Data analysis

Faculty of
Mathematics and Physics
University of Ljubljana

(from raw data to physics results)

- From raw data to summary data
  - ("Raw data -> DST")
  - track fitting
  - momentum determination
  - ~~calorimetry (cluster reconstr.)~~
  - particle identification (Cherenkov angle)

- Calibration
  - tracking detectors
  - data (RICH) and MC (tracking) calibration

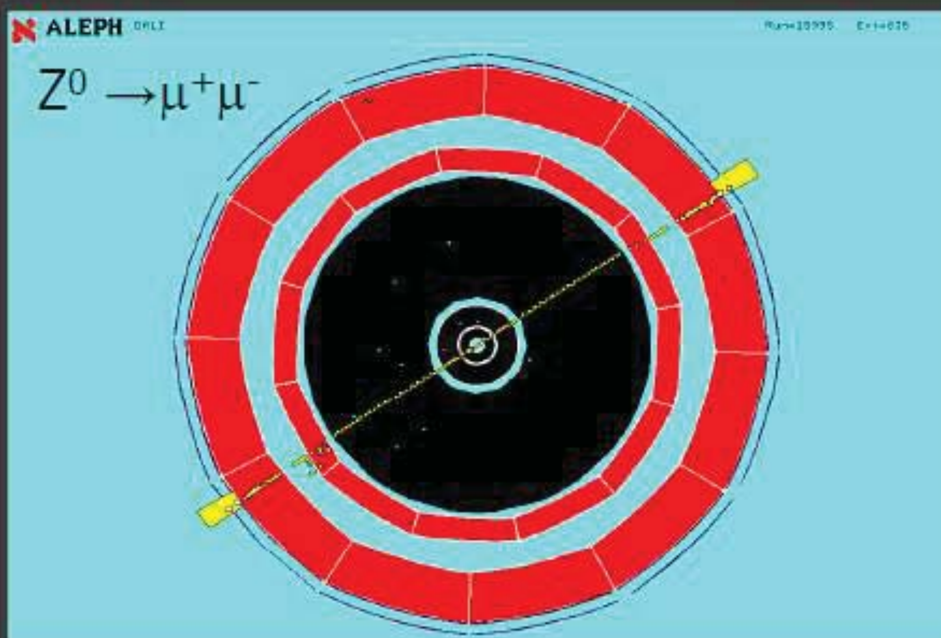
- Analysis
  - ~~stat. methods~~
  - jet reconstruction
  - b-quark tagging
  - flavour tagging
  - ~~fitting using kinematic constraints~~
  - ~~exclusive/inclusive channels~~
  - ~~neural networks~~

From raw data to summary data

- Raw data: digitized record of detector electronic signals;

  directly used for graphical presentation;

detector part	signal value



for statistical analysis: need physics quantities
p, E, q, m, ...

- processed data, summary data, Data Summary Tape (DST)

example of graphical presentation: Aleph detector, LEP, $e^+e^- \rightarrow Z^0 \rightarrow \mu^+\mu^-$

From raw data to summary data

reconstruction

- Procedure of processing raw data to summary data: reconstruction

example: to conclude about $Z^0 \to \mu^+\mu^-$ decay one needs to

establish two tracks of corresponding **p**

↓

association of signals in tracking det. into tracks; track fitting; determination of **p**

determine small energy deposited in EM calorimeter(μ)

↓

association of signals in calorim. into clusters; association of clusters to tracks

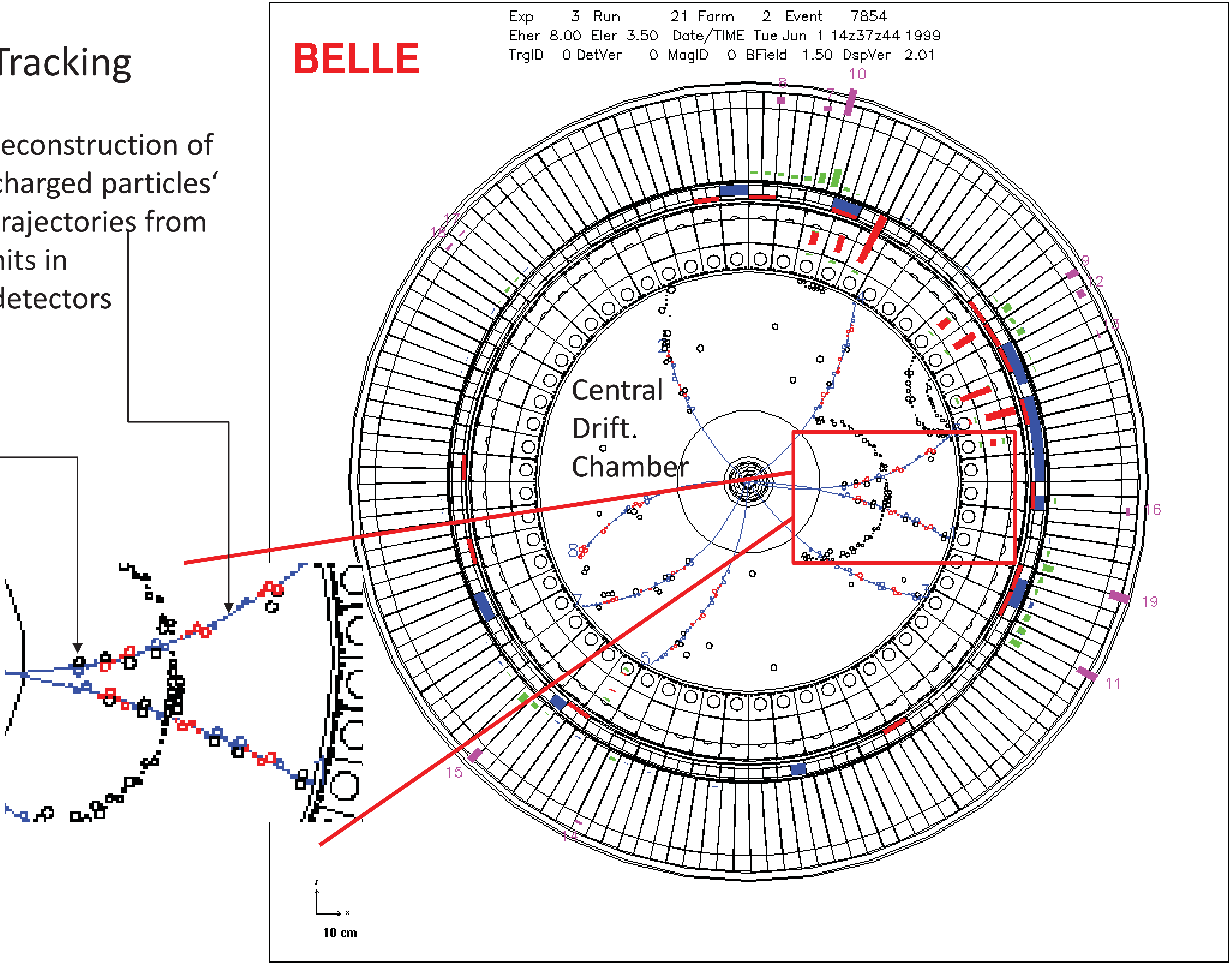
identify μ

↓

hits in μ det.; association to tracks (different procedures for hadron ident.)

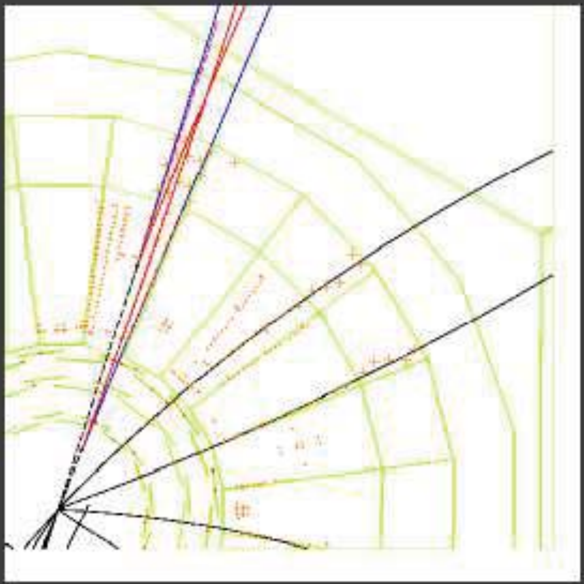
Tracking

reconstruction of charged particles‘ trajectories from hits in detectors



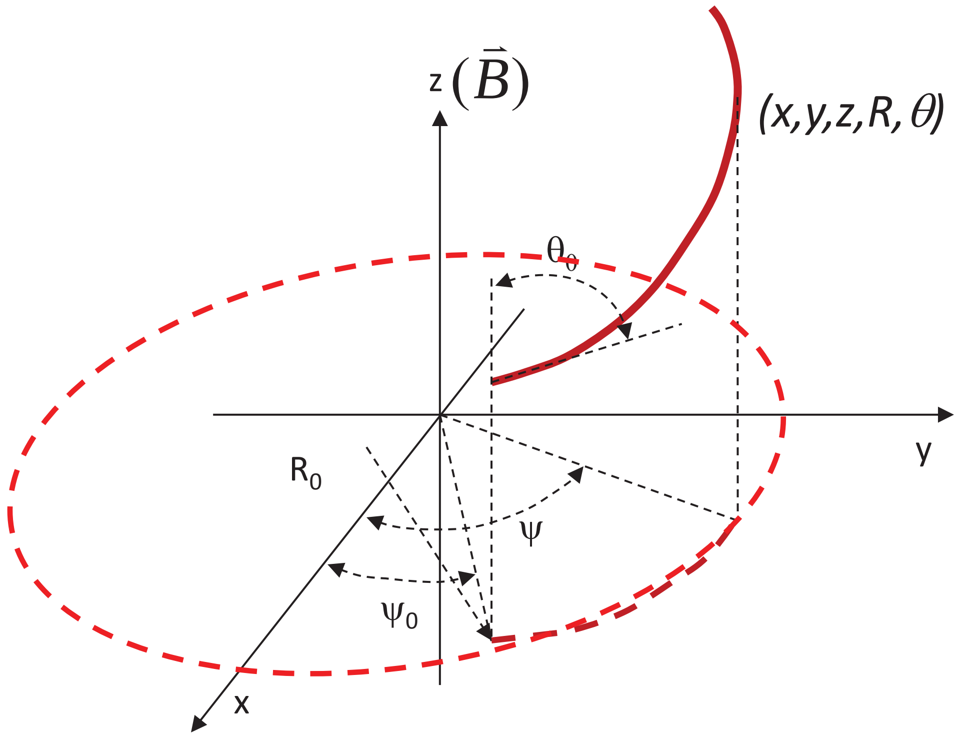
From raw data to summary data

track fitting



- charged track in **B** $\Rightarrow$ helix
- association of electronic signals in tracking detectors into groups - tracks
  pattern recognition
- fitting of helix parameters to associated hits
  track fitting

Helix parametrization



$$x = x_0 + R(\sin\psi - \sin\psi_0)$$

$$y = y_0 - R(\cos\psi - \cos\psi_0)$$

$$z = z_0 + (\psi - \psi_0)R\cot\vartheta$$

$$R = R_0$$

$$\vartheta = \vartheta_0$$

helix defined by 5 parameters:

$$y_0, z_0, \psi_0, \vartheta_0, 1/R$$

$$(x_0 = y_0 / \tan\psi_0)$$

Pattern recognition

projection of helix:

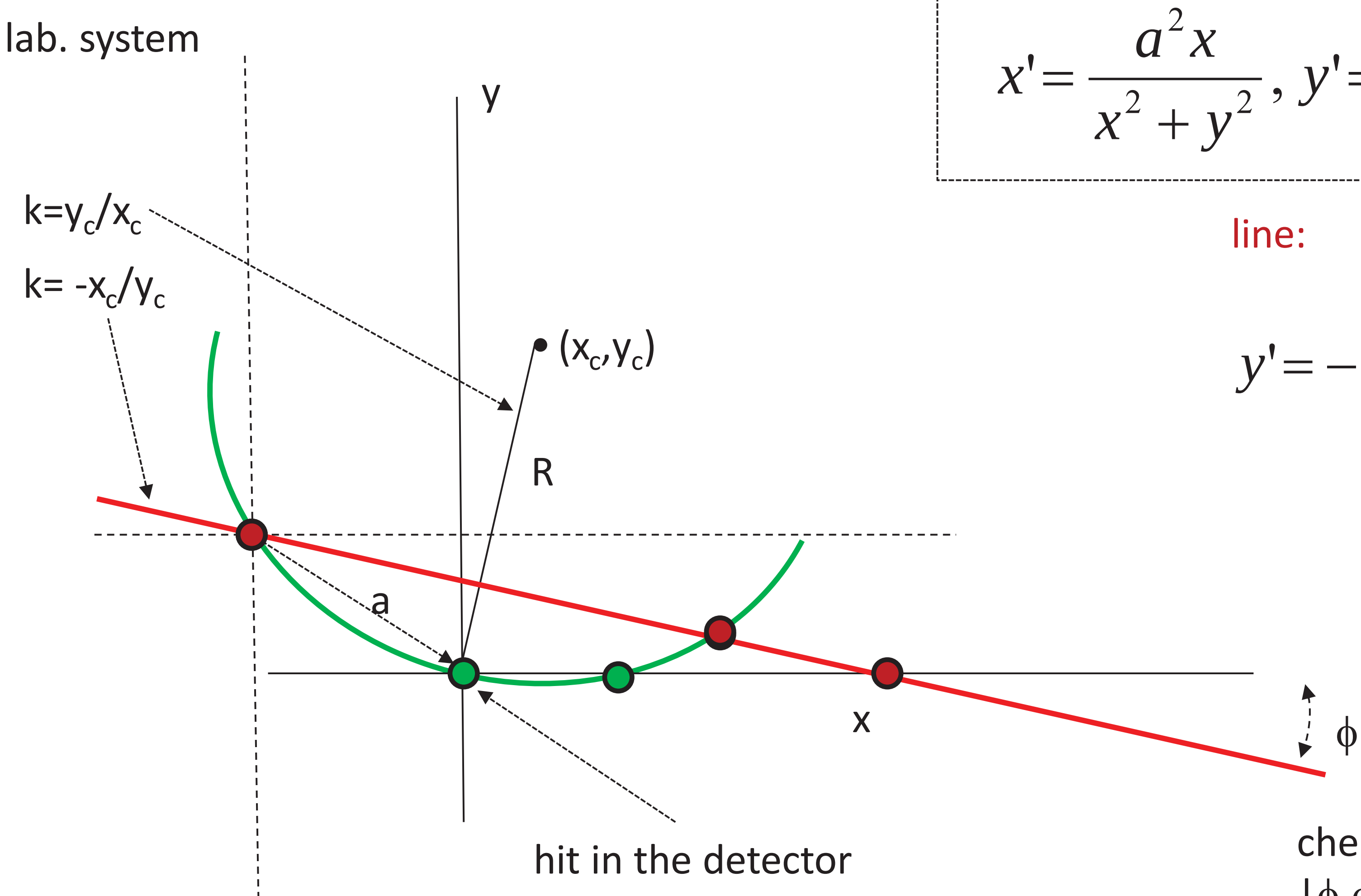
$$(x-x_c)^2+(y-y_c)^2=R^2$$

transformation:

$$x'=\frac{a^2x}{x^2+y^2},\ y'=\frac{a^2y}{x^2+y^2}$$

line:

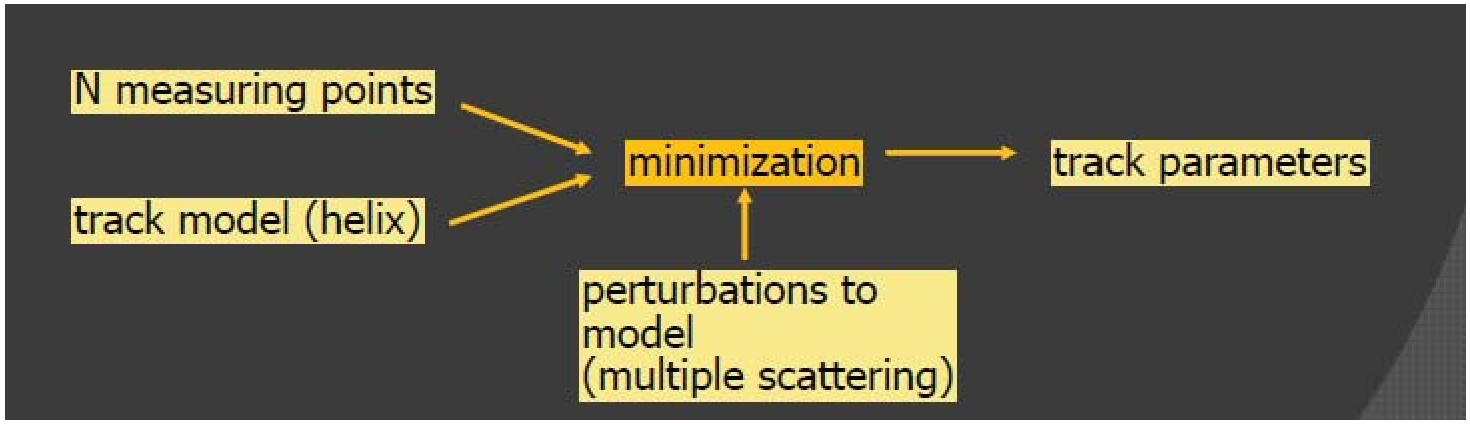
$$y'=-\frac{x_c}{y_c}x'+\frac{a^2}{2y_c}$$



check for each hit:

$|\phi-\phi_0|<\alpha$?

Track fit



Track fitting algorithms:
divided according to
track model usage, inclusion of model distortions (mult. scatt., energy losses)

Global Methods   Progressive Methods   Break Point Methods

Global Methods:
simultaneous minimization of χ^2 of all measurement points;
mult. scatt. included in the error matrix

properties:
- all meas. points used simultaneously;
- simultaneous pattern recognition not possible (as opposed to Progressive methods);
- calculation expensive (NxN matrix inversion);

From raw data to summary data
track fitting

Global method - track model:
expected coordinate values

$$\begin{pmatrix} x^n_{\rm exp} \\ y^n_{\rm exp} \\ z^n_{\rm exp} \end{pmatrix} = \begin{pmatrix} x_0 + R_0^{-1}\left[\sin\psi_n - \sin\psi_0\right] \\ y_0 - R_0^{-1}\left[\cos\psi_n - \cos\psi_0\right] \\ z_0 + R_0^{-1}\cot\theta_0\left[\psi_n - \psi_0\right] \end{pmatrix}$$

5 free parameters: $\boldsymbol{p_0}=(y_0, z_0, \psi_0, \theta_0, 1/R)$
$(x_0=y_0/\tan\psi_0)$

N measured 3-dimensional points ⇒ N 3-dimensional functions depending on 5 parameters $\boldsymbol{f}(\boldsymbol{p_0})$

global χ^2 minimization:

$$\chi^2(\vec{p_0}) = \left(\vec{f}(\vec{p_0}) - \vec{m}\right)^T \vec{C}^{-1}\left(\vec{f}(\vec{p_0}) - \vec{m}\right)$$

From raw data to summary data
track fitting

Global method - example:
straight line fit

model: $y_n = kx_n + y_0$
N meas. of y at x_n

N	$k\Delta x$	$\sigma_k \Delta x$
2	y_2-y_1	$\sqrt{2}\sigma$
3	$(y_3-y_1)/2$	$\sigma/\sqrt{2}$
4	$(3y_4+y_3-y_2-3y_1)/10$	$\sigma/\sqrt{5}$

$$\chi^2 = \sum_{n=1}^{N} \frac{(y_n - kx_n - y_0)^2}{\sigma_n^2}$$

minimization yields

$$k\sum_{n=1}^{N}\frac{x_n^2}{\sigma_n^2} + y_0\sum_{n=1}^{N}\frac{x_n}{\sigma_n^2} - \sum_{n=1}^{N}\frac{y_n x_n}{\sigma_n^2} = 0$$

$$k\sum_{n=1}^{N}\frac{x_n}{\sigma_n^2} + y_0\sum_{n=1}^{N}\frac{1}{\sigma_n^2} - \sum_{n=1}^{N}\frac{y_n}{\sigma_n^2} = 0$$

for $x_n = n\Delta x$ and $\sigma_n = \sigma \Rightarrow$

$$k = \frac{1}{\Delta x}\,\frac{N\sum n y_n - \sum n \sum y_n}{N\sum n^2 - (\sum n)^2}$$

From raw data to summary data

track fitting

Progressive method:

- vector of parameters after n measurement points → $\vec{p}_n$
- error matrix after n measurement points → W_n
- vector of extrapolated parameters → $\vec{p}_n^{\,e}$
- extrapolated error matrix → W_n^e

$$W_n^e = D^T W_n D, \quad D = \frac{\partial \bar{p}}{\partial \bar{p}^e}$$

$$W_{n+1} = W_n + U$$

vector of measured points → $\vec{p}_{n+1}^{\,m}$

χ^2: sum of contribution from extrapolation and measurement:

n-th point   extrapolation to (n+1)st point   (n+1)st point

$$\chi^2(\vec{p}_{n+1}) = \chi^2(\vec{p}_n) + \left[\vec{p}_{n+1} - \vec{p}_n^{\,e}\right]^T W_n^e \left[\vec{p}_{n+1} - \vec{p}_n^{\,e}\right] + \left[\vec{p}_{n+1} - \vec{p}_{n+1}^{\,m}\right]^T U \left[\vec{p}_{n+1} - \vec{p}_{n+1}^{\,m}\right]$$

From raw data to summary data
track fitting

Progressive method:

- vector of parameters after n measurement points → $\vec{p}_n$
- error matrix after n measurement points → W_n
- vector of extrapolated parameters → $\vec{p}_n^{\,e}$
- extrapolated error matrix → W_n^e

$$W_n^e = D^T W_n D, \quad D = \frac{\partial \bar{p}}{\partial \bar{p}^e}$$

$$W_{n+1} = W_n + U$$

vector of measured points → $\vec{p}_{n+1}^{\,m}$

χ^2: sum of contribution from extrapolation and measurement:

n-th point | extrapolation to (n+1)st point | (n+1)st point

$$\chi^2(\vec{p}_{n+1}) = \chi^2(\vec{p}_n) + \left[\vec{p}_{n+1} - \vec{p}_n^{\,e}\right]^T W_n^e \left[\vec{p}_{n+1} - \vec{p}_n^{\,e}\right] + \left[\vec{p}_{n+1} - \vec{p}_{n+1}^{\,m}\right]^T U \left[\vec{p}_{n+1} - \vec{p}_{n+1}^{\,m}\right]$$

after minimization: set of equations for $\mathbf{p}_{n+1}^F$;
if χ^2 from extrapol. larger than chosen value for specific point
$\Rightarrow$ point not assigned to track

} pattern recognition

From raw data to summary data
track fitting

Progressive method:

- vector of parameters after n measurement points → $\vec{p}_n$
- error matrix after n measurement points → W_n
- vector of extrapolated parameters → $\vec{p}_n^{\,e}$
- extrapolated error matrix → W_n^e

$$W_n^e = D^T W_n D, \quad D = \frac{\partial \bar{p}}{\partial \bar{p}^e}$$

$$W_{n+1} = W_n + U$$

vector of measured points → $\vec{p}_{n+1}^{\,m}$

χ^2: sum of contribution from extrapolation and measurement:

already known (green) · to be determined (red) · calculated (blue) · measured in detector (brown)

$$\chi^2(\vec{p}_{n+1}) = \chi^2(\vec{p}_n) + \left[\vec{p}_{n+1} - \vec{p}_n^{\,e}\right]^T W_n^e \left[\vec{p}_{n+1} - \vec{p}_n^{\,e}\right] + \left[\vec{p}_{n+1} - \vec{p}_{n+1}^{\,m}\right]^T U \left[\vec{p}_{n+1} - \vec{p}_{n+1}^{\,m}\right]$$

after minimization: set of equations for $\mathbf{p}_{n+1}^F$;
if χ^2 from extrapol. larger than chosen value for specific point
$\Rightarrow$ point not assigned to track

} pattern recognition

Track fit

progressive method, example: straight line fit

N	$k^F \Delta x$	$\sigma_k^F \Delta x$	$\sigma_k \Delta x$	
2	$y_2 - y_1$	$\sqrt{2}\sigma$	$\sqrt{2}\sigma$	
3	$(3y_3 - y_2 - 2y_1)/5$	$\sqrt{(14/25)}\sigma = 0.748\sigma$	$\sigma/\sqrt{2}$	$=0.707\ \sigma$
4	$(30y_4 - y_3 - 18y_2 - 11y_1)/70$	0.524σ	$\sigma/\sqrt{5}$	$=0.447\ \sigma$

↑ global method

global method: better precision; CPU extensive (NxN matrix inversion), simultaneous patt. recognition not possible

From raw data to summary data

track fitting

Global method – multiple scattering:

error matrix:

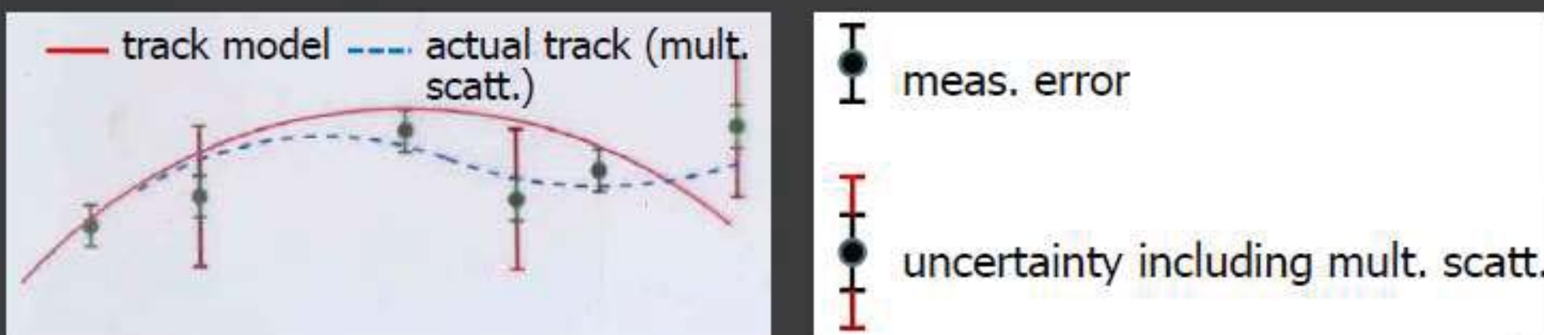
$$C_{ij} = \sigma_i \sigma_j \delta_{ij} + \overline{\epsilon_i^{MS} \epsilon_j^{MS}}$$

σ_i: uncertainty of ind. measurem.;
ε_i: contr. to uncertainty due to mult. scatt.
(Molièr formula:

$$\overline{\theta_i^{MS}} = 0$$

$$\sqrt{\overline{(\theta_i^{MS})^2}} = \frac{13{,}6\ MeV}{cp\beta} \sqrt{\frac{L}{X_0}} [1 + 0.038 \ln \frac{L}{X_0}]$$

)



distribution of $(y_{meas}-y_{fit})/\sigma_y$ ("pull") is a measure of understanding the effect of mult. scatt. rather than of understanding the meas. errors

σ_y: estimated uncertainty of individual measurement; expected distrib. of „pull": Gaussian with unity width; distrib. width > (<) 1 $\Rightarrow$ σ_y under-(over-)estimated

From raw data to summary data
track fitting

- **Progressive method – multiple scattering:**
  mult. scatt. between n^{th} and $(n+1)^{st}$ point:

$$W_n^e = \left[\left[D^T W_n D \right]^{-1} + W_{\mathrm{MS}}^{-1} \right]^{-1}$$

  included in the error matrix extrapolation;

  using a corresponding mult. scatt. matrix W_{MS} one can include specifics of material between n^{th} and $(n+1)^{st}$ point

- **Break points method:**
  appropriate for detectors with a limited number of regions with significant scattering;

  scattering angles included in χ^2 as free parameters

  $\chi^2(p_n^F) \rightarrow \chi^2(p_n^F, \theta_n)$

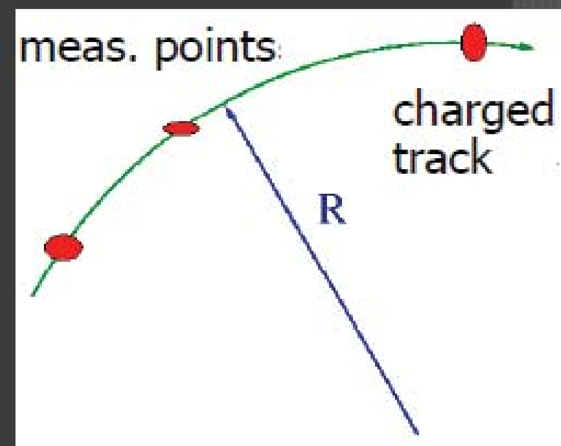
From raw data to summary data

momentum measurement

- Magnetic field:
  $p_t = qBR$;
  from curvature R one determines the transverse (w.r.t. **B**) component of **p**;
  actual meas. is curvature R;

  accuracy depends on:
  # of meas. points;
  spatial resolution of each point;
  mag. field integral BL;
  momentum p;

  multiple scattering;

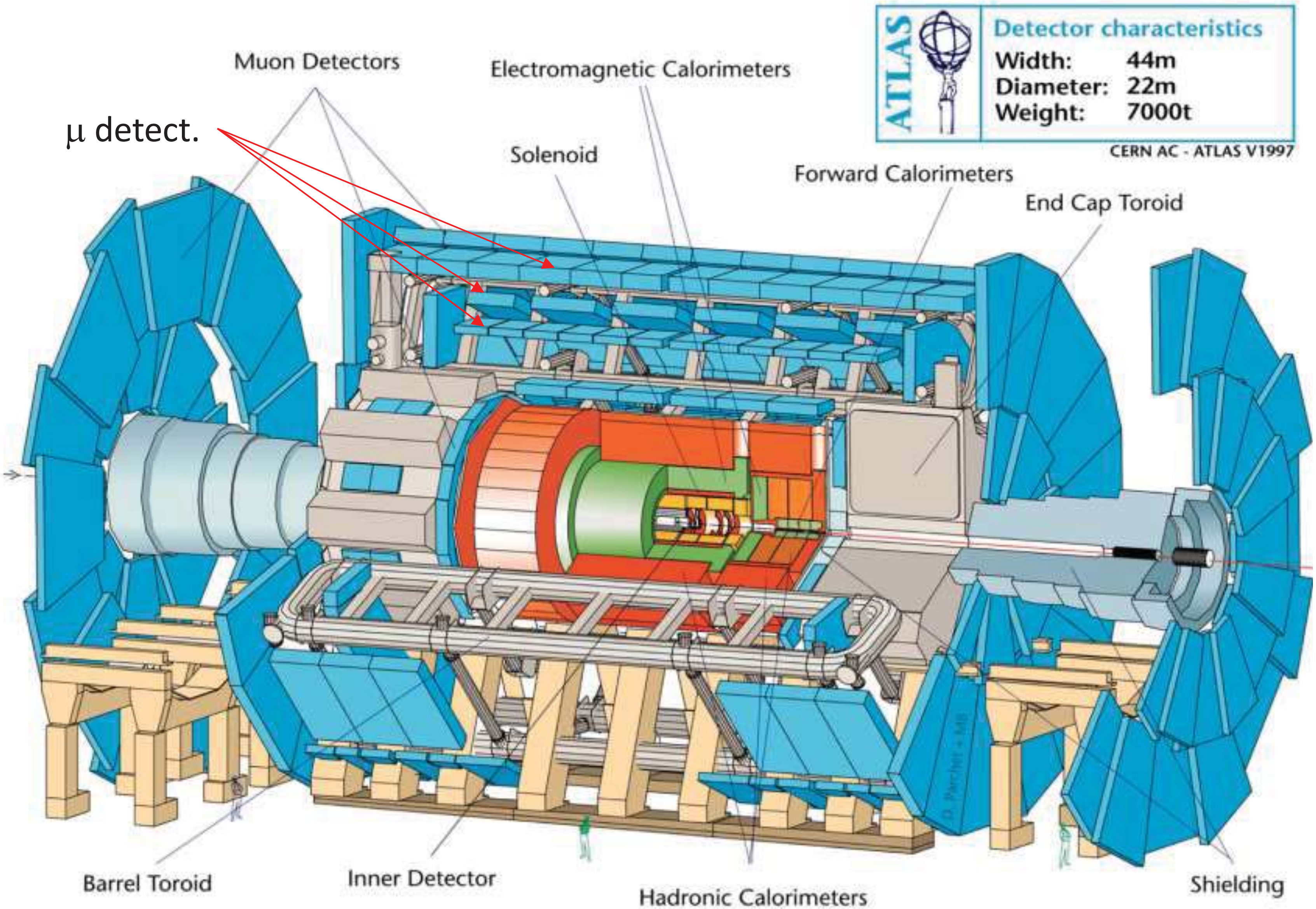


$$\frac{\sigma_{p_t}}{p_t} = \sqrt{a p_t^2 + b}$$

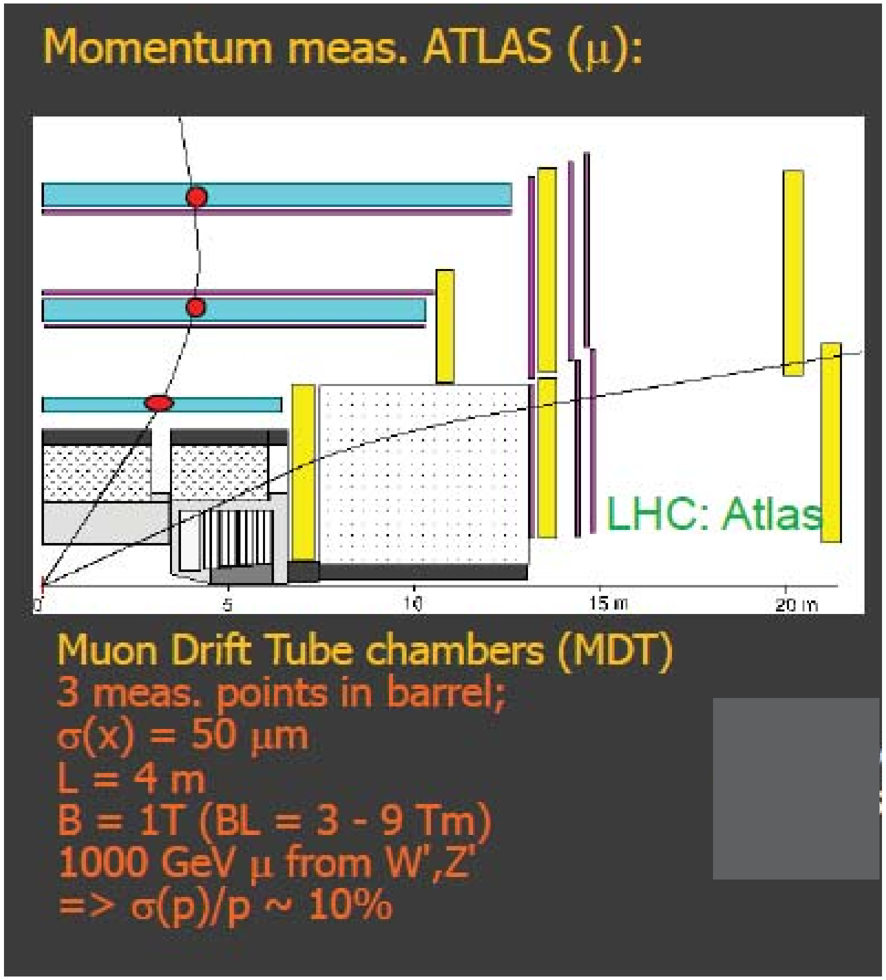
intrinsic resol. mult. scatt.

Momentum measurement

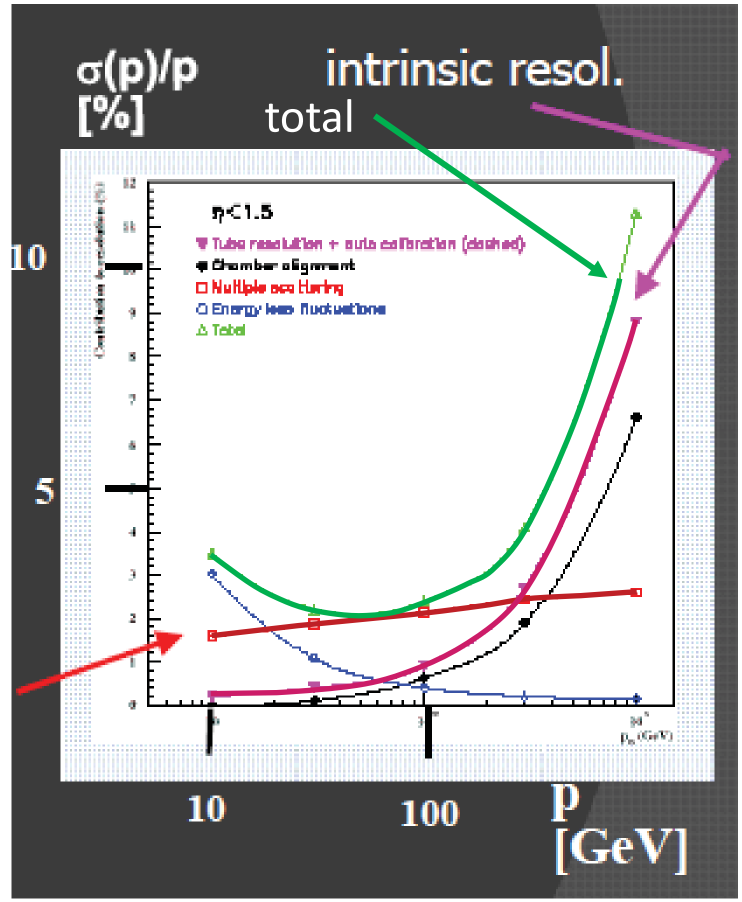
μ detect.



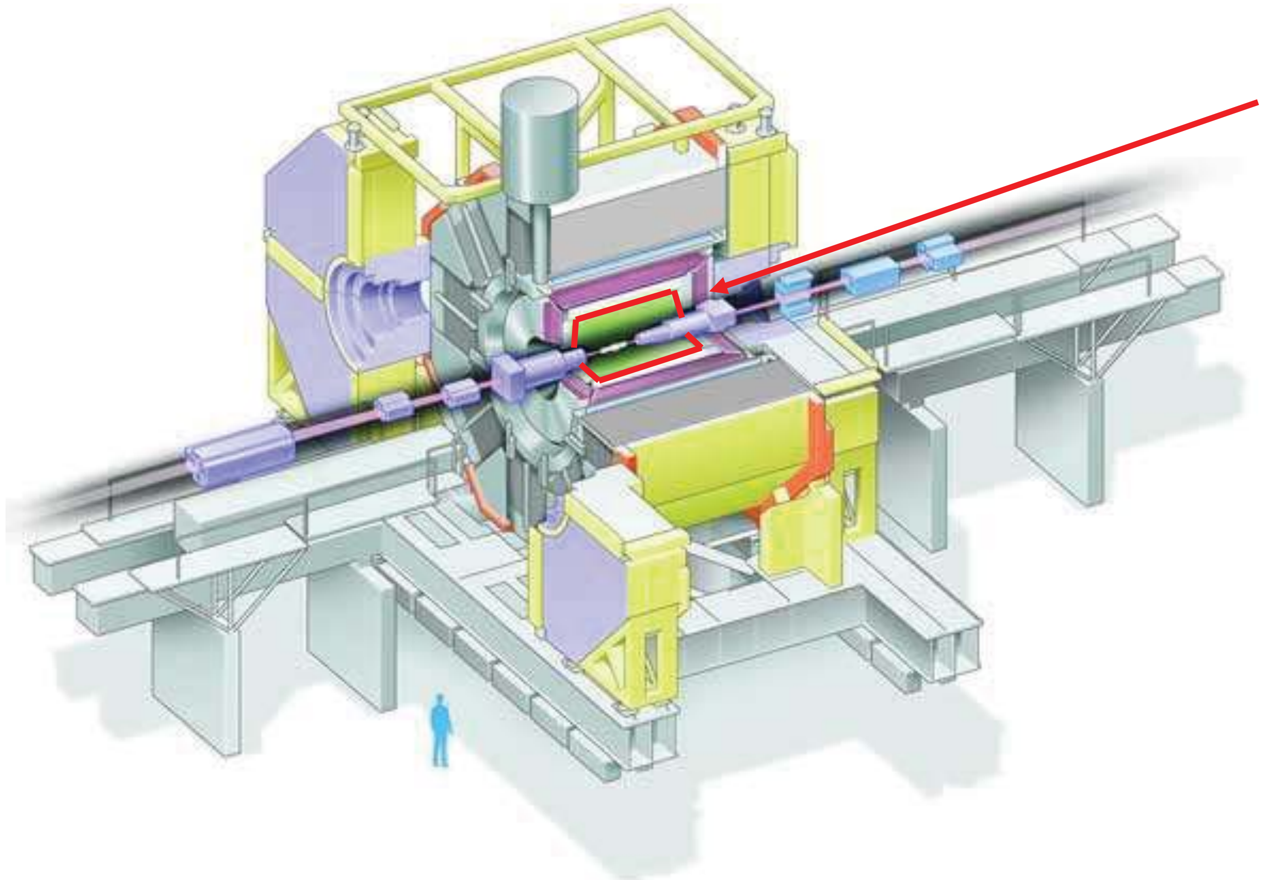
Momentum measurement



MS
(~ const.)



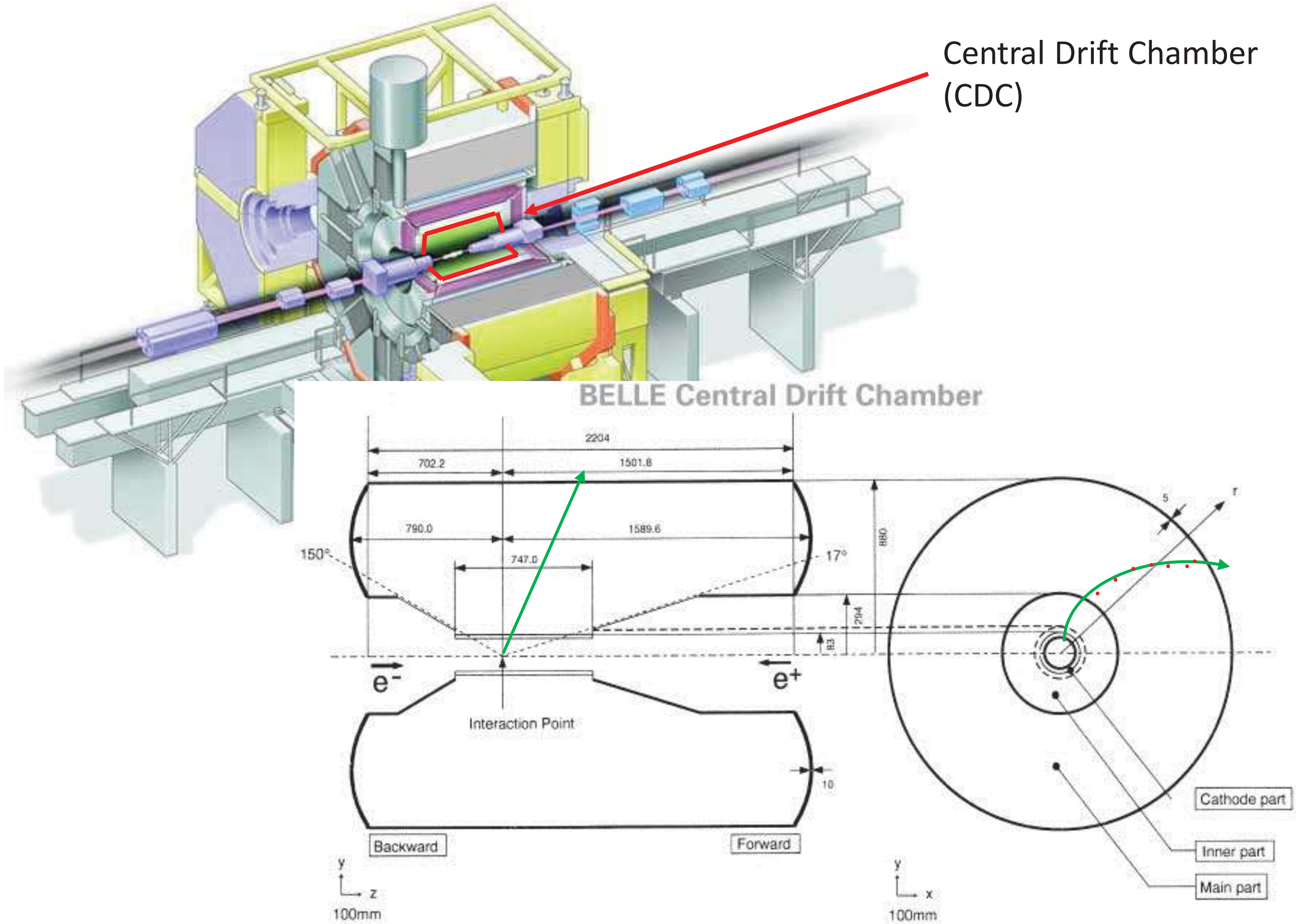
Momentum measurement



Central Drift Chamber
(CDC)

Momentum measurement

Central Drift Chamber (CDC)



p_t~1 GeV/c

B=1.5 T

L~1m

N~50

X_0~$2.9\cdot10^5$ cm

estimate:

σ_{pt}/p_t ~

$\sqrt{[(8\cdot10^{-3})^2+(0.6\cdot10^{-3})^2]}$

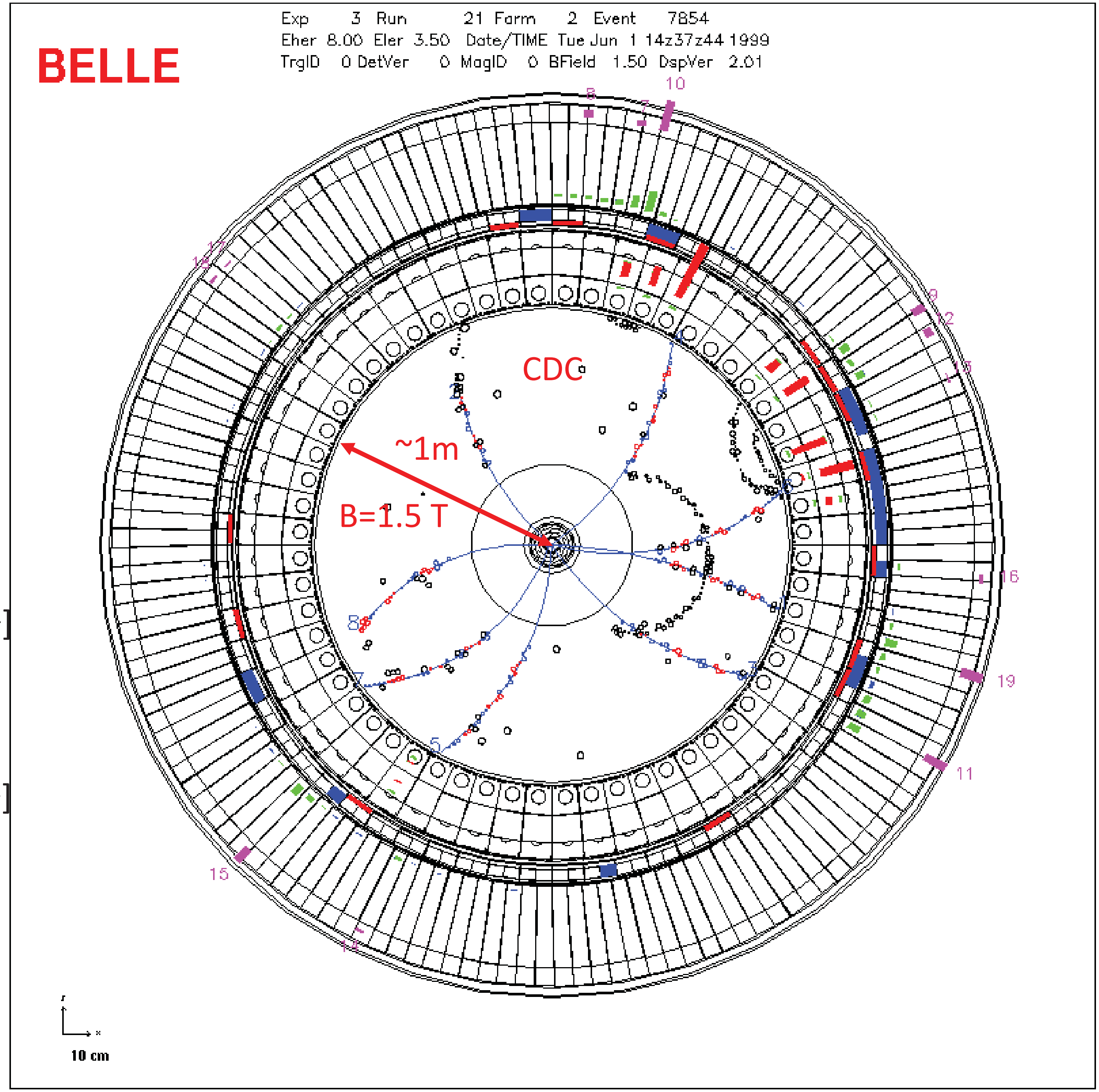
measured:

σ_{pt}/p_t ~

$\sqrt{[(3\cdot10^{-3})^2 p_t+(3\cdot10^{-3})^2]}$

how to measure?

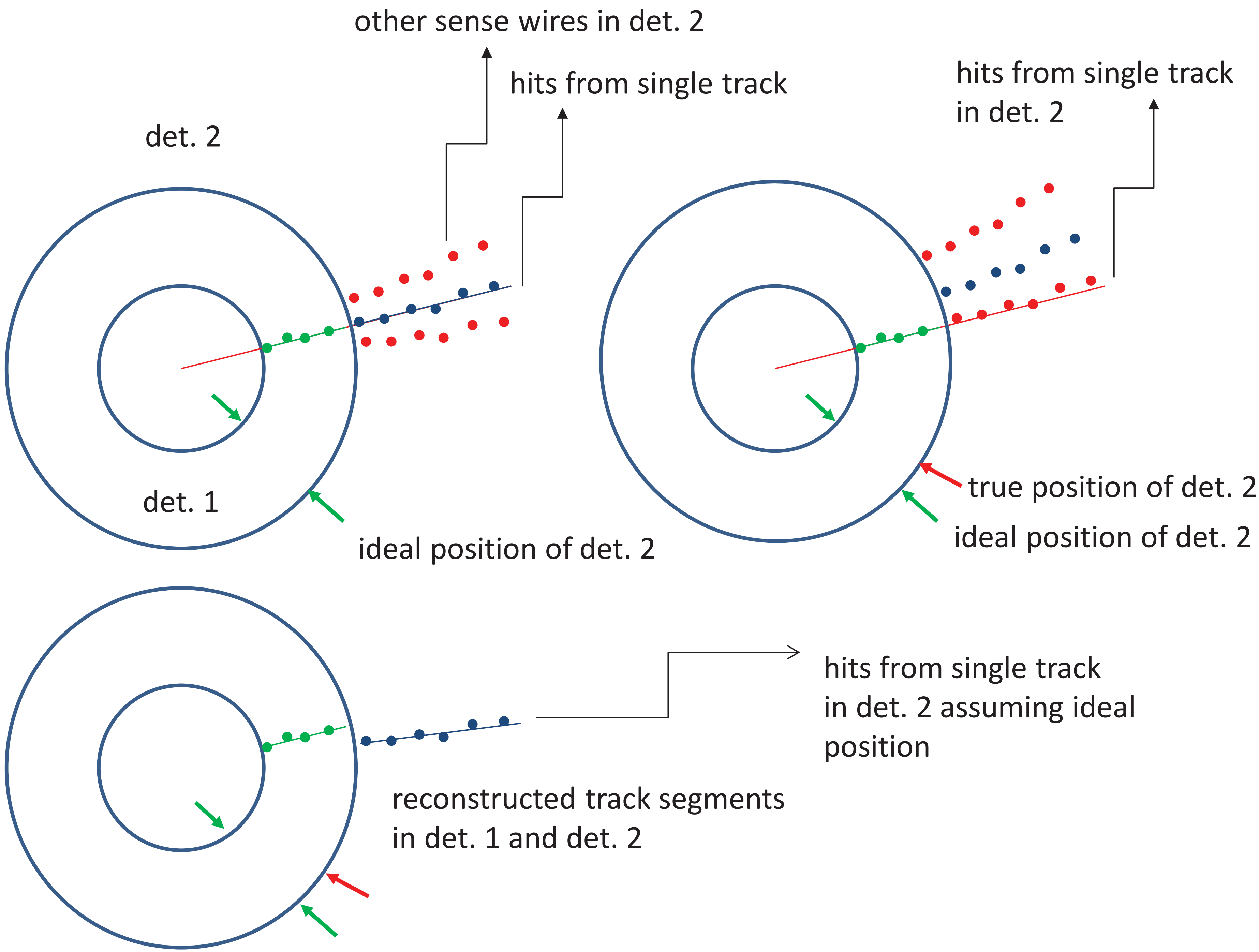
→ calibration!



Calibration
Tracking detectors

- Tracking detectors calibration
  individual subdetectors must be properly inter-orineted, otherwise tracks distorted;

  for any calibration need
    sample (tracks, decays, ...) with precisely known detector response


other sense wires in det. 2
hits from single track
det. 2
det. 1
ideal position of det. 2
hits from single track
in det. 2
true position of det. 2
ideal position of det. 2
hits from single track
in det. 2 assuming ideal
position
reconstructed track segments
in det. 1 and det. 2

Calibration
Tracking detectors

Description of detector (mis)alignment

position of individual subdetector w.r.t. reference (most precisely mechanically positioned detector)

described by set of small parameters α (translation, rotation, t-delay,...)

assume linear relation

$$\vec{q}^{\,meas} - \vec{q}^{\,ext} = S\vec{\alpha}$$

q^{meas}: vector of measured coordinates
q^{ext}: vector of extrapolated coord. (from the reference detector)
S: matrix depending on measuring coord., track model, detector geometry

simplest case:
α composed of 3 translations and 3 rotations
$\alpha = (\eta_x, \eta_y, \eta_z, \varepsilon_x, \varepsilon_y, \varepsilon_z)$

Calibration

tracking detectors

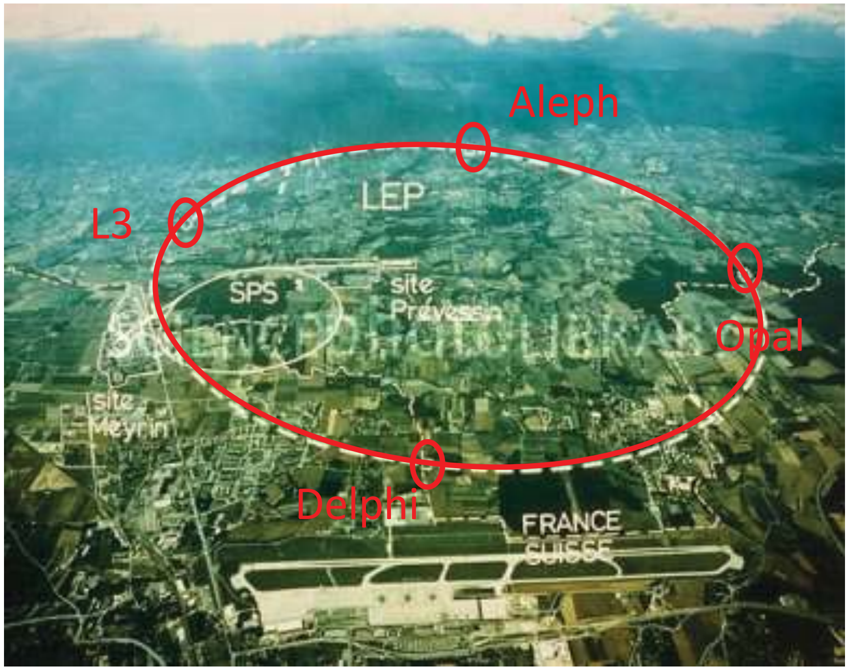
$$\chi^2 = \sum_k [\vec{q}_k^{\,meas} - \vec{q}_k^{\,ext} - S_k \vec{\alpha}]^T W_k^{-1} [\vec{q}_k^{\,meas} - \vec{q}_k^{\,ext} - S_k \vec{\alpha}]$$

$$\frac{\partial \chi^2}{\partial \vec{\alpha}} = 0 \Rightarrow (\sum_k S_k^T W_k^{-1} S_k) \vec{\alpha} = \sum_k S_k^T W_k^{-1} (\vec{q}_k^{\,meas} - \vec{q}_k^{\,ext})$$

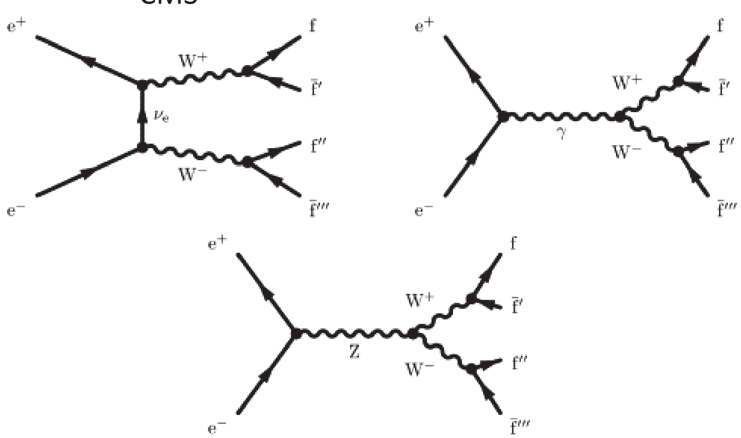
$$\Rightarrow \vec{\alpha}$$

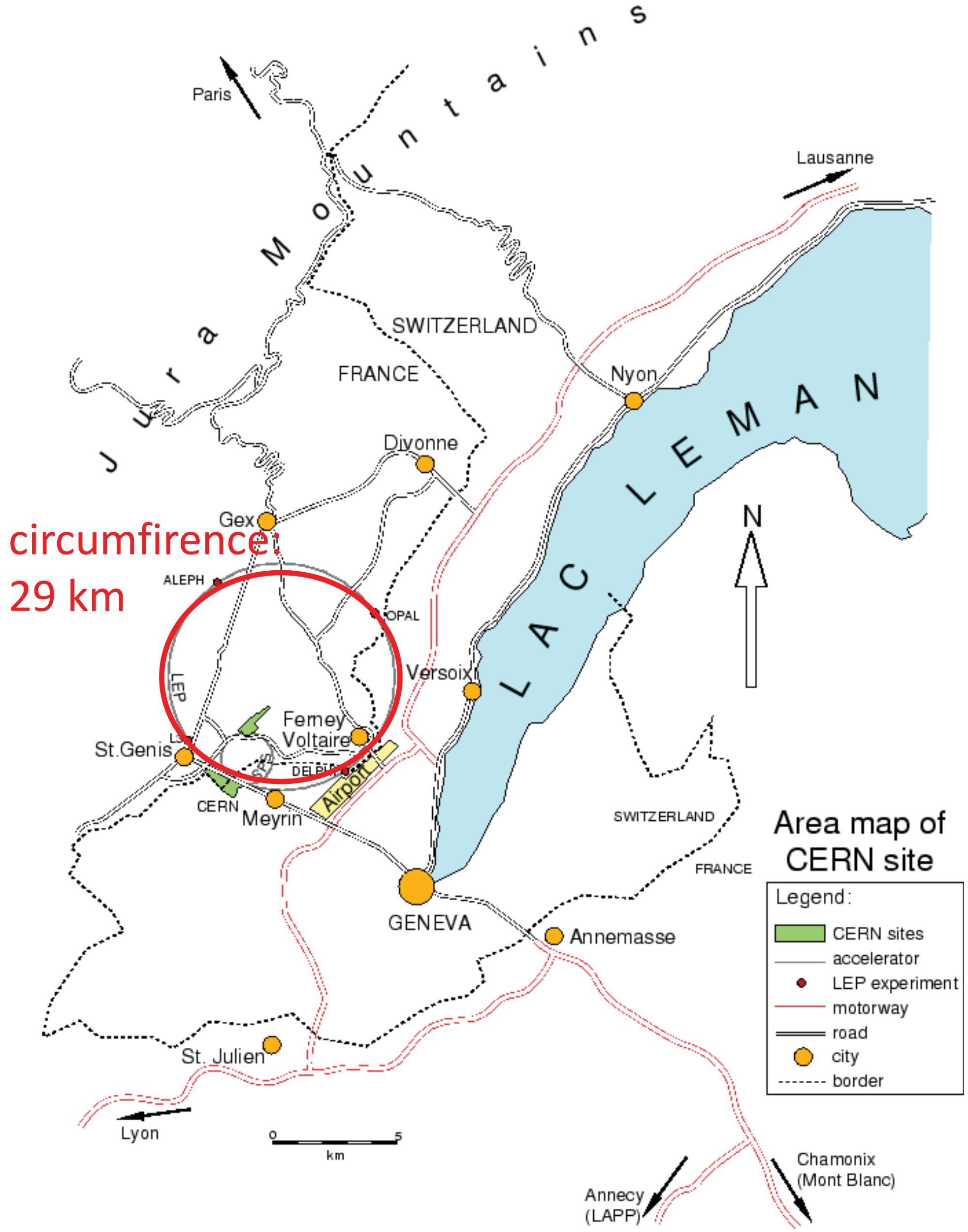
Calibration

nowadays the tunnel is occupied by the Large Hadron Collider (LHC)



Large Electron Positron (LEP) collider:
$e^+ e^-$, E_{CMS}=90-170 GeV

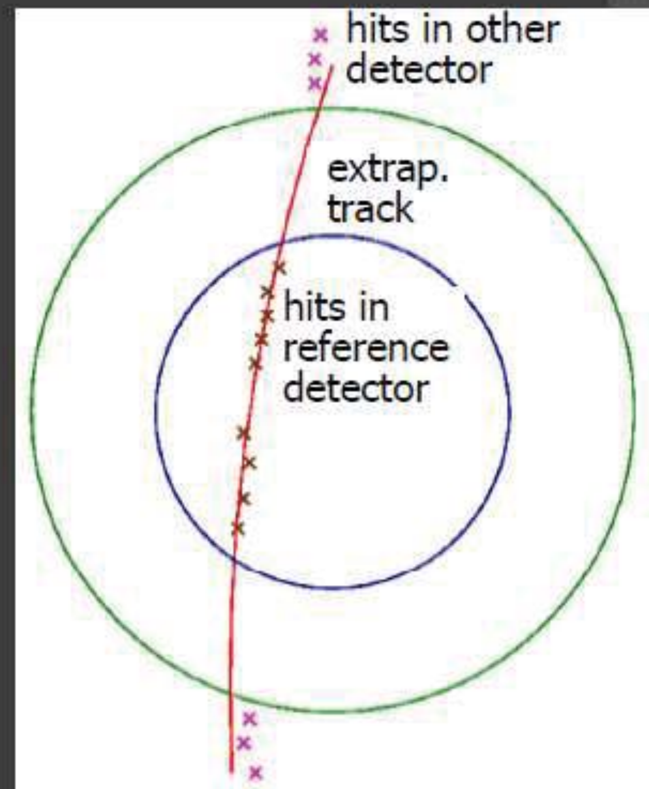




Calibration
Tracking detectors

Appropriate sample
often cosmic rays;
other decays observed,
e.g. $Z^0 \rightarrow \mu^+\mu^-$ (LEP);

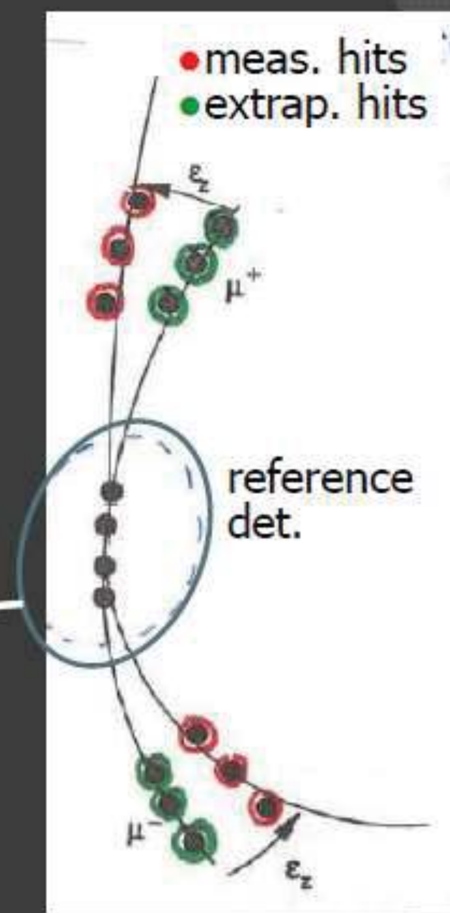
(needed also to check
the alignment method)



Calibration
Tracking detectors

Appropriate sample
e.g. $Z^0 \rightarrow \mu^+\mu^-$ (LEP);

extrapolations of
meas. tracks
do not intersect
in interaction point

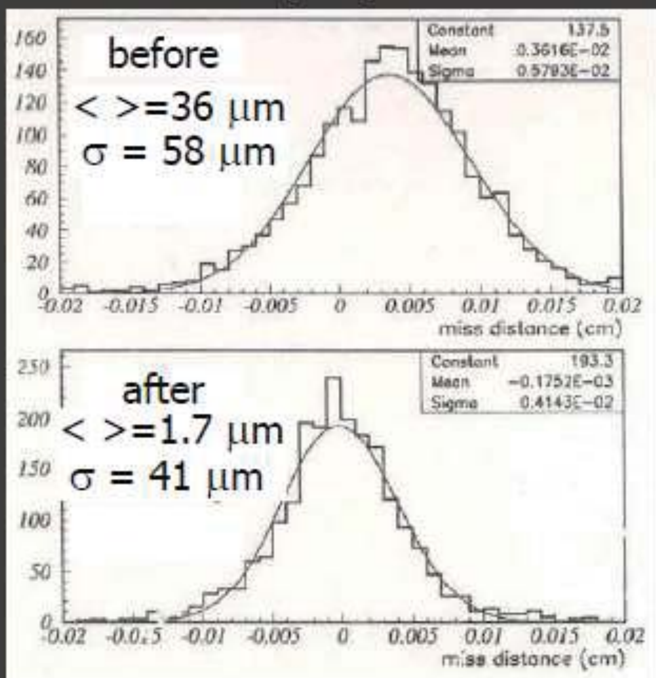


Calibration
Tracking detectors

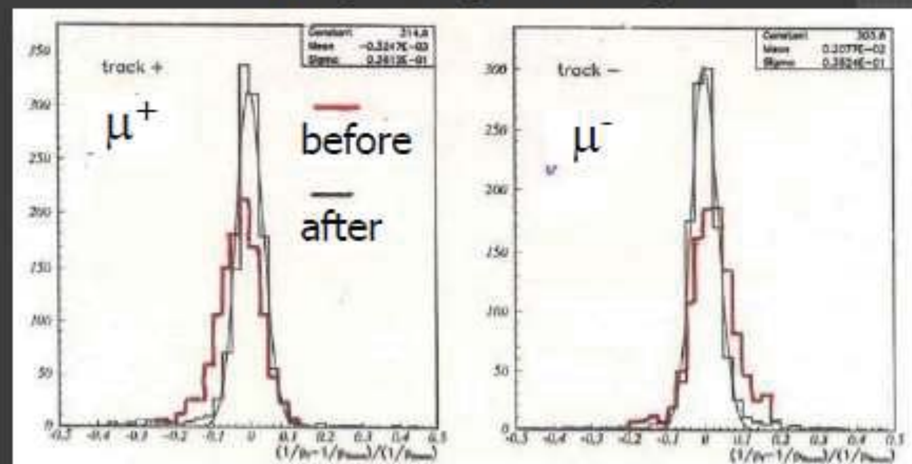
Example

Delphi detector at LEP

δ [cm]



$[(1/p_t)-(1/p_t^{ext})]/(1/p_t^{ext})$



Analysis of data
Summary

Path from electronic signal detection to result for measured physical quantities involves a number of steps

Each of those represents a specific problem and requires specific methods and solutions (some of those illustrated here)

Quality (correctness and accuracy) of the final results depends crucially on the quality of reconstruction of raw data

Analysis of data, part 2: particle identification

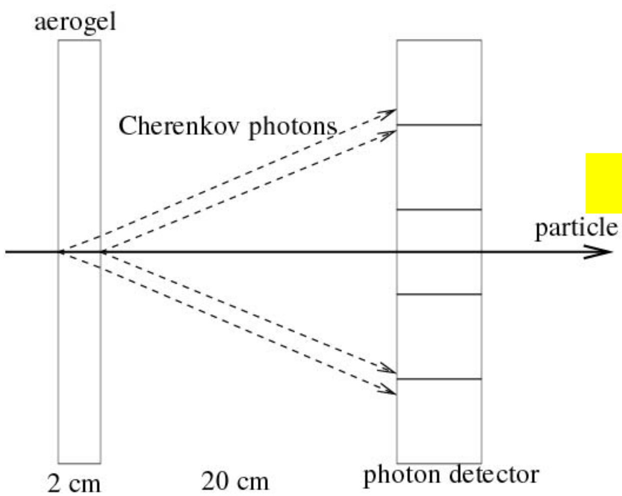
Identification

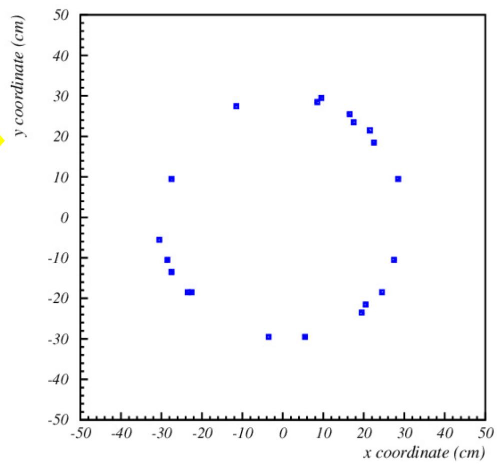
Measuring the Cherenkov angle

Particles above threshold: measure θ

Idea: transform the direction into a coordinate → ring on the detection plane

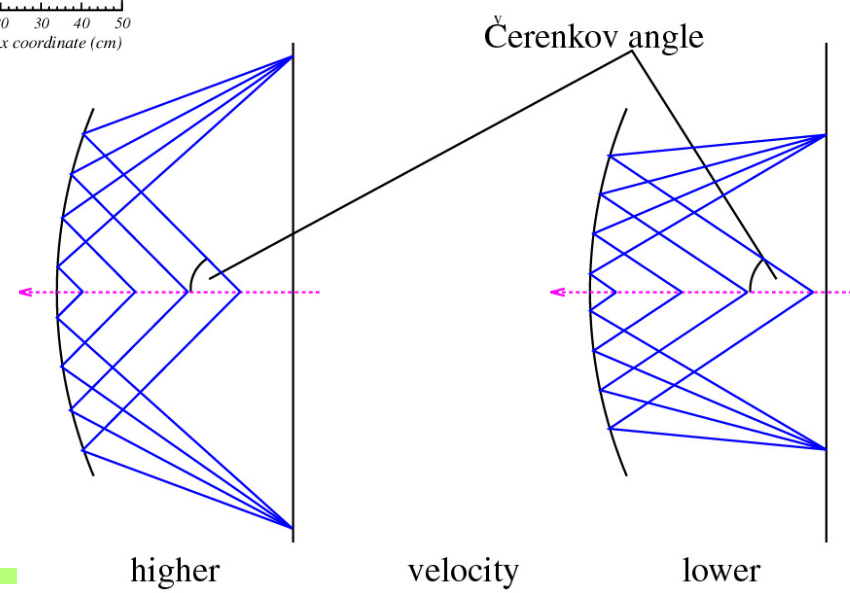
→ Ring Imaging Cherenkov (RICH) counter



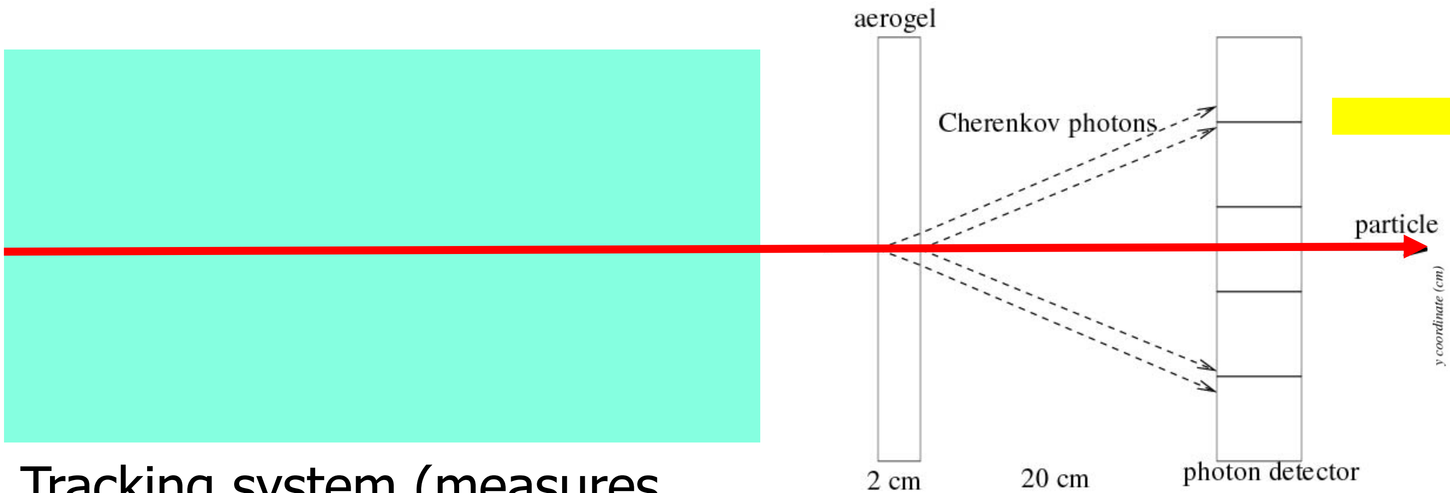


Proximity focusing RICH

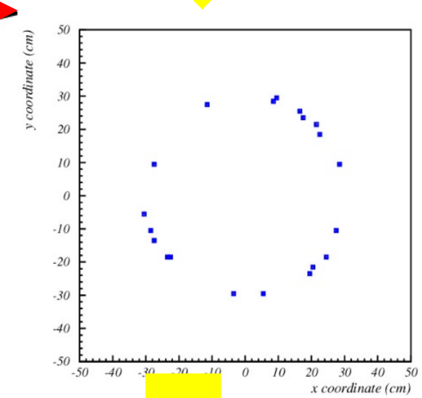
RICH with a focusing mirror

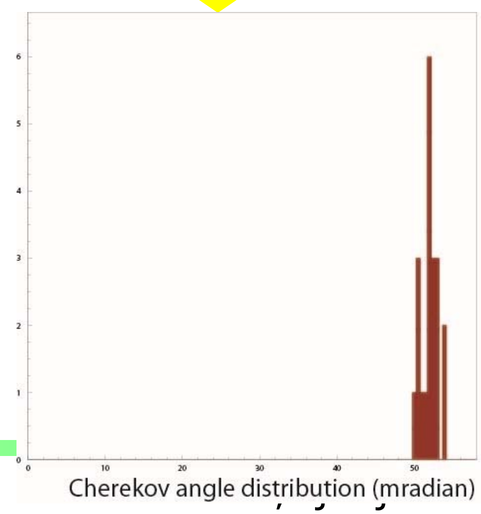


Measuring the Cherenkov angle



Tracking system (measures direction of the particle)

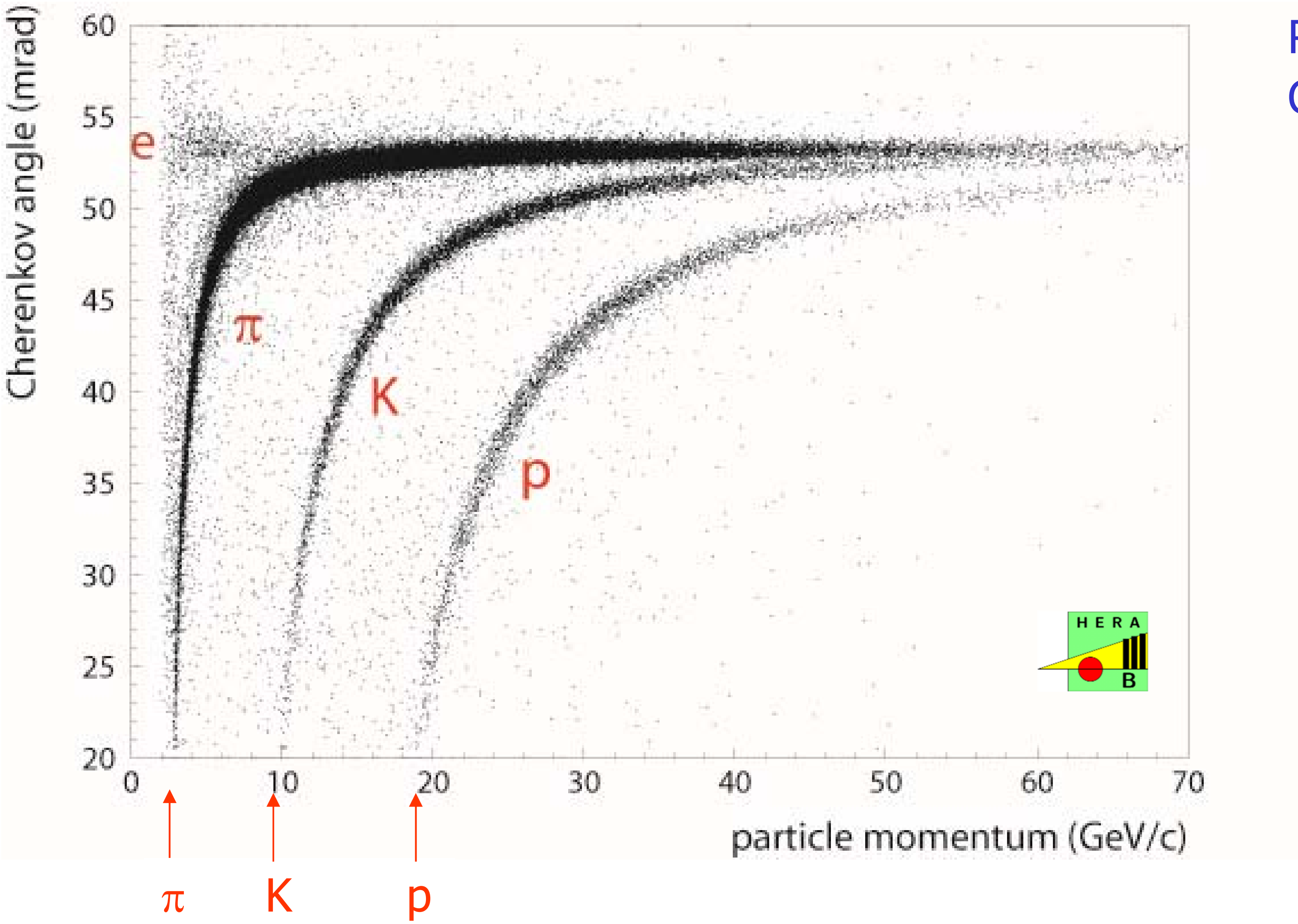




Tracking system tells us where the particle hit the radiator, and at which angle.

Use this information to calculate the Cherenkov angle for each individual detected photon

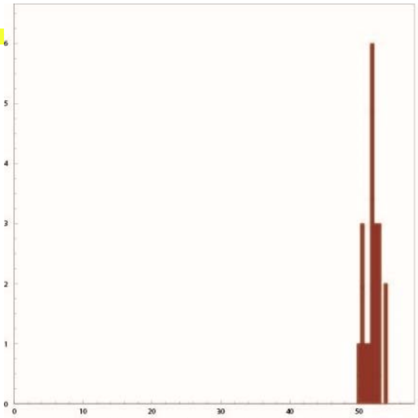
Measuring Cherenkov angle



thresholds

Radiator:
C_4F_{10} gas

Likelihood for a given PID hypothesis



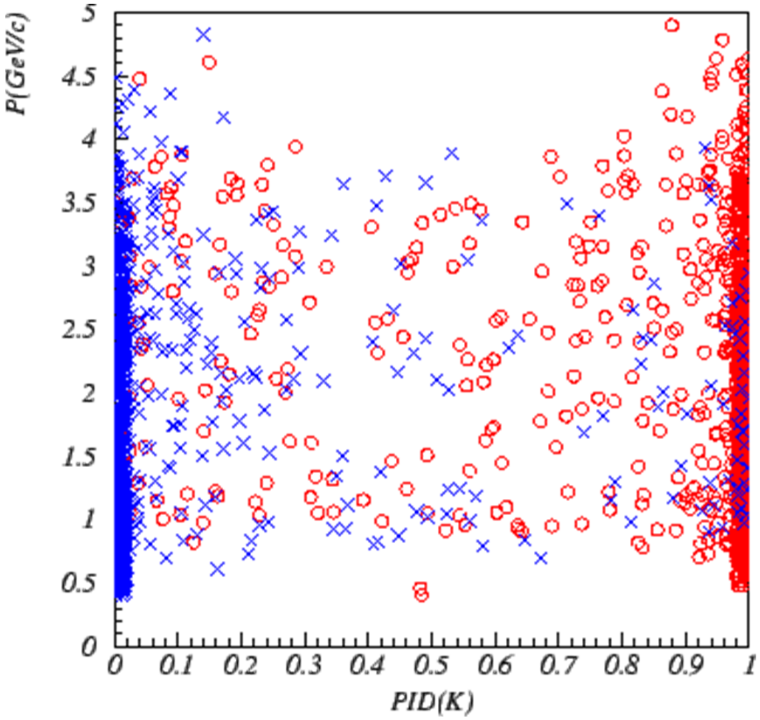
Simplest version:

- Measure the Cherenkov angle for a given particle, Θ_e = average of Cherenkov angles for all photons on the ring
- Calculate the expected values of Cherenkov angles Θ_h for all possible hypotheses h and the corresponding uncertainties σ_h
  (taking into account the momentum as determined in the tracking system)
- Likelihood for a given hypothesis

  $L_h =$ $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

  with $x = \Theta_e$ and $\mu = \Theta_h$
- For a specific case, e.g., pion-kaon separation, form ratio of log-likelihoods,

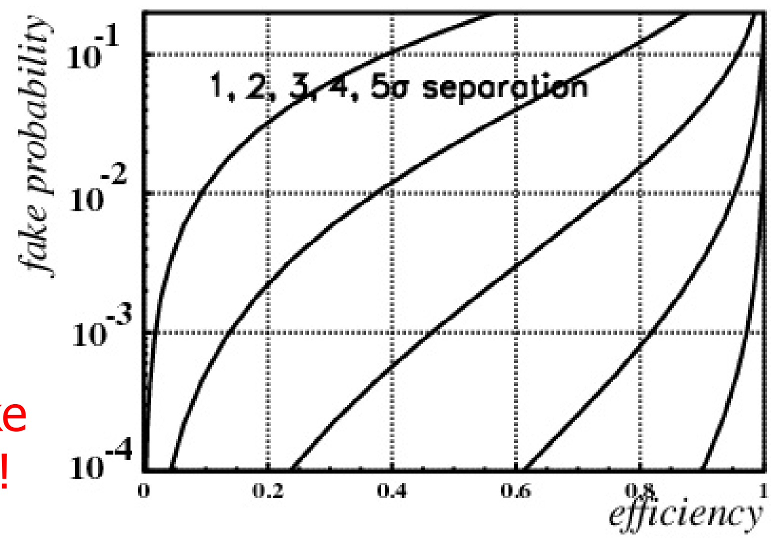
$$R_K = \ln L_K / (\ln L_\pi + \ln L_K)$$



$R_K = \ln L_K / (\ln L_\pi + \ln L_K)$

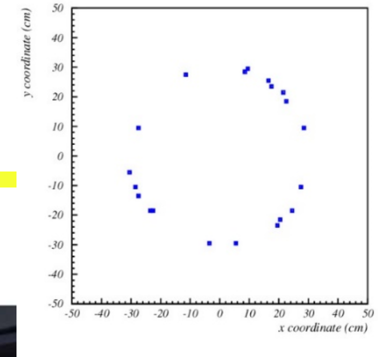
for kaons (red) and pions (blue)

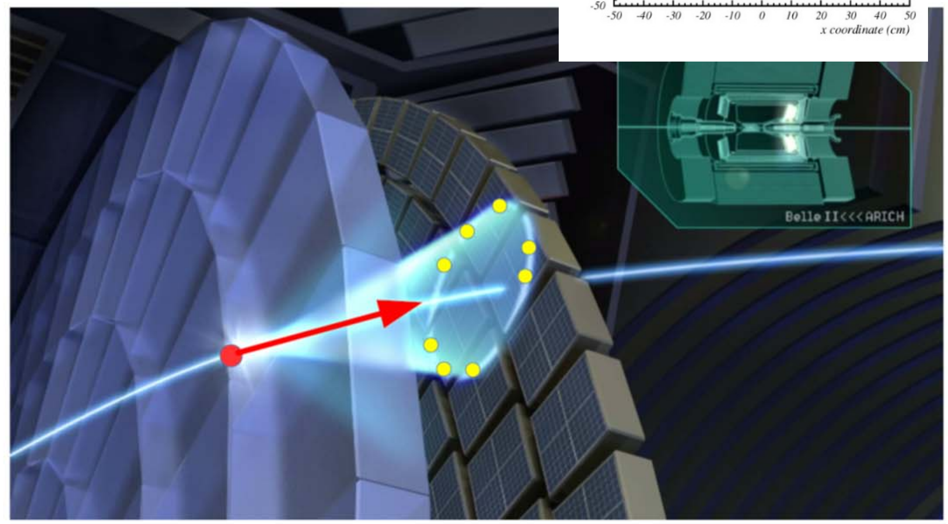
A reminder: efficiency and fake probability are tightly coupled!



Next level: detailed analysis of the image

Improve separation between particle species: add more details to the likelihood function → take each individual pixel on the photon detector and evaluate the probability that there is a hit (from the Cherenkov photons of the particle and from background sources)





Likelihood function

$$\mathcal{L} = \prod_{i}^{pixels} p_i$$

$$p_i = e^{-n_i} n_i^{m_i} / m_i!$$

For each particle hypothesis h

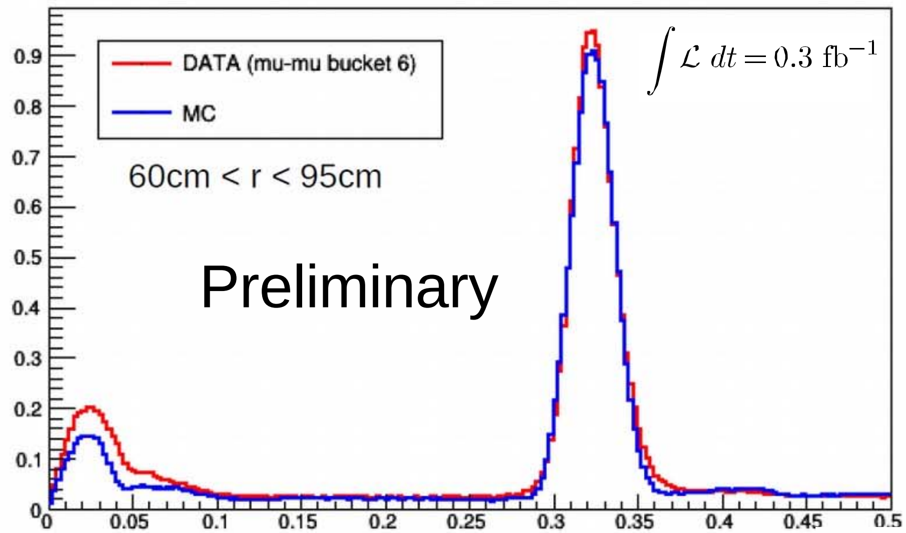
$$\ln \mathcal{L}^h = -N^h + \sum_{\text{hit } i} \left[n_i^h + \ln(1 - e^{-n_i^h}) \right]$$

N^h: Expected total number of hits

n_i^h: Expected number of hits on pixel i

Crucial: understading of the details in the image – try to model as precisely as possible

Cherenkov angle distribution in $e^+e^- \to \mu^+\mu^-$



DATA

$N_{sig} = 11.38/\mathrm{track}$

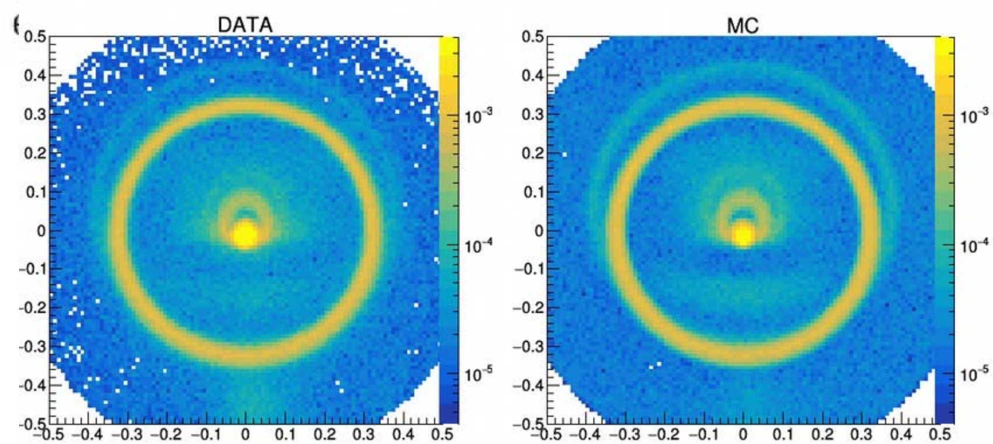
$\sigma_c = 12.7$ mrad

MC

$N_{sig} = 11.27/\mathrm{track}$

$\sigma_c = 12.75$ mrad

Overall a very good DATA/MC agreement !



Chrenkov ring (accumulated)

Estimation of π/K separation capabilities using $D^{*\pm}$ decays

- Identify K, π based on track charge in association with the charge of π_{slow}

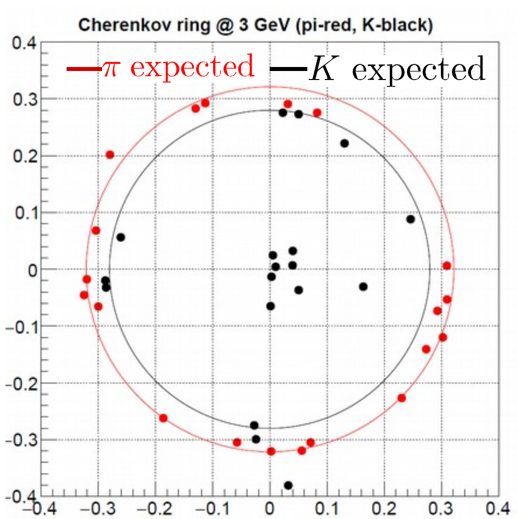
$D^{*+} \to D^0\ \pi^+_{\text{slow}}$, $D^0 \to K^-\ \pi^+$

$D^{*-} \to \bar{D}^0\ \pi^-_{\text{slow}}$, $\bar{D}^0 \to K^+\ \pi^-$

- Apply selection criteria on

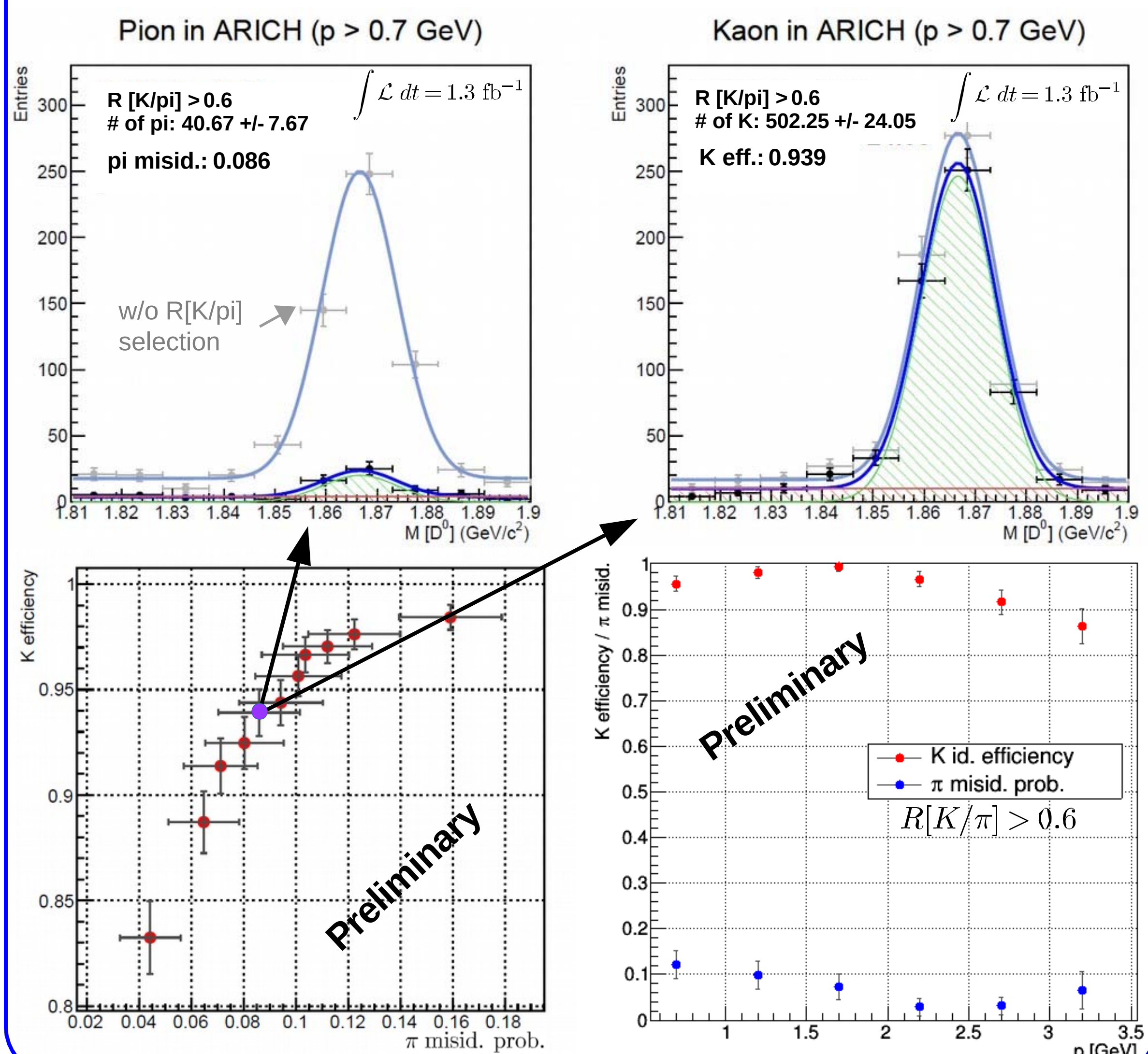
$$R[K/\pi] = \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi}$$

$\mathcal{L}$ - likelihood for given id. hypothesis



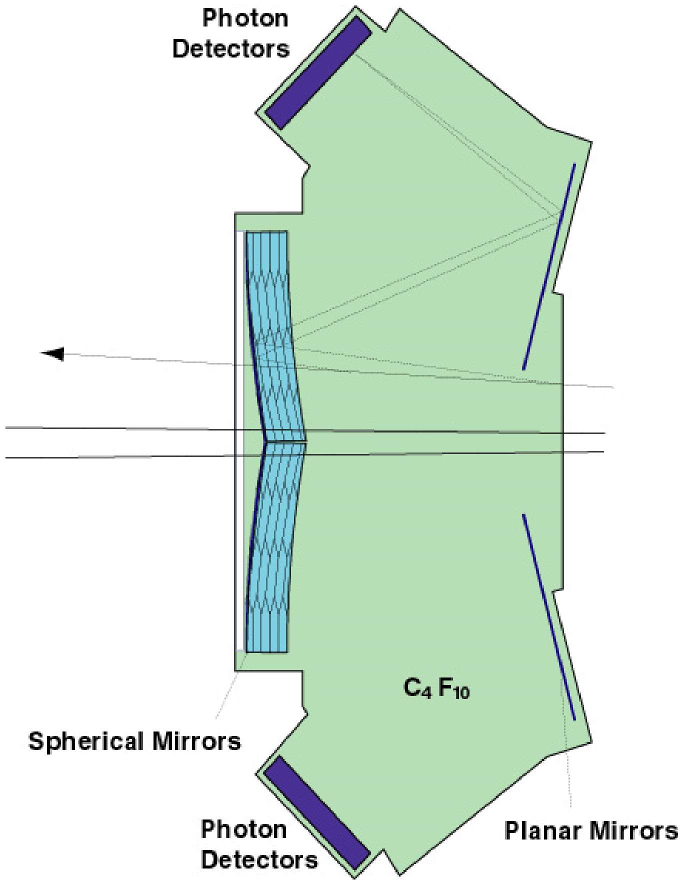
- Only coarse/preliminary calibrations included
  → further improvements expected

D^0 mass peak

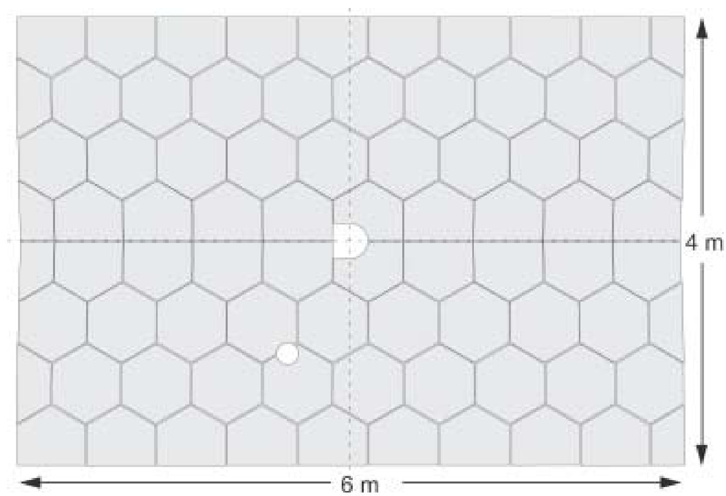


Alignment

Mirror alignment



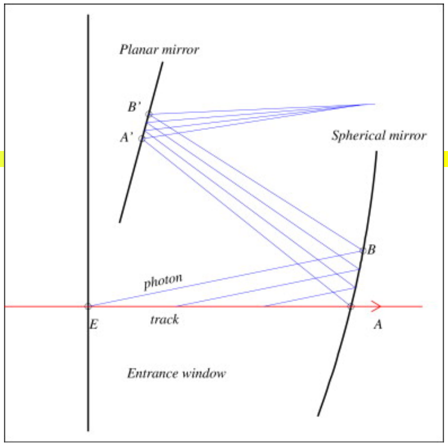
Gas radiator RICHes: large mirrors → tens of mirror segments with individual mounting → need relative alignment



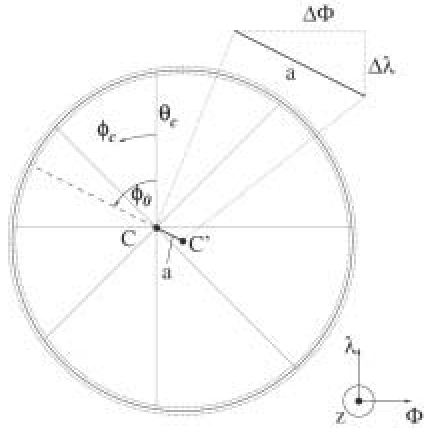
- Spherical mirror: 80 hexagonal segments
- Planar mirrors: 2x 18 rectangular segments

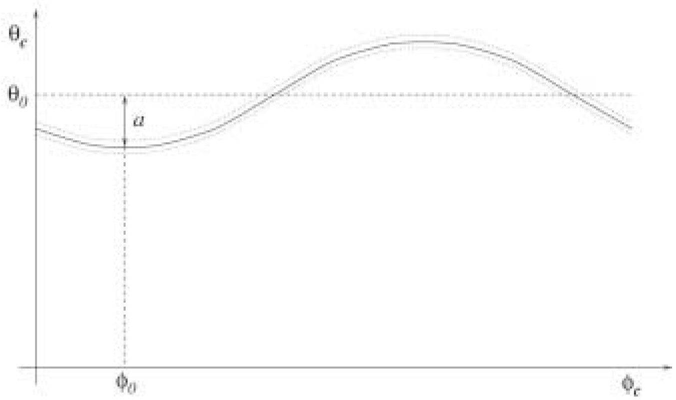
Aligning pairs of spherical and planar segments by using Cherenkov photons.

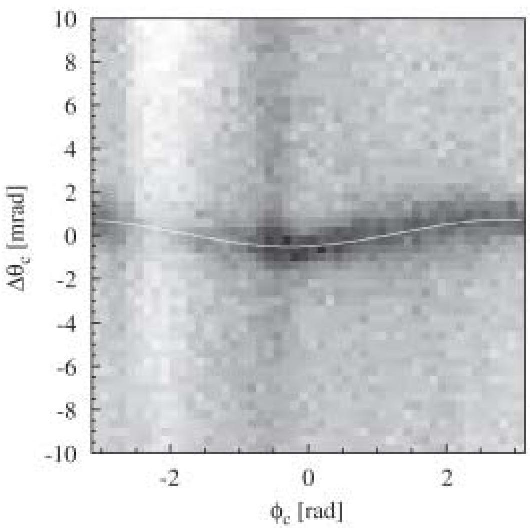
Mirror alignment



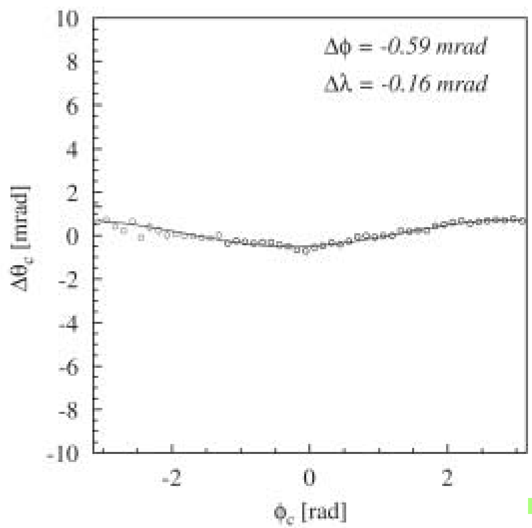
Misalignment: ring center (C ') not where expected (C) → measured Cherenkov angle depends on the azimuthal angle around the track

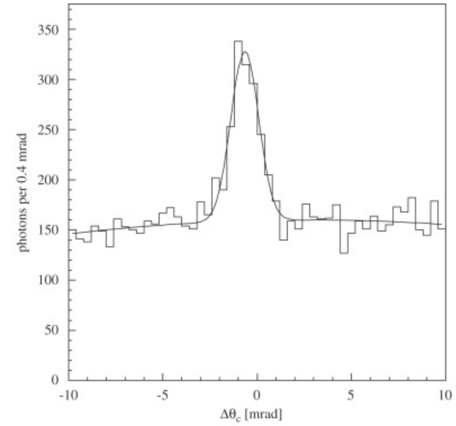






mirrors 34 14





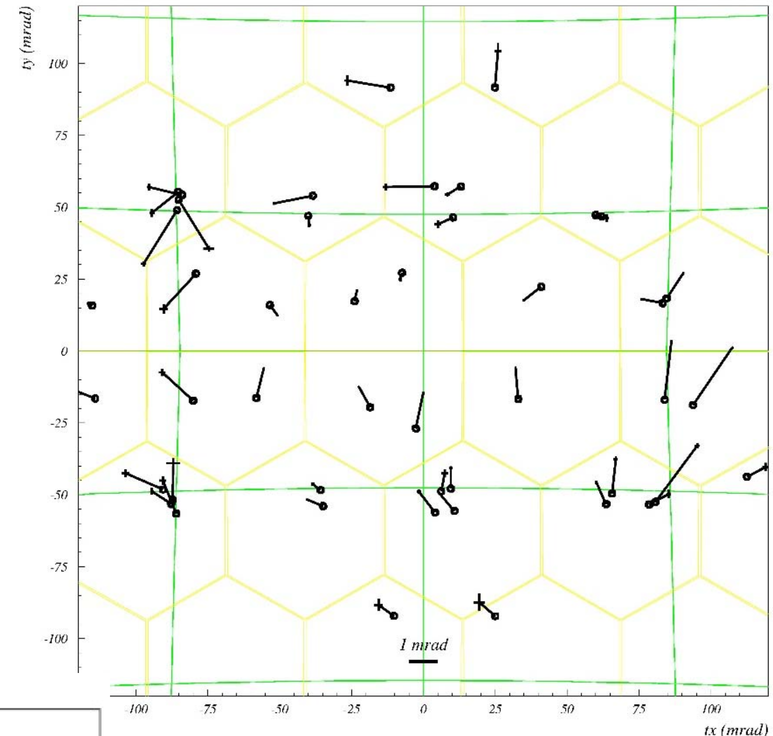
Slice in ϕ_c

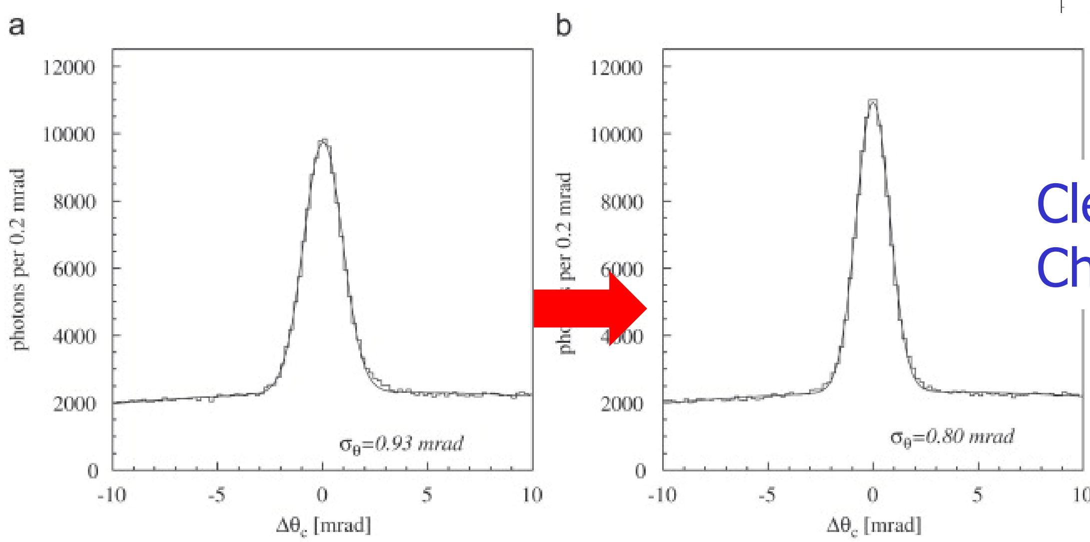
Mirror alignment

Initial mirror system alignment: with optical methods, theodolite.

Alignment with data: tells us the ultimate truth…

Combine all alignment data for all (possible) pairs of mirror segments → solve a huge system of linear equations



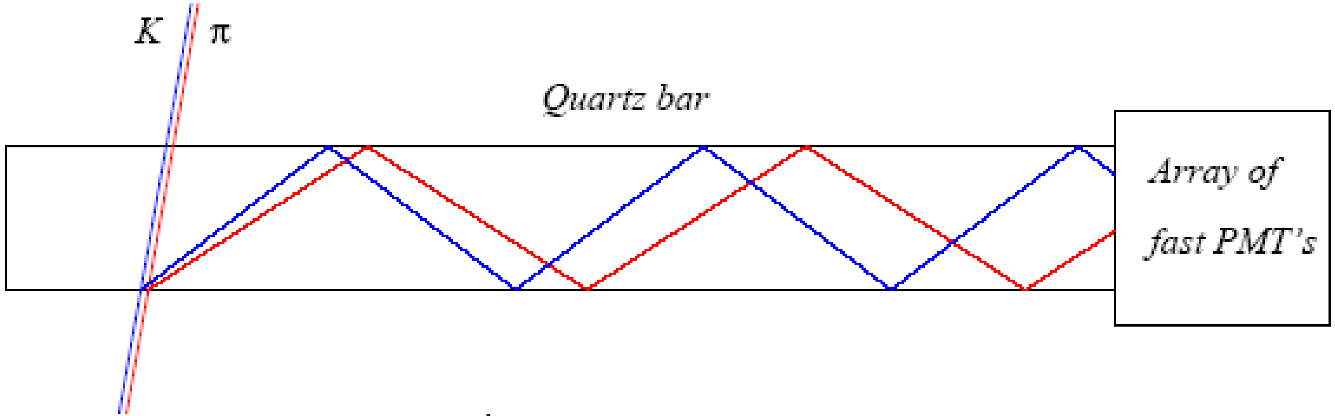


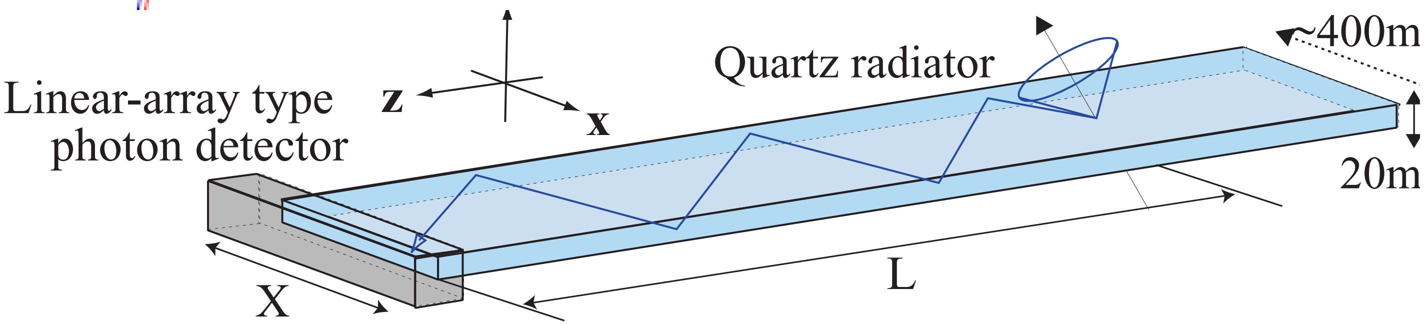
Clear improvement in Cherenkov angle resolution

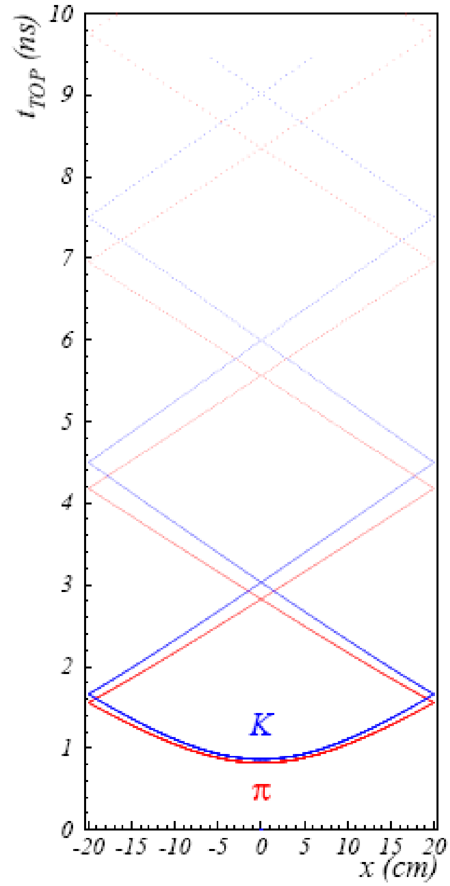
→NIMA 586 (2008) 174

More slides

Time-Of-Propagation (TOP) counter



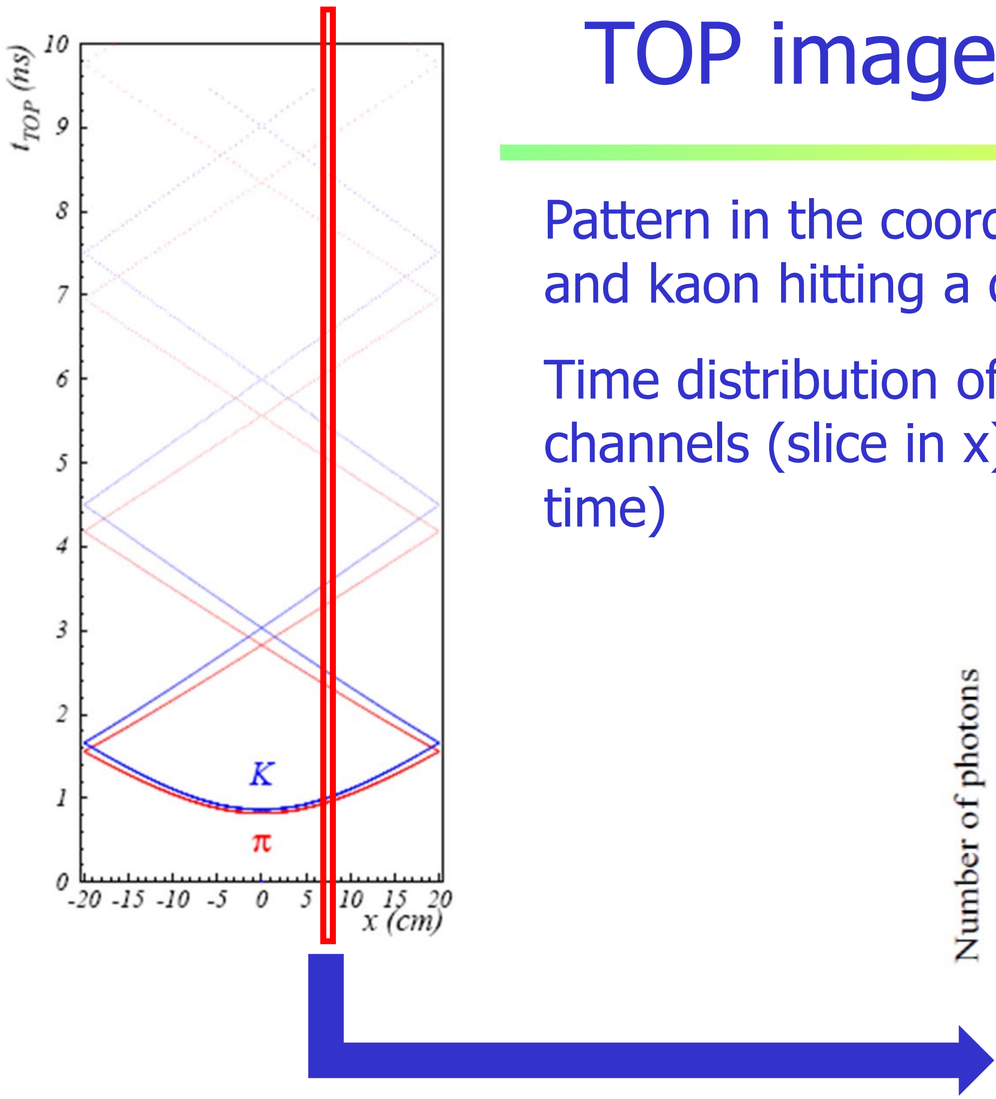




Similar to DIRC, but instead of two coordinates measure

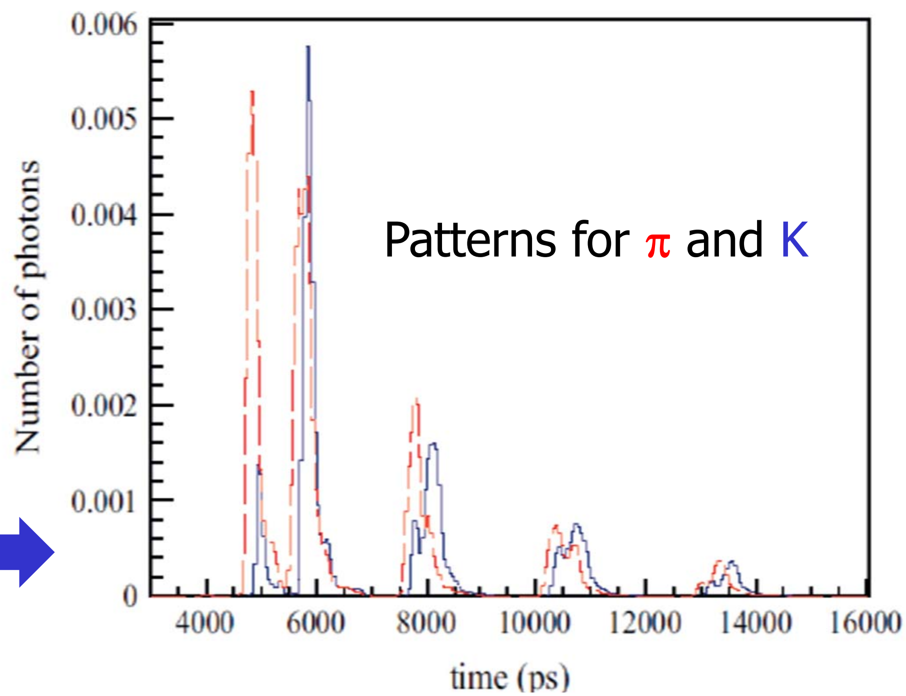
- One (or two coordinates) with a few mm precision
- Time-of-arrival

TOP image reconstruction



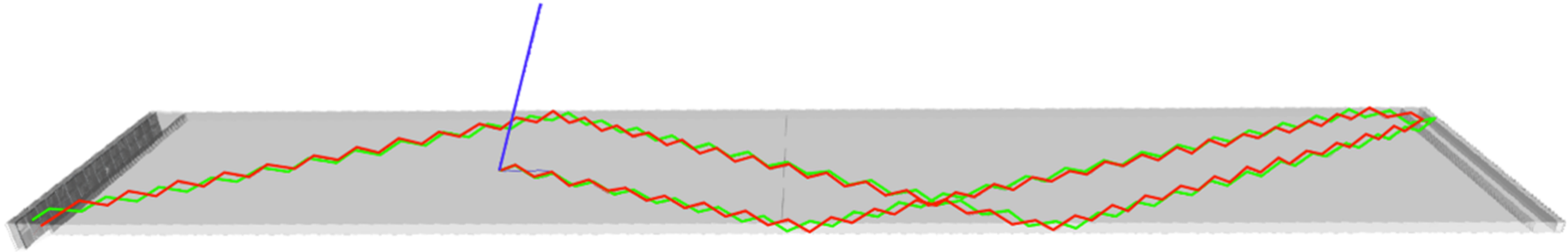
Pattern in the coordinate-time space ('ring') of a pion and kaon hitting a quartz bar

Time distribution of signals recorded by one of the PMT channels (slice in x): different for π and K (~shifted in time)

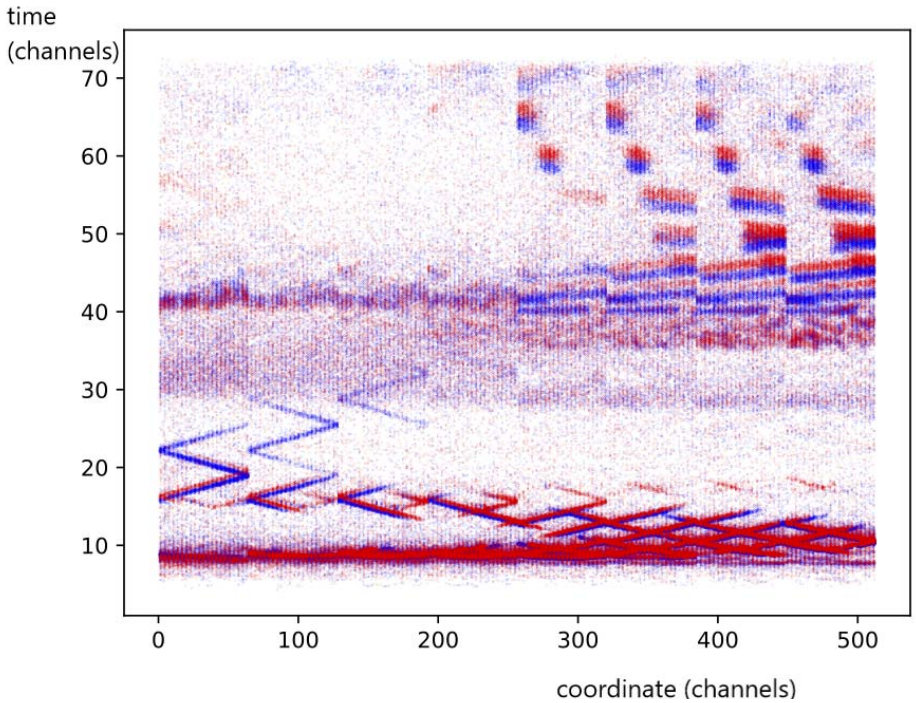


The name of the game: analytic expressions for the 2D likelihood functions →M. Starič et al., NIMA A595 (2008) 252-255

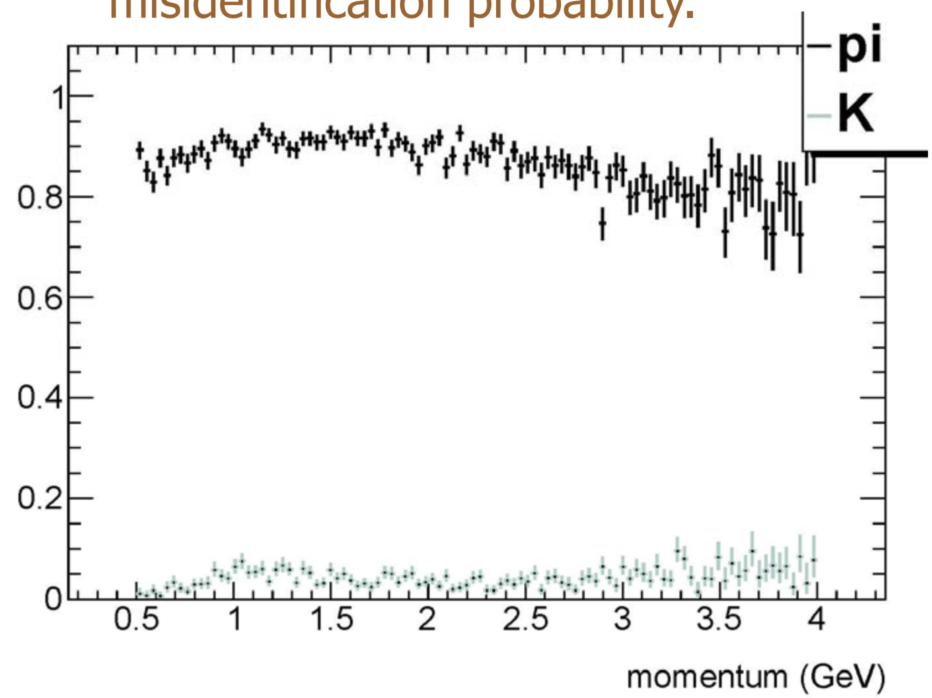
Separation of kaons and pions



Pions vs kaons in TOP:
different patterns in the time vs PMT impact point coordinate

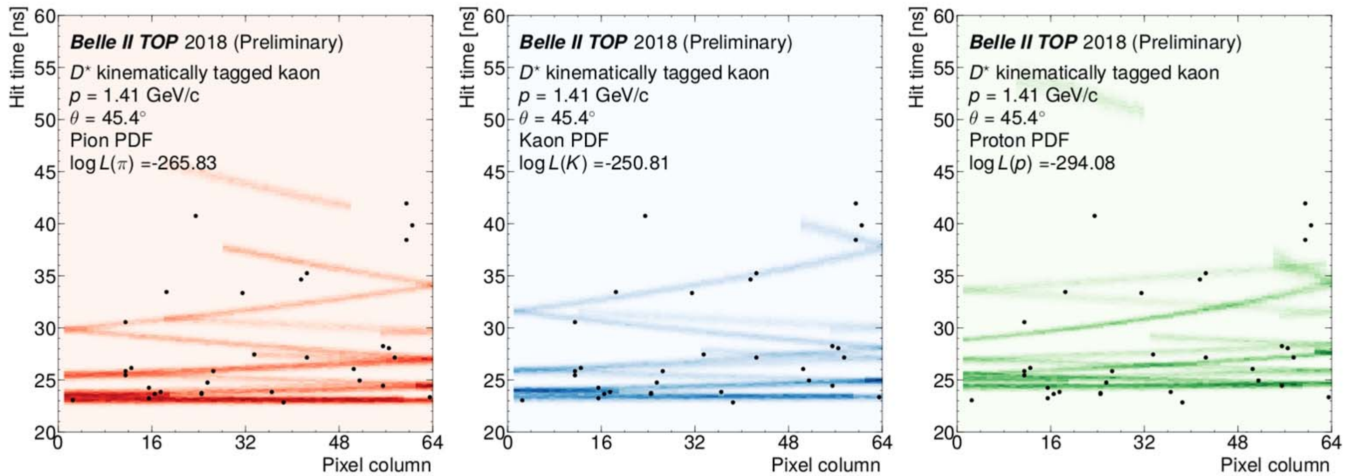


Pions vs kaons:
Expected PID efficiency and misidentification probability.



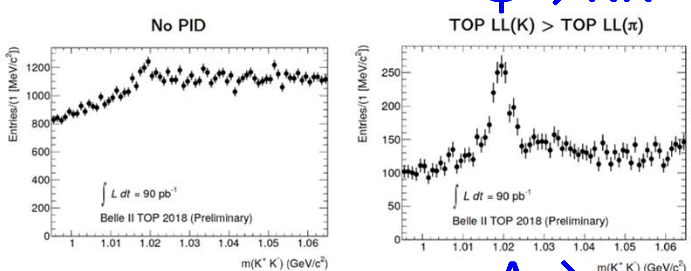
TOP first events

The early data demonstrated that the TOP principle is working



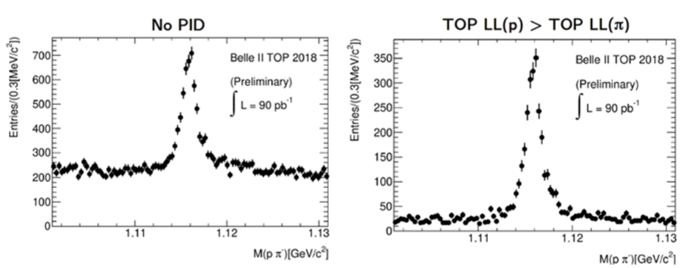
φ → K⁺K⁻ with both the tracks in the TOP acceptance

φ→KK

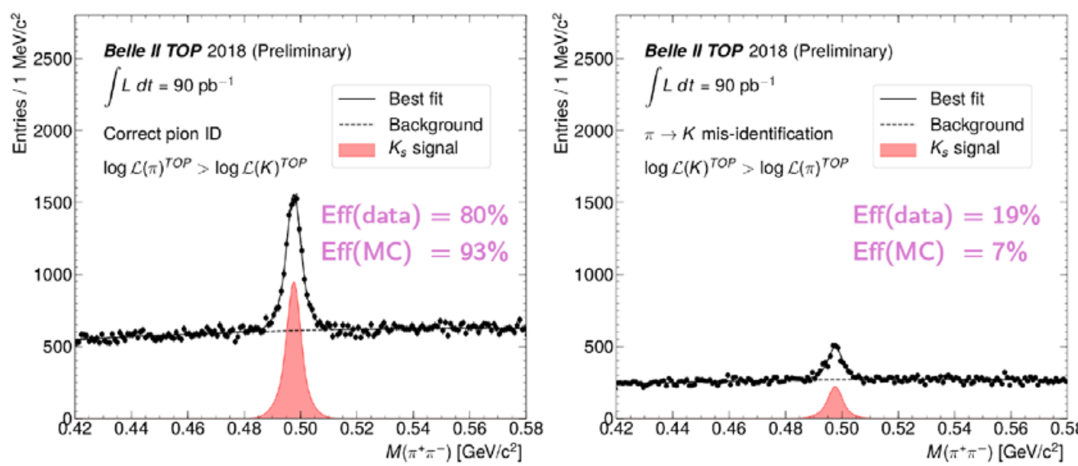


Λ → pπ with the **proton candidate** in the TOP acceptance

Λ→pπ



$K_s \rightarrow \pi\pi$

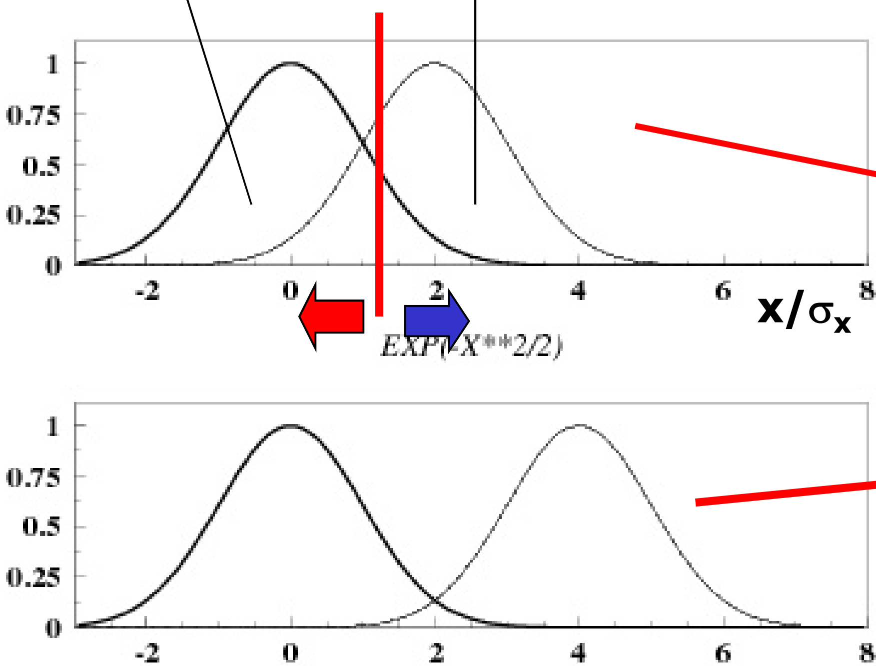


Efficiency and purity in particle identification

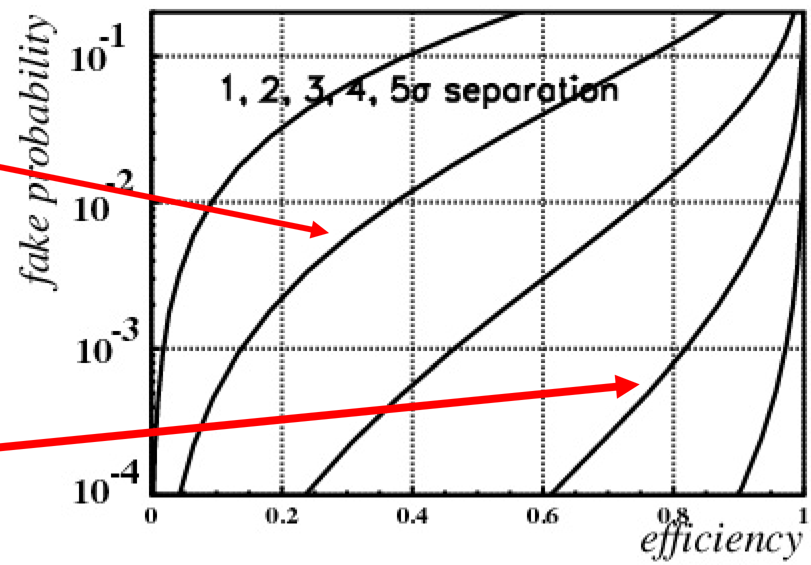
Efficiency and purity are tightly coupled!

Two examples:

particle type 1   type 2



eff. vs fake probability
(for Gaussian distributions)



- some discriminating variable **x**, scaled to the resolution σ_x