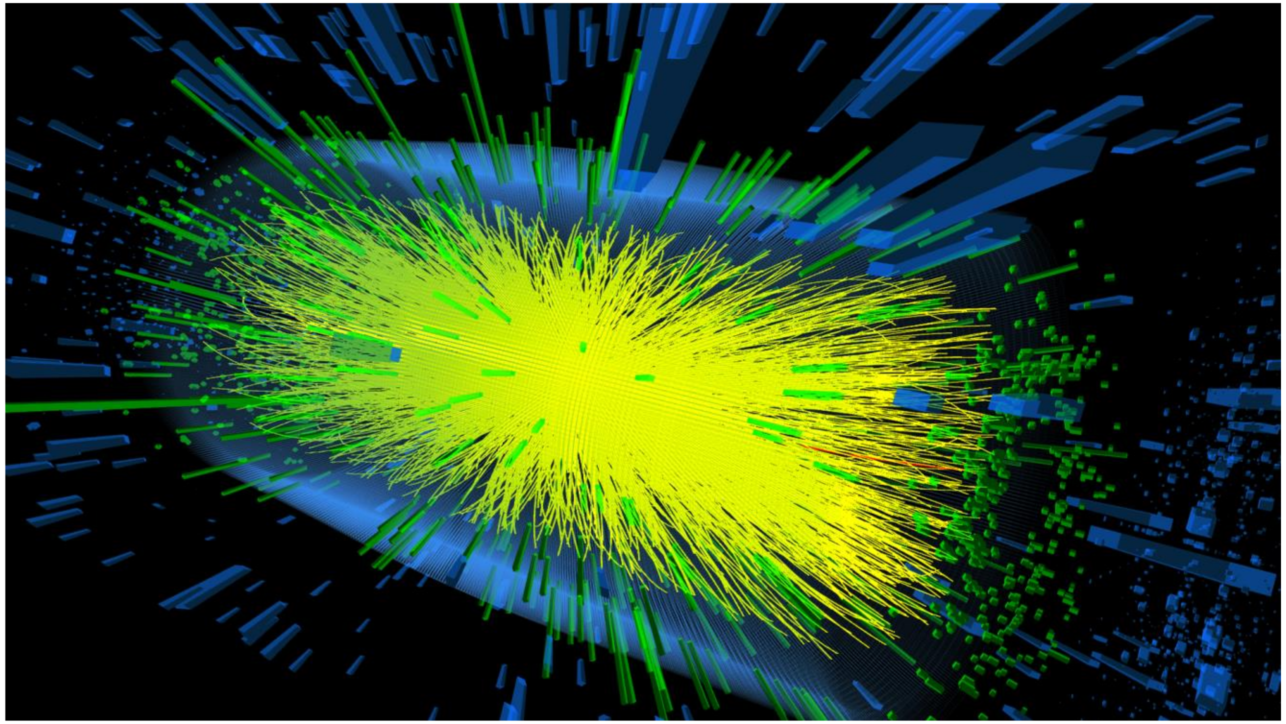
The alignment challenge in complex high resolution trackers

KIT Colloquium



Rainer Mankel (DESY)

Karlsruhe, 12 May 2022

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Outline

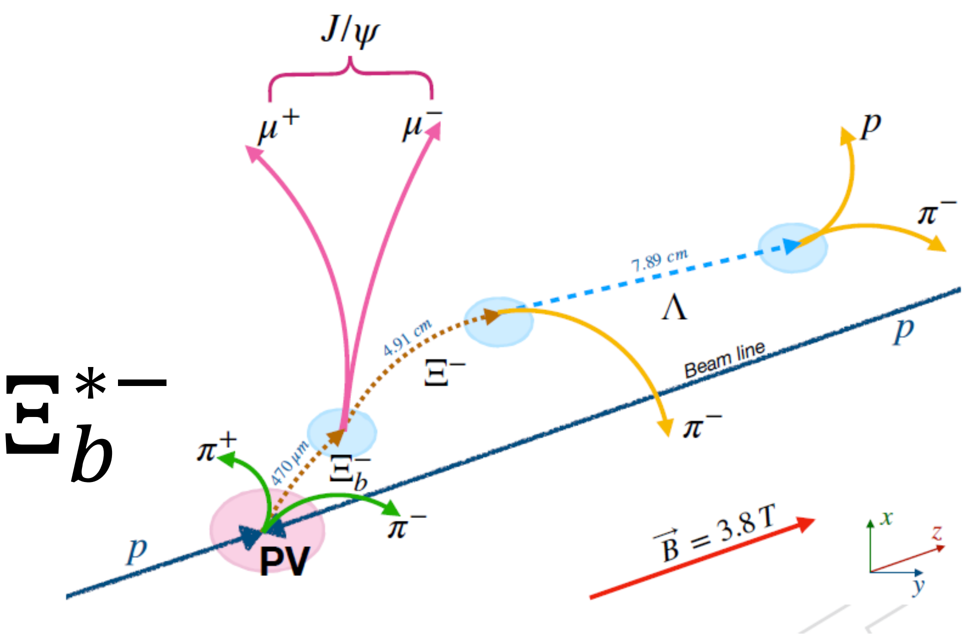
- Motivation
- Alignment challenges in complex tracking systems
- Methodology & solutions
- Practical experience with alignment in a huge tracking system
  - inputs & workflows
  - evolution with time
  - systematic effects & their impact on physics performance

Includes already results from our upcoming new paper:

Strategies and performance of the CMS silicon tracker alignment during LHC Run 2, arxiv:2111.08757
(will appear soon in Nucl. Instr. Meth. A)

Motivation: tracking in the LHC era

- How about discovering a new beauty-strange baryon…
  - possibly an excitation of Ξ_b^-, quark content (bsd)
  - produced at the interaction point → expect complex decay cascade

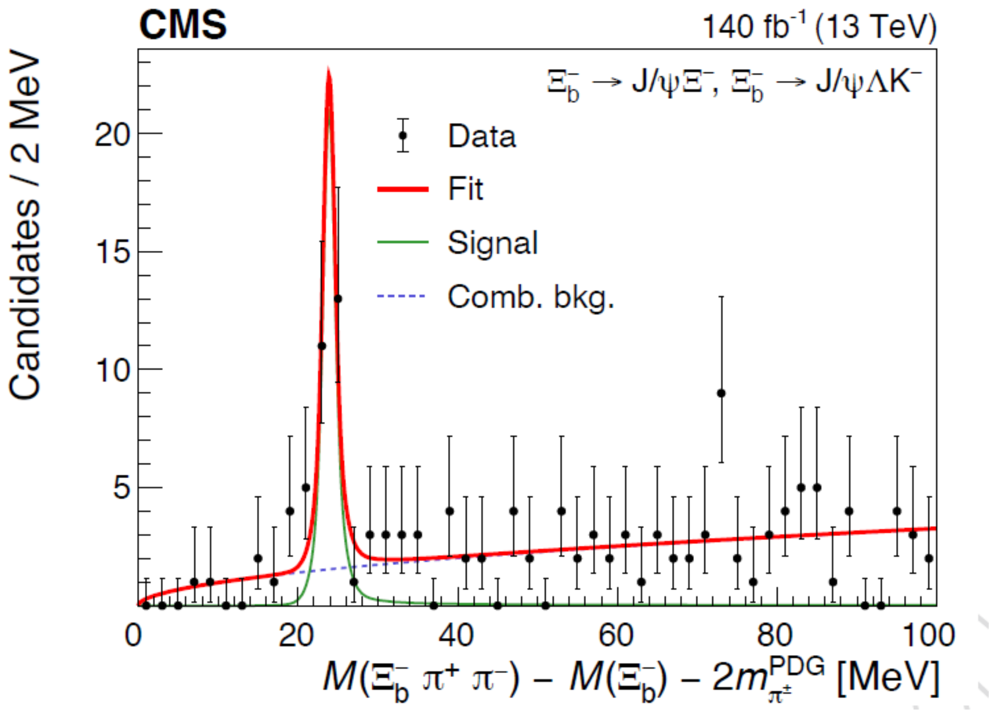


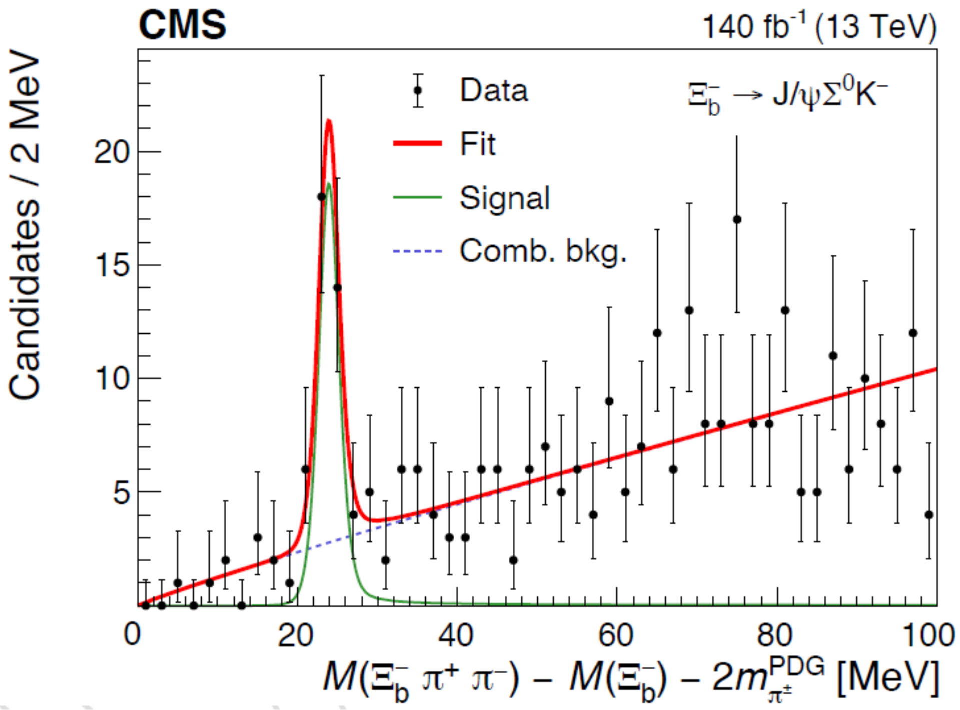
- Reminiscent of bubble chamber physics in the 60’s
- But can we do this in presence of a pileup of 60 and more inelastic interactions in the detector for each event…?

Observation of a new excited beauty strange baryon

Phys. Rev. Lett. 126, 252003 (June 2021)

- First observation of $\Xi_b(6100)^-$
  - orbital excitation of Ξ_b^-, $J^P = {}^3/_2{}^-$





$6.2 - 6.7\sigma$

- Very low background due to lifetime signature. Excellent mass resolution
  - precision tracking at the LHC

Why precision tracking matters

Tracking is more important than ever

Secondary vertices

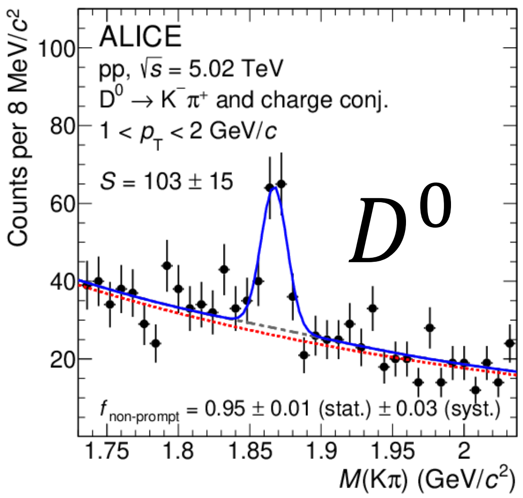
Example: properties of charm and beauty hadrons

Flavor tagging of jets

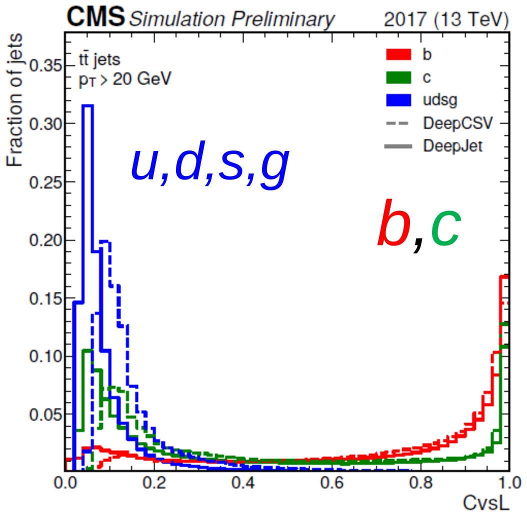
Example: H→bb, H→cc

Improved jet reconstruction & pileup mitigation

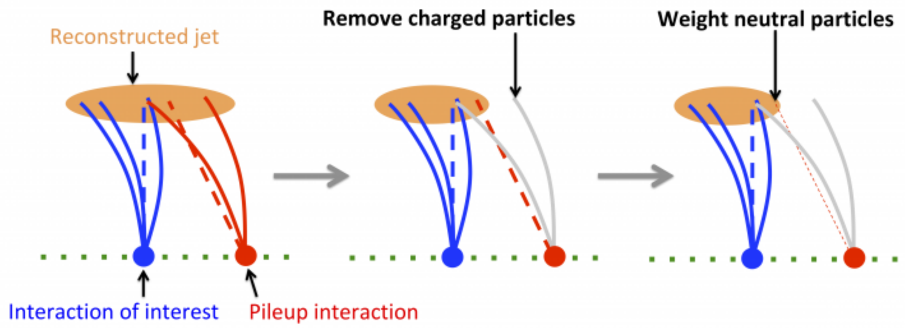
Key technology for HL-LHC



JHEP 2105 (2021) 220



JINST **13** (2018) P05011



- Precision tracking and alignment are key drivers of physics performance

Why is alignment important?

- Intrinsic coordinate resolution:
  - σ_{hit} ~9 μm (pixel), σ_{hit} ~20-60 μm (strip)
- The effective coordinate resolution emerges from combination of **intrinsic resolution** and **alignment**

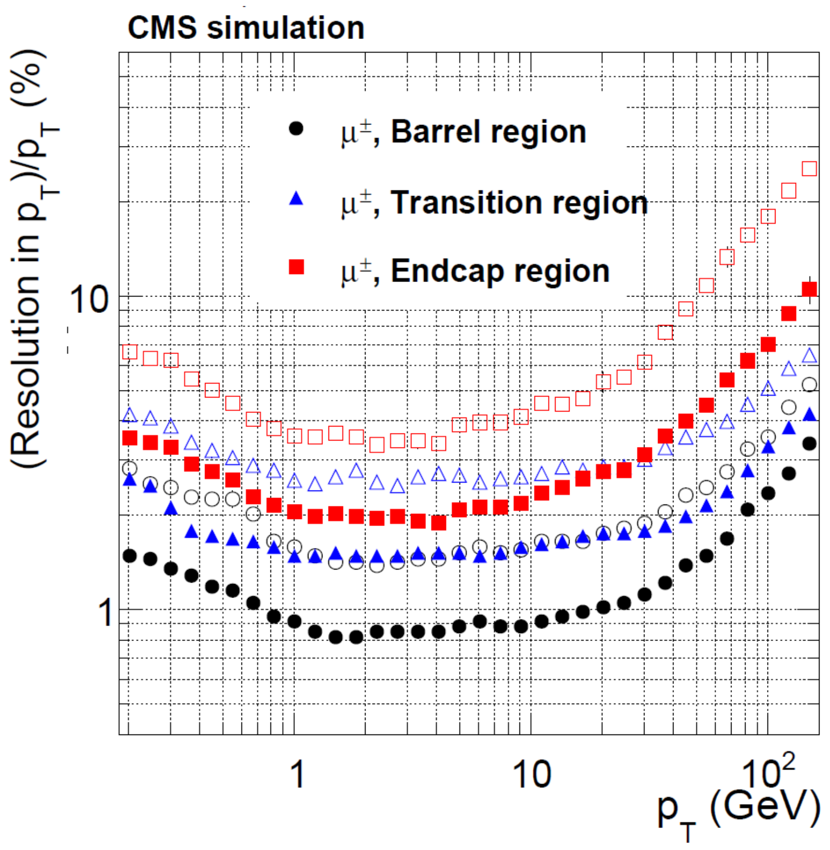
$$\sigma_{meas} \sim \sqrt{\sigma_{hit}^2 + \sigma_{alignment}^2}$$

➔ In a simplified model, the relative momentum resolution is the combined effect of **coordinate resolution** and **multiple scattering**

$$\frac{\delta p_T}{p_T} = C_1 \cdot p_T \oplus C_2 \qquad \text{where } C_1 \propto \sigma_{meas}$$

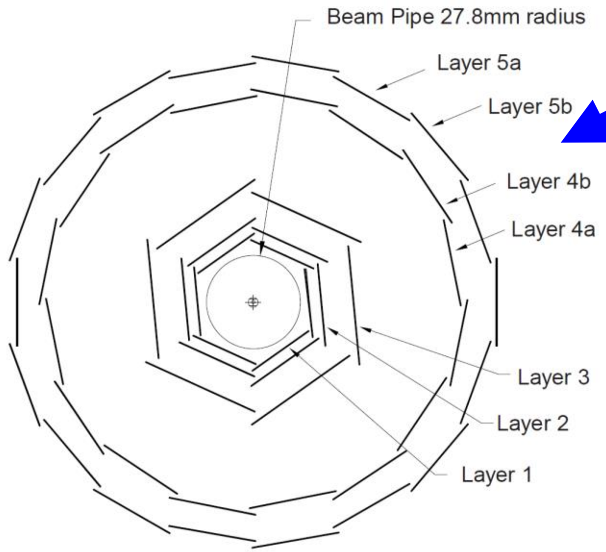
➔ **Need to keep $\sigma_{alignment}$ << σ_{hit}**

JINST **9** (2014) P10009



Complexity evolution of silicon trackers

A very arbitrary selection

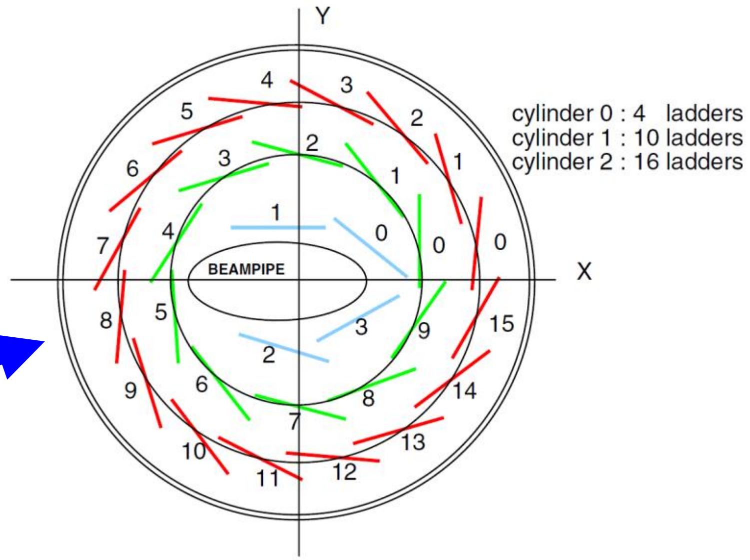


BaBar SVT

- 340 wafers

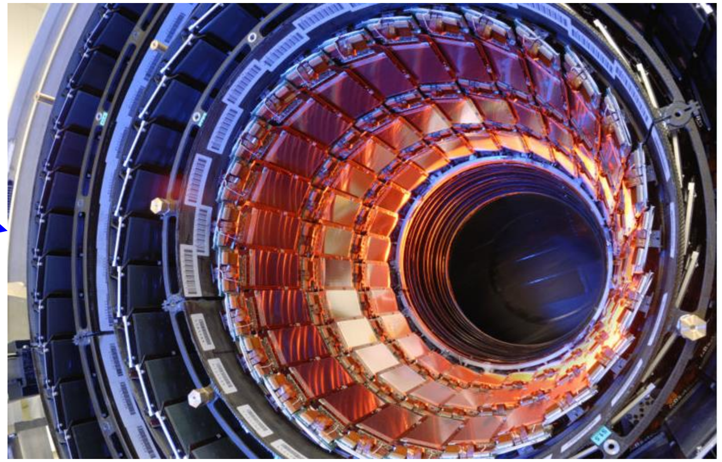
ZEUS MVD

- 712 single-sided strip detectors



CMS tracker

- 15,148 strip modules, 1,852 pixel modules (Phase 1)
  → ~17,000 modules

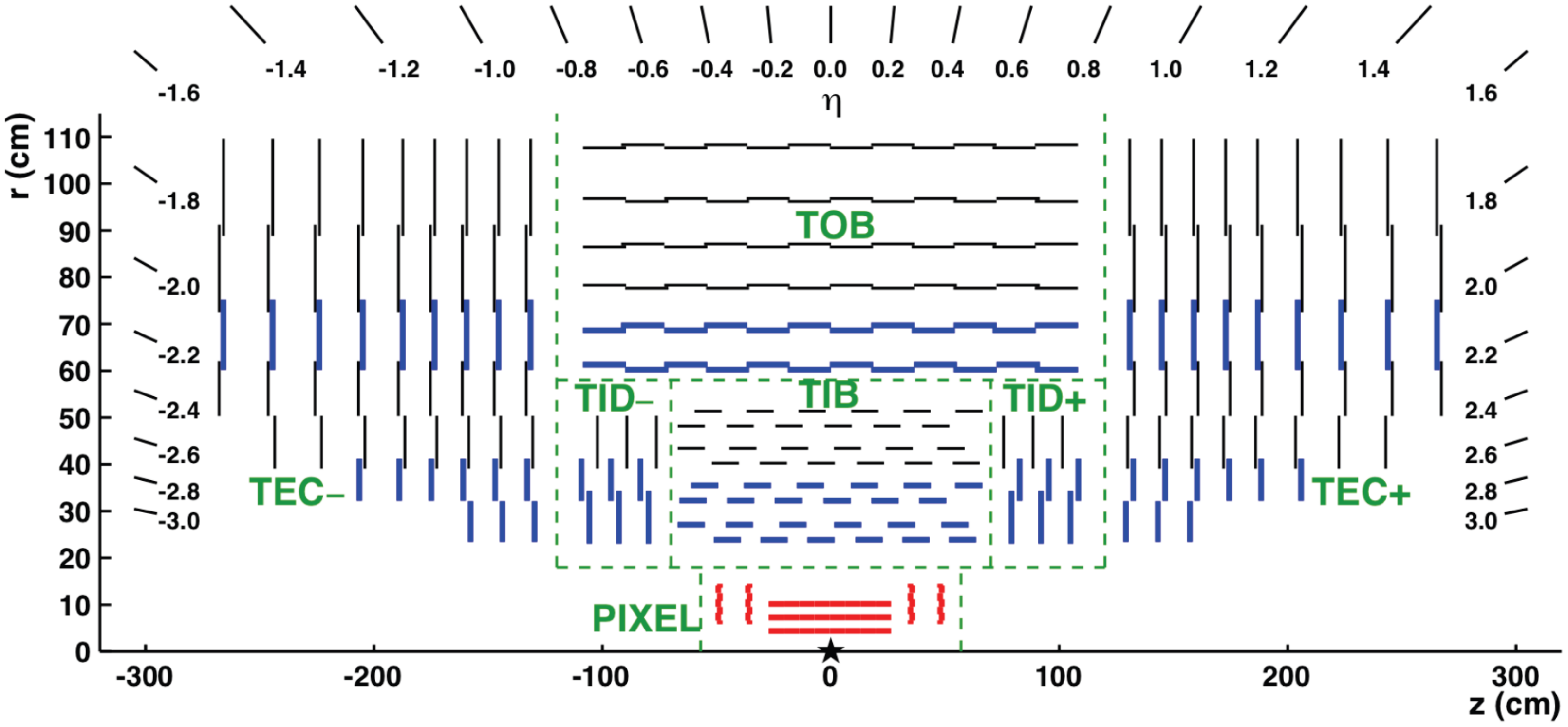


The CMS all-silicon tracker

The largest silicon tracker ever built

- Si-Pixel Detector
  ("Phase 1 upgrade" in 2017)
  - 66 M (124 M) pixels
  - 100 x 150 µm^2
  - 3 (4) barrel layers
  - 2x 2 (2x 3) endcap wheels
  - 4.7 < r < 10.2 cm
    (2.9 < r < 16 cm)

- Si-Strip Detector
  - 10 M strips in 10 layers
  - > 200 m^2 of silicon
  - 20 < r < 116 cm
  - 80—184 µm pitch

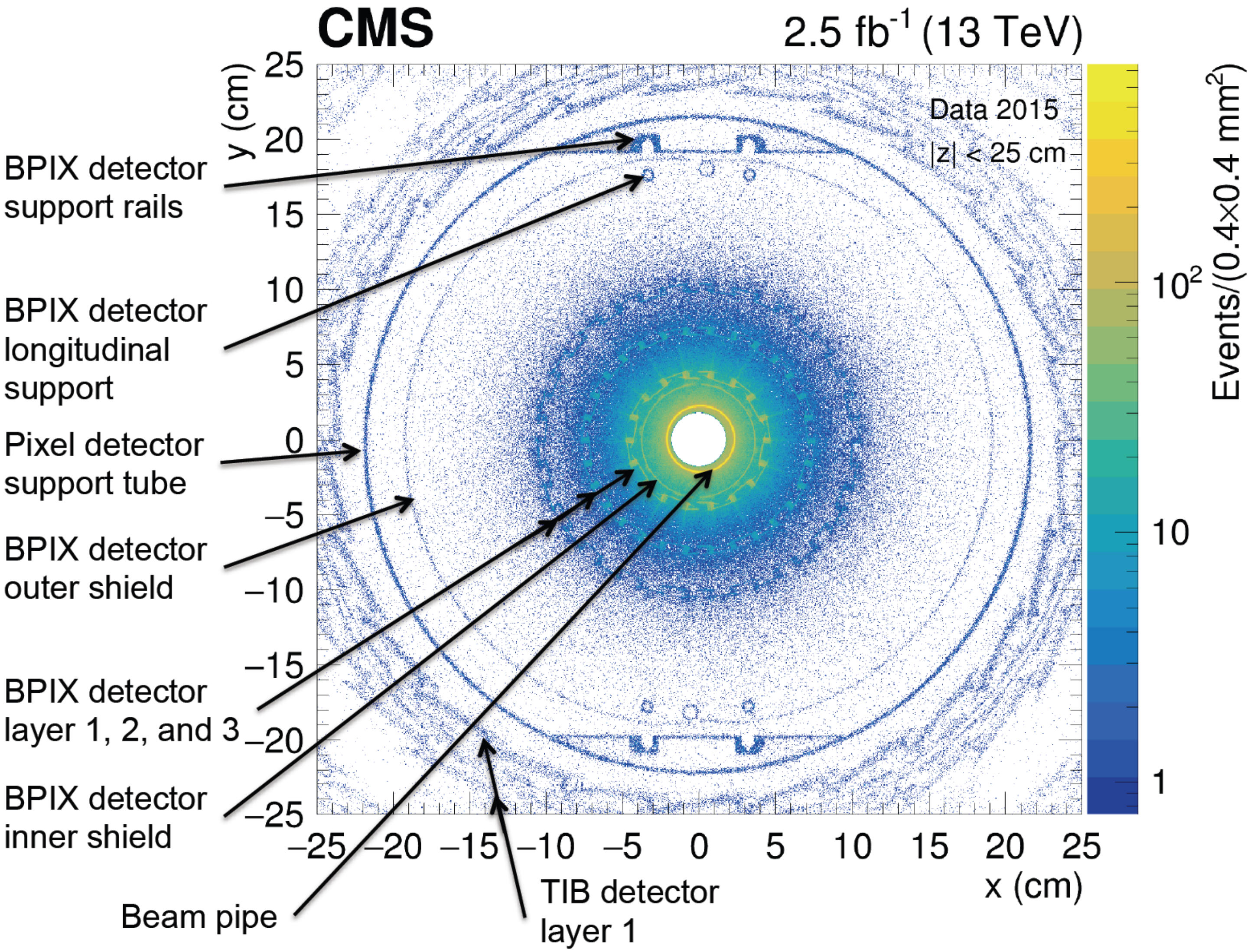


An “X-ray view” of the tracker in operation (2015 data)

Hadrography

- Based on reconstructed vertices from nuclear interactions in the material

➔ Detailed map of both sensitive and “dead” material



The LHC: a new level of challenge for detector alignment

- In the beginning, we were entering new territory in terms of tracker complexity. **Even in 2008, it was not entirely clear if/how the problem could be managed**

- Very clearly, major methodological developments were necessary

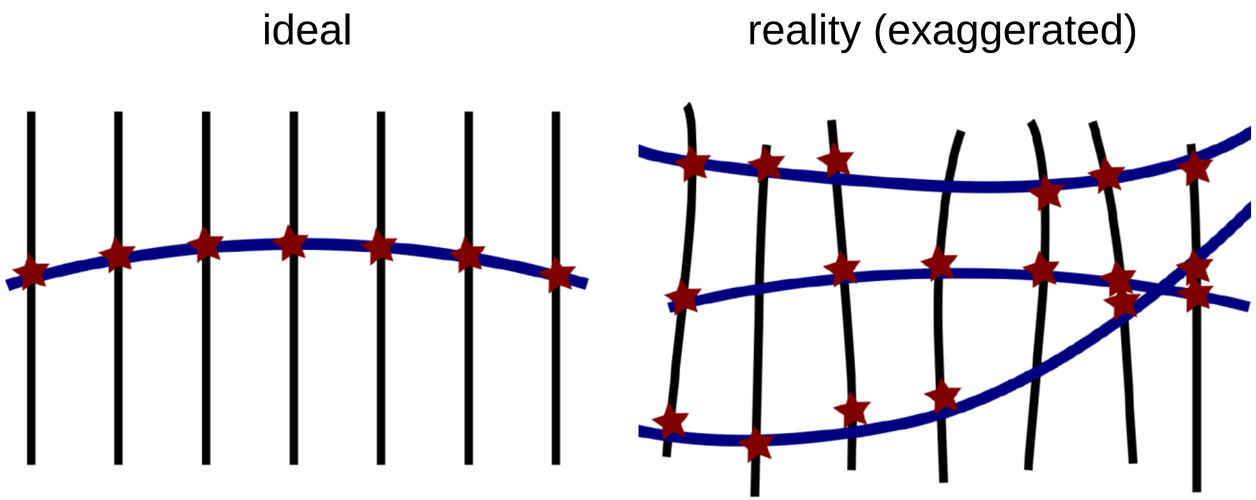
➔ A series of three LHC alignment workshops, with experts also from previous experiments, were organized to address these problems

LHC Detector Alignment Workshop 2009

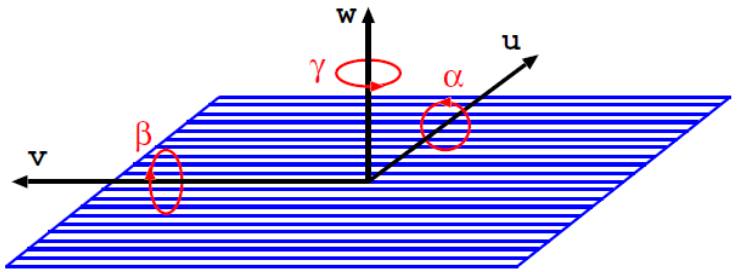
https://indico.cern.ch/event/50502/

Alignment basics

- For track-based alignment, we use many millions of tracks and study how they match to the hits in the detector modules
  - distance between track and hit: **"residual"**



- We introduce **corrections** to the module geometry (alignment parameters) such that they match well with the tracks
- Typically, there are three translational and three rotational alignment parameters per module (assuming planar shape)
  - corrections assumed to be relatively small
- But in practice, things are less simple…



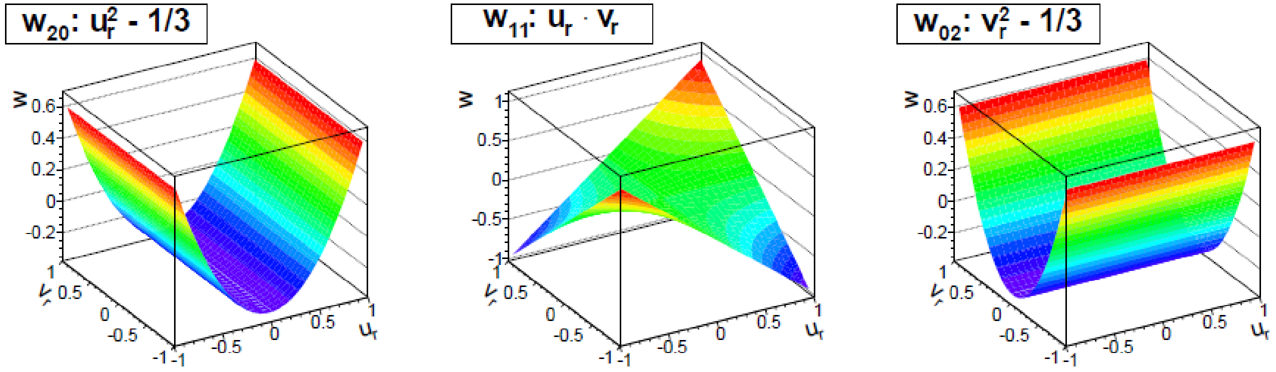
Sensor shape parameters

- In real life, sensors are not planar
  - ➔ without correction, coordinate measurement of non-perpendicular tracks is biased

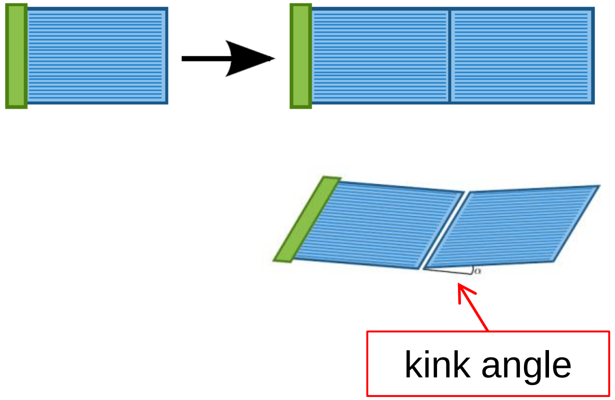
➔ Introduce three additional **curvature parameters** per sensor

➔ In addition, "**kink angles**" and offsets are introduced between daisy-chained sensors in TOB modules

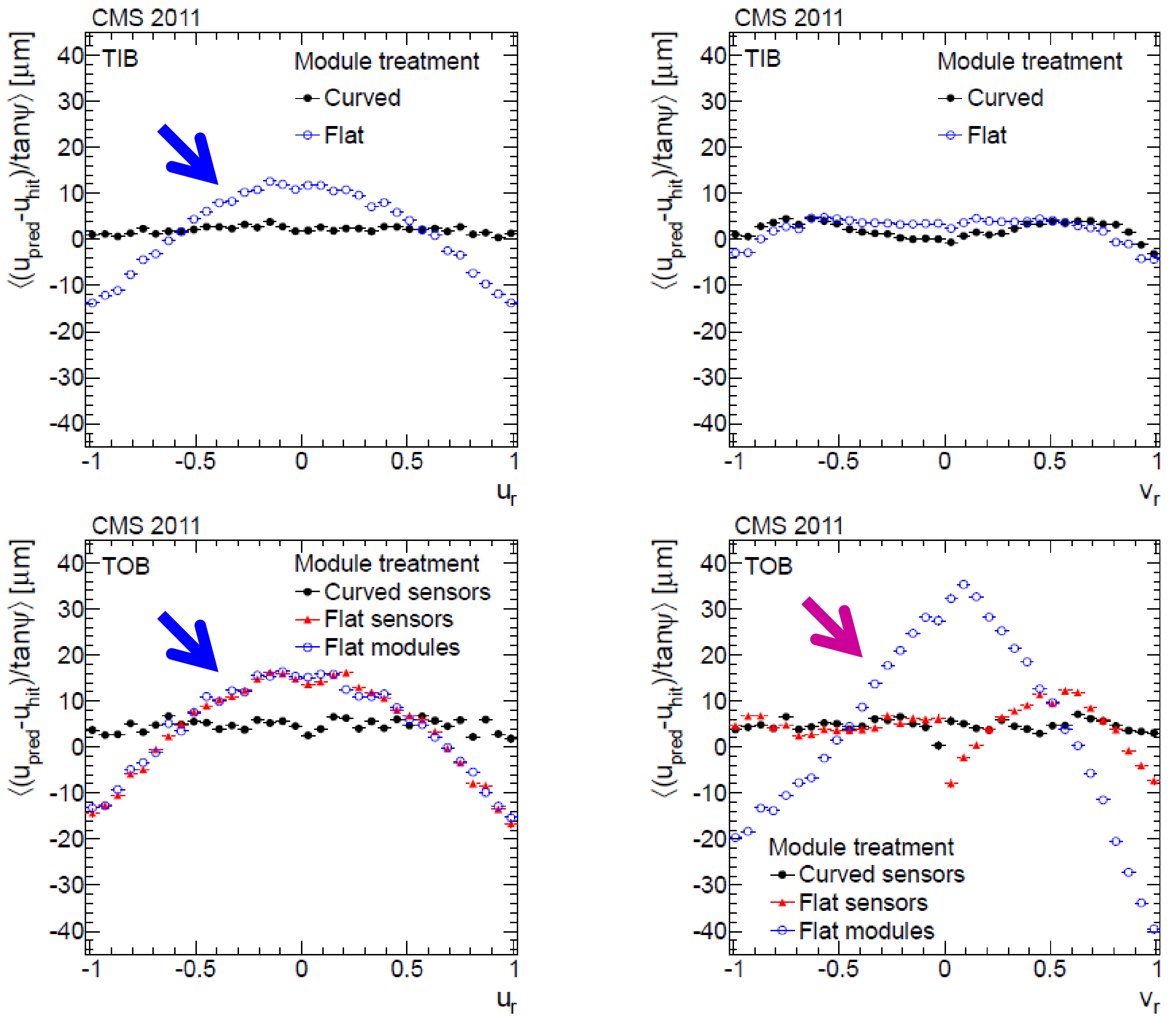
➔ Increases the number of alignment parameters 80,000 → 200,000

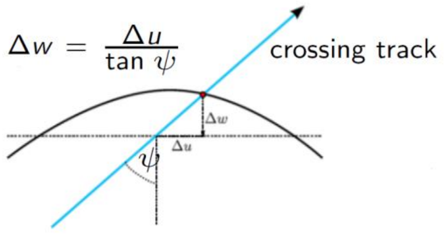


(strongly exaggerated)



Sensor shape parameters (cont'd)

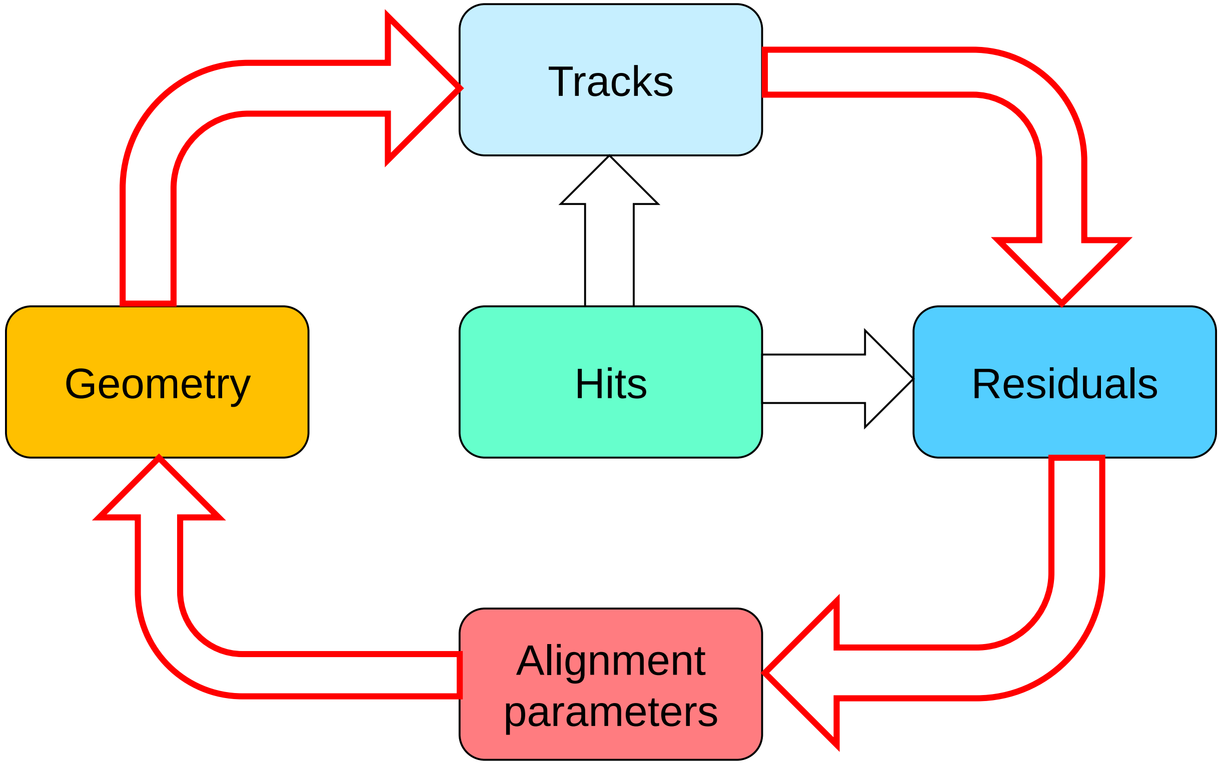




- Curvature in TIB and TOB modules (in direction transverse to strips)
- Kink between sensors in TOB modules (in direction parallel to strips)

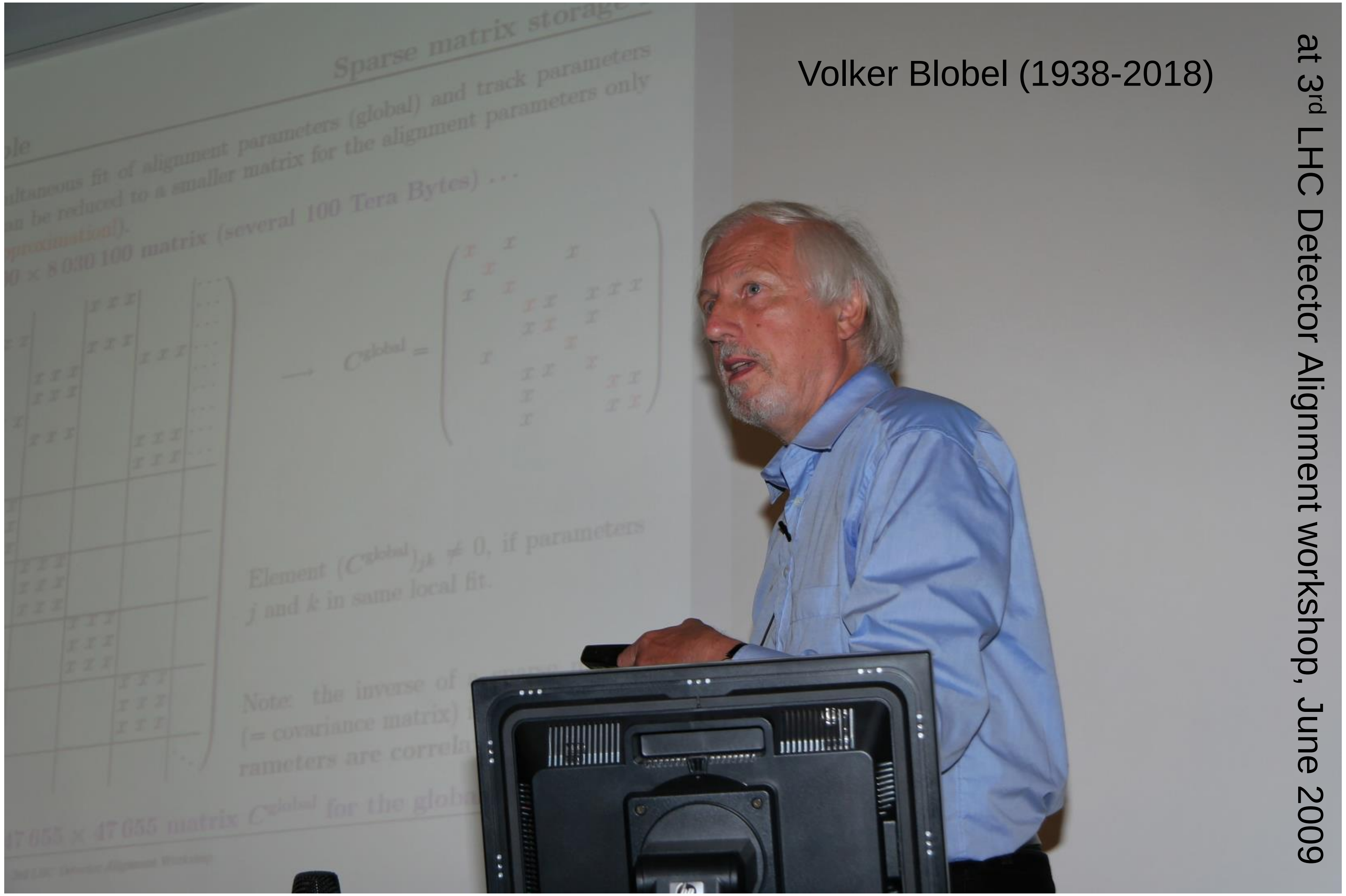
Alignment with residuals

- Straight-forward approach:
  - for each alignable object, evaluate track-hit residuals for all tracks, and compute alignment corrections by means of a least-squares fit
  - this leads to an updated geometry
- The problem:
  - also **tracks will change** when updating geometry
  - need to iterate (this procedure is actually applied in various experiments. in CMS: “HipPy” algorithm)
  - but convergence not guaranteed!
  - in a fit, correlations are important, and no good to ignore them
- The rigorous solution:
  - **simultaneous fit** of all tracks and all alignment parameters



The Millepede idea

A rigorous solution that is computationally manageable

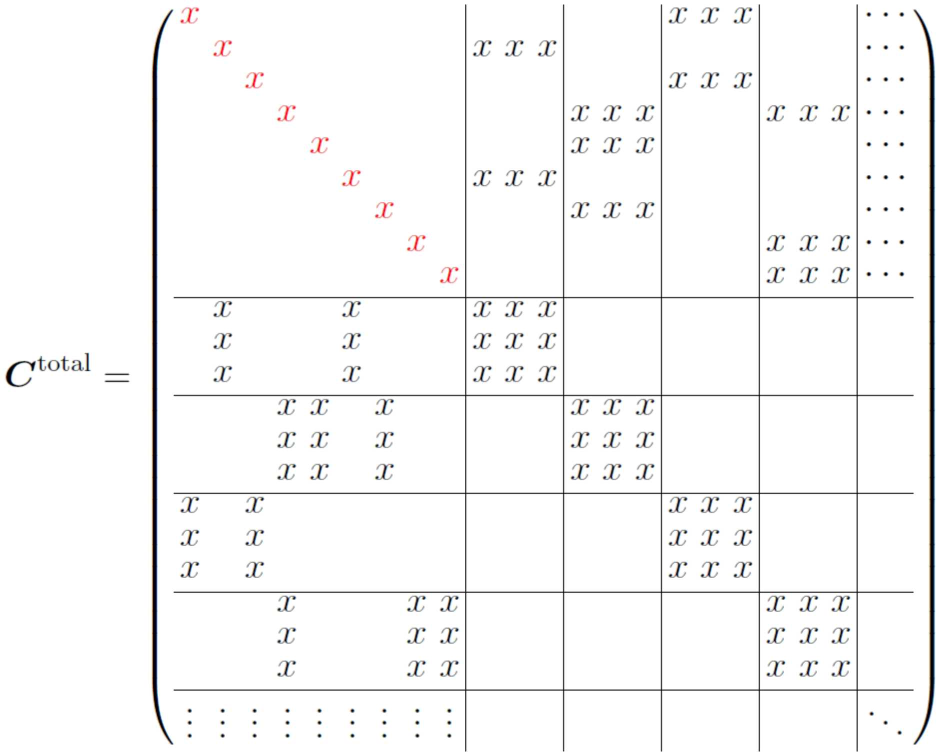


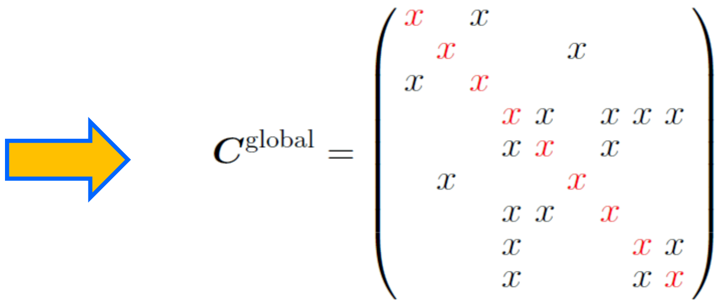
Volker Blobel (1938-2018)

at 3rd LHC Detector Alignment workshop, June 2009

The Millepede idea (cont‘d)

- Blobel’s example: 1,596,489 tracks (@ 5 parameters); 47,655 alignment parameters
  - >8 M free parameters to be determined → equation system characterized by 8M x 8M matrix (several 100 TB!)





- With a smart transformation, using Schur complements, this problem can be reduced to one with a much smaller matrix for the alignment parameters only
  - 47,655 x 47,655
  - no approximation involved
  - this is a sparse (!) matrix

https://indico.cern.ch/event/50502/contributions/1183071/attachments/964111/1368903/cernali.pdf

The Millepede program

- Millepede (I):
  - since 1998 used in H1 for vertex detector and central jet chamber
  - since 2000 downloadable from the web… adopted by many experiments, still used today
  - used for up to 4,800 alignment parameters

- With LHC on the horizon it became clear that this program could not meet the highest demands
  - for example, CMS:
    - 17,000 modules → ~100,000 alignment parameters in straight-forward implementation
      - number of matrix elements → exceeds largest possible 4 byte integer
      - numerical methods for solving in Millepede I not adequate
    - today's CMS alignment campaigns even exceed 200,000 parameters

- **Development of Millepede-II** → cutting-edge solving of massive linear problems

Millepede-II: computational/numerical technology

- Simply speaking, track-based alignment can be described as solving a huge linear equation system:

$$\underbrace{\begin{pmatrix} \square & \cdots & \square \\ \vdots & \ddots & \vdots \\ \square & \cdots & \square \end{pmatrix}}_{C'} \underbrace{\begin{pmatrix} \Delta p_1 \\ \vdots \\ \Delta p_n \end{pmatrix}}_{\text{alignment parameters}} = \underbrace{\begin{pmatrix} \square \\ \vdots \\ \square \end{pmatrix}}_{b'}$$

In CMS, C' is typically a matrix with 50,000 - 200,000 rows and columns

- Straight-forward solution (= inversion of the matrix C’) only possible for “small” number of parameters

Method	Computing time	Solution type	Error calculation
Inversion (Gauss–Jordan)	$\sim n^3$	Exact	Yes
Cholesky decomposition	$\sim n^3$	Exact	Skipped (for speed)
MINRES [24, 25]	$\sim n^2 \times n_{\text{it}}$	Approximate	No

- Very good turnaround thanks to exploitation of matrix sparsity, multithreading, and dedicated large-memory machines

Millepede-II: further information

- Millepede-II is maintained & further developed by Claus Kleinwort (DESY)
  - under the umbrella of the Helmholtz alliance “Physics at the Terascale”
- https://gitlab.desy.de/claus.kleinwort/millepede-ii

Contact

For information exchange the **Millepede** mailing list anacentre-millepede2@desy.de should be used.

References

1. A New Method for the High-Precision Alignment of Track Detectors, Volker Blobel and Claus Kleinwort, Proceedings of the Conference on Adcanced Statistical Techniques in Particle Physics, Durham, 18 - 22 March 2002, Report DESY 02-077 (June 2002) and hep-ex/0208021
2. Alignment Algorithms, V. Blobel, Proceedings of the LHC Detector Alignment Workshop, September 4 - 6 2006, CERN
3. Software alignment for Tracking Detectors, V. Blobel, NIM A, 566 (2006), pp. 5-13, doi:10.1016/j.nima.2006.05.157
4. A new fast track-fit algorithm based on broken lines, V. Blobel, NIM A, 566 (2006), pp. 14-17, doi:10.1016/j.nima.2006.05.156
5. Millepede 2009, V. Blobel, Contribution to the 3rd LHC Detector Alignment Workshop, June 15 - 16 2009, CERN
6. General Broken Lines as advanced track fitting method, C. Kleinwort, NIM A, 673 (2012), pp. 107-110, doi:10.1016/j.nima.2012.01.024

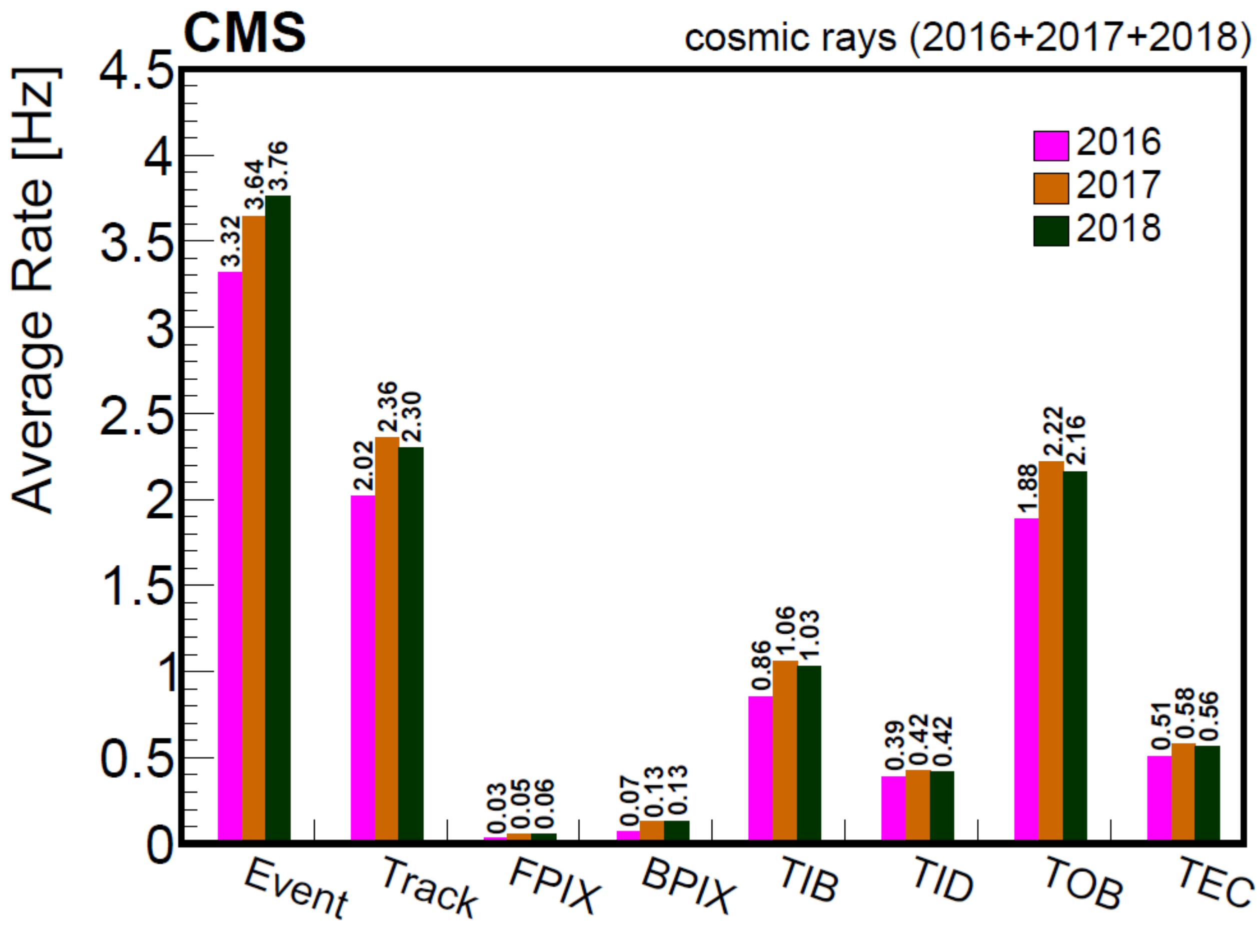
Track inputs used for CMS alignment

- "Minimum bias" events
  - inclusive trigger
  - dominantly QCD, p > 8 GeV, p_T >1 GeV

- Isolated muons
  - mostly W decays

- Di-muon resonances
  - $Z \rightarrow \mu\mu$, $\Upsilon(1S) \rightarrow \mu\mu$

- Cosmic ray muons
  - with / without magnetic field
  - dedicated, interfill & during collisions



Generation cycles of CMS alignment

- Alignment during data-taking
  - automated, unsupervised alignment with limited number of degrees of freedom

↓

- Interim and end-of-year (EOY) campaigns & reprocessing

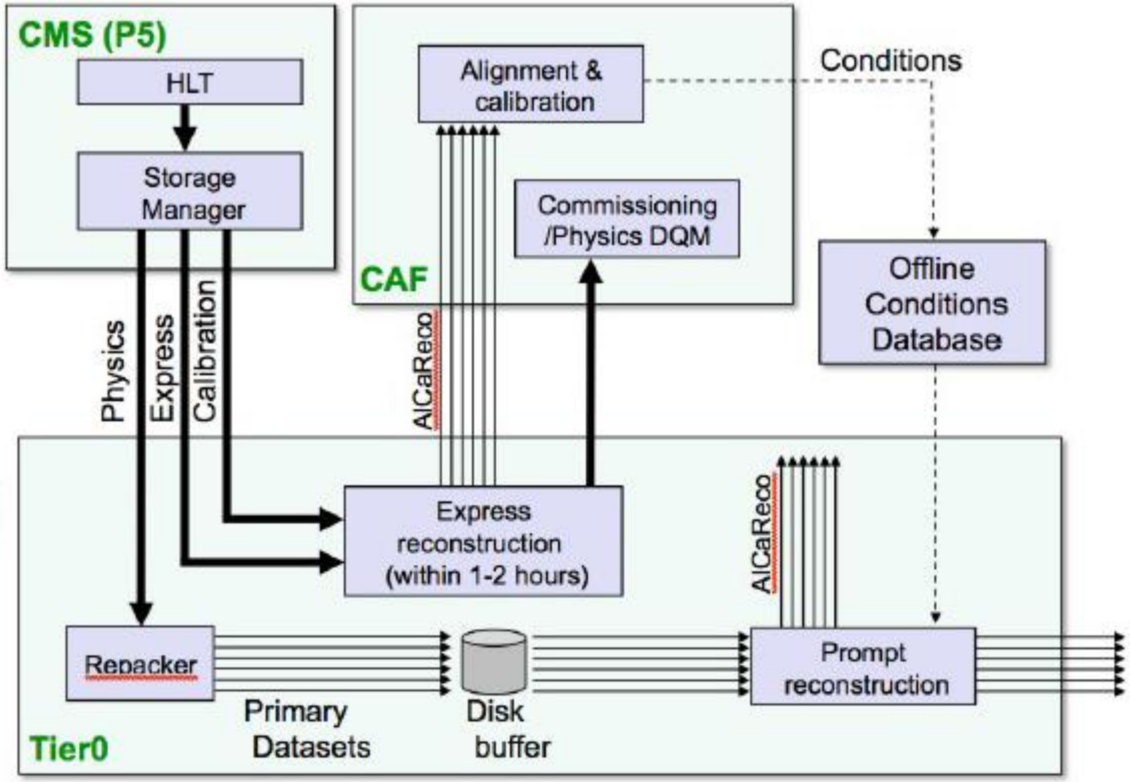
↓

- Legacy alignment, calibration & reprocessing
  - a special effort after the end of Run 2
  - bring all physics data in 2016-18 to the optimal level

Automated alignment

Prompt calibration loop (PCL)

- Restricted to parameters of very high level structures
- Focuses on offsets and angles of pixel tracker:
  - two half-barrels
  - two half cylinders in each endcap
  - 36 parameters in total
- Part of **prompt calibration**, which operates on stream from express reconstruction at the CAF
- **Fast updates** of alignment constants can be provided within 48 hours
  - in time for prompt reconstruction

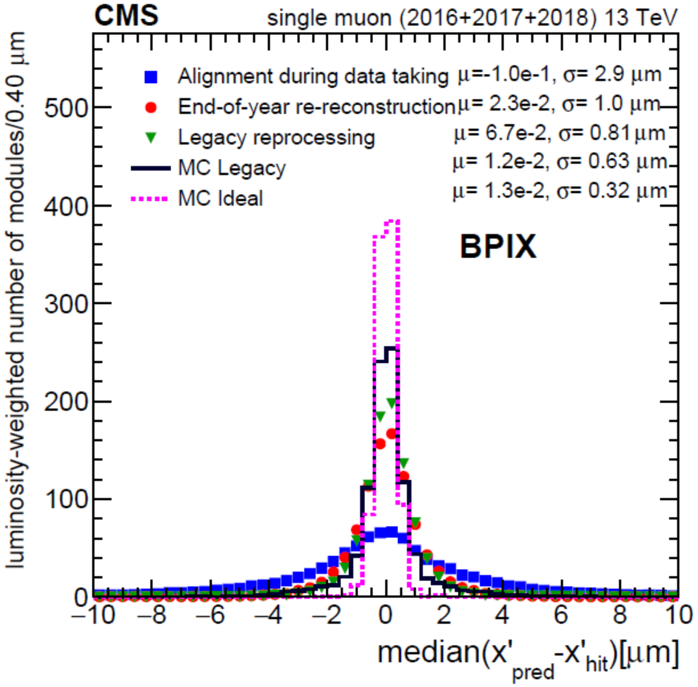


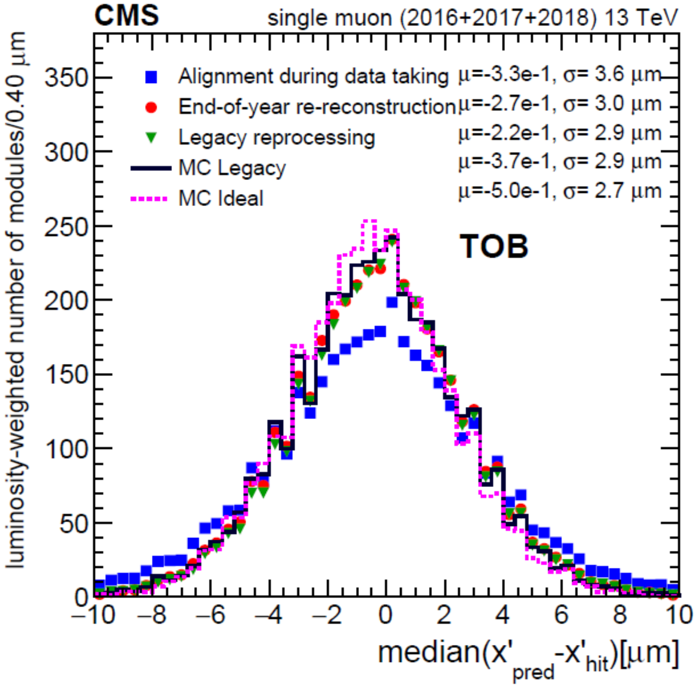
General quality of the alignment

DMR = distributions of the medians of the residual distributions

- Misalignment shows by de-centered distributions of hit residuals → visible in **median**
  - put medians of all residual distributions into one plot → representative of alignment precision
  - expect **narrow peak** for perfect alignment

barrel pixel detector



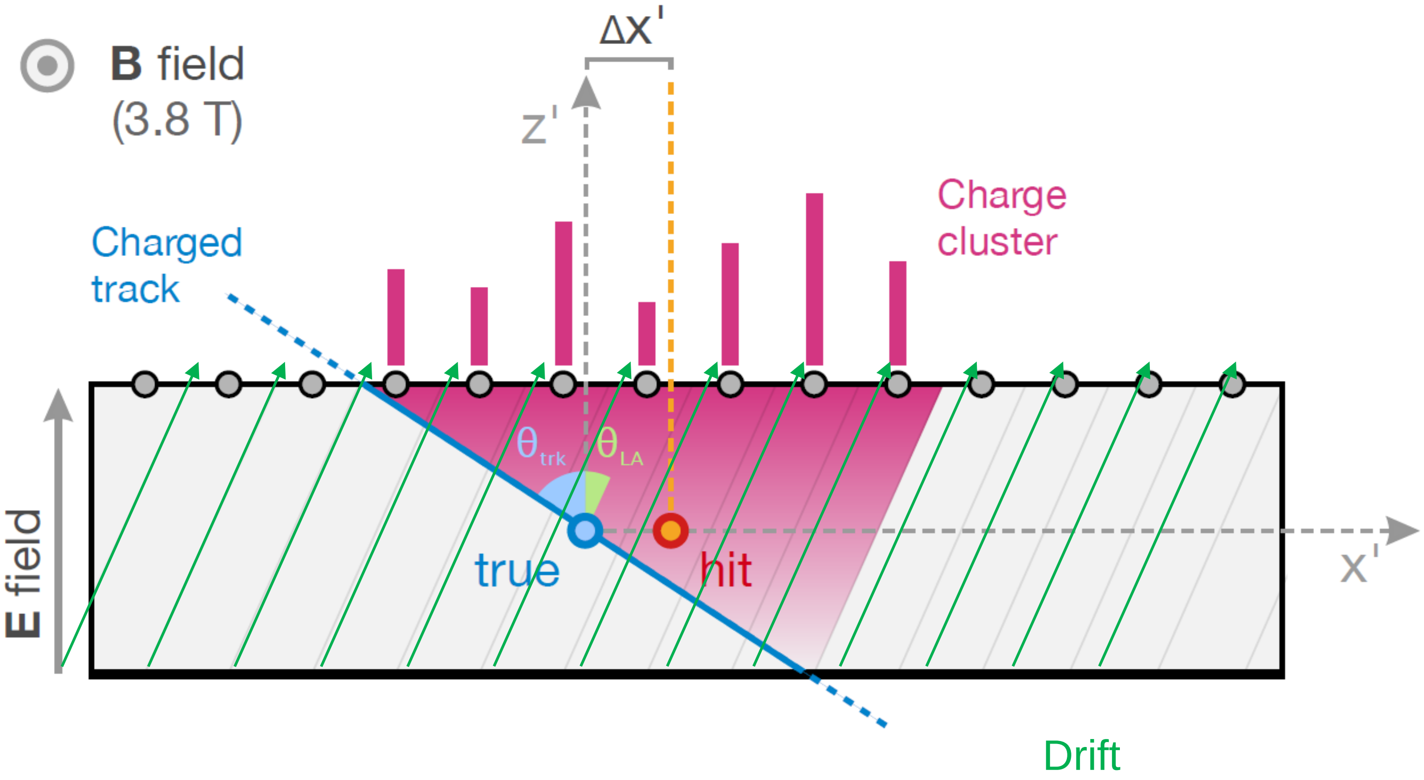


outer barrel strip detector

→ After legacy alignment, **close to ideal**. Also very decent description in MC

Lorentz angle effects

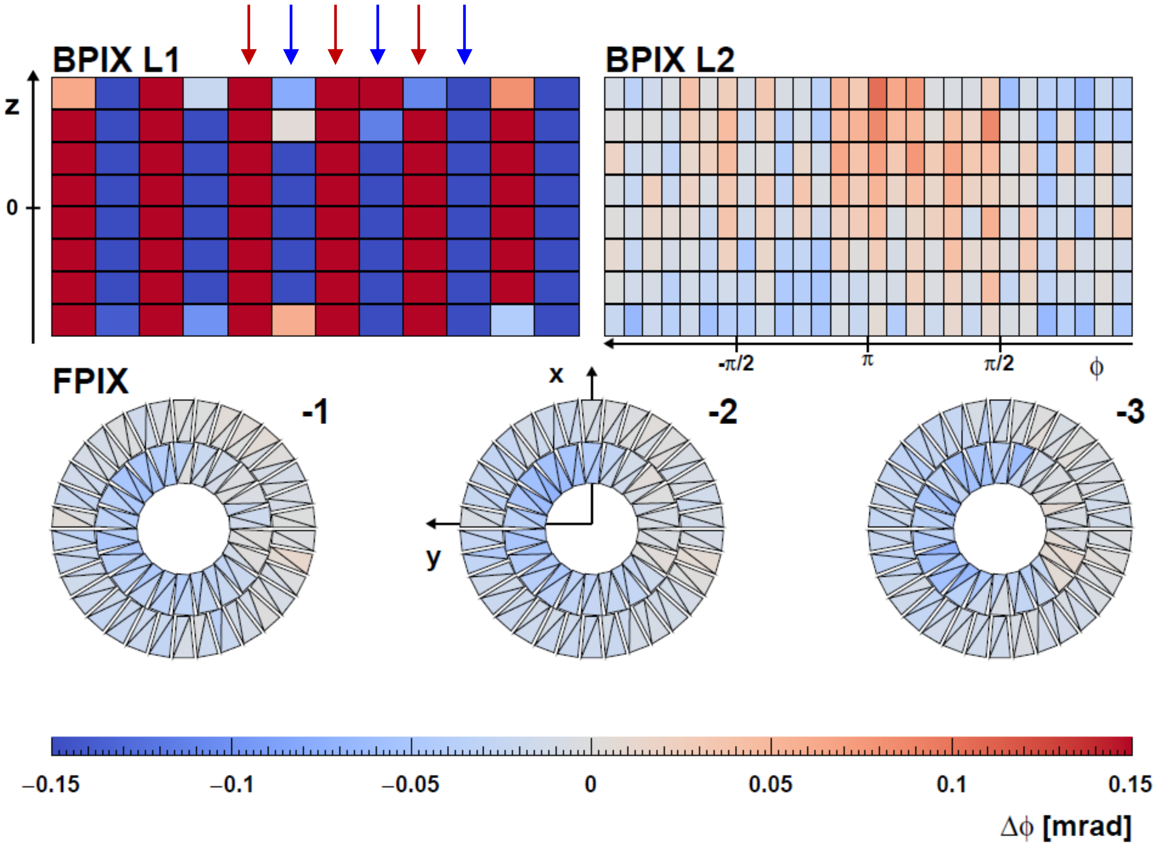
- Inside the silicon volume, the drift of the charge carriers is **deflected** by the Lorentz angle
  - shifts the apparent cluster position
- While this is addressed in first order by a dedicated Lorentz angle calibration, variations of the Lorentz angle as a function of location and time may result in effects that “look” like a misalignment of the sensor

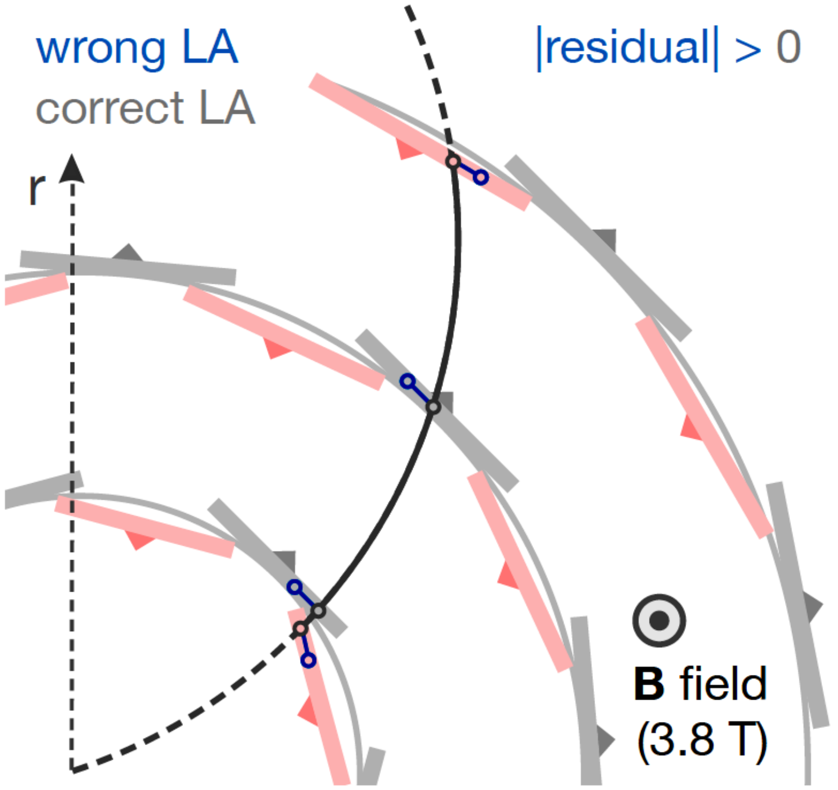


- Radiation damage may have impact after accumulation of 1 fb^{-1} , while pixel local reconstruction calibration can only be performed after 10 fb^{-1}
- **De-facto corrected** by the alignment procedure

Lorentz angle effects (cont'd)

- Large alignment corrections in innermost barrel pixel layer, alternating between adjacent ladders
  - explained by alternating orientations of pixel modules

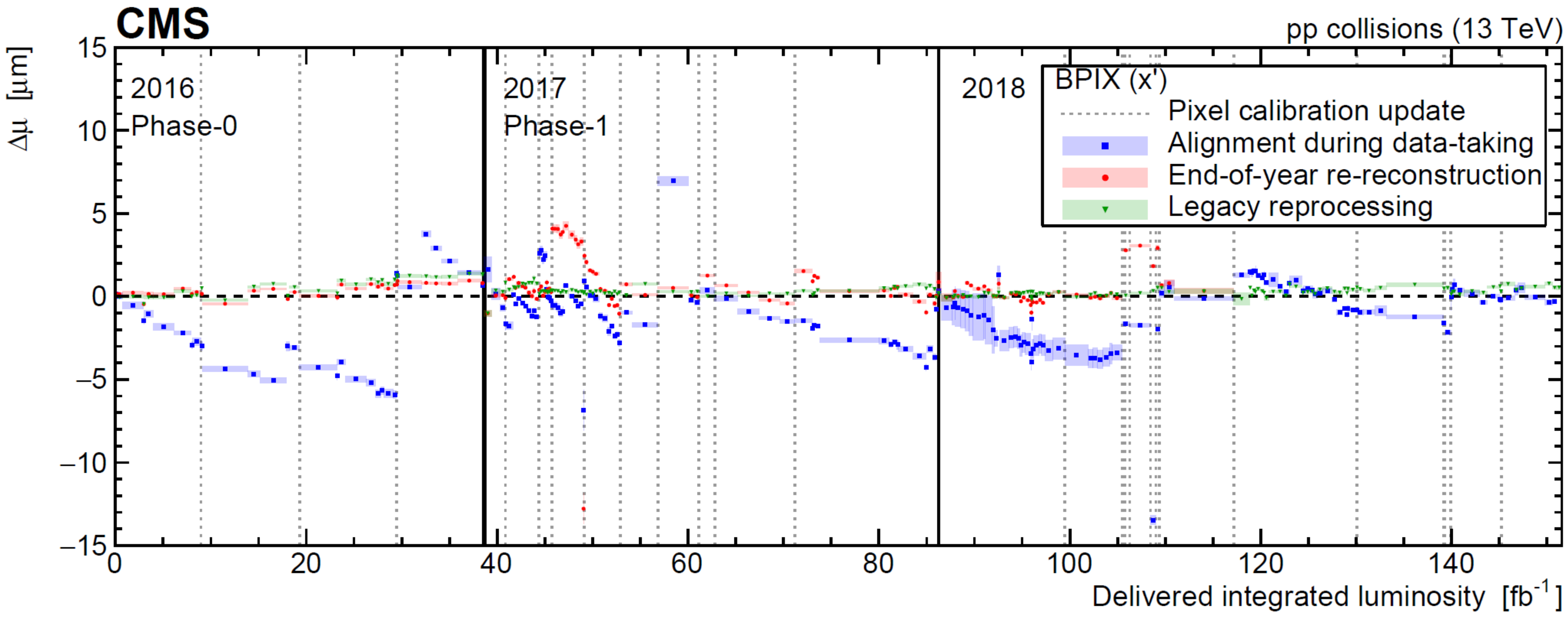




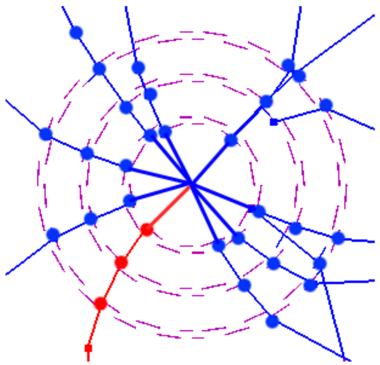
Lorentz angle effects (cont'd)

Effect of radiation damage

- Can we see this effect building up? Compare mean values of DMR for modules with electric field pointing inwards and outwards: $\Delta\mu = \mu_{inwards} - \mu_{outwards}$



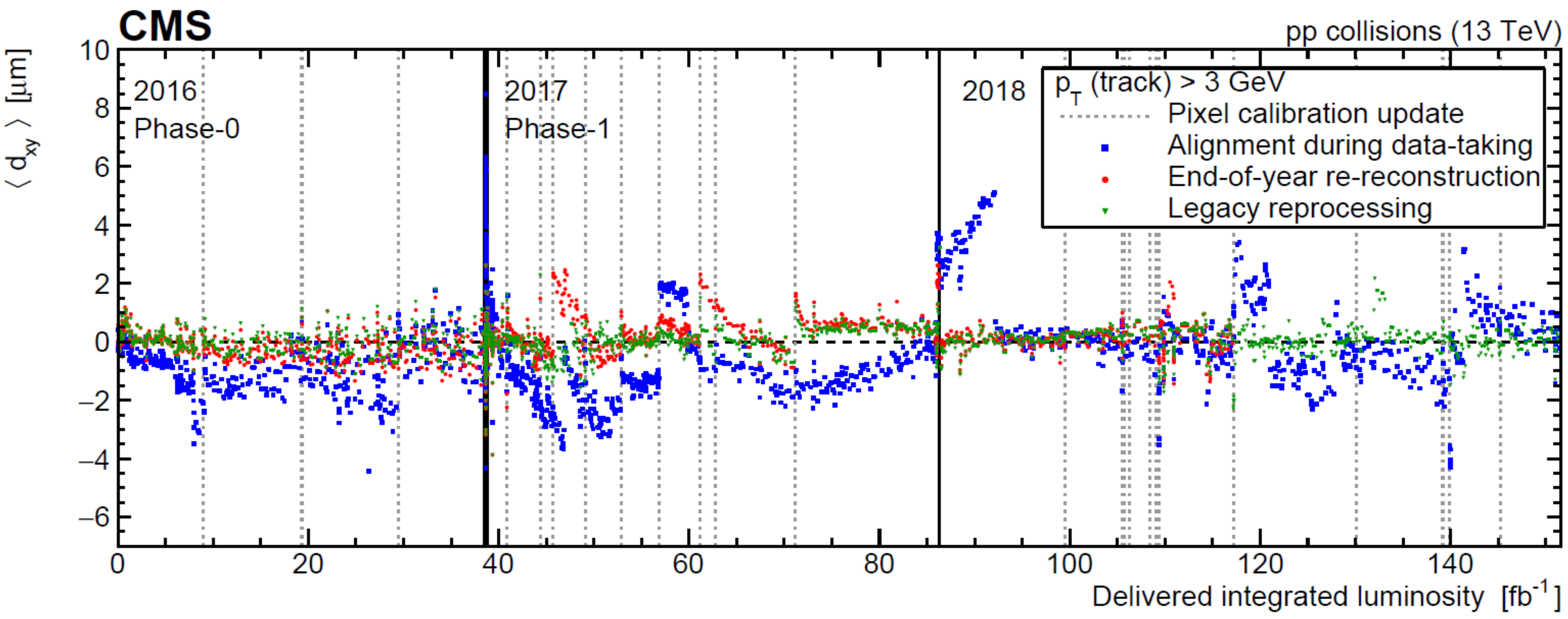
Impact parameter monitoring



- Measured by refitting a primary vertex with one track excluded, and evaluating the latter's impact parameter

- Initially, in early 2017 suboptimal tracking performance due to commissioning of new pixel tracker

➔ Generally very good performance after **legacy alignment**

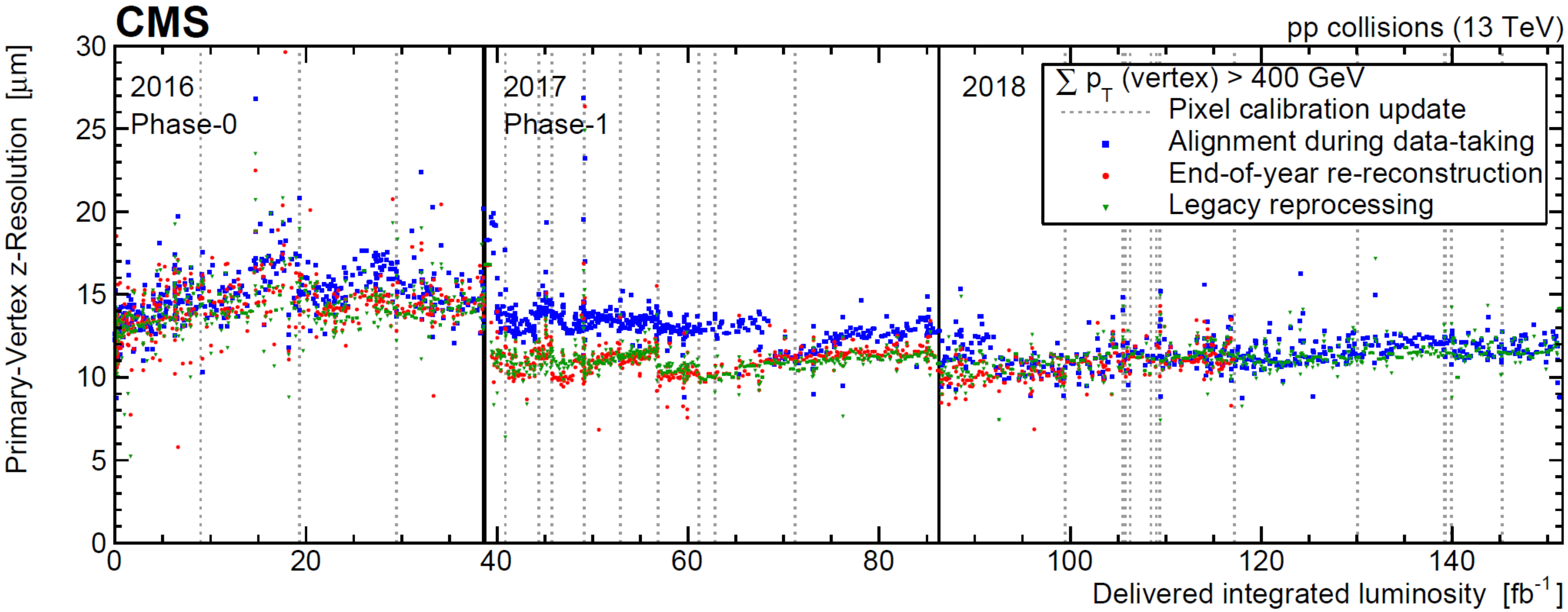


Primary vertex reconstruction performance

- Measured by splitting a primary vertex into two sub-vertices and studying the residuals

→ After proper alignment, visible improvement due to the new pixel tracker

- Outliers in prompt alignment: short IOV → suboptimal local pixel reconstruction configuration



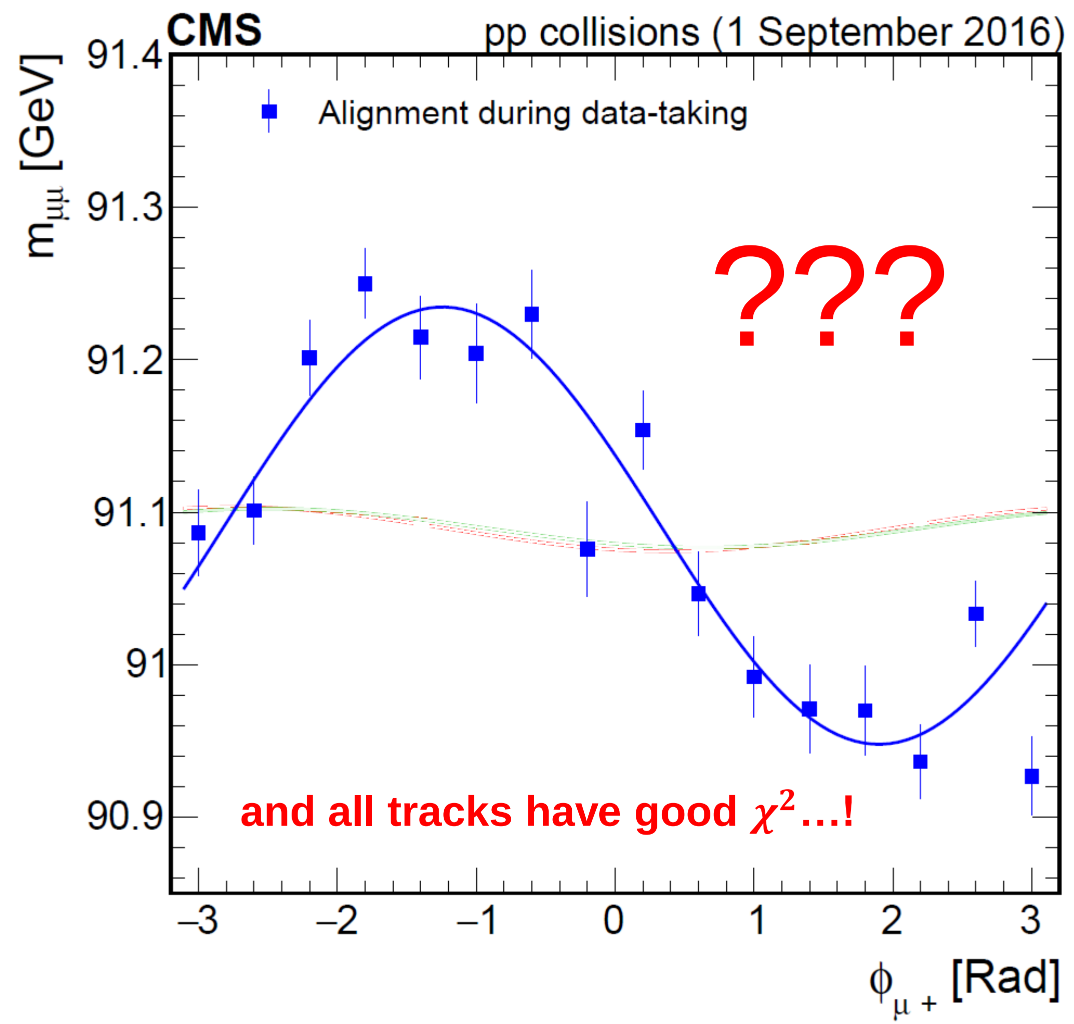
Systematics of misalignment: weak modes

What is going on here?

- Track-based alignment of trackers with a large number of individual modules (~17,000 in case of CMS) has potential for large systematic effects

- For example, in reconstructed $Z \rightarrow \mu\mu$ decays, position of the mass peak should not (!) depend on azimuth angle of a muon
  - “weak modes”

- **Control of weak modes is one of the greatest challenges in alignment**



What are weak modes?

- As mentioned, track-based alignment can be described as solving a huge linear equation system:
  - $$\underbrace{\begin{pmatrix} \square & \cdots & \square \\ \vdots & \ddots & \vdots \\ \square & \cdots & \square \end{pmatrix}}_{C'} \underbrace{\begin{pmatrix} \Delta p_1 \\ \vdots \\ \Delta p_n \end{pmatrix}}_{\text{alignment parameters}} = \underbrace{\begin{pmatrix} \square \\ \vdots \\ \square \end{pmatrix}}_{b'}$$

In CMS, C' is typically a matrix with 50,000 - 200,000 rows and columns

- The matrix C' reflects also the (inverse) **covariance matrix** of the alignment parameters
- In practice, we may find that some of the **eigenvalues** of this matrix are close to zero → infinite uncertainty
- The eigenvalues are associated to **eigenvectors** ("modes"), i.e. linear combinations of alignment parameters, that are only weakly constrained by our computation

➔ "weak modes"

➔ total χ^2 remains (almost) unchanged when this parameter combination is varied

But why does this happen…?

- A weak mode corresponds to a certain **geometry transformation** (= coherent set of alignment corrections)
- In track-based alignment, we detect misalignment by **incompatibility** of the reconstructed hit positions with the **track model**
- The geometry transformation of a weak mode is such that it transforms all valid tracks into other valid tracks

  ➔ track sample is **invariant** under this transformation

  ➔ no change of total χ^2

Helix track parameters:

QR: signed curvature radius
$\cot\theta$: dip angle
d_0: transverse impact parameter
z_0: longitudinal " "

- The helix trajectory in cylindrical coordinates (track from origin, assuming $d_0 = z_0 = \phi_0 = 0$):

  $r = -2\,QR\,\sin\phi \approx -2\,QR\,\phi$

  $z = -2\,QR\,\phi\,cot\,\theta$

➔ Within validity of $\sin\phi \approx \phi$ approximation, **any linear transformation** in (r, ϕ, z) space results in a weak mode

Classification and diagnosis of weak modes

For collision tracks

	Δz	Δr	$\Delta \phi$
vs. z	*z expansion* $\Delta z = \epsilon z$ overlap	*bowing* $\Delta r = \epsilon r(z_0^2 - z^2)$ overlap	*twist* $\Delta \phi = \epsilon z$ $Z \rightarrow \mu\mu$
vs. r	*telescope* $\Delta z = \epsilon r$ cosmics	*radial* $\Delta r = \epsilon r$ overlap	*layer rotation* $\Delta \phi = \epsilon r$ cosmics
vs. ϕ	*skew* $\Delta z = \epsilon \cos(\phi + \phi_0)$ cosmics	*elliptical* $\Delta r = \epsilon r \cos(2\phi + 2\phi_0)$ cosmics	*sagitta* $\Delta \phi = \epsilon \cos(\phi + \phi_0)$ cosmics

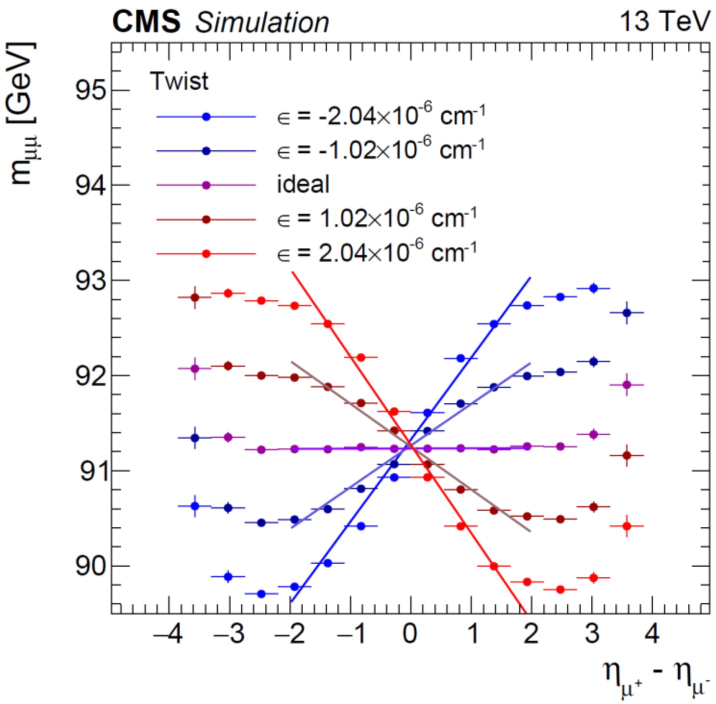
- Overlap validation: check relative hit positions in sensor overlaps (not shown)
- Cosmics validation: split cosmic muon track by hemispheres, compare parameters of sub-tracks
- $Z \rightarrow \mu\mu$ validation: check for dependence of Z mass peak on muon parameters

Classification and diagnosis of weak modes (cont'd)

Test with simulation: a few examples

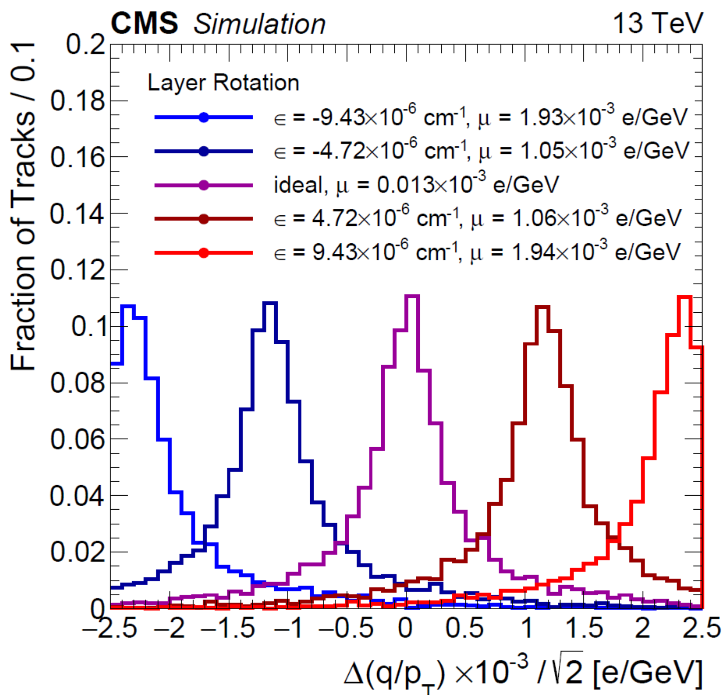
$Z \to \mu\mu$ validation

Twist: $m_{\mu\mu}$ vs $\Delta\eta$

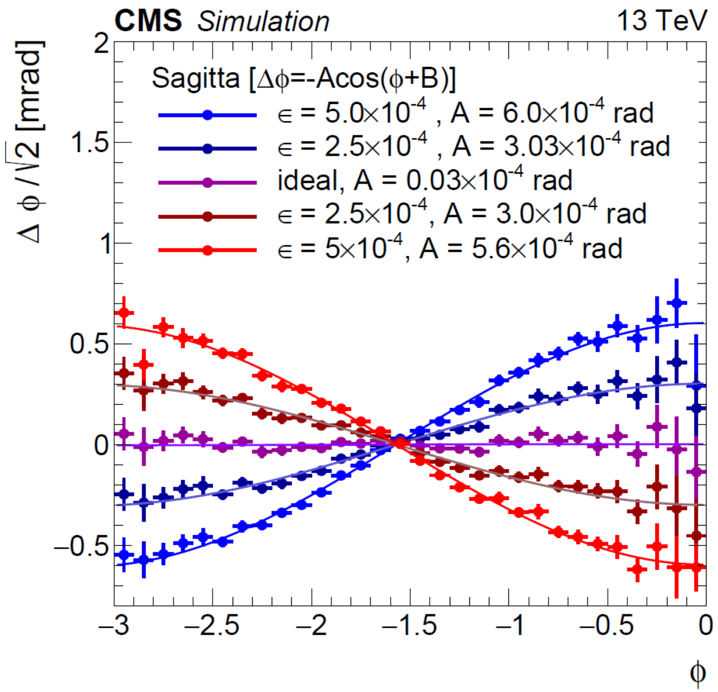


Cosmic ray track validation („cosmic splitting")

Layer rotation: distribution of $\Delta(\frac{q}{p_T})$



Sagitta: $\Delta\phi$ vs ϕ



➔ Demonstrates the power of $Z \to \mu\mu$ and cosmic ray events to identify & control weak modes

How to control weak modes: a strategy

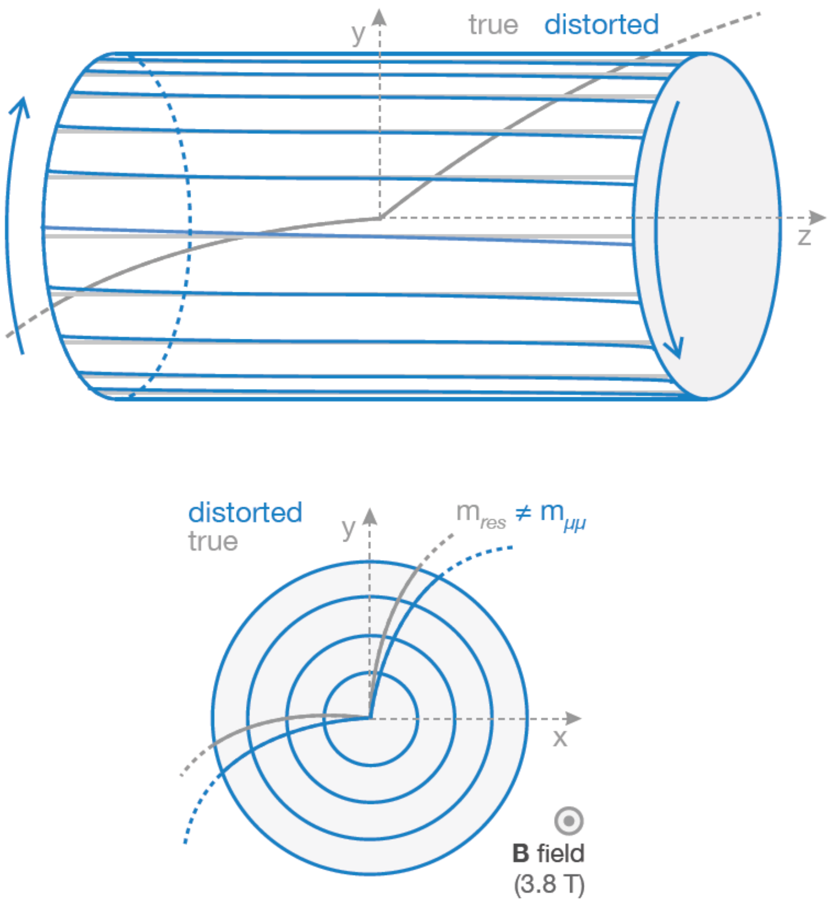
- Include tracks in the alignment which do not pass through the detector center

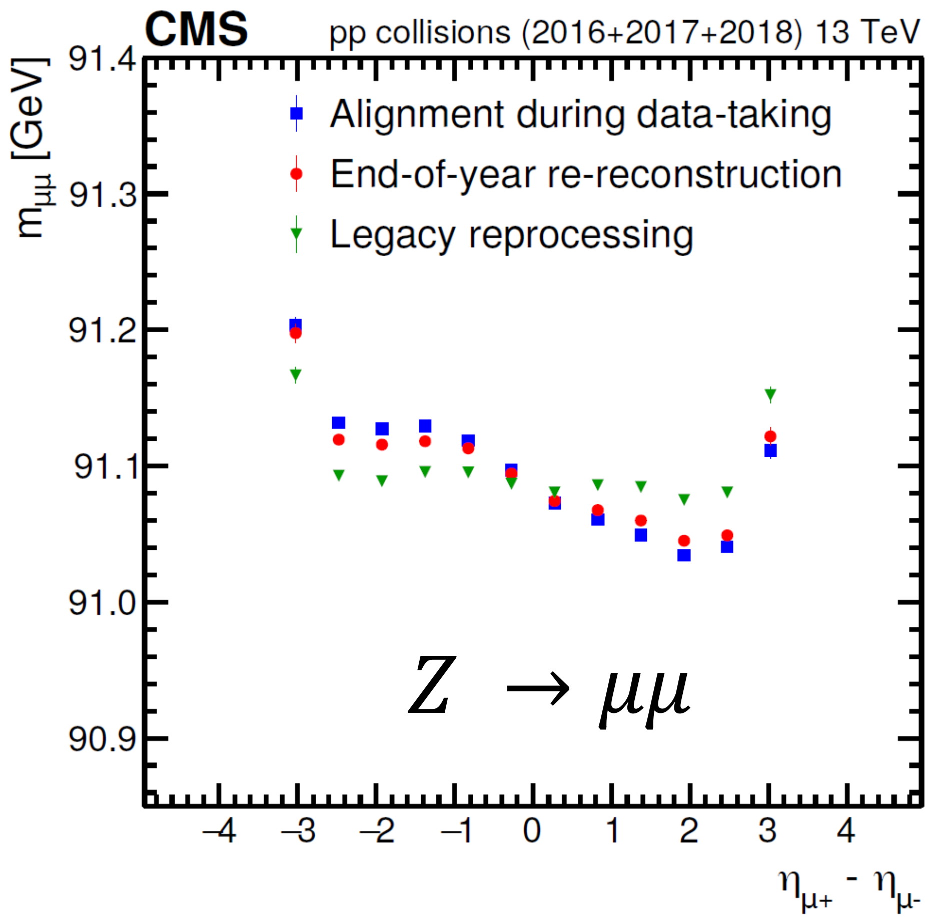
  ➔ cosmic muons, recorded both with magnetic field on and off

- Include track combinations having mass and vertex constraints
  - $Z \to \mu\mu$
  - $\Upsilon(1S) \to \mu\mu$

- If all else fails: apply counter-transformation in form of a constraint

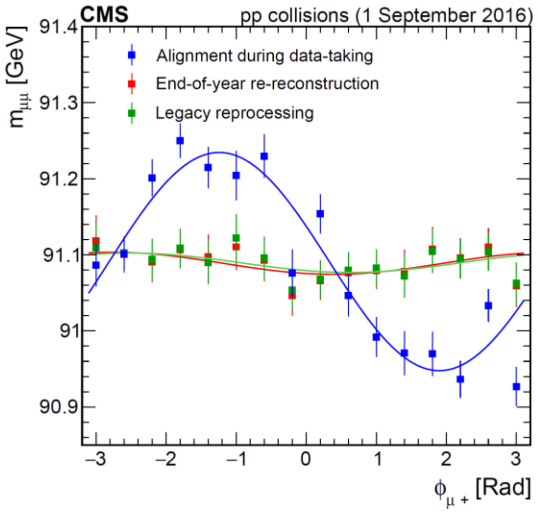
Practical example: correction of a twist weak mode

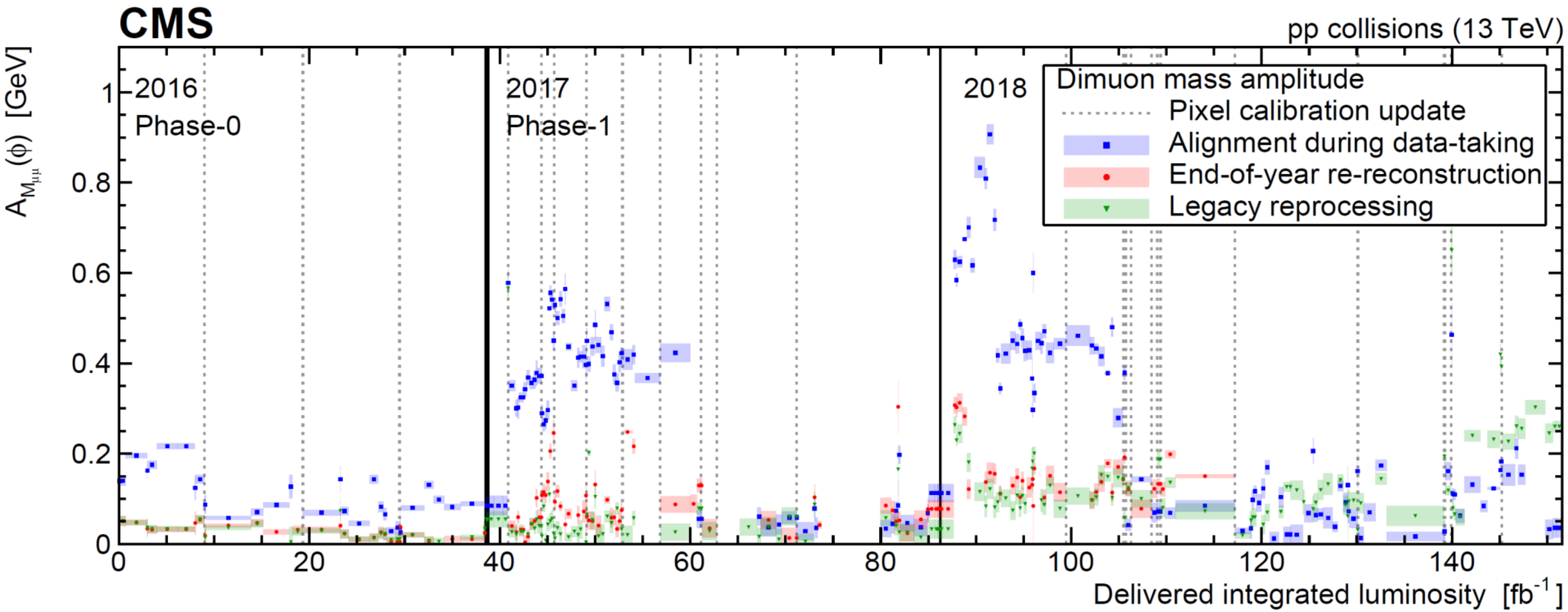




- Sizable twist in alignment during data-taking → resolved in legacy alignment

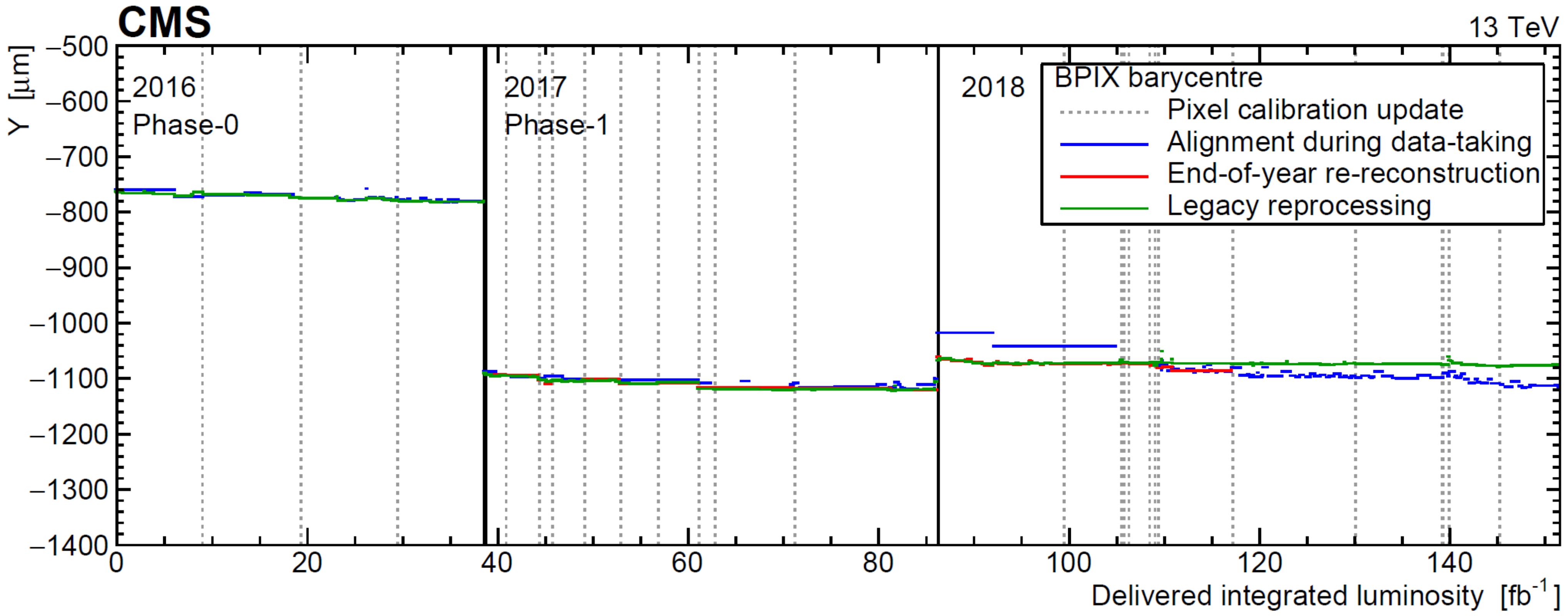
Dimuon mass validation: evolution with time





➔ Large initial amplitudes in data-taking alignment are resolved in the end-of-year and legacy alignment

Barycenter of barrel pixel detector



➔ Very good stability (at level of few microns)

➔ Changes in winter shutdowns due to (re-)insertions of pixel tracker

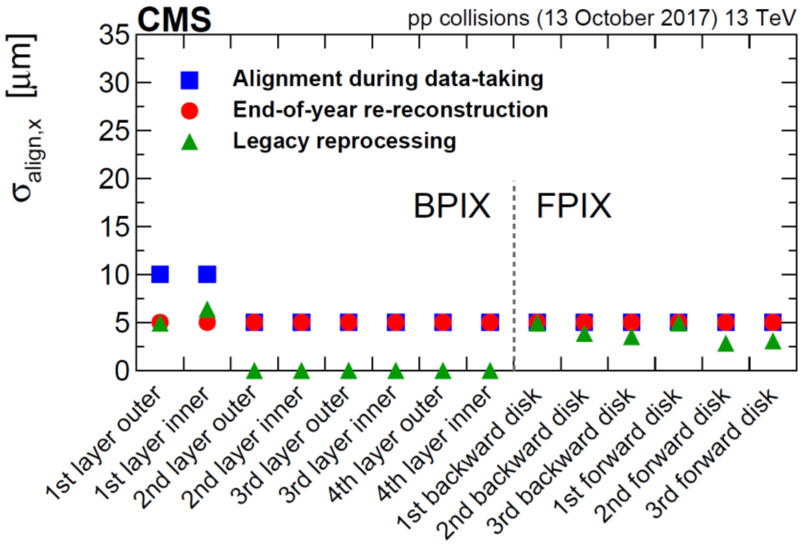
➔ Reprocessing cures an artificial drop due to radiation damage effects

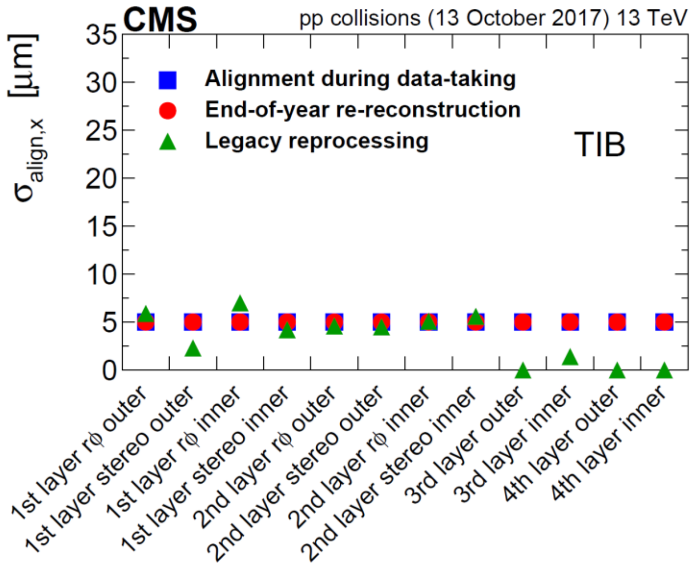
How precise are the alignment parameters?

APU = Alignment parameter uncertainty

- Direct error estimation by matrix inversion usually not feasible, since matrix too large

- Obtained by studying distributions of normalized residuals: $\frac{x'_{hit} - x'_{track}}{\sigma}$, where $\sigma = \sqrt{\sigma_{hit}^2 + \sigma_{track}^2 + \sigma_{align}^2}$
  - adjust σ_{align} such that distributions become unit normal → iterative procedure

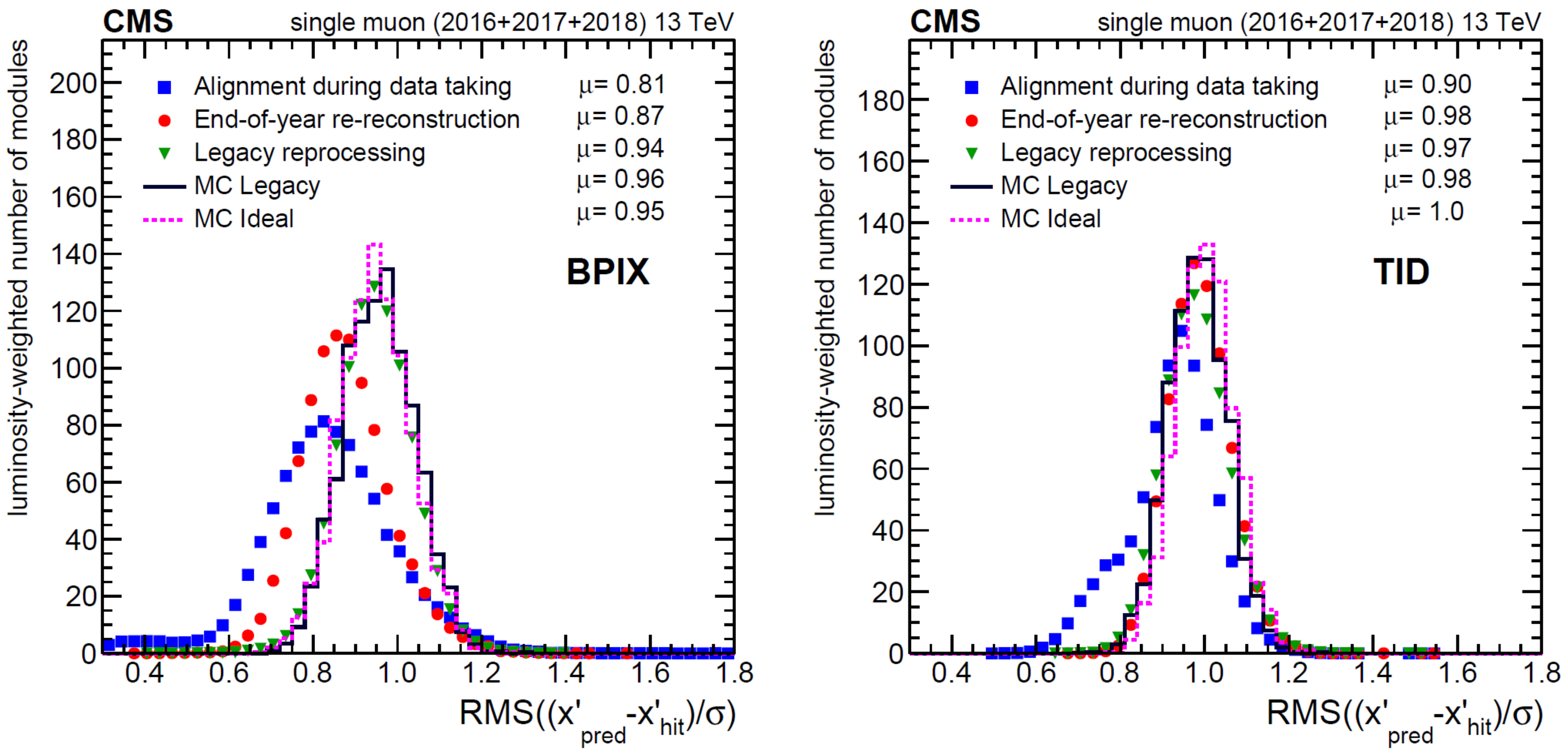




→ Very good control of alignment precision

A direct look at normalized residuals

- Important test: check RMS width of the normalized residuals
- After the legacy alignment, it is centered close to 1, and agrees well with MC
  - ➔ shows both correct alignment and correct assignment of alignment parameter uncertainties



Summary

- Alignment is a key driver for physics performance
- Methodology has evolved enormously to a new level, to meet LHC challenges
- Powerful alignment workflows are in place
  - still a huge effort year by year; always new challenges surfacing
- For Run 3, first alignments have already been produced from cosmic runs, and even first collisions
  - start thinking about alignment Phase 2 tracker

⇨ Alignment is not static…

it continues to be challenging… *and interesting!*

Further reading

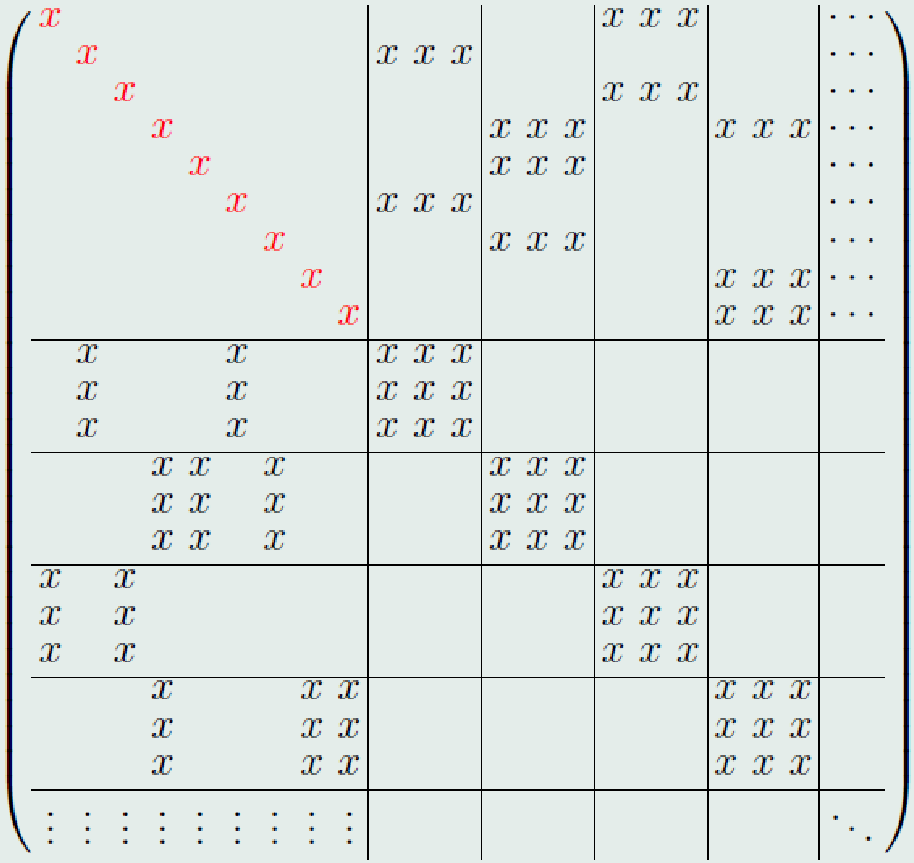
- V. Blobel and C. Kleinwort, "A New method for the high precision alignment of track detectors", https://inspirehep.net/conferences/973991, https://arxiv.org/abs/hep-ex/0208021
- CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", 2014 *JINST* **9** P10009
- CMS Collaboration, "Alignment of the CMS tracker with LHC and cosmic ray data", 2014 JINST 9 P06009
- CMS Collaboration, "Strategies and performance of the CMS silicon tracker alignment during LHC Run 2", arxiv:2111.08757 (2021), accepted for publication in NIM A
- R. Mankel, "Pattern recognition and event reconstruction in particle physics experiments", Rept.Prog.Phys. 67 (2004) 553

Backup

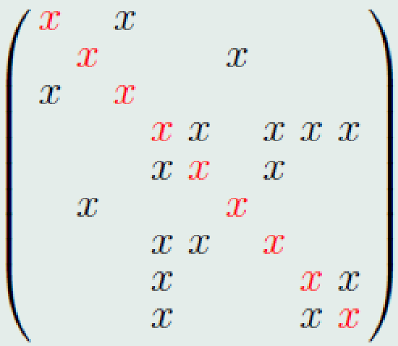
The sparse matrix C of a simultaneous fit of alignment parameters (global) and track parameters (local) is a large matrix, that can be reduced to a smaller matrix for the alignment parameters only using Schur complements (no approximation!).

- **The matrix C^{total}, a 8 030 100 × 8 030 100 matrix (several 100 Tera Bytes) . . .**

$C^{\text{total}} =$



$\longrightarrow \quad C^{\text{global}} =$



Element $(C^{\text{global}})_{jk} \neq 0$, if parameters j and k in same local fit.

Note: the inverse of a sparse matrix (= covariance matrix) is dense; **all parameters are correlated!**

. . . is reduced to a (sparse) 47 655 × 47 655 matrix C^{global} for the global parameters.

NIM A 461 (2001) 162–167

•

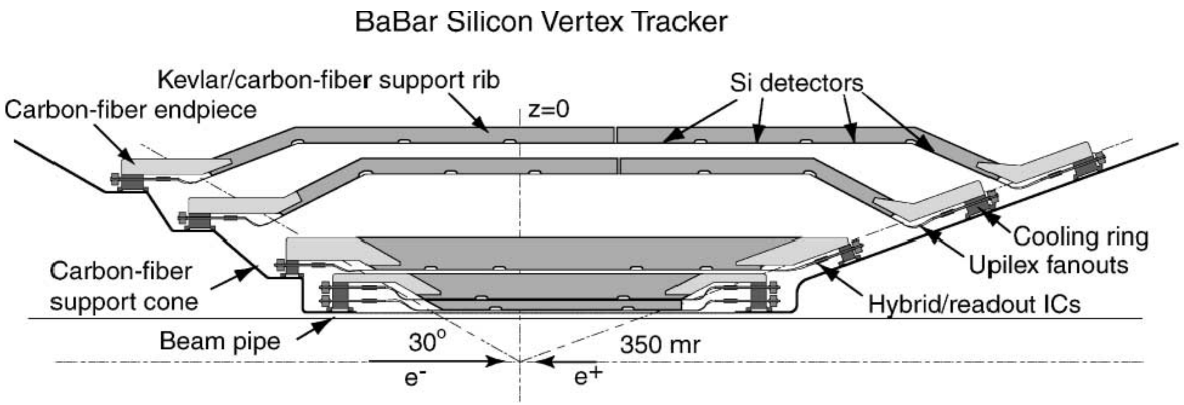


Fig. 1. SVT layout: *rz* cross-sectional view. The modules of layers 4 and 5 are "bent" towards the beam axis to increase angular coverage and to reduce the crossing angle of low-angle tracks. Note the asymmetry of the detector with respect to $z = 0$.

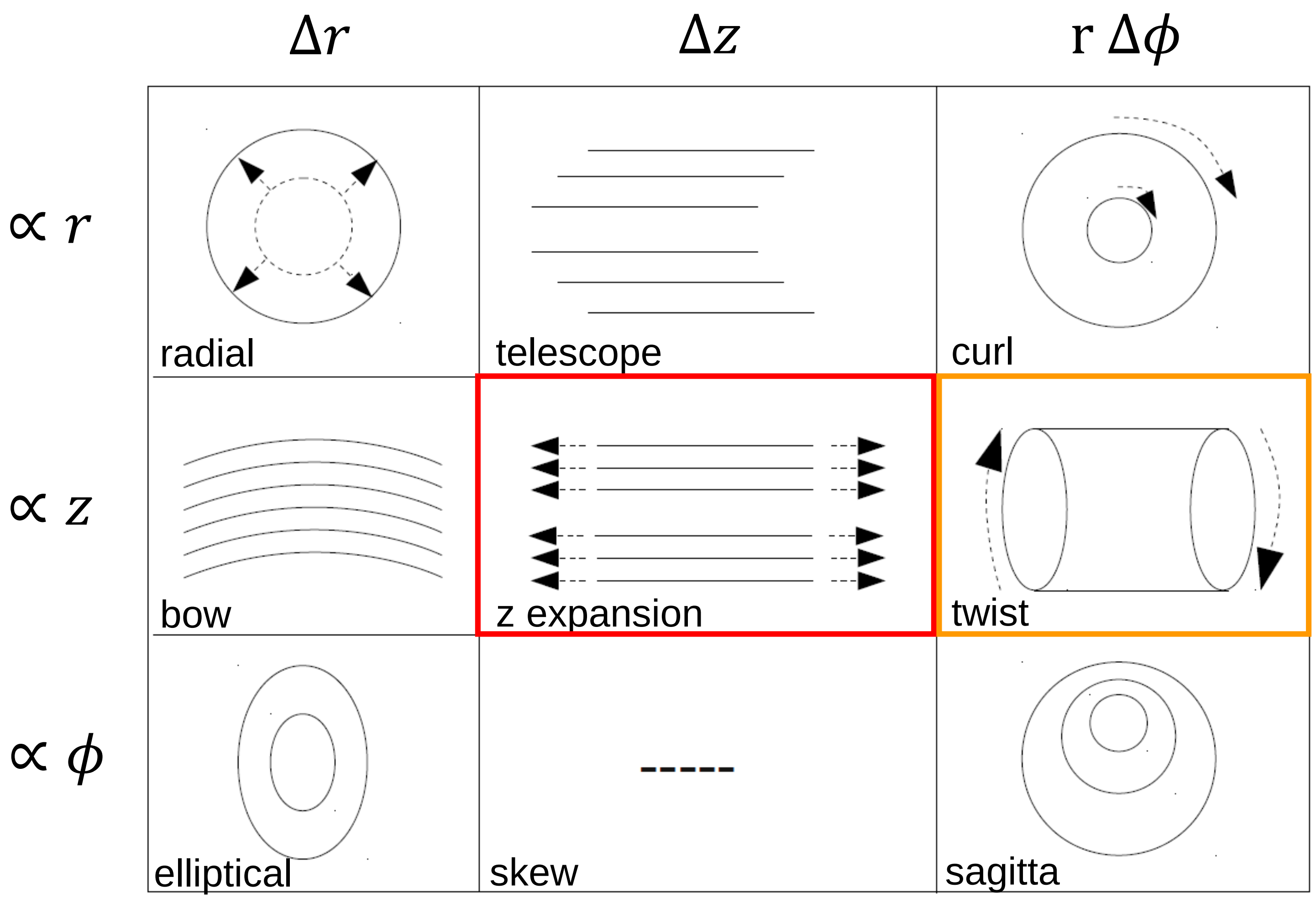
Table 1
Layer structure of the BaBar SVT

Layer	Radius (mm)	Modules/layer	Si Wafers/module	ϕ pitch (μm)	z pitch (μm)
1	32	6	4	50 or 100	100
2	40	6	4	55 or 110	100
3	54	6	6	55 or 110	100
4a	124	8	7	100	210
4b	127	8	7	100	210
5a	140	9	8	100	210
5b	144	9	8	100	210

Classification of weak modes

For collision tracks

-



Adapted from: Alessio Bonato, https://indico.cern.ch/getFile.py/access?contribId=11&sessionId=2&resId=0&materialId=slides&confId=137973

Weak modes and track parameter transformations

Twist:

- $\phi \quad \rightarrow \phi + k\,z$

➔ $\cot\theta \rightarrow \cot\theta$

➔ $\frac{1}{QR} \quad \rightarrow \frac{1}{QR} - 2k\ \cot\theta$

r expansion:

- r $\quad \rightarrow r + k\,r$

➔ $\cot\theta \rightarrow \frac{1}{1+k}\cot\theta$

➔ $QR \quad \rightarrow (1+k)\,QR$

z expansion:

- z $\quad \rightarrow z + k\,z$

➔ $\cot\theta \rightarrow (1+k)\ \cot\theta$

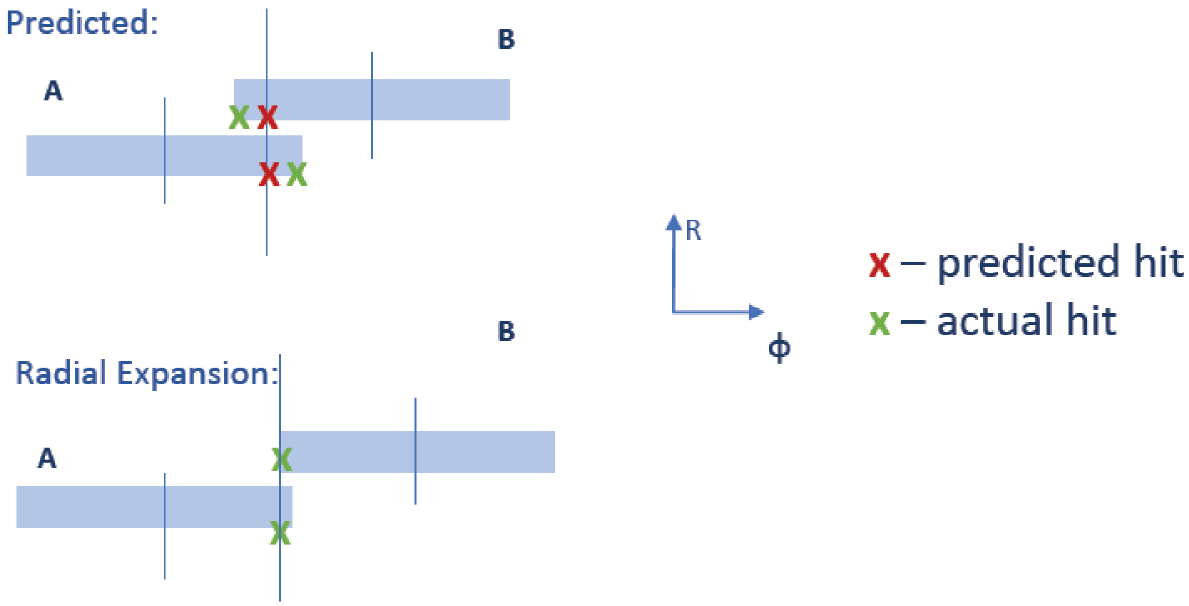
➔ $QR \quad \rightarrow QR$

- In general, weak modes cause track parameters (momentum, direction) to change

➔ **affect physics**

Overlap validation

Overlap validation (radial and z expansion, bowing)



- Millepede-II timing

Table 3: Examples of PEDE wall time (time taken from start of the program to end) for some larger alignment campaigns on a dedicated test machine (Intel Xeon E5-2667 @ 3.2 GHz, 256 GB memory @ 51 GB/s).

Number of global parameters	Number of constraints	Number of records	Matrix size [GB] (sparse)	Wall time [s] (10 threads)
217500	138	4.46×10^7	44	8.4×10^3
213900	1782	2.90×10^7	85	6.8×10^3
576000	942	5.20×10^7	218	4.4×10^4