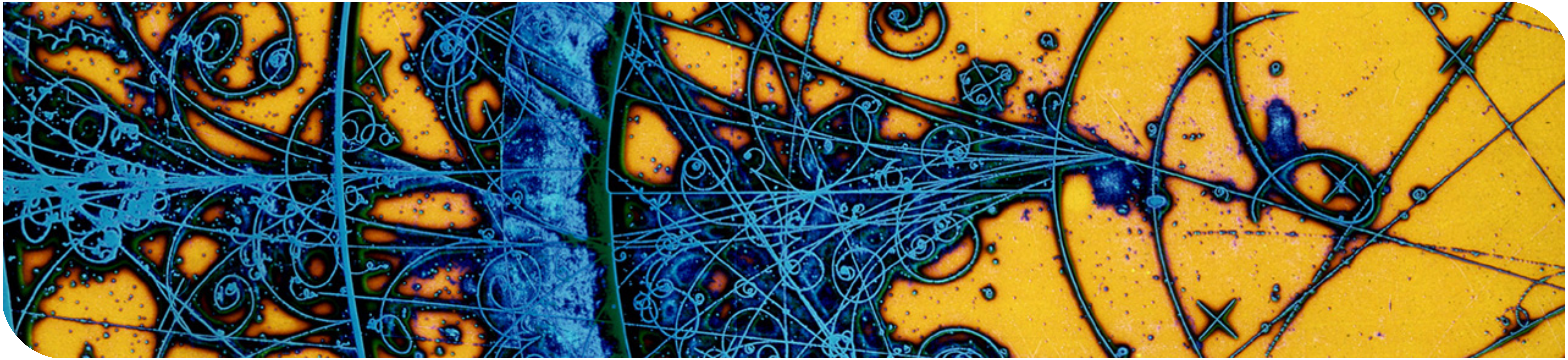
Particle Physics 1
Lecture 5: Reconstruction and calibration

Prof. Dr. Torben FERBER (torben.ferber@kit.edu, he/him), Institut für Experimentelle Teilchenphysik (ETP)
Wintersemester 2023/2024



Credit: CERN

Questions from past lectures

Learning goals

- Basic understanding of tracking performance
- Basic understanding of calorimeters

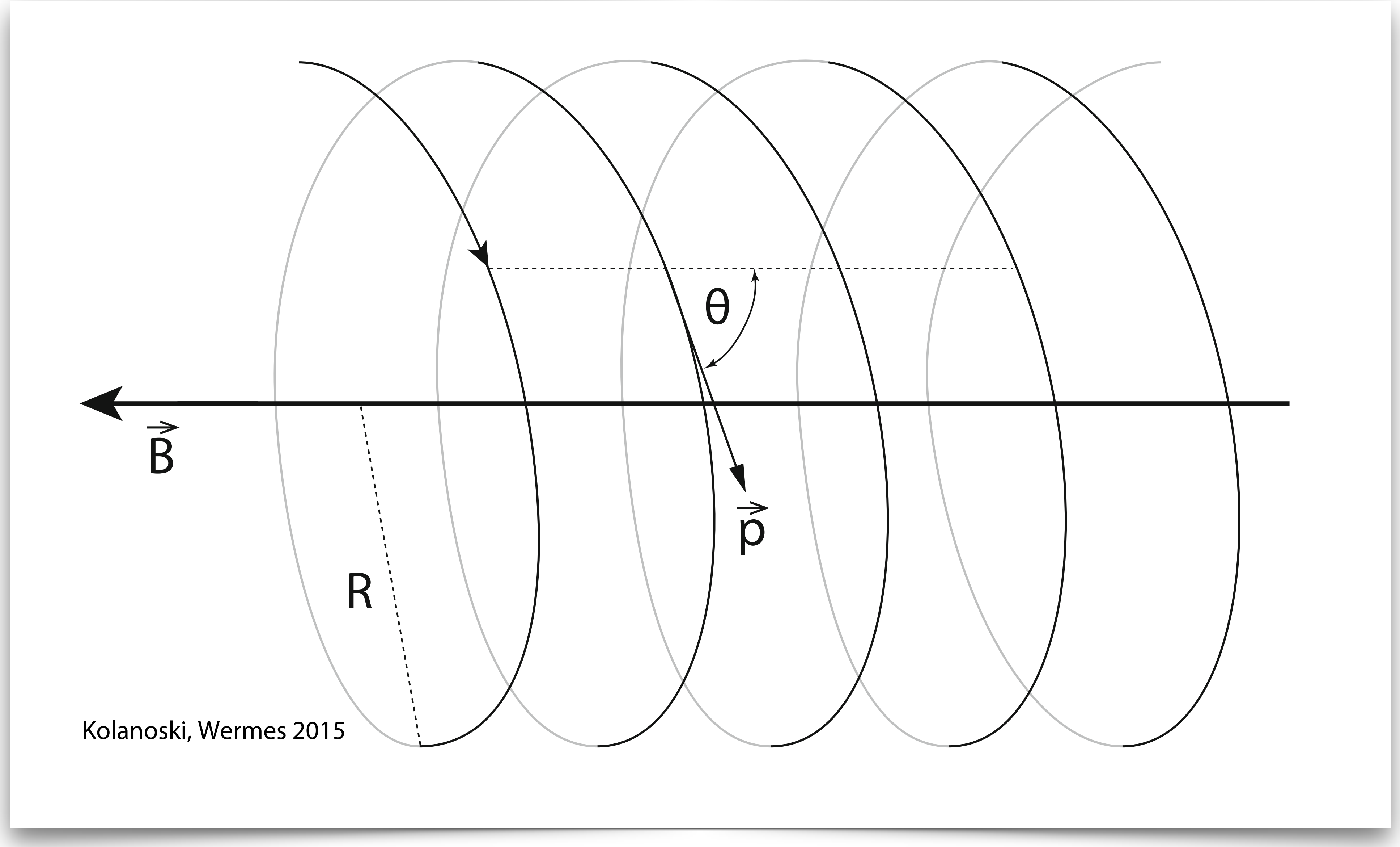
- Have a basic overview about analysis workflow in HEP
- Understand the need for data reduction
- Give examples for reconstruction algorithms (tracking, particle identification)
- Understand event selections, be able to describe own selections
- Give examples for calibration procedures
- Give examples for background modelling

Momentum resolution

- Tracks from charged track with $v = \beta\gamma$ form arcs with bending radius R in a magnetic field B:
  - Lorentz force: $F_L = q(\beta\gamma)B$
  - Centripetal force: $F_Z = m(\beta\gamma)^2/R$

$$\boxed{\rightarrow R = \frac{\gamma m v_T}{cqB} = \frac{p_T}{cqB} \approx \frac{p_T\,[\text{GeV}]}{0.3qB\,[\text{T}]}}$$

  - Note: The actual tracks forms a helix for non-zero p_z
  - Especially for low momentum tracks, energy loss along the helix is non-negligible

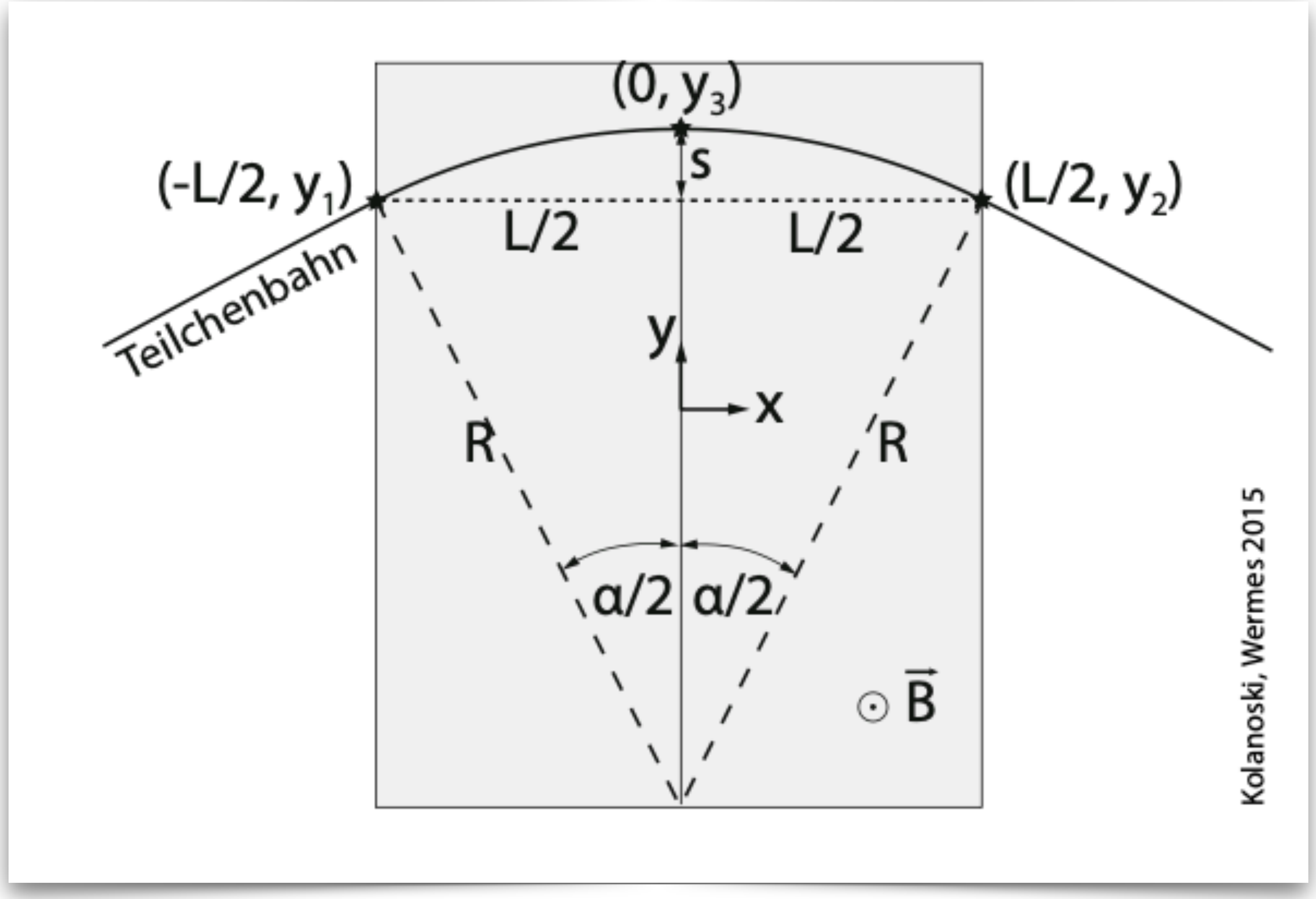


Track curvature

- In reality one measures a few points (need at least three to measure a curvature) along a particle track to get the **sagitta s**:

$$\left(\frac{L}{2}\right)^2 + (R - s)^2 = R^2$$

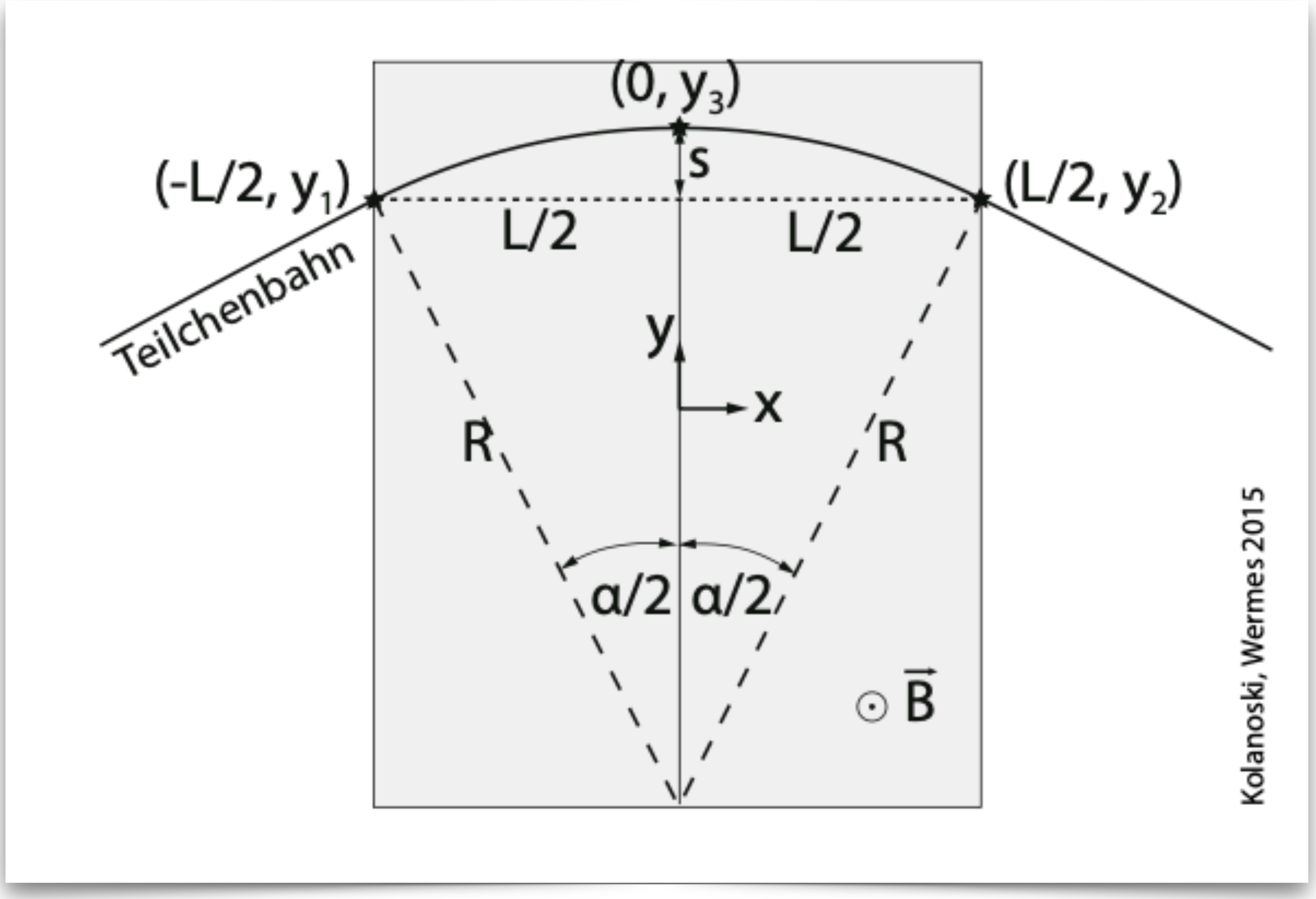
$$\rightarrow R = \frac{L^2}{8s} + \frac{s}{2} \rightarrow s \approx \frac{L^2}{8R}$$



- Sagitta is proportional to curvature 1/R
  - → Uncertainty on s is proportional to uncertainty on R

Example: Momentum resolution

- Typical scenario (e.g. ATLAS):
  - p = 1 TeV
  - B = 1 T
  - L = 4 m
  - q=±1
- $R \approx \dfrac{p[\mathrm{GeV}]}{0.3qB[\mathrm{T}]} = 1000\,\mathrm{GeV}/0.3\,\mathrm{T} \approx 3300\,\mathrm{m}$
- $s \approx \dfrac{L^2}{8R} = 16\,\mathrm{m}^2/(8 \cdot 3300\,\mathrm{m}) \approx 0.6\,\mathrm{mm}$
- If you want a 10% momentum resolution, you need
  $\Delta p/p = \Delta s/s \approx 60\mu\mathrm{m}$



Momentum resolution

- Uncertainty on s for three equidistant points with same uncertainty σ:

$$s = y_3 - \frac{y_1 + y_2}{2} \rightarrow \sigma_s = \sqrt{\sigma^2 - \frac{1}{4}2\sigma^2} = \sqrt{\frac{3}{2}}\sigma$$

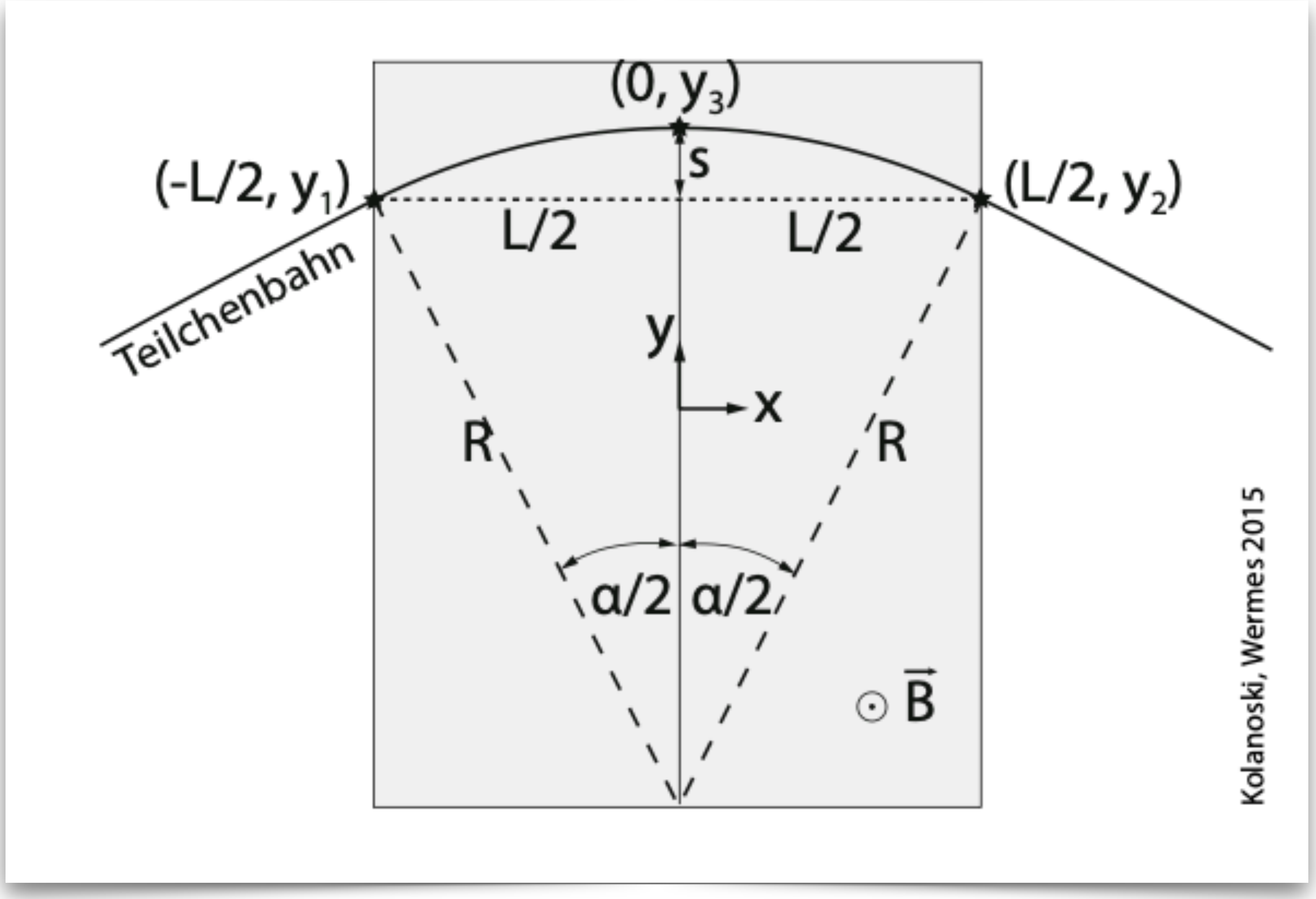
- Uncertainty on momentum:

$$\frac{\sigma_s}{s} = \frac{\sigma_{p_T}}{p_T} = \sqrt{\frac{3}{2}}\sigma \frac{8 p_T\,[\mathrm{GeV}]}{0.3 q B\,[\mathrm{T}] L^2}$$

- Generalization to large number of equidistant points (N>10), “Gluckstern-formula”:

$$\boxed{\frac{\sigma_{p_T}}{p_T} = \sqrt{\frac{720}{N+4}}\sigma \frac{p_T\,[\mathrm{GeV}]}{0.3 q B\,[\mathrm{T}] L^2}}$$

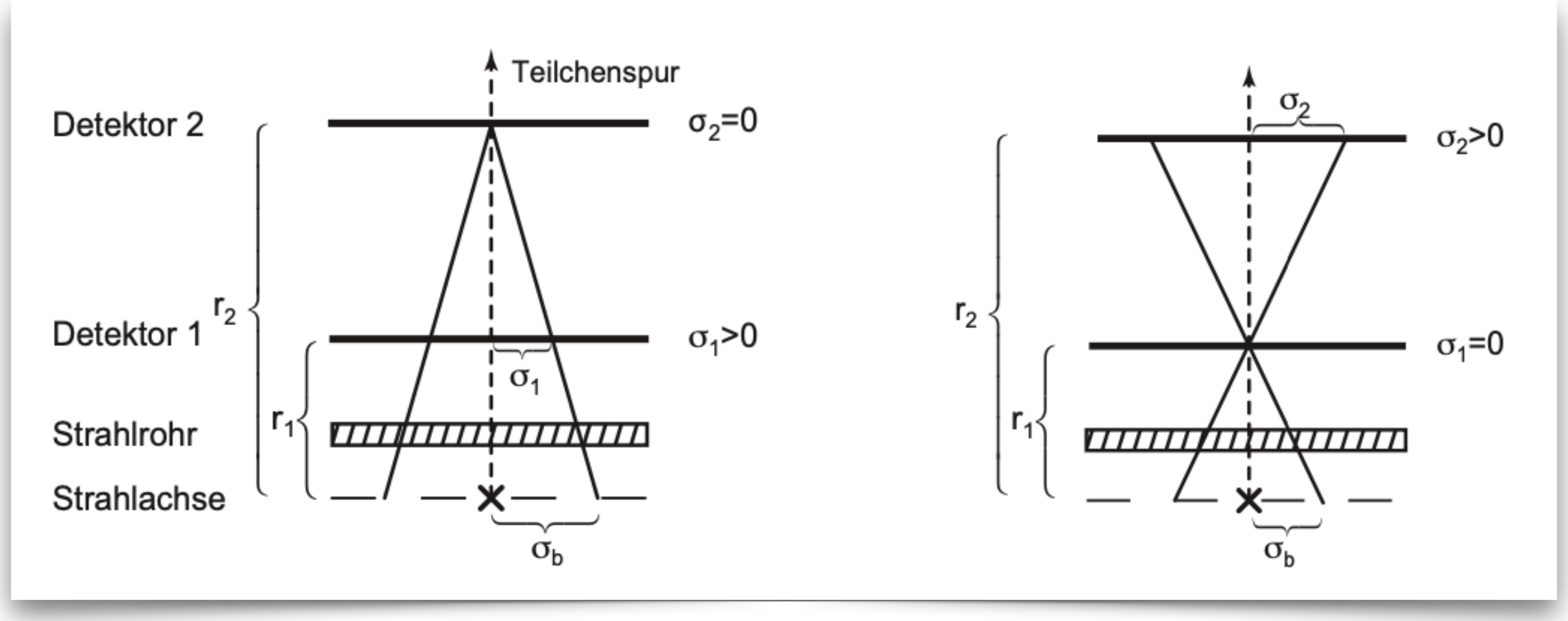
NIM 24 (1963) 381



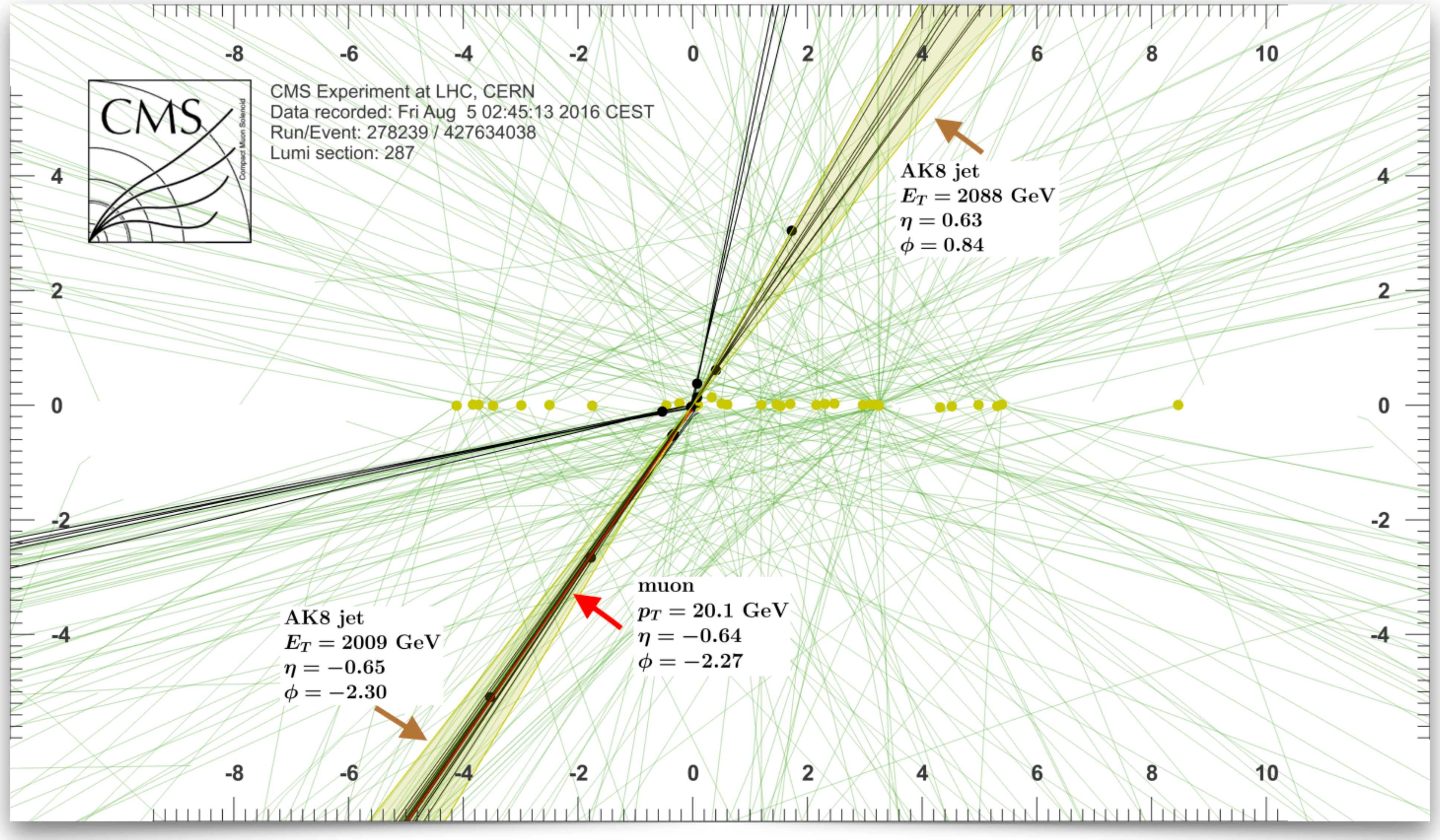
Position resolution with multiple scattering

- Simplified two-layer vertex detector:

$$\sigma_b^2 = \left(\frac{r_1}{r_2 - r_1}\sigma_2\right)^2 + \left(\frac{r_2}{r_2 - r_1}\sigma_1\right)^2 + \sigma_{\mathrm{MS}}^2$$



- Minimize σ_b:
  - Since $r_2 > r_1 \rightarrow \sigma_1$ must be small
  - Make $r_2 - r_1$ large
  - Make r_1 as small as possible
  - Make r_1 and beampipe as thin as possible to reduce σ_{MS}



PINGO

- **PINGO:**
  - Umfrage: Teilchenphysik 1 (WS 23/24)
  - Zugangsnummer: 434521
  - Link: https://pingo.coactum.de/events/434521

PINGO: Detector optimization

- In order to improve the momentum resolution for high p_T tracks of a detector, would you invest in:
  - Making the tracking detector larger
  - Adding more detector layers
  - Increasing the magnetic field
  - Making the sensors thicker (increase efficiency)
  - Remove passive material (reduce multiple scattering)

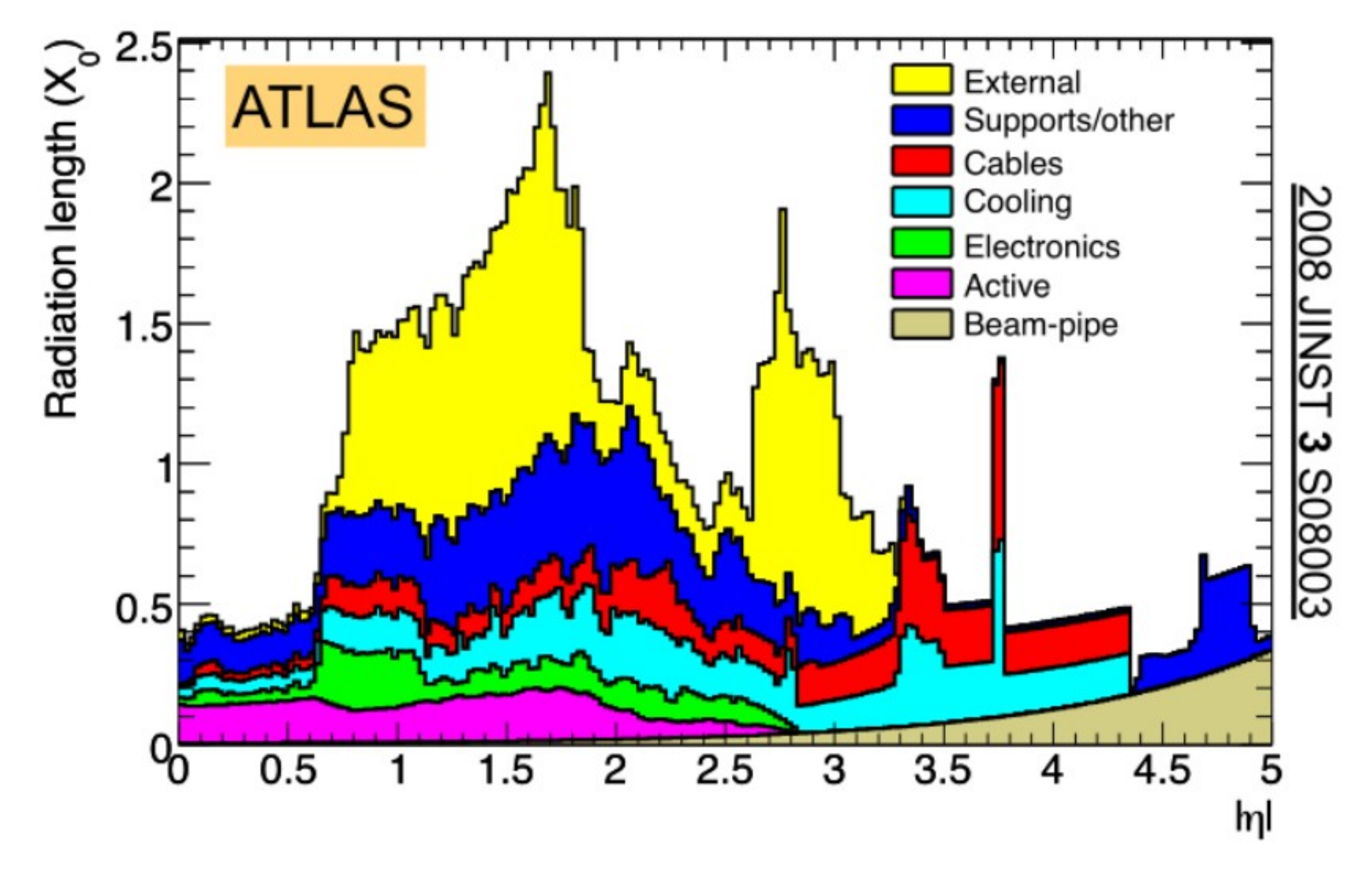
$$\frac{\sigma_{p_T}}{p_T} = \sqrt{\frac{720}{N+4}}\sigma\frac{p_T\,[\text{GeV}]}{0.3qB\,[\text{T}]L^2} \oplus \frac{\text{const.}}{B\sqrt{LX_0}}$$

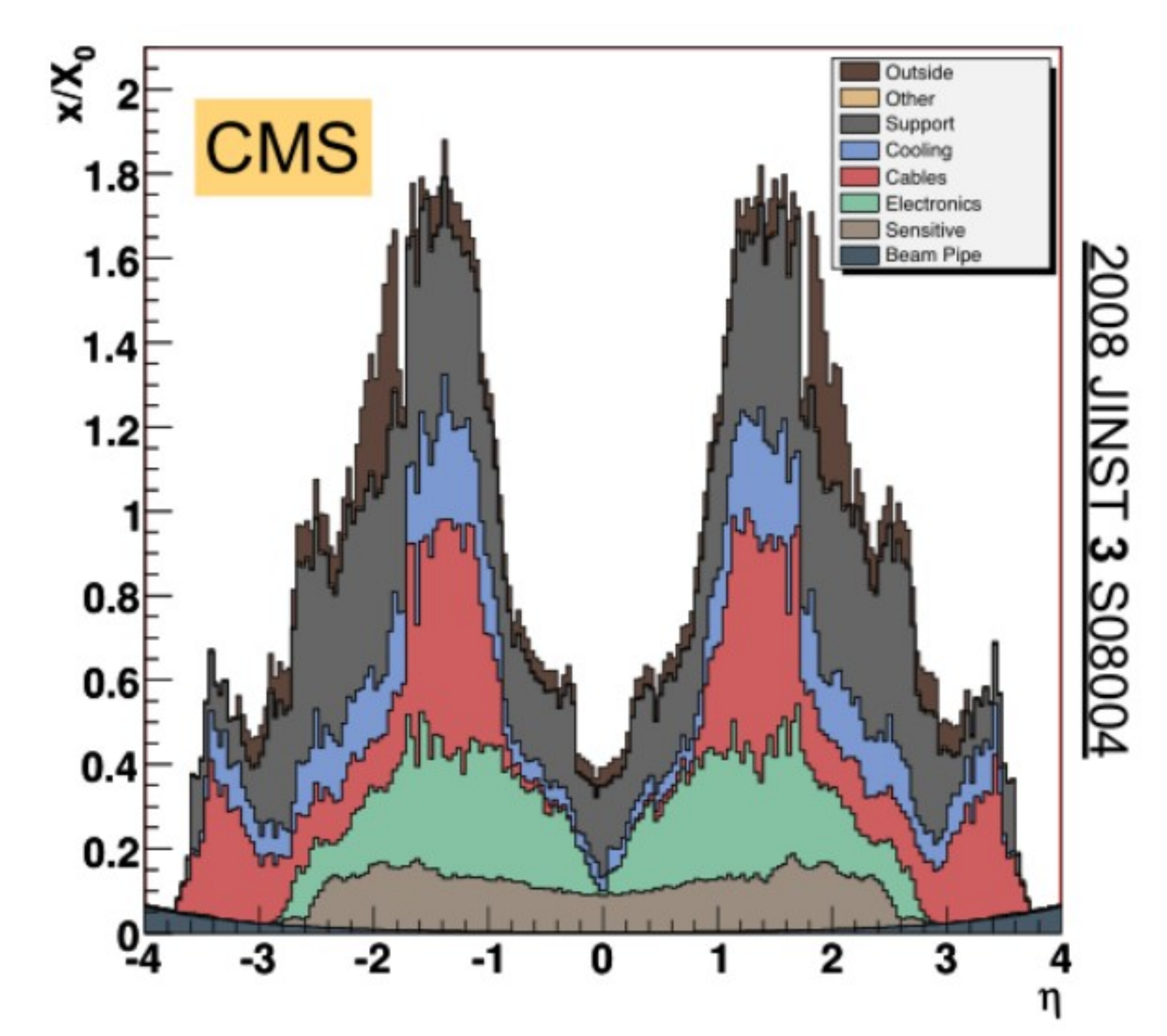
PINGO: Detector optimization

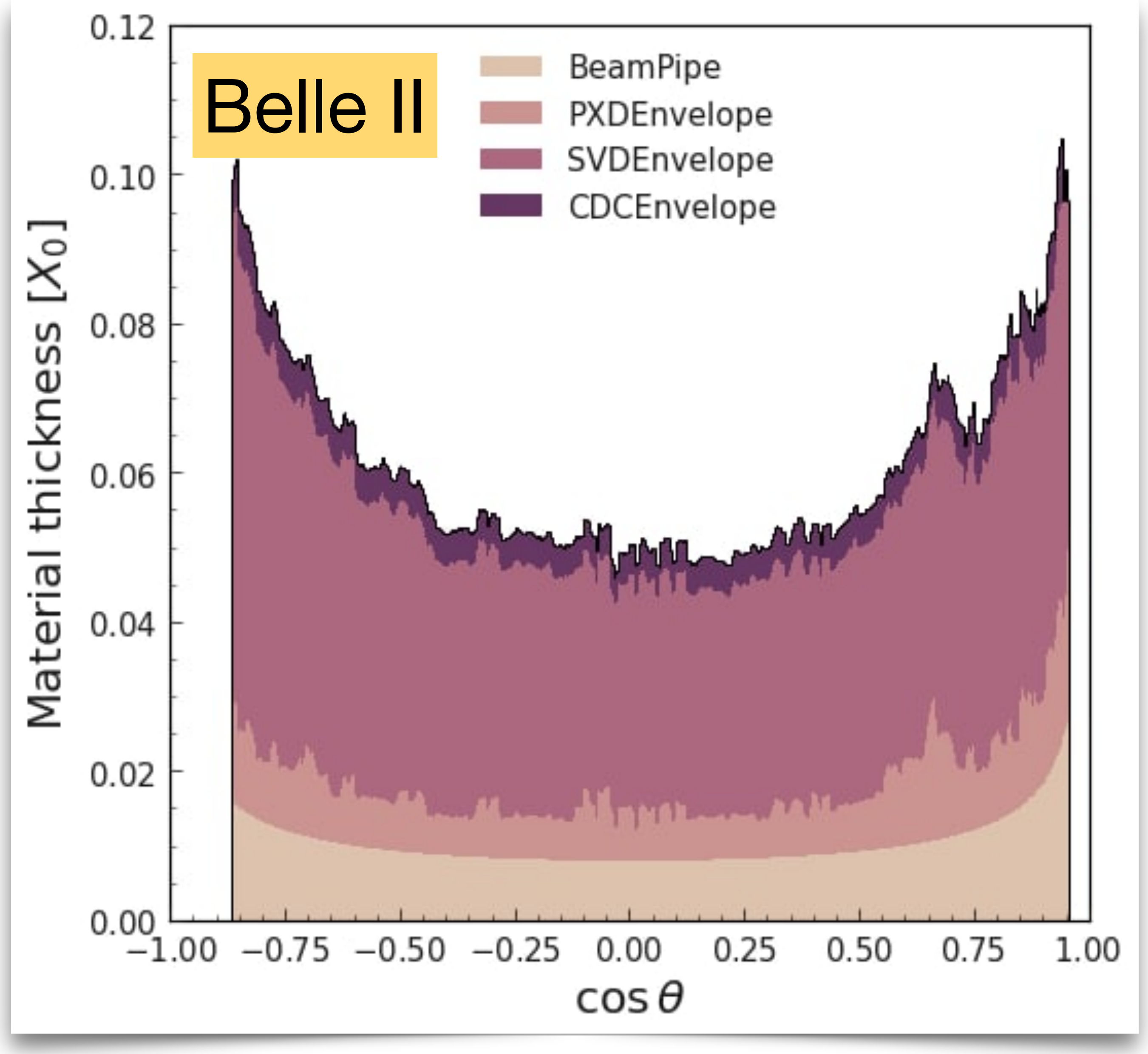
- In order to improve the momentum resolution for high p_T tracks of a detector, would you invest in:
  - Making the tracking detector larger (correct answer)
  - Adding more detector layers
  - Increasing the magnetic field
  - Making the sensors thicker (increase efficiency)
  - Remove passive material (reduce multiple scattering)

$$\frac{\sigma_{p_T}}{p_T} = \sqrt{\frac{720}{N+4}}\sigma \frac{p_T\,[\mathrm{GeV}]}{0.3qB\,[\mathrm{T}]L^2} \oplus \frac{\mathrm{const.}}{B\sqrt{LX_0}}$$

Example Material in front of tracking detectors

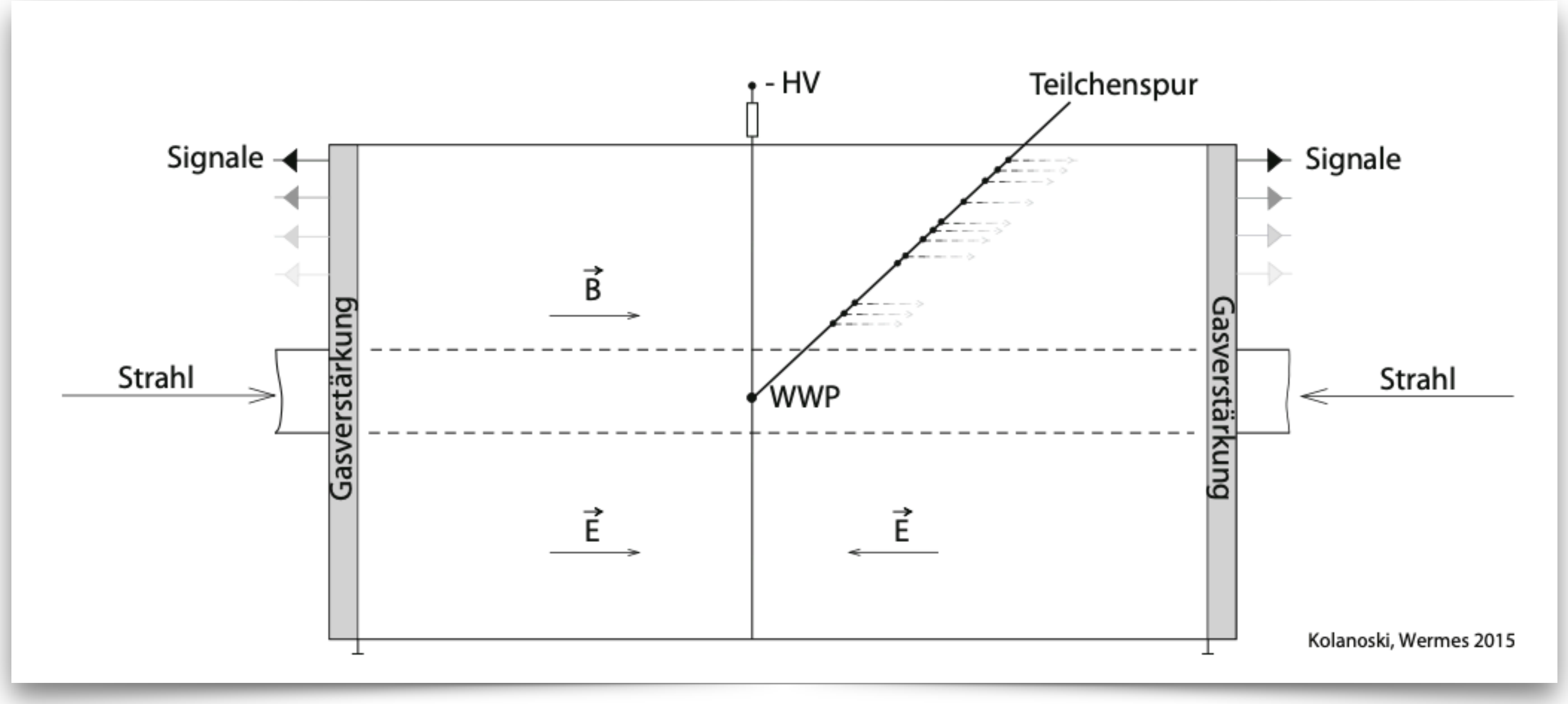






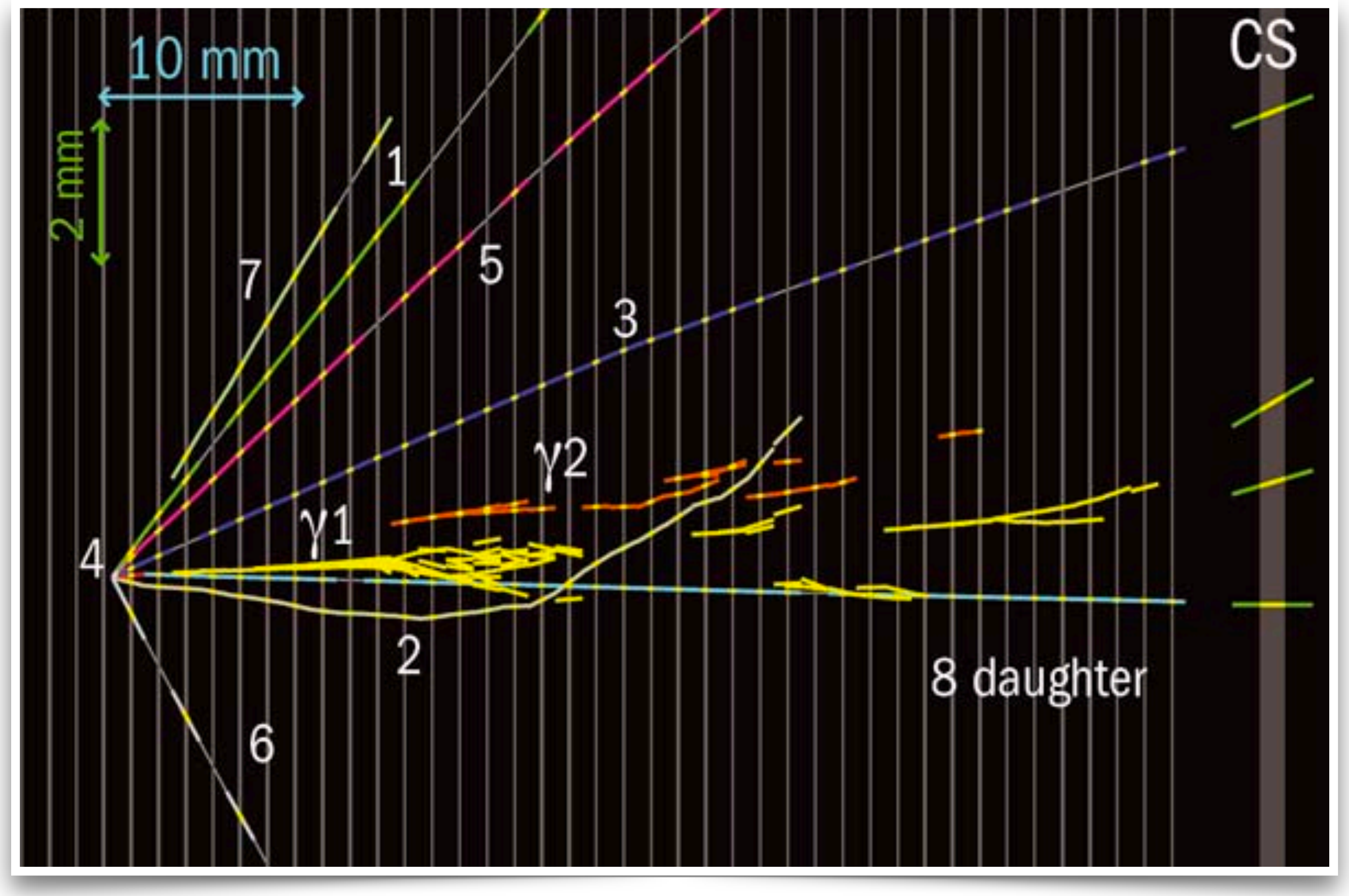
Tracking detectors: Time-projection chamber (TPC)

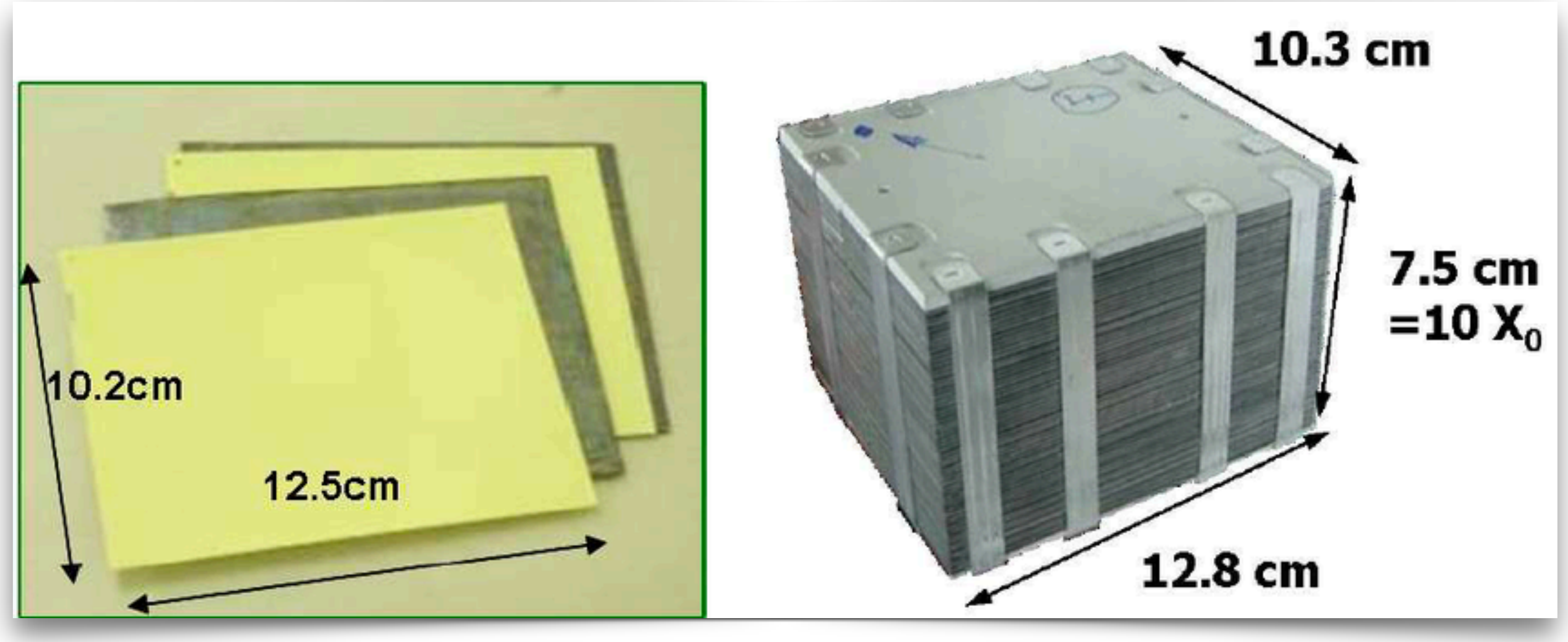
- Charged particles create electrons along path in gas-filled medium
- Electrons drift long distances along homogenous E-field to readout planes
- Readout-planes determine both x-y coordinates and arrival time → 3D track
- Instead of gas, also liquid Argon is used for example in the ICARUS neutrino detector



Tracking detectors: More…

- Resistive plate chambers:
  - Simple gas-filled detectors with very good timing resolution but worse position resolution (often used in muon chambers)
- Emulsion cloud chambers (ECC):
  - Photoemulsions with not timing resolution, but excellent 3D position resolution and very large target mass (popular in neutrino detectors)
- Bubble chambers
- Spark chambers
- Cloud chambers
- …





Source: OPERA

Calorimeters

- Mainly used interaction: pair conversion, bremsstrahlung, nuclear interactions
- Two main detector types, both work by absorbing all particles:
  - Electromagnetic calorimeters (for photons and electrons) → ECAL
  - Hadronic calorimeters (well, for hadrons…) → HCAL
- Performance metrics:
  - Relative energy resolution $\Delta E/E$
  - Position resolution of shower center (2D or 3D)
  - Missing energy reconstruction
  - Particle identification (later)

Calorimeter energy resolution

- High intrinsic energy resolution (large light yield, low electronics noise)
- High hermiticity (no “dead” material between active detector)
- Very deep calorimeters to avoid longitudinal leakage
- Simplified resolution (often needs modification for very low energies and if reconstruction algorithms are not just suming up energy depositions)

quadratic sum: $a \oplus b = \sqrt{a^2 + b^2}$

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E[\mathrm{GeV}]}} \oplus \frac{b}{E[\mathrm{GeV}]} \oplus c$$

Stochastic term: Fluctuations

Noise term

Constant term

Calorimeter energy resolution

- Stochastic term:
  - Sampling calorimeter: Fluctuations of energy deposited in passive and active layers $\propto 1/\sqrt{n_{\text{visible}}}$, plus:
    - low E photon interaction cross sections energy-dependent
    - energy threshold
    - scattering angle and resulting path length in active material
    - …
  - Homogenous calorimeters: Usually small (often $\propto a/\sqrt[4]{E}$) unless photon-statistics is very small (e.g. lead-glass)

homogenous: absorber and signal generation in the same material

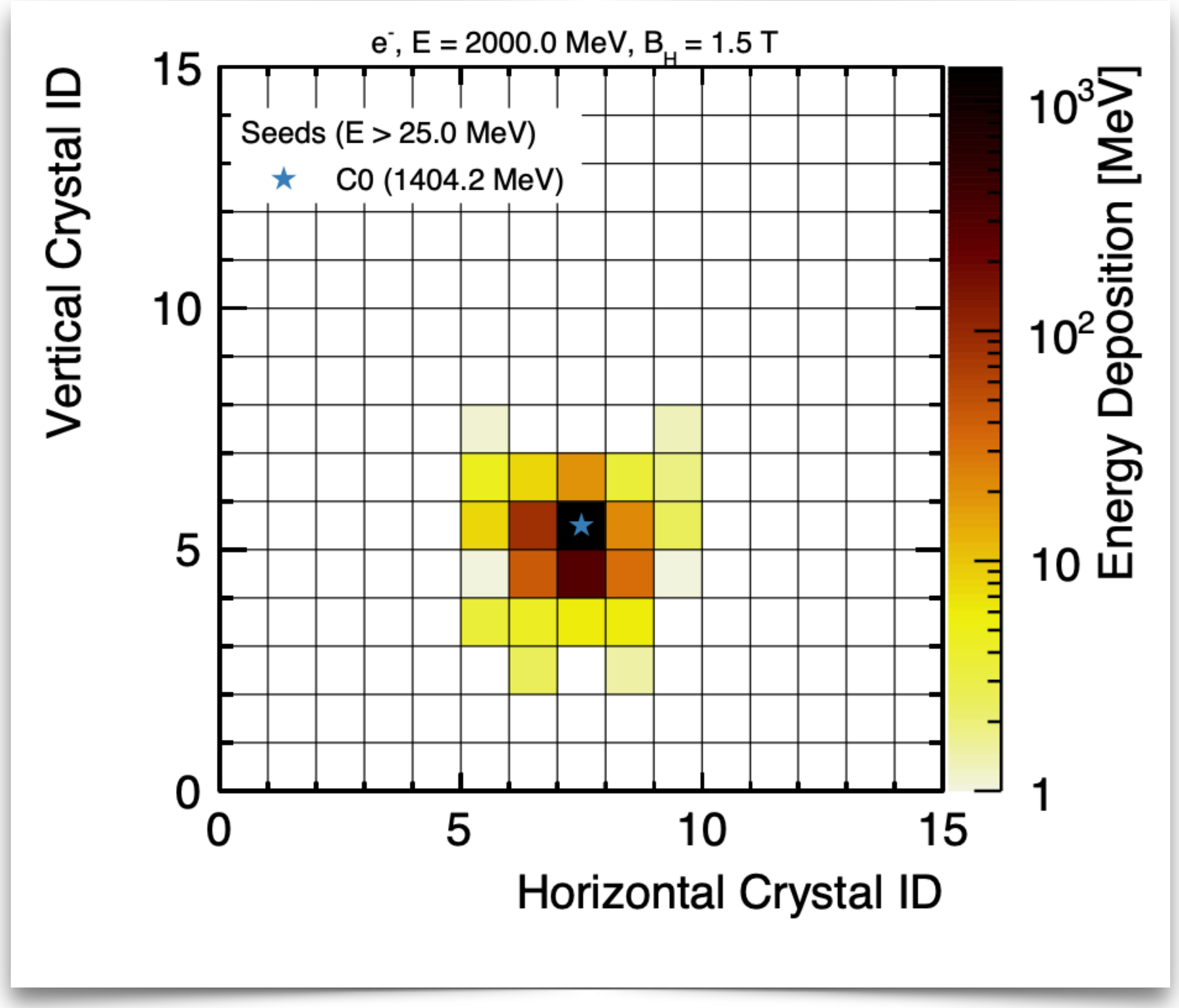
sampling: absorber (passive) and detector (active) in different materials

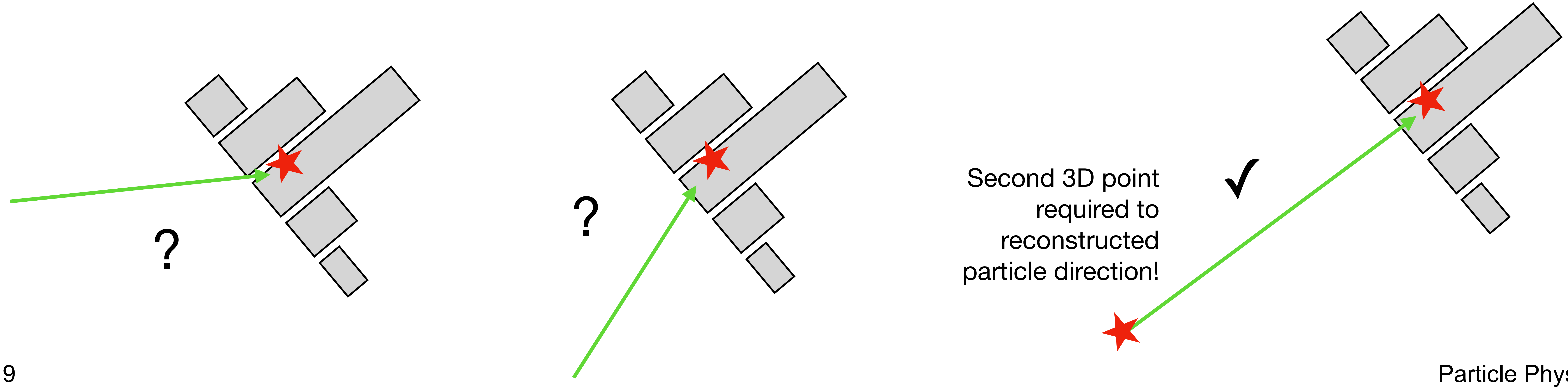
Calorimeter energy resolution

- Noise term:
  - "equivalent noise energy" (ENE) determines minimal detectable energy determined by readout electronics and radiation backgrounds
    - liquid-argon calorimeter in ATLAS: b ~ 190 MeV
    - CsI(Tl) calorimeter in Belle II: b ~ 0.2 MeV
- Constant term:
  - Mechanical imperfections
  - (Longitudinal) leakage (i.e. not all energy contained) (actually scales $\propto \ln(E)$, not exactly constant)
  - …

Position resolution

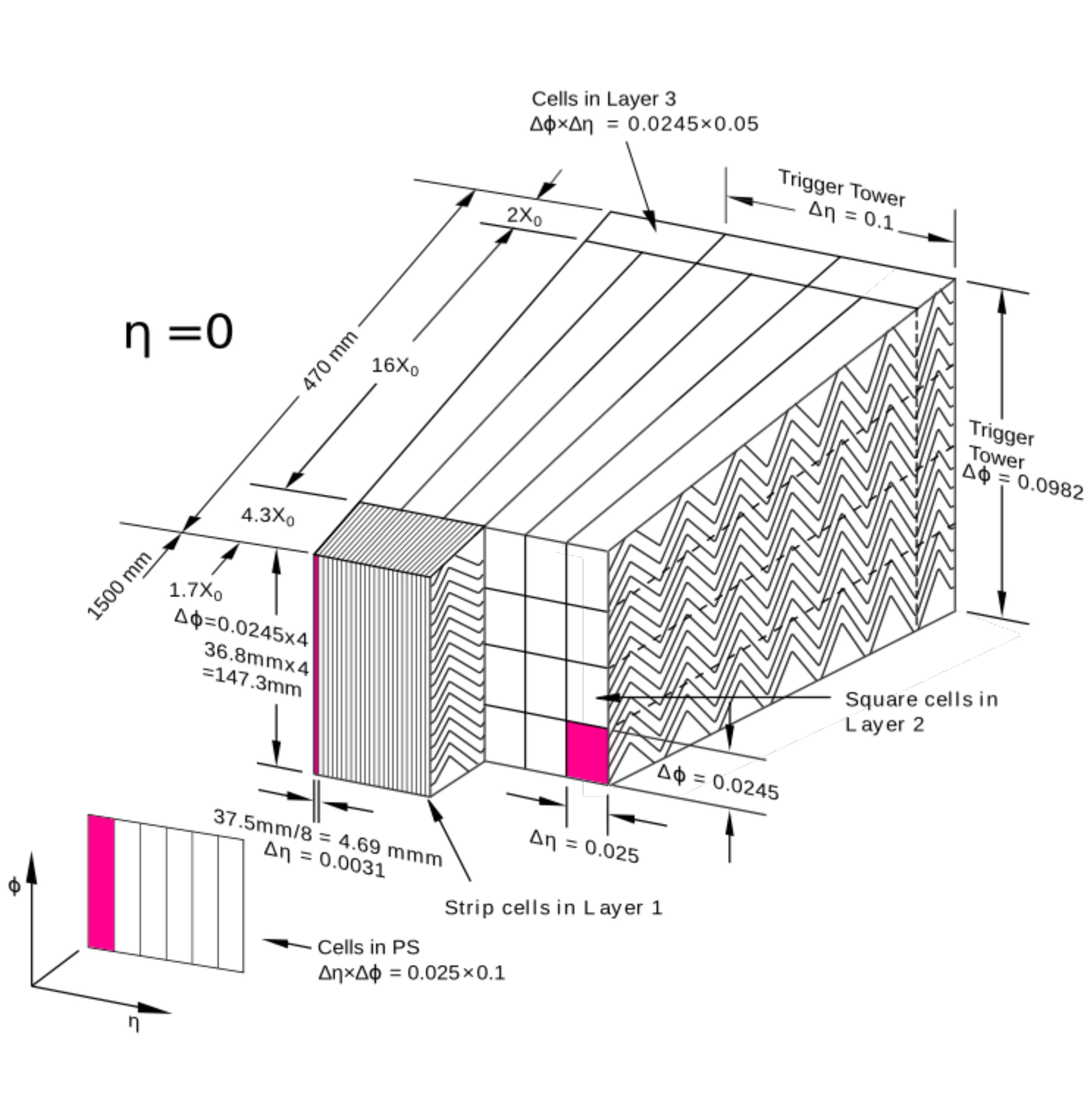
- Position of a cluster is determined by the center of gravity of the energy depositions (i.e. much better than $d/\sqrt{12}$
- The particle direction is typically not determined by the calorimeter alone but by combining the cluster position and a known (or assumed) origin of the particle





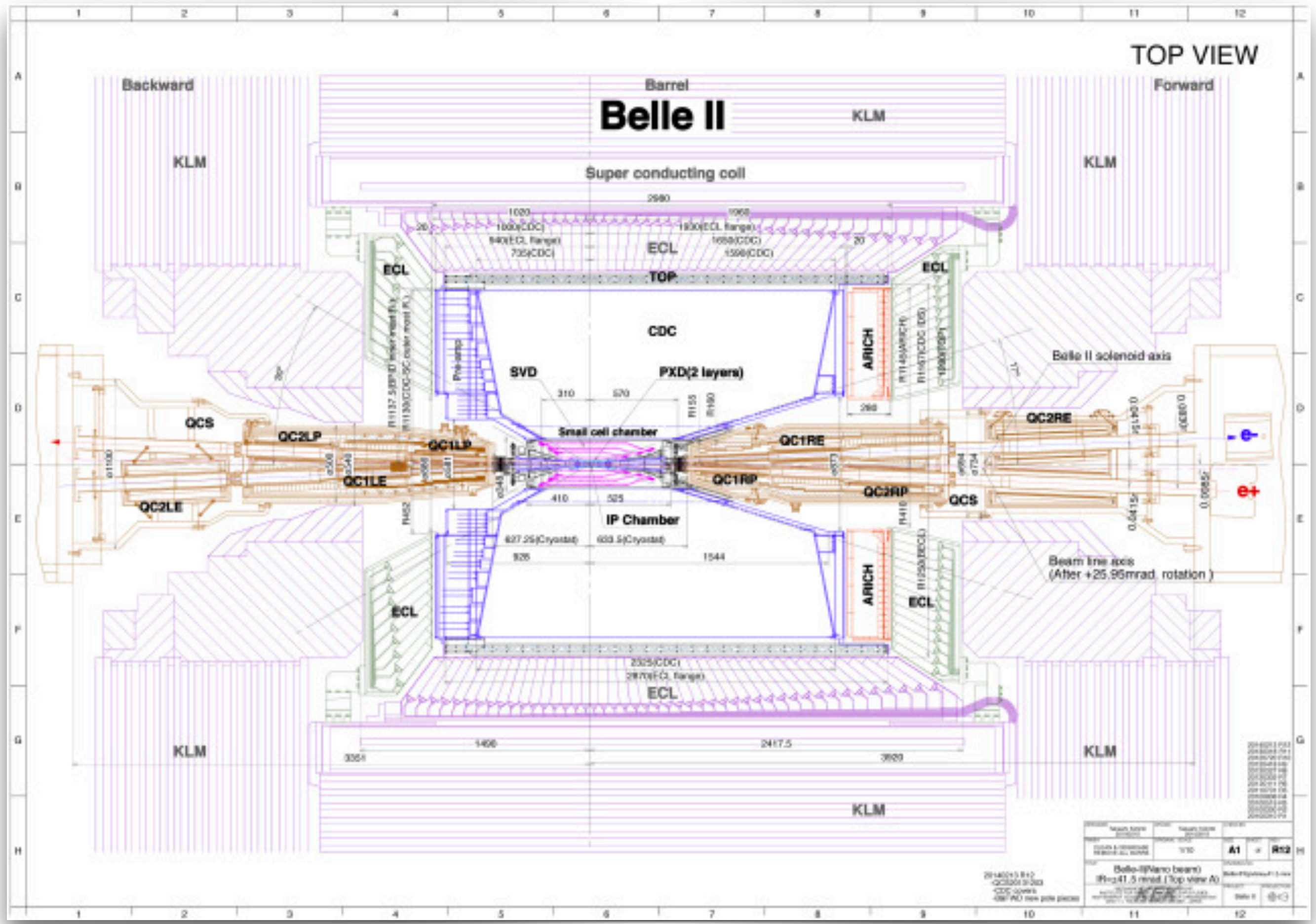
Example: Liquid-Argon ECAL in ATLAS

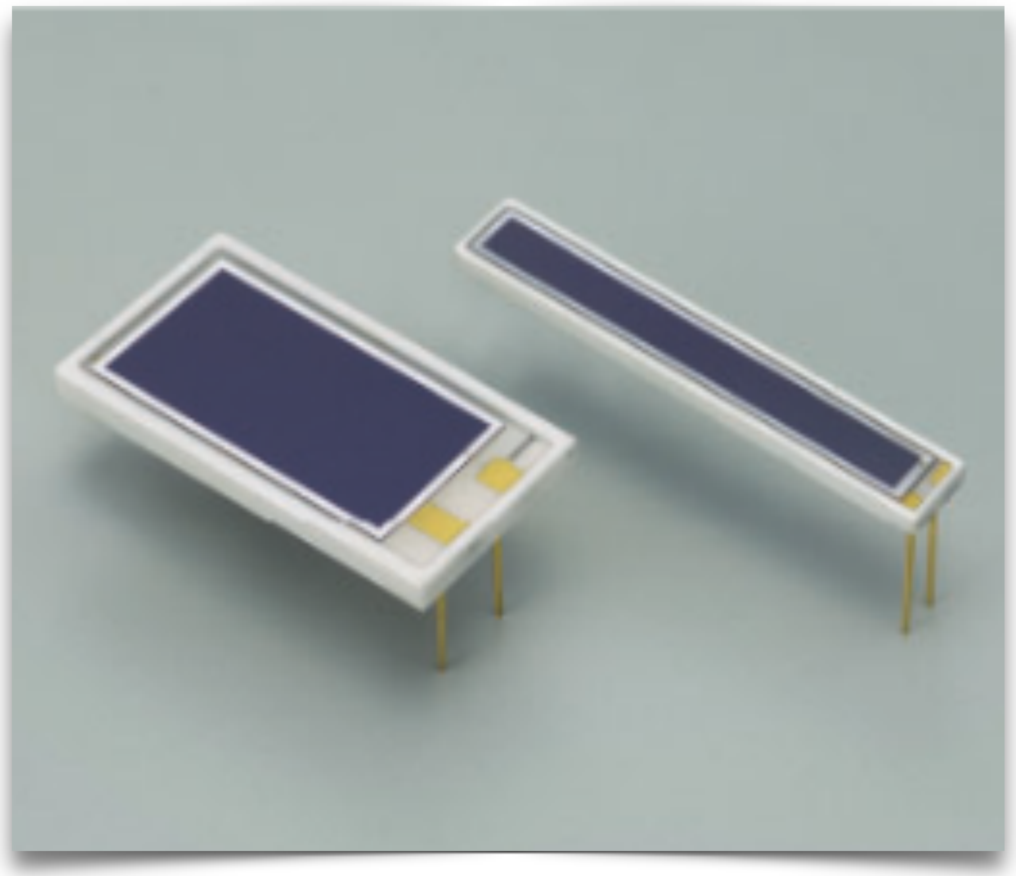
- Sampling calorimeter: Lead (passive) and liquid argon at 80K (active)
  - Ionization in pure liquid Argon (26 eV per e/ion pair)
  - Ions drift to electrodes (~1000V)
  - Accordion structure to avoid gaps in detector coverage
  - several layers with different segmentations
- ATLAS ECAL is behind the solenoid magnet: Additional $3 - 6X_0$ in front of calorimeter!



Example: CsI(Tl) Crystal ECAL in Belle II

- Homogeneous calorimeter: Thallium-doped CsI crystals:
  - Non-projective geometry
  - Very high light yield
  - Rather slow light collection (~20μs)
- Positioned inside the solenoid magnet
  - Readout with Si photodiodes
- Optimized to measure very low energy photons (down to 20 MeV) → needs small noise term, dominated by <1 MeV photons





Source: Hamamatsu S2744-08

Hadron calorimeters (HCAL)

- Three main challenges:
  - Hadronic interaction length is (much) larger than the radiation length

    → need a deeper calorimeter to collect total hadron energy
  - The hadronic shower consists to about 1/3 of neutral pions that decay into two photons

    → part of the hadronic shower is purely electromagnetic
  - The hadronic shower contains a large fraction (20-40%) of “invisible” particles

    → much worse energy resolution compared to ECAL

Hadron calorimeters: Compensation

- Unknown electromagnetic fraction of the hadronic shower is a challenge:

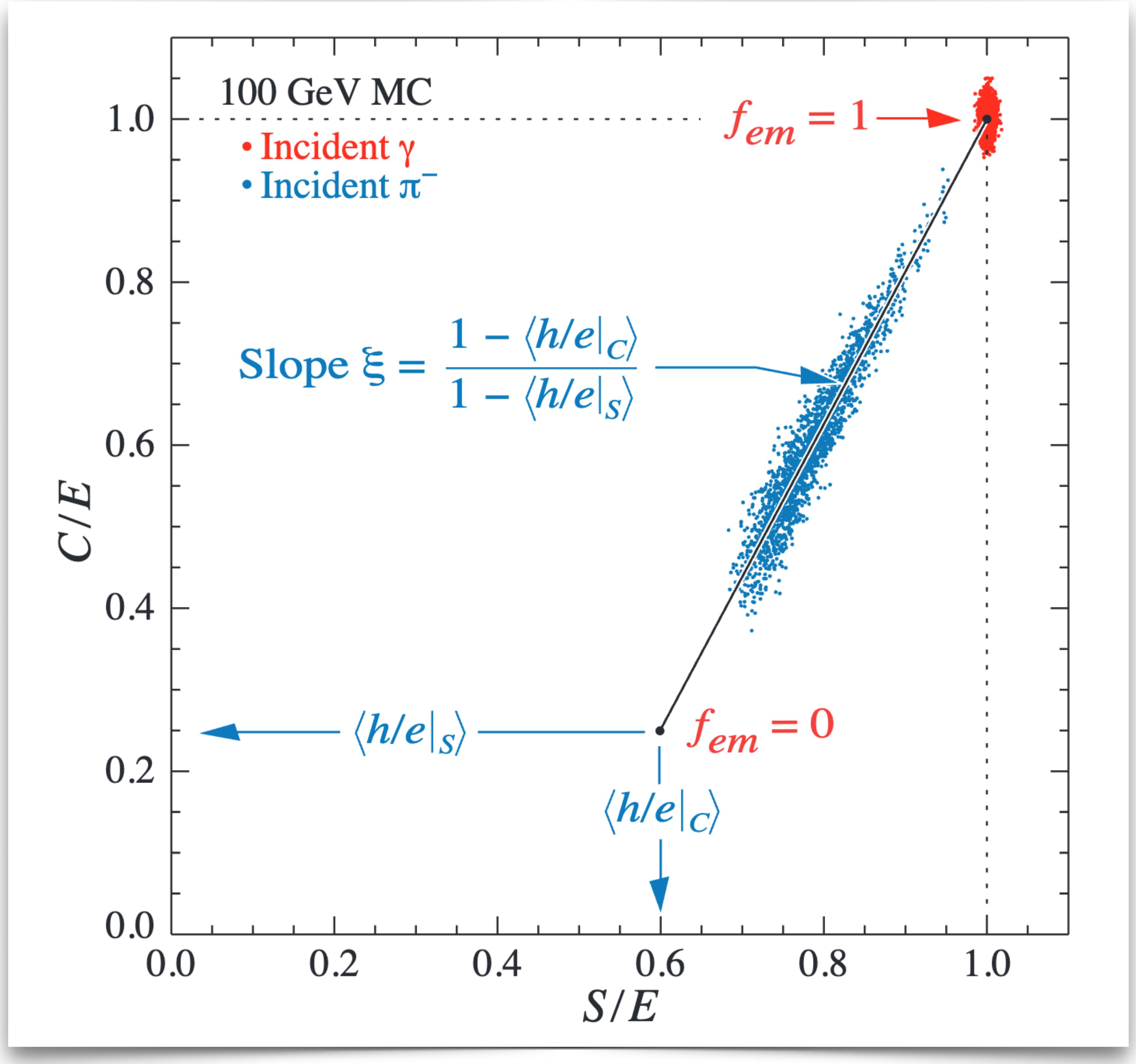
  $$E_{\text{measured}} = \left(f_{em}e + (1 - f_{em})h\right)$$

  where e and h are is the respective responsive to an EM or an hadronic energy deposition, and f_{em} is the electromagnetic shower fraction

  - if $e \neq h$, the calorimeter itself is sensitive to fluctuations of f_{em}
  - since $e = e(E)$ the response of the calorimeter becomes non-linear

Hadron calorimeters: Compensation

- Solution is conceptually simple: make e=h ("compensation")
  - **Hardware compensation**:
    - Decrease of em sensitivity, e. g. thicker absorber
    - Use hydrogen-rich active detector to increase neutron interaction
    - Increase visibly energy of neutrons by spallation (e.g. ZEUS Uranium HCAL)
  - **Software compensation**:
    - Highly segmented calorimeter that can identify cells with low and high local energy depositions and weight them accordingly
  - **Design compensation** ("the dream"):
    - Dual read-out with very different values of e/h exploiting different kinematics of electrons and positrons in the EM and hadronic part of the shower using detector that are sensitive to particle velocity β



Credit: M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018).

$$S = E\left[f_{em} + \left(\tfrac{h}{e}\right)_S (1 - f_{em})\right]$$
$$C = E\left[f_{em} + \left(\tfrac{h}{e}\right)_C (1 - f_{em})\right]$$

$$E = \frac{\xi S - C}{\xi - 1}$$

Hadron Comparison: ATLAS and CMS

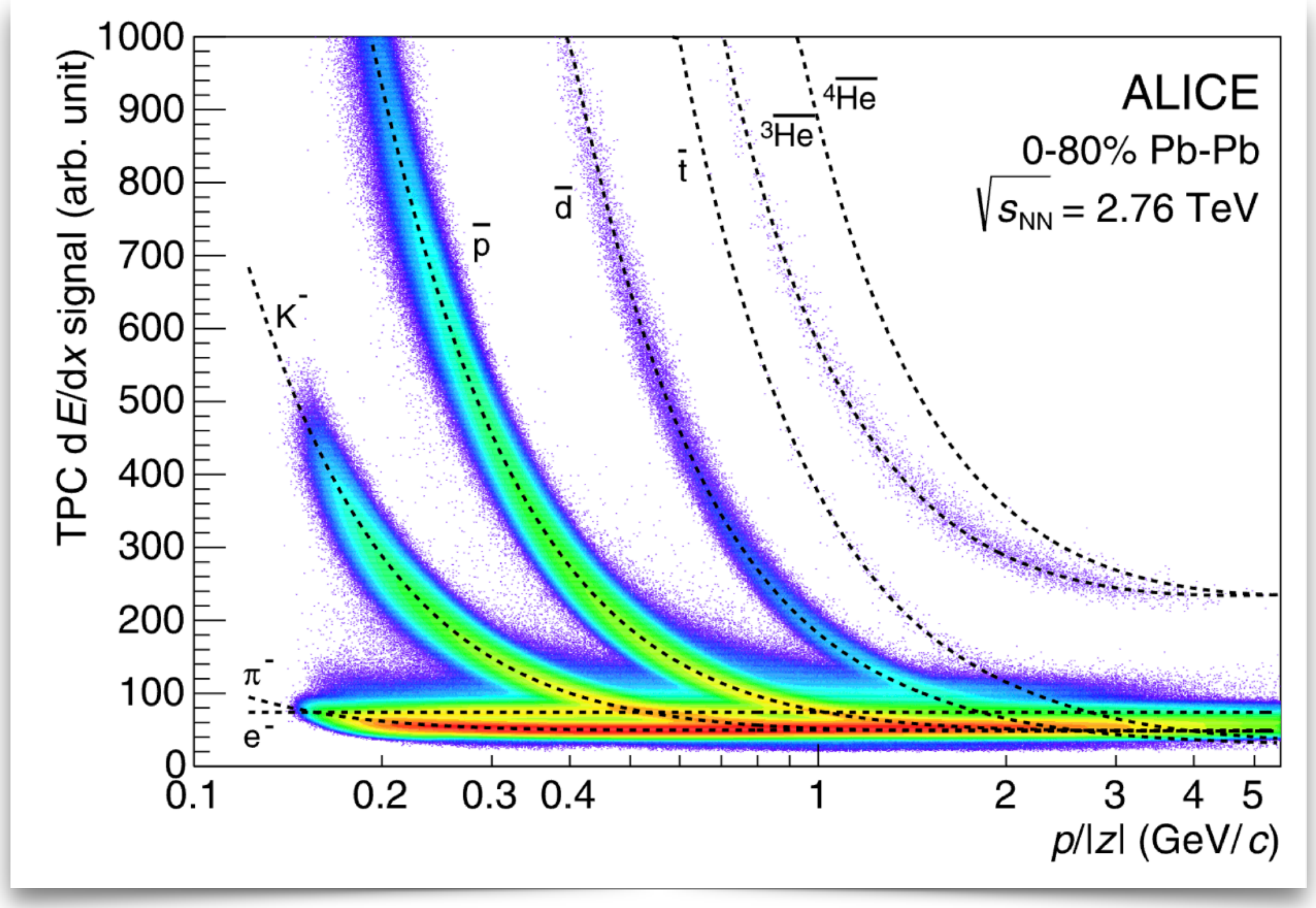
	ATLAS	CMS
Position	***Outside*** of magnet coil: 2–4 X_0 additional material in front of ECAL	***Inside*** magnet coil: Limited depth (HCAL only 7.2 λ_I at $\eta = 0$: additional 'tail catcher')
ECAL	Lead/liquid-argon (LAr) ***sampling calorimeter***: high ***granularity*** and longitudinal resolution	***Homogeneous*** crystal calorimeter ($PbWO_4$): high ***intrinsic energy resolution***
HCAL	***sampling calorimeters***: iron+scintillator (barrel) copper+LAr (endcap)	***sampling calorimeters***: brass+scintillator (barrel+endcap) iron+quartz fibres (forward)

Calorimeter	Term	ATLAS	CMS
ECAL (barrel)	stochastic (a)	$10\,\%\sqrt{\text{GeV}}$	$3\,\%\sqrt{\text{GeV}}$
	noise (b)	250 MeV	200 MeV
	constant (c)	0.2 %	0.5 %
ECAL+HCAL (barrel)	stochastic (a)	$55\,\%\sqrt{\text{GeV}}$	$70\,\%\sqrt{\text{GeV}}$
	noise (b)	3.2 GeV	1.0 GeV
	constant (c)	2.3 %	8 %

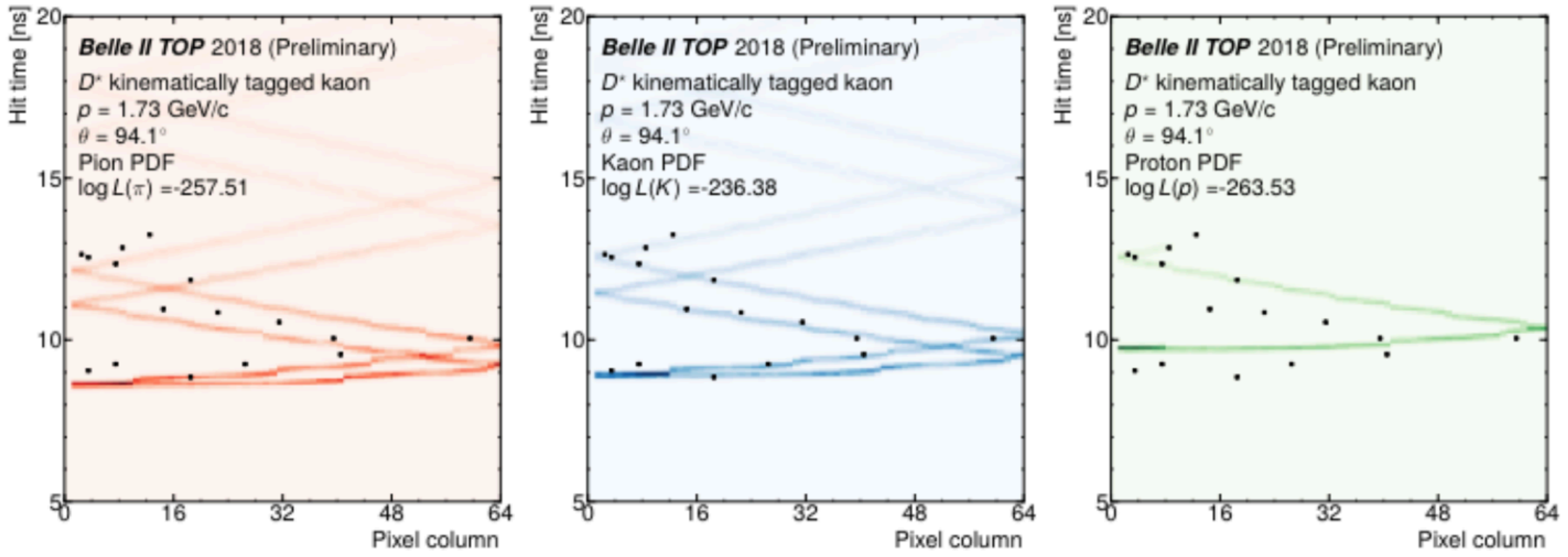
Source: Annu. Rev. Nucl. Part. Sci. 56 (2006) 375

Particle Identification: Charged Particles

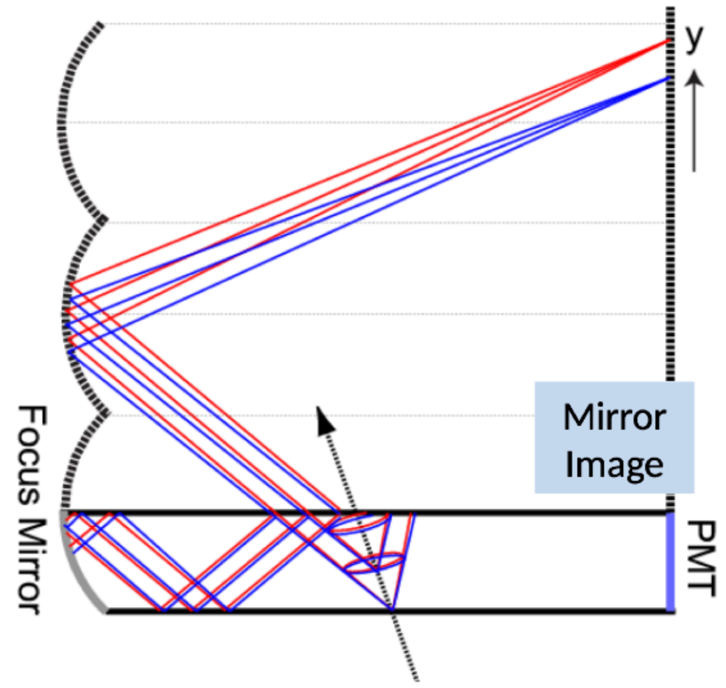
- In addition to momentum, energy and position measurements, detectors must identify particle species:
  - electrons: deposit most energy in the ECAL: E/p ≈ 1 (in reality machine learning methods exploiting shower shapes in the ECAL)
  - muons: cross the muon chambers with little multiple-scattering within the muon chambers
  - pions vs kaons: Cerenkov angle ((A)RICH detectors) or time-of-propagation (TOP)
  - protons: very high energy loss per unit distance ("high dE/dx")

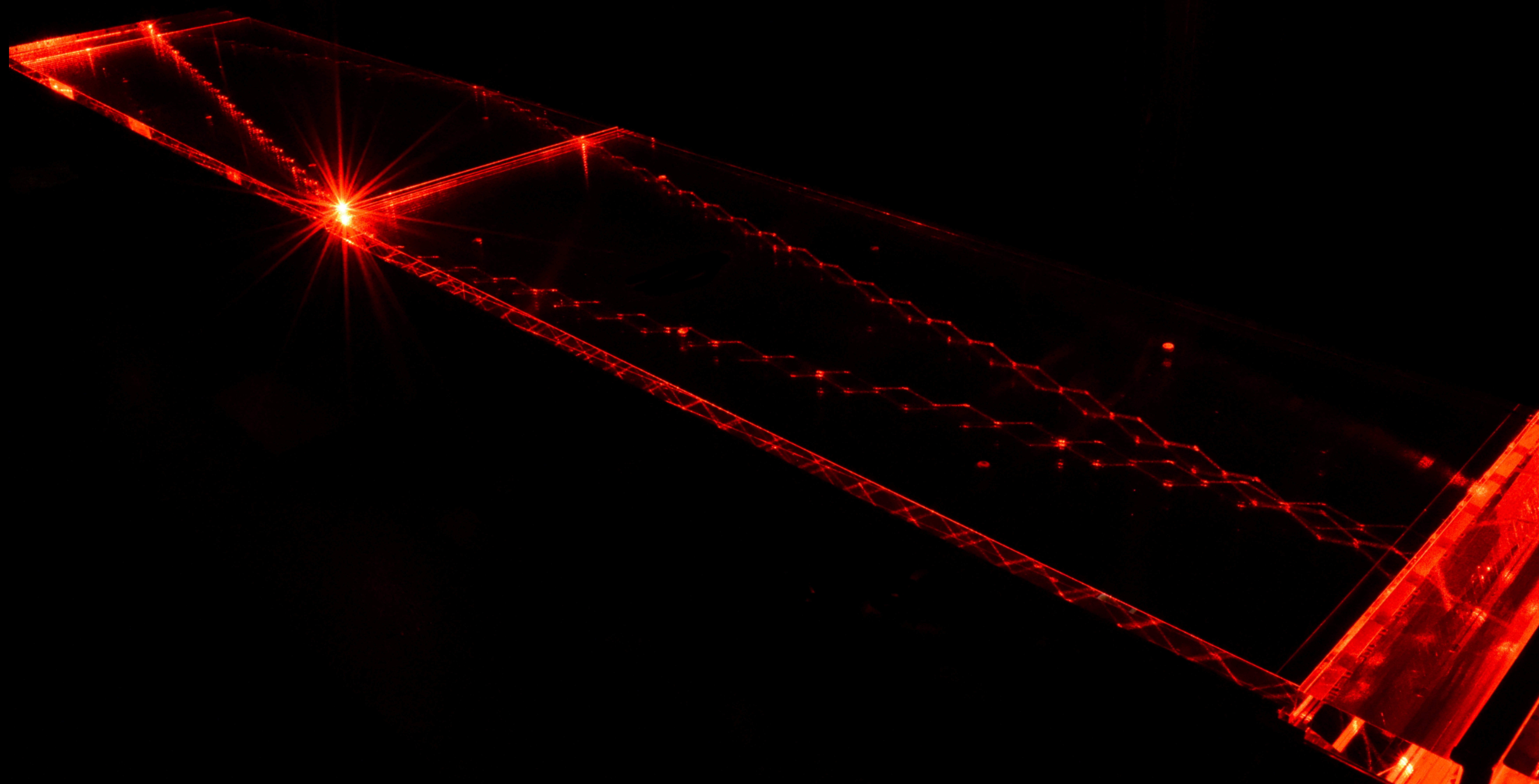


Source: Physics Letters B 752(C) (2015)



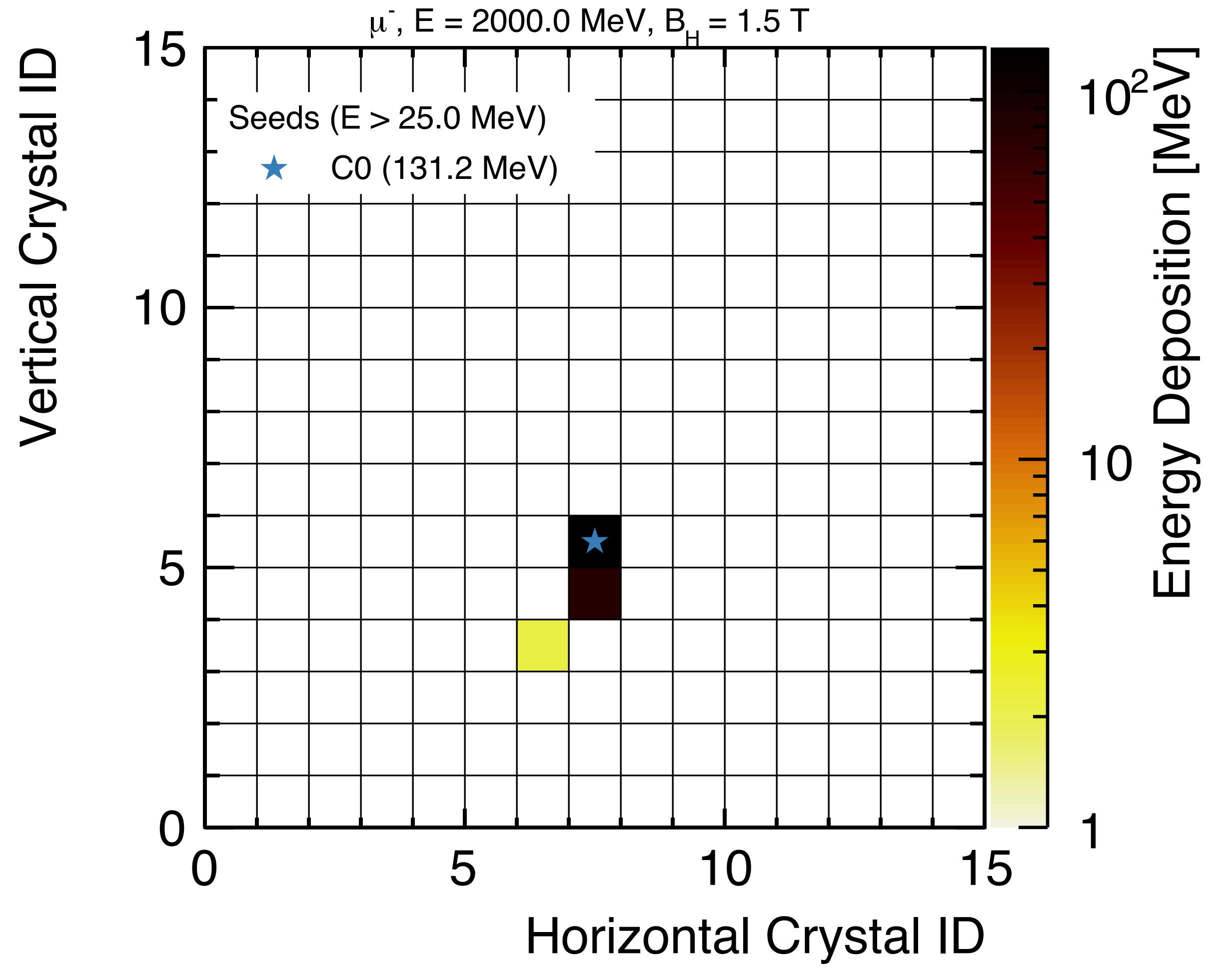
Source: Belle II TOP



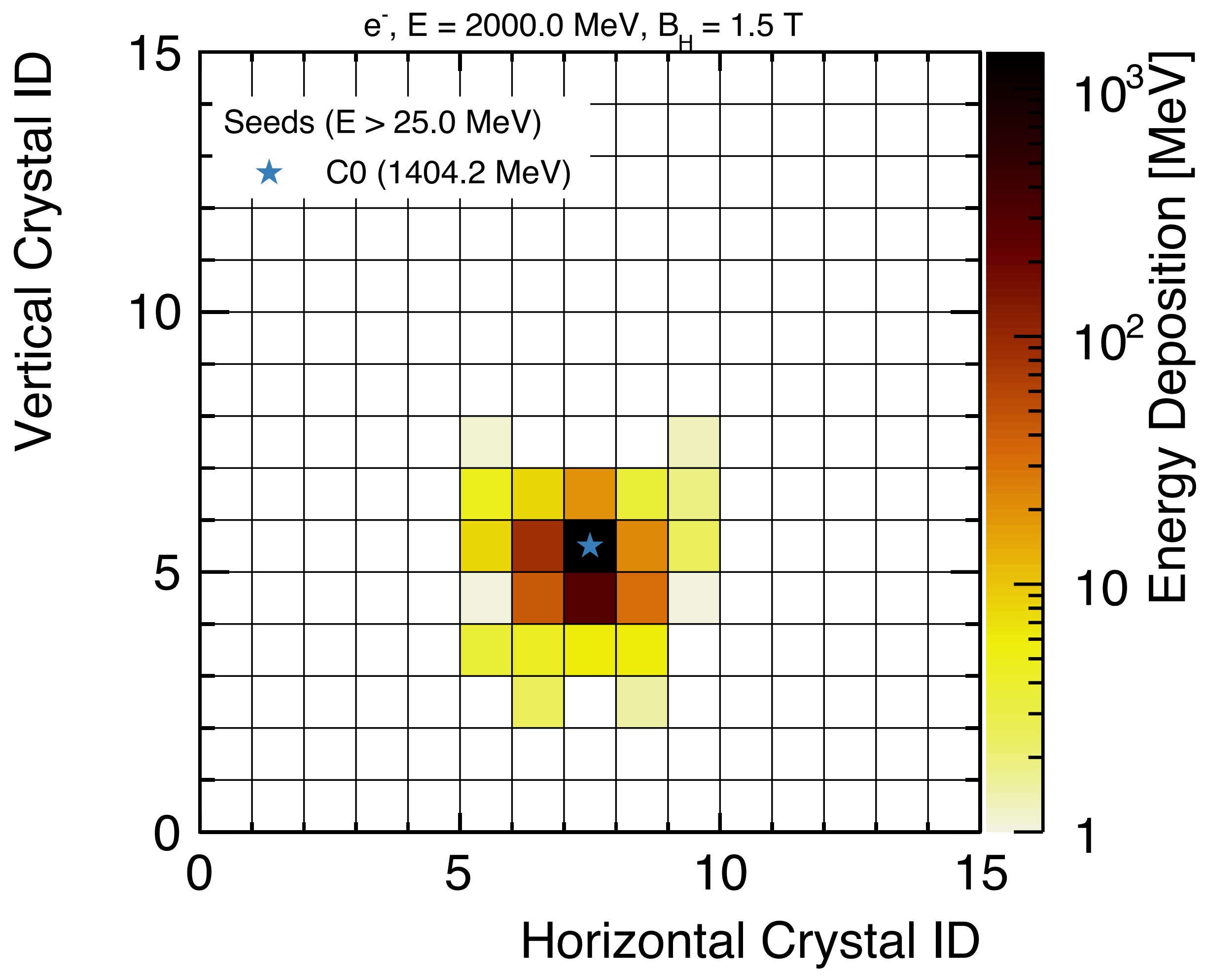


Source: Belle II TOP

Particle Identification: Neutral Particles

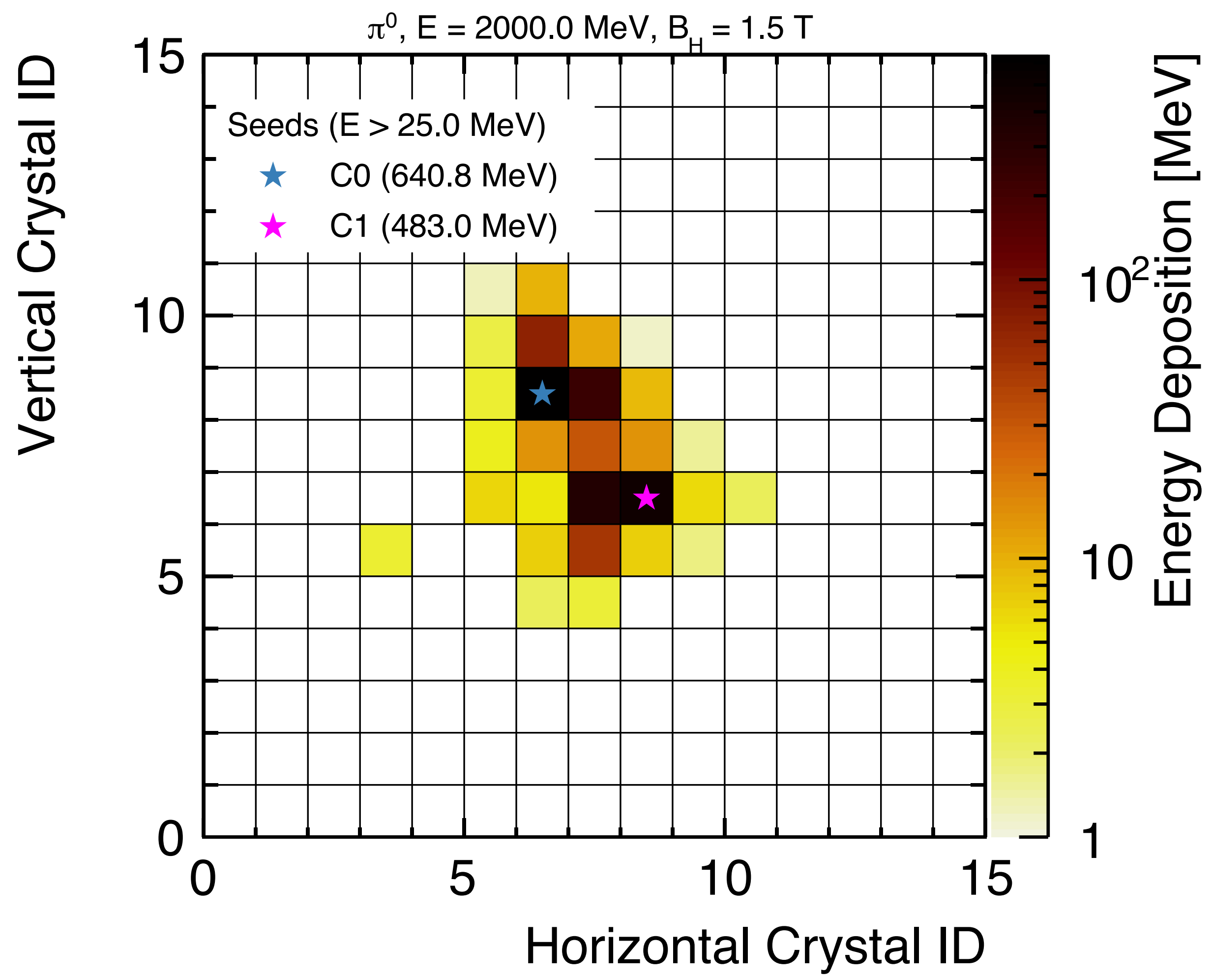


"MIPs" deposits a known amount of energy in a very localized area.

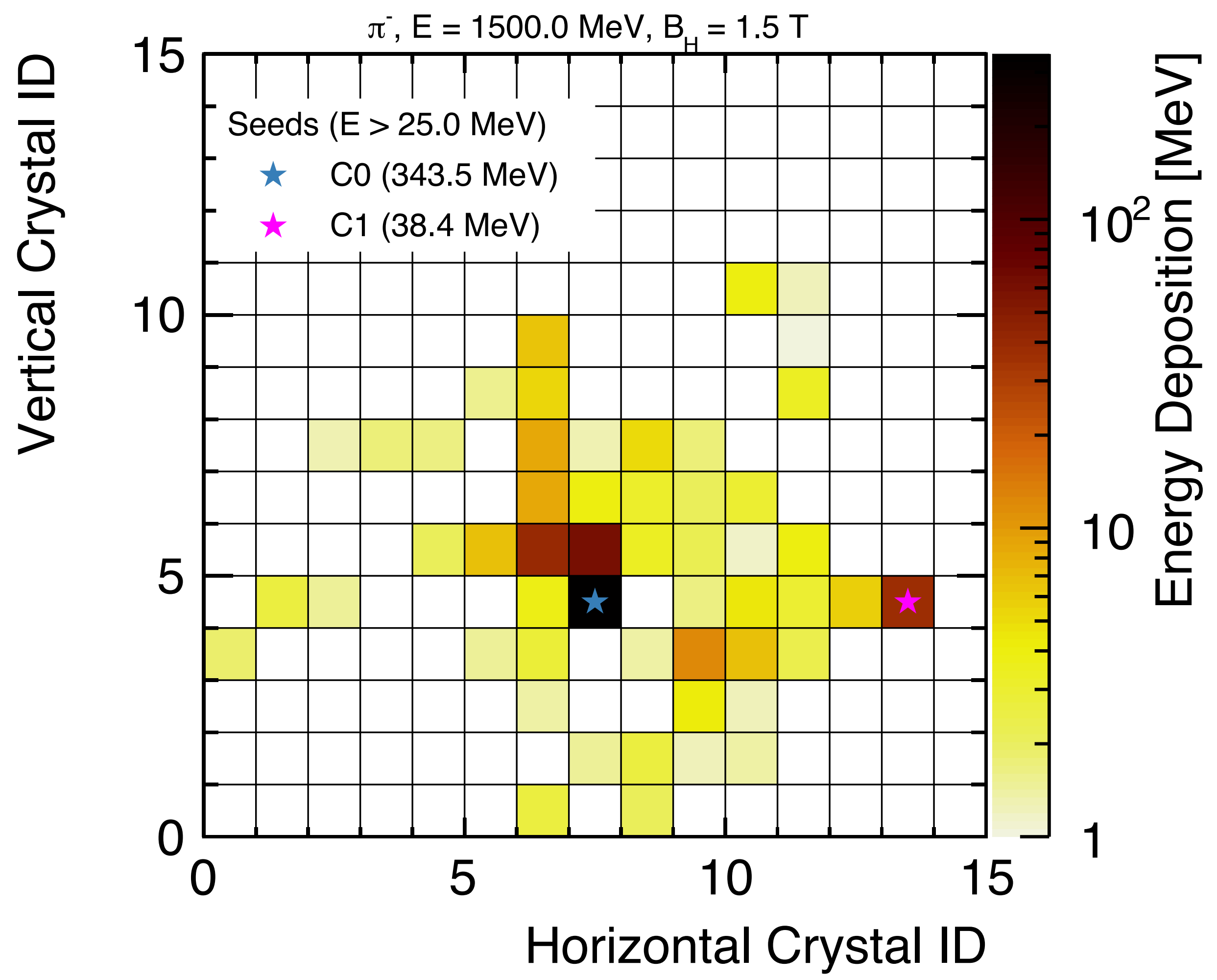


Electrons (and photons) create a well defined electromagnetic shower

π^0 decay into two photons that can overlap ("boosted objects")



Hadrons create distributed showers with large fluctuations and a lot of missing energy



Summary

- Tracking detectors are used to determine momentum and track origins
  - Gas detectors provide low multiple scattering and many track hit
  - Solid state detectors provide superior hit resolution and radiation hardness
- Calorimeters are used to measure energy depositions
  - Sampling calorimeters: absorber ≠ active material
  - Homogeneous calorimeters: absorber = active material
  - In typical detector: separate electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL)
- Particle identification detector separate particle species by exploiting different interactions of different particles (e, μ, π, K, p) and (γ, hadrons)

unnamed - looking for adoption!

(by Sarah Untereiner)

Simulation vs reality

compare with theory

Reality

real experiment

detector readout

theory calculation

event generator

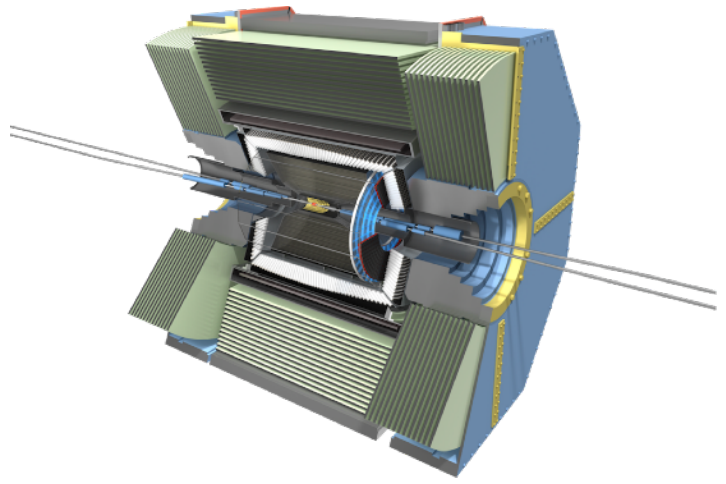
hadronization
(quarks → hadrons)

propagation through detector

simulated electronics
response

Simulation

software “twin”



digitized and calibrated
detector hits

reconstruction

object calibration

user analysis

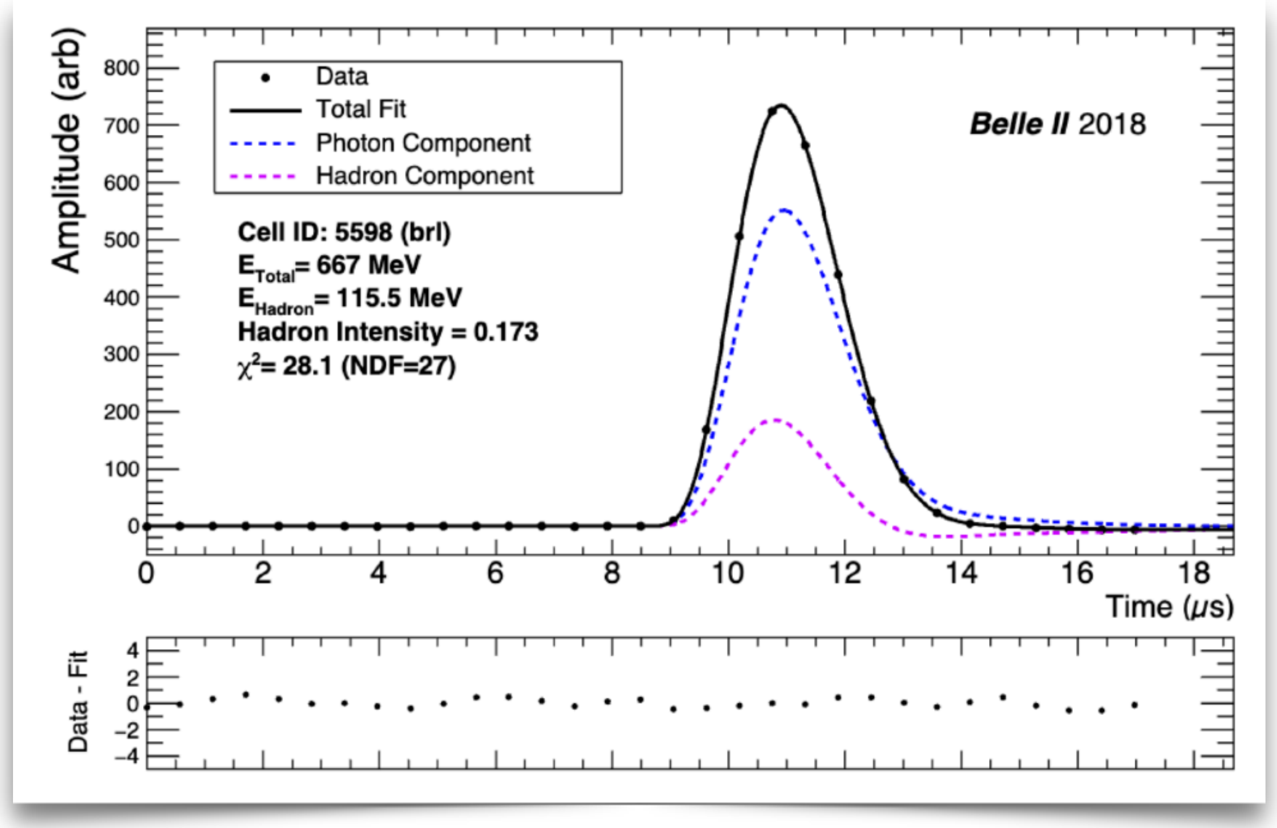
Digitized raw data

Reality

real experiment

detector readout

digitized and calibrated detector hits



Example: Belle II ECAL single crystal raw data
(one such entry for each readout crystal (~4000 per event, 30.000 events per second)

b[31..30] – quality flags
b[29..18] – peak time
b[17..0] – peak amplitude

b[17..0] – sample amplitude

stored for high energy crystals to allow hadron identification

```
 id   packet_size            header
0010.0024 011F.0152 0001.5EC7 0000.0001
 fit_result
CAB8.0157
               31 sample of ADC data
0000.0DD6 0000.0D86 0000.0D5D 0000.0D69
0000.0D9E 0000.0DEF 0000.0E1D 0000.0E12
0000.0DC1 0000.0D60 0000.0D24 0000.0D4C
0000.0DA9 0000.0E00 0000.0E33 0000.0E31
0000.0DF9 0000.0DA2 0000.0D83 0000.0DD0
0000.0E5E 0000.0EC9 0000.0ED0 0000.0E6E
0000.0DCE 0000.0D3F 0000.0D17 0000.0D62
0000.0DDD 0000.0E37 0000.0E44
```

Time to publication is critical

- Scientific results should be published timely after data taking
  - Results will inform planning for future projects (e.g. “How heavy is the lightest SUSY particle?”)
  - Results are needed as input for other results (e.g. precision branching fraction measurements)
  - External competition (same results can not be published twice)
  - No results (negative or positive!) → no funding
- Turnaround time for data analysis is key and a limiting factor in high energy physics: Analysing (reading, processing, …) petabytes* of data is non-trivial and can take years

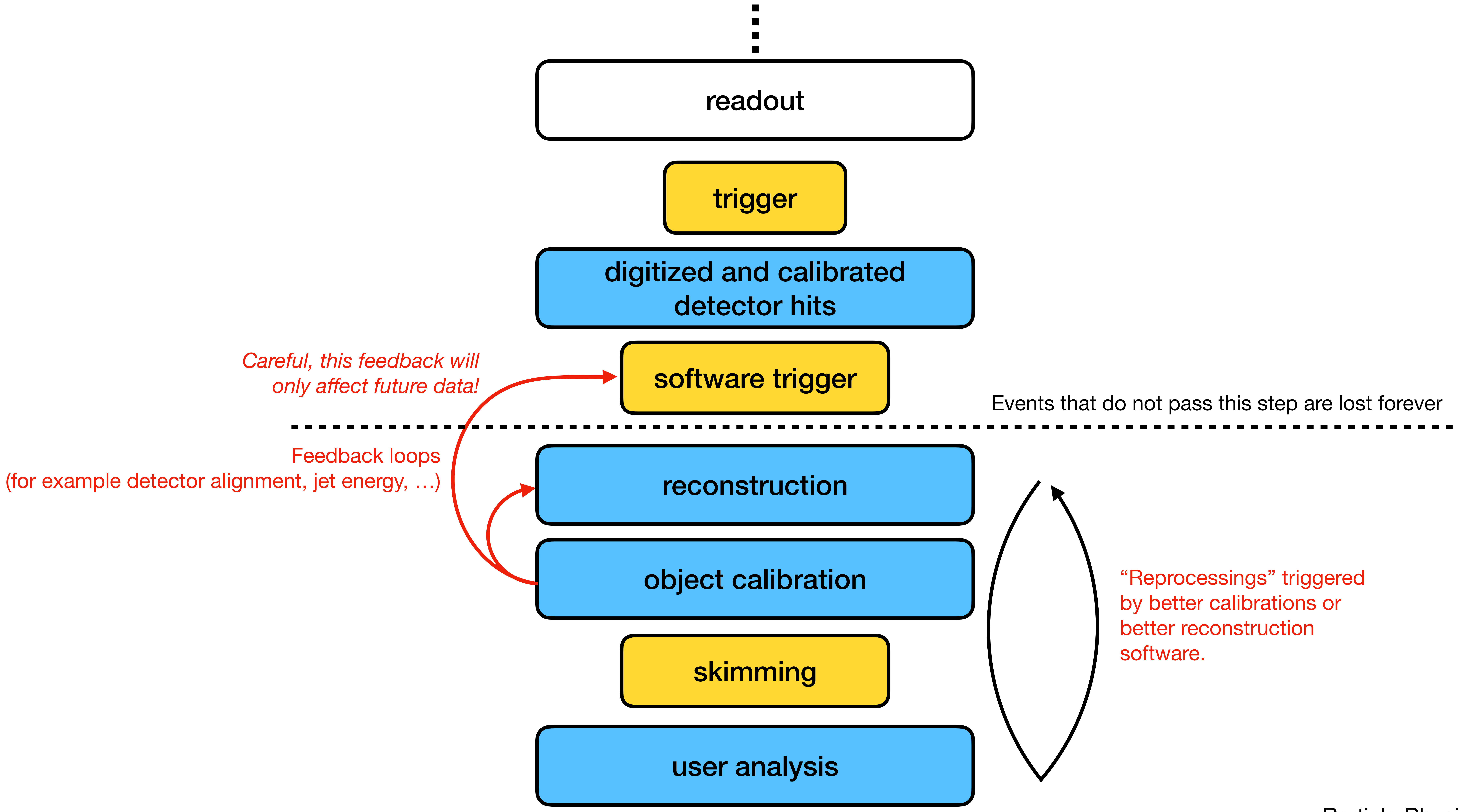
*LHC dataset in 2025 will approach 1 Exabyte.

Analysis workflow*

- Fast multi-stage **triggers** decide which events are recorded
- **Reconstruction** algorithms convert detector hit into physics objects (e.g. “tracks” or “clusters”)
- **Object calibration** algorithms determine efficiency, fake rates, resolutions, systematic uncertainties, …
- **Skimming** algorithms perform high efficiency, low purity preselections
- **Analysis** algorithms optimise signal vs background for a specific analysis and extract results
- **Systematic uncertainties** are needed for final results

* Actual names of the steps differ slightly between experiments

Data reduction



Data reduction

* Actual names of the data formats differ slightly between experiments

example: Belle II
several MHz

readout

30 kHz output rate
typically FPGAs or ASICS
reconstruction time <2μs

trigger

→ rather new developments: turbo streams with extremely high data rates but no digit-level information

digitized and calibrated detector hits

→ DST (digital summary table, typically ROOT)

10 kHz output rate
typically CPUs or GPUs
reconstruction time <1s

software trigger

---

typically CPUs
reconstruction time <2s

reconstruction

→ mDST (mini-DST)*
(data contains only reconstructed objects, typically ROOT)

object calibration

physics analysis dependent preselections that keep a few percent of the data without significant loss in efficiency, e.g. “tau pair skim”

skimming

→ μDST (micro-DST)*
(data contains physics objects, typically ROOT)

user analysis

→ ntuples
(custom data formats, today’s standard are so-called flat arrays HDF5 or ROOT)

Analysis workflow: who does what?

- **Triggers**

Only performed once in real-time. Events that are not kept are lost forever.

- **Reconstruction**
- **Object calibration**
- **Skimming**

Performed centrally by experiment's experts on massive pre-reserved computing resources. Depending on the computing model and data size, so-called "reprocessings" happen about once per year.

- **Analysis**
- **Analysis systematic uncertainties**

Performed locally by analysts (PhD students, post docs, …) using batch-clusters, HPCs or the Worldwide LHC Computing Grid (WLCG).

Analysis workflow: Simulation usage

- **Triggers**

  Development of new algorithms

- **Reconstruction**

  Development of new algorithms, training of multivariate classifiers or regressors

- **Object calibration**

  Determination of acceptance and kinematics inputs

- **Skimming**

  Development of new selections

- **Analysis**

  Development of event selections, background shapes, signal efficiencies, …

- **Analysis systematic uncertainties**

  Find sources of systematics uncertainties, for example by varying B-fields in simulations and study the effects

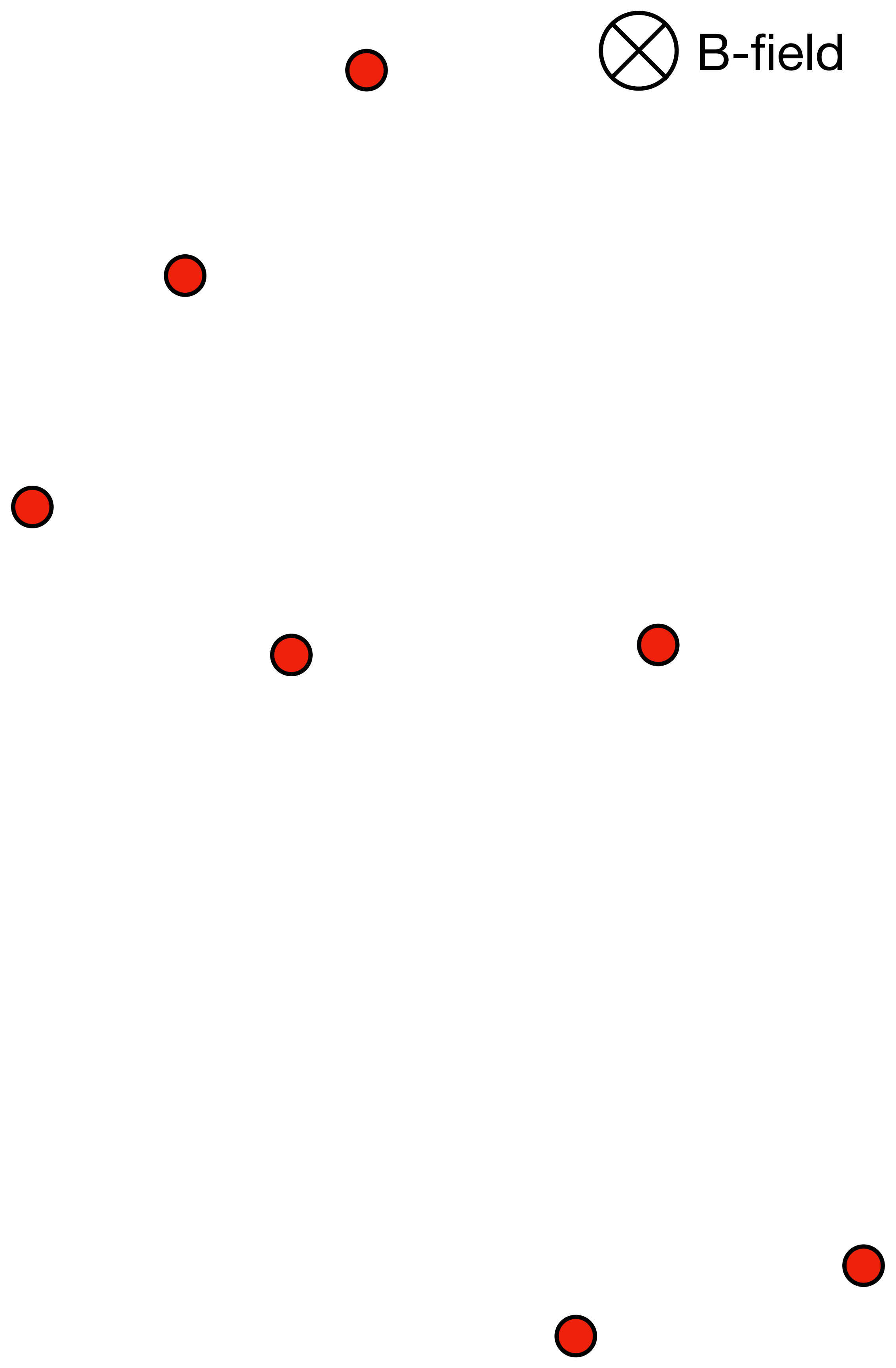
Since simulation is used in so many places, the differences between simulation and reality is often a critical aspect of each analysis

Digits to objects

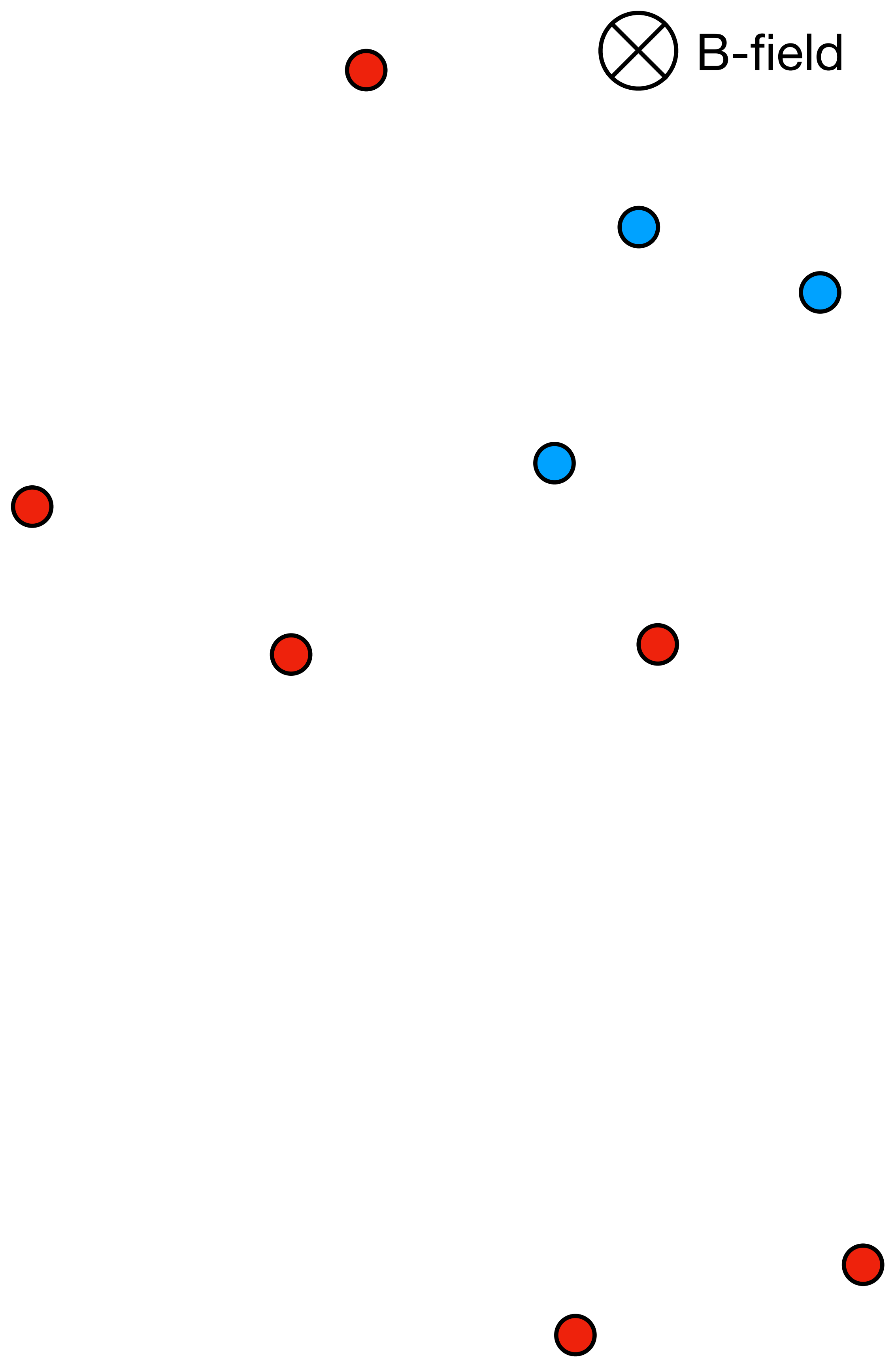
- Detector readout **digits** describe individual measurements:
  Pixel detector hits, drift-times, calorimeter cell energies, …
- Reconstruction algorithms are used to combine digits into objects (tracks, clusters, V0s) and to assign identification likelihoods to them (electron, hadron, photon, neutron, …)
- Different philosophies in CMS and Belle II (to cope with much larger data size in CMS):
  - Belle II provides multiple hypotheses with probabilities for each object, final user decides using additional constraints (track fit results are different for different particle masses, calorimeter clustering is different for photons and hadrons)
  - CMS provides best hypothesis only
- Some special analyses (magnetic monopoles, millicharged particles, long-lived particles, …) may require special reconstruction

Typical reconstruction task: Track finding

Ideal

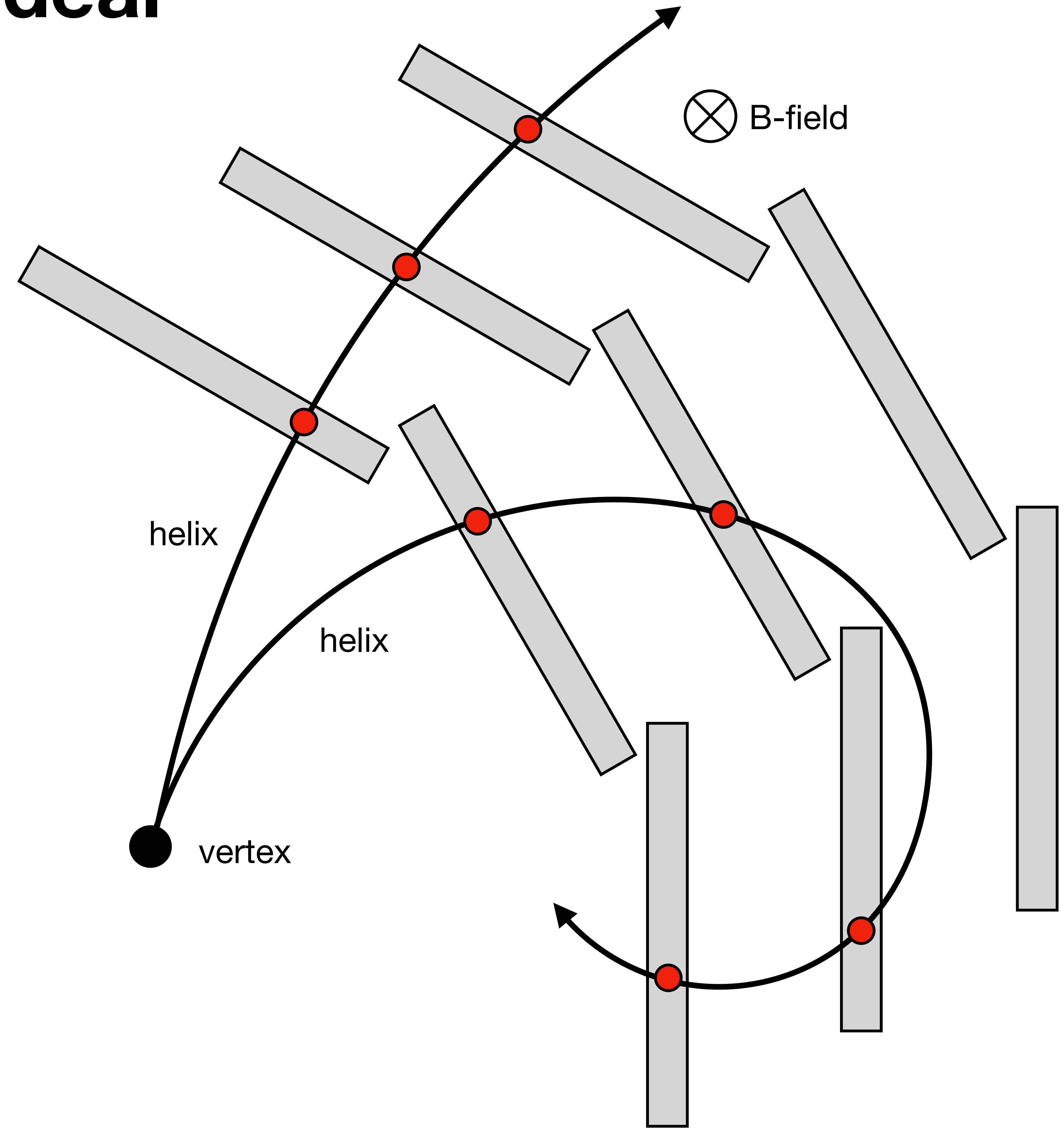


Realistic

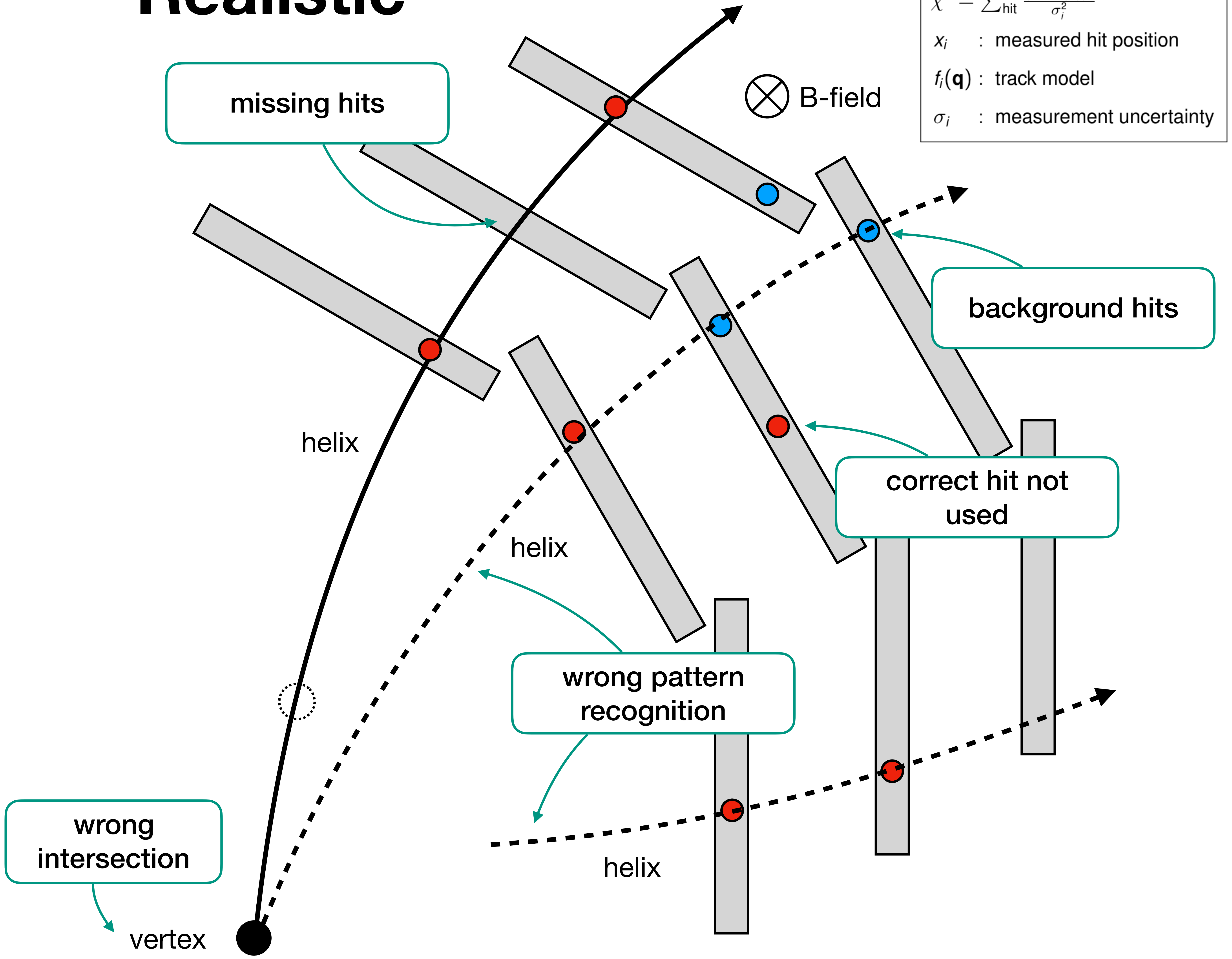


Typical reconstruction task: Track finding

Ideal



Realistic



Track fitting: minimise

$$\chi^2 = \sum_{\text{hit}} \frac{(x_i - f_i(\mathbf{q}))^2}{\sigma_i^2}$$

x_i : measured hit position

$f_i(\mathbf{q})$: track model

σ_i : measurement uncertainty

Reconstruction: Particle identification

- Typically using multiple subdetectors **d** (energy loss in trackers, energy deposition in ECAL, track propagation in muon system, …)

- Information is combined in likelihoods for each particle hypothesis **i**

$L = \prod_{d}^{d \epsilon D} L^d(\mathbf{x} \,|\, i)$ for a given set of observables **x**.

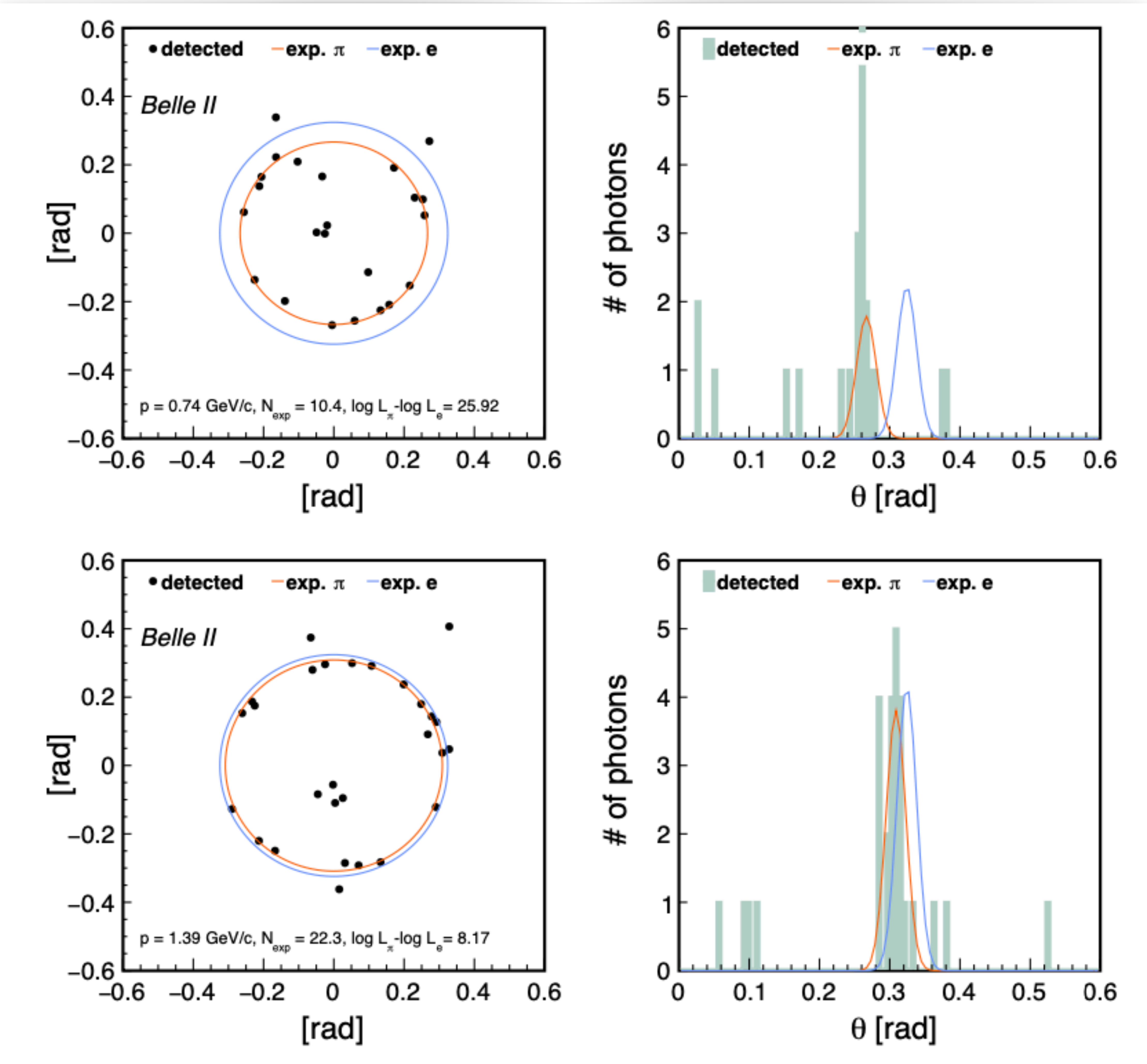
- **Global likelihood** using Bayes'-theorem and a total probability of one for $A_i = \{e, \mu, \dots\}$

$$P(A_i \,|\, \mathbf{x}) = \frac{P(\mathbf{x} \,|\, A_i) P(A_i)}{\sum_j P(\mathbf{x} \,|\, A_j) P(A_j)} = \frac{L_i}{\sum_j L_j}$$

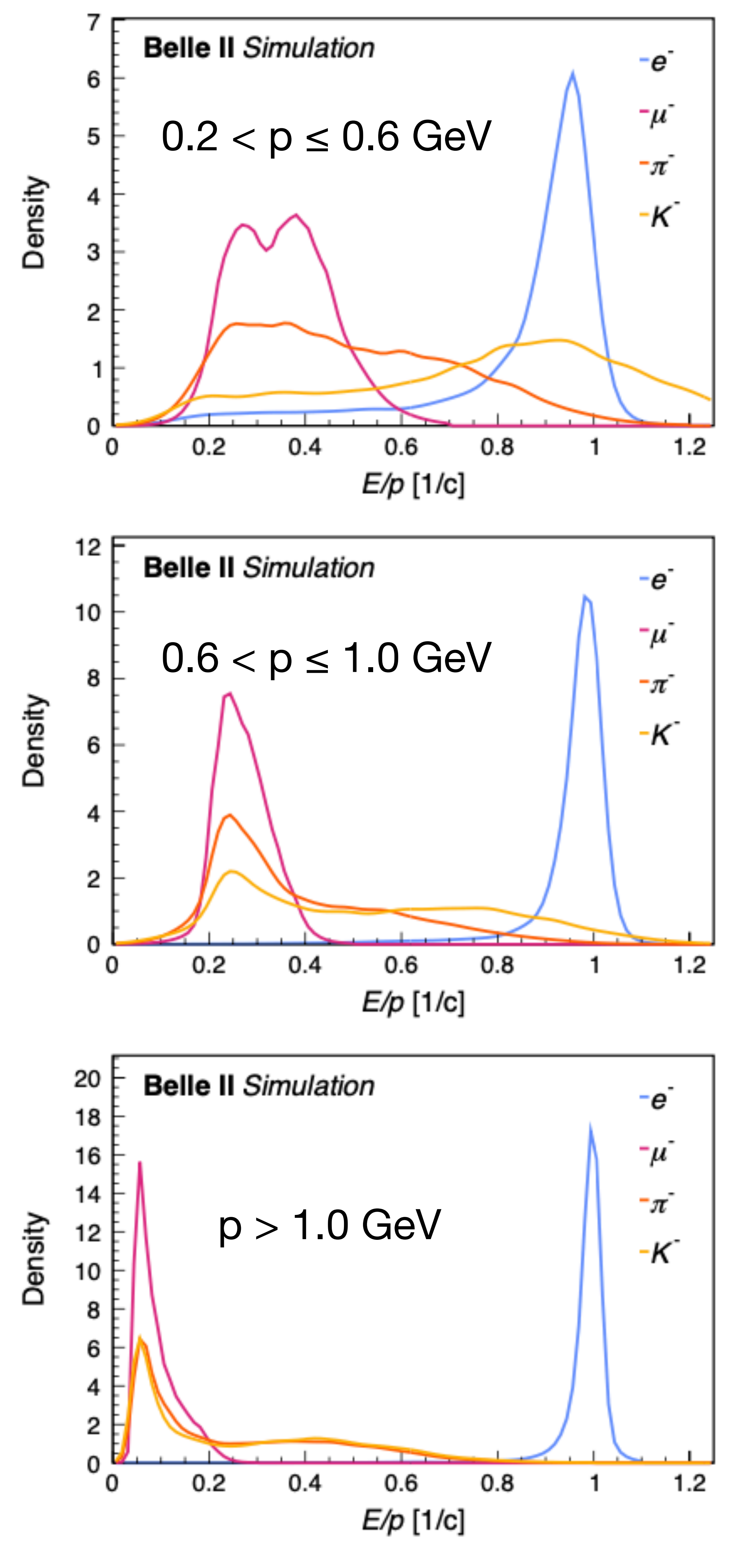
- Depending on the analysis, a **binary likelihood** $P(i/j \,|\, \mathbf{x}) = \dfrac{L_i}{L_i + L_j}$ can be sufficient

Reconstruction: Particle identification examples

Example: Belle II ARICH electron vs pion

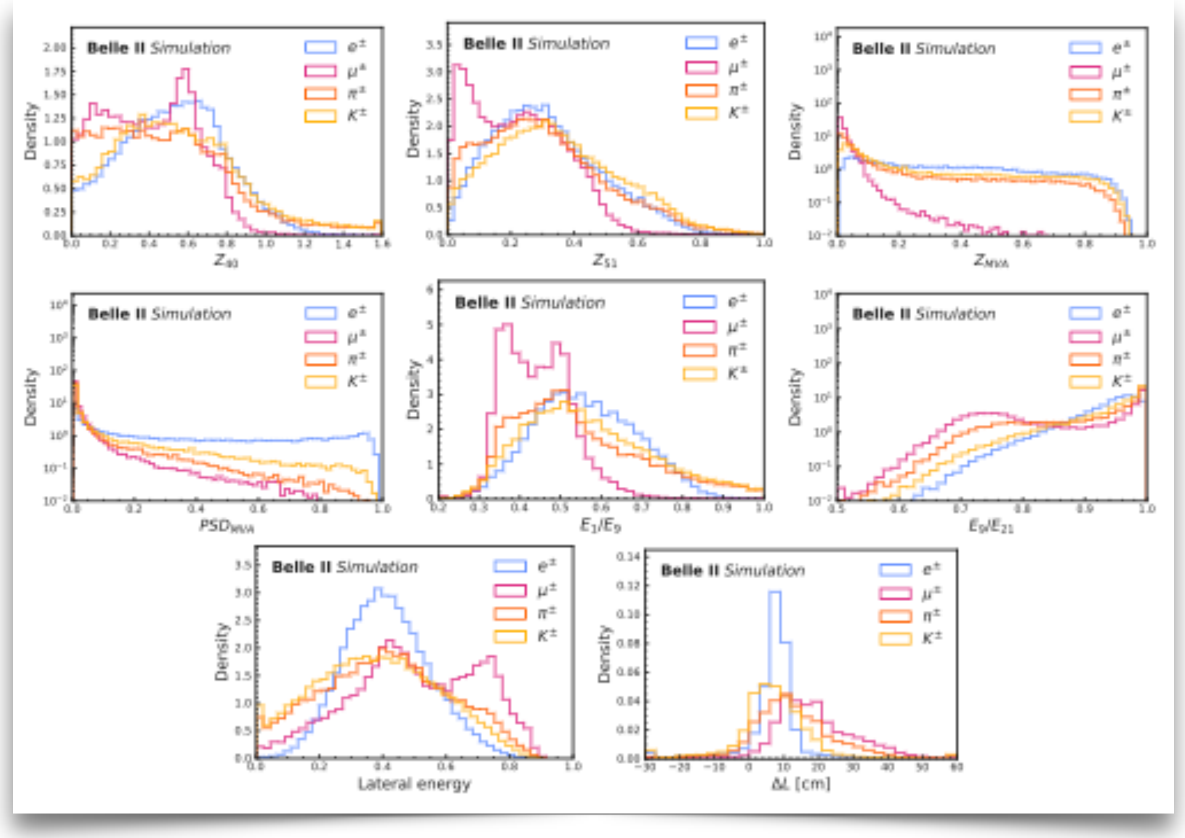


https://github.com/belle2/basf2/blob/main/arich/modules/arichReconstruction/src/ARICHReconstruction.cc



Example: ECAL

0.2 < p ≤ 0.6 GeV

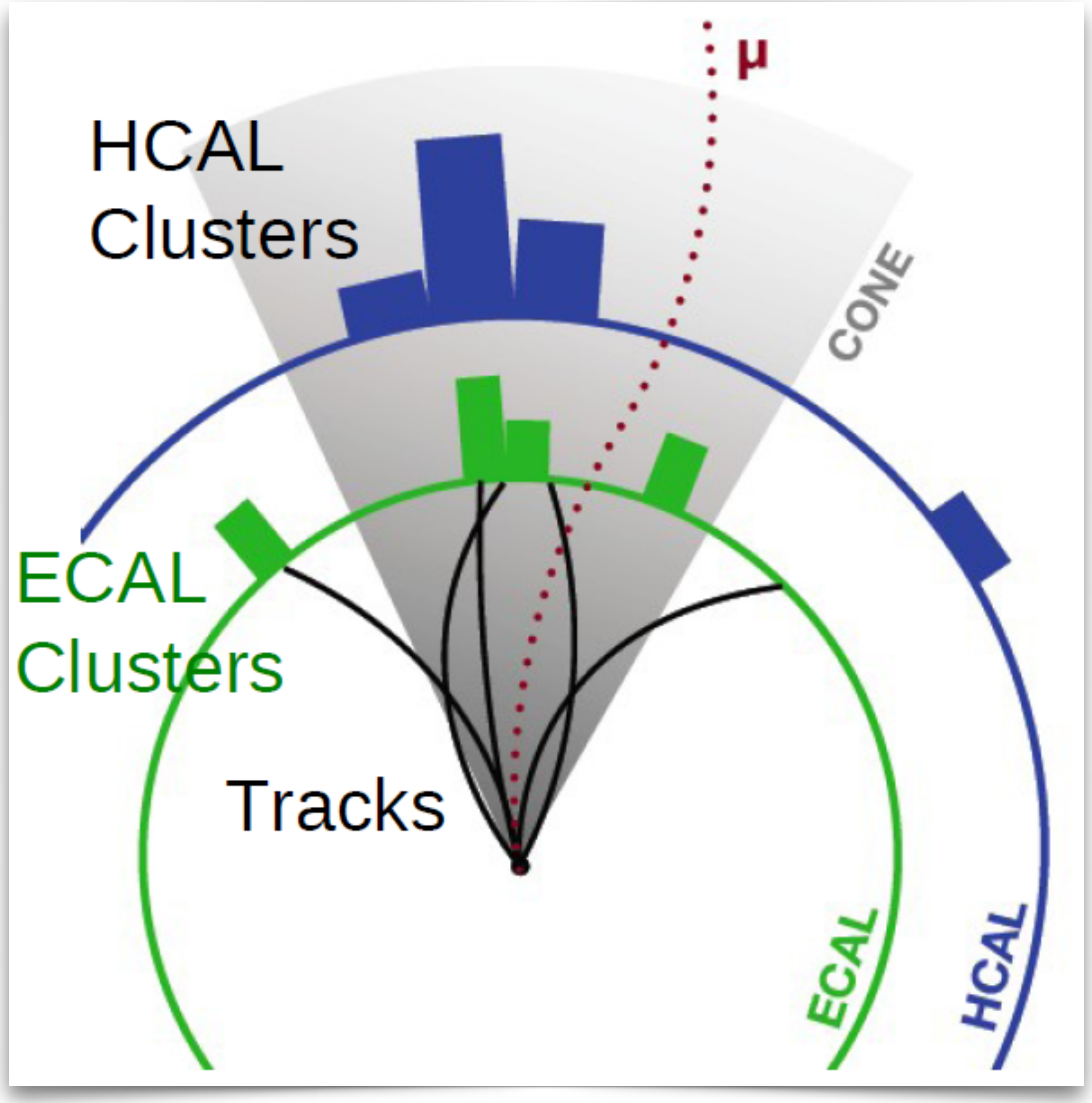


for low momentum electrons, a BDT-based classifier improves the PID

https://github.com/belle2/basf2/blob/main/ecl/modules/eclChargedPID/src/ECLChargedPIDModule.cc

Particle Flow

Credit: F. Beaudette at ICHEP2010



- Traditional jet reconstruction (later) was done using calorimeters only:
  - $\rightarrow E_{jet} = E_{ECAL} + E_{HCAL}$
  - poor energy resolution because 70% of jet energy in HCAL
- Particle flow uses each subdetector optimally:
  - charged tracks in tracker, "matched" to calorimeter clusters
  - muon momentum in muon system
  - photon energy in ECAL
  - neutral hadron energy in HCAL (10% of jet energy)

  $\rightarrow E_{jet} = E_{Tracks} + E_{\gamma} + E_{neutral\,hadrons}$

  - Requires high granularity (tracker and calorimeter) and strong B-field

Event selection

Read input file and load lists of reconstruction objects (with particle hypothesis!)

```python
#!/usr/bin/env python3

##########################################################################
# basf2 (Belle II Analysis Software Framework)                           #
# Author: The Belle II Collaboration                                     #
#                                                                        #
# See git log for contributors and copyright holders.                    #
# This file is licensed under LGPL-3.0, see LICENSE.md.                  #
##########################################################################

##########################################################################
#                                                                        #
# Stuck? Ask for help at questions.belle2.org                            #
#                                                                        #
# This tutorial demonstrates how to reconstruct the                      #
# following decay chain (and c.c. decay chain):                          #
#                                                                        #
# D*+ -> D0 pi+                                                          #
#        |                                                               #
#        +-> K- pi+                                                      #
#                                                                        #
##########################################################################

import basf2 as b2
import modularAnalysis as ma
import variables.collections as vc
import variables.utils as vu
import stdCharged as stdc

# create path
my_path = b2.create_path()

# load input ROOT file
ma.inputMdst(filename=b2.find_file('Dst2D0pi.root', 'examples', False),
             path=my_path)


# use standard final state particle lists
#
# creates "pi+:all" ParticleList (and c.c.)
stdc.stdPi(listtype='all', path=my_path)
# creates "pi+:loose" ParticleList (and c.c.)
stdc.stdPi(listtype='loose', path=my_path)
# creates "K+:loose" ParticleList (and c.c.)
stdc.stdK(listtype='loose', path=my_path)
```

Reconstruct particles out of these charged particles

```python
# reconstruct D0 -> K- pi+ decay
# keep only candidates with 1.8 < M(Kpi) < 1.9 GeV
ma.reconstructDecay(decayString='D0:kpi -> K-:loose pi+:loose', cut='1.8 < M < 1.9', path=my_path)

# reconstruct D*+ -> D0 pi+ decay
# keep only candidates with Q = M(D0pi) - M(D0) - M(pi) < 20 MeV
ma.reconstructDecay(decayString='D*+ -> D0:kpi pi+:all', cut='0.0 < Q < 0.2', path=my_path)
```

If we run on simulated data, find the corresponding MC particles

```python
# perform MC matching (MC truth association)
ma.matchMCTruth(list_name='D*+', path=my_path)
```

Store information to flat arrays (here: ROOT format)

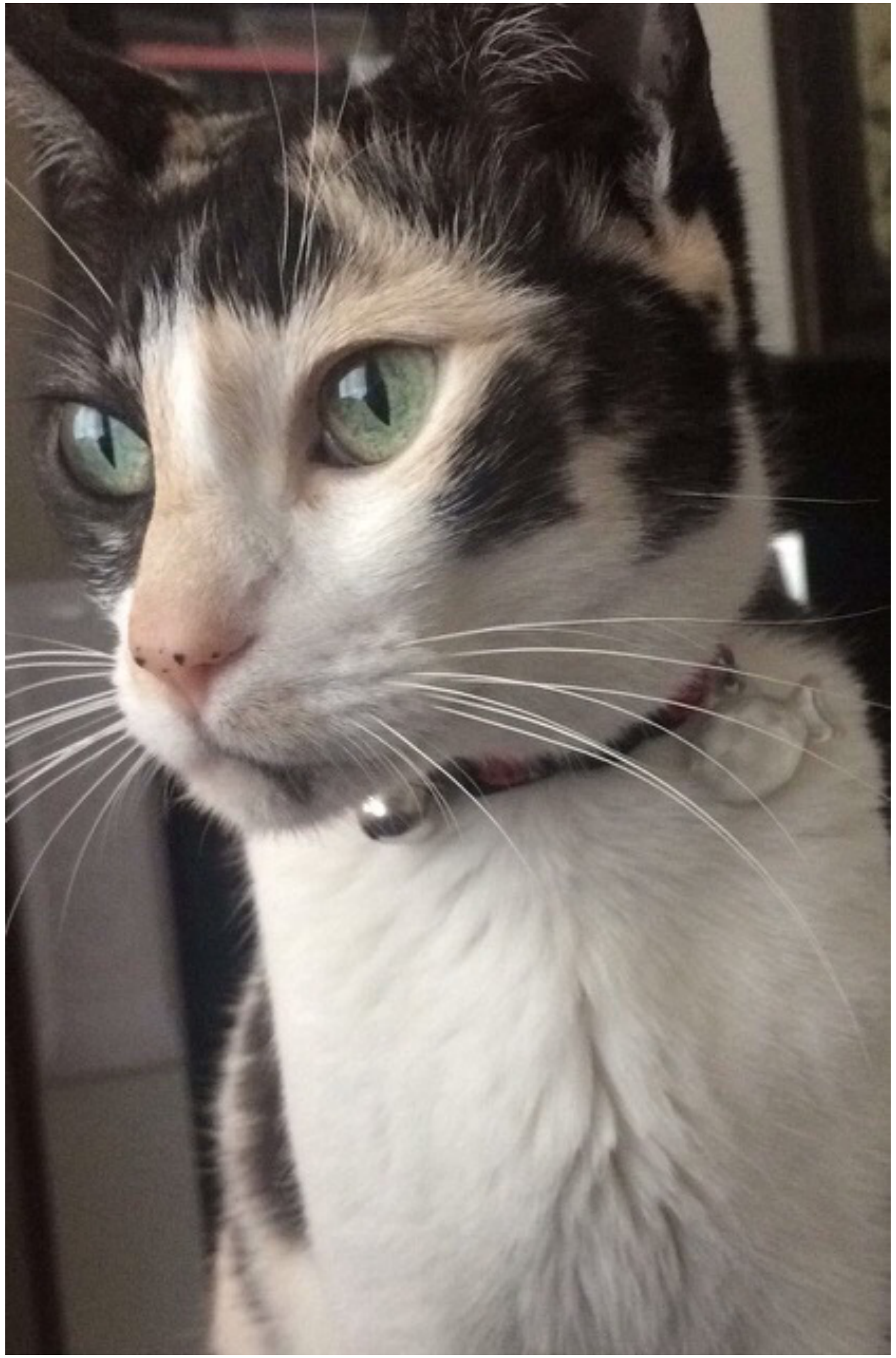
```python
fs_hadron_vars = vu.create_aliases_for_selected(
    list_of_variables=vc.pid + vc.track + vc.mc_truth,
    decay_string='D*+ -> [D0 -> ^K- ^pi+] ^pi+')

d0_vars = vu.create_aliases_for_selected(
    list_of_variables=vc.inv_mass + vc.mc_truth,
    decay_string='D*+ -> ^D0 pi+',
    prefix='D0')


# Saving variables to ntuple
output_file = 'B2A301-Dstar2D0Pi-Reconstruction.root'
ma.variablesToNtuple('D*+', dstar_vars + d0_vars + fs_hadron_vars,
                     filename=output_file, treename='dsttree', path=my_path)

# Process the events
b2.process(my_path)
```

https://github.com/belle2/basf2/blob/main/analysis/examples/tutorials/B2A301-Dstar2D0Pi-Reconstruction.py



“Sandy” (by Sarah Alshamaily)

Simulation vs reality

compare with theory

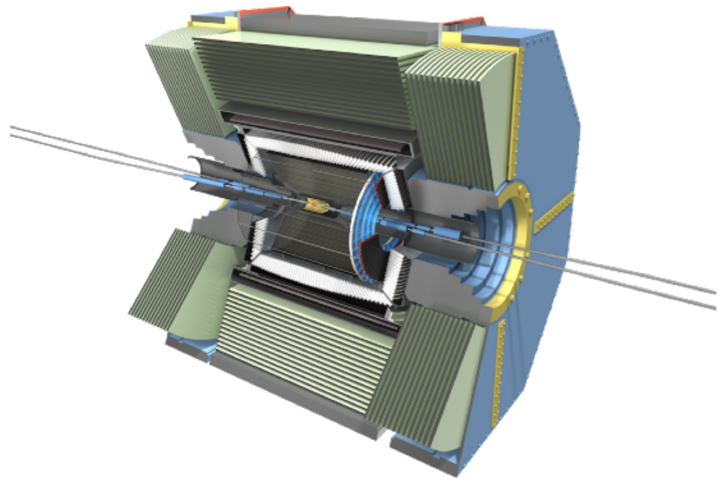
Reality

real experiment

detector readout

Simulation

software “twin”



theory calculation

event generator

hadronization
(quarks → hadrons)

propagation through detector

simulated electronics
response

digitized and calibrated
detector hits

reconstruction

object calibration

user analysis

How to get σ?

$$\sigma_{exp} = \frac{N_{\text{signal}} - N_{\text{background}}}{\epsilon \int \mathscr{L}}$$

Calibration: Luminosity

- Recall definition of luminosity: $\frac{dN}{dt} = \mathscr{L}\sigma \rightarrow \int \mathscr{L} = \frac{N_{exp}}{\sigma_{theory}}$

- Measured experimentally by using a physics process with very precisely known cross section σ that can be counted (dN/dt) with small statistical and systematic uncertainty

- At e^+e^- colliders a combination of $e^+e^- \rightarrow \gamma\gamma$ and $e^+e^- \rightarrow e^+e^-$ (Bhabha scattering) is used, typically selected using calorimeter information only to reduce systematic uncertainties

- At e^+e^- colliders integrated luminosity typically known to <0.5% (offline) and <2% (almost real-time every second)

How to get σ?

$$\sigma_{exp} = \frac{N_{\text{signal}} - N_{\text{background}}}{\epsilon \int \mathscr{L}}$$

Efficiency

- Very general:
  - Measurements of (differential) cross sections need shape and absolute value
  - Measurements of mass or lifetimes usually only need efficiency shape
  - Measurements of ratios (e.g. charge asymmetry) usually do not need efficiencies
    - However, some extreme precision measurements, e.g. input for g-2 are dominated by systematic uncertainties from efficiency differences in ratios

Efficiency and purity

		Estimated state P	Estimated state N
True state	P	TP (true positive) "hit"	FN (false negative) type 2 error "miss"
True state	N	FP (false positive) type 1 error "false alarm"	TN (true negative)

- Definition efficiency (or sensitivity):
  - TP / (TP + FN)
  - "Number of all correctly reconstructed particles out of all real particles" → ideally 100%
  - Trivial solution for high efficiency: Very loose selection
- Definition purity (or precision):
  - TP / (TP + FP)
  - "Number of all correctly reconstructed particles out of all reconstructed particles" → ideally 100%
  - Trivial selection for very high purity: Very tight selection

Efficiency determination

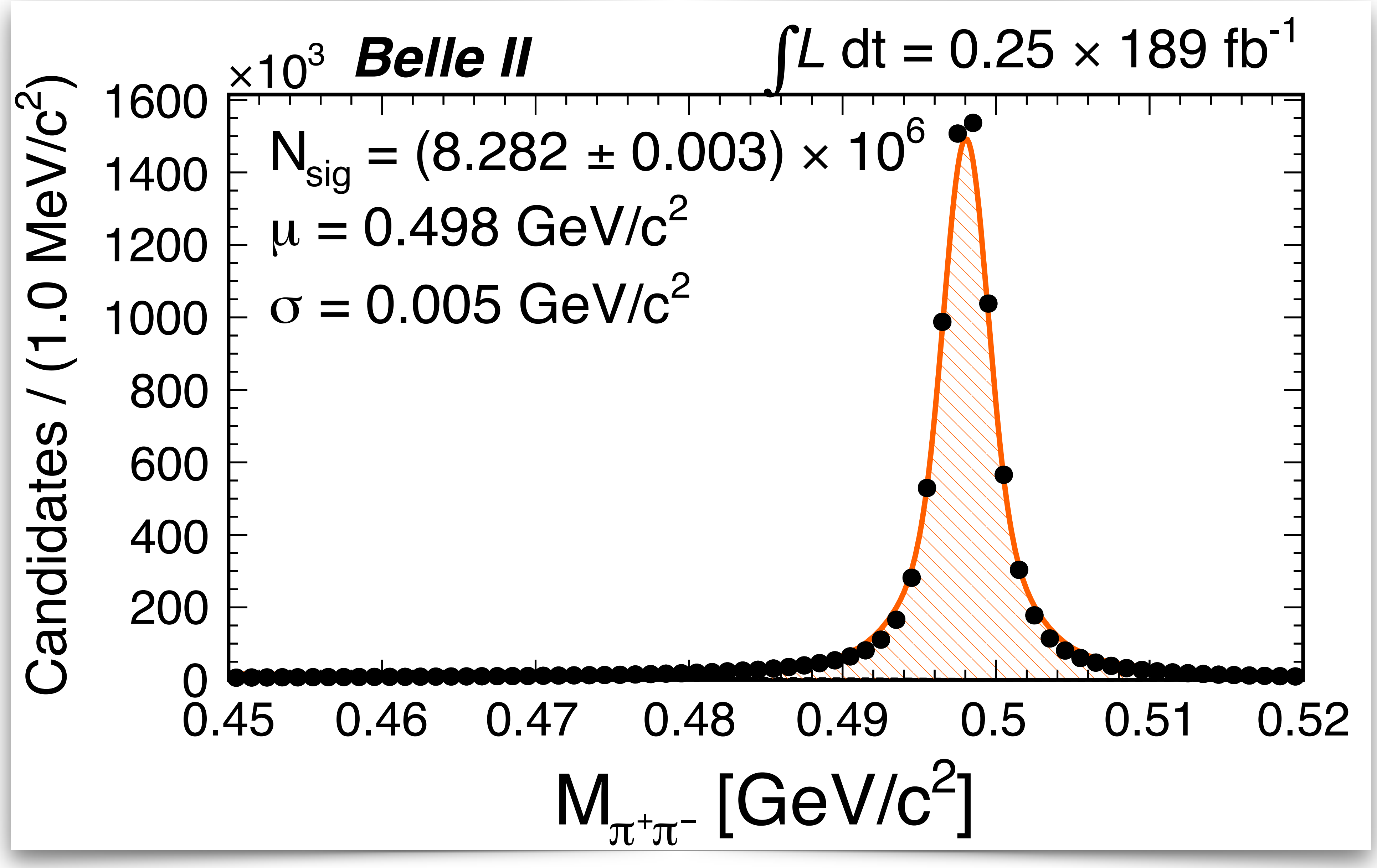
- Efficiency can be determined for signal but also for different backgrounds (sometimes called “fake rate”)

- Determine the efficiency **from simulation** (the correct particle is known here, often called “MC truth”)
  - if the efficiency is high, it is often easier to determine the correction to simulation, i.e. the efficiency difference between simulation and data

- Gold standard: **Data-driven techniques**
  - Tag&Probe: Use well-known particles decaying into two particles (e.g. $Z^0 \rightarrow \mu^+\mu^-$ or $K_S^0 \rightarrow \pi^+\pi^-$)
  - Orthogonal selection: Use two different selection variables A and B
  - Kinematic selection: Use kinematic constraints and charge conservation to infer true particle type

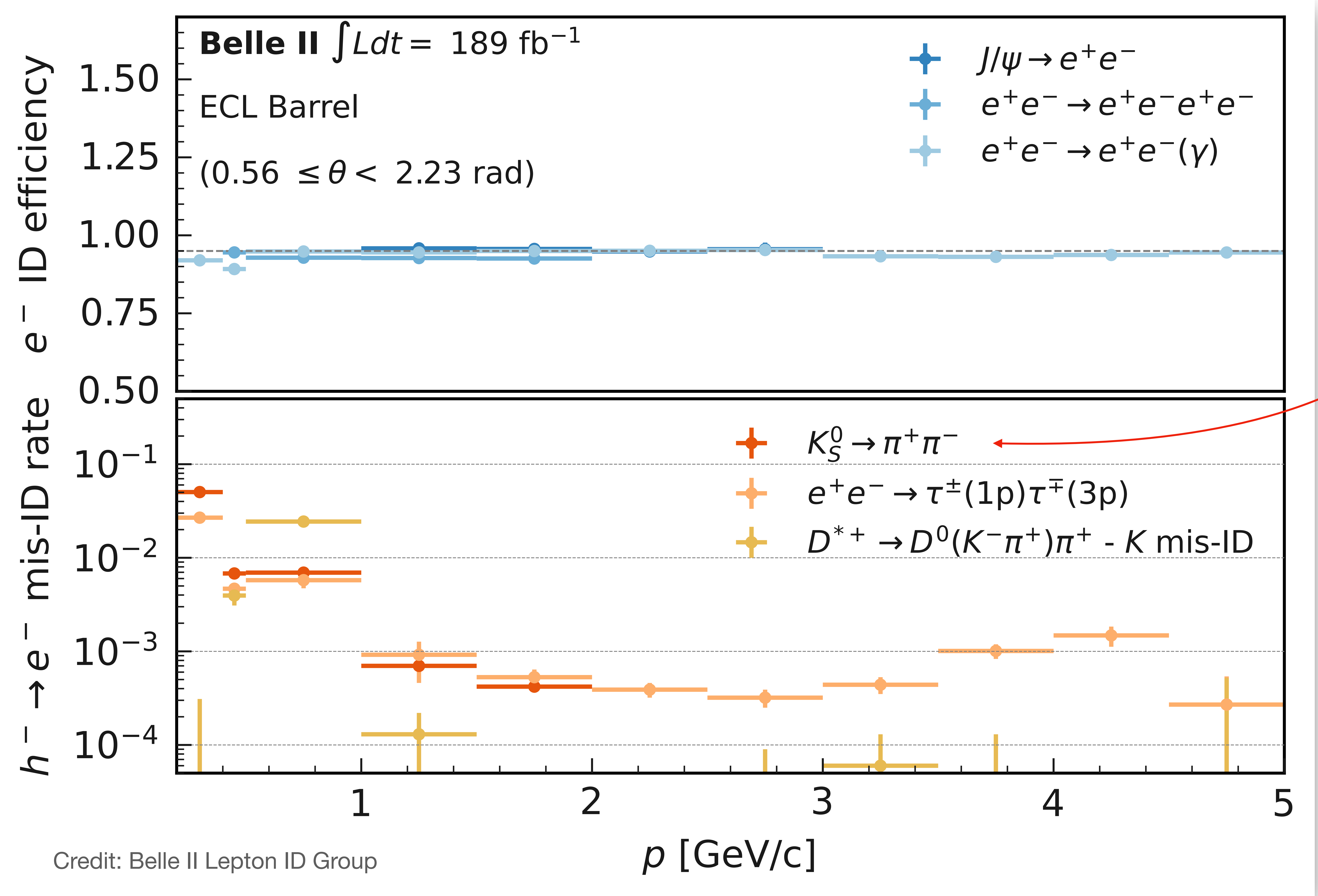
Efficiency determination: Tag and probe

- Powerful tool to determine e.g. particle identification efficiencies
- Select a very pure (usually not very efficient) sample where one of the selected particles has no particle identification criteria applied
- Example:
  - $K_S^0 \to \pi^+\pi^-$: Two tracks, **one** with pure pion identification, invariant mass of the two tracks very close to the known K_S^0 mass, two tracks coming from the same vertex, vertex is rather displaced, …
  - Next check if the track without pure particle identification is correctly identified as pion
  - Studies are often performed as function of momentum, direction, separation to other particles, data taking period, …



Credit: Belle II Lepton ID Group

Efficiency determination: Tag and probe example



Tag&Probe pions that are misreconstructed as electrons

Efficiency determination: Orthogonal selection

- Example: Determine ECAL trigger efficiency for events with a high energetic photon (i.e. something reconstructed in the ECAL)

- Strategy: Select events based on information that does not rely on the ECAL at all*
  - $e^+e^- \rightarrow \mu^+\mu^-\gamma$, but only select the two muons that do not sum up to the full event energy
  - Require that the event is triggered by a track trigger, T_{track}

- The ECAL efficiency is given by $\epsilon_{ECAL} = \dfrac{n(T_{ECAL} \,\&\, T_{Track})}{n(T_{Track})}$

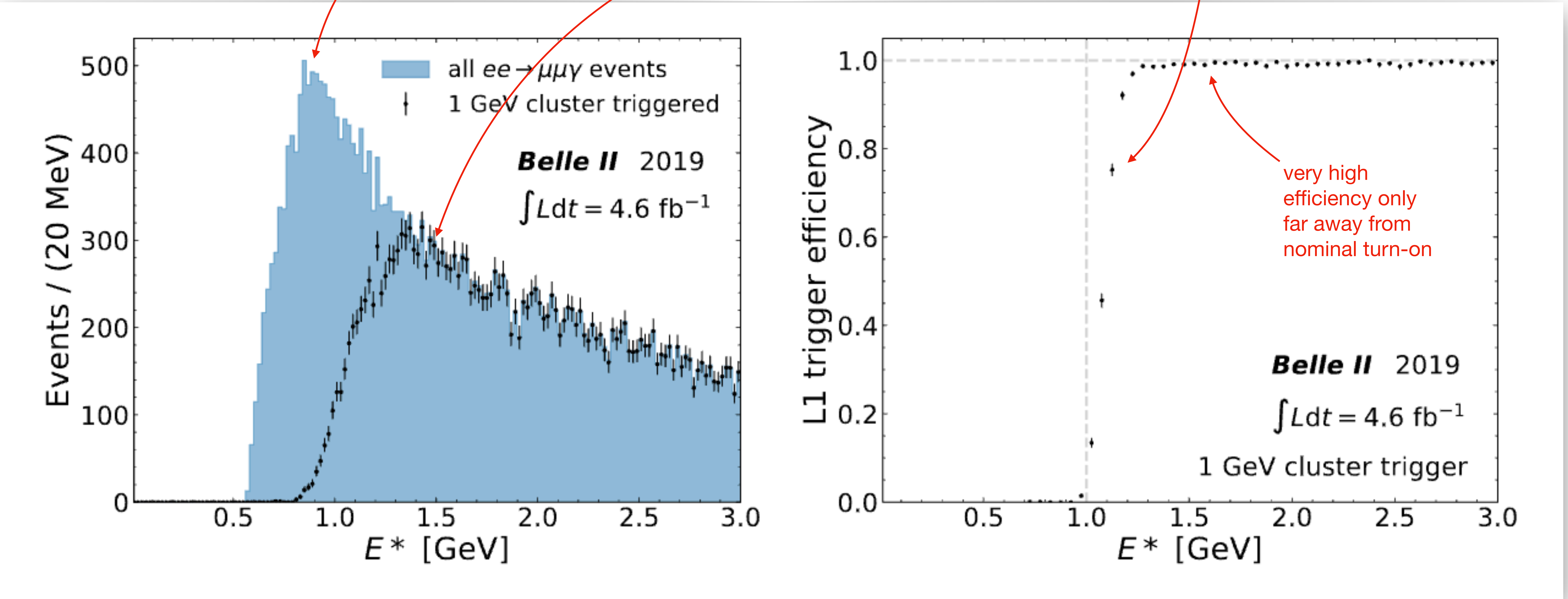
* If this requirement is violated (it almost always is somewhat violated), corrections have to be applied for correlations (often from simulation)

Efficiency determination: Orthogonal selection (example)

all events selected by two track trigger

all events selected by a cluster trigger
and the two track trigger

“turn on” curve



Credit: Belle II L1 Group

Efficiency determination: Kinematic tagging

- Rule of thumb: HEP detectors are very good measuring charges of particles (+ or -) to 0.1% or better

- “Nature’s gift” *(test your PDG skills if you don’t believe me!)*

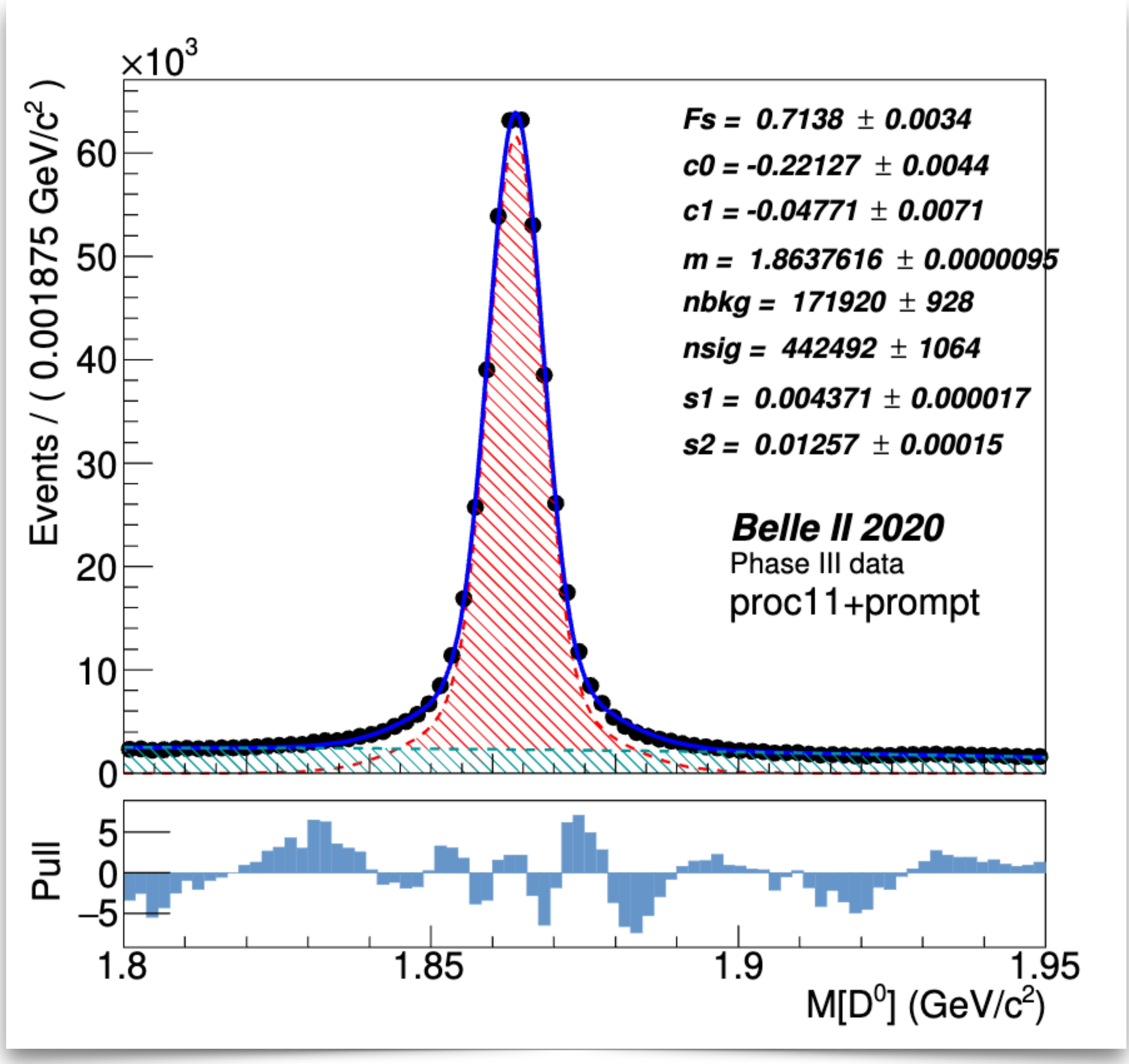
$D^{*+} \to \bar{D}^0[\to K^+\pi^-]\pi^+$ and $D^{*-} \to D^0[\to K^-\pi^+]\pi^-$

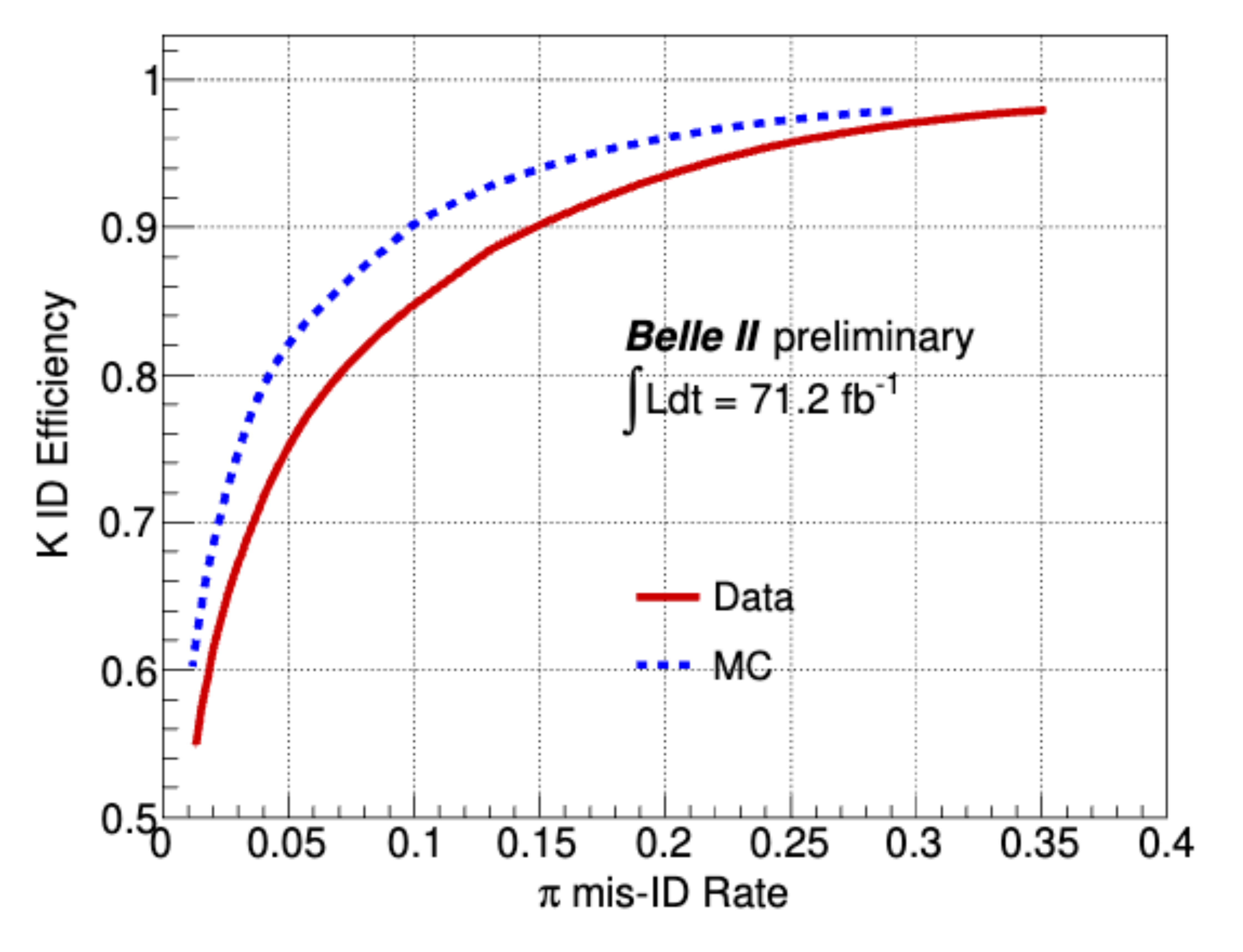
  - Select three charged particles that combine to the $D^{*\pm}$ mass, while two of them also combine to the D^0 mass

  - Determine the charge of the (low momentum) leftover particle, this is pion (with very high purity)

  - If the leftover particle is positively charged, the remaining other positively charged particle is a Kaon; if the leftover is negatively charged, the other positively charged particle is a pion

Efficiency determination: Kinematic tagging





Credit: Belle II Hadron ID Group

Other calibrations

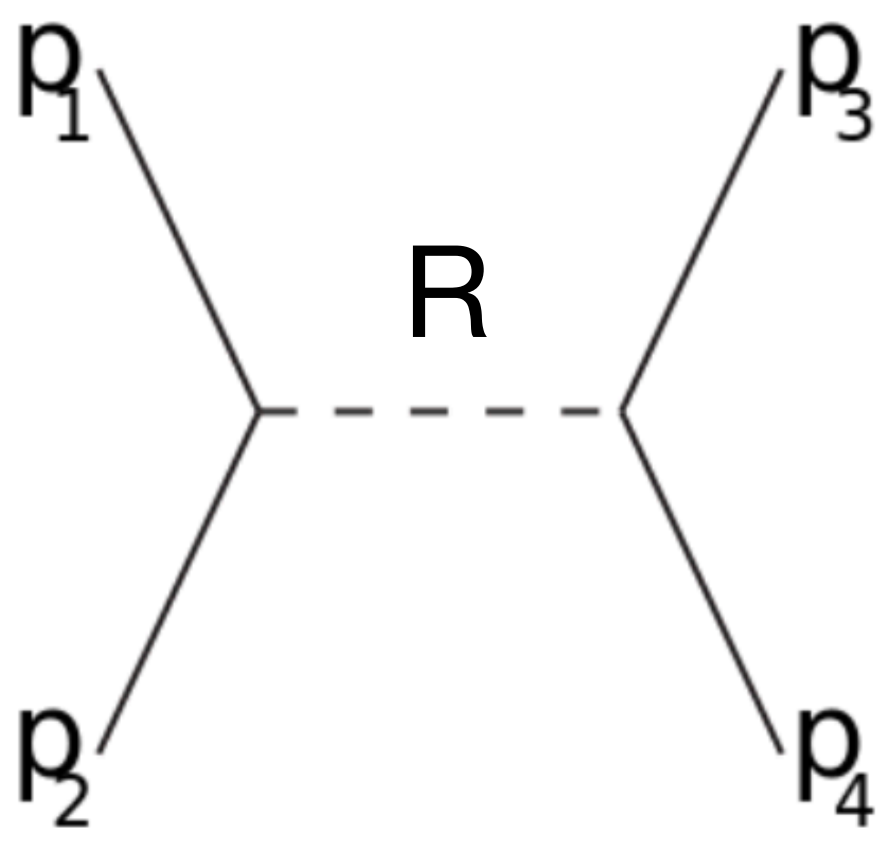
- Many more calibrations are needed with hundreds of people working on this in the large HEP experiments:
  - Tracking efficiency and fake tracks
  - Alignment of detectors
  - Track momentum and direction resolution and bias
  - Photon energy and position resolution and bias
  - Jet energy resolution and bias
  - Collision energy and interaction point
  - Stability of calibrations over time (from seconds to days)
  - …

How to get σ?

$$\sigma_{exp} = \frac{N_{\text{signal}} - N_{\text{background}}}{\epsilon \int \mathscr{L}}$$

Concept: Invariant mass

- If the collision energy is large enough to produce intermediate resonances **R** on-shell (and not just as virtual particles):
  - 4-momentum conservation $M_R = M(P_3 + P_4)$
  - Works also for more than just two particles
- Long-lived resonances are so narrow that they produce visible enhancements of the cross section near their nominal mass (e.g. π^0, K_S^0, Z^0, H, …) → Exercise 2



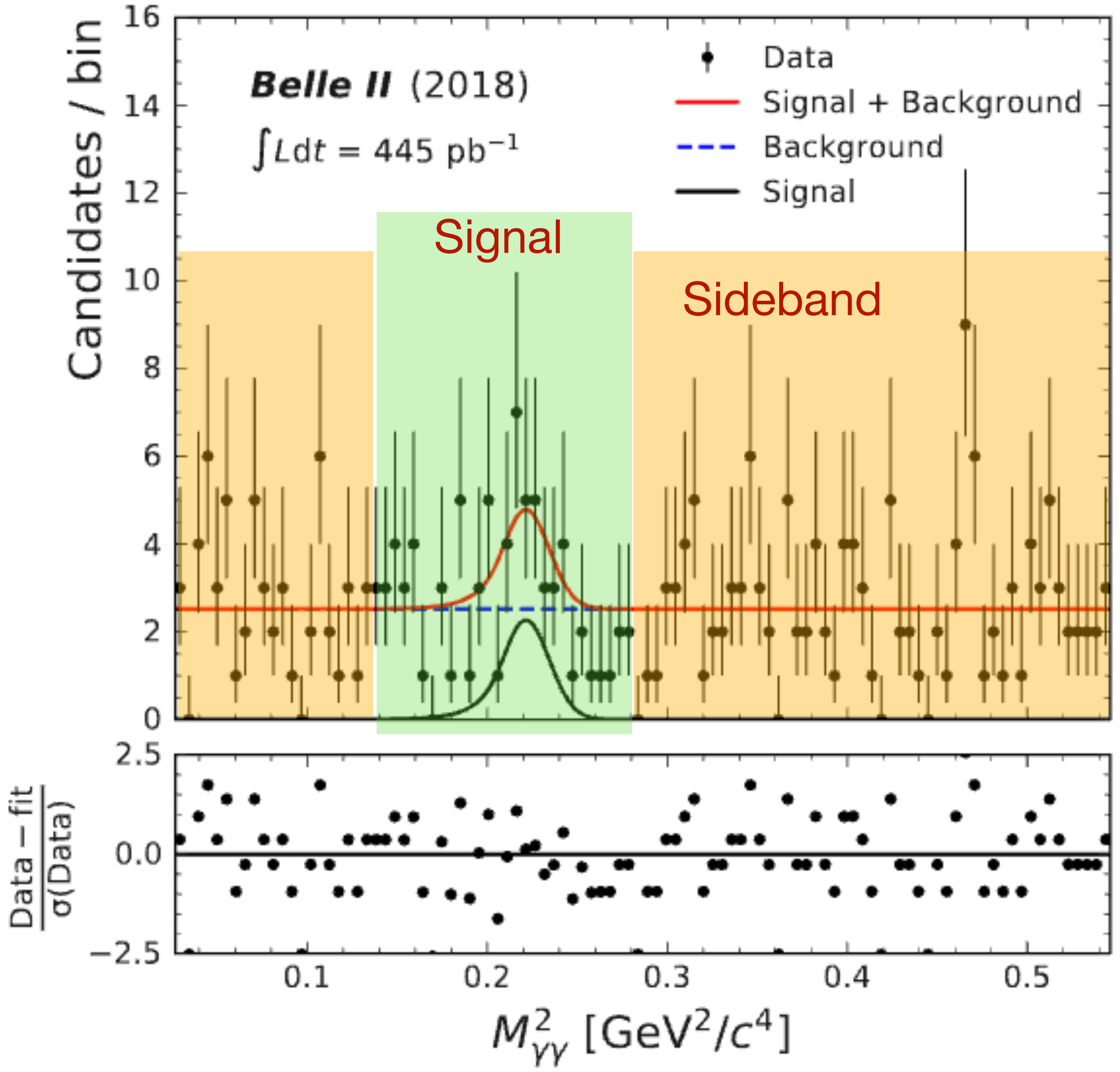
s-channel

Analysis: Background determination

- Background from **simulation**
  - Needs to be corrected for efficiency differences
  - Can not account for unforeseen differences
- Background taken from **control channels**
  - Physics channels that are similar to the channel under study, e.g. different charged hadrons or different charged leptons can be used to constrain overall normalization
  - Sometimes control channels have very low statistics
- Background taken from **event mixing**
  - If the background is coming from random wrong combinations of particles (e.g. $\pi^0 \to \gamma_1\gamma_2$), mixing random particles from different events can yield a pure background sample (there is no signal in particle from different events per definition): $\pi^0_{mixed} \to \gamma_{event\,A}\gamma_{event\,B}$

Analysis: Background determination

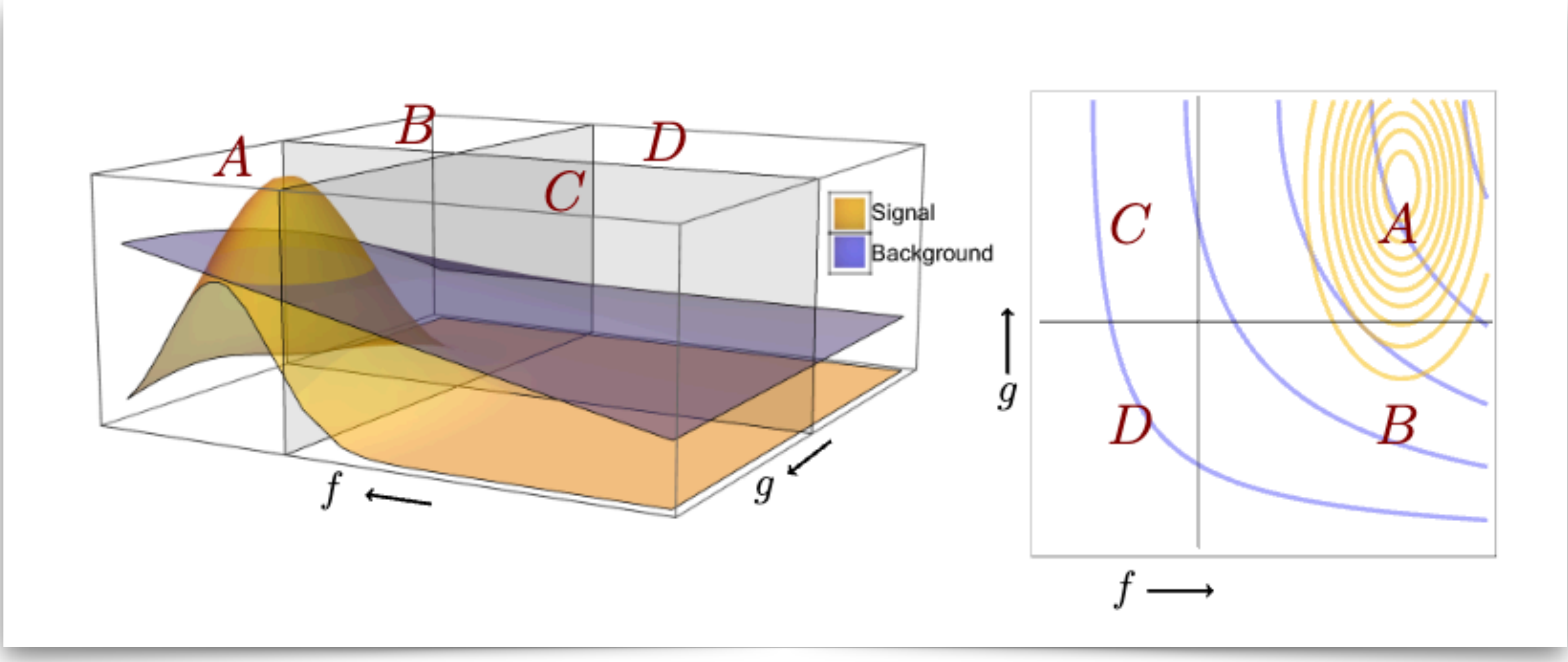
- Data-driven background from the **sidebands** near the signal peak (same variable)
  - Shape of background sometimes determined in simulation
  - Only works if sidebands exist (and are not zero)
  - Combined fit to signal+background extrapolates background into the signal region



- More advanced method: **sPlot** (https://root.cern/doc/master/classRooStats_1_1SPlot.html)

Analysis: Background determination via ABCD method

- Assumption:
  - Two statistically independent variables f and g exist (for example the invariant mass of a two jet system and the rapidity)
  - Apply selection cuts on the two variables that split the parameter space in four regions
  - Requires low signal contamination in regions B, C, and D (absolute and relative)
- Number of events in region D: $n_A = \dfrac{n_B n_C}{n_D}$
- Residual correction for correlations and signal contribution in background regions from simulation



Credit: https://arxiv.org/pdf/2007.14400.pdf

Simulation vs reality

compare with theory

Reality

real experiment

detector readout

theory calculation

event generator

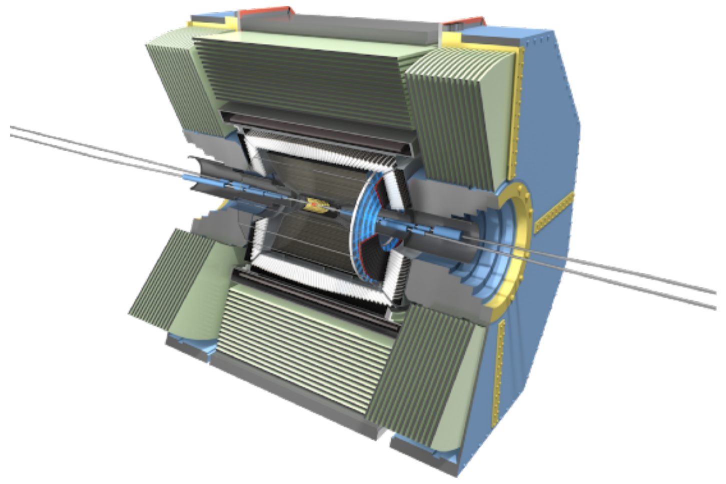
hadronization
(quarks → hadrons)

propagation through detector

simulated electronics response

Simulation

software “twin”



digitized and calibrated detector hits

reconstruction

Examples:
- Tracking
- Particle Flow
- Particle identification

$$\sigma_{exp} = \frac{N_{\text{signal}} - N_{\text{background}}}{\epsilon \int \mathscr{L}}$$

object calibration

Examples:
- Efficiencies

user analysis

Examples:
- Background modelling

What questions do you have?