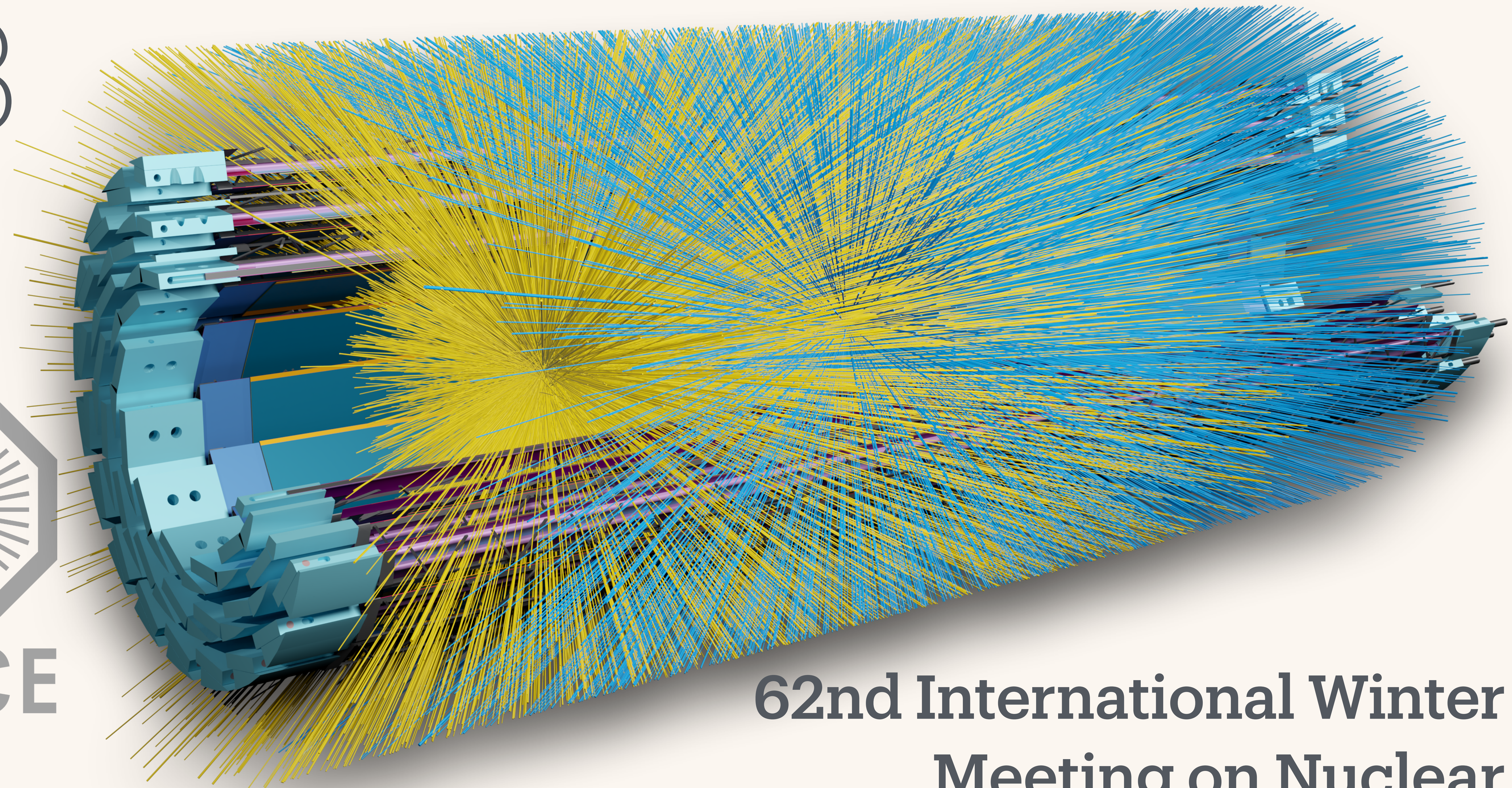
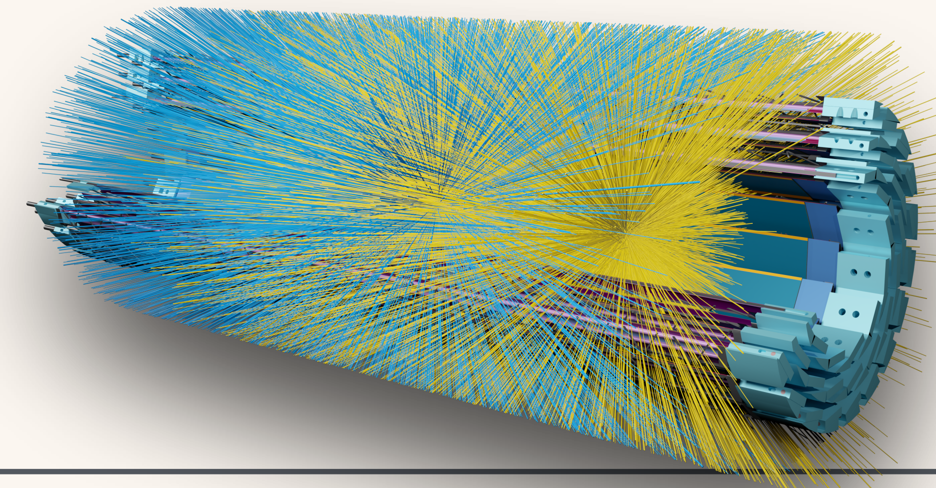
3+1D reconstruction of the Inner Tracking System 2 for ALICE Run 3

On behalf of the
ALICE Collaboration



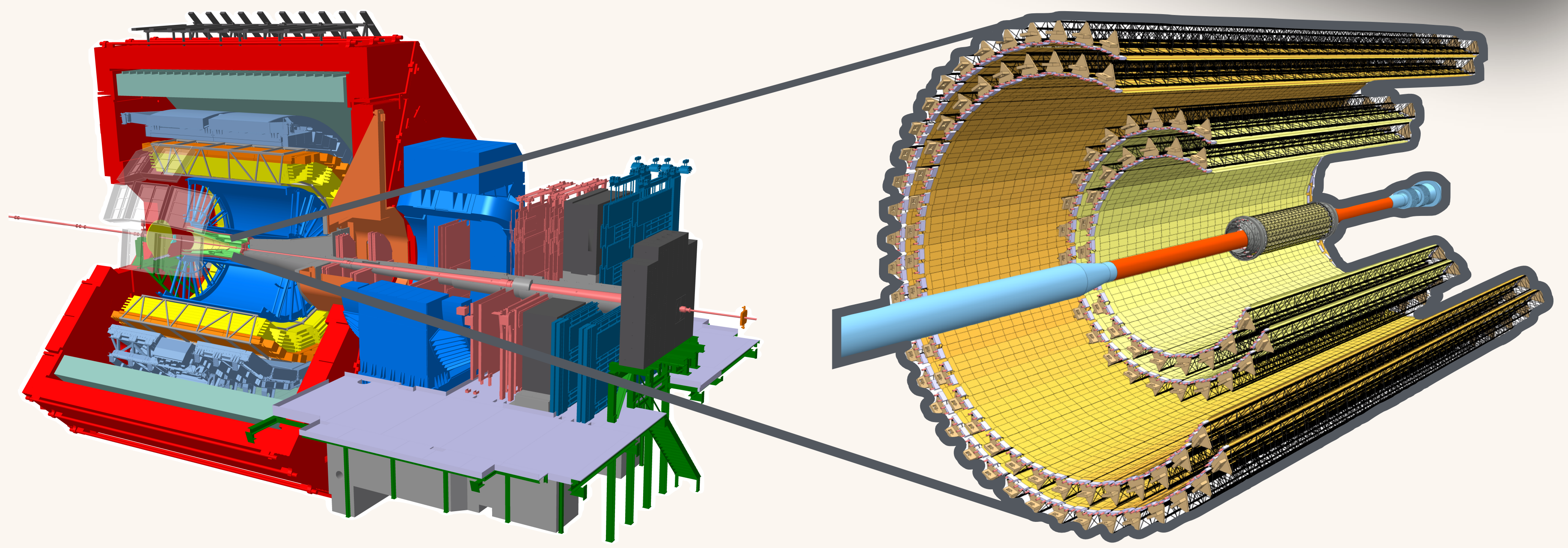
Felix Schlepper
(CERN / Heidelberg University)

62nd International Winter Meeting on Nuclear Physics Bormio 2026

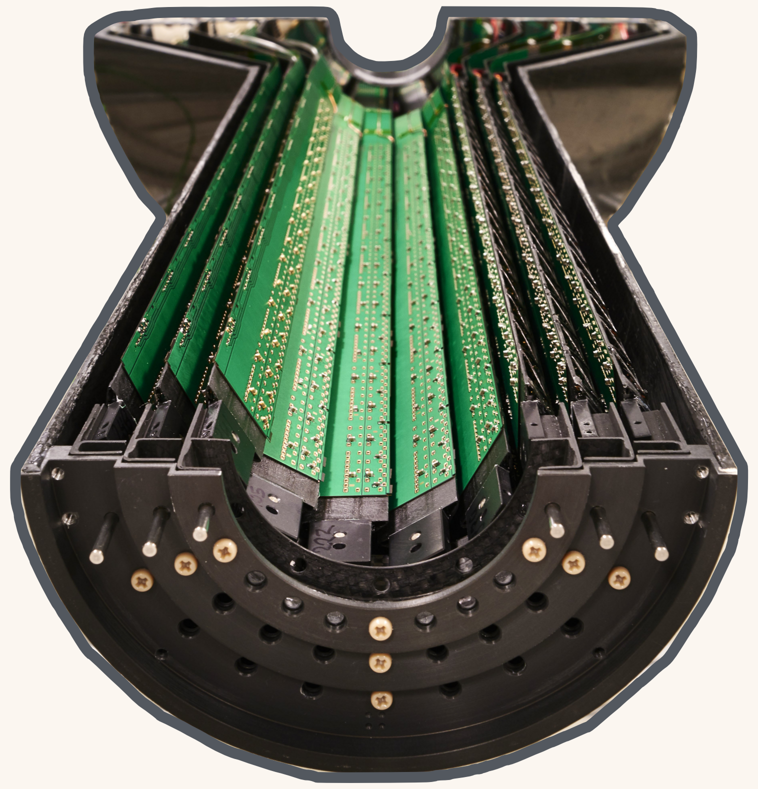


Inner Tracking System 2

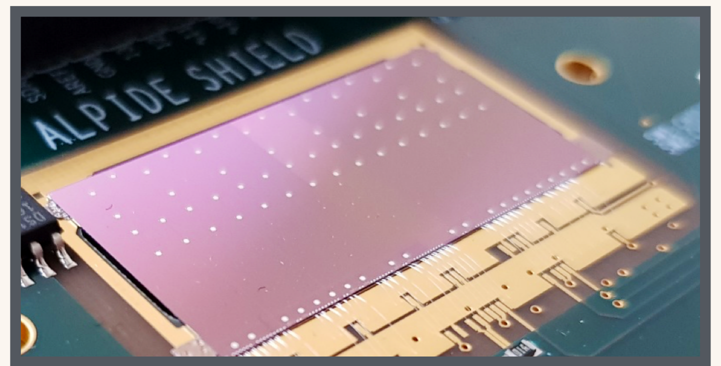
ALICE Upgrade for LS2



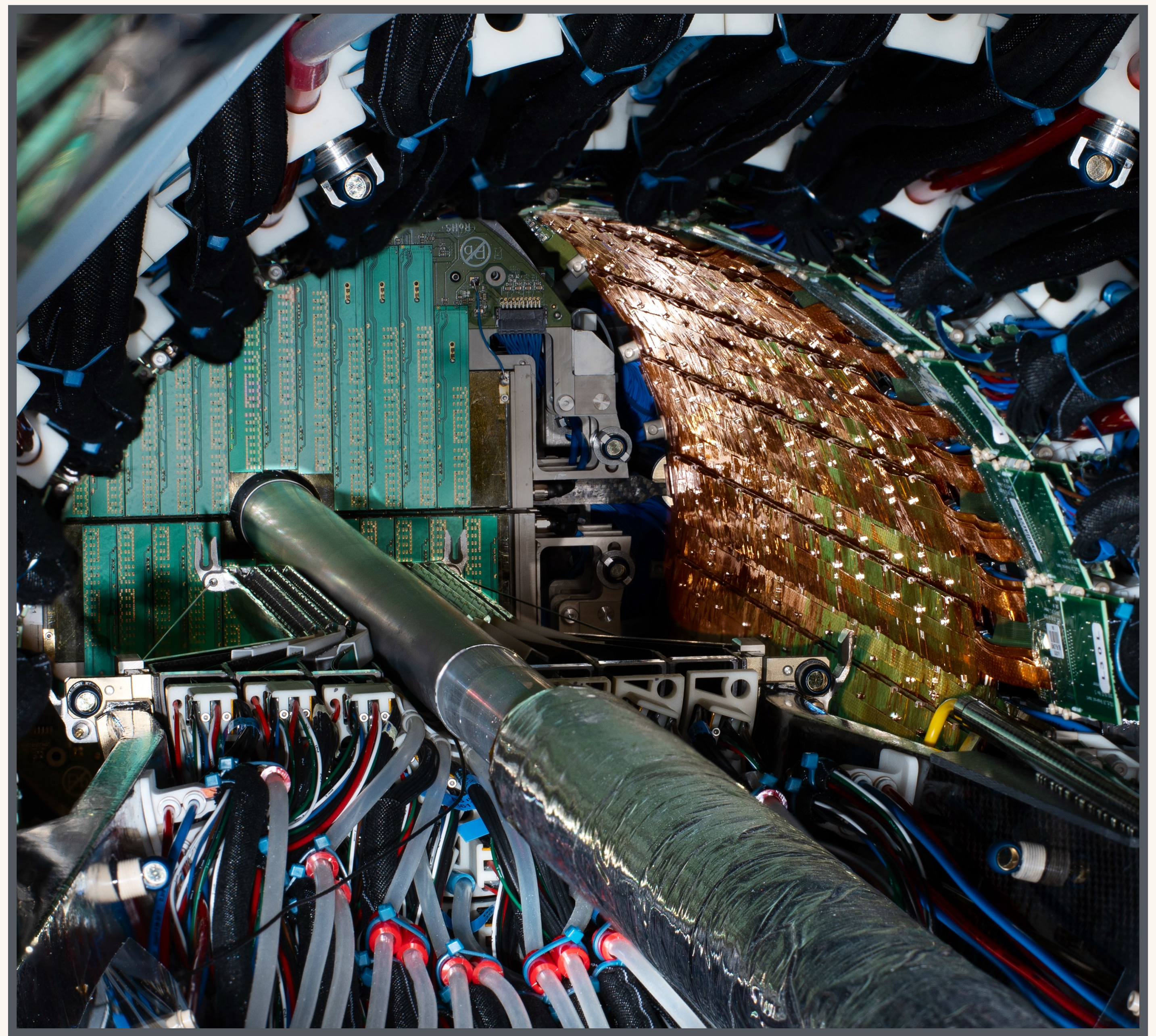
- Enabling high-precision measurements of rare probes down to zero momentum
  - Continuous readout of 50 kHz Pb-Pb collisions
- 7 layers starting at R=2.2 cm up to 40 cm
- 10 m² active silicon area, 12.5 GigaPixels
- 180 nm CMOS MAPS (Monolithic Active Pixel Sensors) 15x30 mm², 512x1024 pixels, O(30μm)



Inner Barrel

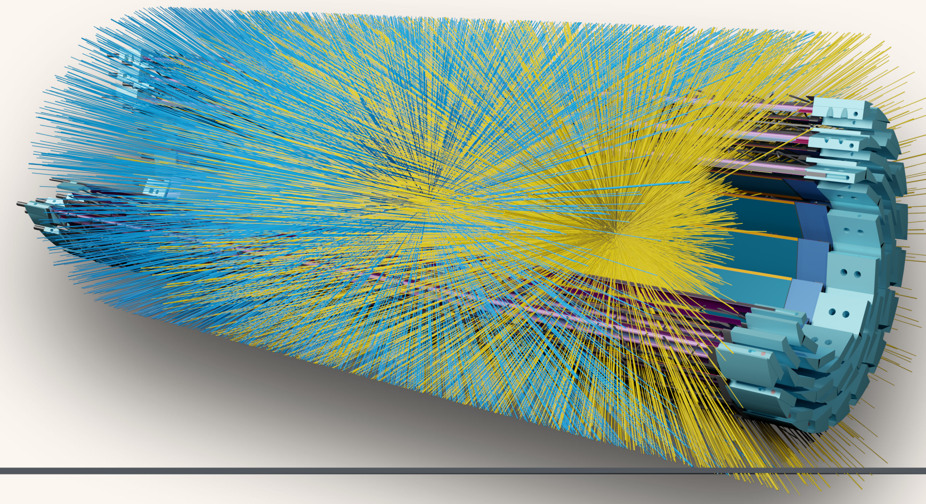


ALPIDE

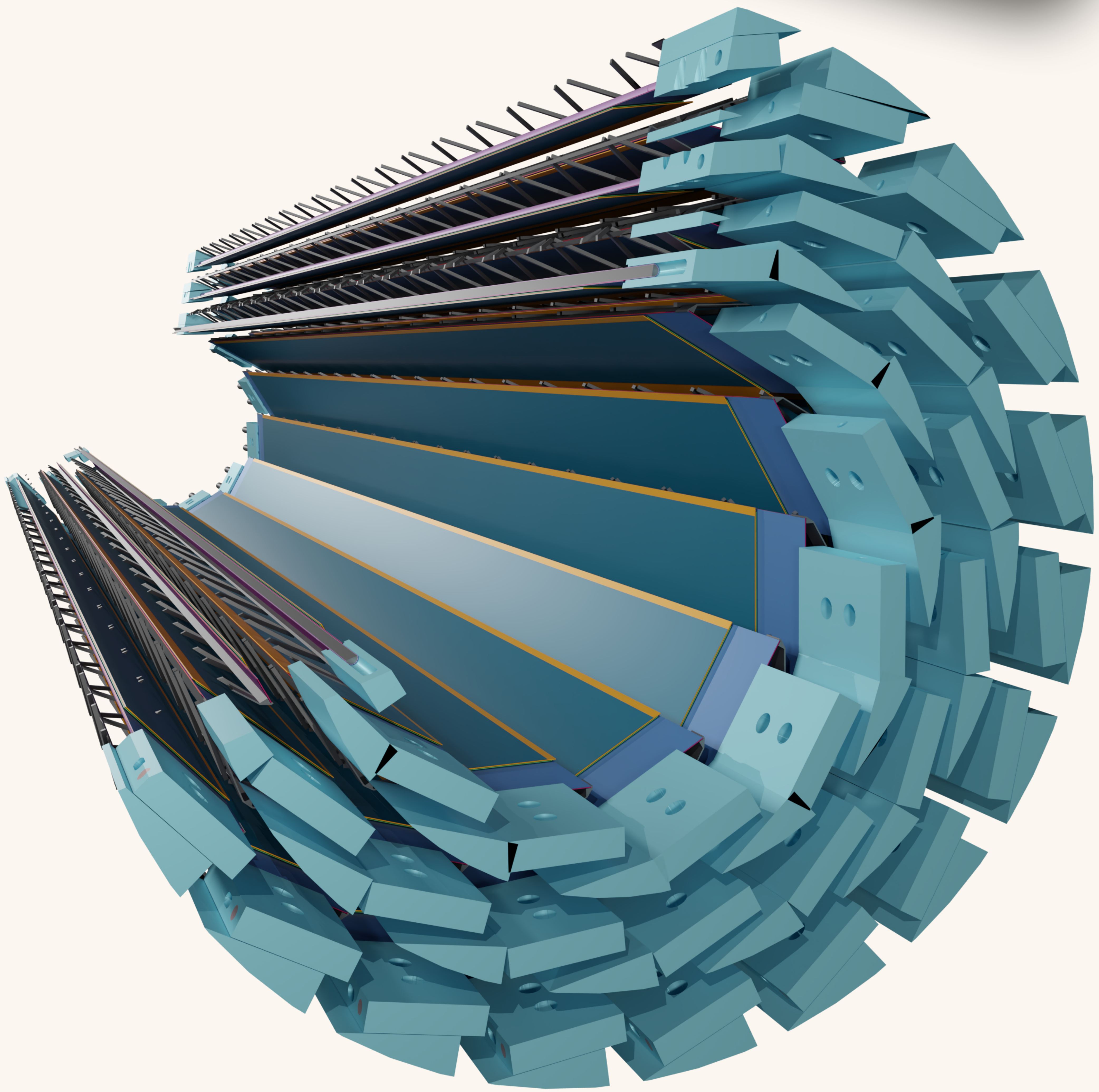


Inner Tracking System 2

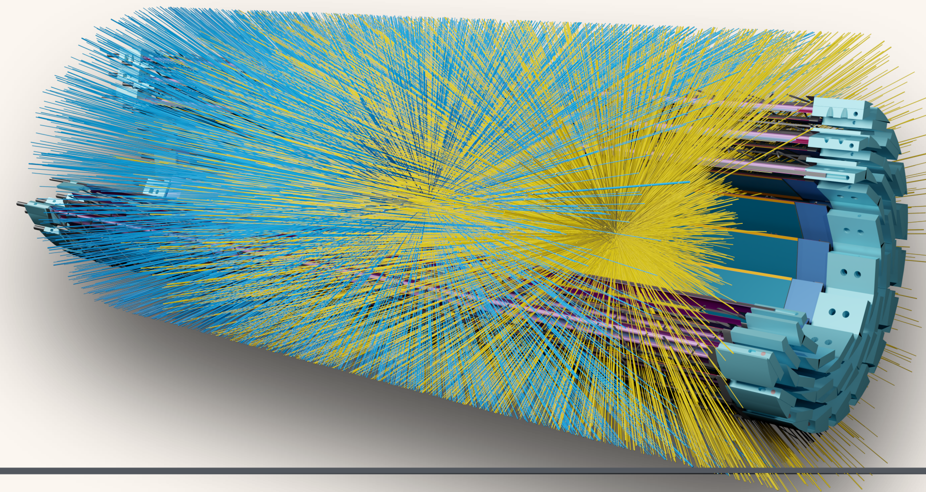
Goal of reconstruction



Recipe for pixel detectors

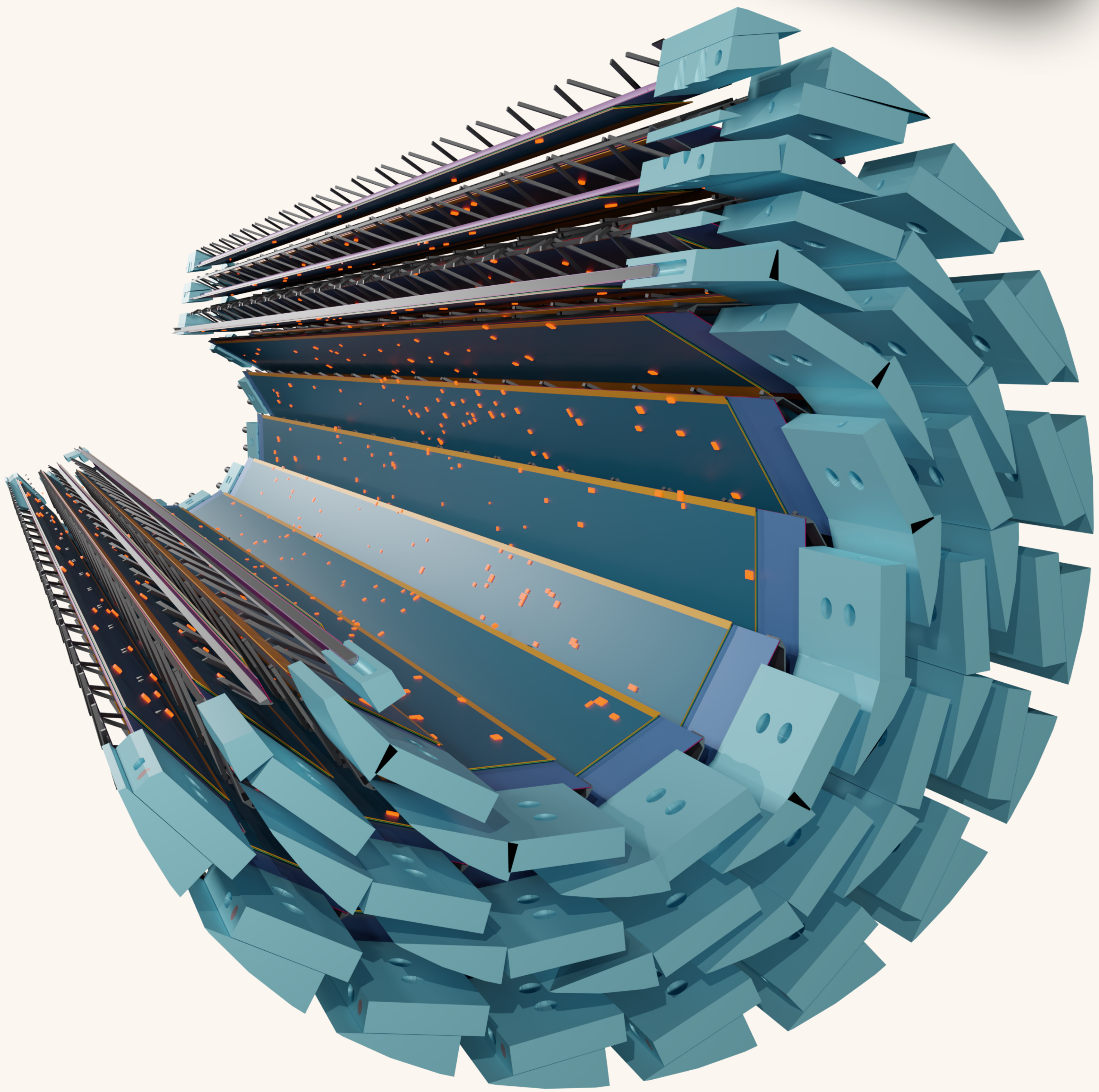


Goal of reconstruction

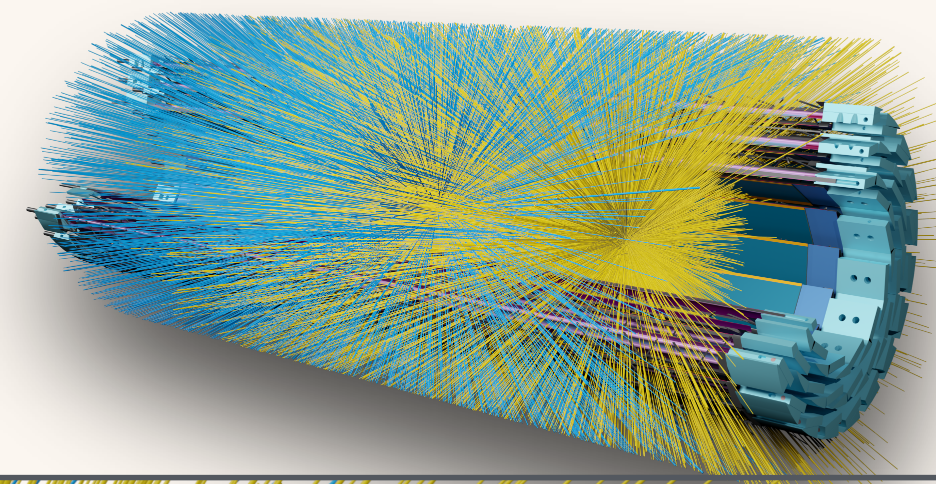


Recipe for pixel detectors

- From digits to clusters: group individual digits into clusters (2D clustering)
- Transform clusters into 3D space points
  - Assuming precisely known position of a chip (alignment)
  - Assign timestamp (from the readout) + position resolution to the point

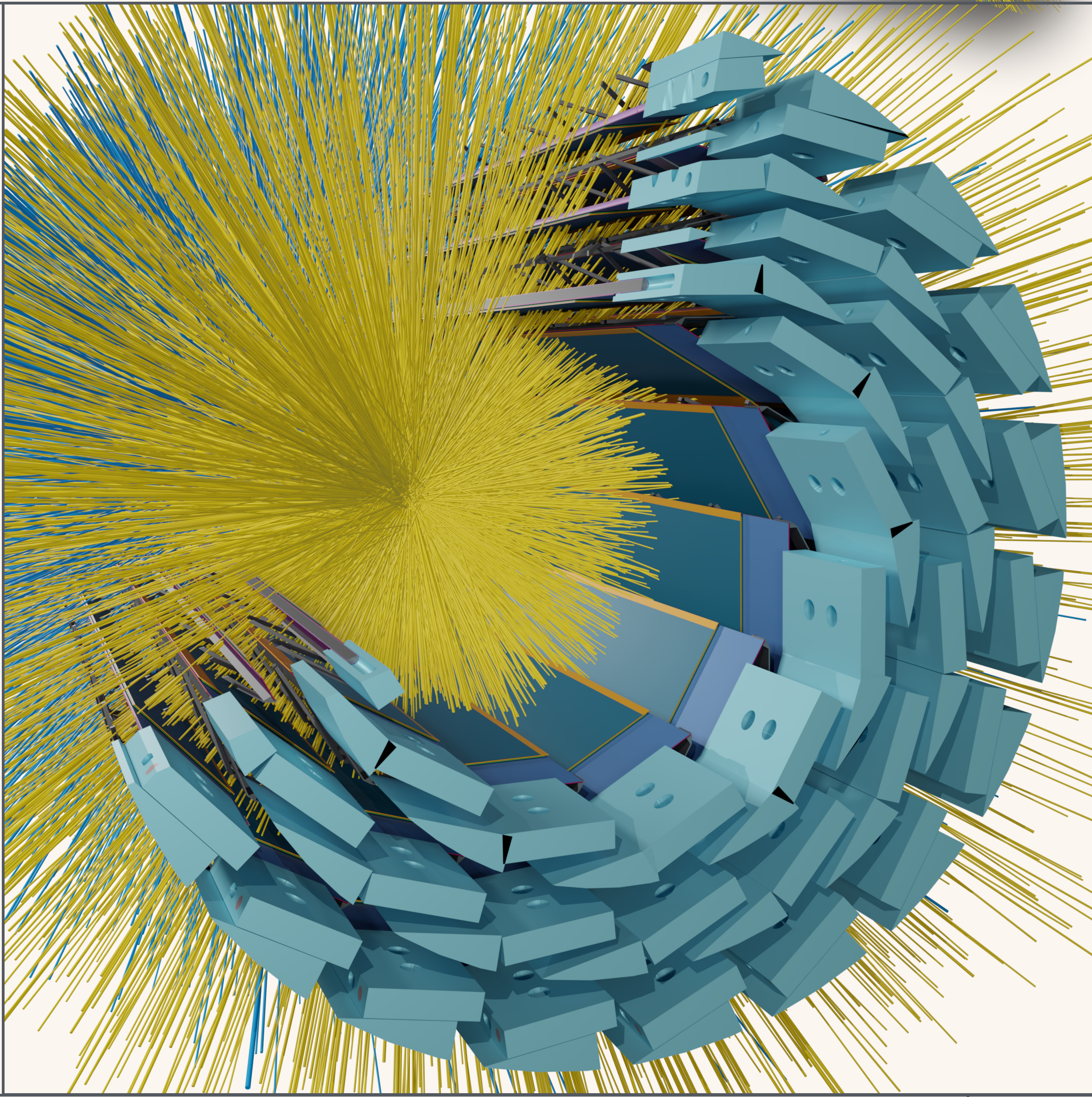


Goal of reconstruction



Recipe for pixel detectors

- From digits to clusters: group individual digits into clusters (2D clustering)
- Transform clusters into 3D space points
  - Assuming precisely known position of a chip (alignment)
  - Assign timestamp (from the readout) + position resolution to the point
- Connect clusters across detector layers to form tracks

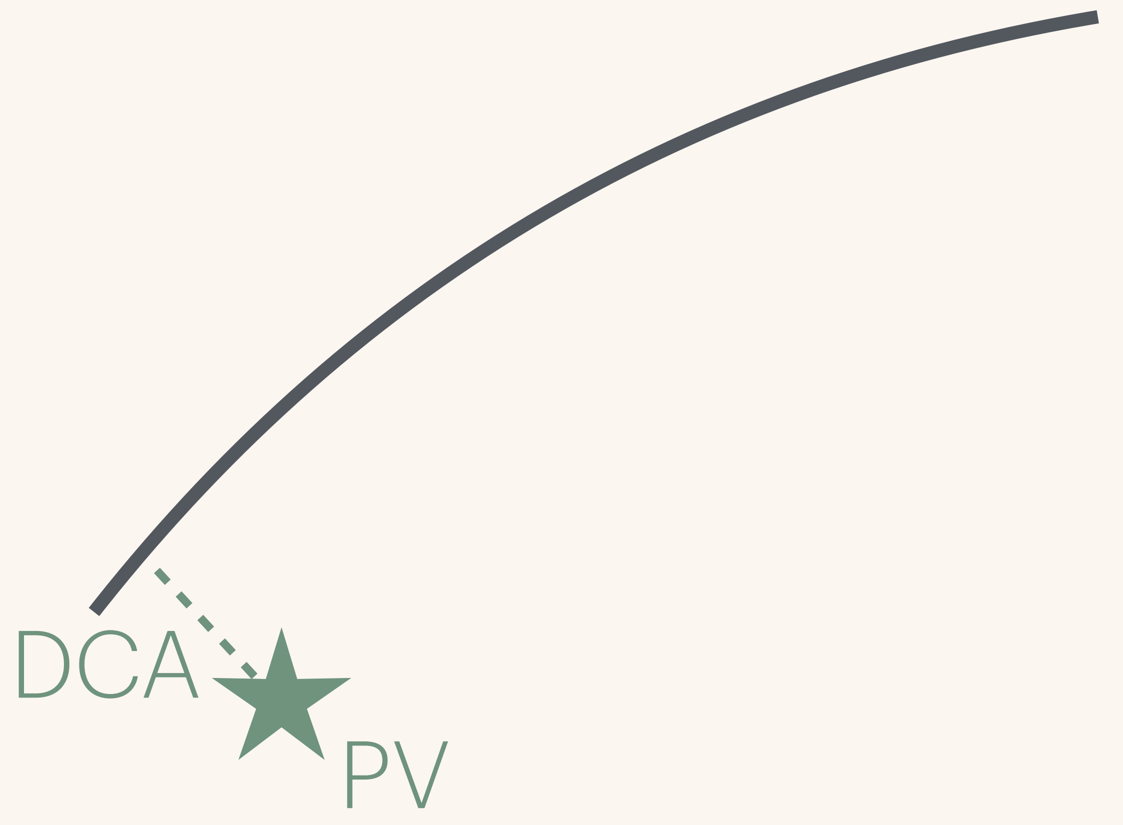


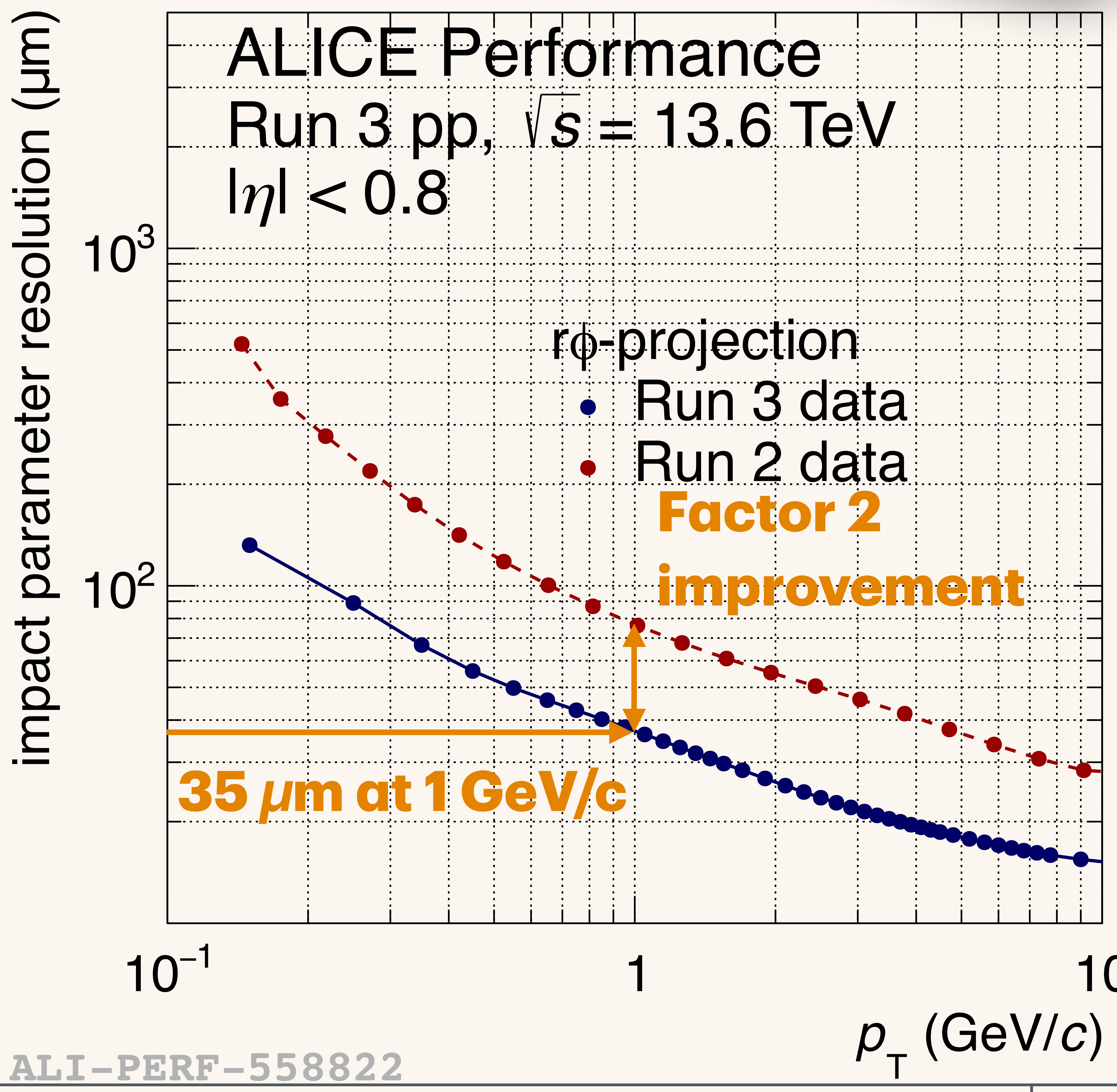
Principle goal is to determine the kinematic properties of the particles with good precision!

Performance

Single track resolution

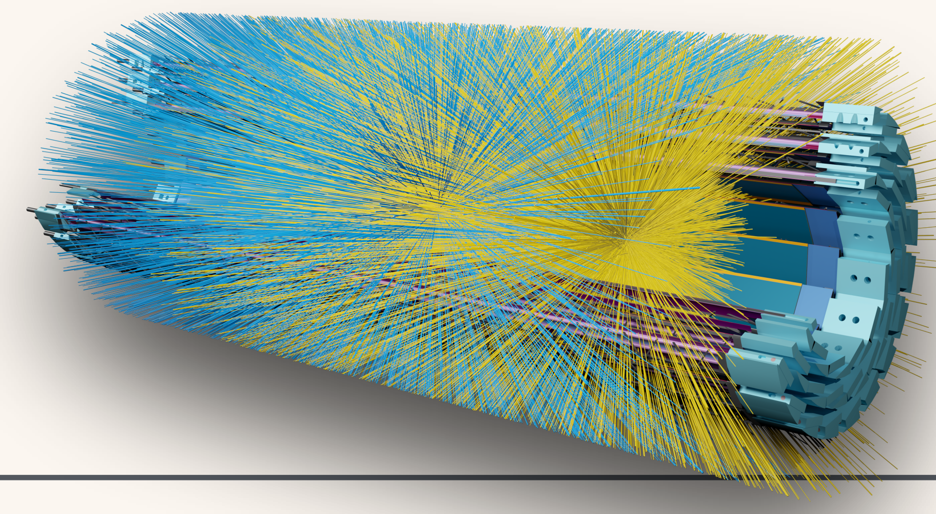
- First layer starts at innermost radius 22 mm from the nominal interaction point (39 mm for ITS1)





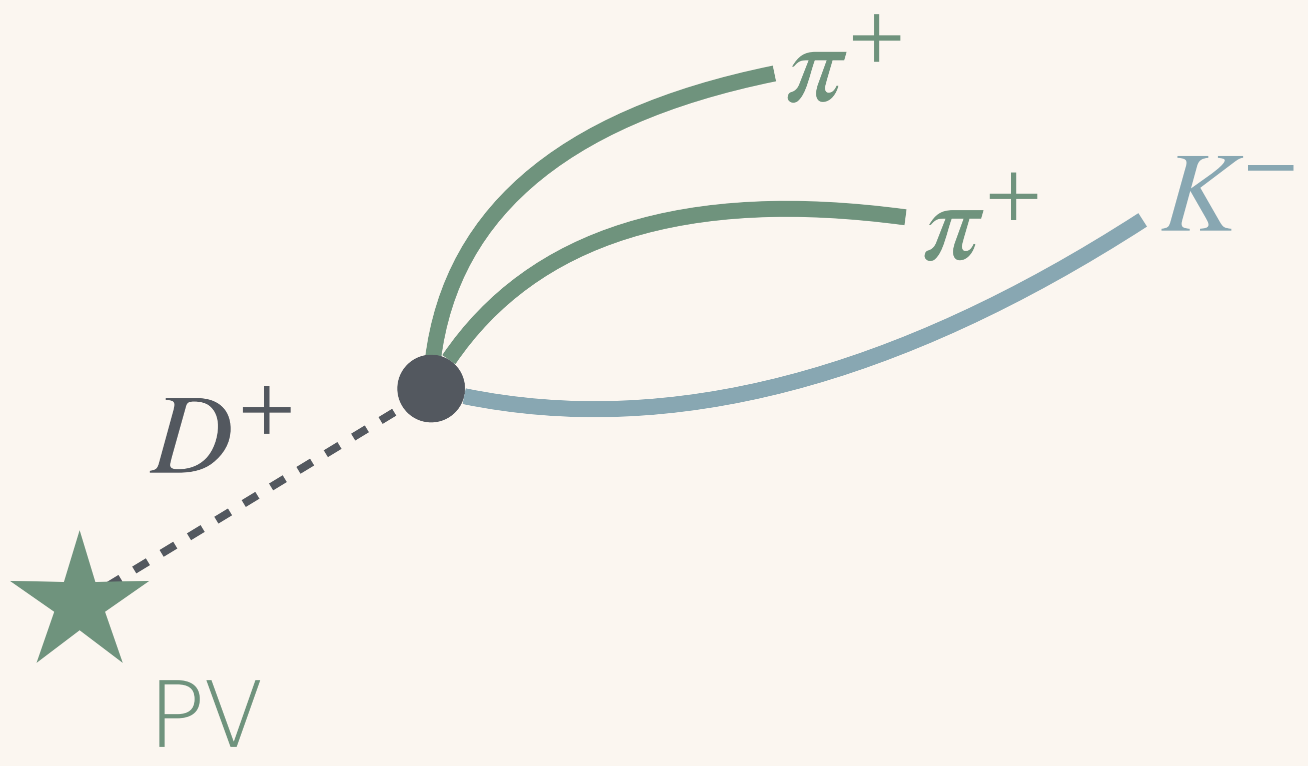
Factor of 2-5 in impact parameter resolution with ITS2!

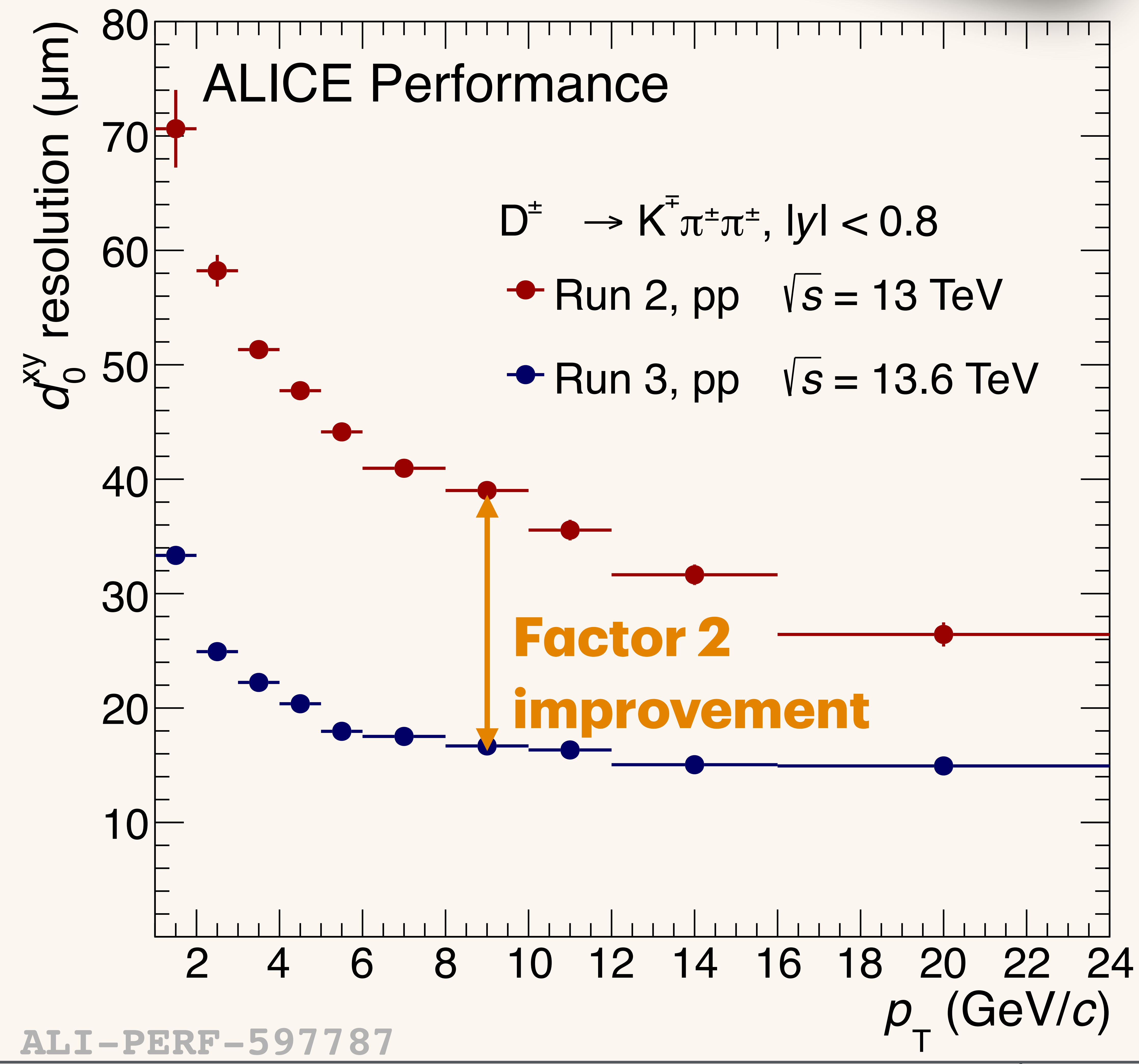
Performance



Multi track resolution

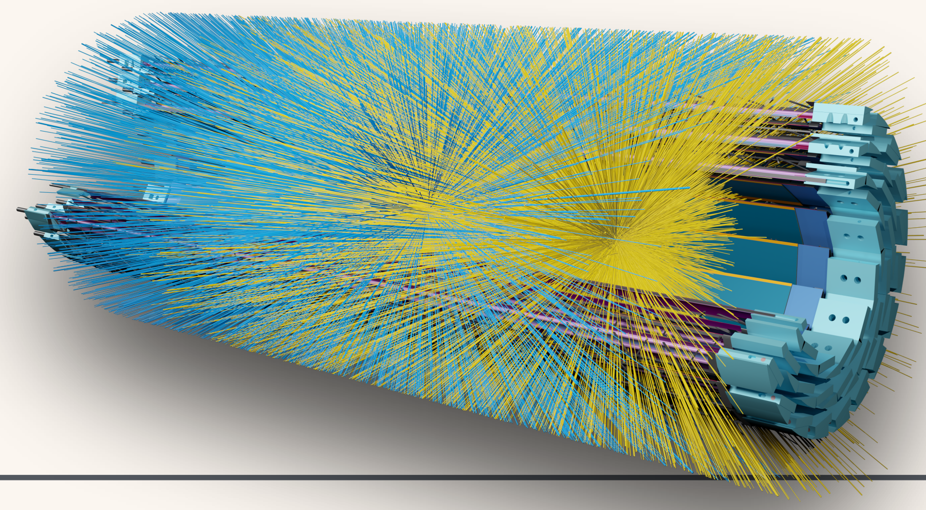
- First layer starts at innermost radius 22 mm from the nominal interaction point (39 mm for ITS1)
- Measurement of impact parameter with charmed $D^{\pm}$ mesons ($c\tau \approx 300$ µm)



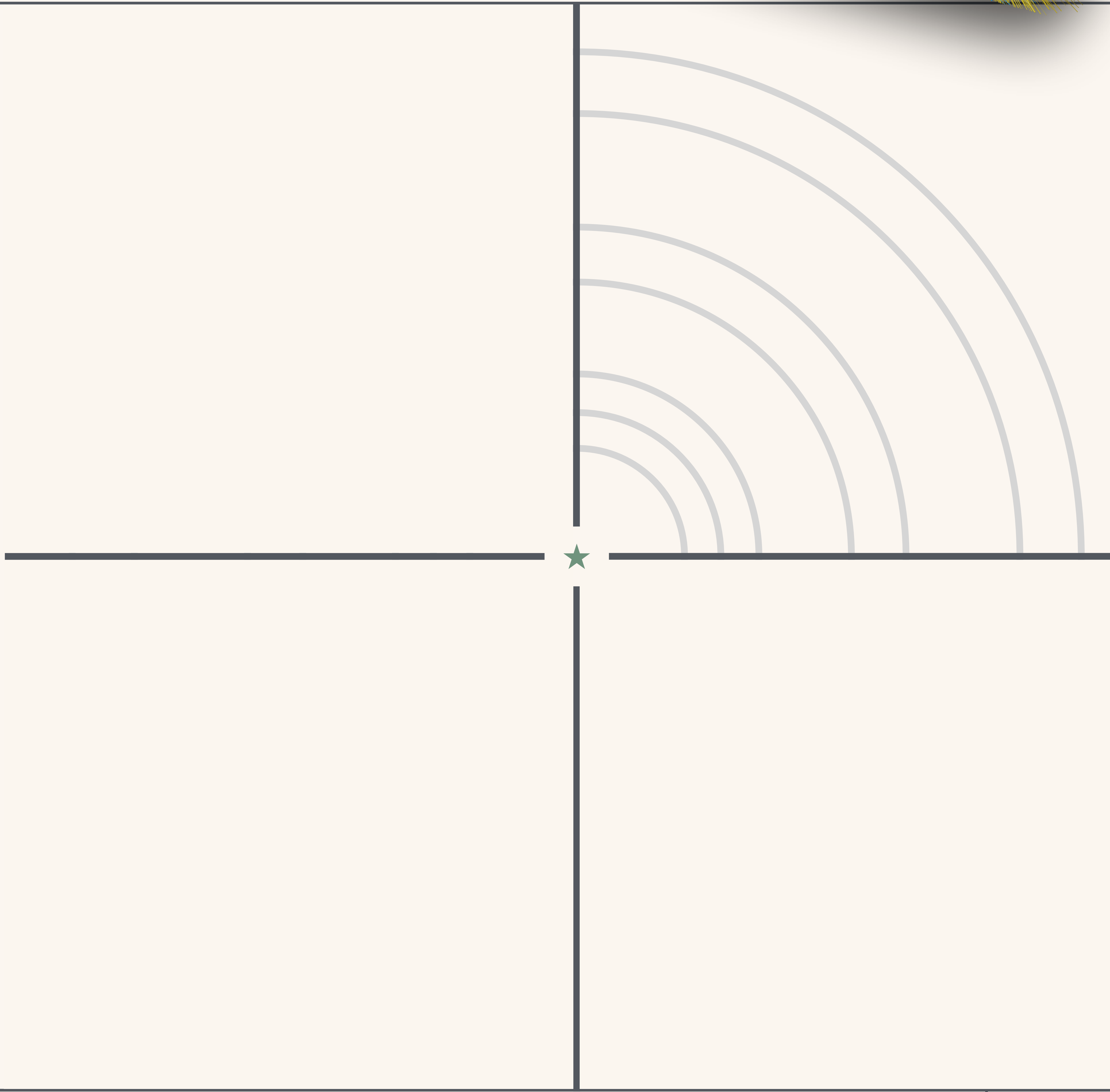


Factor of 2-5 in impact parameter resolution with ITS2!

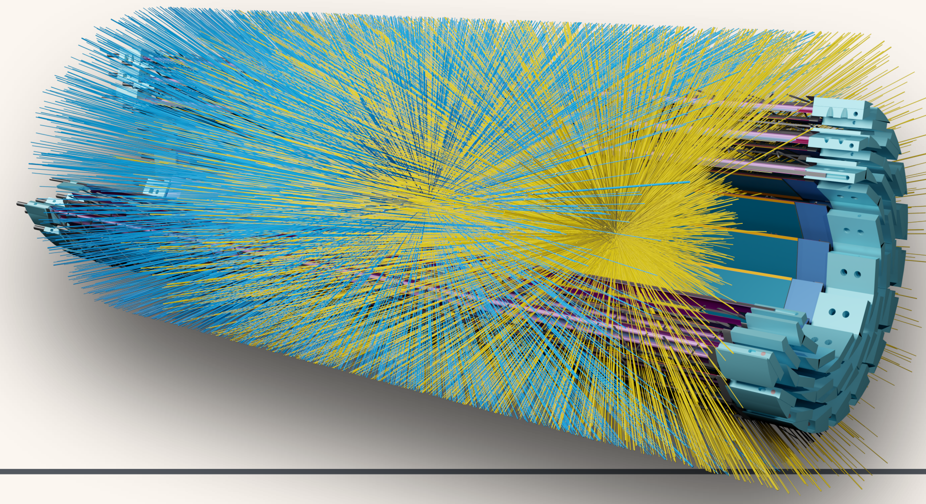
Cellular Automata reconstruction



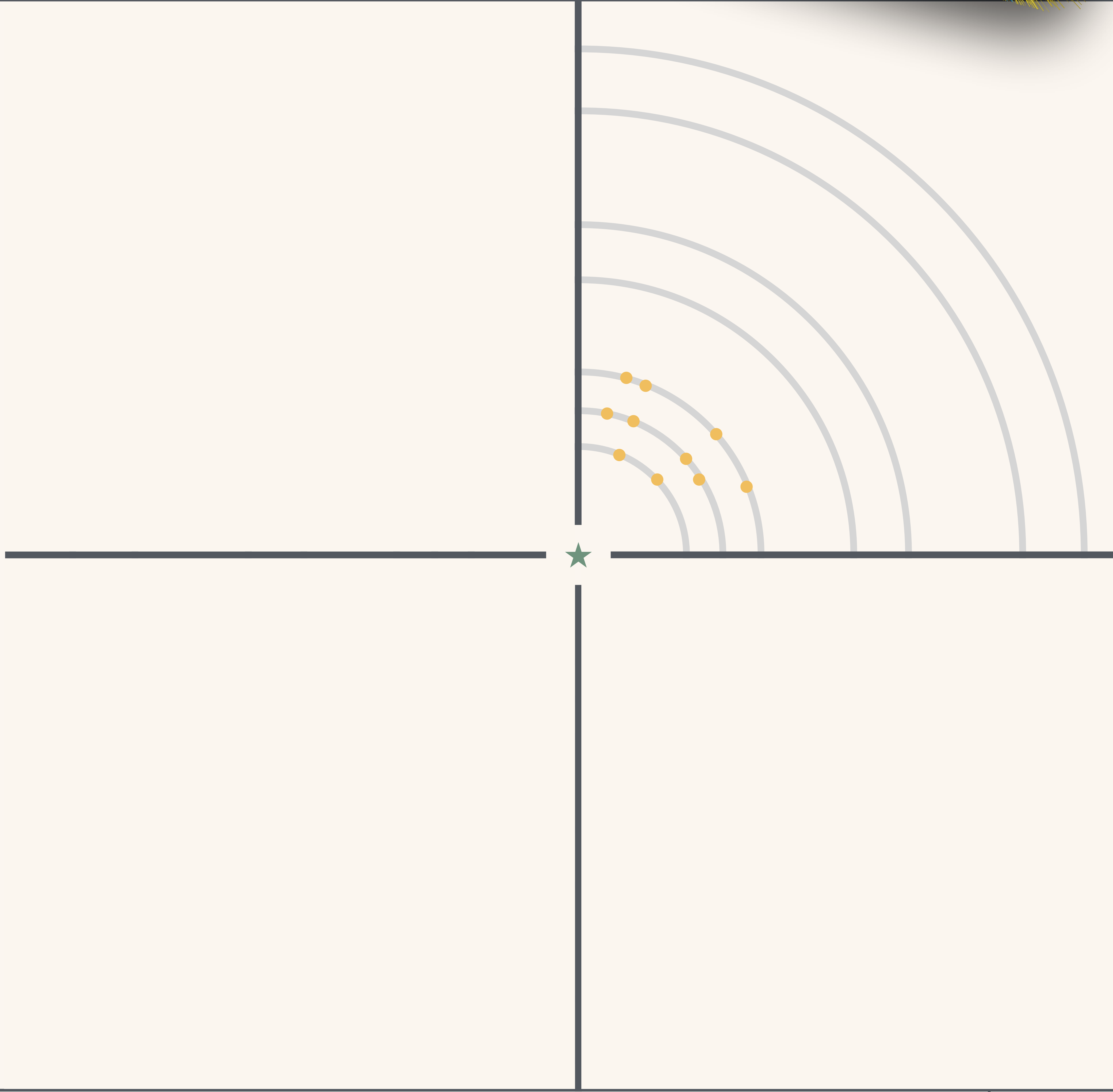
How to do it in practice



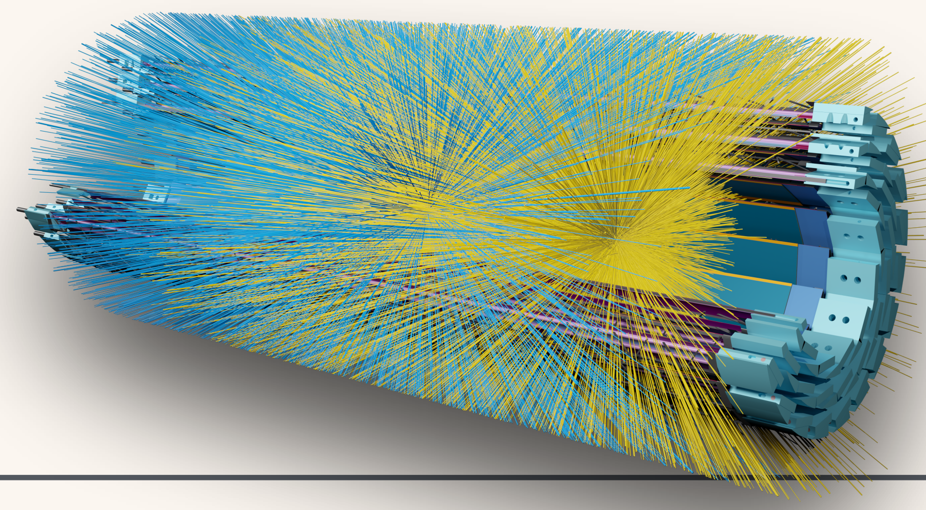
Cellular Automata reconstruction



How to do it in practice

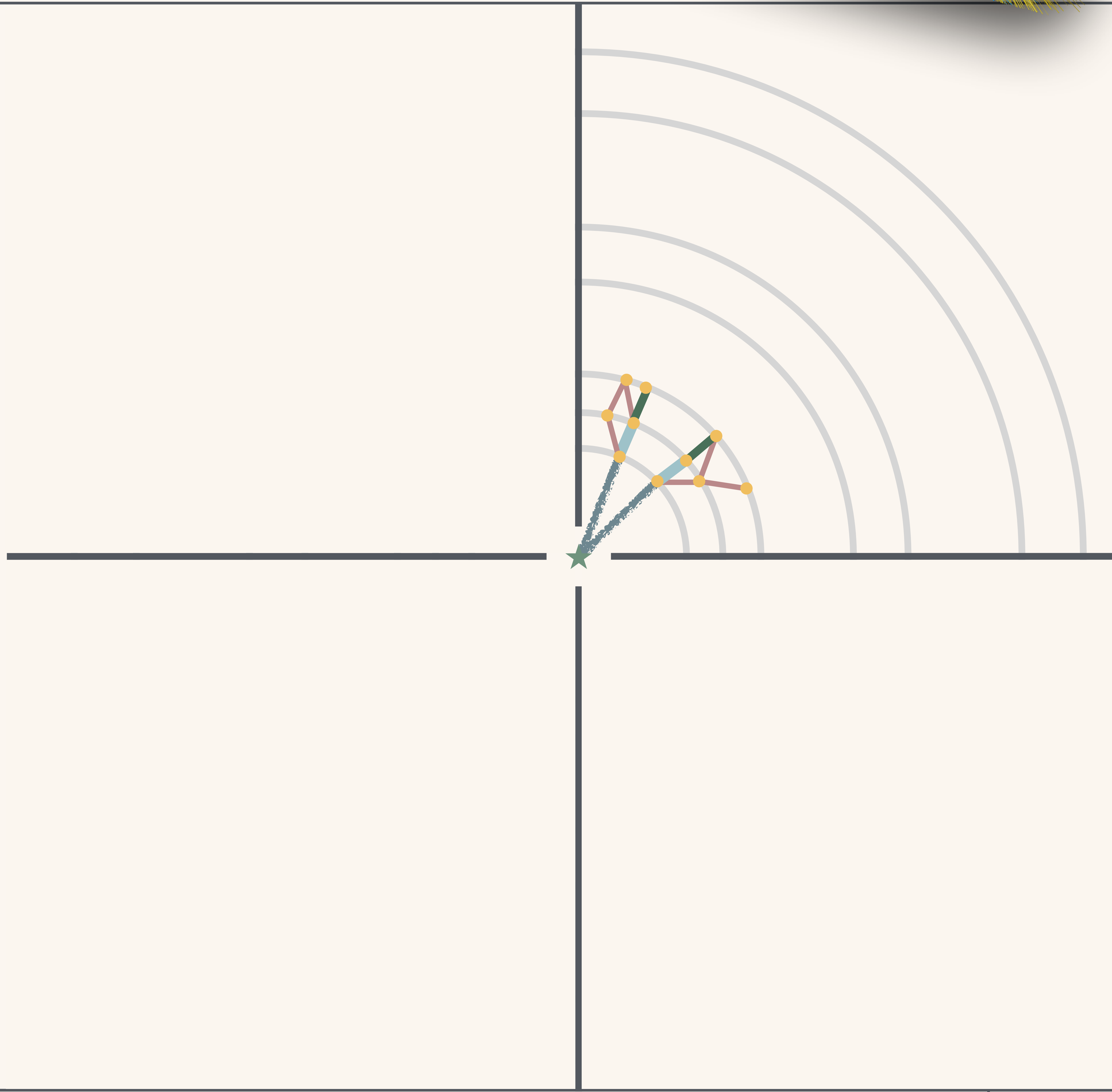


Cellular Automata reconstruction



How to do it in practice

- Build tracklets in the first three layers compatible with luminous region (σ=6cm)
- Use straight-line approximation to fit seeding vertices

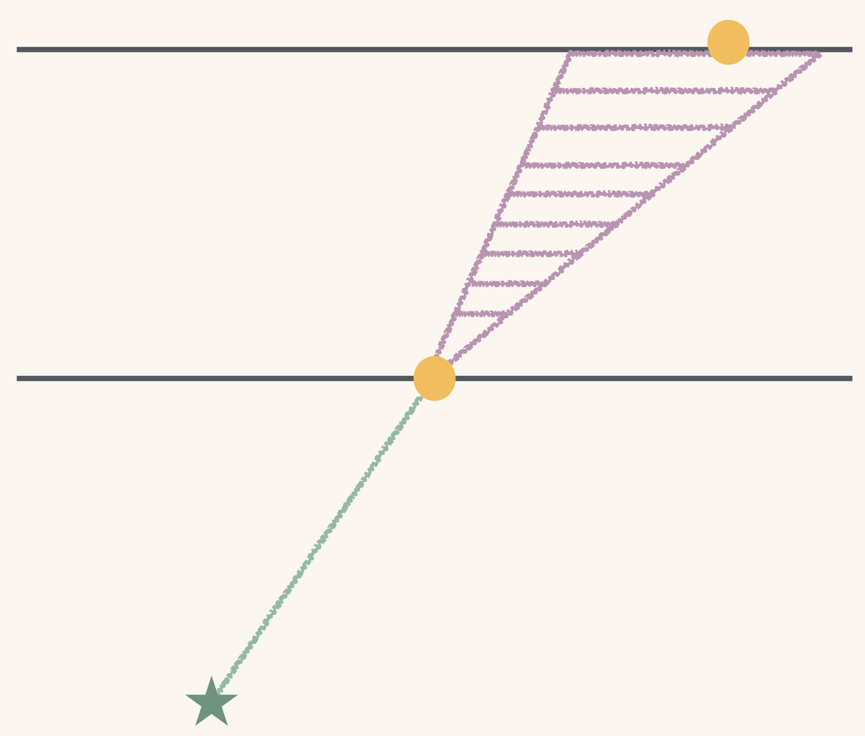


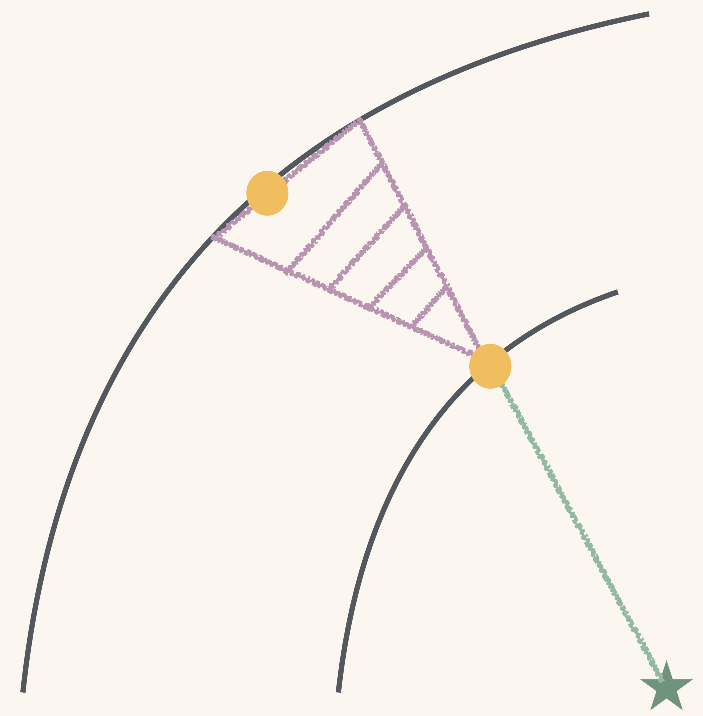
fschlepp@cern.ch

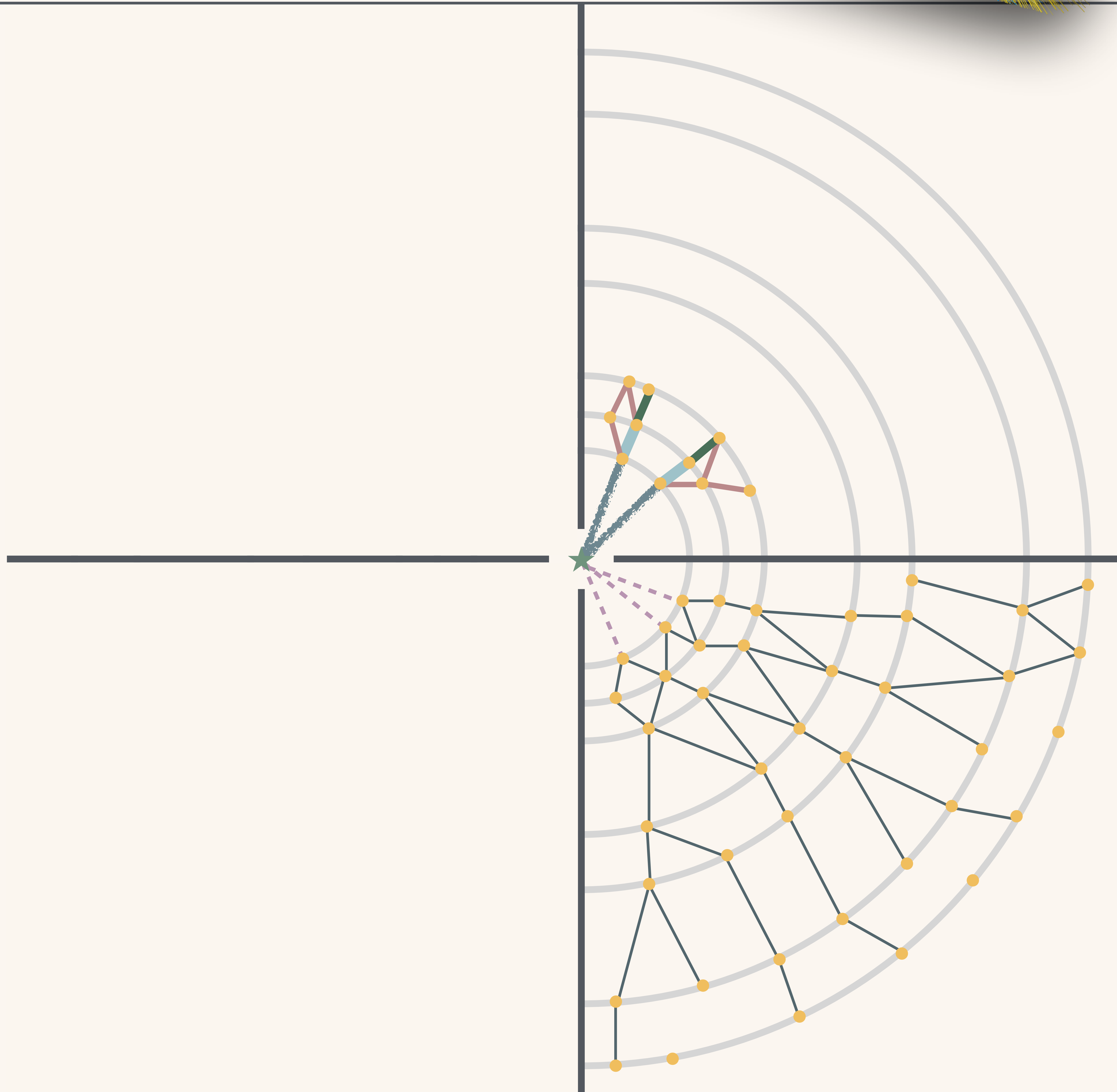
Cellular Automata reconstruction

How to do it in practice

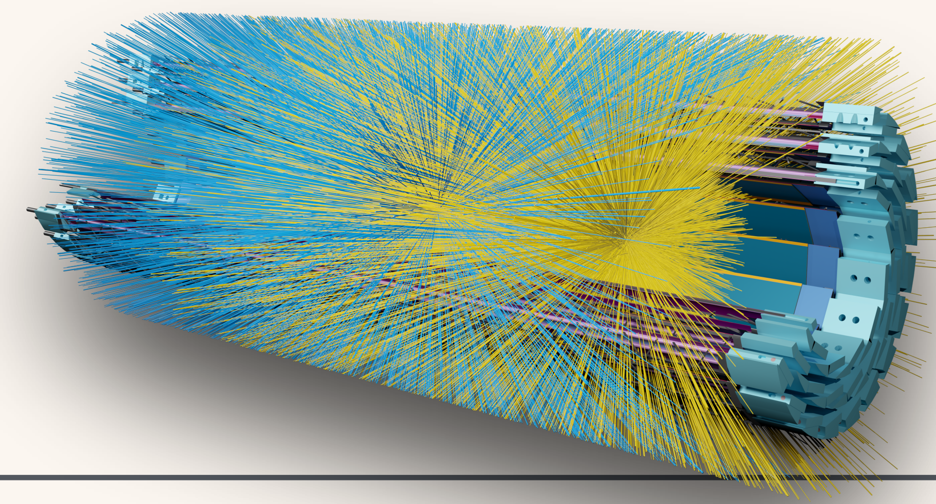
- Use seeding vertices as constraints to form tracklets across all layers
- Use simple straight-line extrapolation to define search windows in consecutive layers





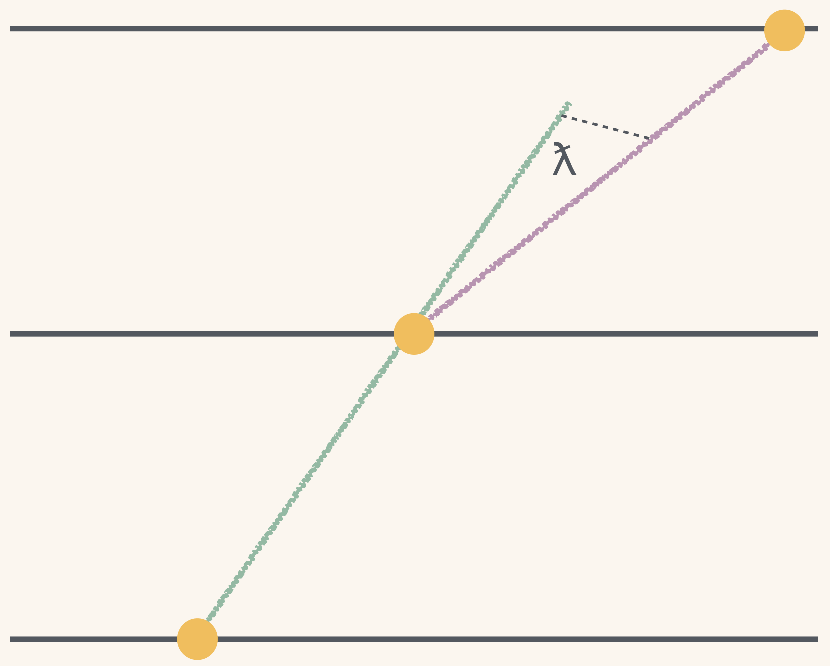


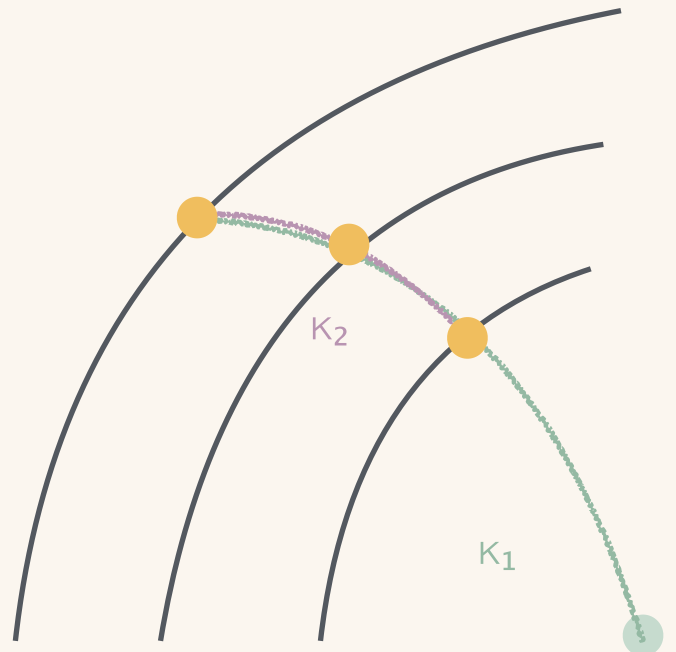
Cellular Automata reconstruction



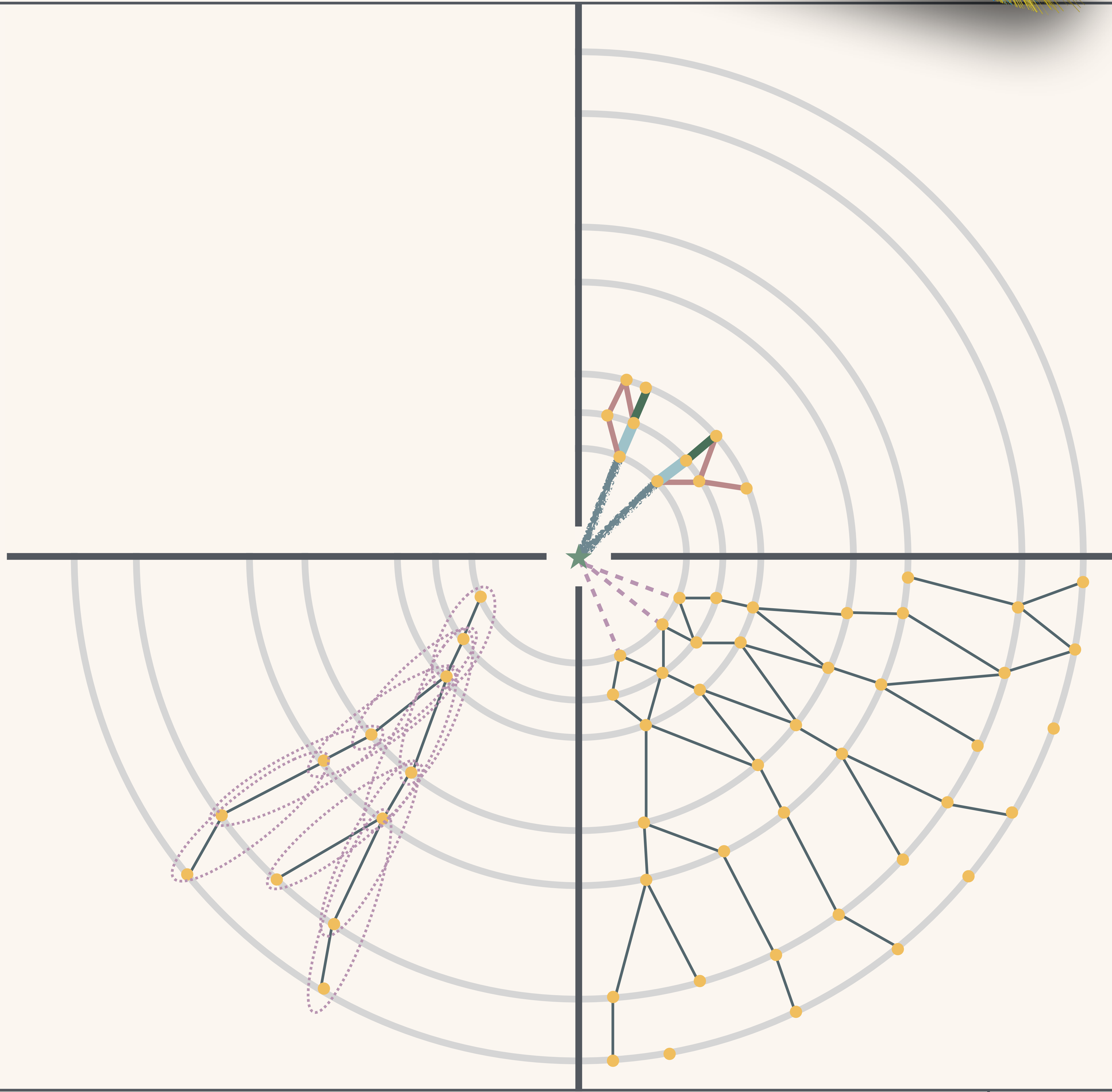
How to do it in practice

- Combine connected tracklets into cells by using a Kalman Filter fit:
  - Constraints are given both in the bending plane as a 'curvature consistency cut' and in non-bending plane as a 'slope consistency cut'





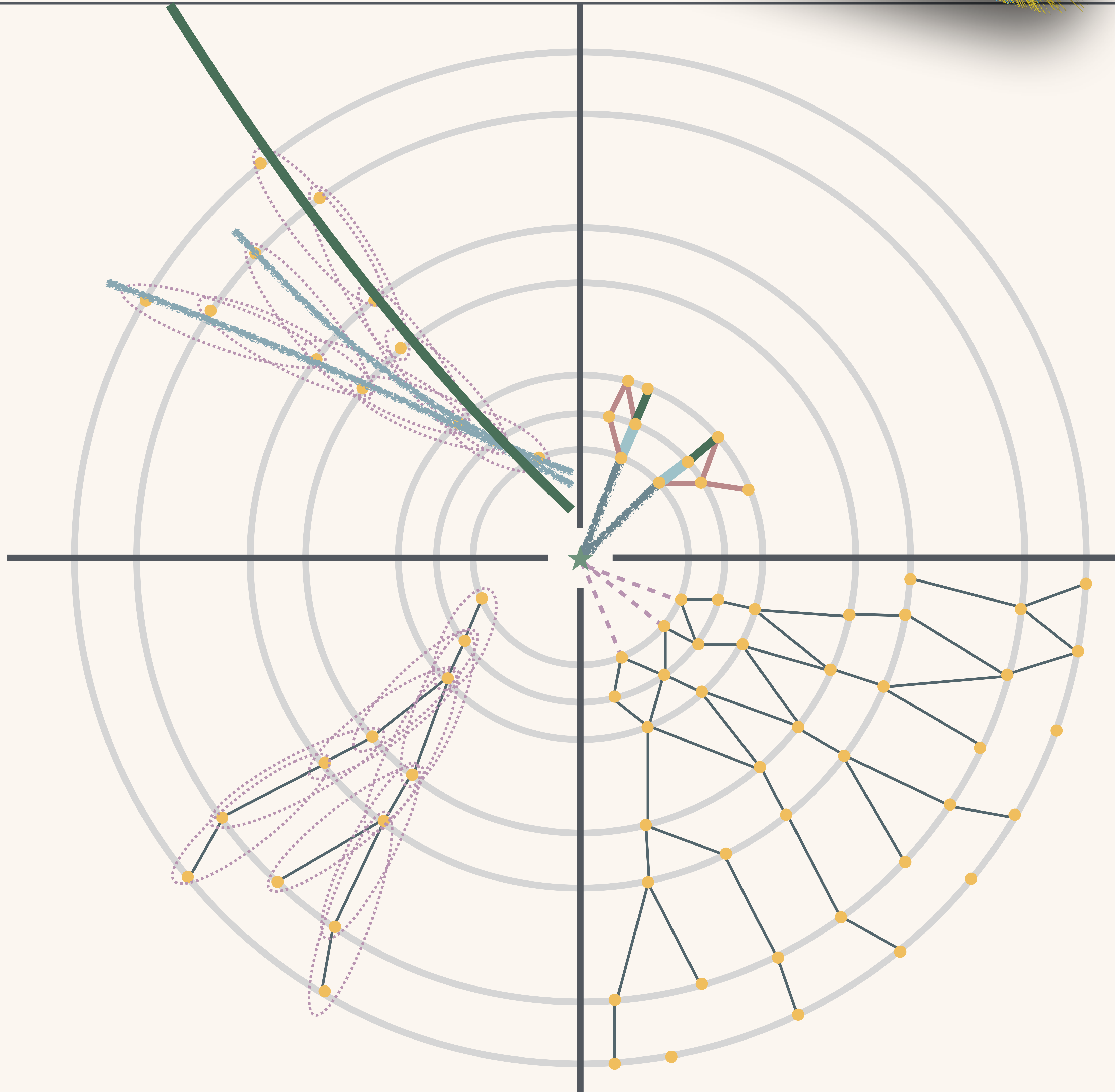
  - Cells have to pass a χ^2-cut



Cellular Automata reconstruction

How to do it in practice

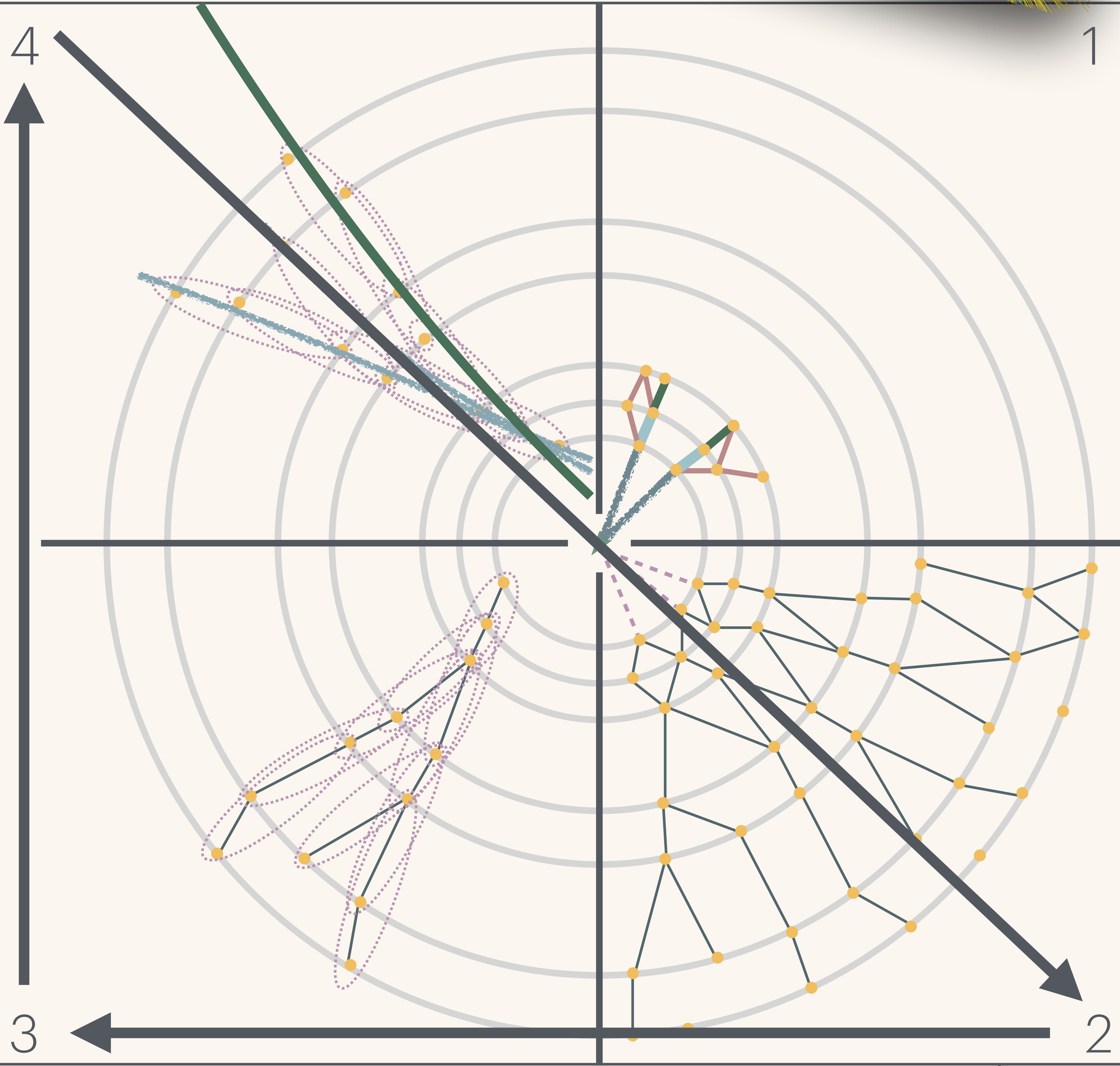
- Use a Kalman Filter fit for connected cells to form all possible roads
  - Disambiguate roads sharing clusters based on quality (χ^2-based) and length (more clusters)
  - The winner amongst all roads is promoted to a track



Cellular Automata reconstruction

How to do it in practice

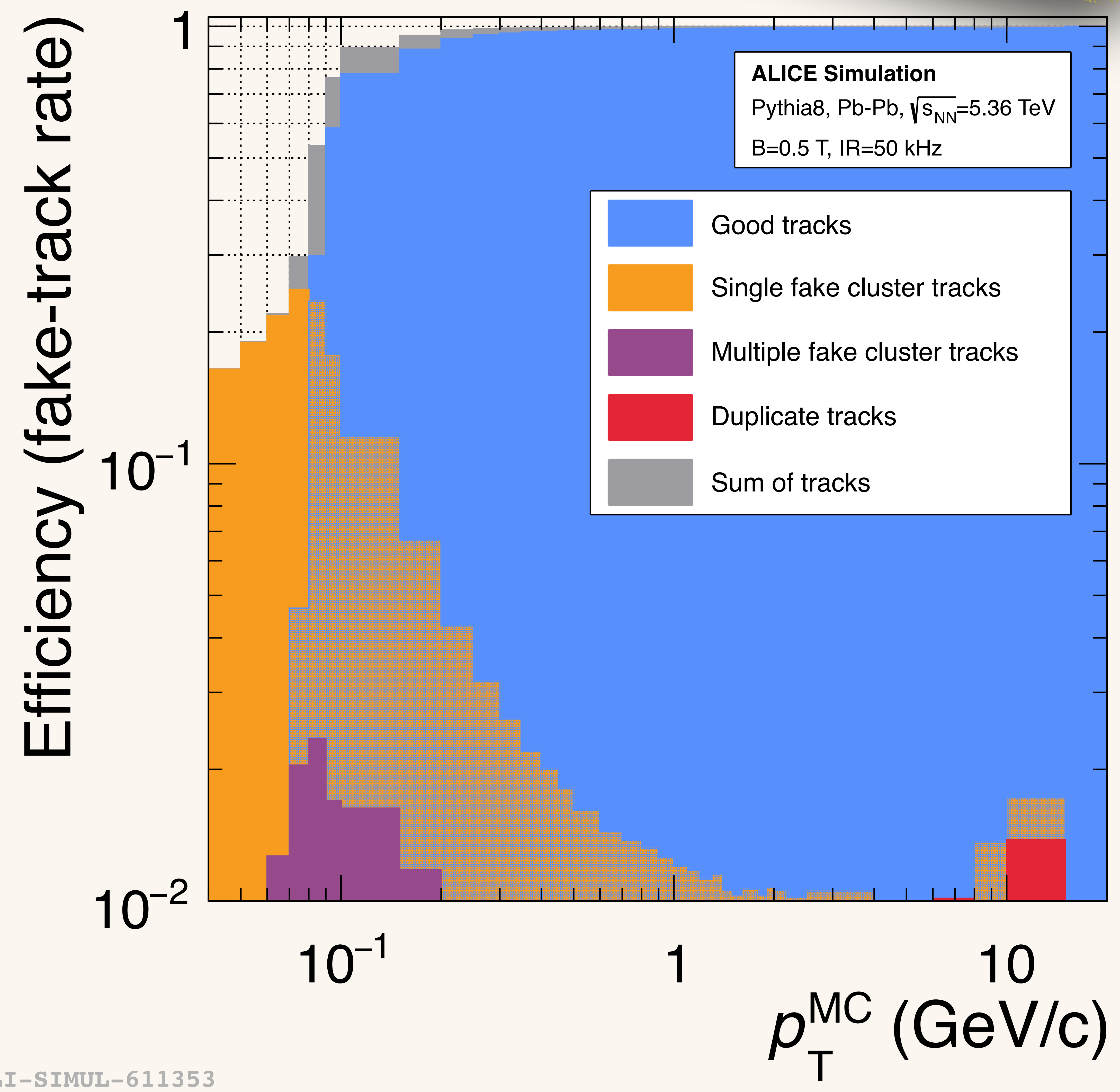
- Steps 2 to 4 are run iteratively
  - Clusters used by tracks are marked and cannot be reused in subsequent iterations
  - Idea is to find straight tracks first, then gradually relax criteria to capture shorter, low-momentum tracks (easiest to hardest)
  - Currently, we have 3 or 4 iterations
    - In Pb-Pb, there is a special UPC-iteration



Performance

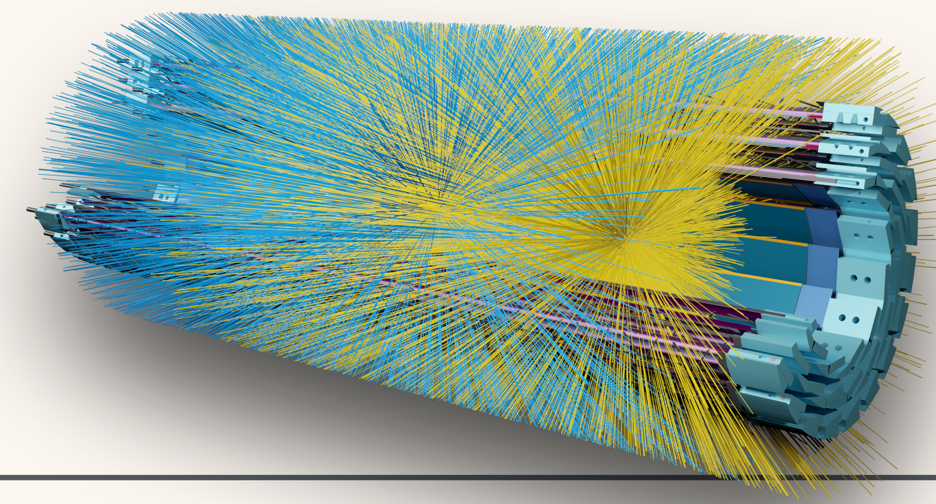
Reconstruction efficiency

- Total efficiency above 98% down to 0.2 GeV/c
- Single fake cluster attachments less than 5% down to 0.2 GeV/c
- Multi fake cluster attachment only below 0.2 GeV/c and less than 3%
- S/B=1 point around 80 MeV/c with total efficiency above 70%



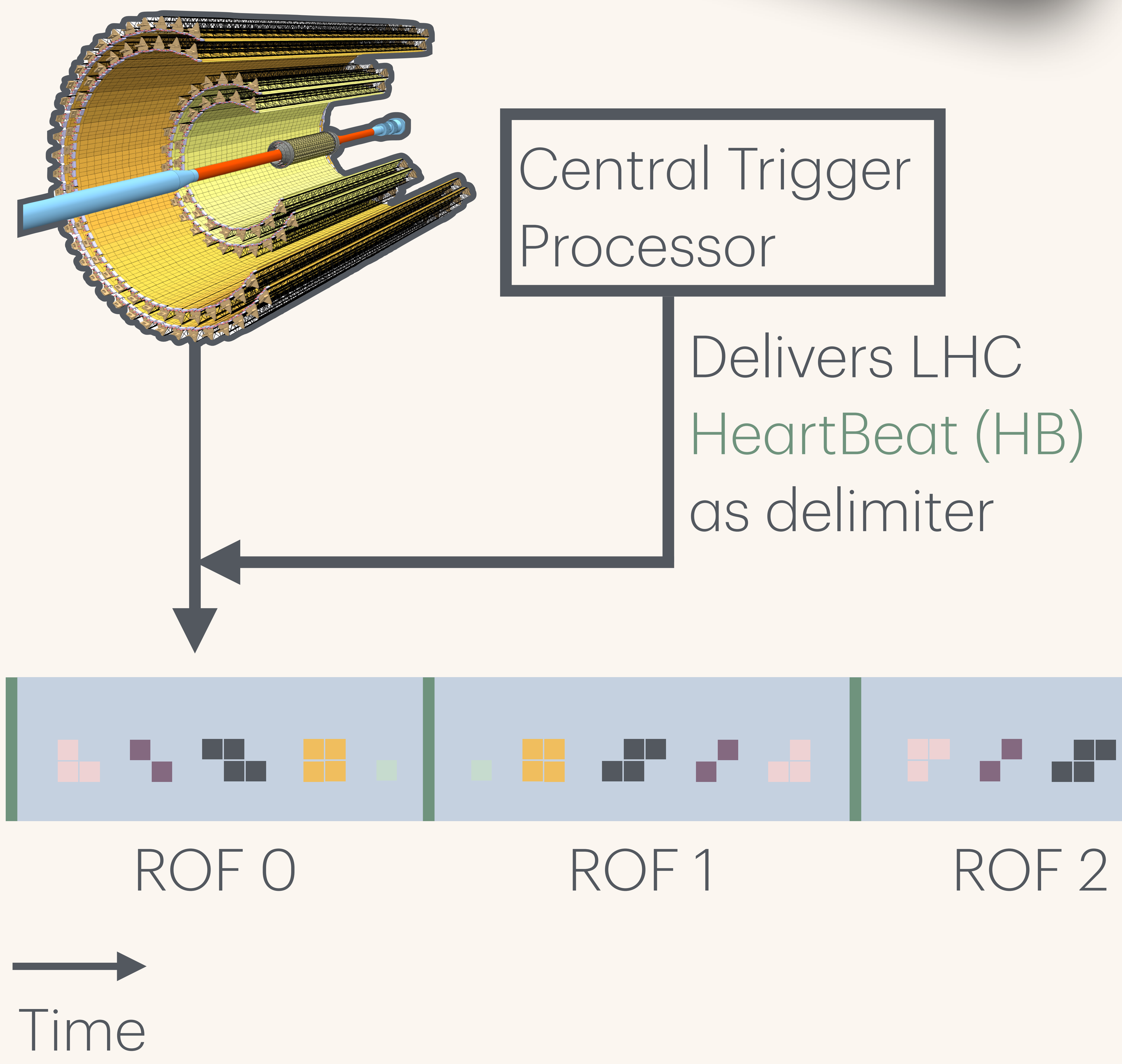
Shown here for primary-like particles

Detector readout

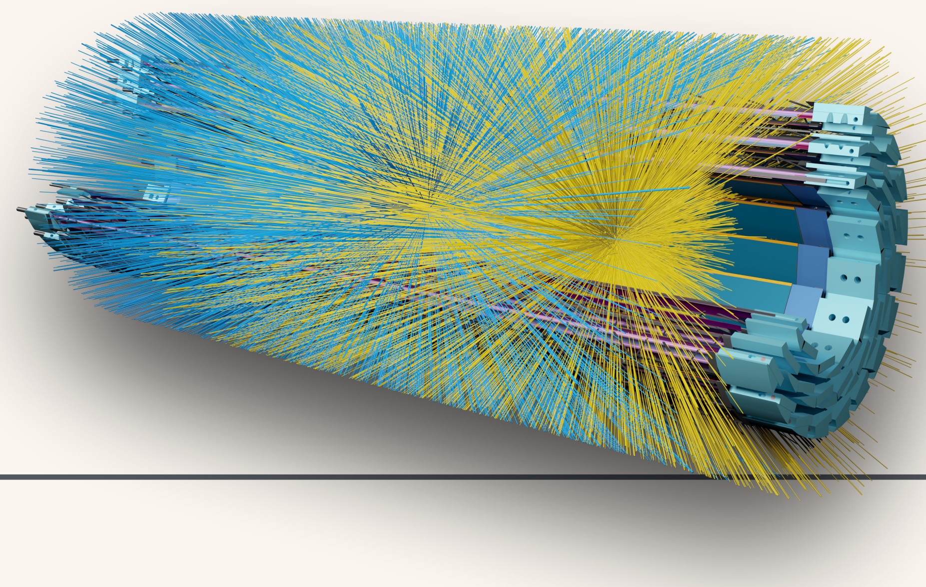


Intermezzo

- Continuous stream of ITS2 data is chopped up into consecutive readout frames (ROFs)
- Each ROF contains data from all layers within the same time interval
- ROF length is configurable and selected depending on the collision system to cope with bandwidth constraints:
  - pp ≈ 5 µs
  - Pb-Pb ≈ 15 µs



Time-aware reconstruction

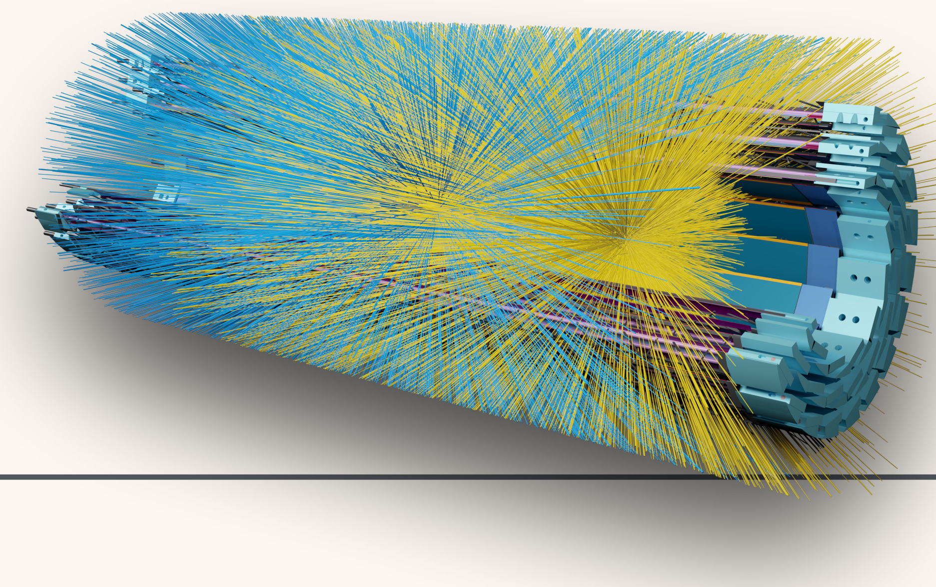


Timing detector effects

- Charge dependent time-of-arrival of ALPIDE chip and energy loss fluctuations lead to partial migration of digits resulting in incomplete clusters

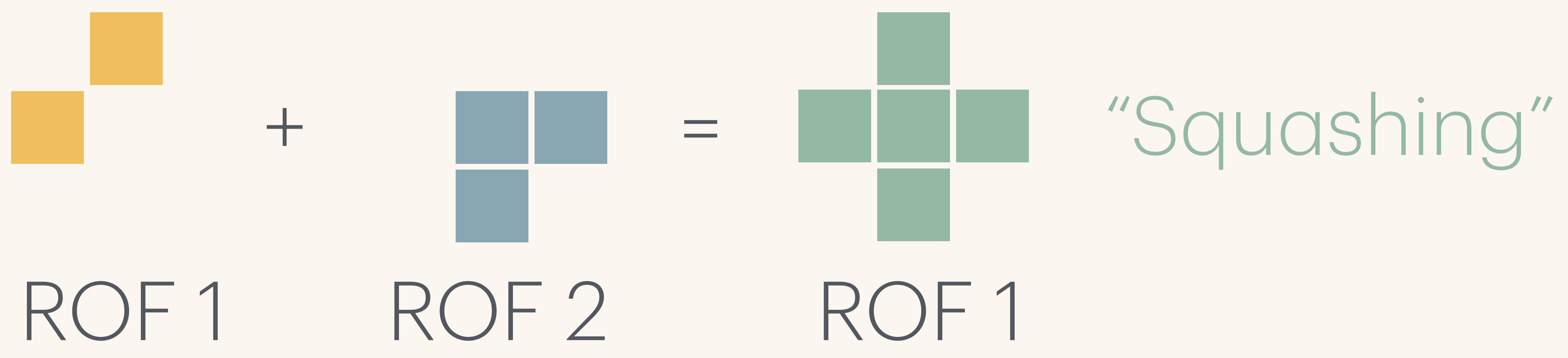
"Squashing"

Time-aware reconstruction

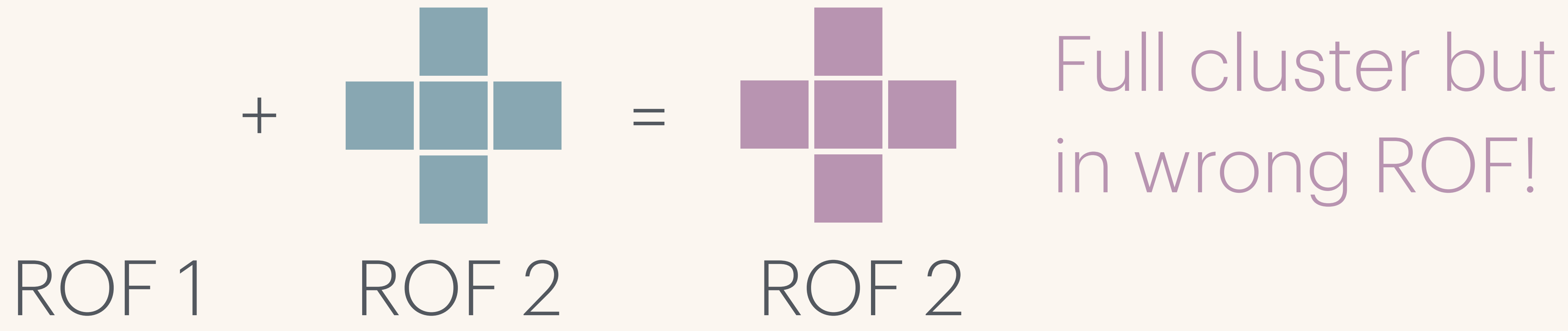


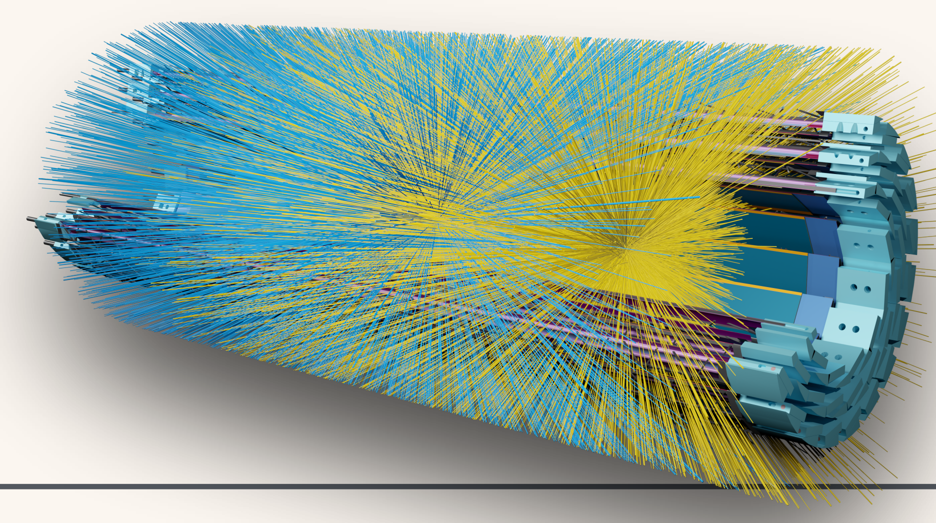
Timing detector effects

- Charge dependent time-of-arrival of ALPIDE chip and energy loss fluctuations lead to partial migration of digits resulting in incomplete clusters



- Extreme case: full migration of clusters into the next readout frame (ROF)





Time-aware reconstruction

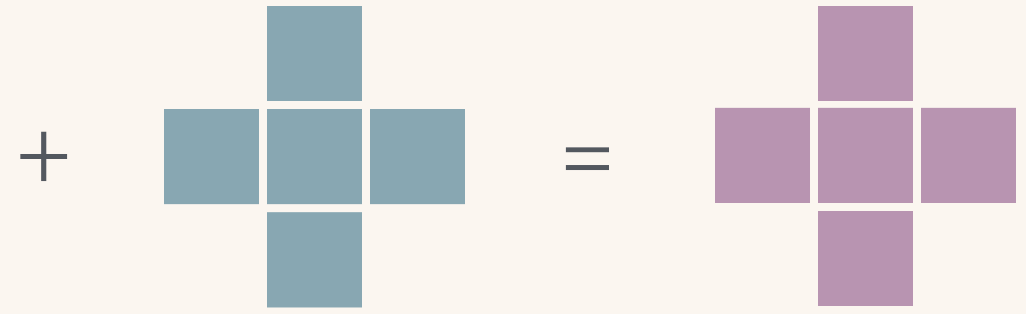
Timing detector effects

- Charge dependent time-of-arrival of ALPIDE chip and energy loss fluctuations lead to partial migration of digits resulting in incomplete clusters

"Squashing"

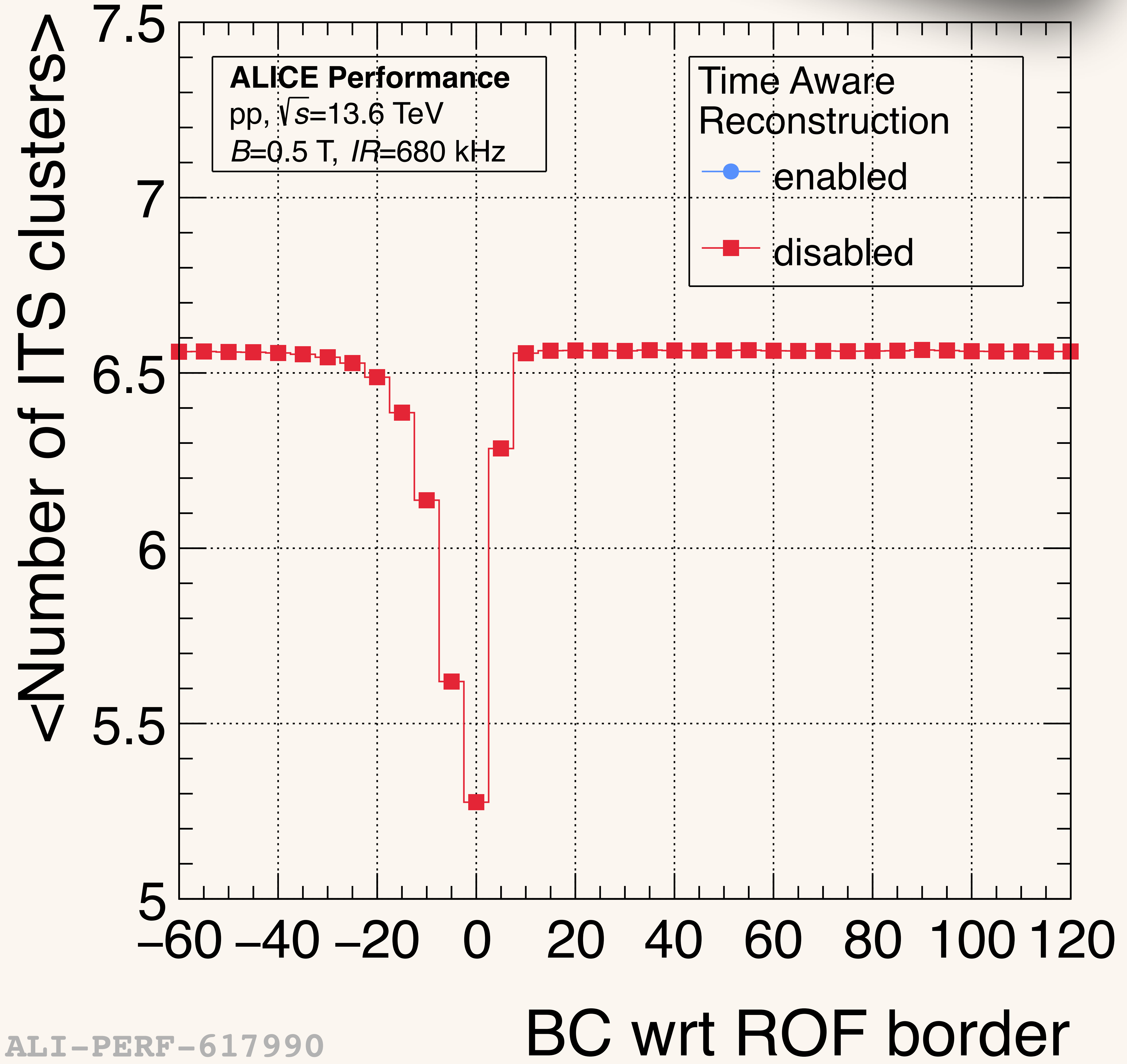
ROF 1 ROF 2 ROF 1

- Extreme case: full migration of clusters into the next readout frame (ROF)

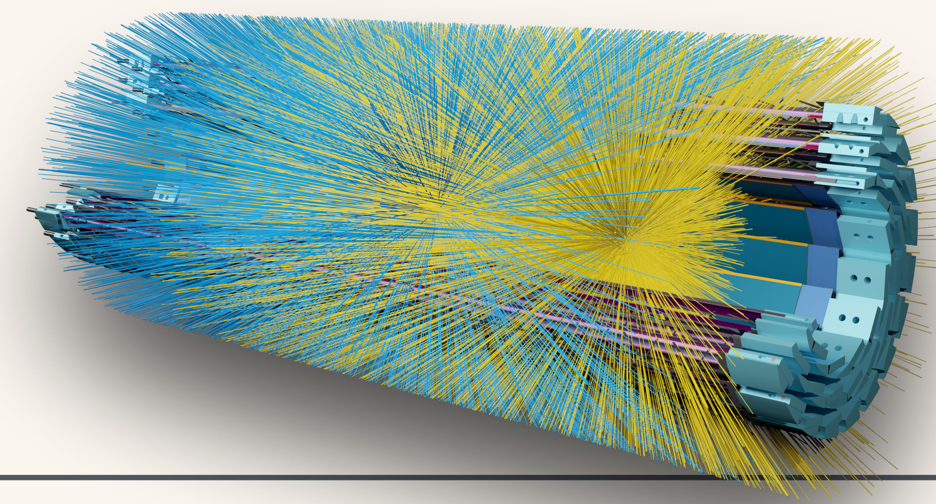


Full cluster but in wrong ROF!

ROF 1 ROF 2 ROF 2



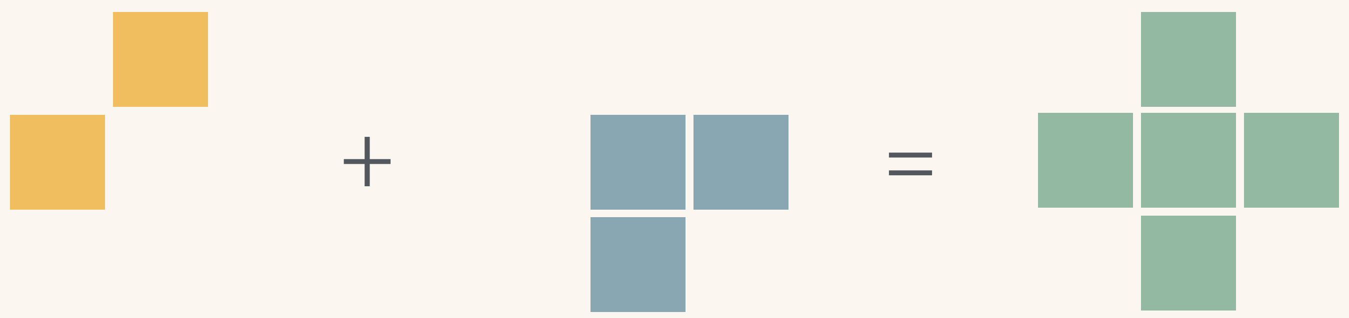
Loss of clusters leads to incomplete track topologies & loss of tracks!



Time-aware reconstruction

Timing detector effects

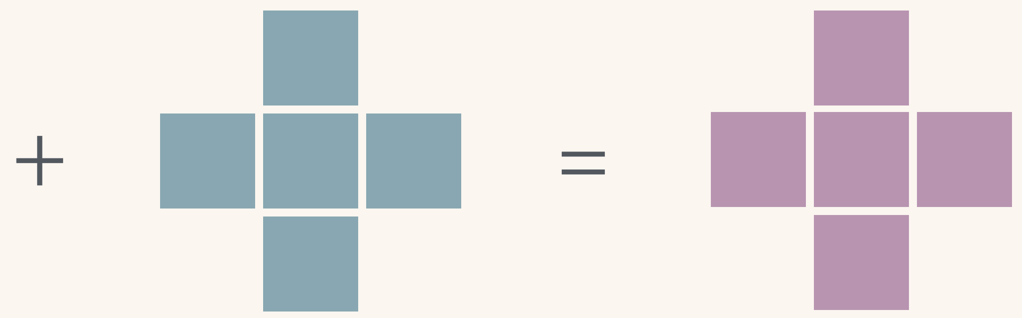
- Charge dependent time-of-arrival of ALPIDE chip and energy loss fluctuations lead to partial migration of digits resulting in incomplete clusters



“Squashing”

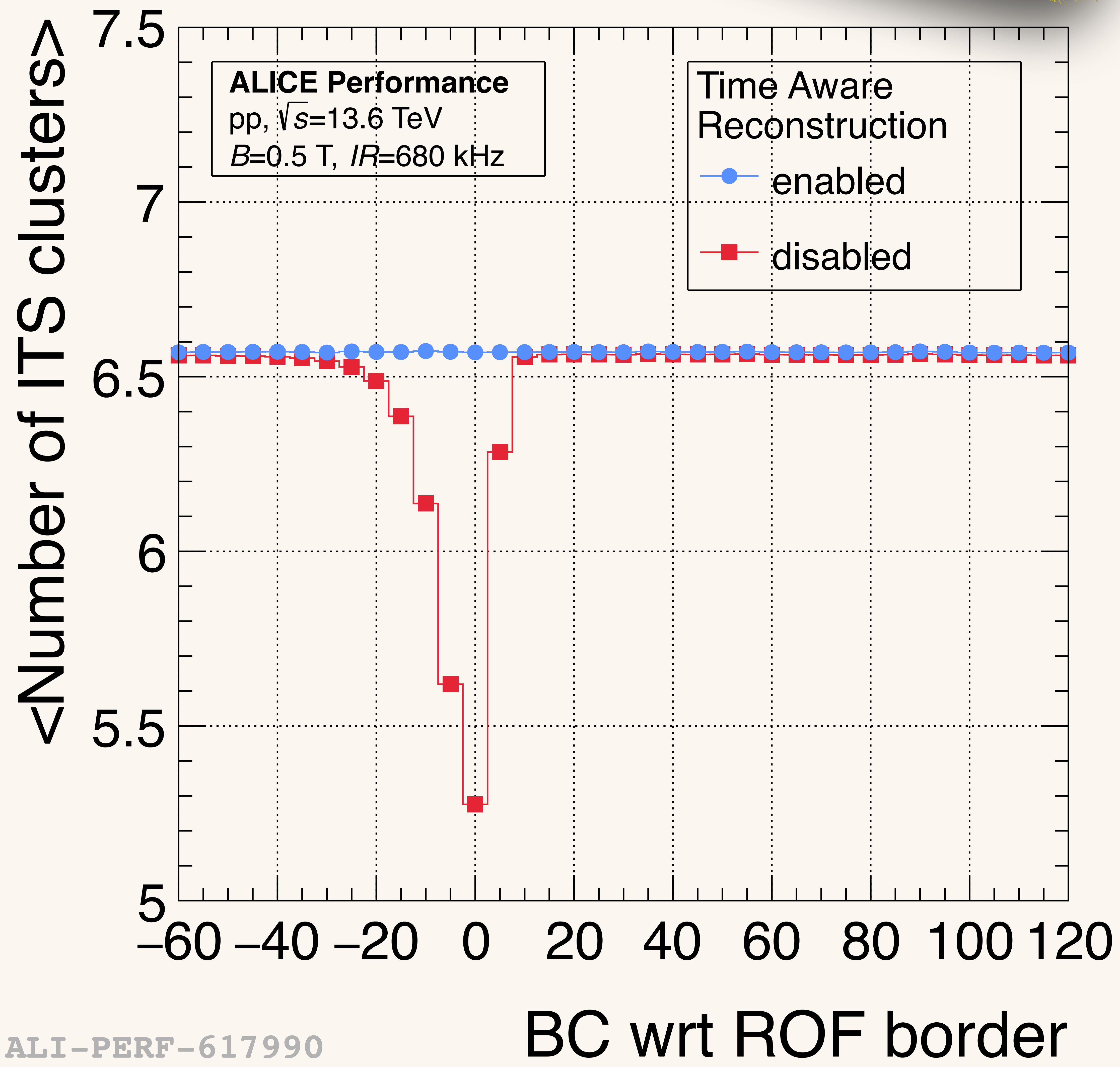
ROF 1 ROF 2 ROF 1

- Extreme case: full migration of clusters into the next readout frame (ROF)

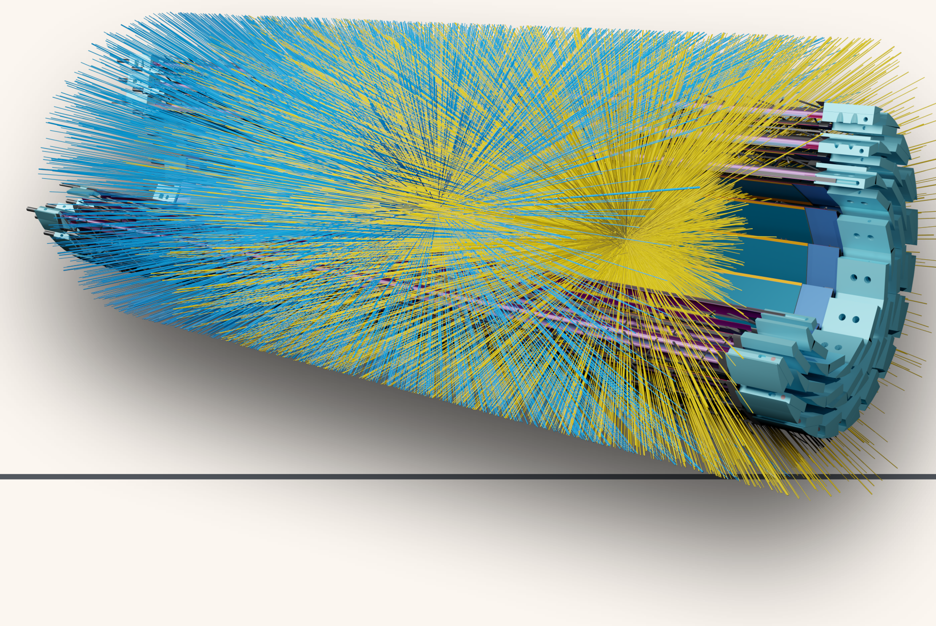


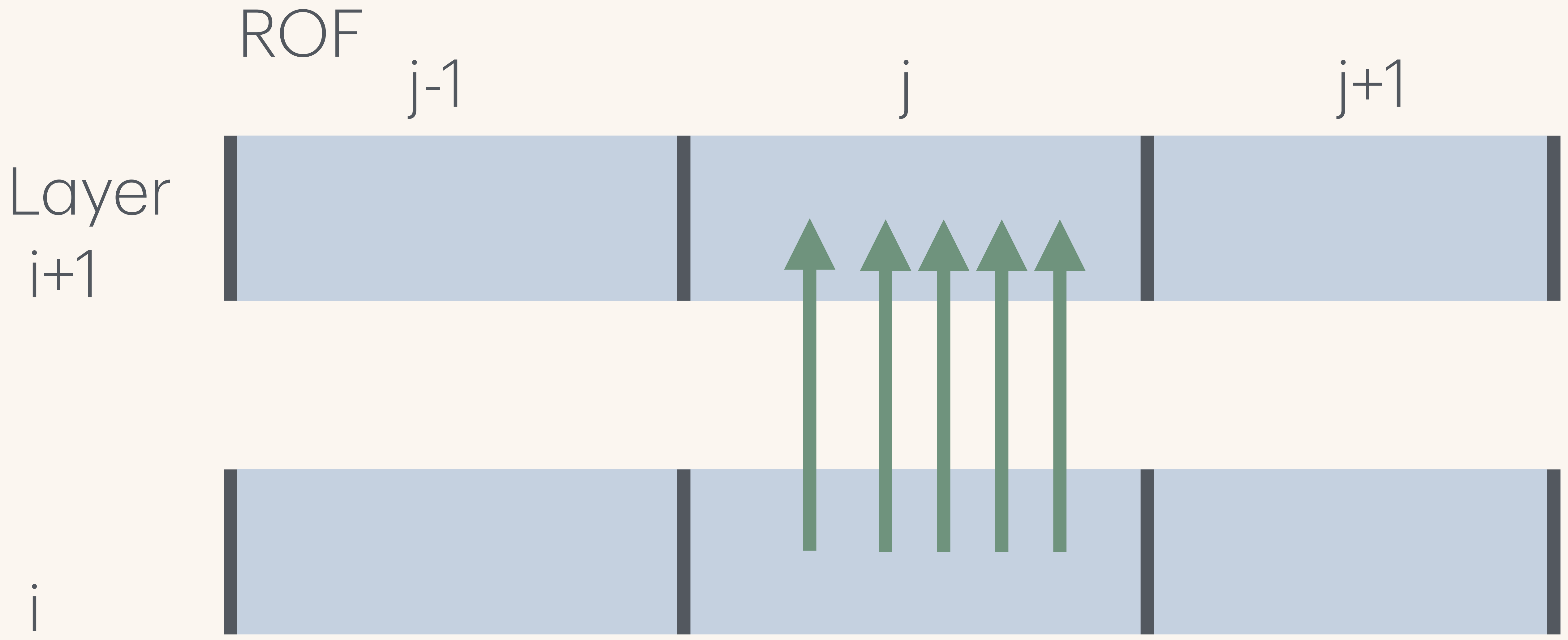
Full cluster but in wrong ROF!

ROF 1 ROF 2 ROF 2

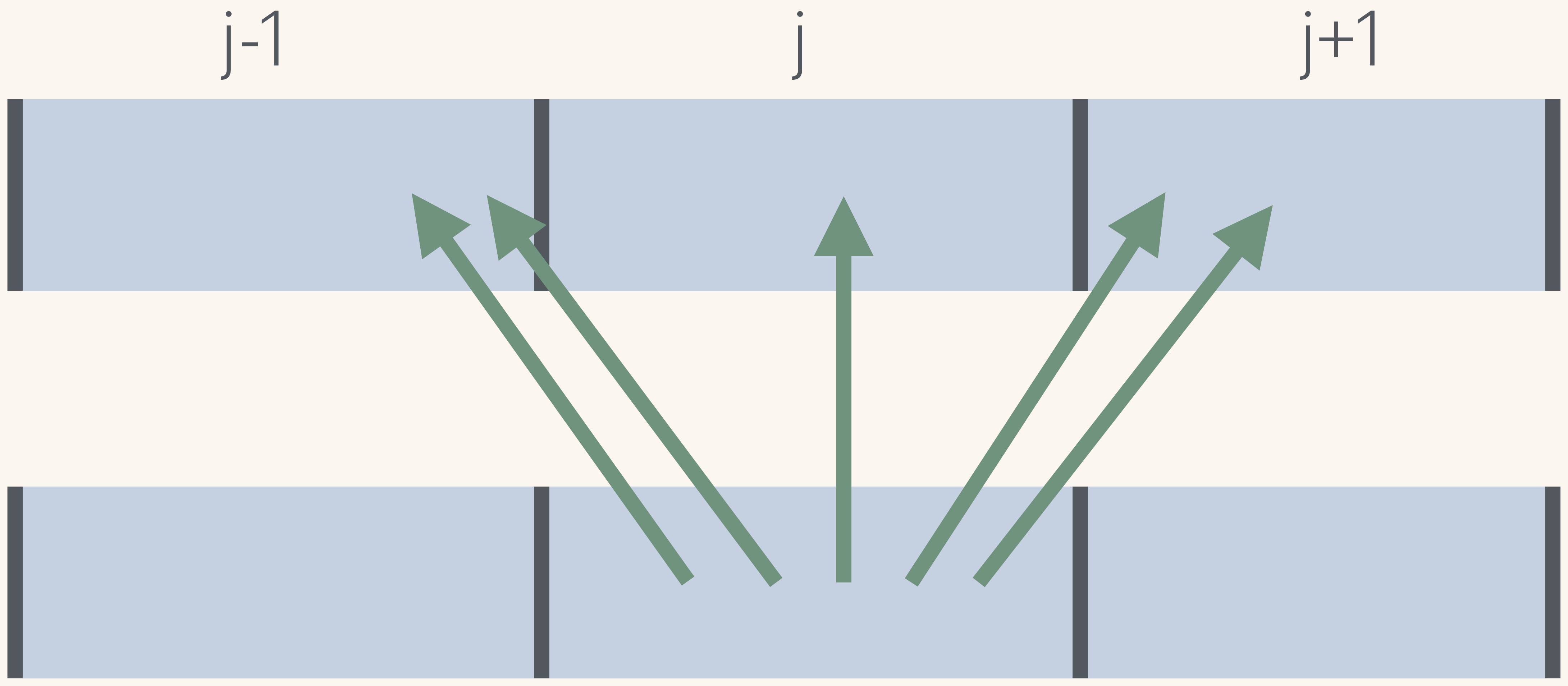


Time aware reconstruction





'Normally' tracklets are only formed with clusters from the same ROF

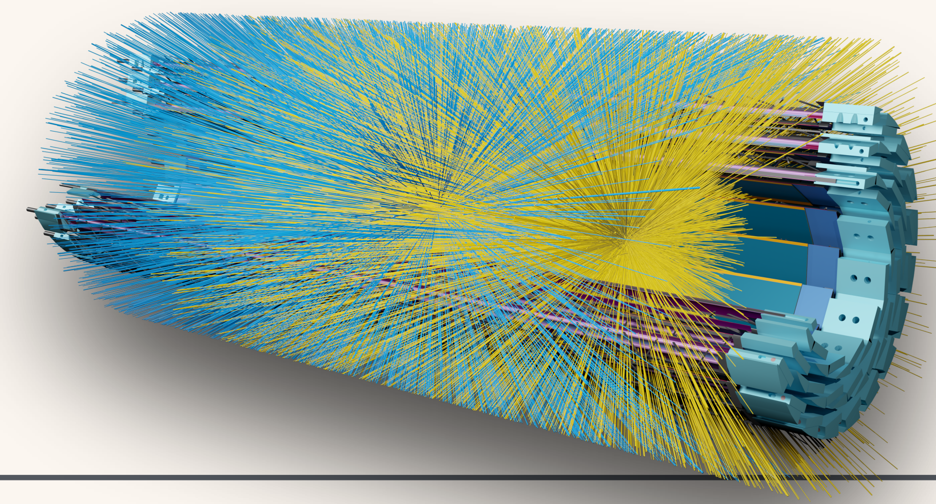


Allow tracklets to be formed with clusters from neighbouring ROFs

Also only allow combination of cells which are compatible in time

Tracks using clusters from neighbouring ROFs have better timing ~1 μs!

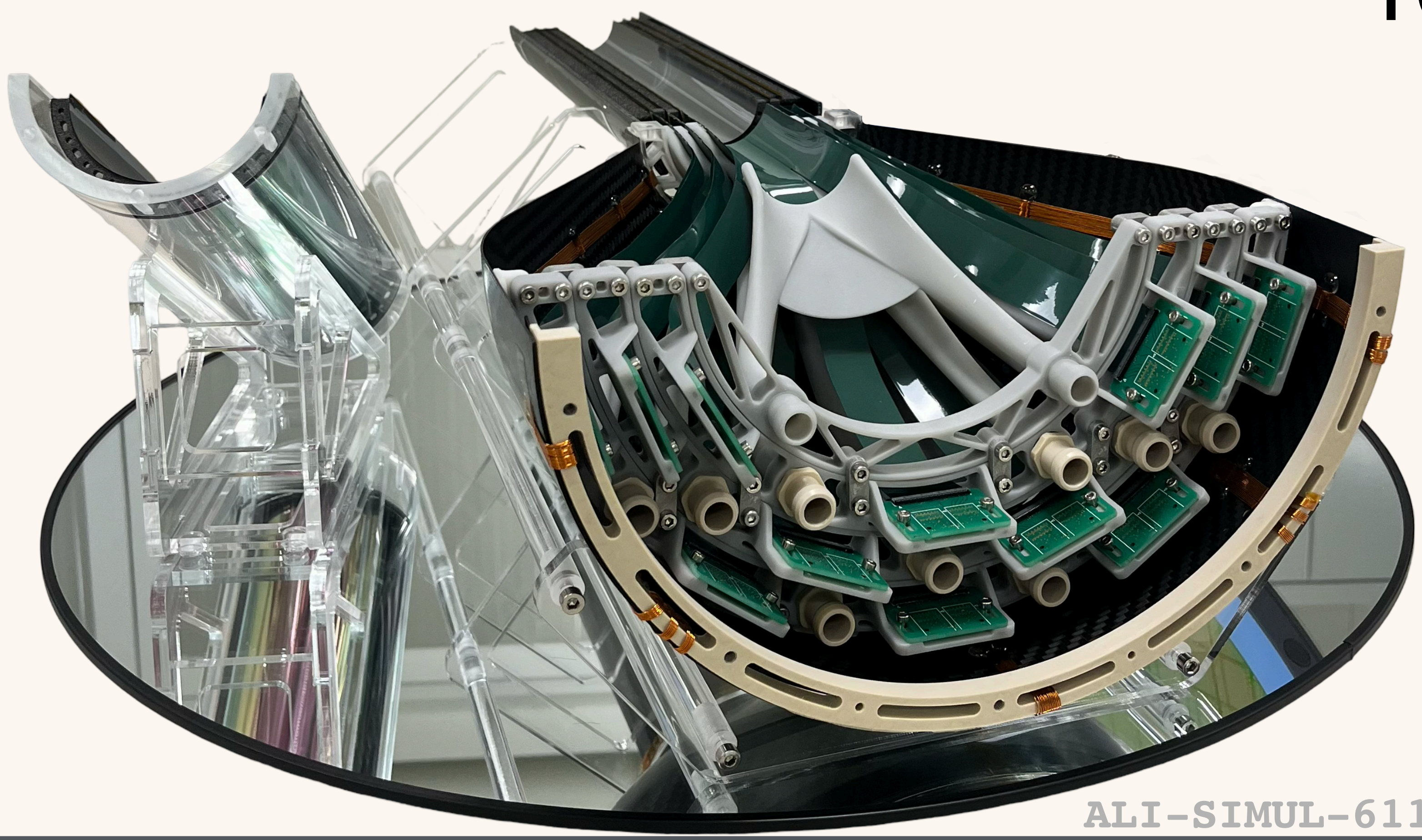
Inner Tracking System 3

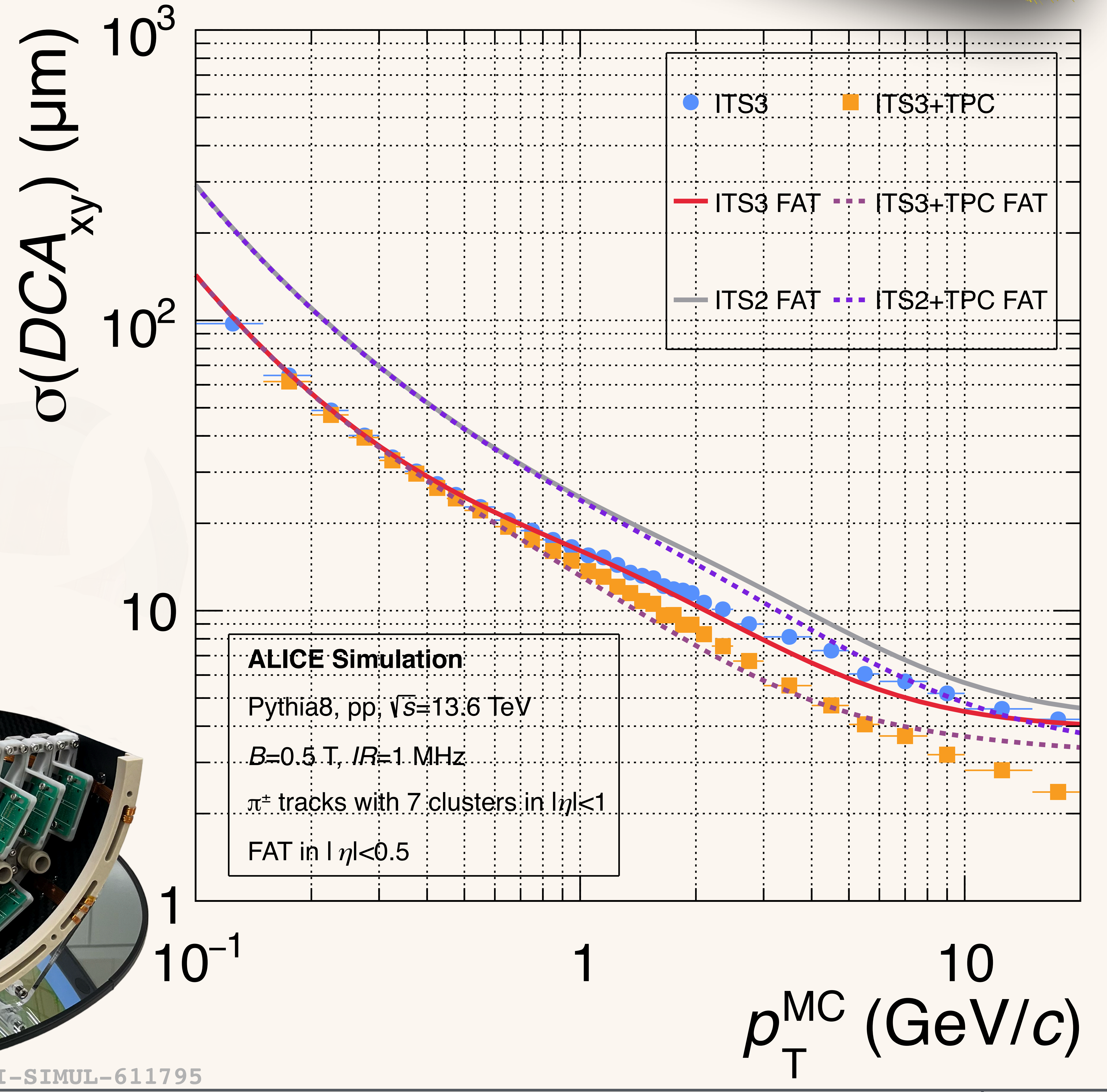


Run 4 upgrade

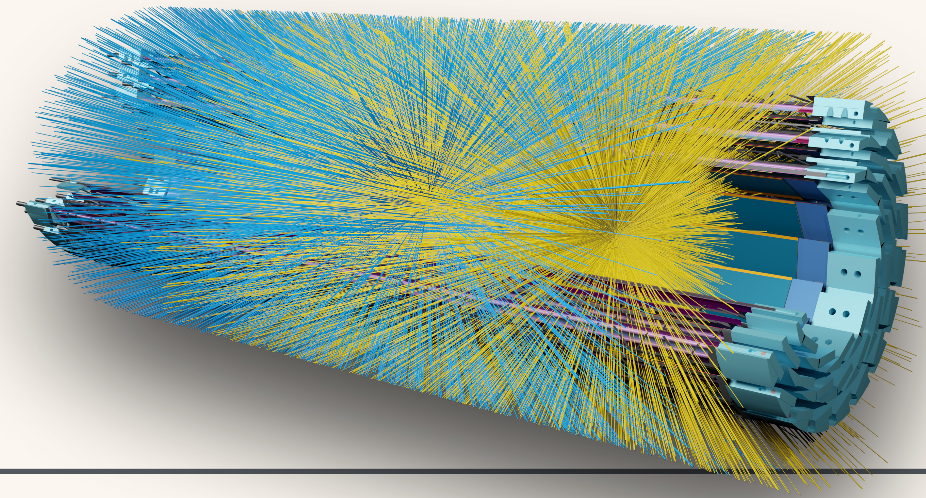
- Bent wafer-scale sensor ASIC
  - 65 nm CMOS MAPS (fabricated using stitching)
  - Even lower material budget 0.36% X_0 ➡ 0.09% X_0
  - A factor of 2 improvement in pointing resolution!
- Will use the same algorithms and reap all benefits from current developments

Need to generalise the time-structure since Inner and Outer Barrel will run at different readout rates

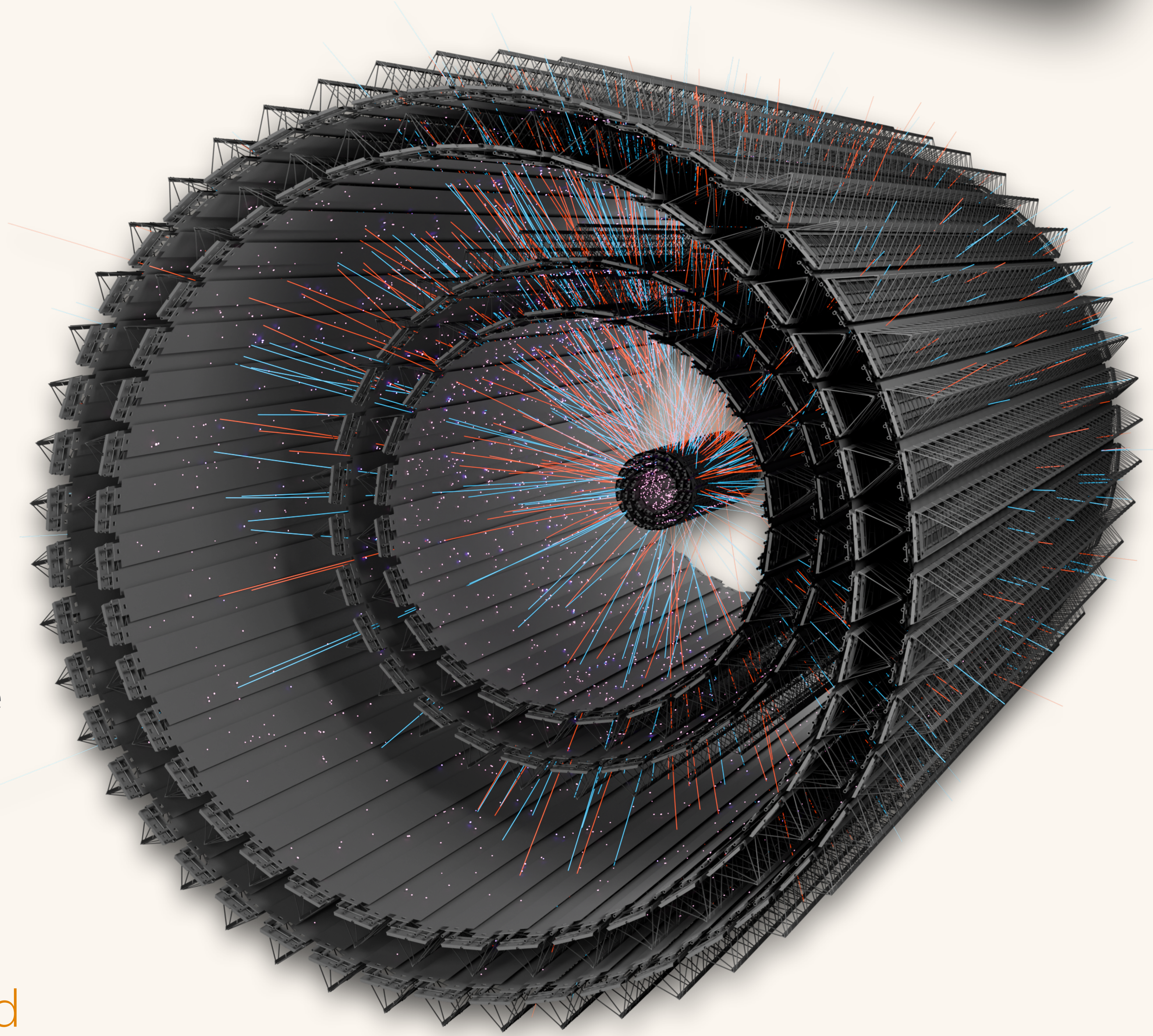


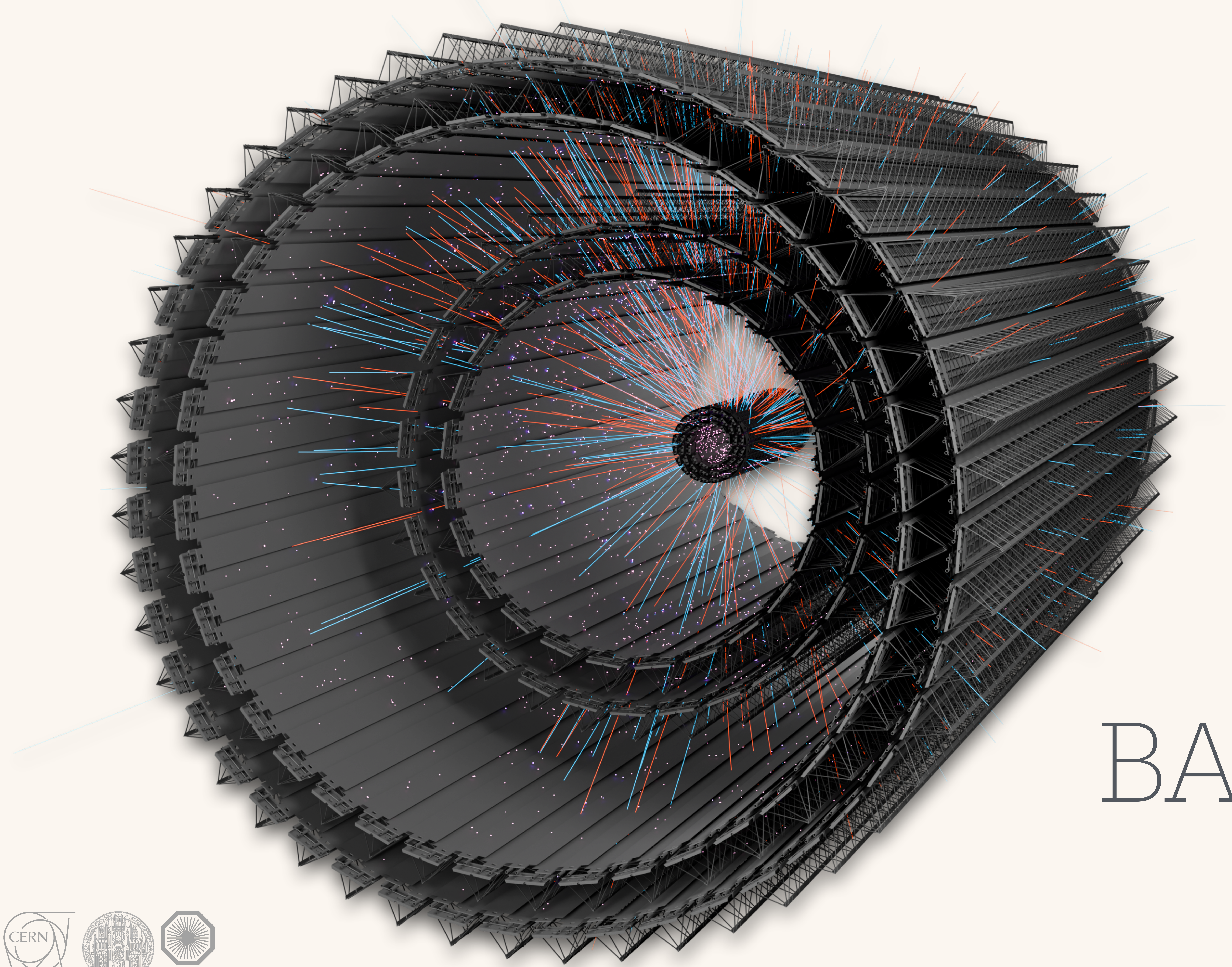


Summary & outlook



- Works well in Run 3!
  - Efficiency above 98% down to 0.2 GeV/c
  - Reconstruction recovers chip-induced inefficiencies providing also better timing information!
- Will be generalised to cope with different readout rate of Inner and Outer Barrel in preparation for Run 4
- Already gives experience with such algorithms now and the ability to estimate computing resource
  - Fully ported to GPUs, speedup factor of 3
  - Needed for future upgrades (ALICE3) in Run 5 and beyond



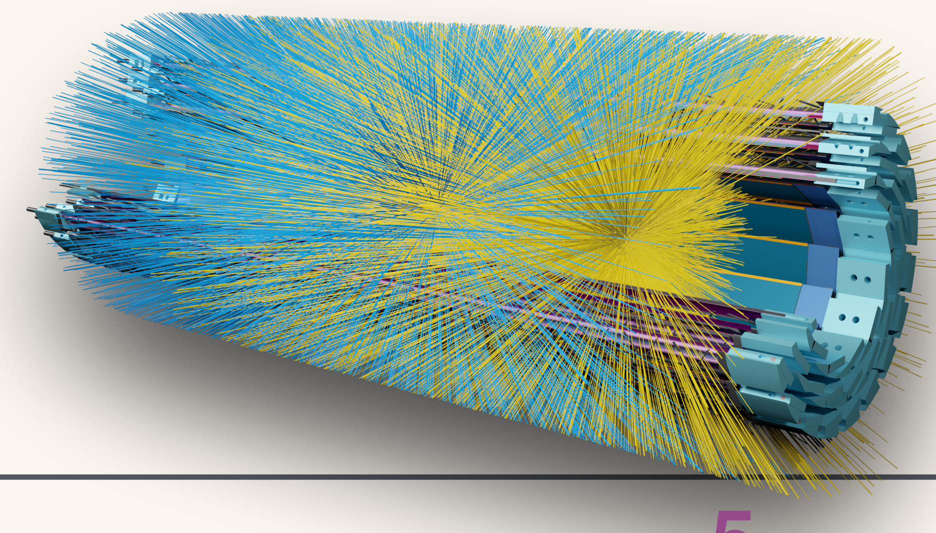


BACKUP

CERN

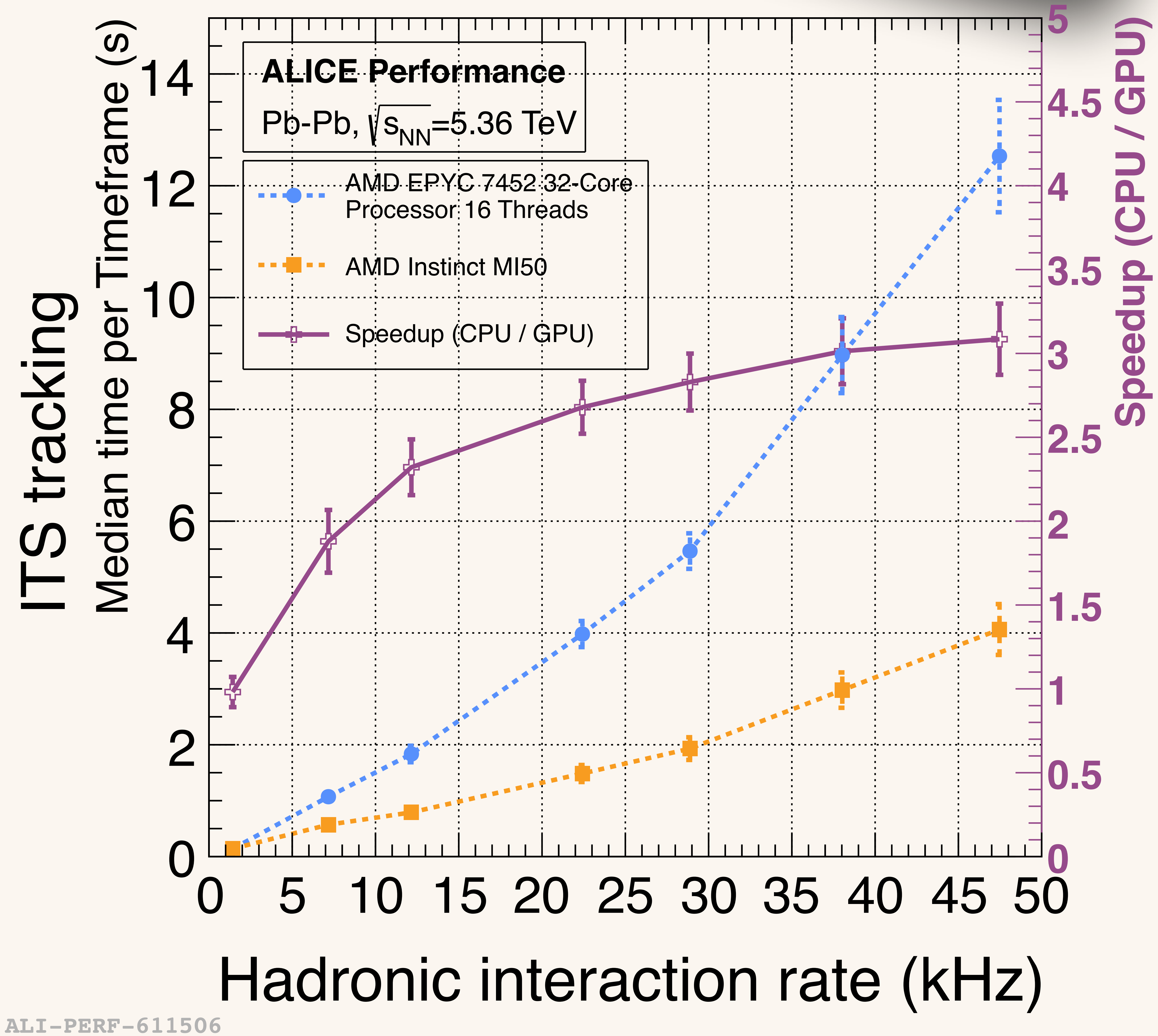
ALICE

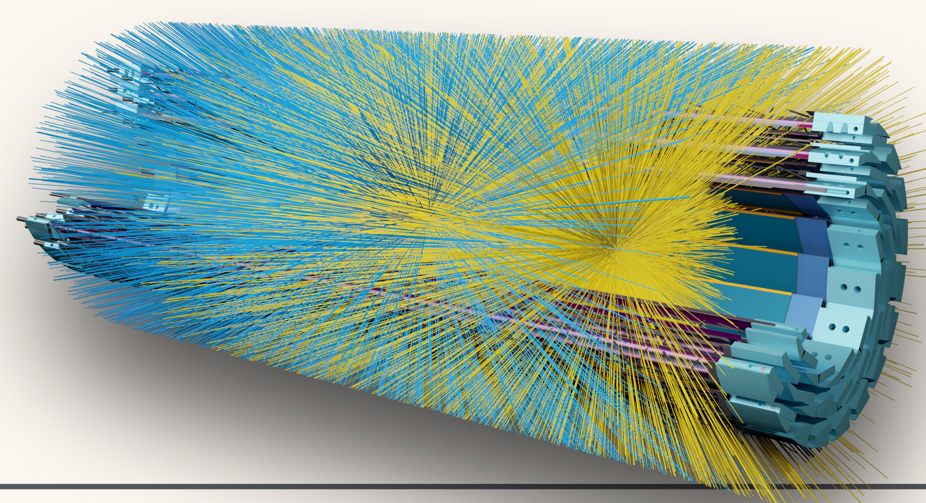
Heterogeneous acceleration



Porting algorithms to GPU

- GPU has much better scaling with interaction rates due to massive parallelism
- Speedup of factor 3 compared to the multi-threaded CPU version with 16 threads
- Speedup of more than 26% for full offline physics production
  - Shortens offline physics production by one week by offloading ITS reconstruction additionally to TPC reconstruction to GPUs!

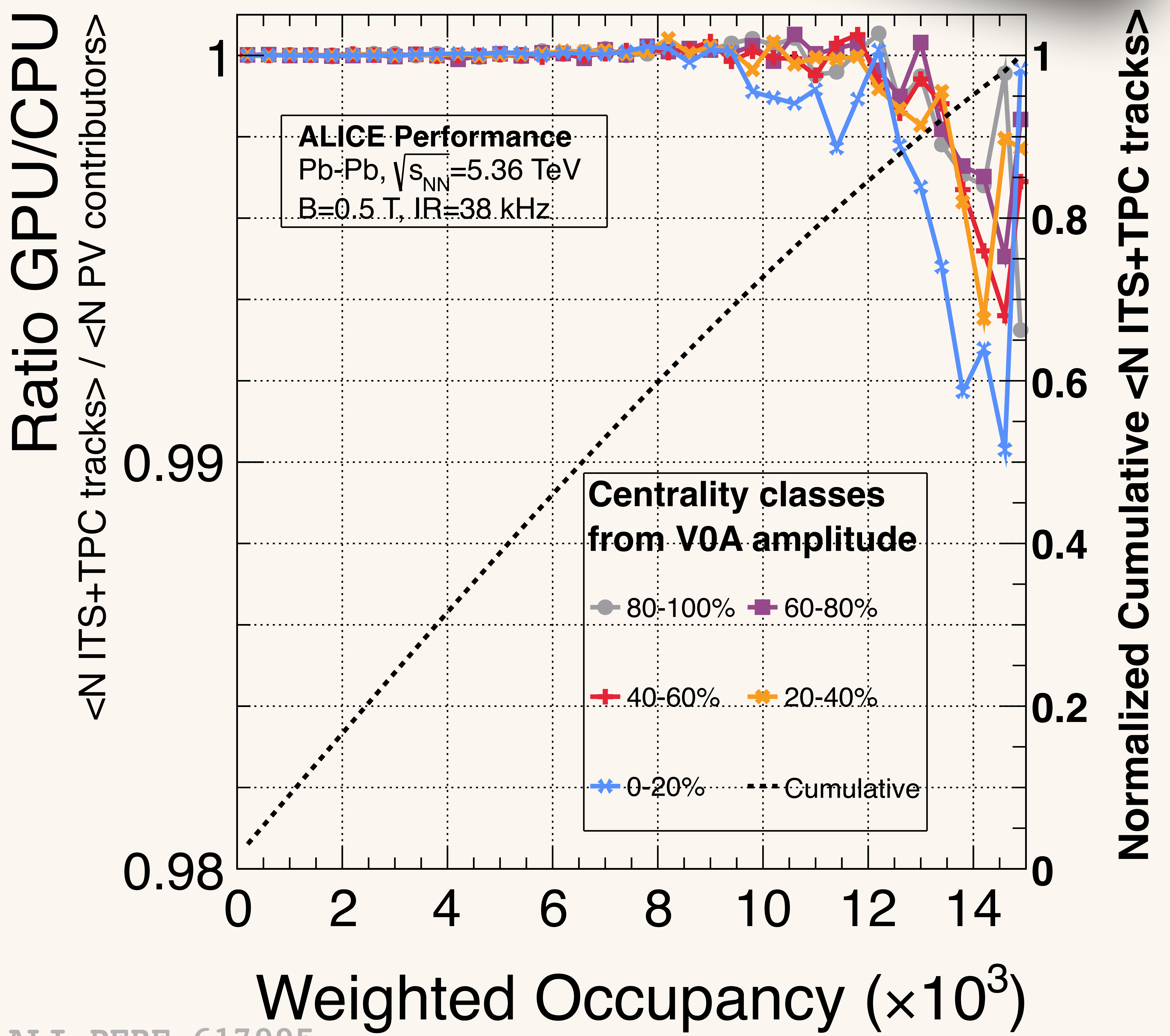




Heterogeneous acceleration

Porting algorithms to GPU

- Comparing the tracking output of the GPU version is almost the same as the CPU version
  - CPU version well validated
  - We have a special mode ‘deterministic mode’ which allows to produce exactly the same output
- In production we observe that due to more memory constraint environment at high occupancy we see very small loss of tracks



Already “good enough” and will be fixed in next iteration