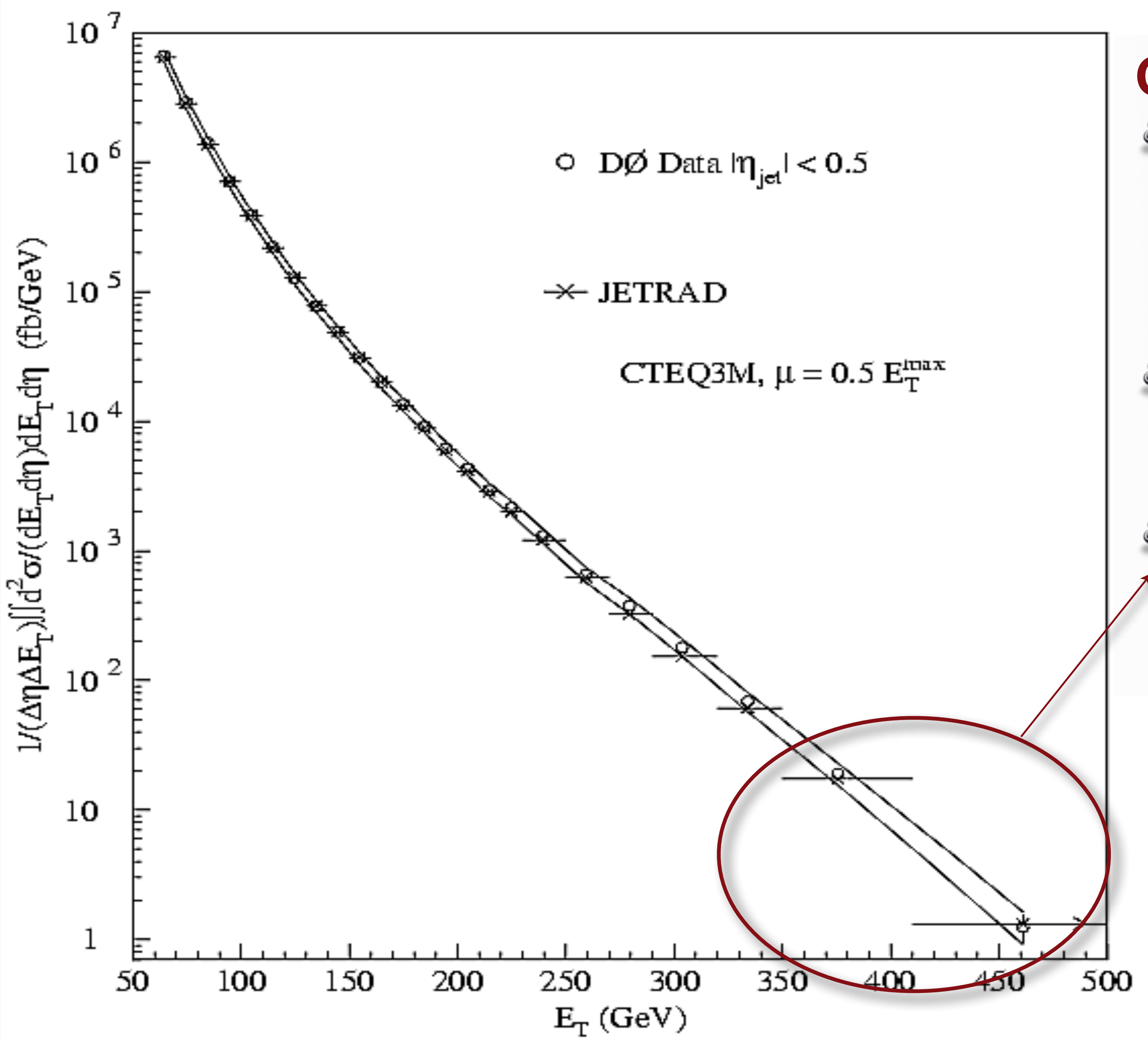
ETH Institute for Particle Physics



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

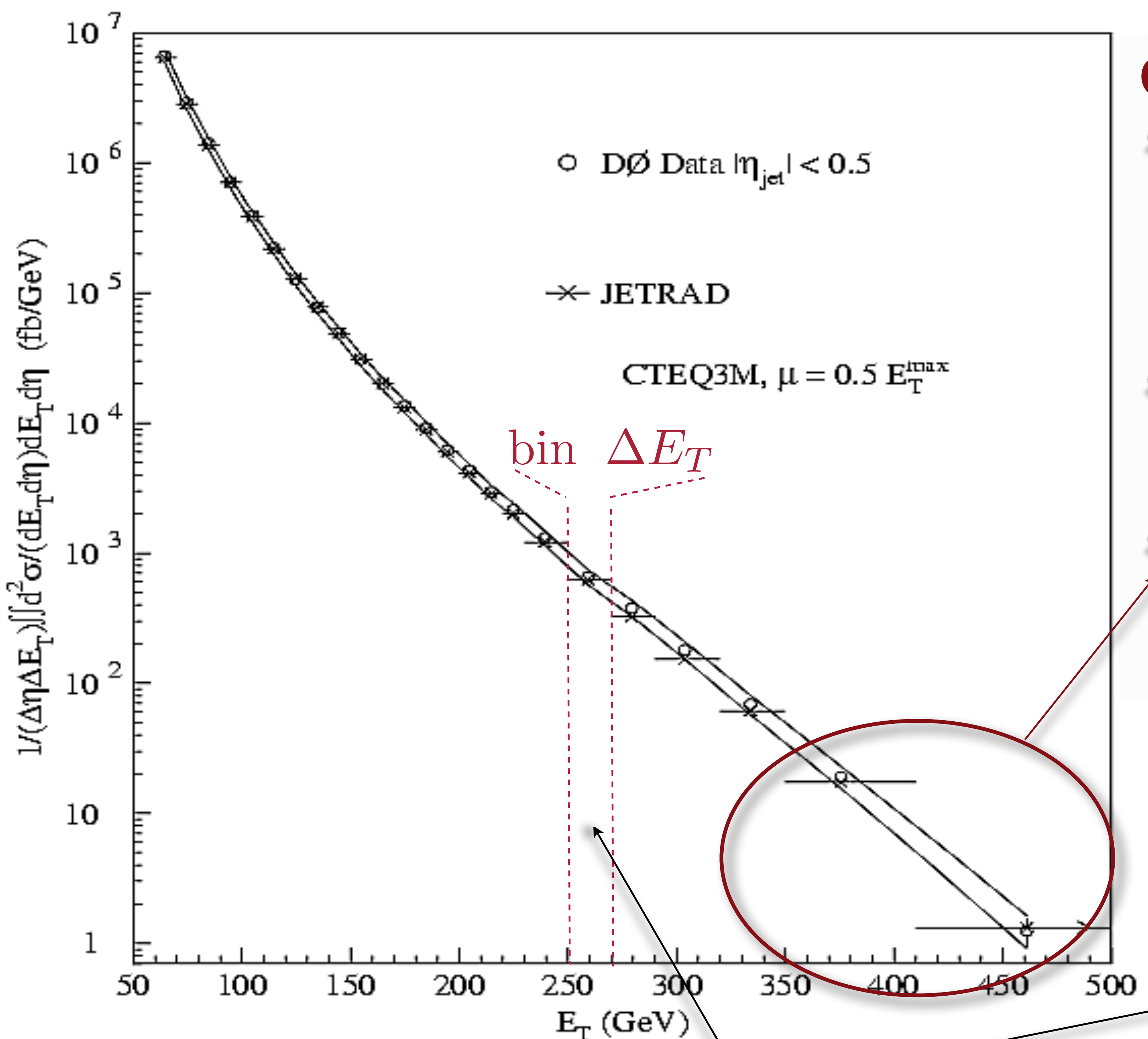
- measure cross section (probability) that jets are produced with a certain transverse energy E_T, within a certain rapidity range
- Test of perturbative QCD, over many orders of magnitude!
- Look at very high energy tail, new physics could show up there in form of excess (eg. sub-structure of quarks?)

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?

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?)

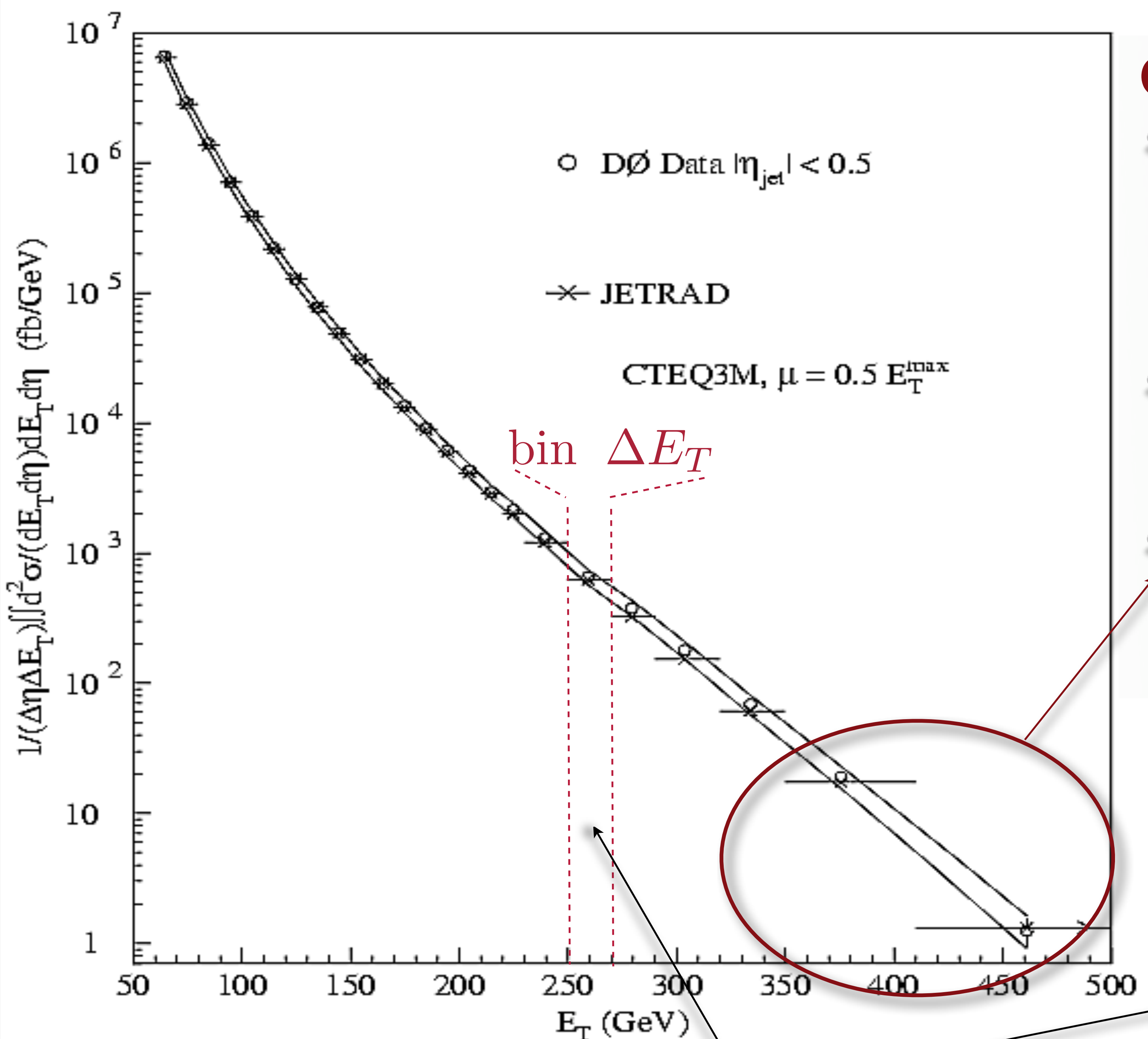
can be calculated in pert. QCD

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

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

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

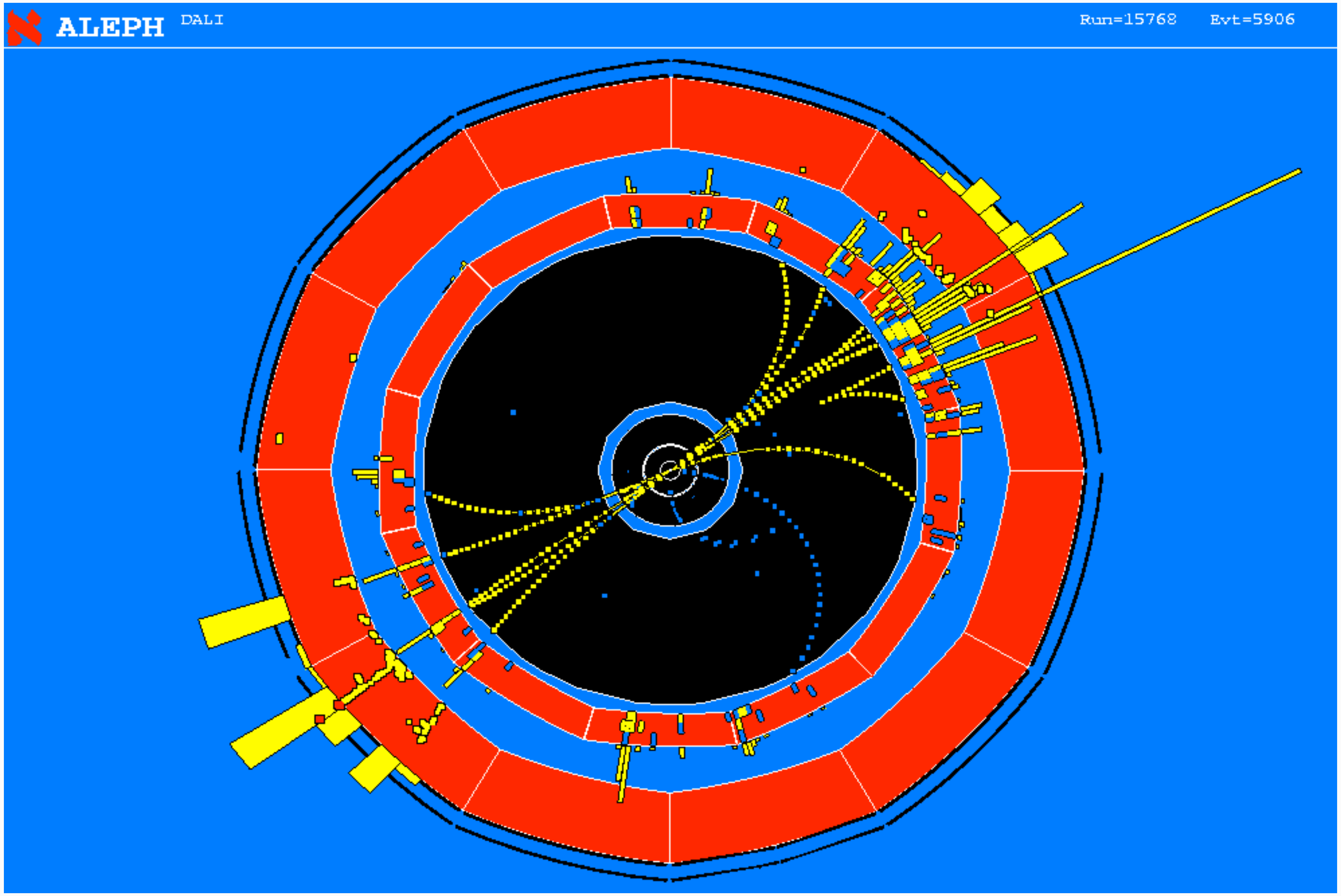
$$\left\langle \frac{d^2\sigma}{dE_T d\eta} \right\rangle = \frac{N}{\Delta E_T \Delta\eta\, \epsilon\, \mathcal{L}}$$

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

efficiency to reconstruct jets

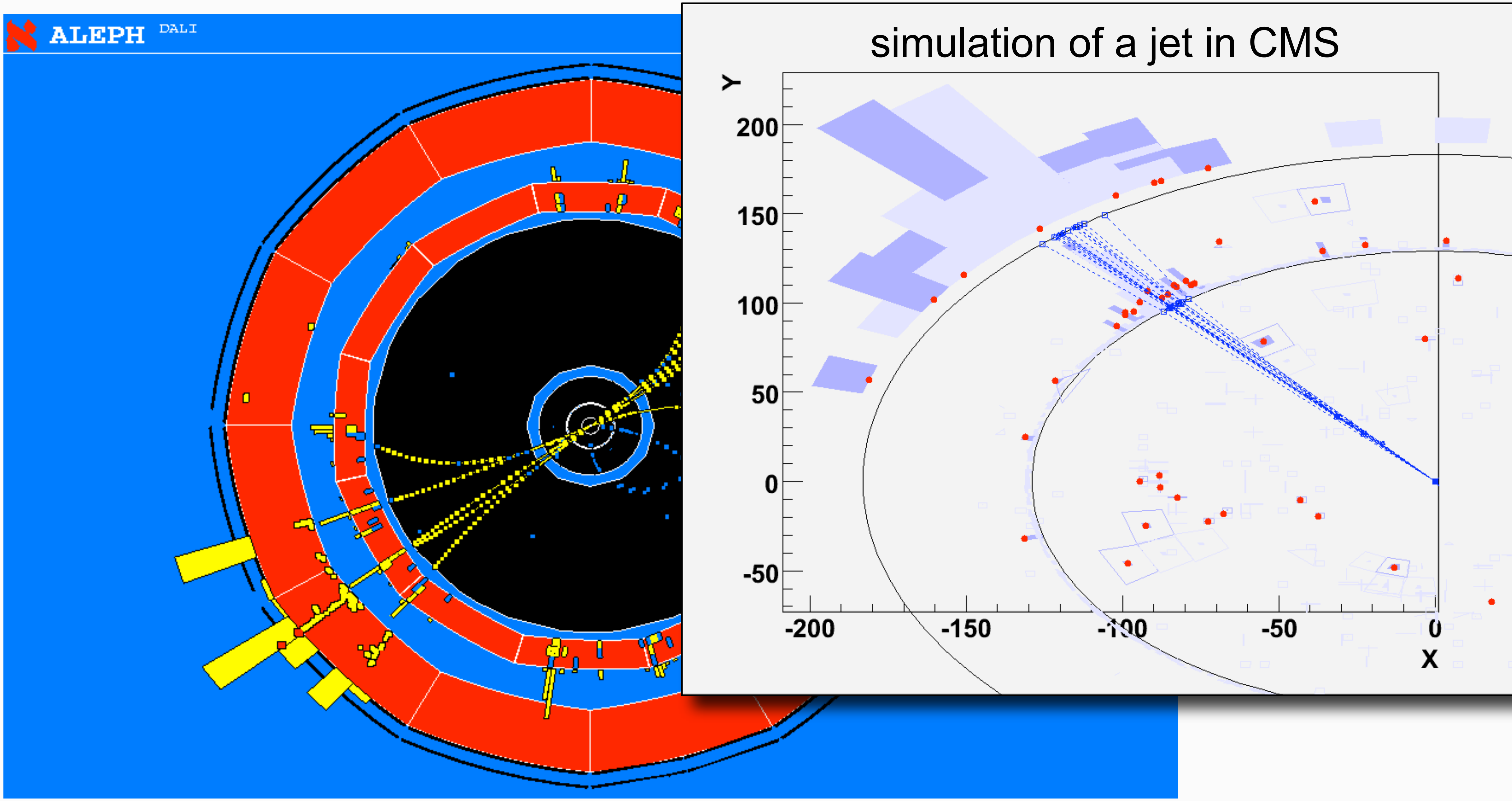
integrated accelerator luminosity

What is a jet?



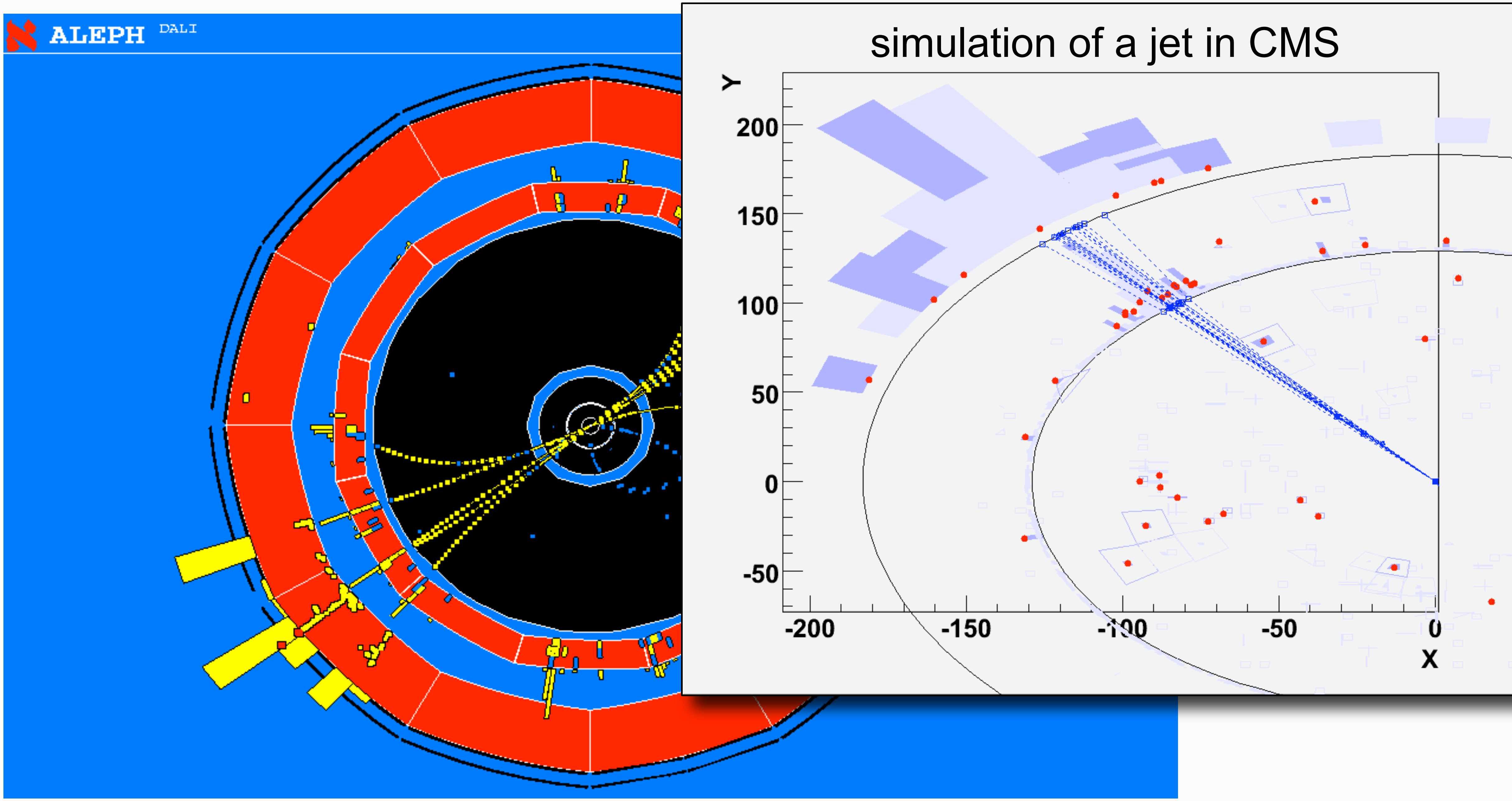
“cluster/spray of particles (tracks, calorimeter deposits) or flow of energy in a restricted angular region”

What is a jet?



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

What is a jet?

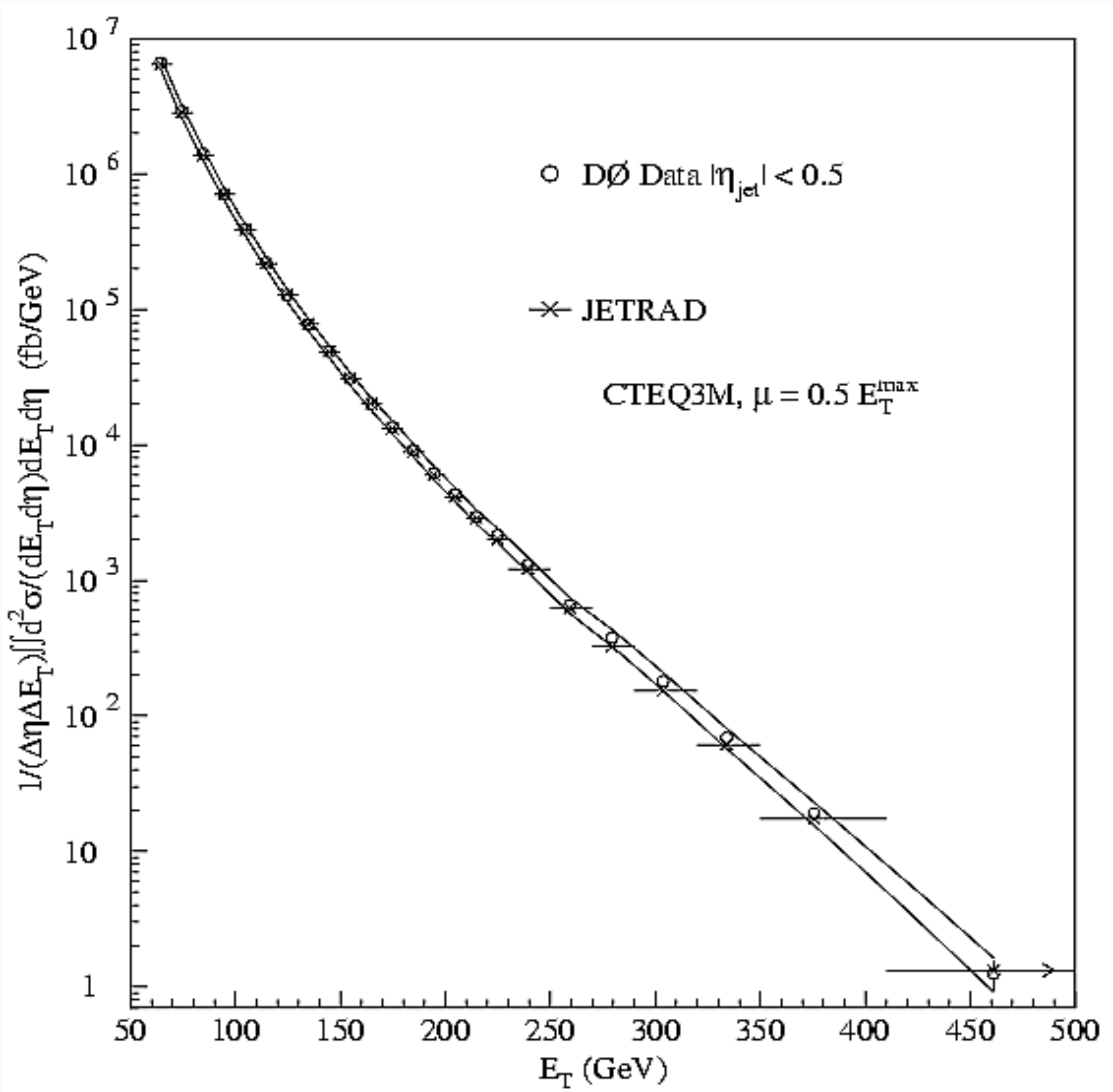


"cluster/spray of particles (tracks, calorimeter deposits) or flow of energy in a restricted angular region"

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

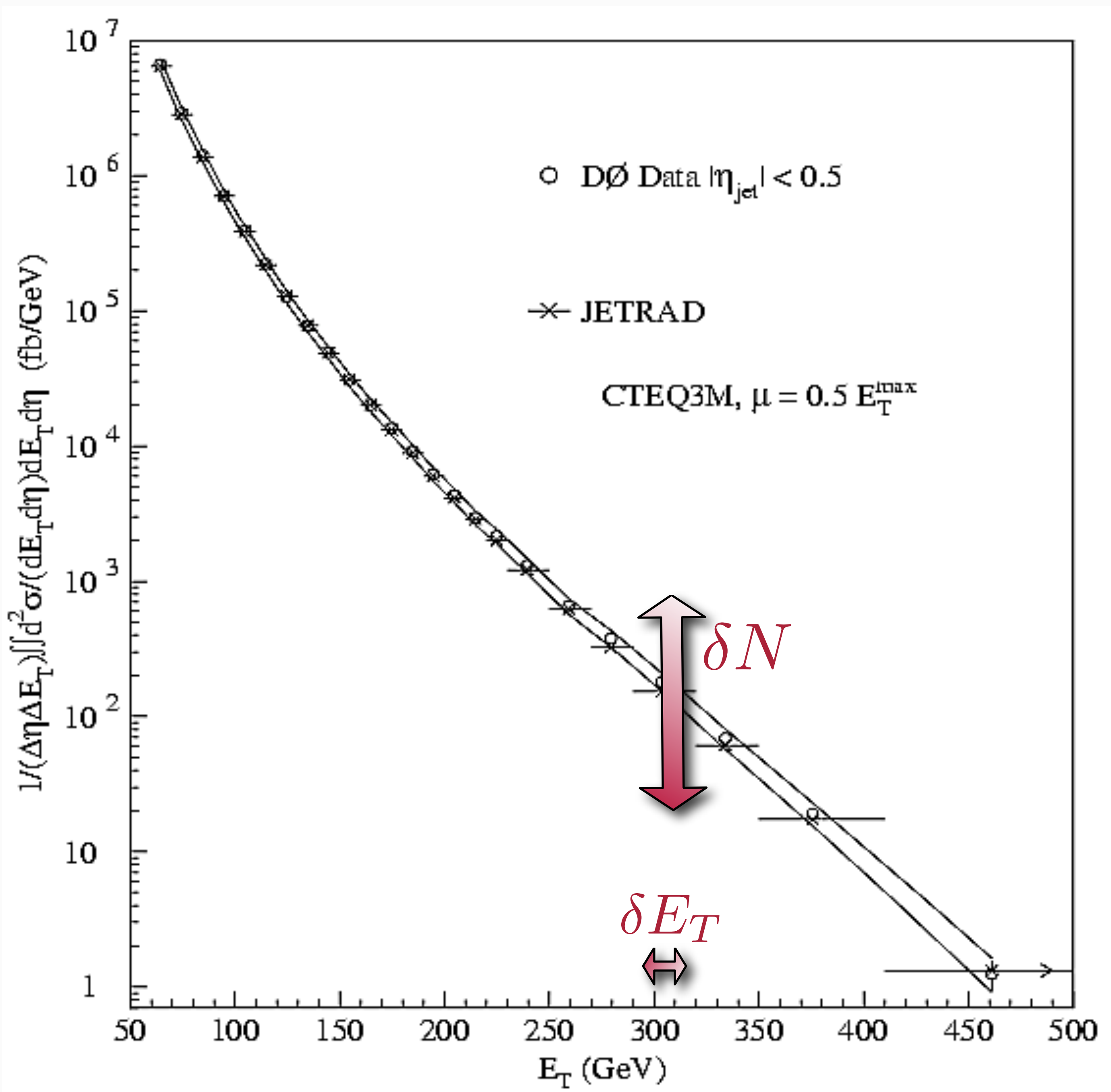
Problem 1 : Energy scale

- 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!



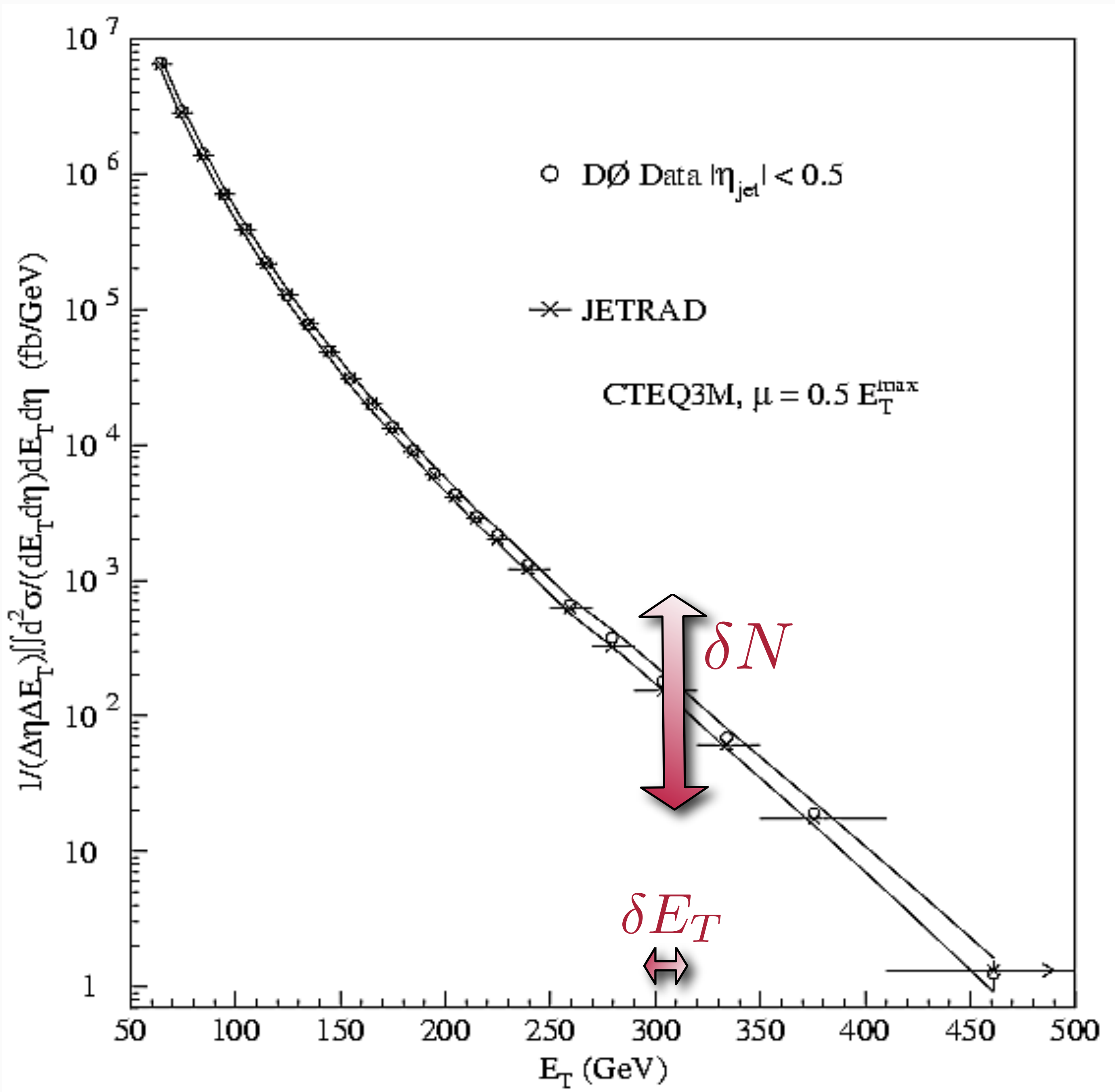
$$\frac{d^2\sigma}{dE_T d\eta} \approx \text{const} \cdot E_T^{\ -6}$$

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!



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

relative uncertainties

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

so beware:

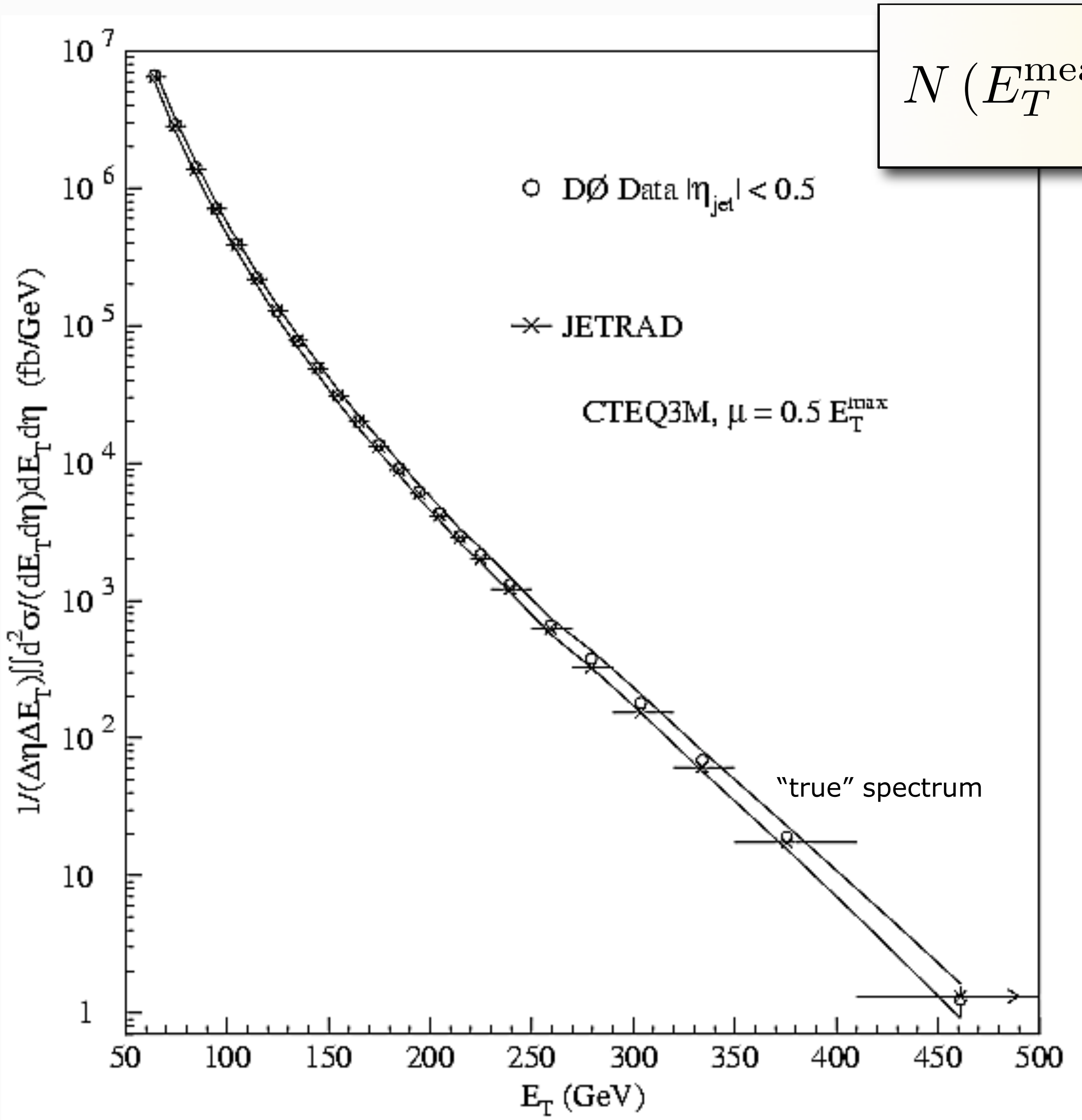
eg. an uncertainty of **5%** on absolute energy scale (calibration)

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

Problem 2 : Energy resolution

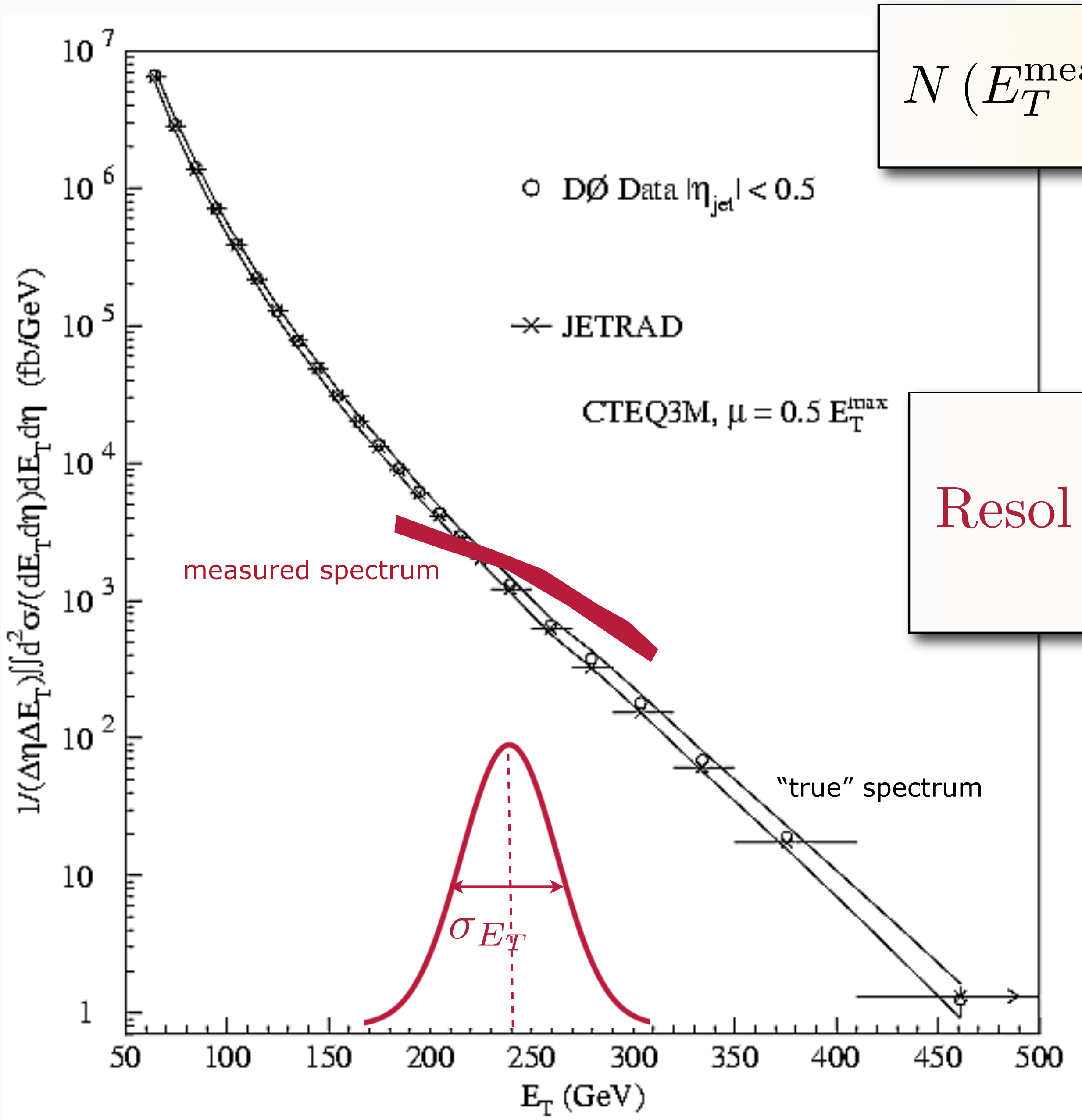
- The energy resolution can distorts the spectrum
- Again : Critical because of very steeply falling spectrum!

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



Problem 2 : Energy resolution

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



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

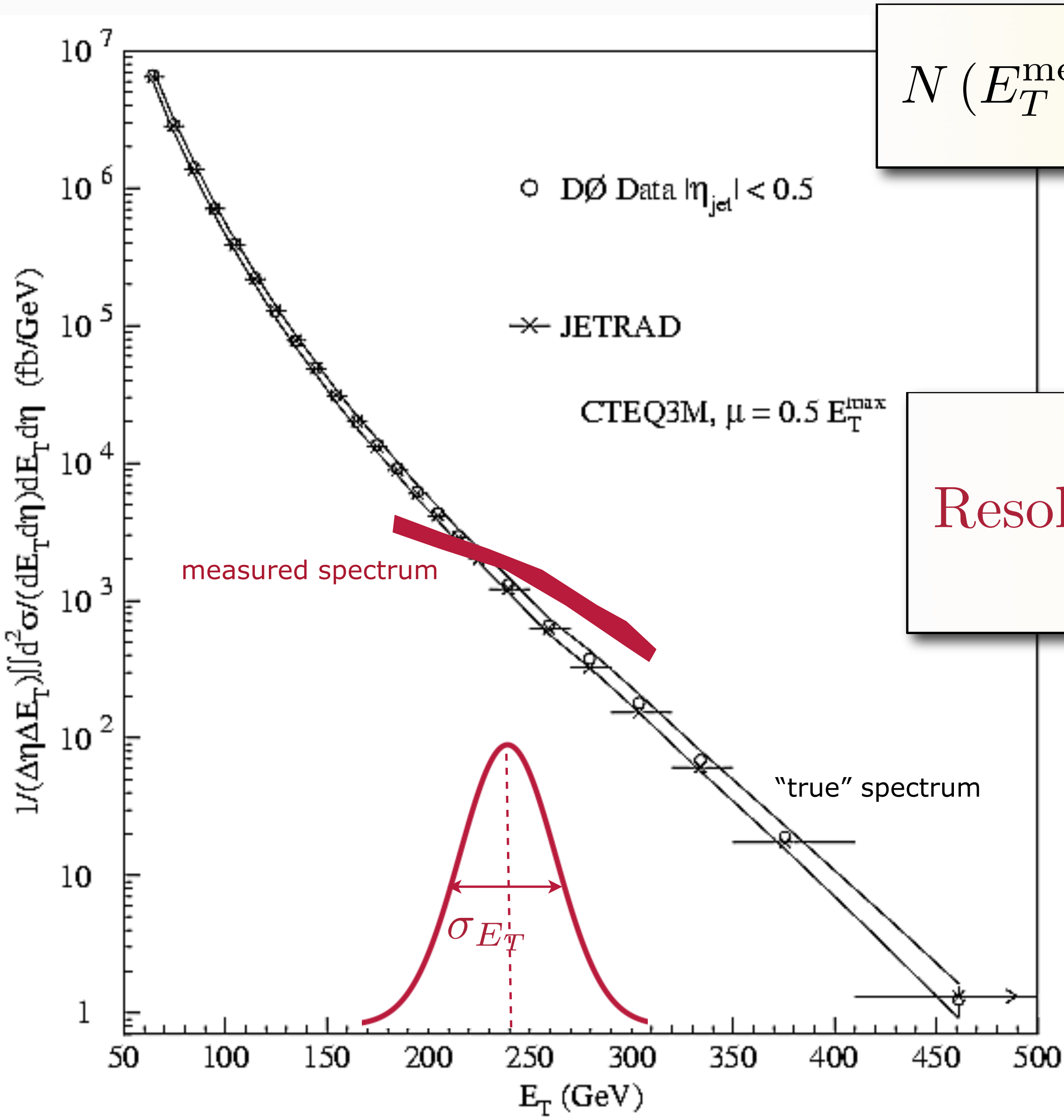
eg. Gaussian resolution function

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

ETH Institute for Particle Physics

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

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

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?

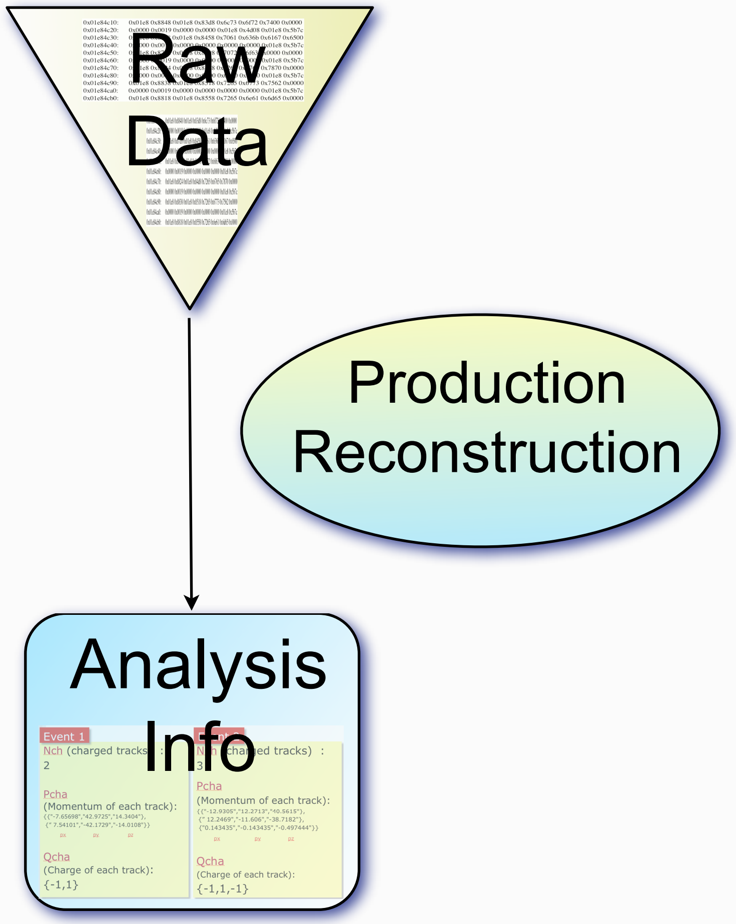
- After data flow from DAQ: data reduction and abstraction
  - reconstruct tracks, energy deposits (clusters) in calorimeters
  - calculate “high-level” physics quantities
    - eg. momentum of charged particles, energy of neutral particles
  - apply even higher-level algorithms, eg. jet finding
  -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

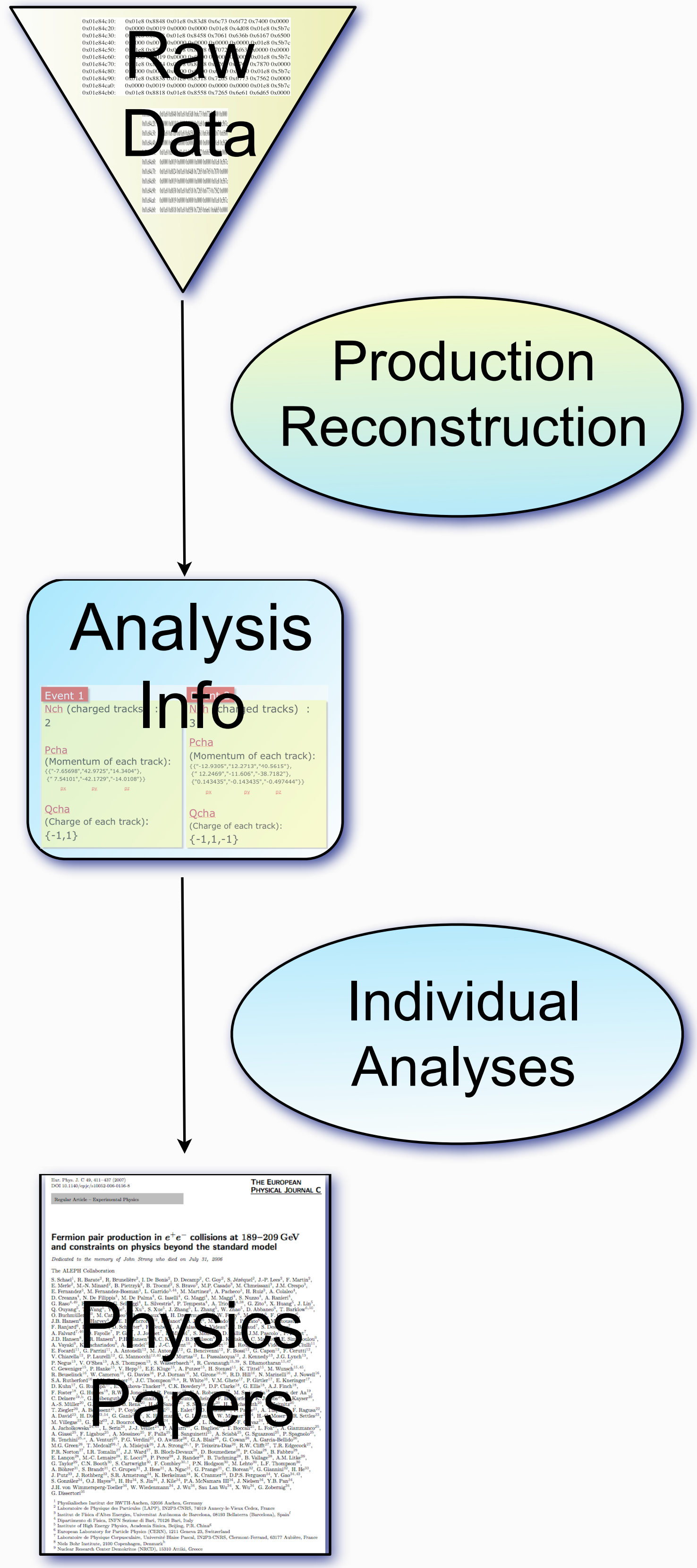
The process in practice

- 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



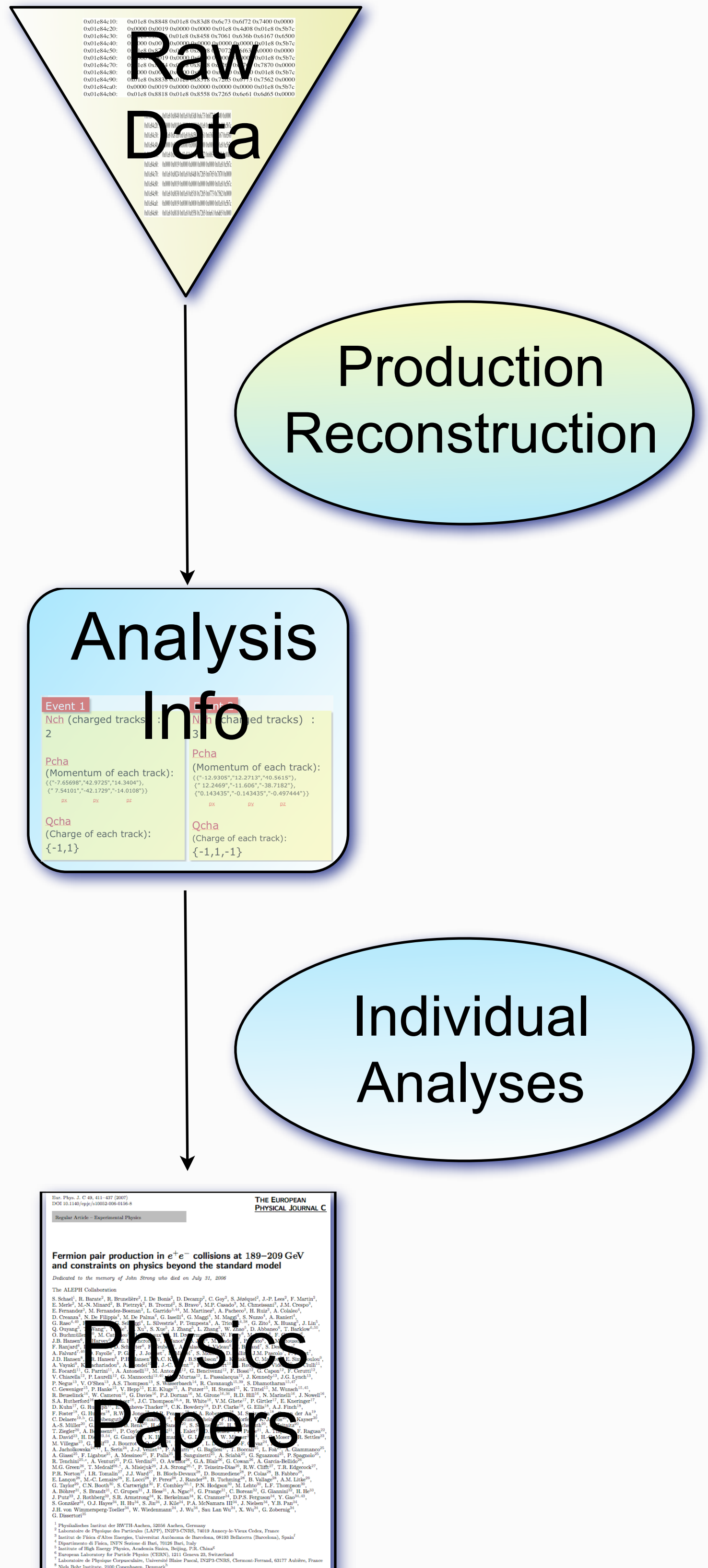
The process in practice

- 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

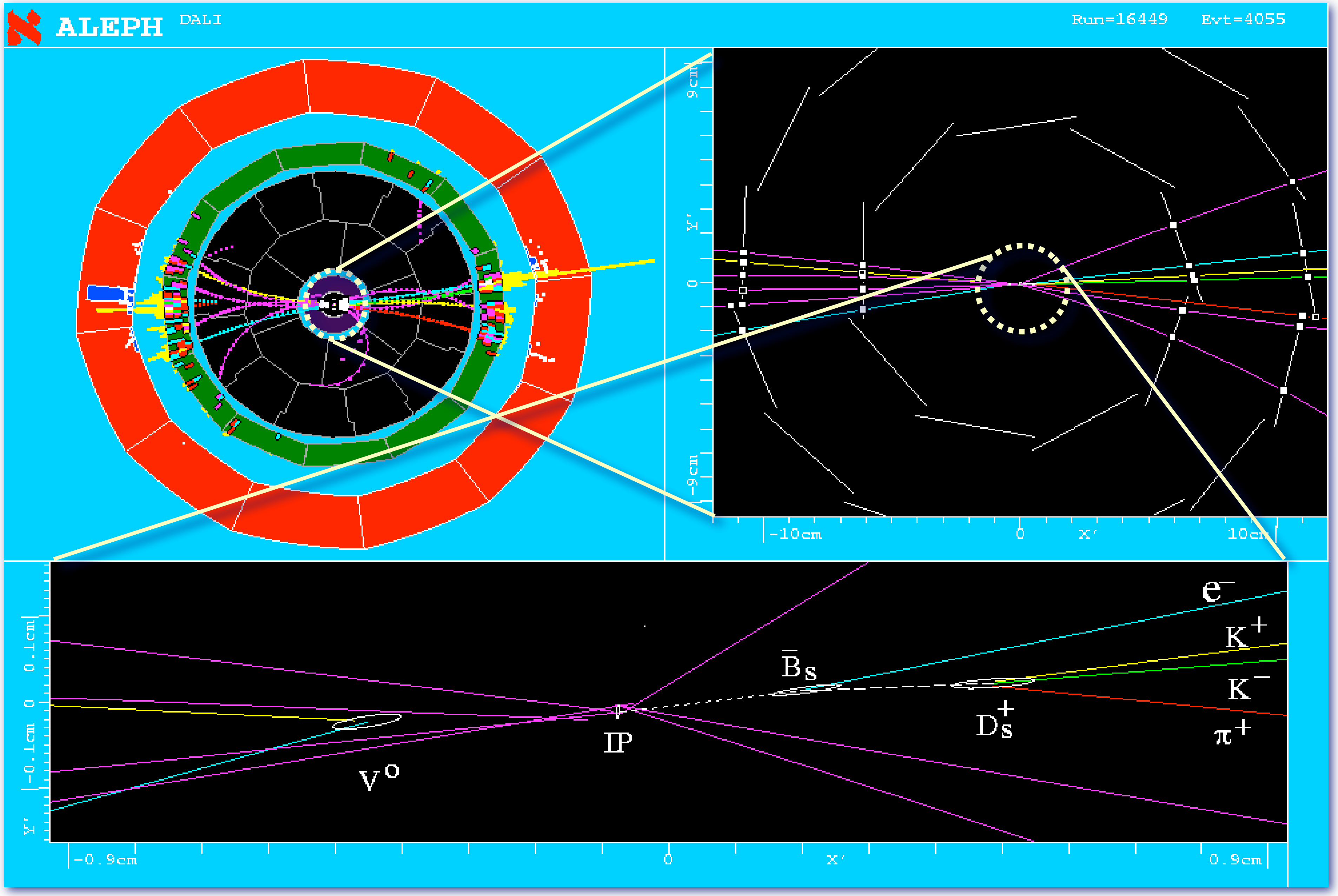
Why does tracking need to be done well?

- 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?

ETH Institute for Particle Physics

- 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

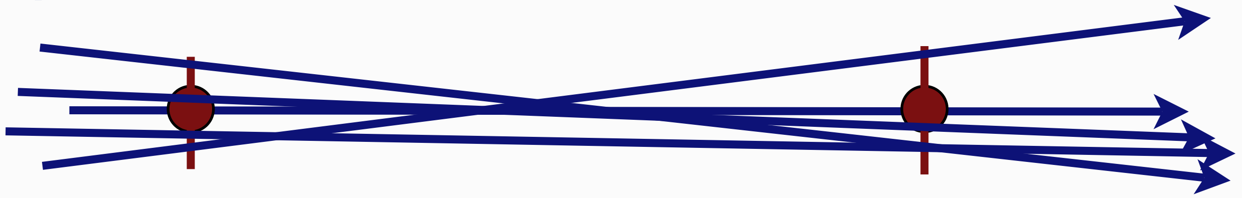
- 1D straight line fit as simple case
- Two perfect measurements
  - away from interaction point
  - no measurement uncertainty
  - just draw a straight line through them and extrapolate

- 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

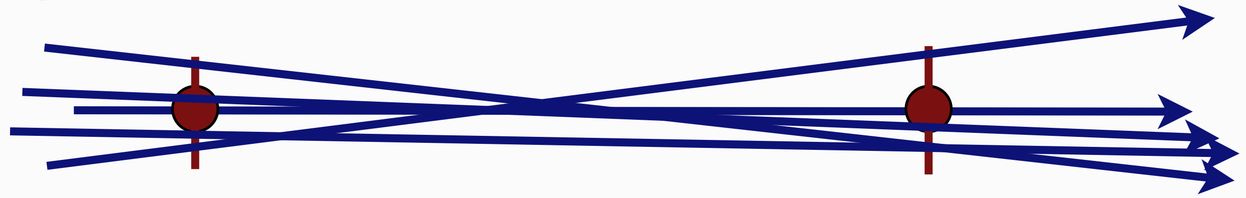
- Imperfect measurements give less precise results
  - the farther you extrapolate, the less you know



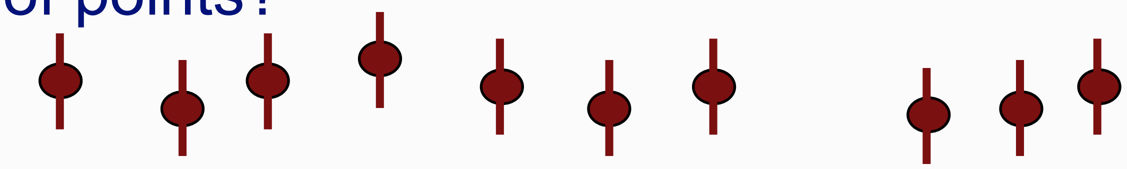
Track Fitting

- 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



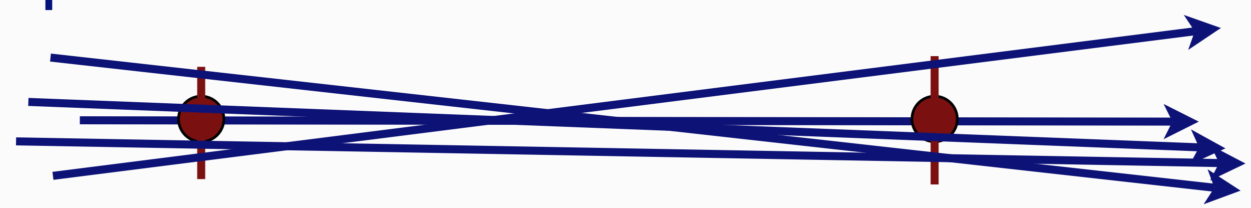
- Smaller errors and more points help to constrain the possibilities.
  But how to find the best point from a large set of points?



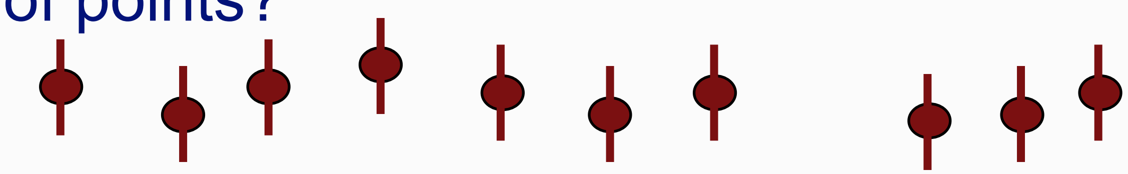
Track Fitting

- 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?

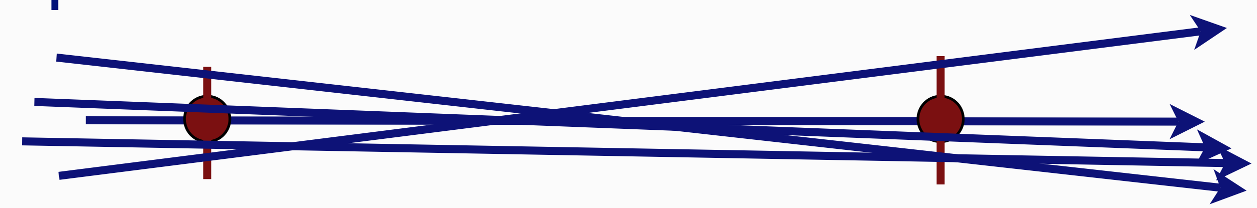


- Quantitatively
  - parameterize a track:
  - In case of straight line $\boxed{y(x) = \theta\, x + d}$ or, eg., helix in case of magnetic field present

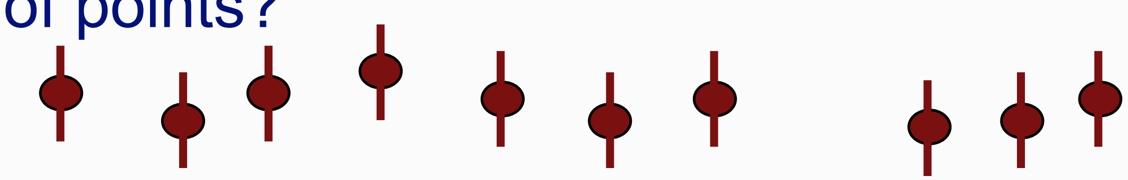
Track Fitting

- 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
  - Find track parameters by Least-Squares-Minimization

predicted track position at i^{th} hit

position of i^{th} hit

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

uncertainty of i^{th} measurement

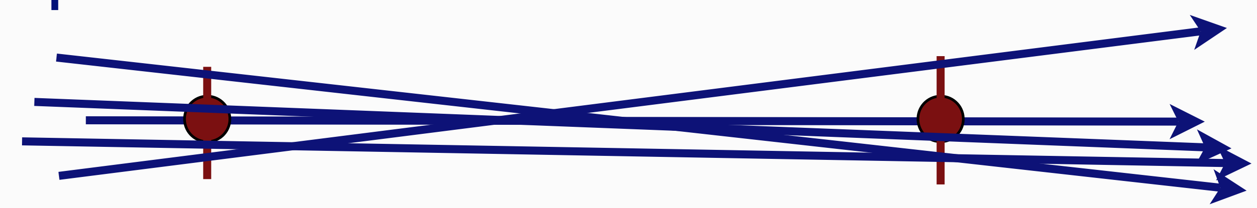
see also lecture by G. Cowan

Track Fitting

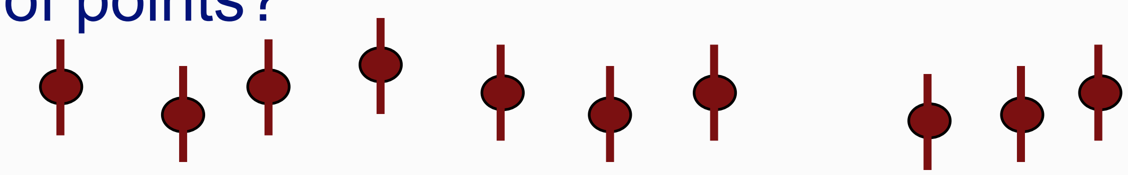
ETH Institute for Particle Physics

- 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

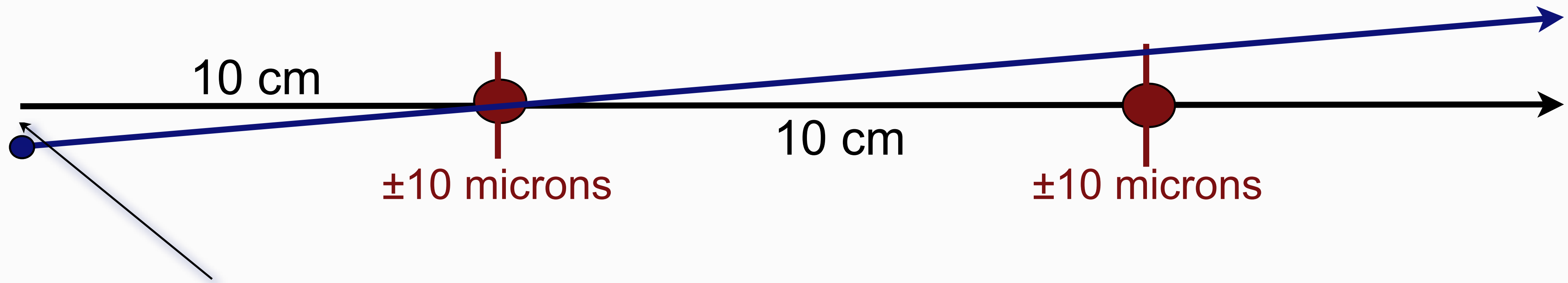
predicted track position at i^{th} hit

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

uncertainty of i^{th} measurement

see also lecture by G. Cowan

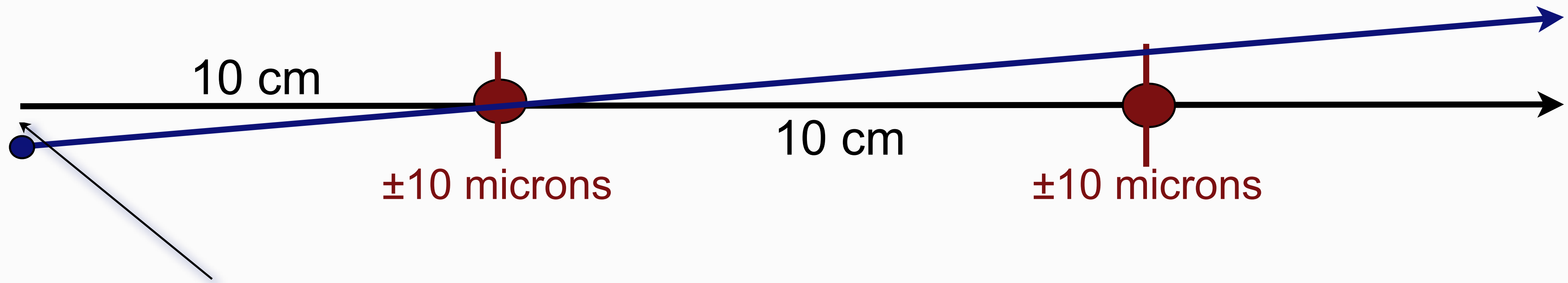
"Typical" size of errors



- 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)

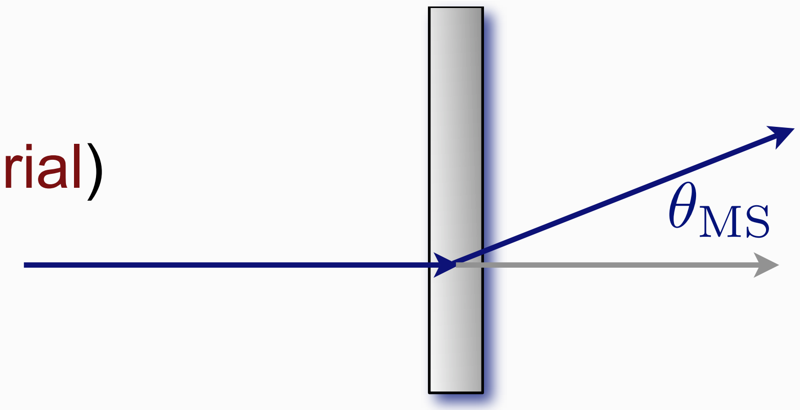
- However
  - we "see" particles by interaction with a detector (=material)

"Typical" size of errors



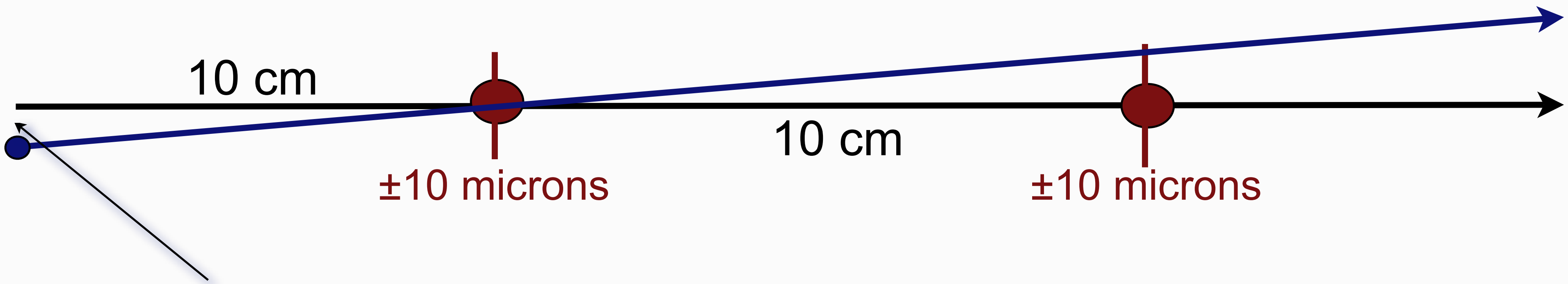
- Error δd on position is about ±10 microns
- Error $\delta\Theta$ on angle is about ±0.1 milliradians (±0.002 degrees)
- Satisfyingly small errors
  - allows separation of tracks that come from different particle decays (which can be separated at the order of mm)

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



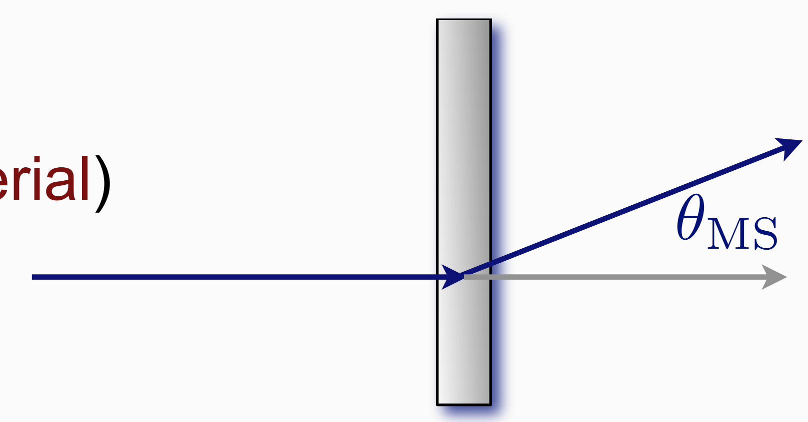
“Typical” size of errors

ETH Institute for Particle Physics



- 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



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

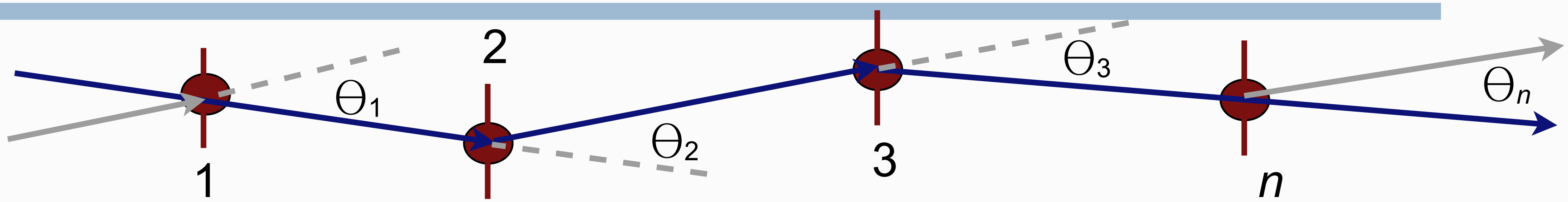
examples:

300 micron Si : RMS = 0.9 mrad / βp

1 mm Be : RMS = 0.8 mrad / βp

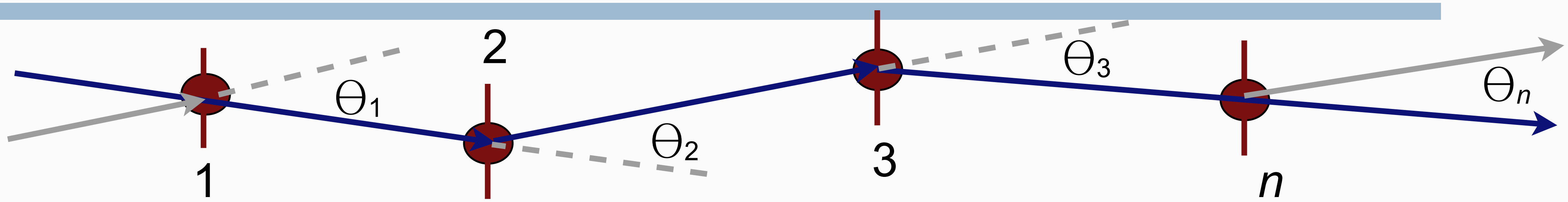
➔ leads to additional position errors

Kalman filter



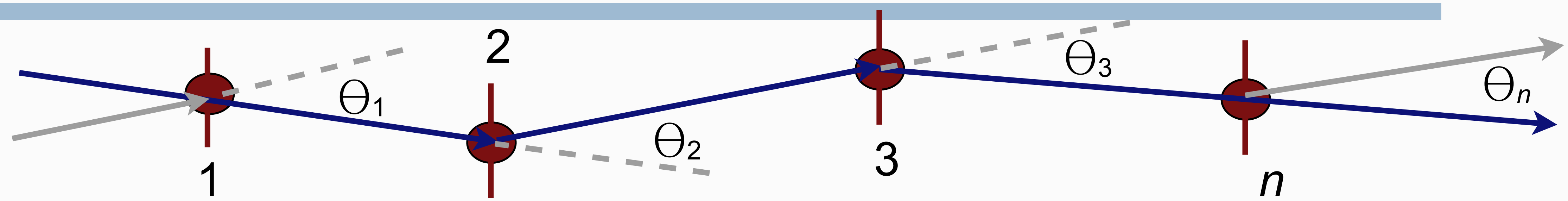
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)
  - 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) $\chi^2 = \chi^2_{\rm old} + \sum_i \frac{\theta_i^2}{\sigma^2_{\rm MS}}$
  - 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

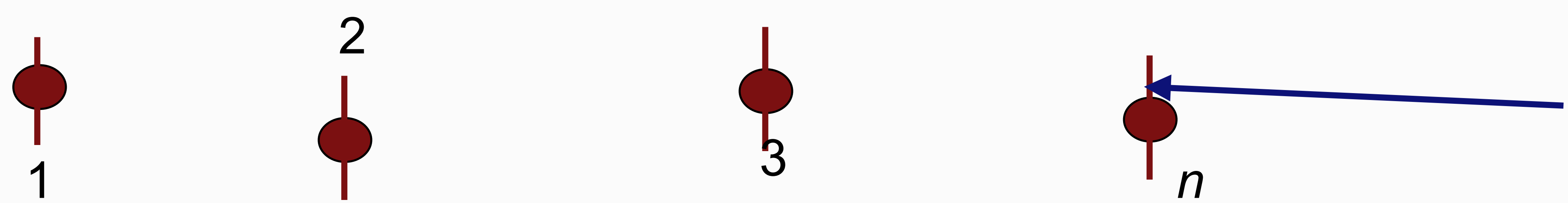
Kalman filter



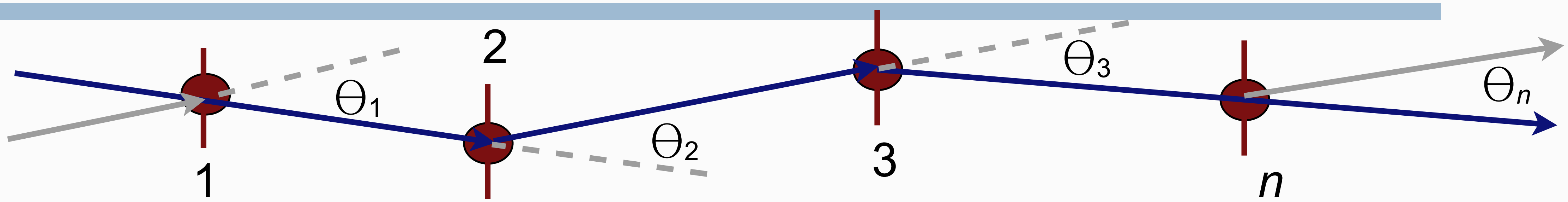
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

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

  - 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



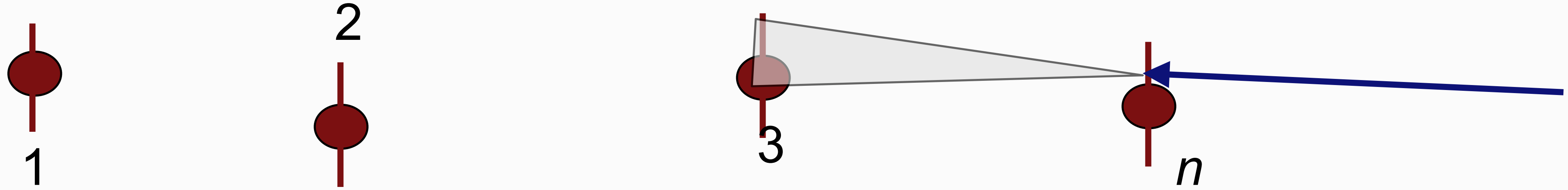
Kalman filter



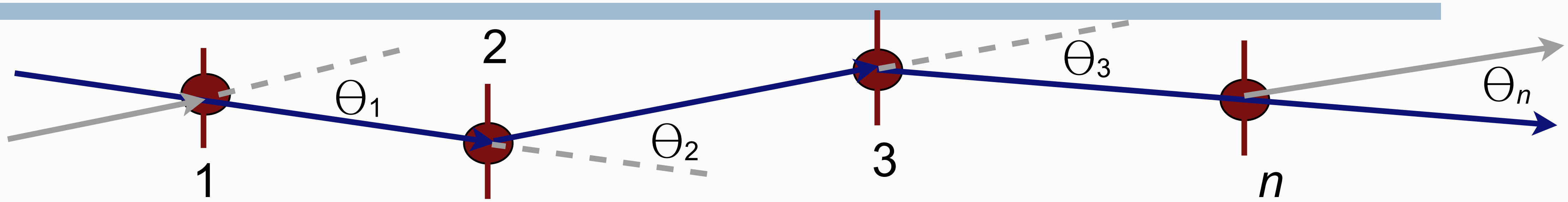
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

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

  - 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



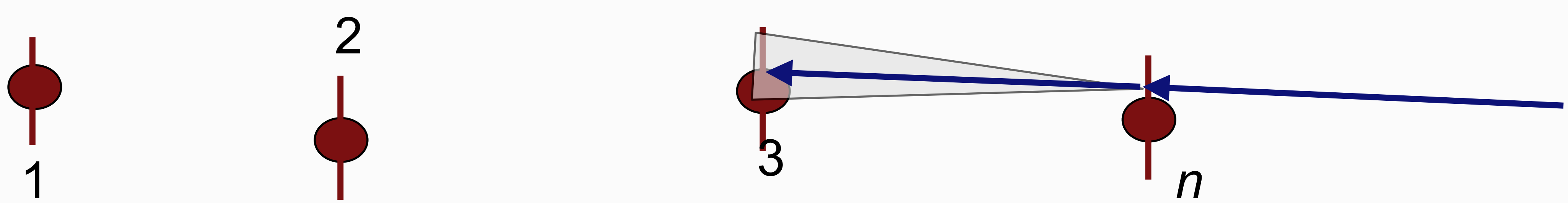
Kalman filter



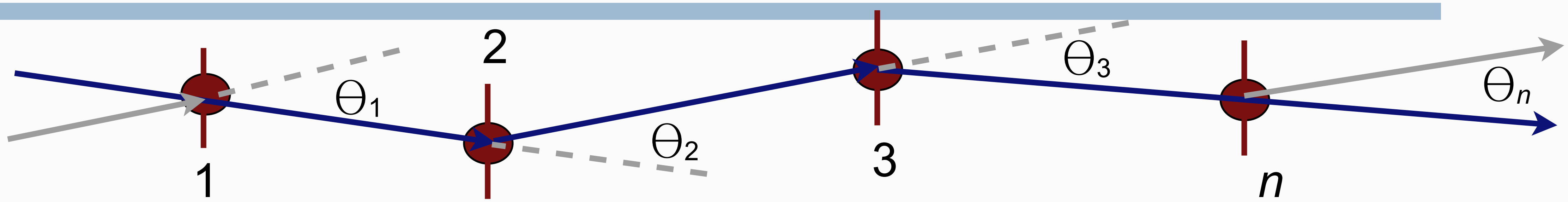
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

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

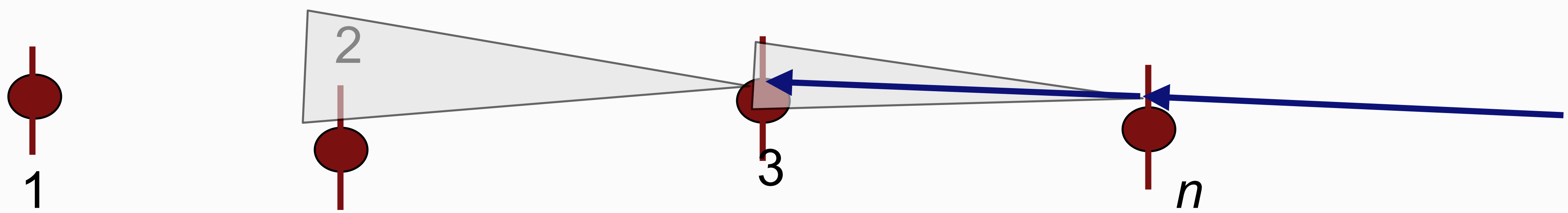
  - For large *n*, computing time grows like O(n^3), quickly un-practicable
  - anyway, not interested in all these angles, only in parameters at the origin
- Instead, approximate, work inward *N* times



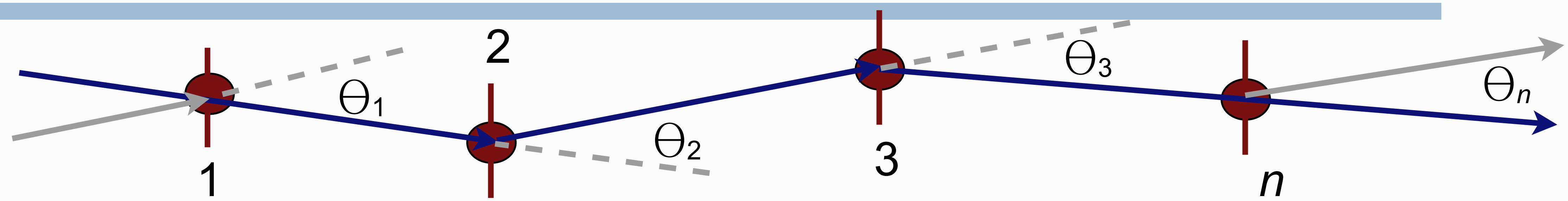
Kalman filter



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



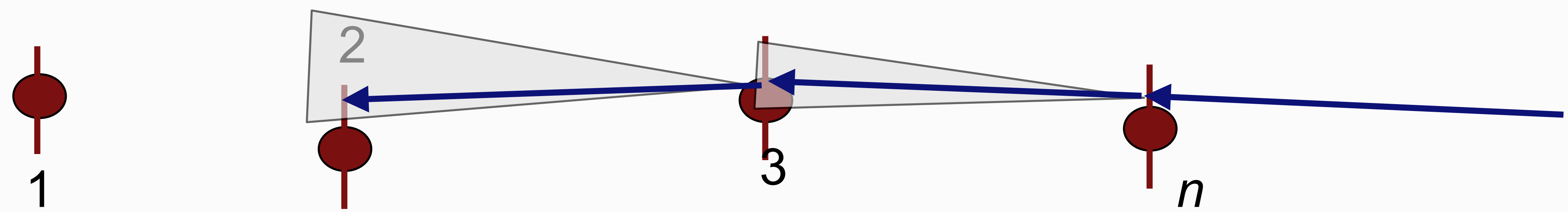
Kalman filter



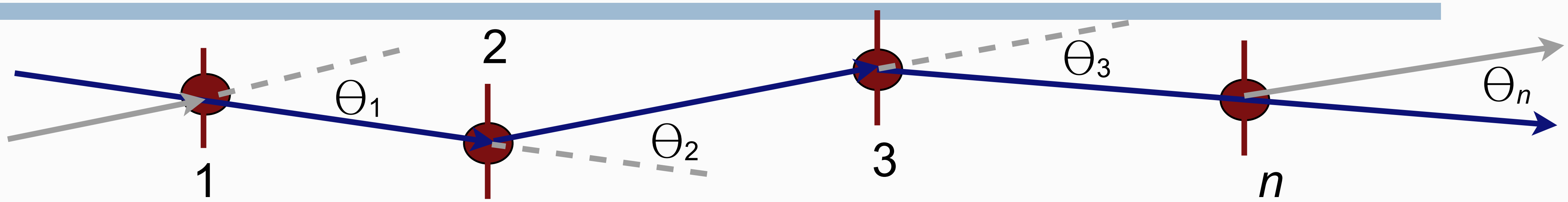
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

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

  - 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



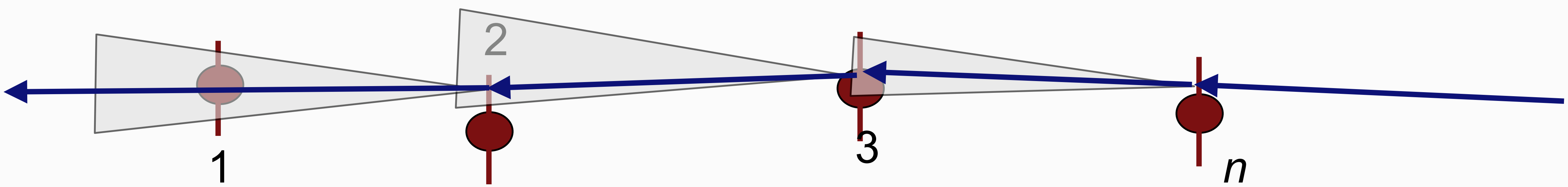
Kalman filter



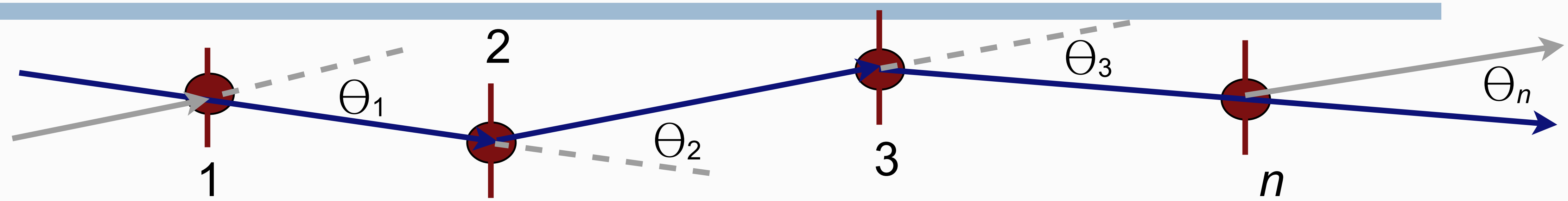
- So? Could extend track parameterization to take this into account
  - *n* additional parameters $y(x) = d + \theta\, x + \Theta(x - x_1)\theta_1(x - x_1) + \Theta(x - x_2)\theta_1(x - x_2) + \cdots$
- and include the multiple scattering information into the Least-Squares (*n* equations, *n* unknowns)

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

  - 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



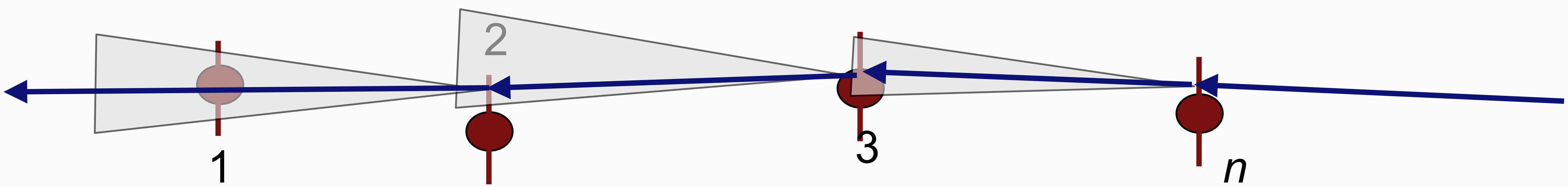
Kalman filter



- 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

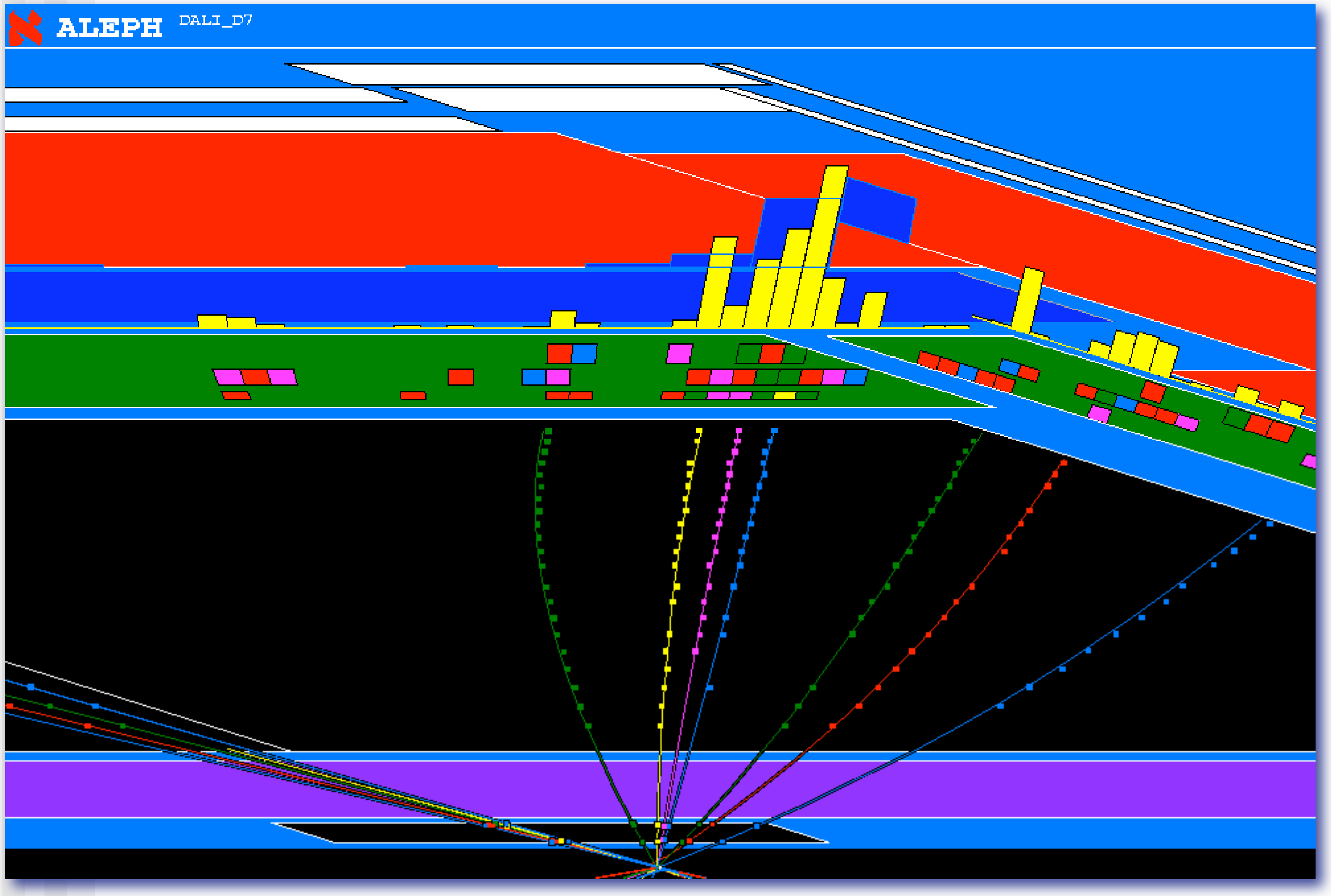


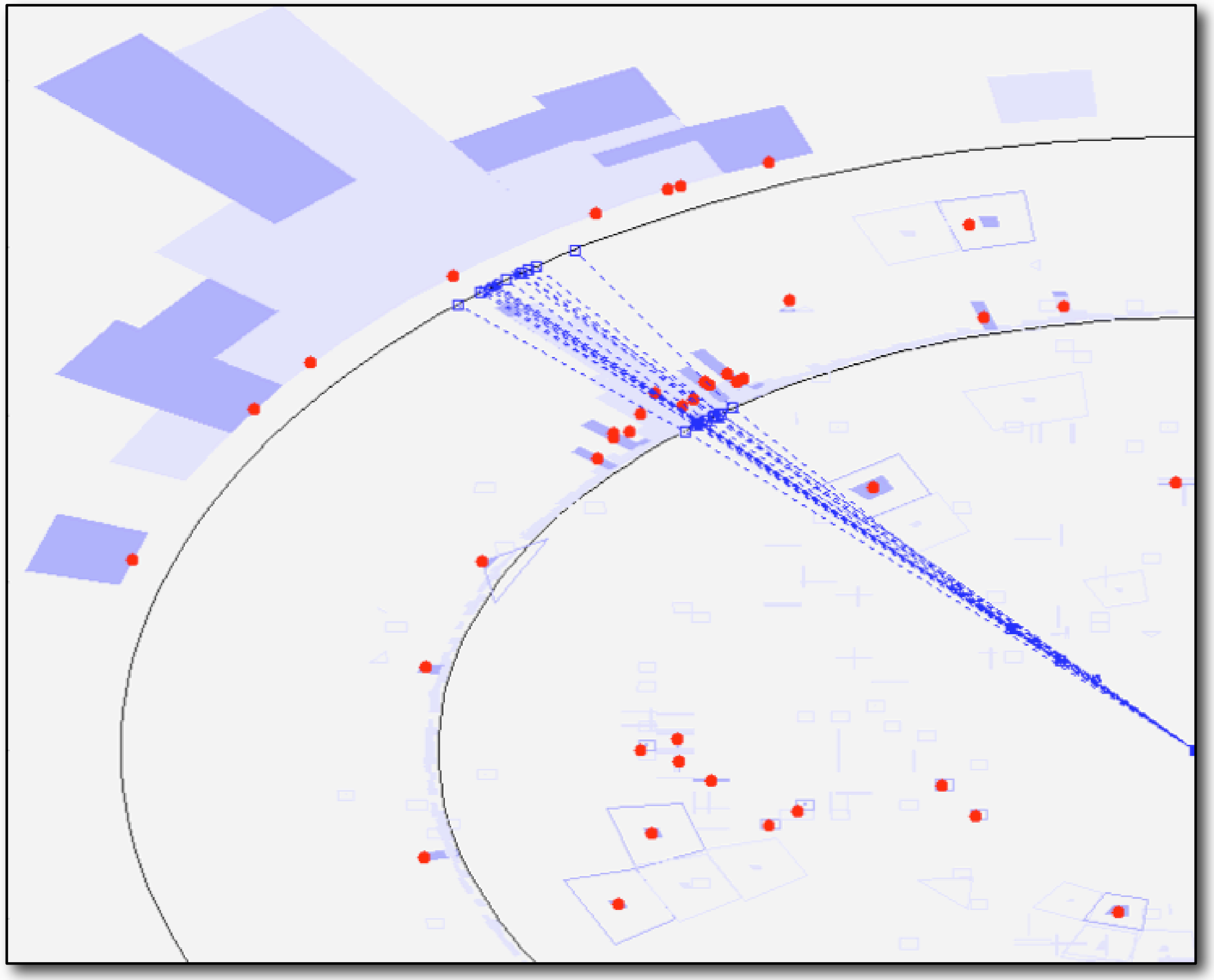
  - leads to O(*n*) computations!
  - in each step, make extrapolation to next layer, using information from current track parameters, expected scattering error, and measurement at next layer
  - Needs a starting estimate (seed) and may need some iterations, smoothing
  - This method is based on theory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





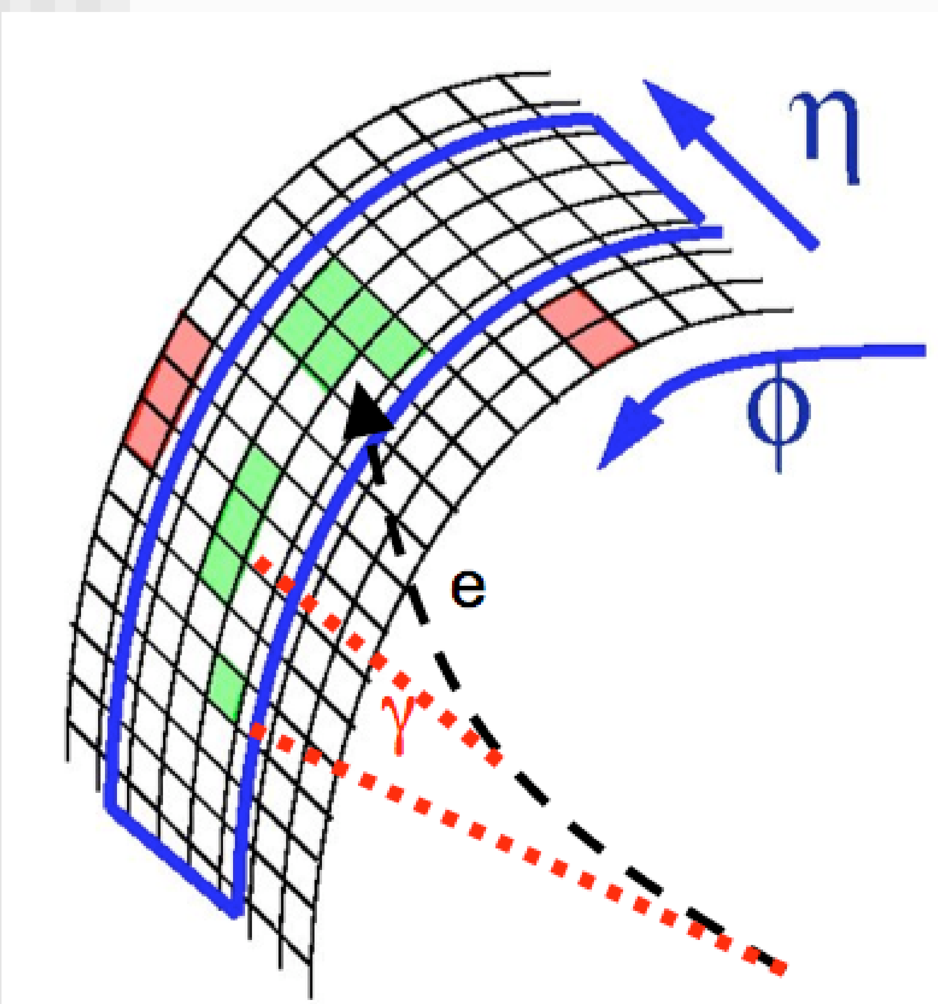
- and obtain the best possible resolution!

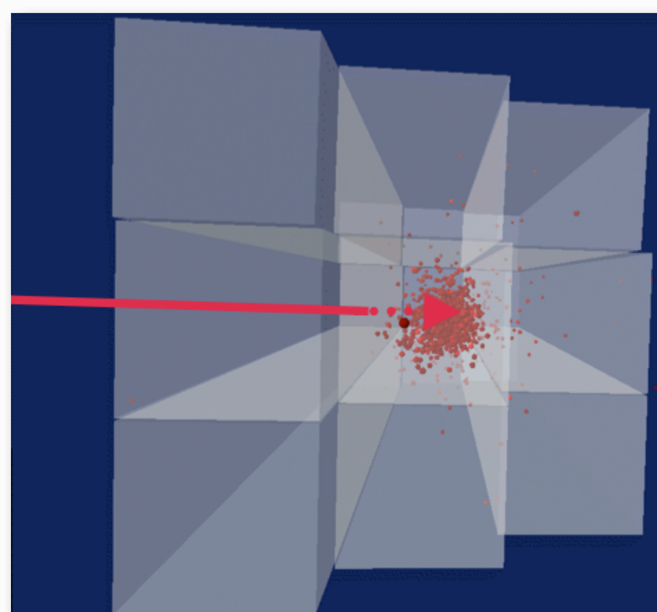
Clusters of energy

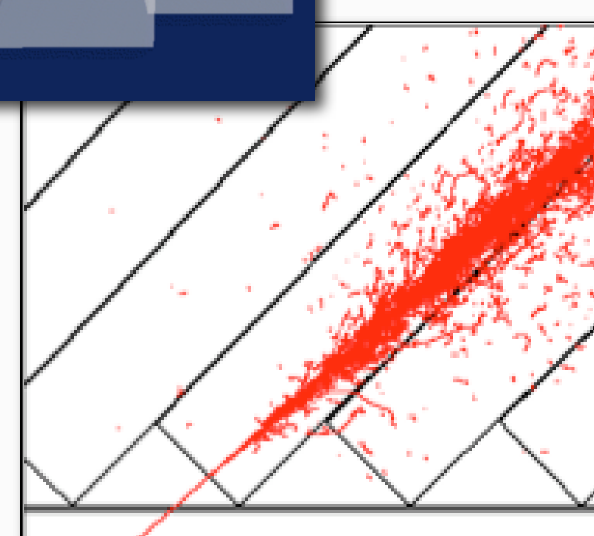
- Calorimeters are segmented in cells
- Typically a shower extends over several cells
  - Useful to reconstruct precisely the impact point from the “center-of-gravity” of the deposits in the various cells

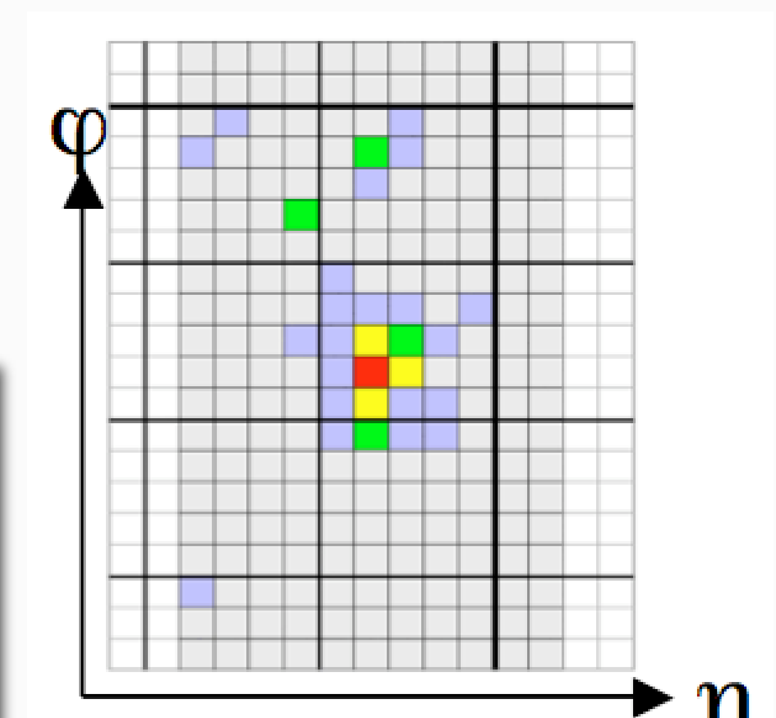
Clusters of energy

- 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

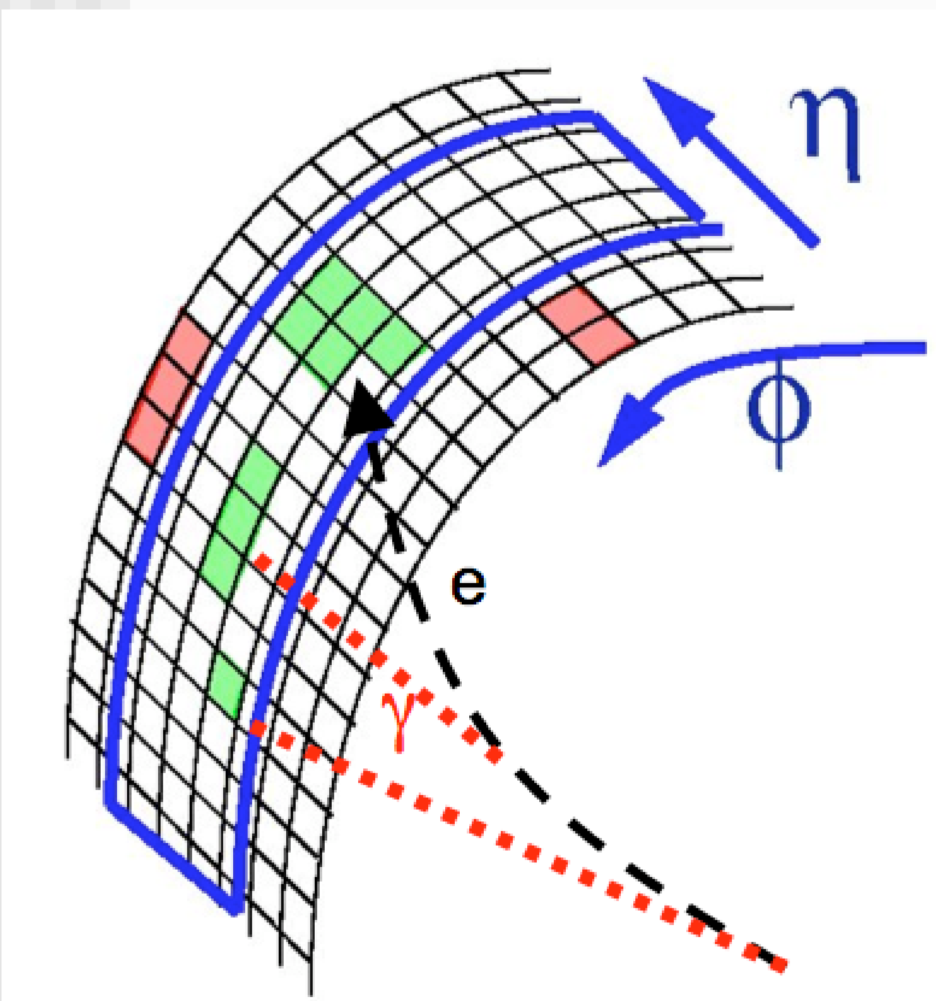

side view

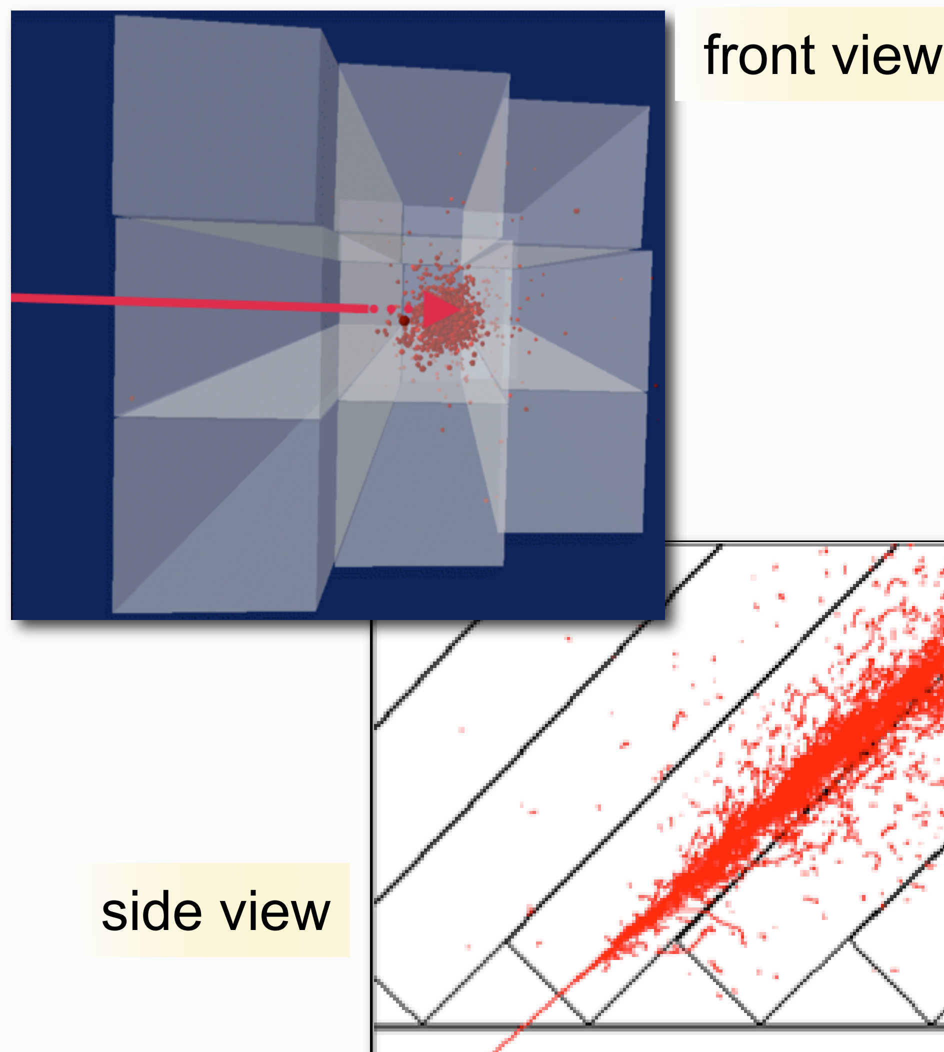


view in (φ,η) cells

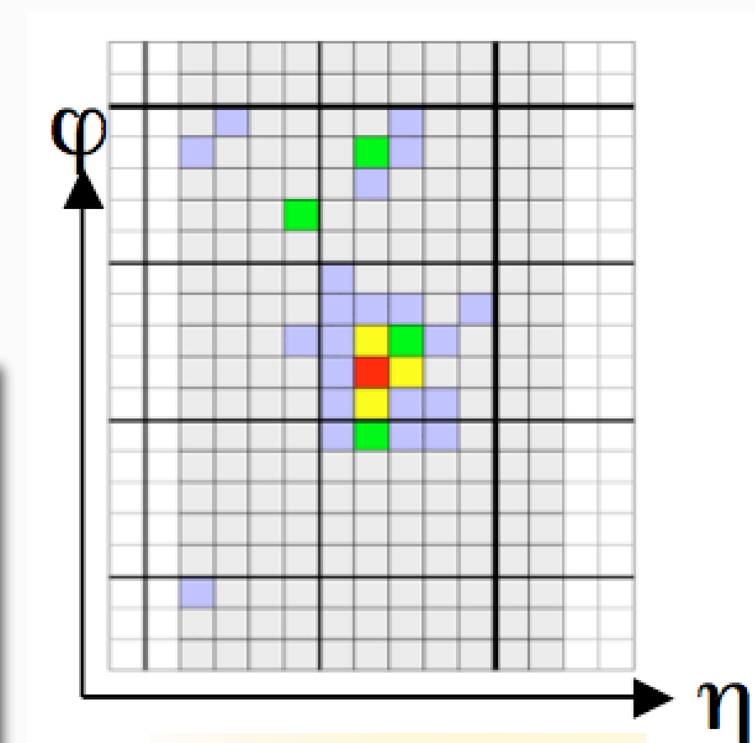
Clusters of energy

- 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

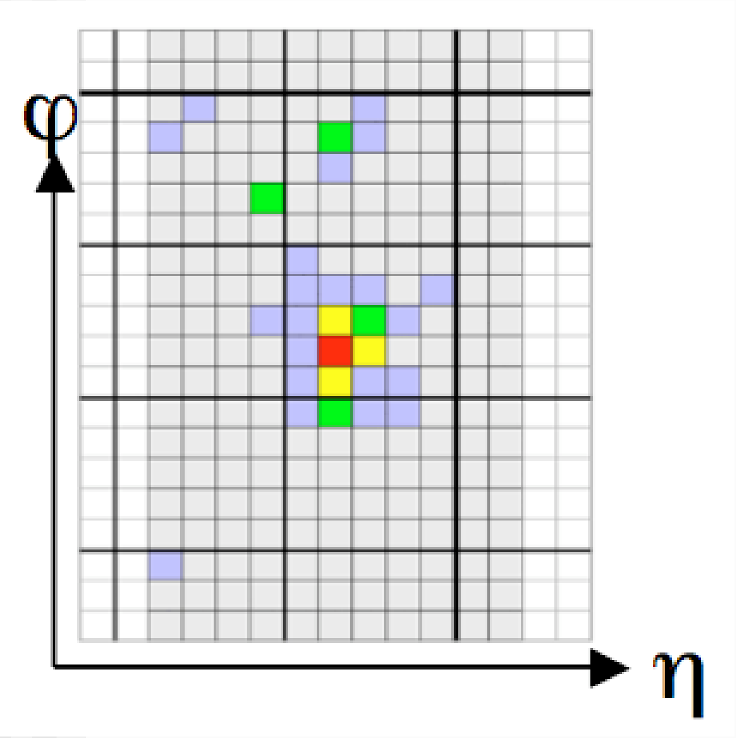
side view



view in (φ,η) cells

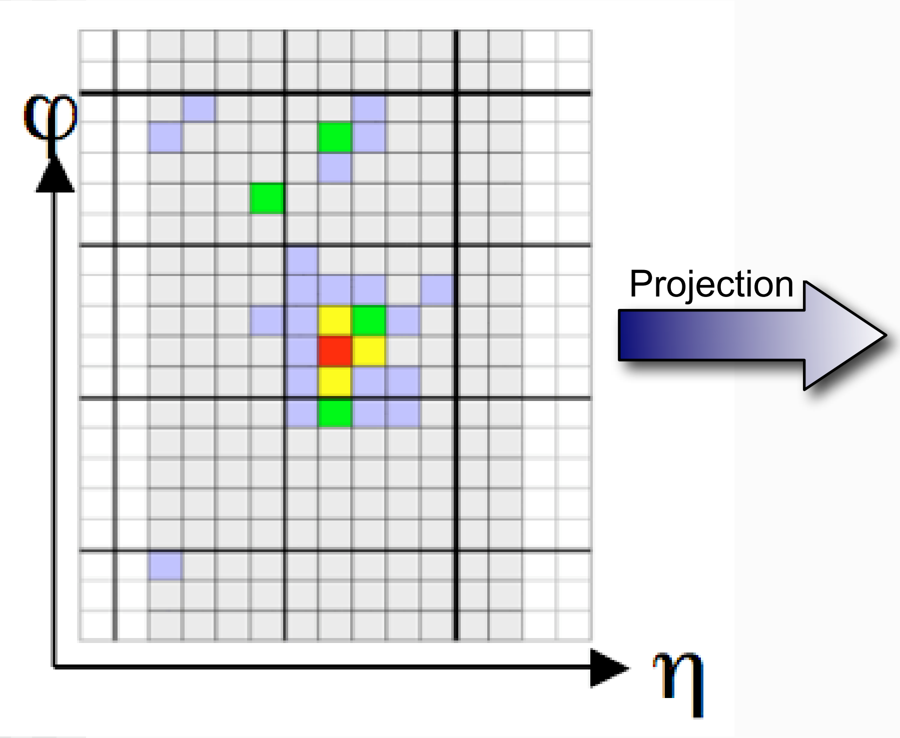
Cluster Finding

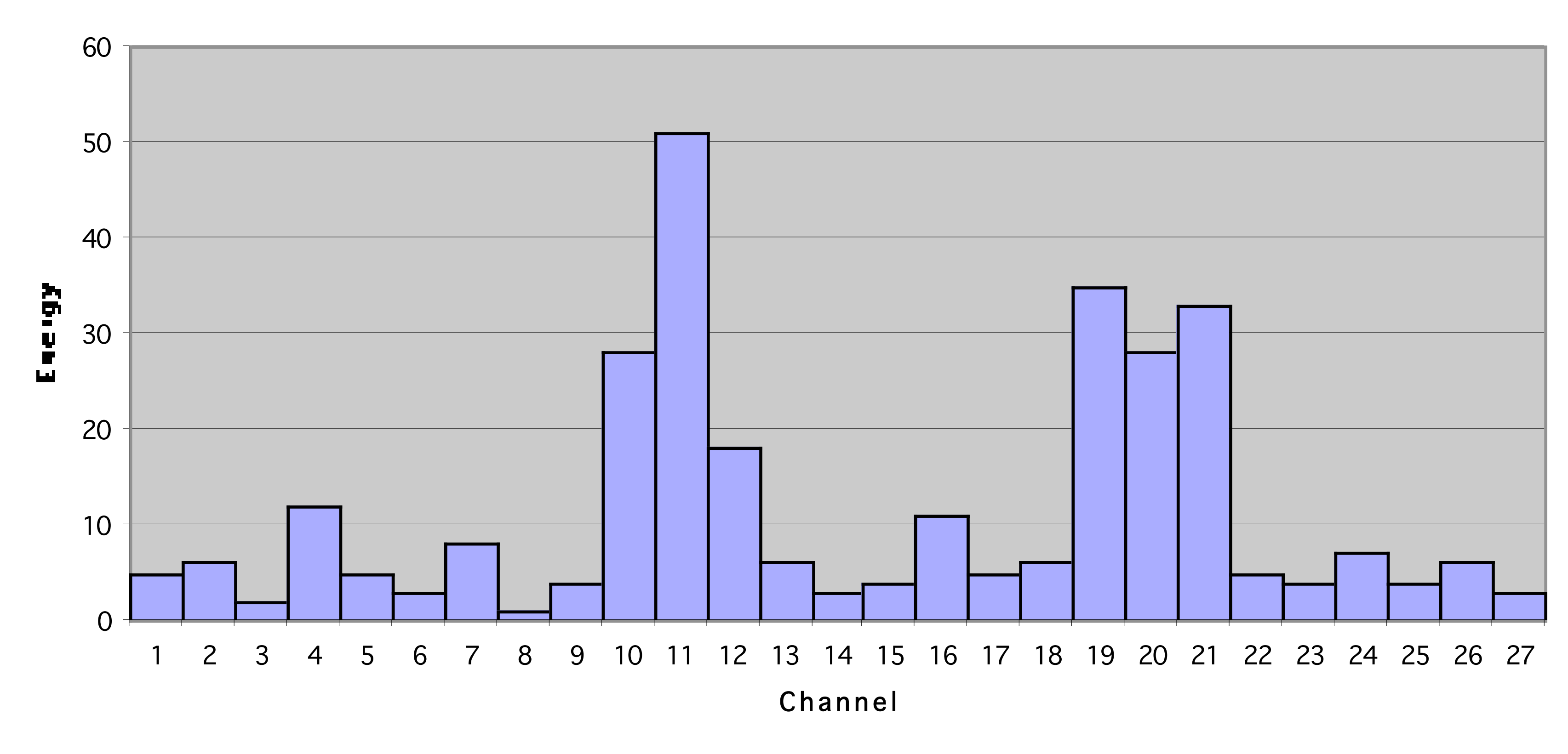
- Clusters of energy in a calorimeter are due to the original particles
  - Clustering algorithm groups individual channel energies
  - Don’t want to miss any; don’t want to pick up fakes



Cluster Finding

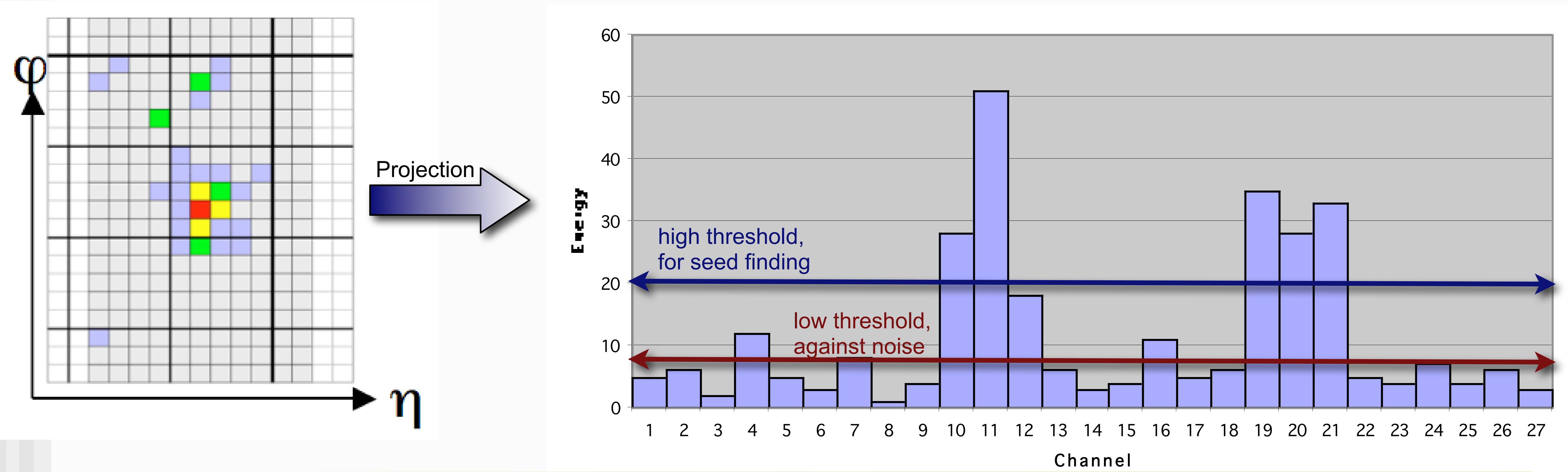
- Clusters of energy in a calorimeter are due to the original particles
  - Clustering algorithm groups individual channel energies
  - Don't want to miss any; don't want to pick up fakes





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



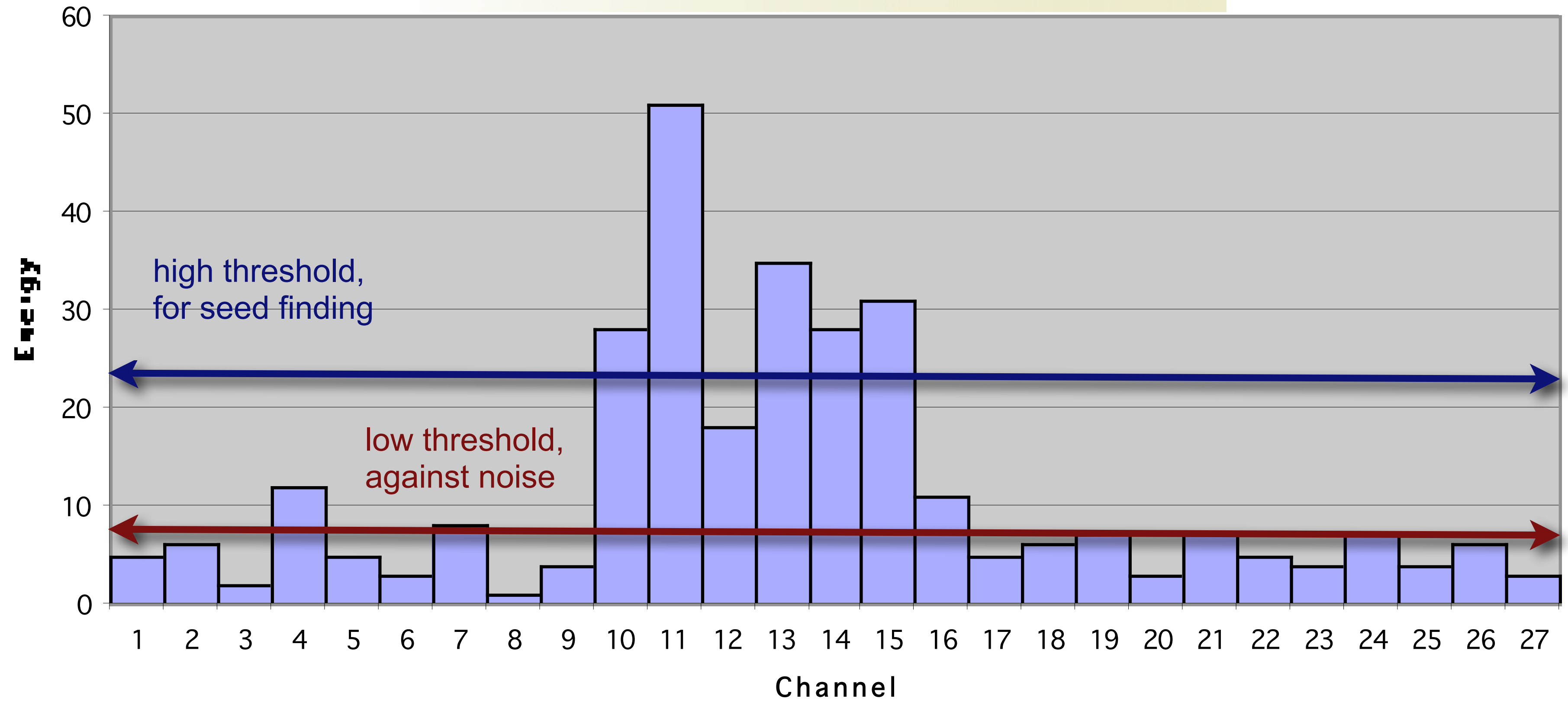
Simple example of an algorithm

- 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

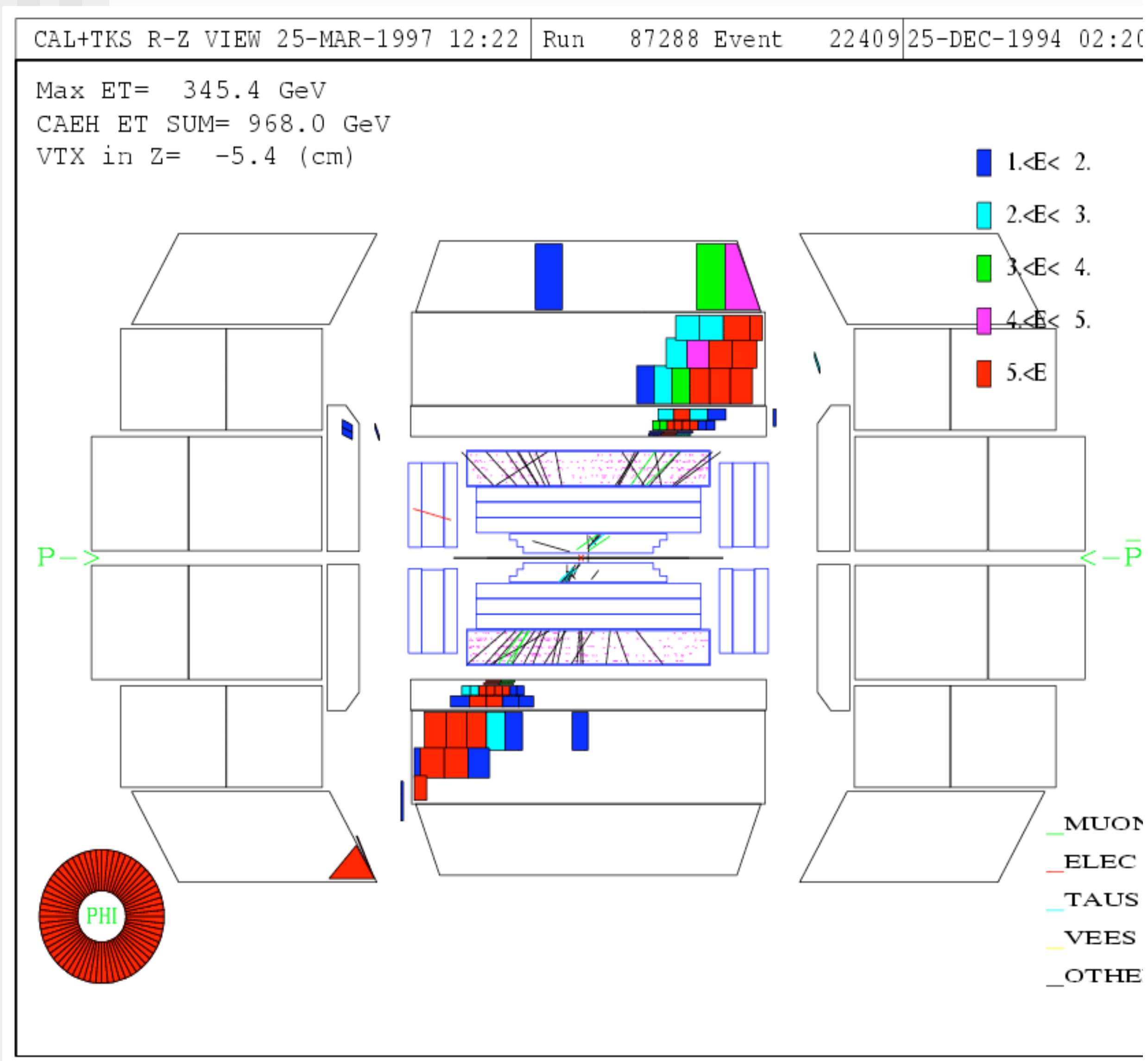
- Careful tuning of thresholds needed
  - needs usually learning phase
  - adapt to noise conditions
  - too low : pick up too much unwanted energy
  - too high : loose too much of “real” energy. Corrections/Calibrations will be larger
- Sometimes several clustering stages, in order separate or combine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

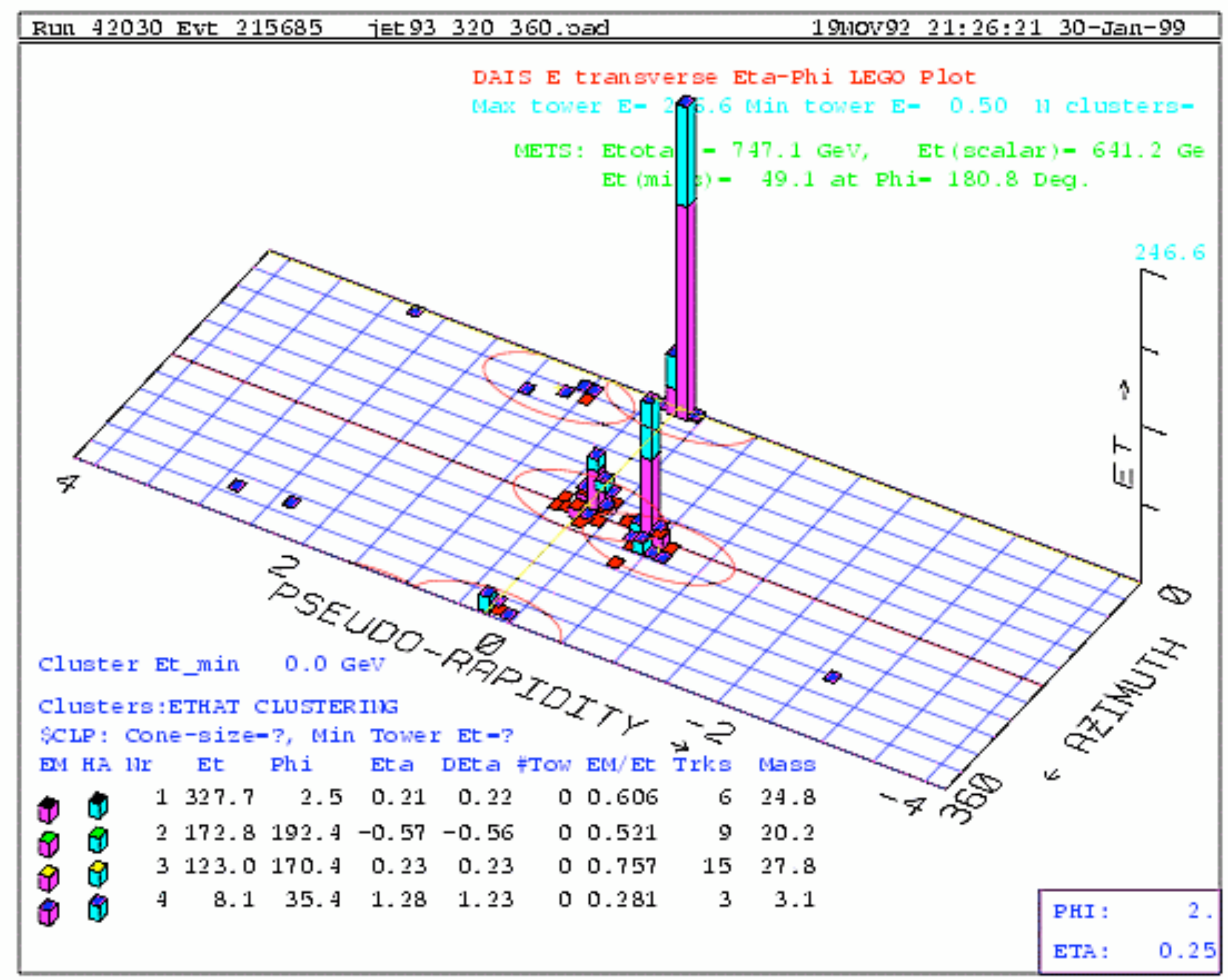
example : one lump or two?



Jet Algorithms

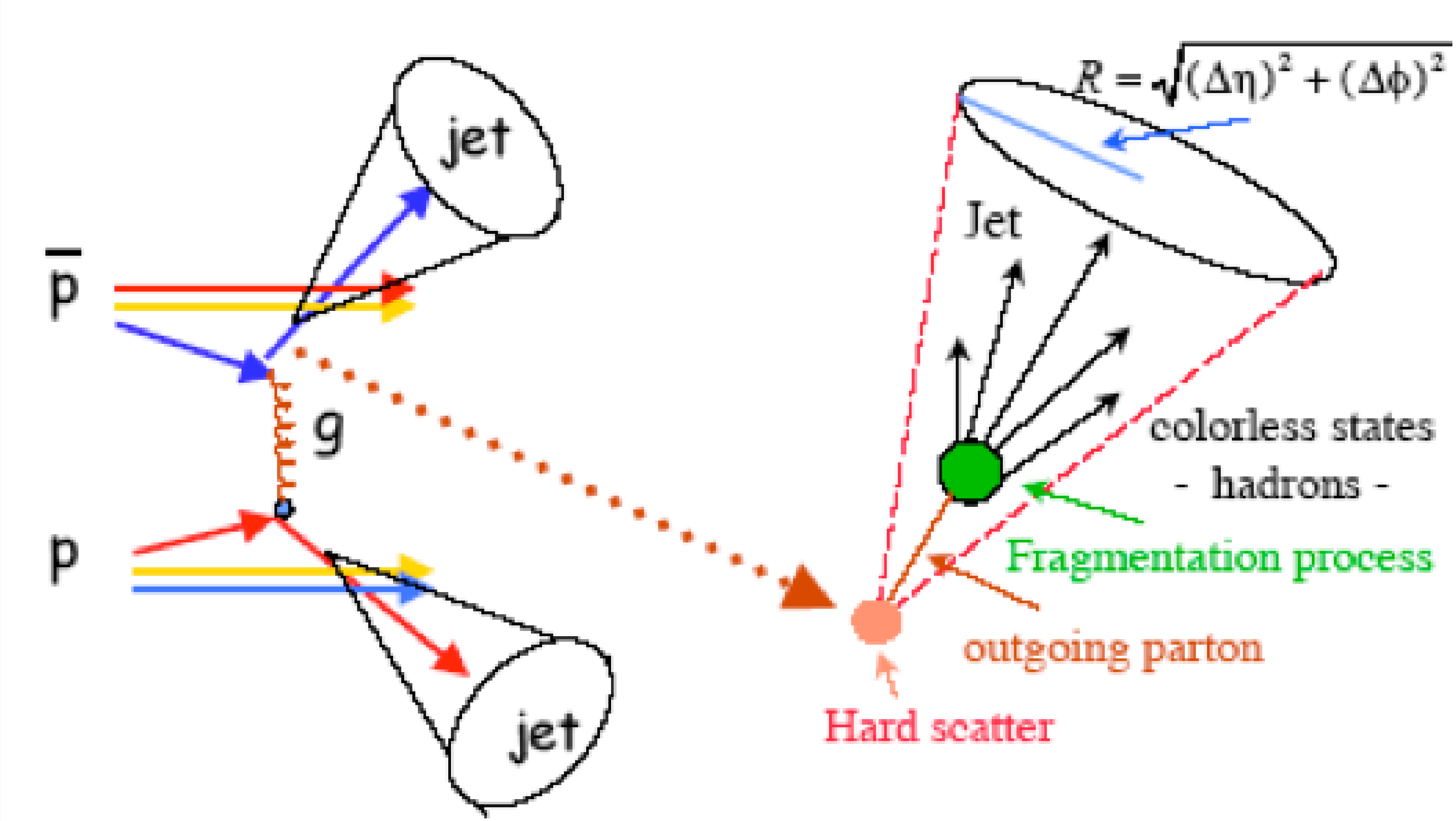
Jets in DØ CDF





- 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...

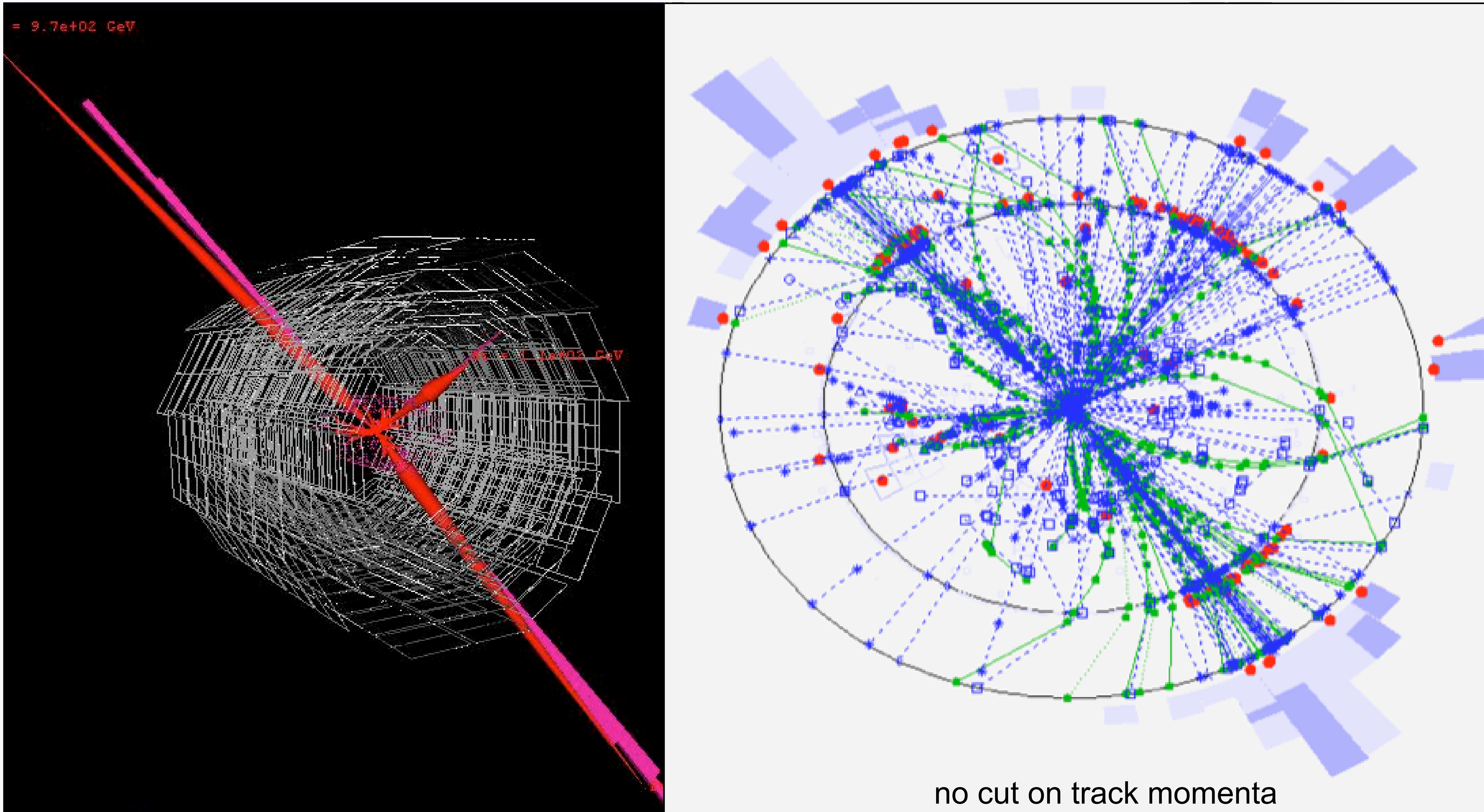


The natural (?) definition of a jet in a hadron collider environment

from J. Huston, CTEQ summer school 2004

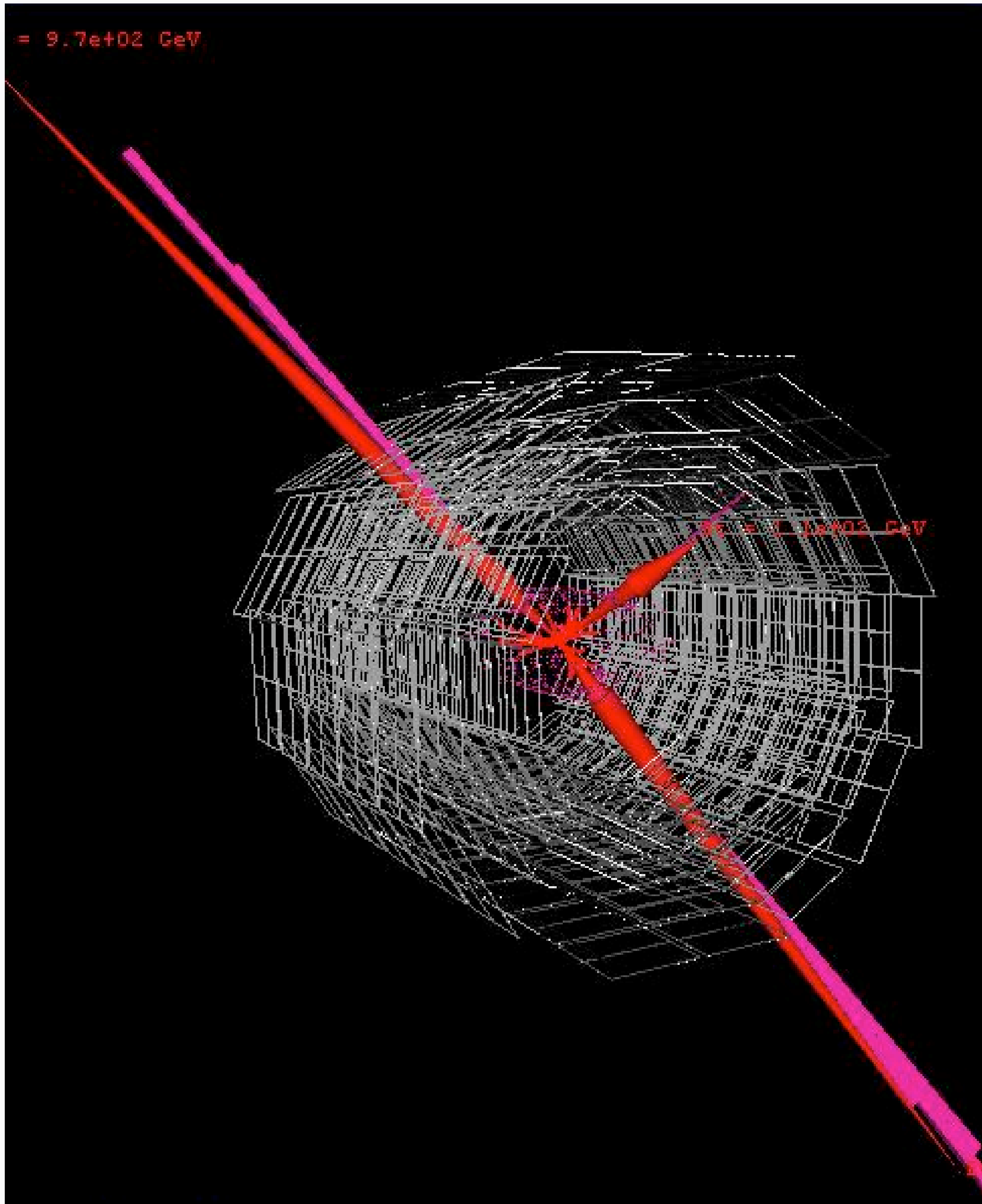
Jets in Hadron Collider Detectors

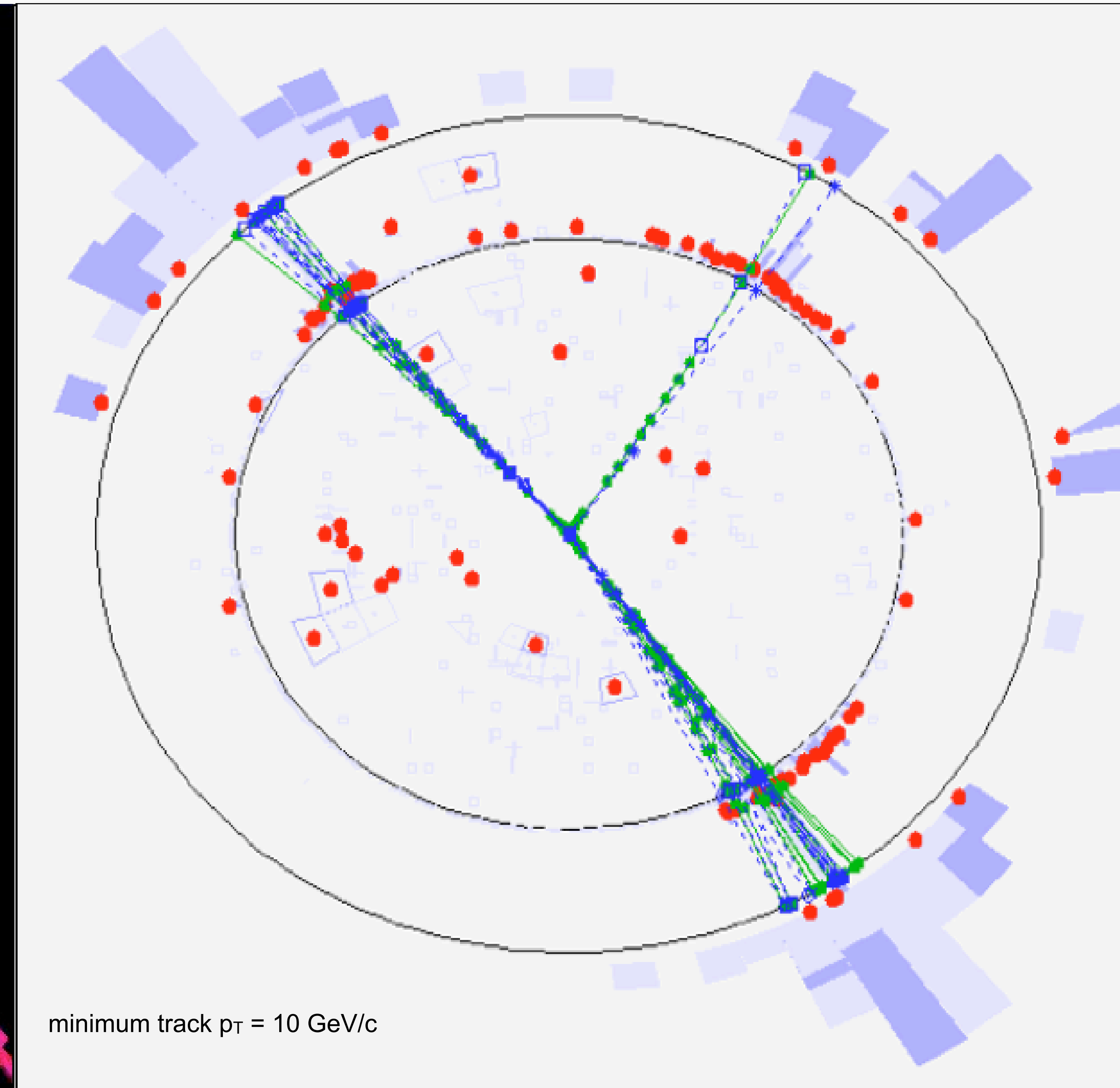
(simulated) Jets in CMS



Jets in Hadron Collider Detectors

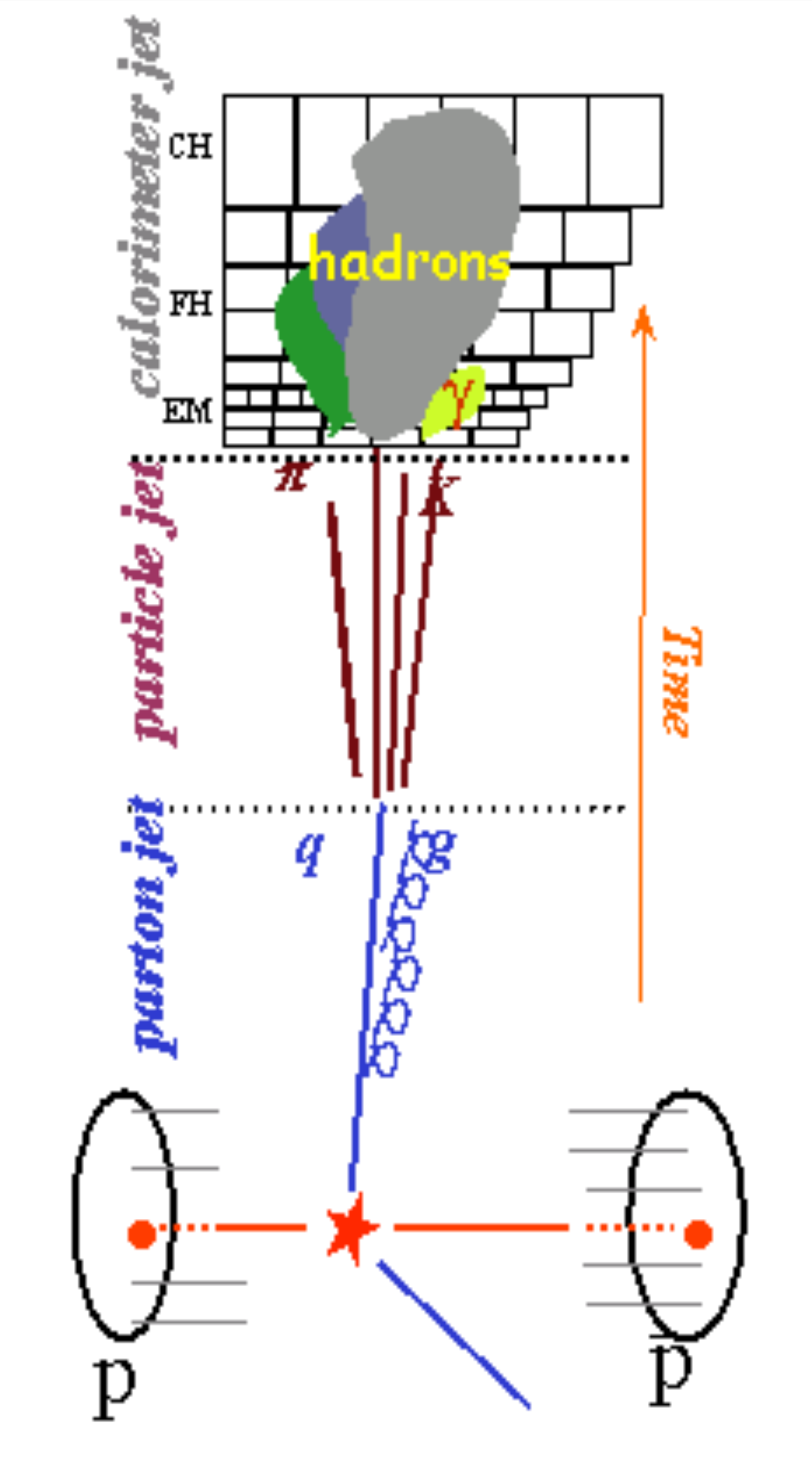
(simulated) Jets in CMS





Requirements

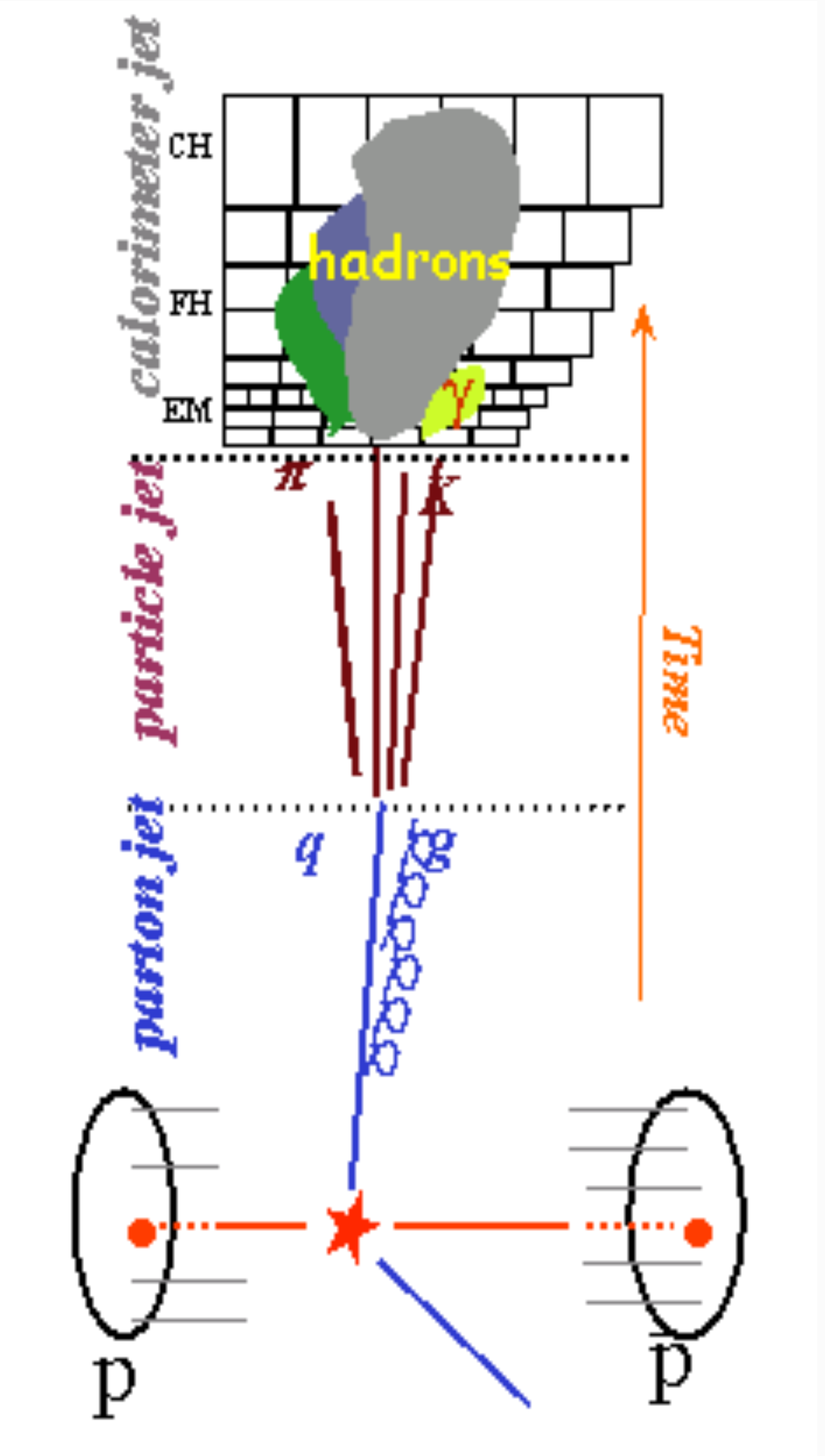
- Applicable at all levels
  - partons, stable particles
    - for theoretical calculations
  - measured objects (calorimeter objects, tracks, etc)
  - and always find the same jet



Requirements

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

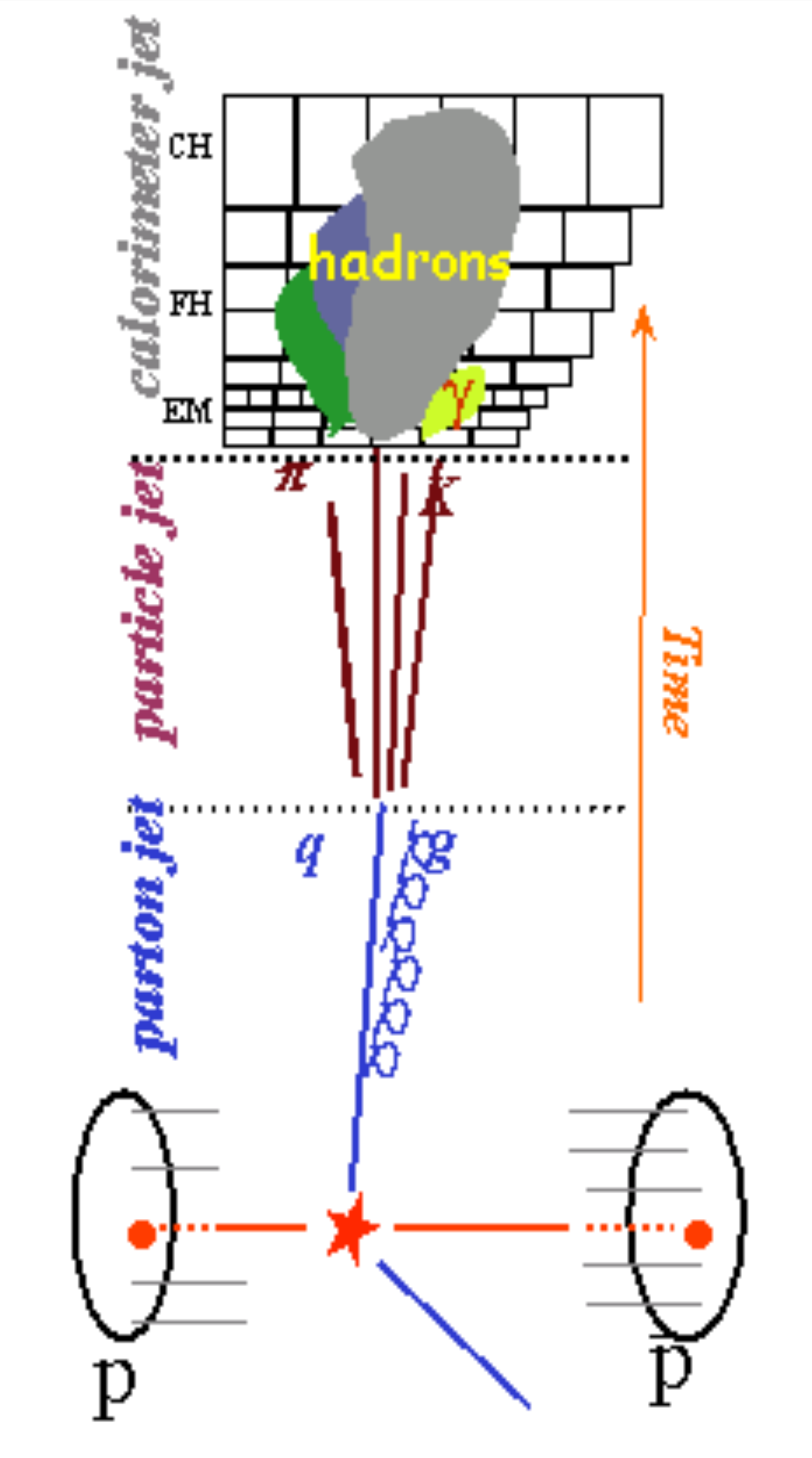
- Independent of the very details of the detector
  - example : granularity of the calorimeter, energy response,...



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,...
- Easy to implement !

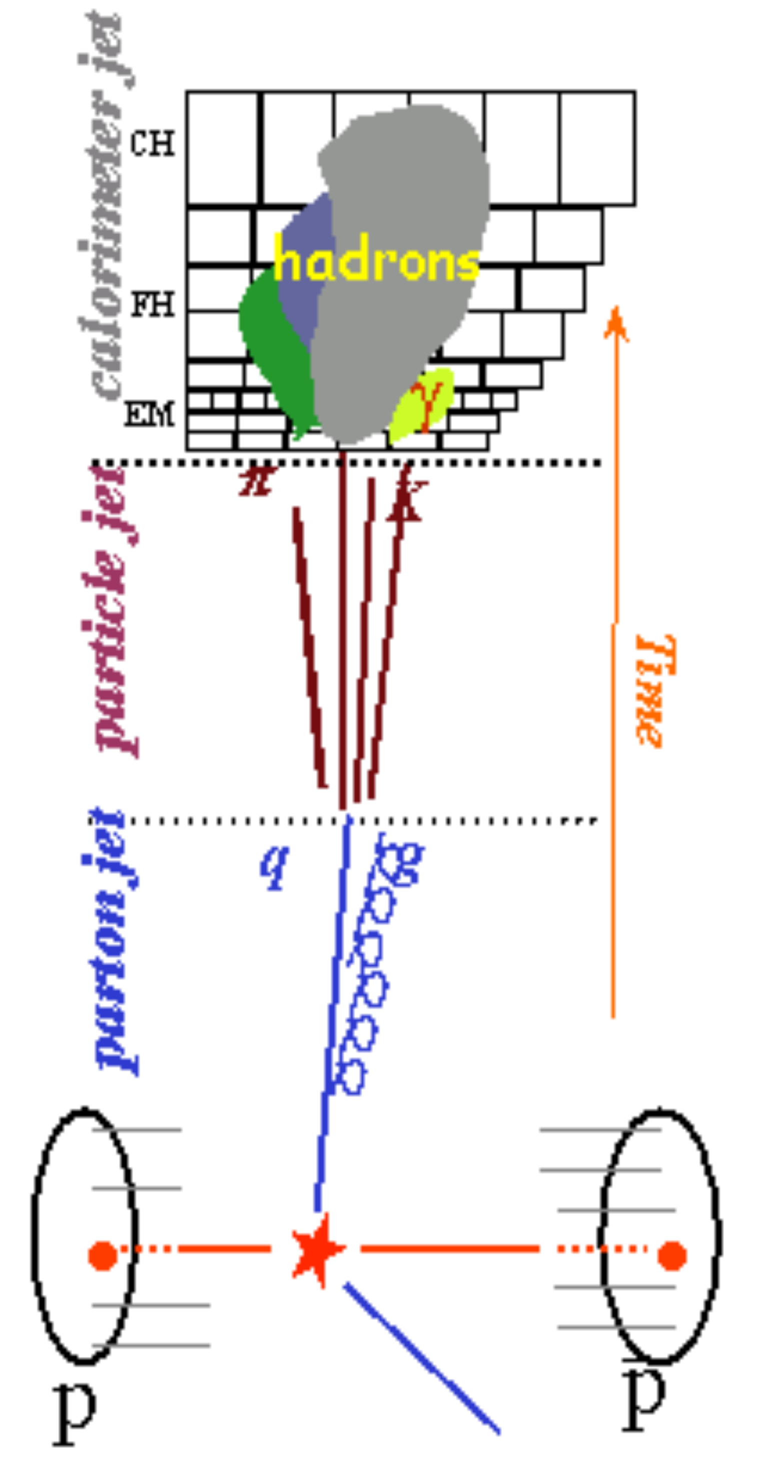


Requirements

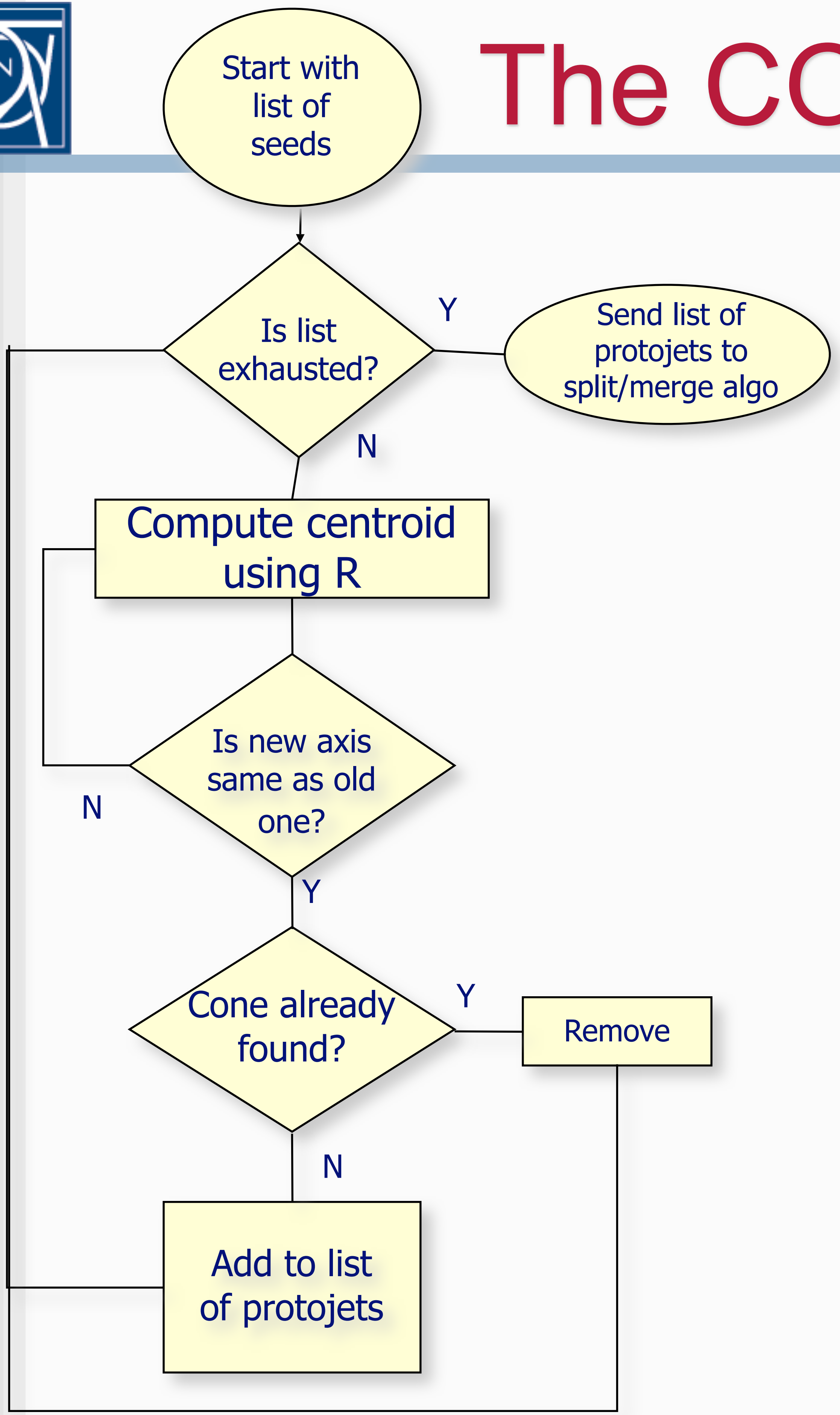
- Applicable at all levels
  - partons, stable particles
    - for theoretical calculations
  - measured objects (calorimeter objects, tracks, etc)
  - and always find the same jet
- Independent of the very details of the detector
  - example : granularity of the calorimeter, energy response,...
- Easy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

$$P_{parton} \Longleftrightarrow P_{jet}$$

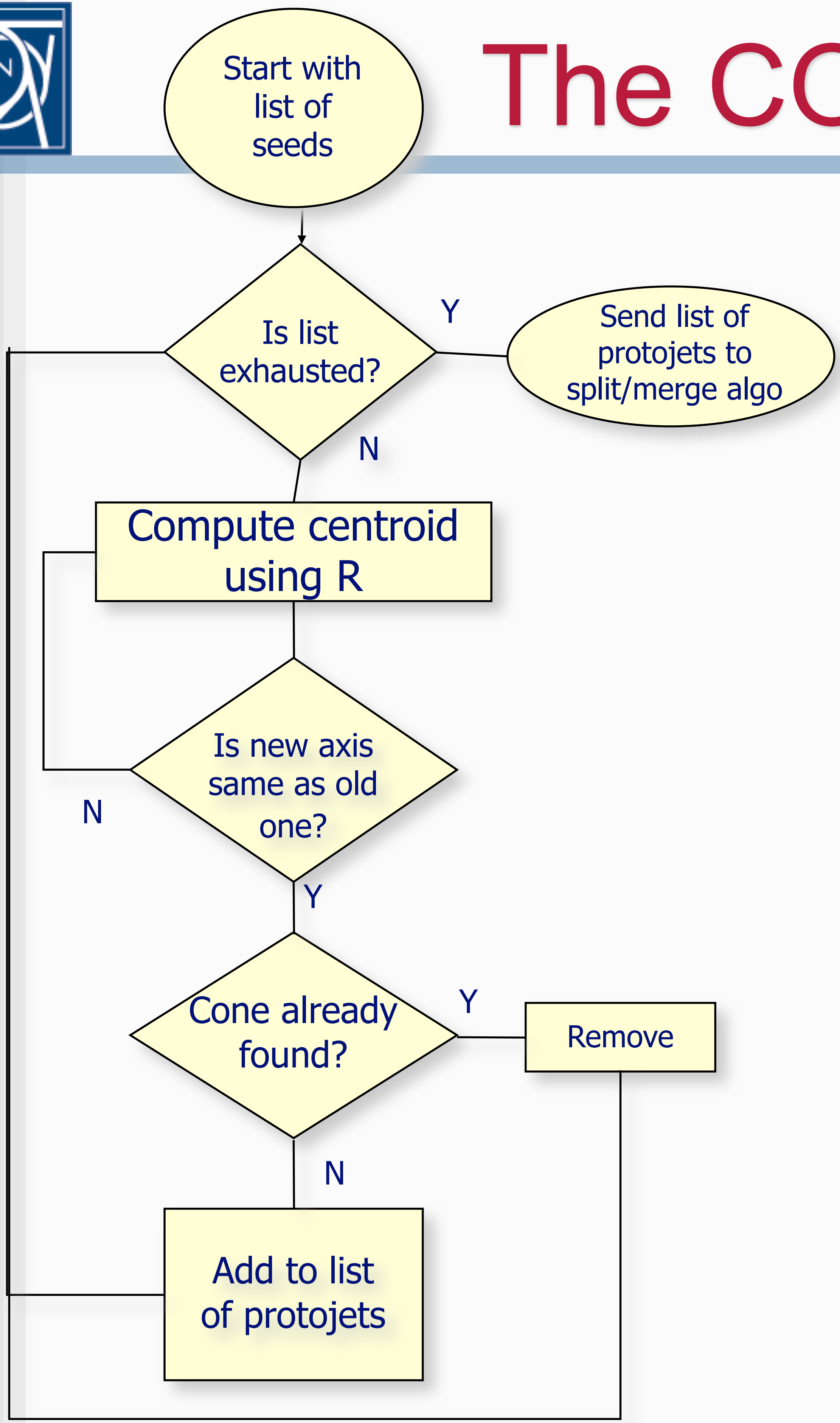
Energy
Momentum
angle



The CONE algorithm

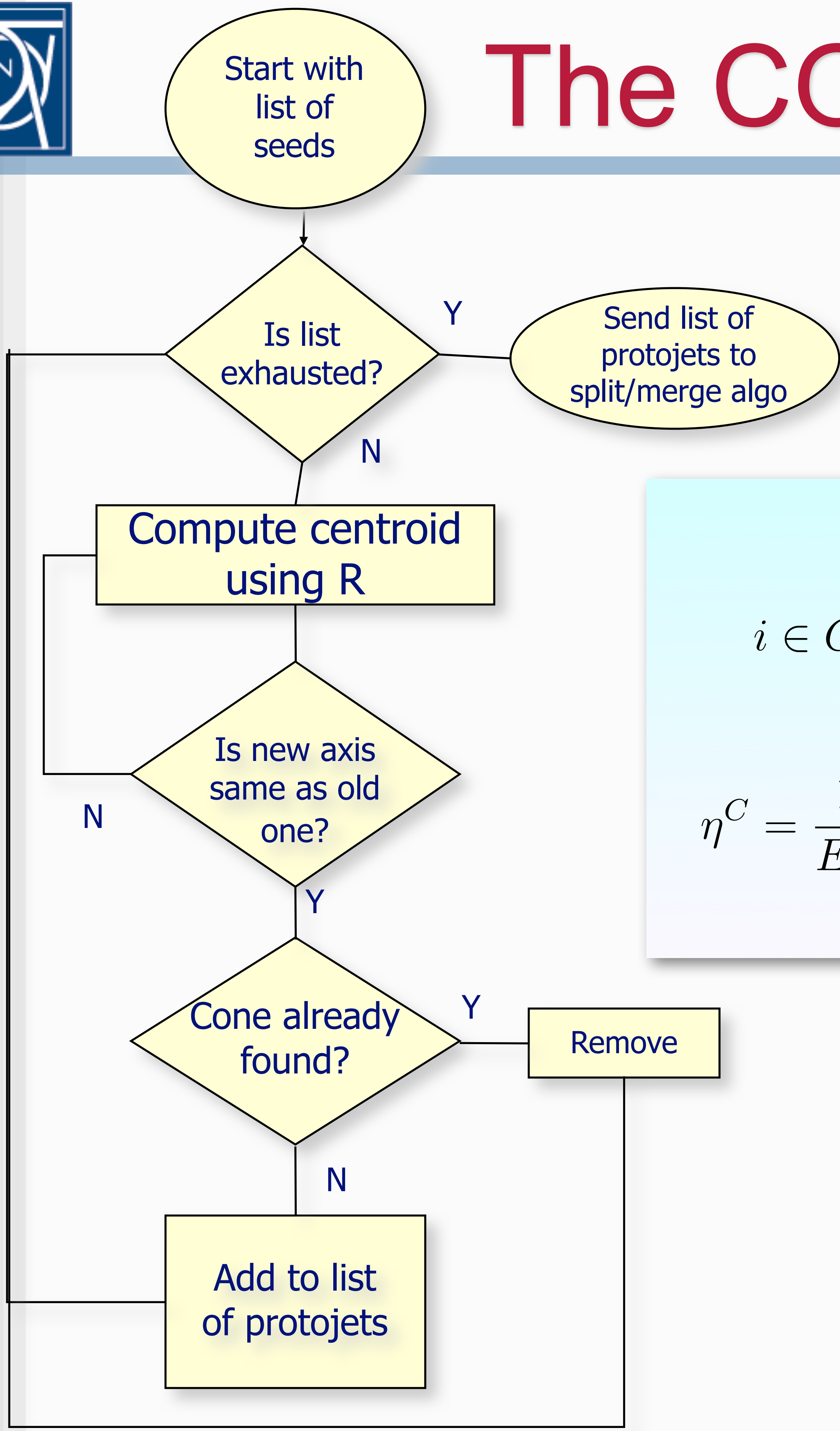


The CONE algorithm



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

The CONE algorithm



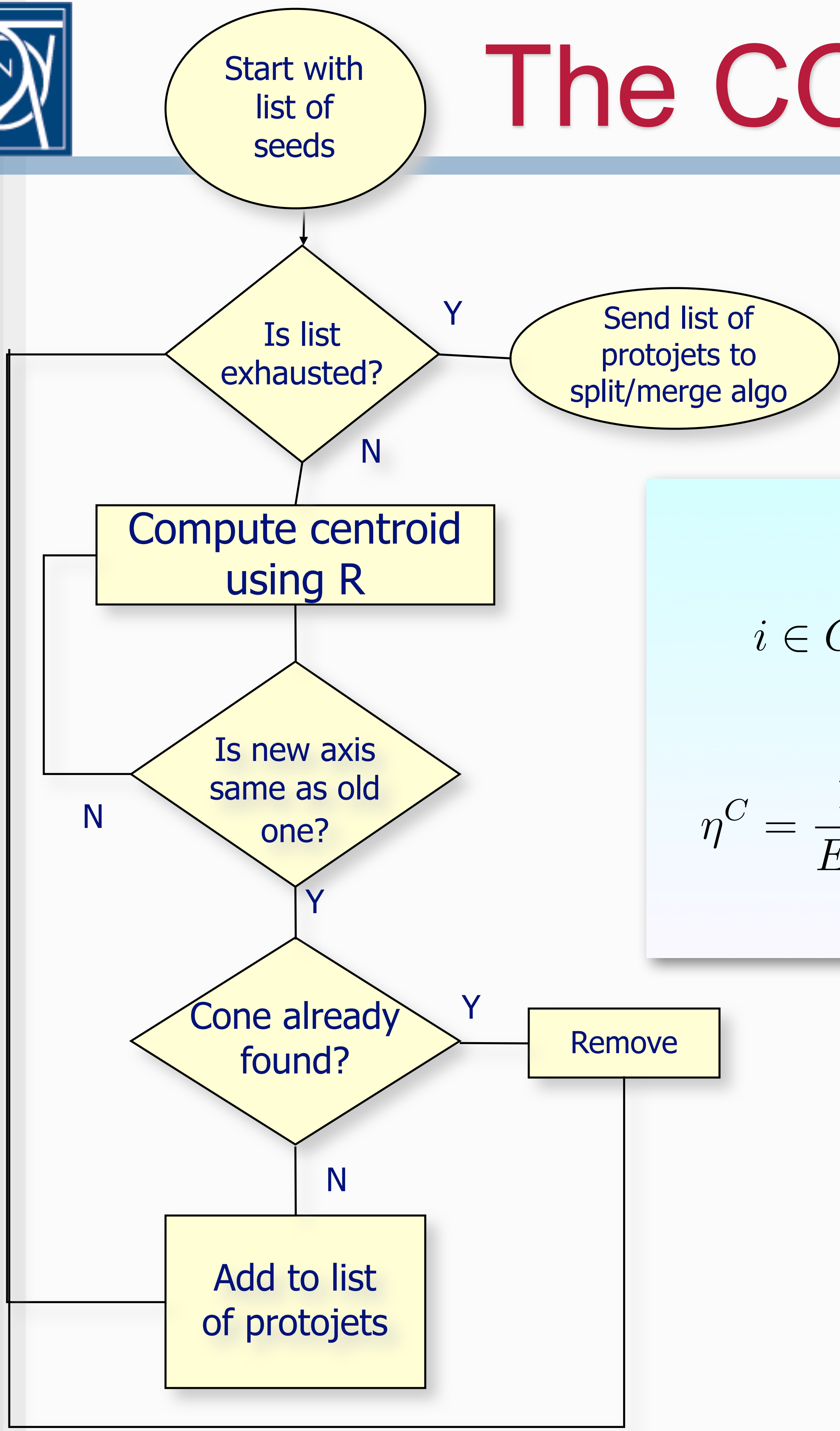
Seeds: for example, energy deposits with transverse energy ($E_T = E \sin\theta$) > 2 GeV in a tower of the calorimeter

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

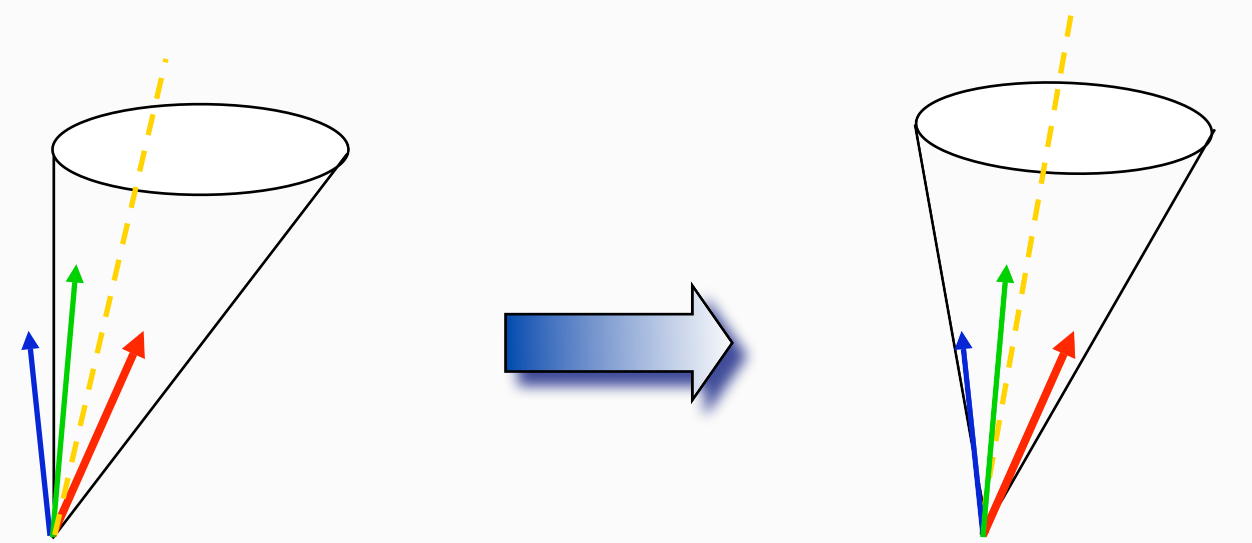


Seeds: for example, energy deposits with transverse energy ($E_T = E \sin\theta$) > 2 GeV in a tower of the calorimeter

Centroid (one possible def) :

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

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



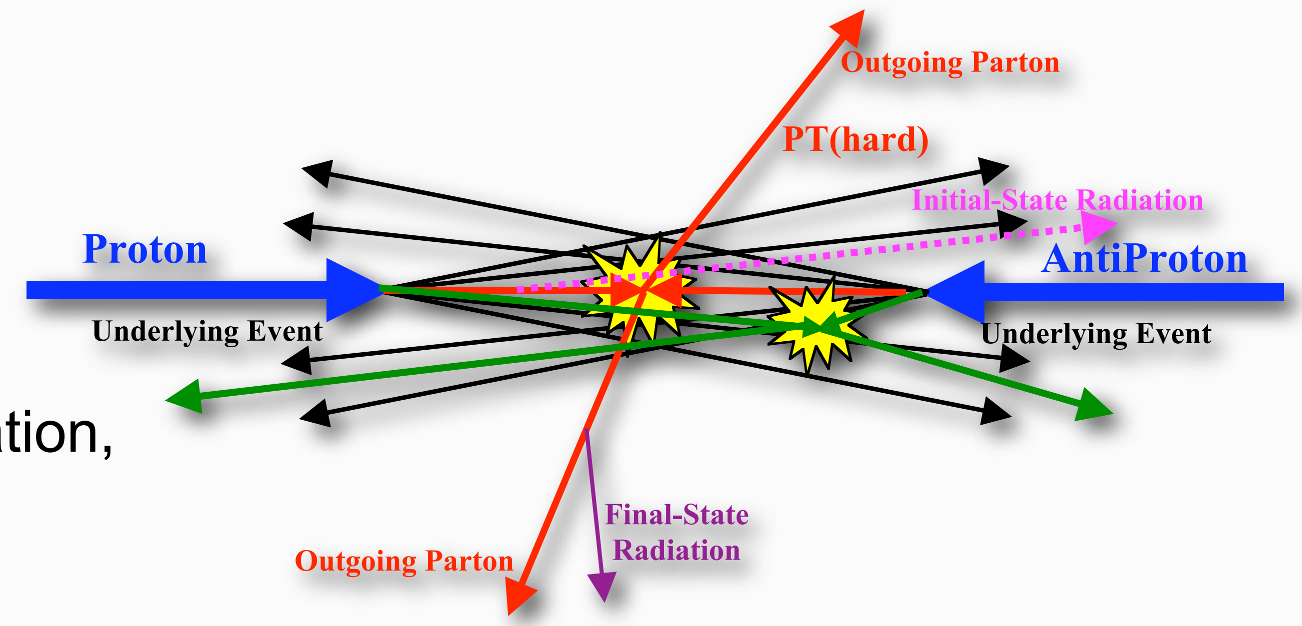
Further difficulties

ETH Institute for Particle Physics

- Pile Up : many additional soft proton-proton interactions
  - up to 20 at highest LHC luminosity

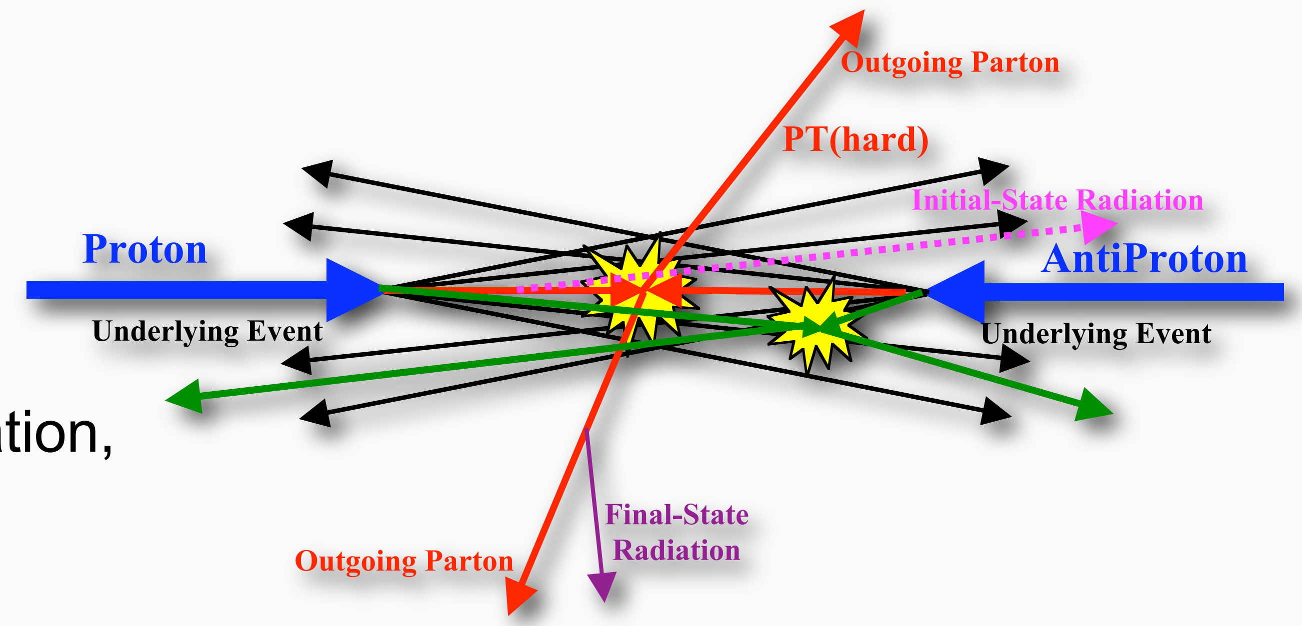
Further difficulties

- Pile Up : many additional soft proton-proton interactions
  - up to 20 at highest LHC luminosity
- Underlying event
  - beam-beam remnants, initial state radiation, multiple parton interactions
  - gives additional energy in the event



Further difficulties

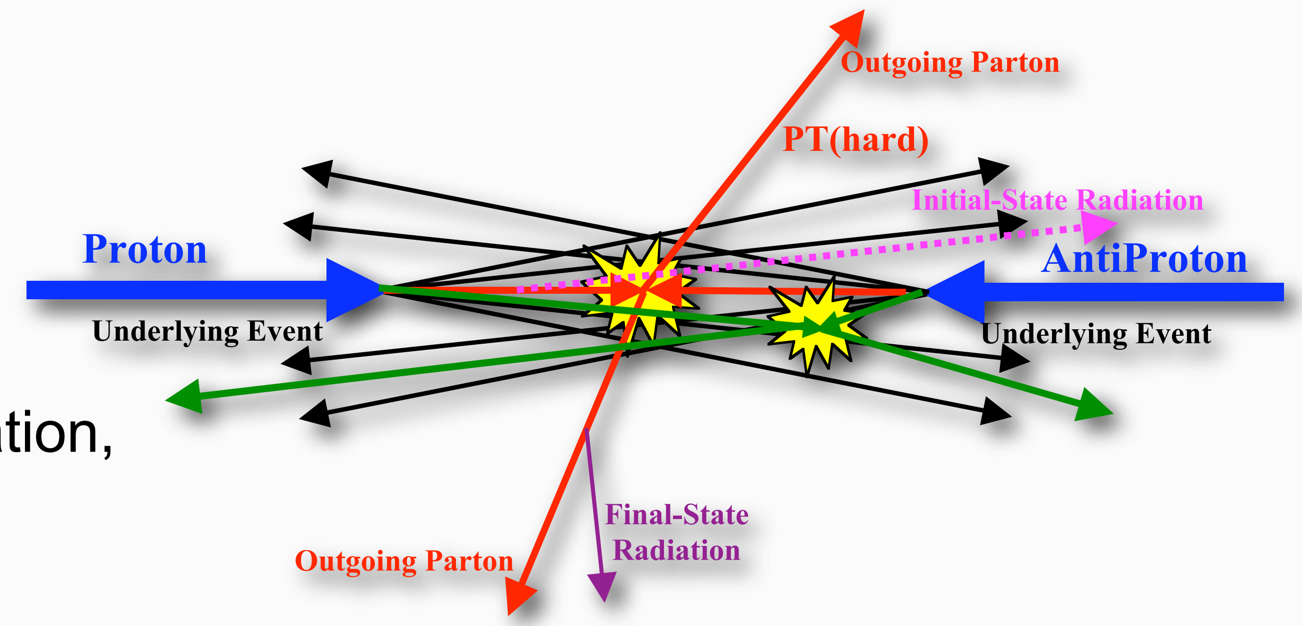
- Pile Up : many additional soft proton-proton interactions
  - up to 20 at highest LHC luminosity
- Underlying event
  - beam-beam remnants, initial state radiation, multiple parton interactions
  - gives additional energy in the event



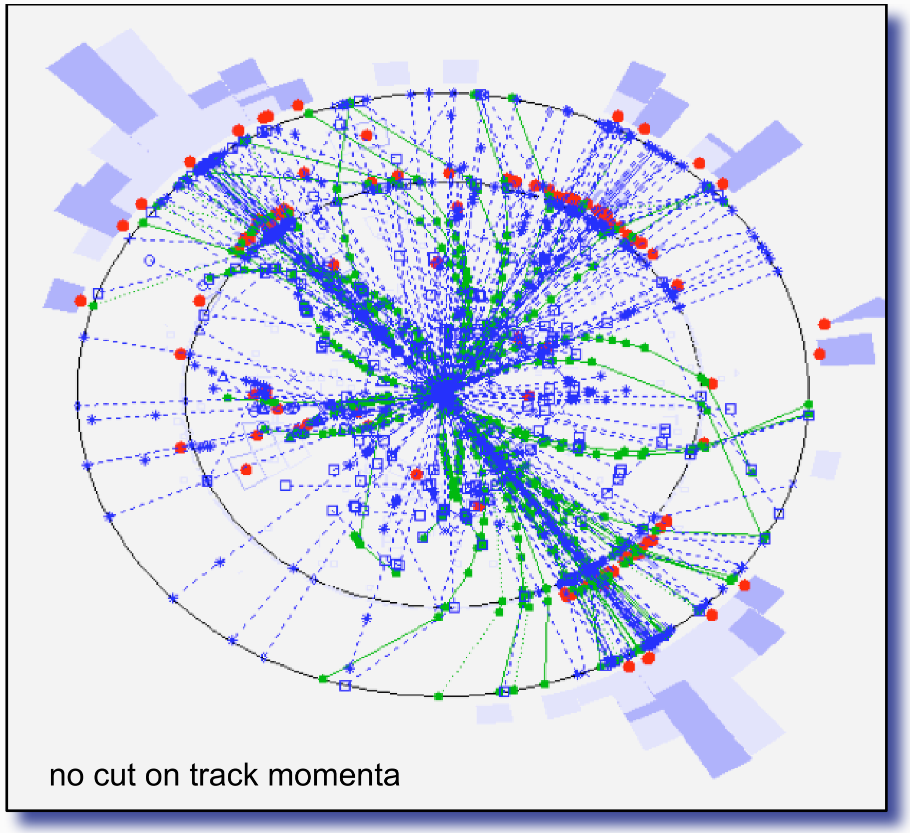
- All this additional energy has nothing to do with jet energies
  - **have to subtract it**

Further difficulties

- Pile Up : many additional soft proton-proton interactions
  - up to 20 at highest LHC luminosity
- Underlying event
  - beam-beam remnants, initial state radiation, multiple parton interactions
  - gives additional energy in the event

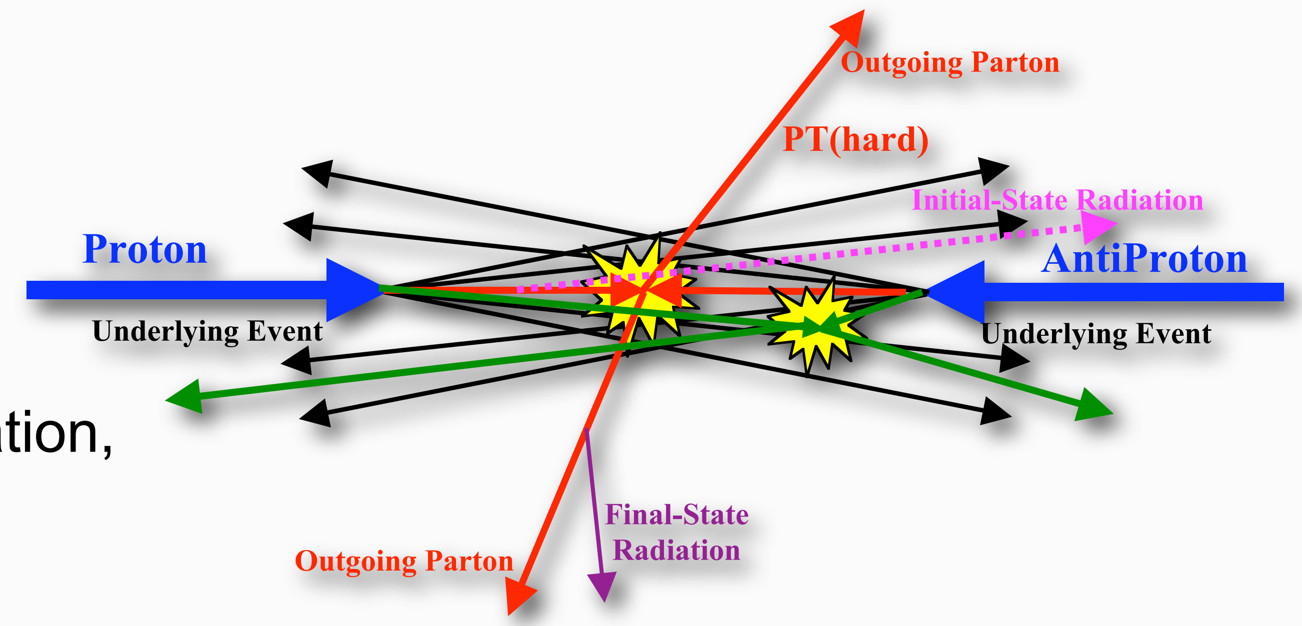


- All this additional energy has nothing to do with jet energies
  - **have to subtract it**

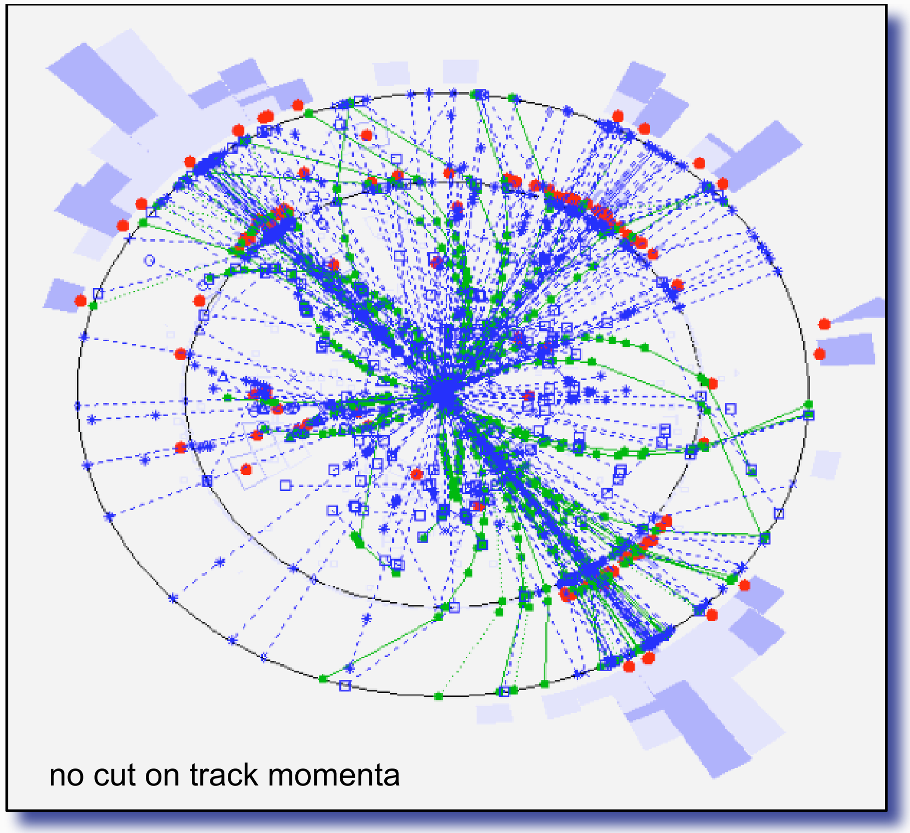


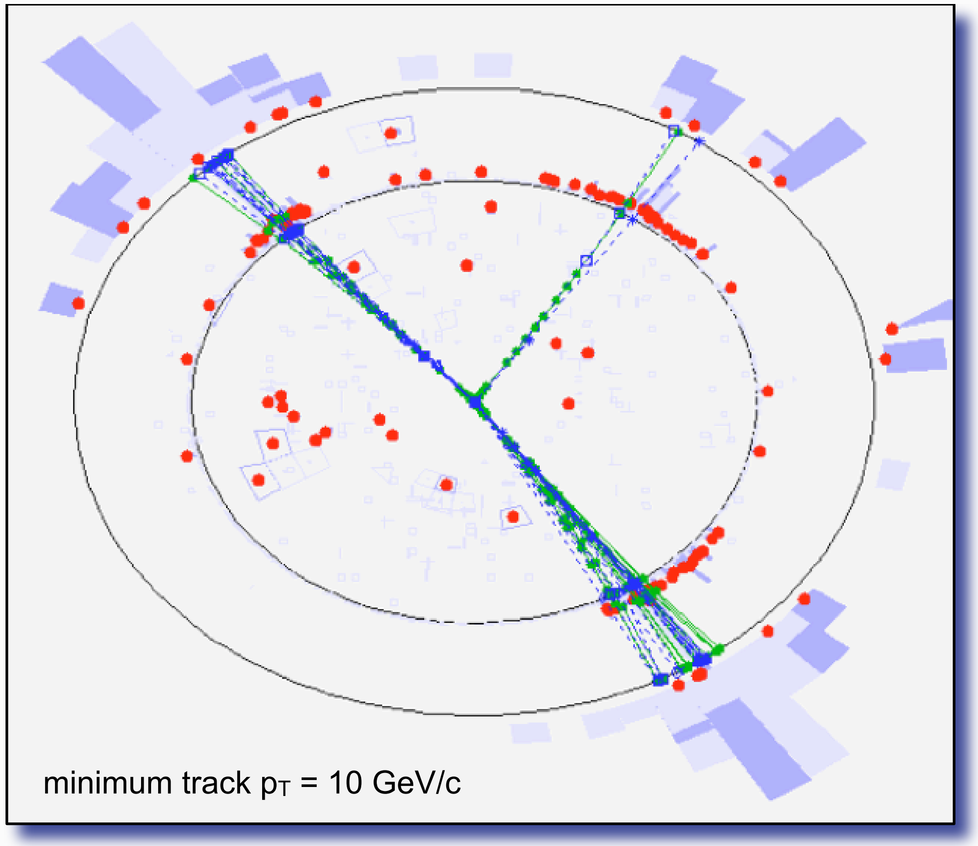
Further difficulties

- Pile Up : many additional soft proton-proton interactions
  - up to 20 at highest LHC luminosity
- Underlying event
  - beam-beam remnants, initial state radiation, multiple parton interactions
  - gives additional energy in the event



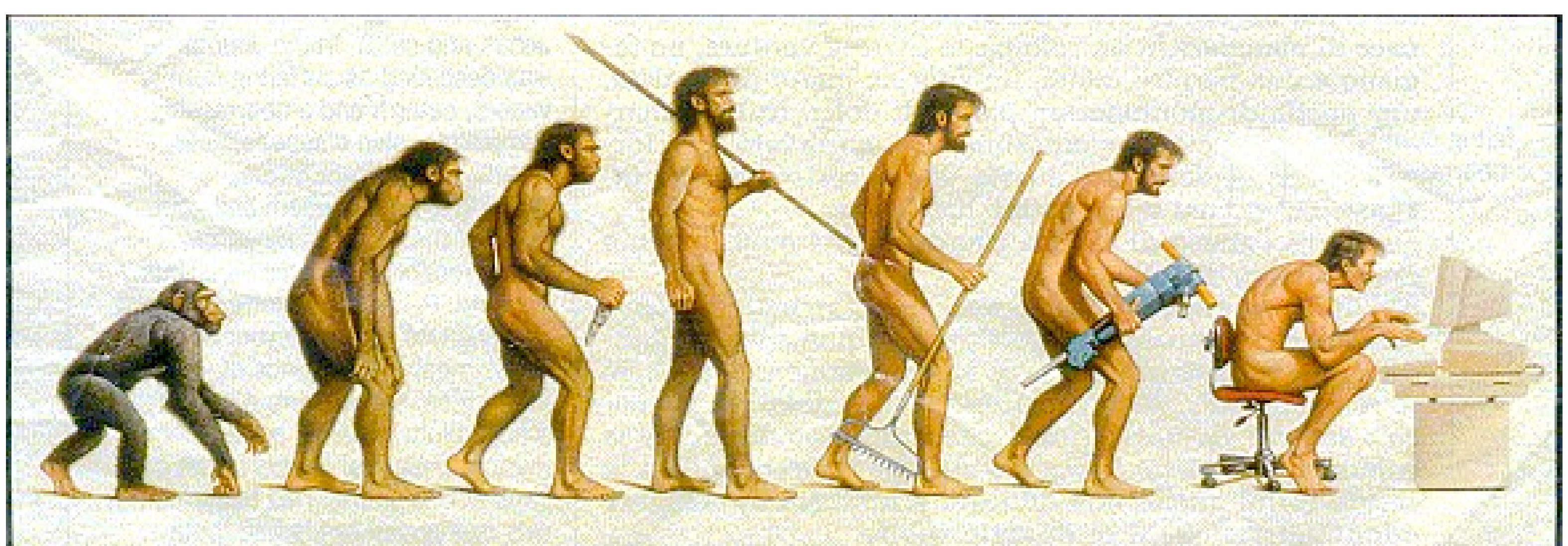
- All this additional energy has nothing to do with jet energies
  - **have to subtract it**





ETH Institute for
Particle Physics

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 1 MByte

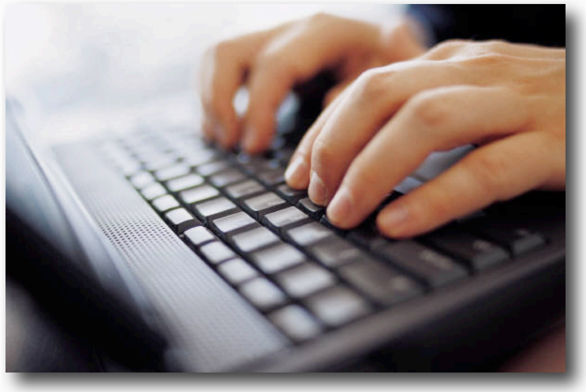
Some numbers

- 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 1 MByte
- 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 1 MByte

- 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

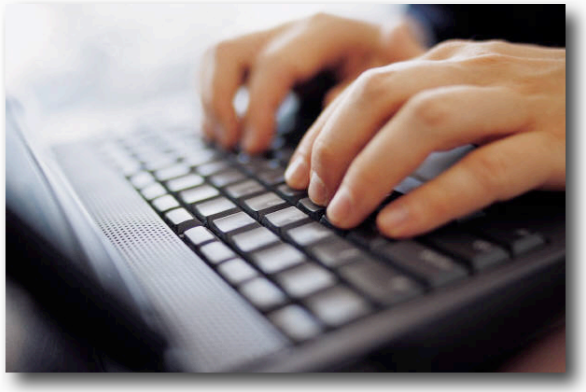


- "prompt" processing
  - Expect to do first reprocessing step within one day
  - Reco time per event on std. CPU: < 5 sec (on lxplus)
  - Note : will have to reprocess several times
    - new/better algorithms, updated calibrations, etc.

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 1 MByte

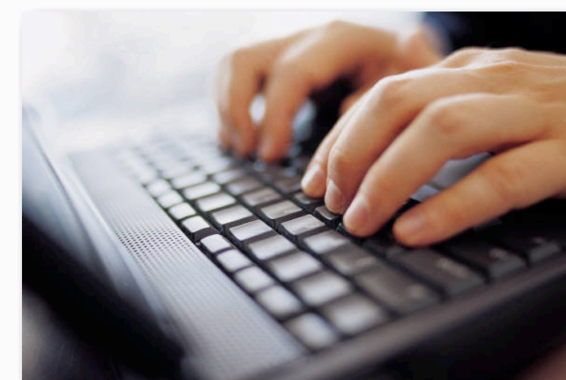
- CMS will record ~100k top-quark events per day
  - among about 10^7 events in total per day
  - will have roughly 150 “physics” days per year
  - thus about 10^9 evts/year, a few Pbyte



- “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

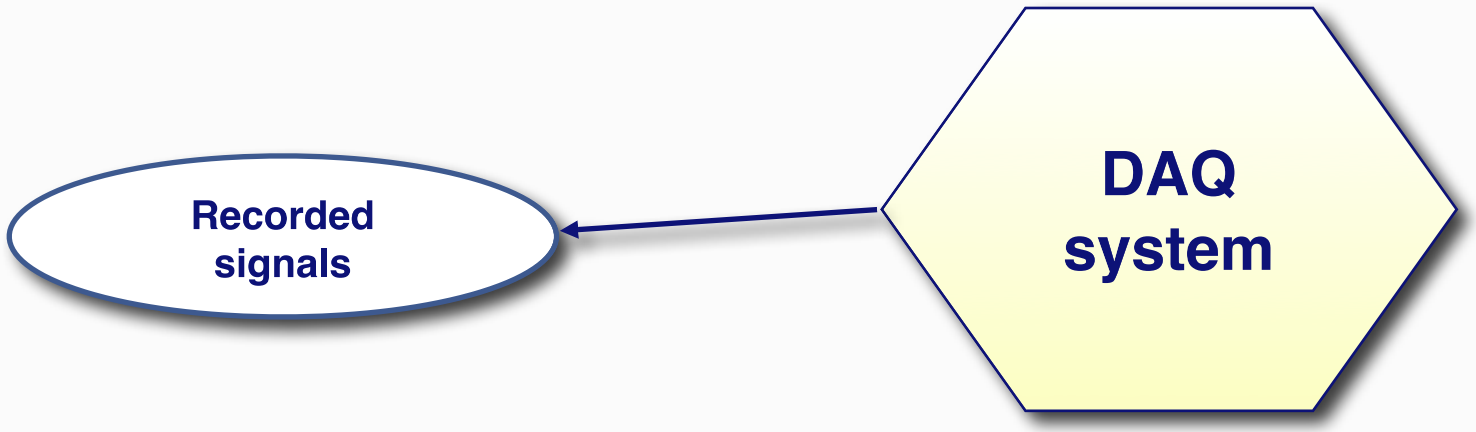
- Examples from CMS, estimates
  - Rate of events streaming out from High-Level Trigger farm ~150 Hz
  - each event has a size of the order of 1 MByte
- CMS will record ~100k top-quark events per day
  - among about 10^7 events in total per day
  - will have roughly 150 “physics” days per year
  - thus about 10^9 evts/year, a few Pbyte



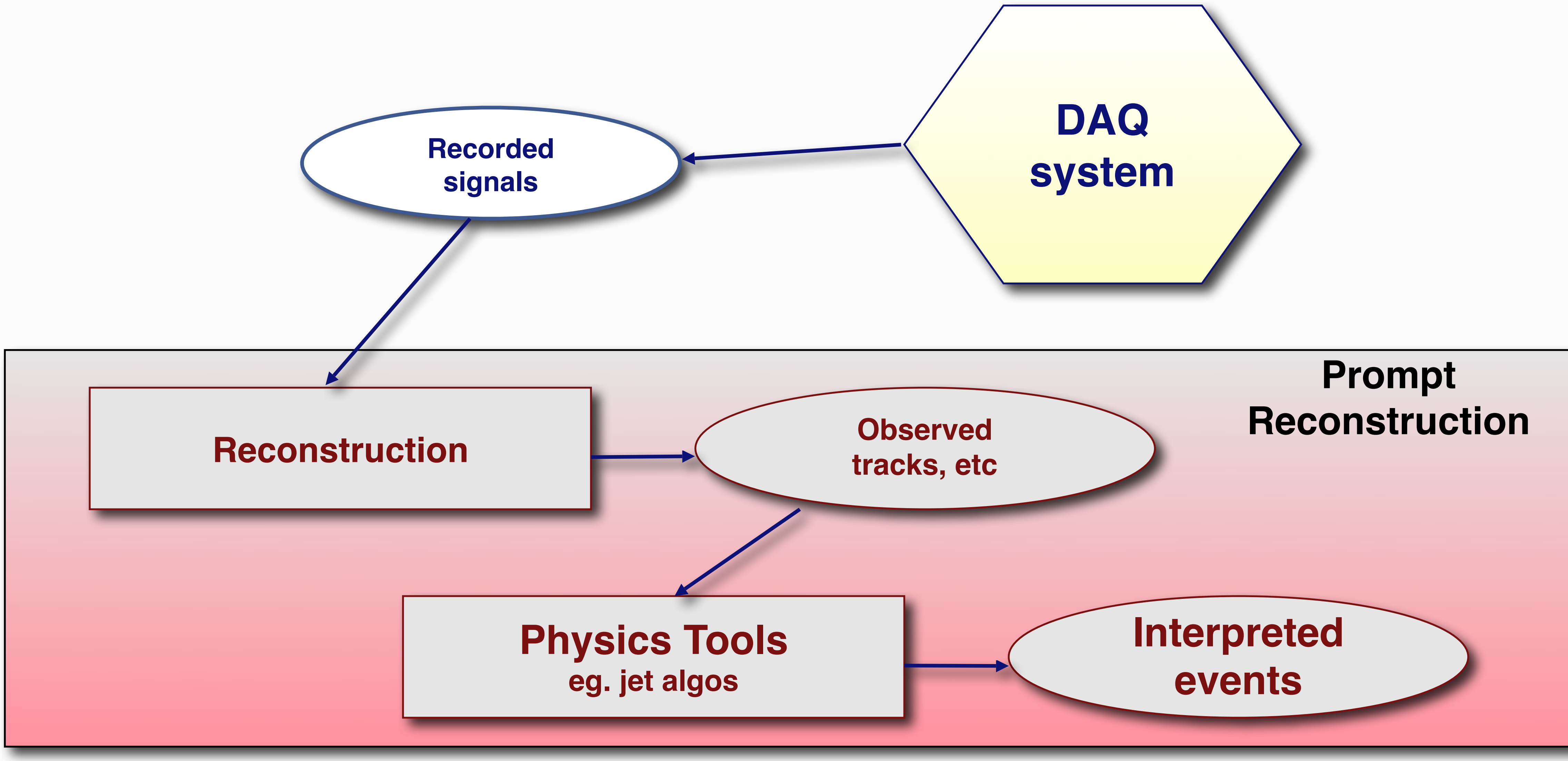
- “prompt” processing
  - Expect to do first reprocessing step within one day
  - Reco time per event on std. CPU: < 5 sec (on lxplus)
  - Note : will have to reprocess several times
    - new/better algorithms, updated calibrations, etc.
- Expect to simulate several 100s to 1000s of millions of events
  - 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

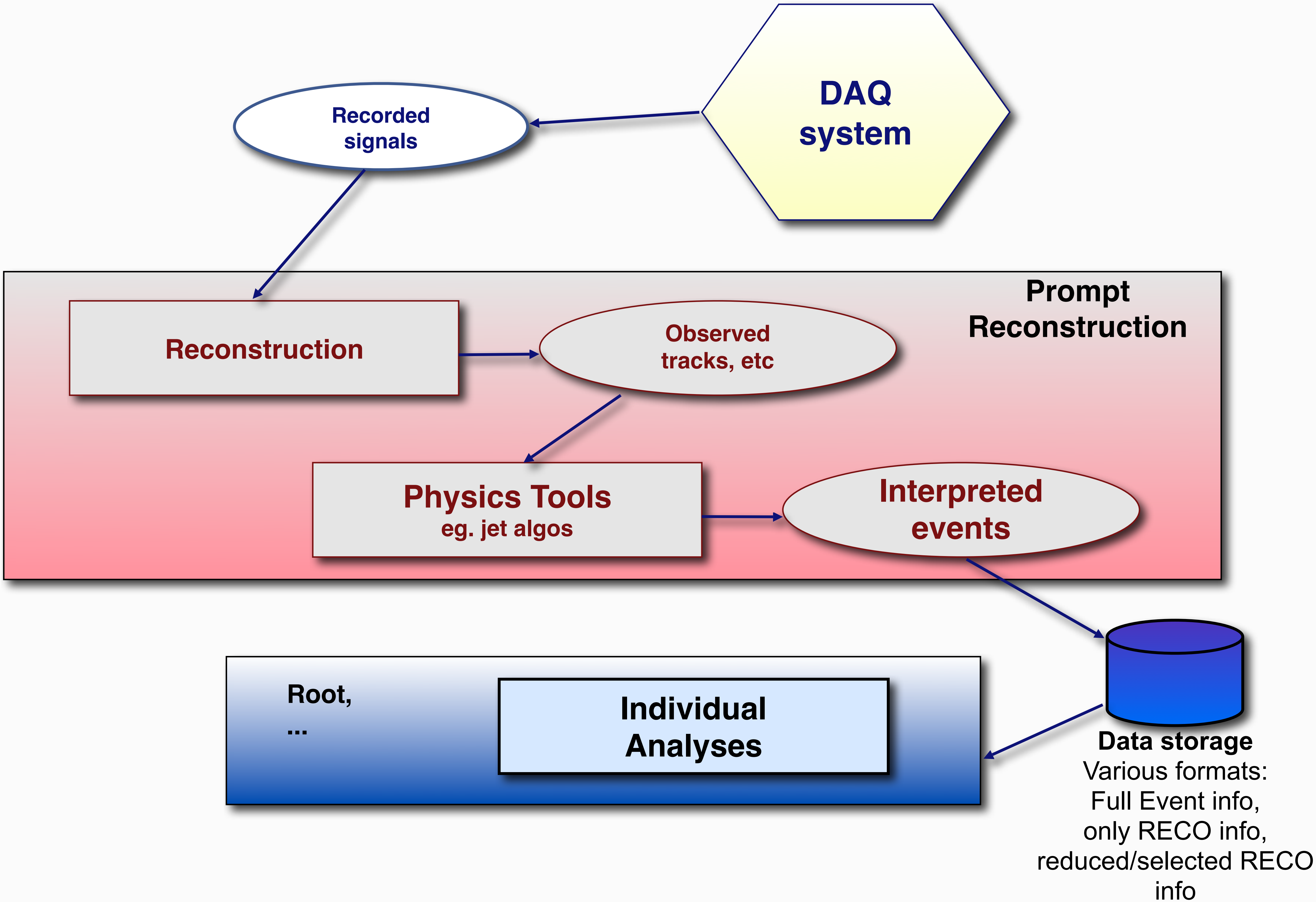


RECO flow



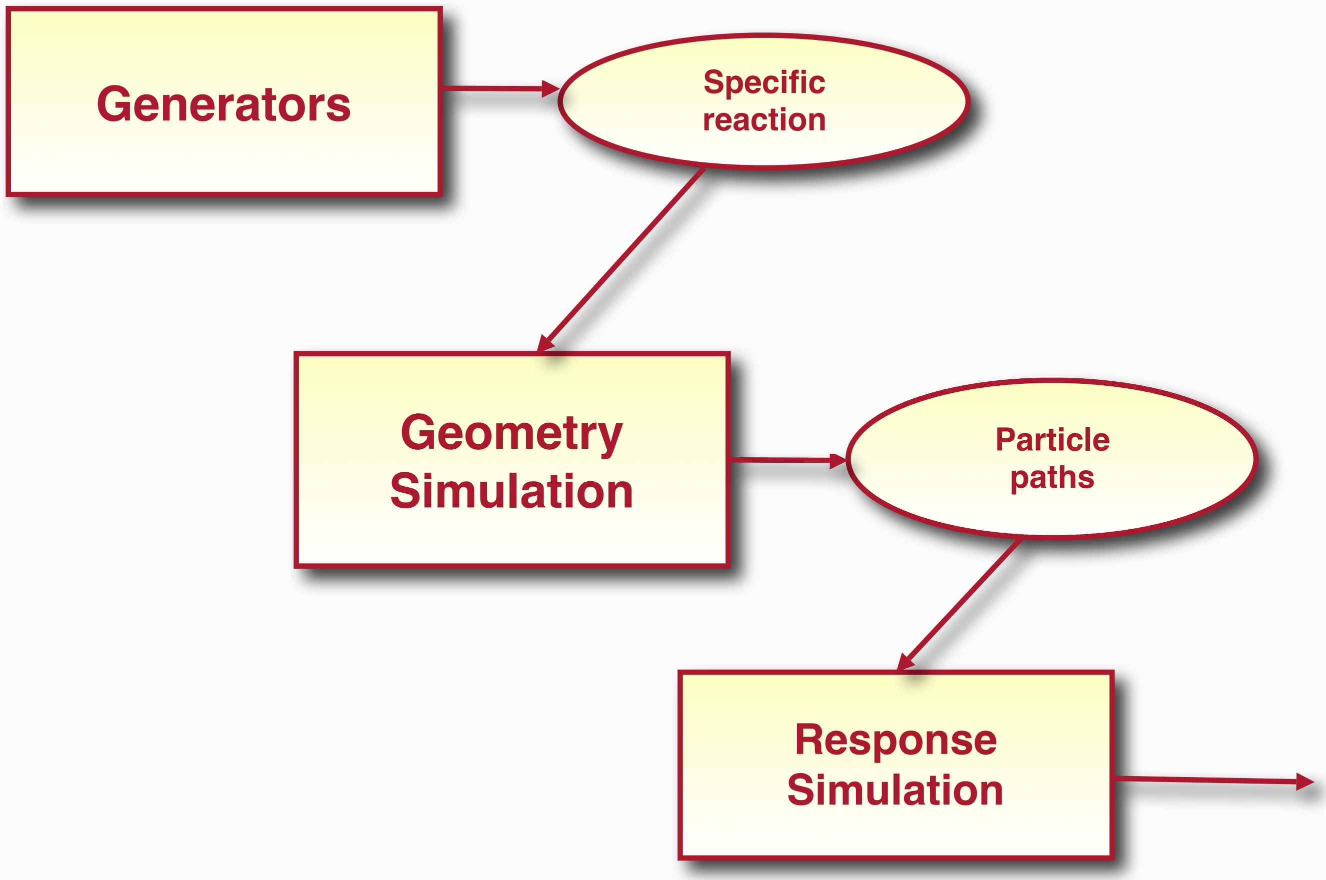
DAQ system
Recorded signals
Prompt Reconstruction
Reconstruction
Observed tracks, etc
Physics Tools
eg. jet algos
Interpreted events

RECO flow

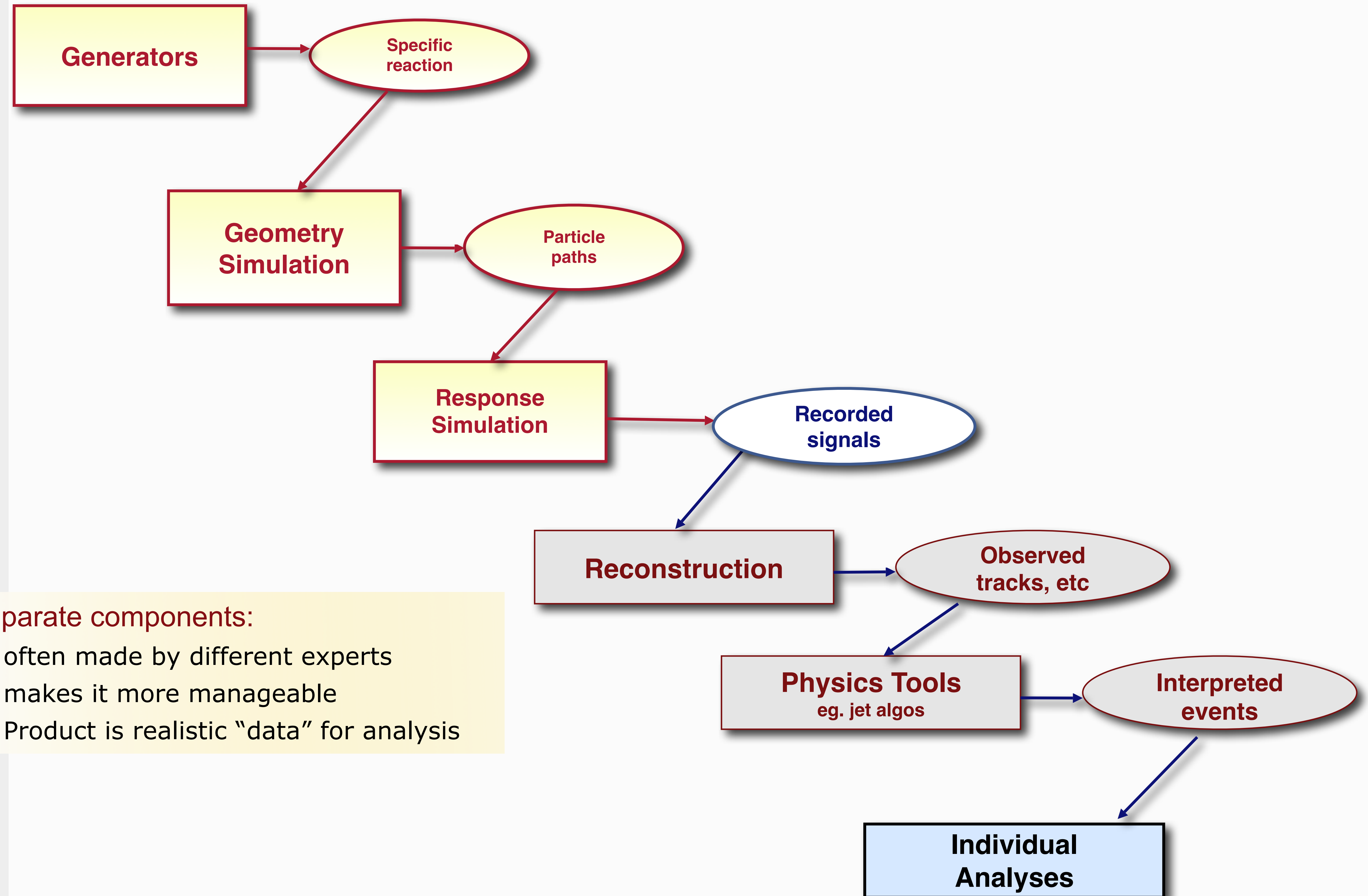


DAQ system
Recorded signals
Prompt Reconstruction
Reconstruction
Observed tracks, etc
Physics Tools
eg. jet algos
Interpreted events
Root,
...
Individual Analyses
Data storage
Various formats:
Full Event info,
only RECO info,
reduced/selected RECO info

Flow of simulated data



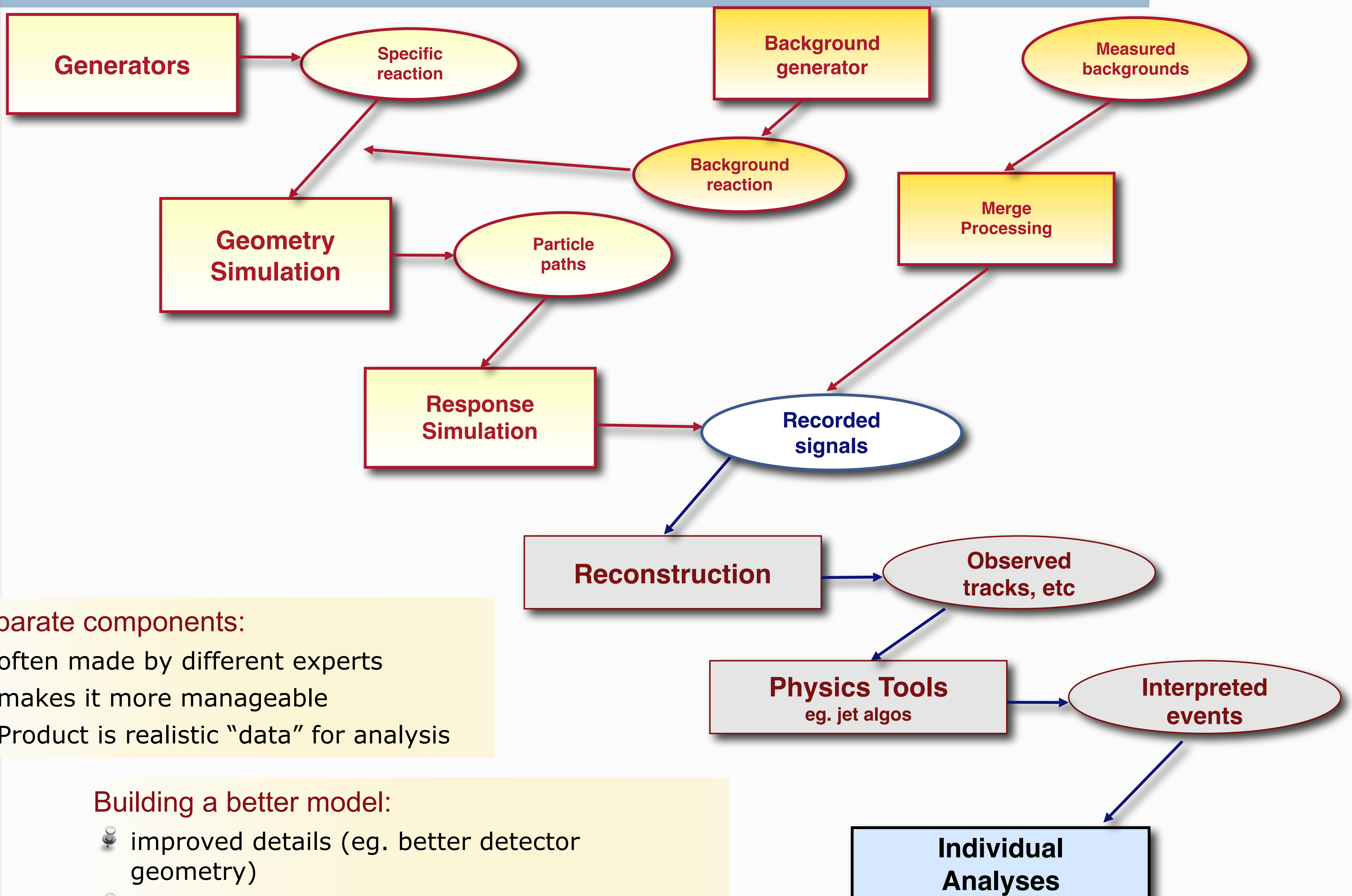
Flow of simulated data



Separate components:

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

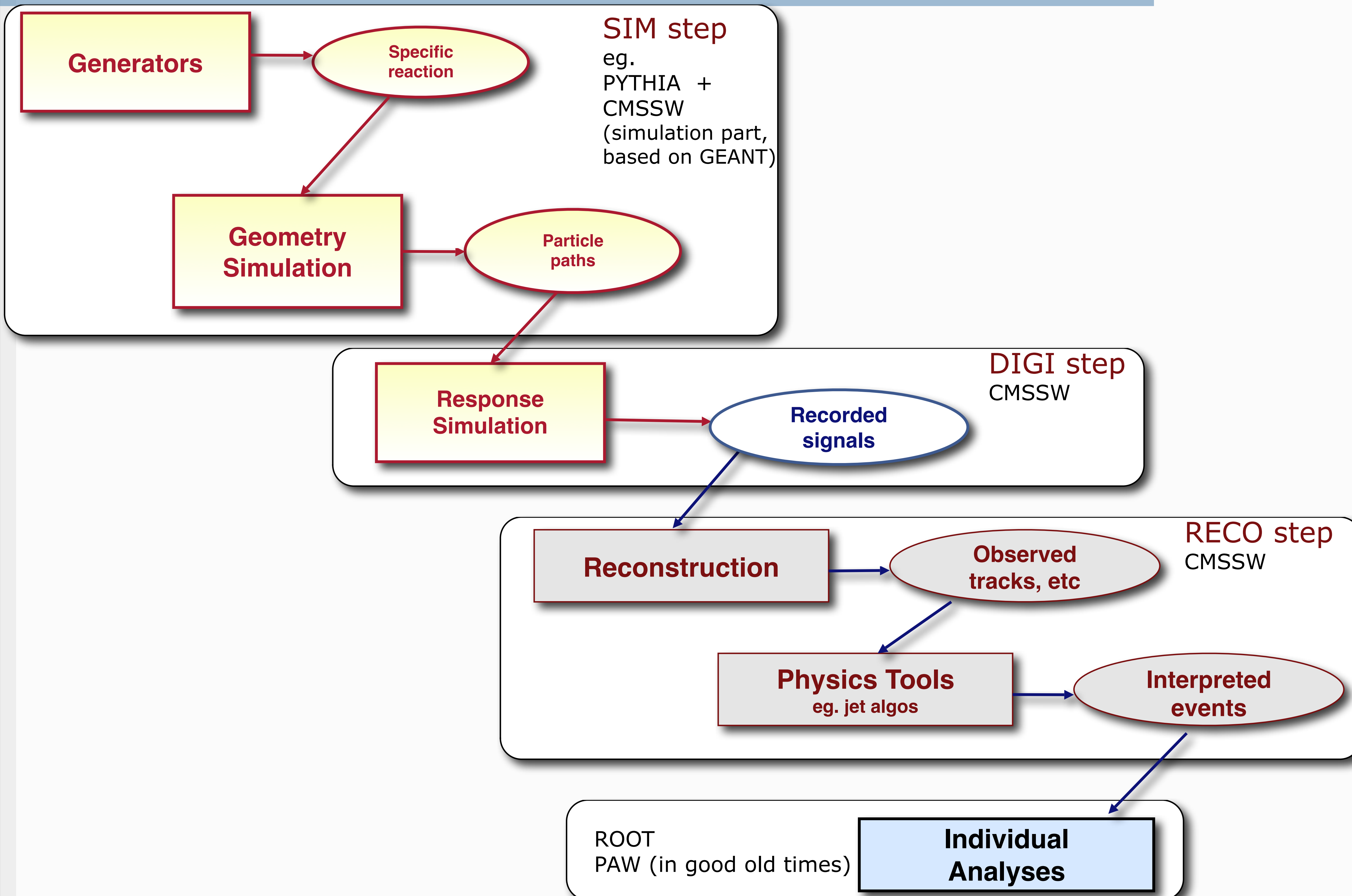
Flow of simulated data



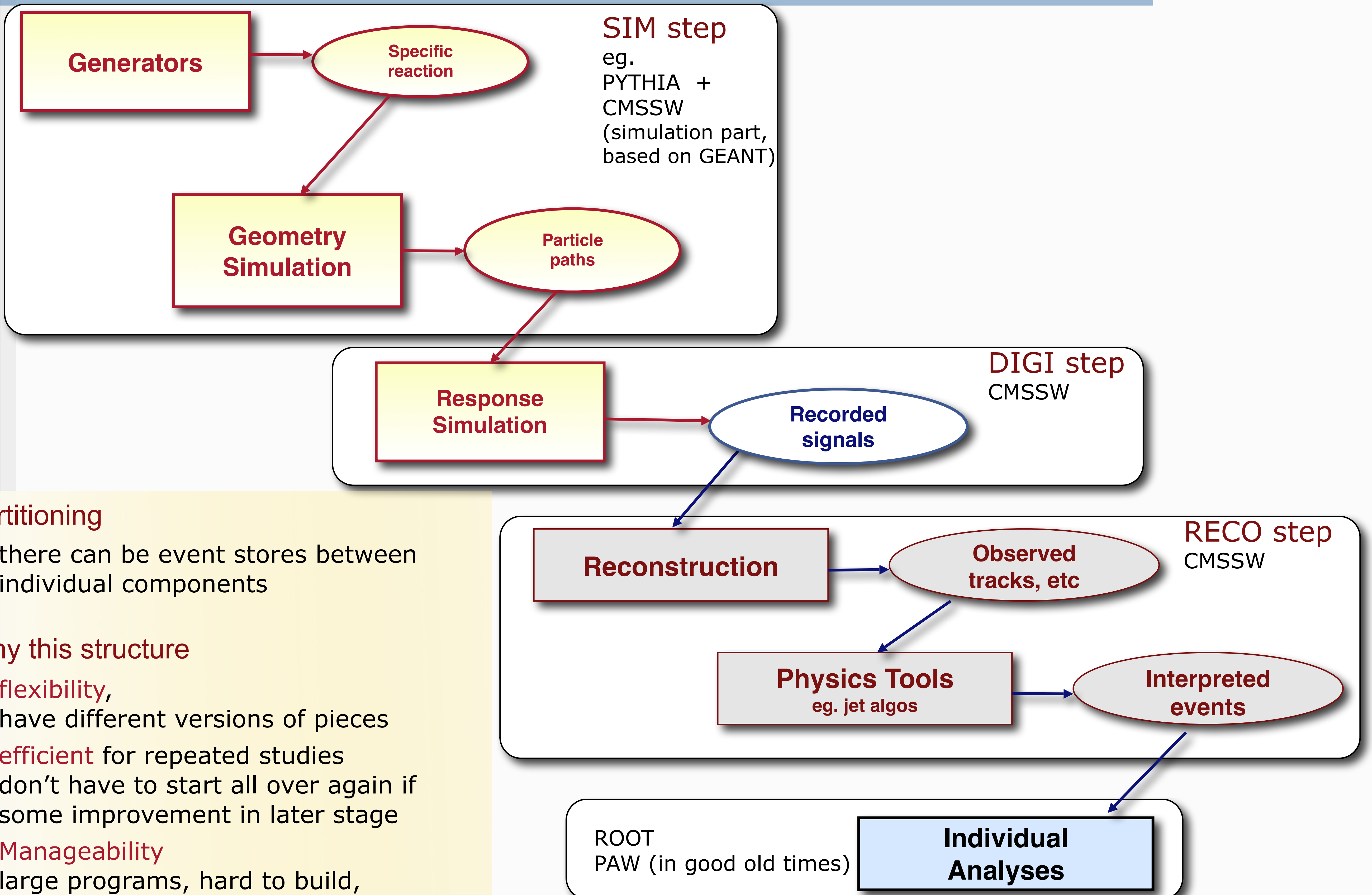
Separate components:
- often made by different experts
- makes it more manageable
- Product is realistic “data” for analysis

Building a better model:
- improved details (eg. better detector geometry)
- real backgrounds

Partitioning production systems



Partitioning production systems



Partitioning

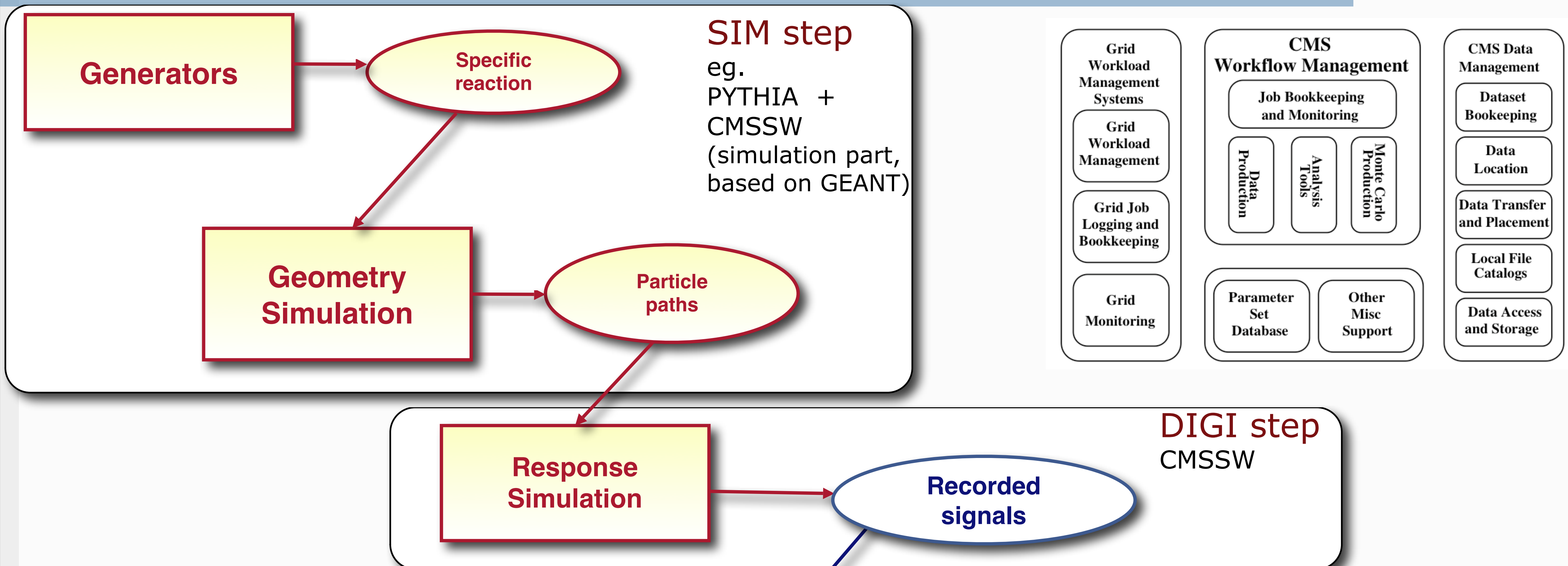
- there can be event stores between individual components

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

ETH Institute for Particle Physics

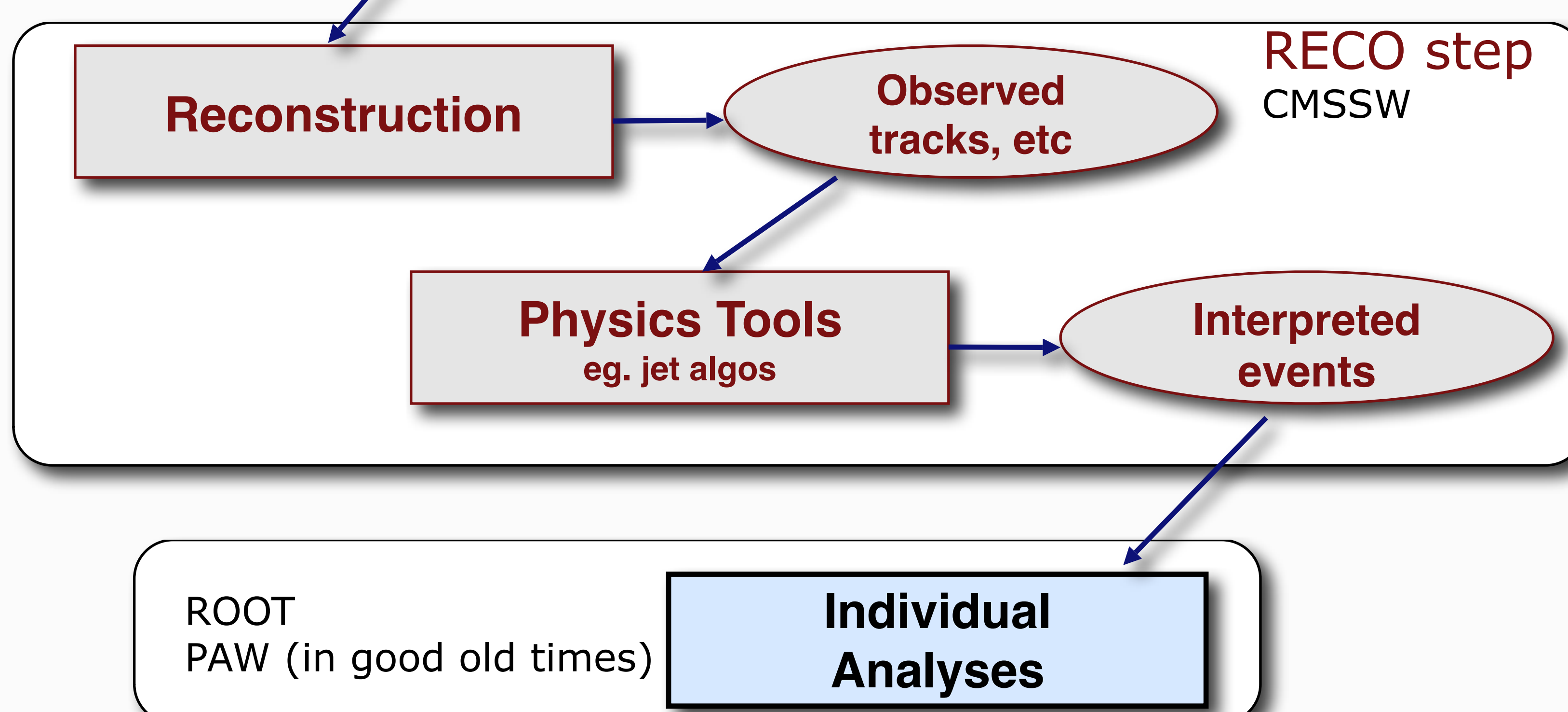


Partitioning

- there can be event stores between individual components

Why this structure

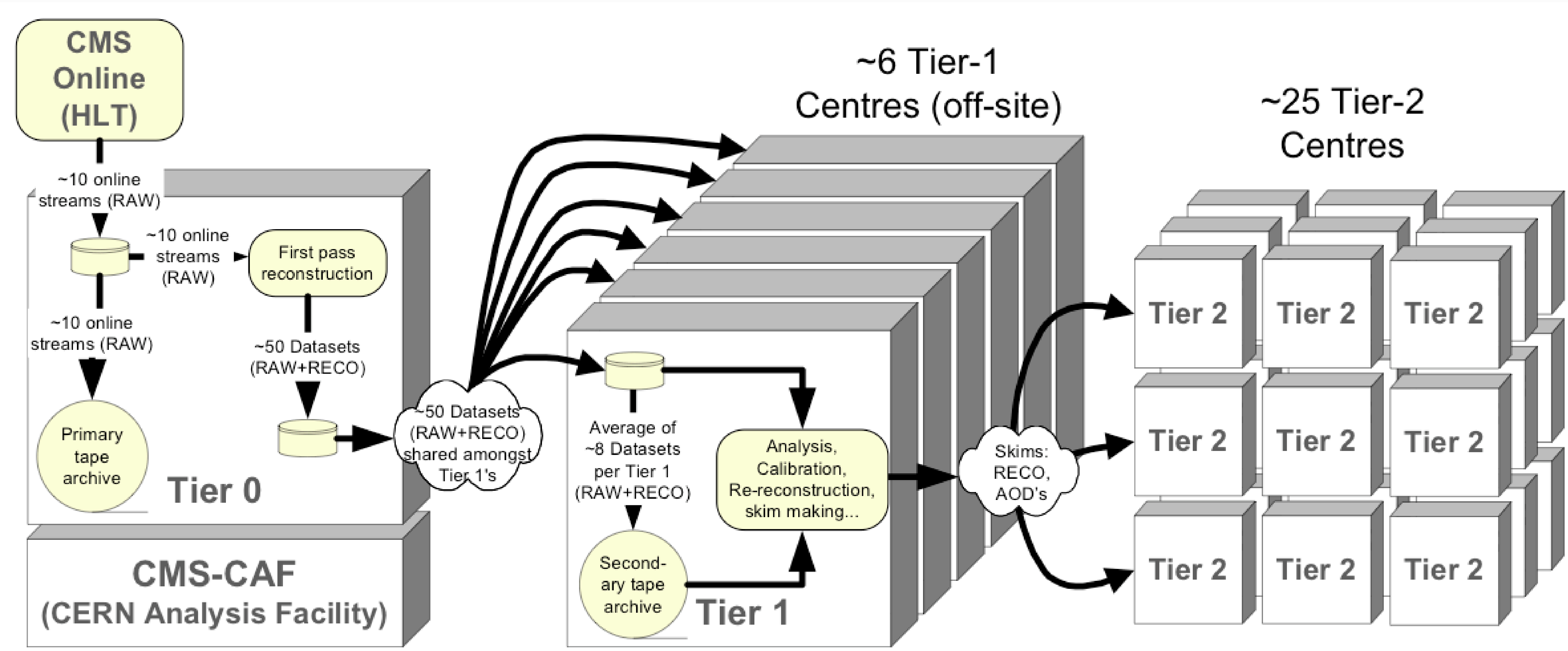
- flexibility,
  have different versions of pieces
- efficient for repeated studies
  don't have to start all over again if some improvement in later stage
- Manageability
  large programs, hard to build, understand, debug, maintain, ...



GRID distribution : example

ETH Institute for Particle Physics

CMS Computing Model: Data Flow



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

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

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

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

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

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

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!

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!