Pattern Recognition at Belle II

International Workshop on Future Linear Colliders 2016 | Morioka, Japan | 8.12.2016

Thomas Hauth for the Belle II Collaboration

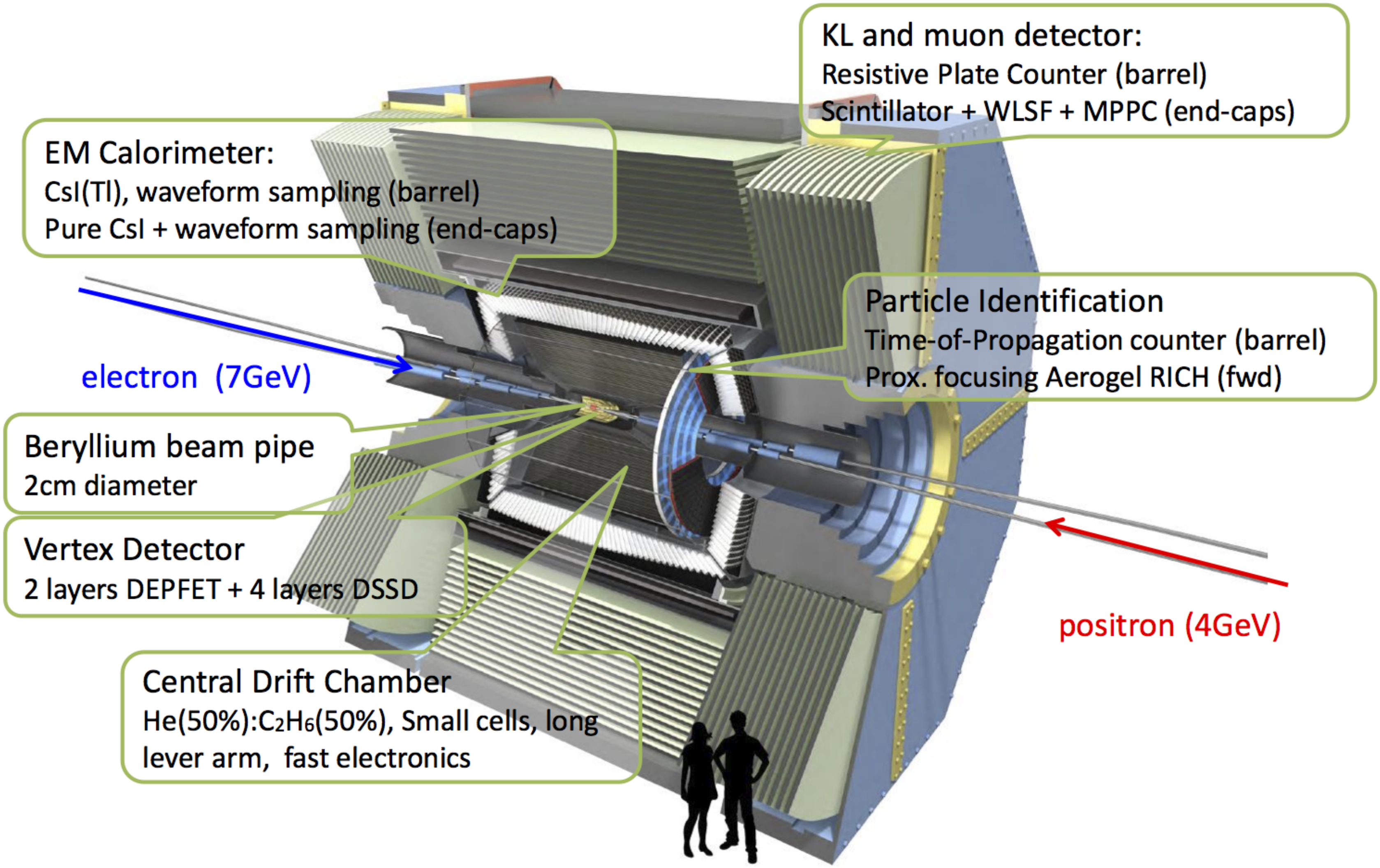
INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK (EKP) · FAKULTÄT FÜR PHYSIK

The Belle II Experiment and its Goal

- KEKB was an electron-positron collider at KEK in Tsukuba/Japan which studied the decay of B mesons at the Y(4S) resonance
- Nobel Prize in Physics 2008 to Kobayashi and Maskawa
- The SuperKEKB collider and the Belle II detector will build on the previous success:
  - Study B meson system in far greater precision
  - Probe for new physics in a wide range of interesting topologies
- The Belle II Collaboration: 681 scientists from 100 institutes in 23 countries !

	KEKB	**Super KEKB**	**Factor**
Instantaneous Luminosity	$2 * 10^{34}$ $cm^{-2}s^{-1}$	$8 * 10^{35}$ $cm^{-2}s^{-1}$	40
Integrated Luminosity	1 ab^{-1}	50 ab^{-1} (projected)	50
Runtime	1998 to 2010	start in 2017	
Detector	Belle	Belle II	
Raw Data	1 PB	100 PB (projected)	100

Belle II Detector

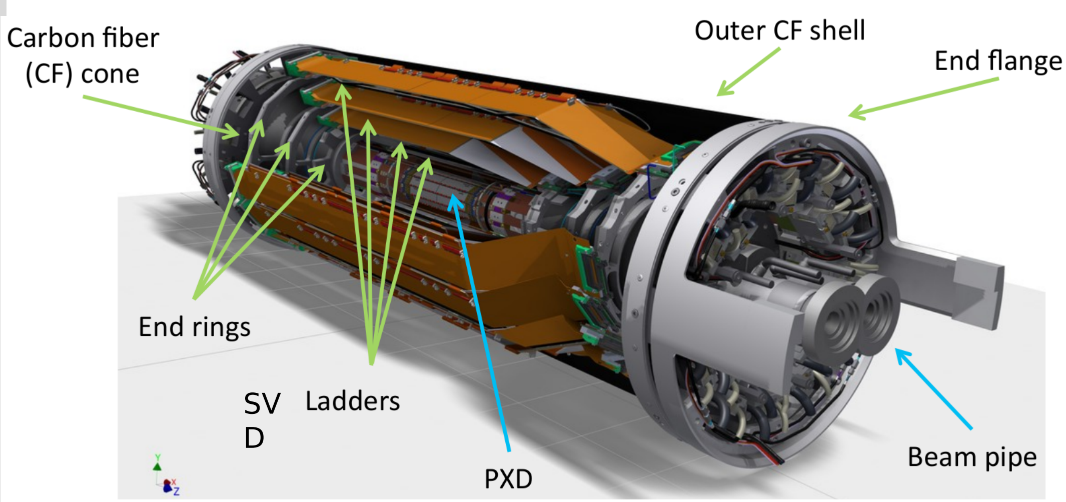


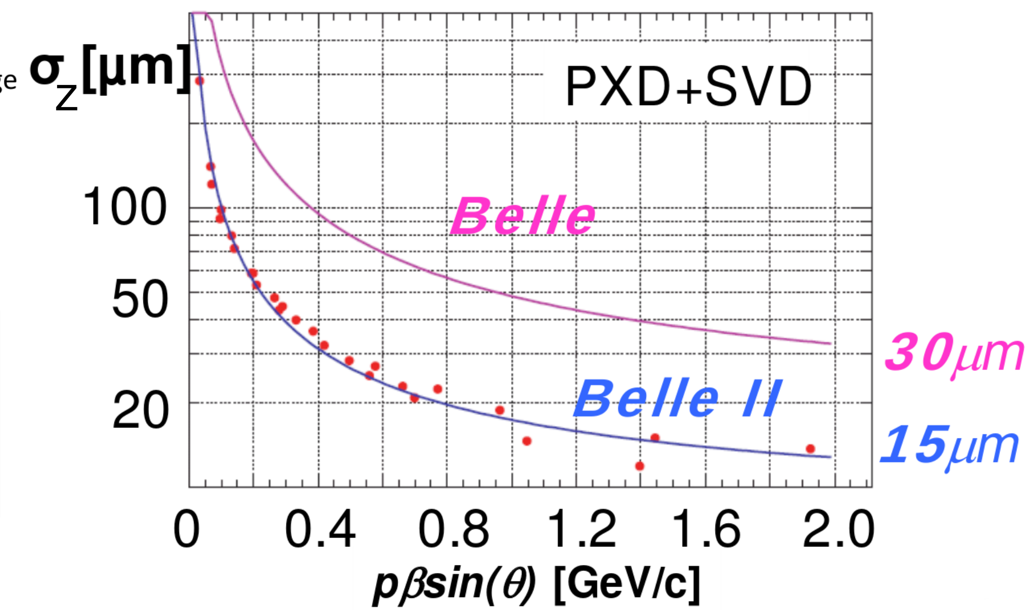
Tracking Detectors - VXD

The Belle II **V**erte**X** **D**etector (VXD) is formed by

- 2 inner DEPFET Pixel Layer (**P**i**X**el **D**etector) : PXD
- 4 outer Silicon Strip Layers (**S**ilicon **V**ertex **D**etector) : SVD

- Very light mechanical structure: ~0.5% X/X_0 per SVD layer
- DEPFET technology - pixel internal amplification allows for very low material budged in PXD layers: ~0.19% X/X_0 per layer
- Compared to Belle: factor 1.5 improvement of the impact parameter over a wide range with the new inner tracking sysem



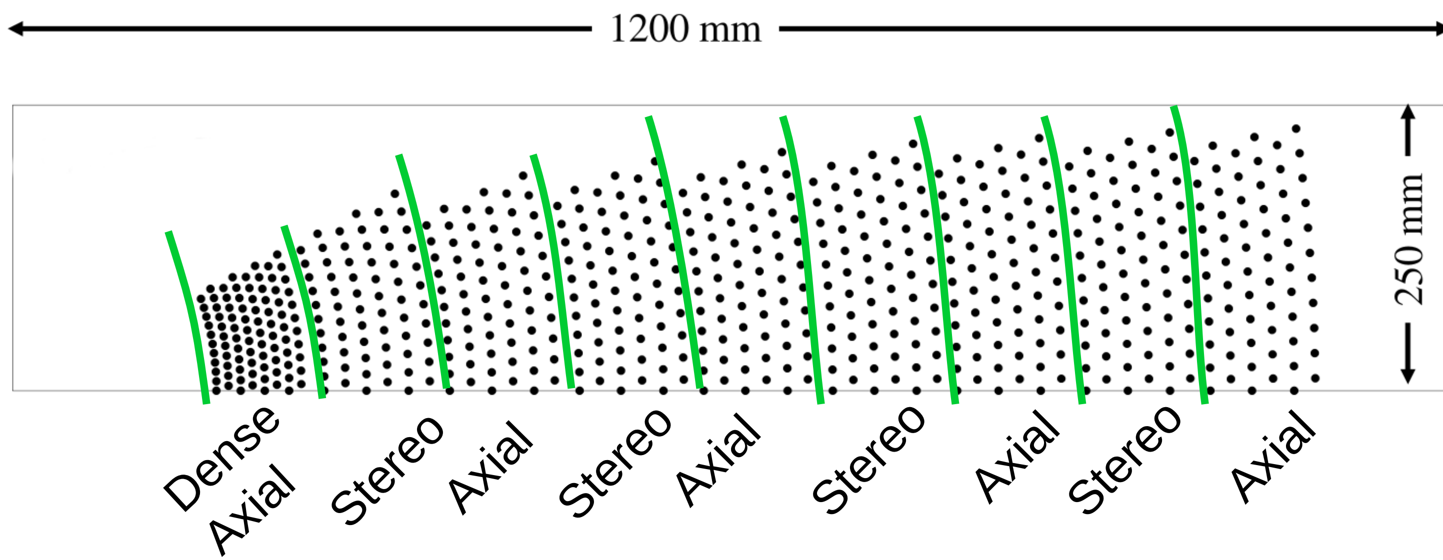


Tracking Detectors - CDC

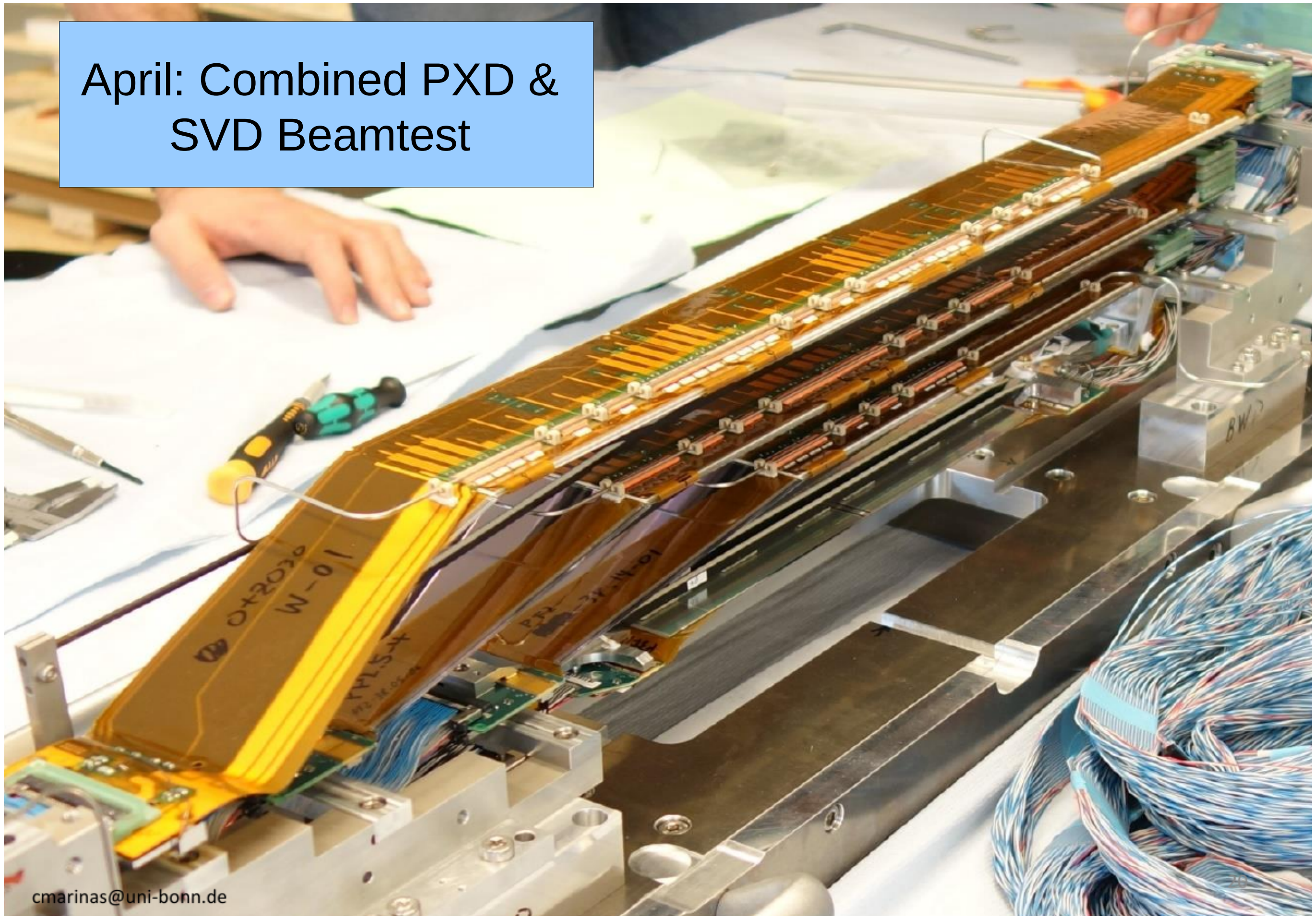
- The Belle II Central Drift Chamber is a significant upgrade compared to the drift chamber of the Belle detector

	Belle	Belle II
Radius of inner cylinder (mm)	77	160
Radius of outer cylinder (mm)	880	1130
Radius of innermost sense wire (mm)	88	168
Radius of outermost sense wire (mm)	863	1111.4
Number of layers	50	56
Number of sense wires	8,400	14,336
Gas	He–C_2H_6	He–C_2H_6
Diameter of sense wire (μm)	30	30

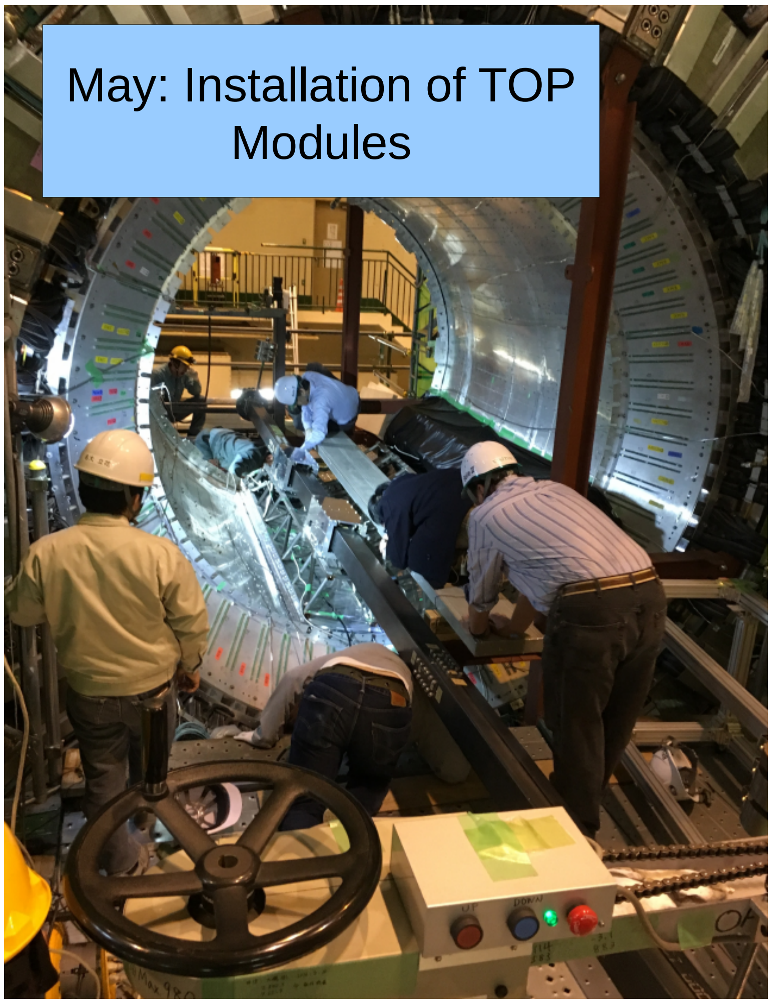
- Larger lever arm for precise momentum estimation
- z-component measurement with stereo layers with wires shifted by 2.6 to 4.2 degrees

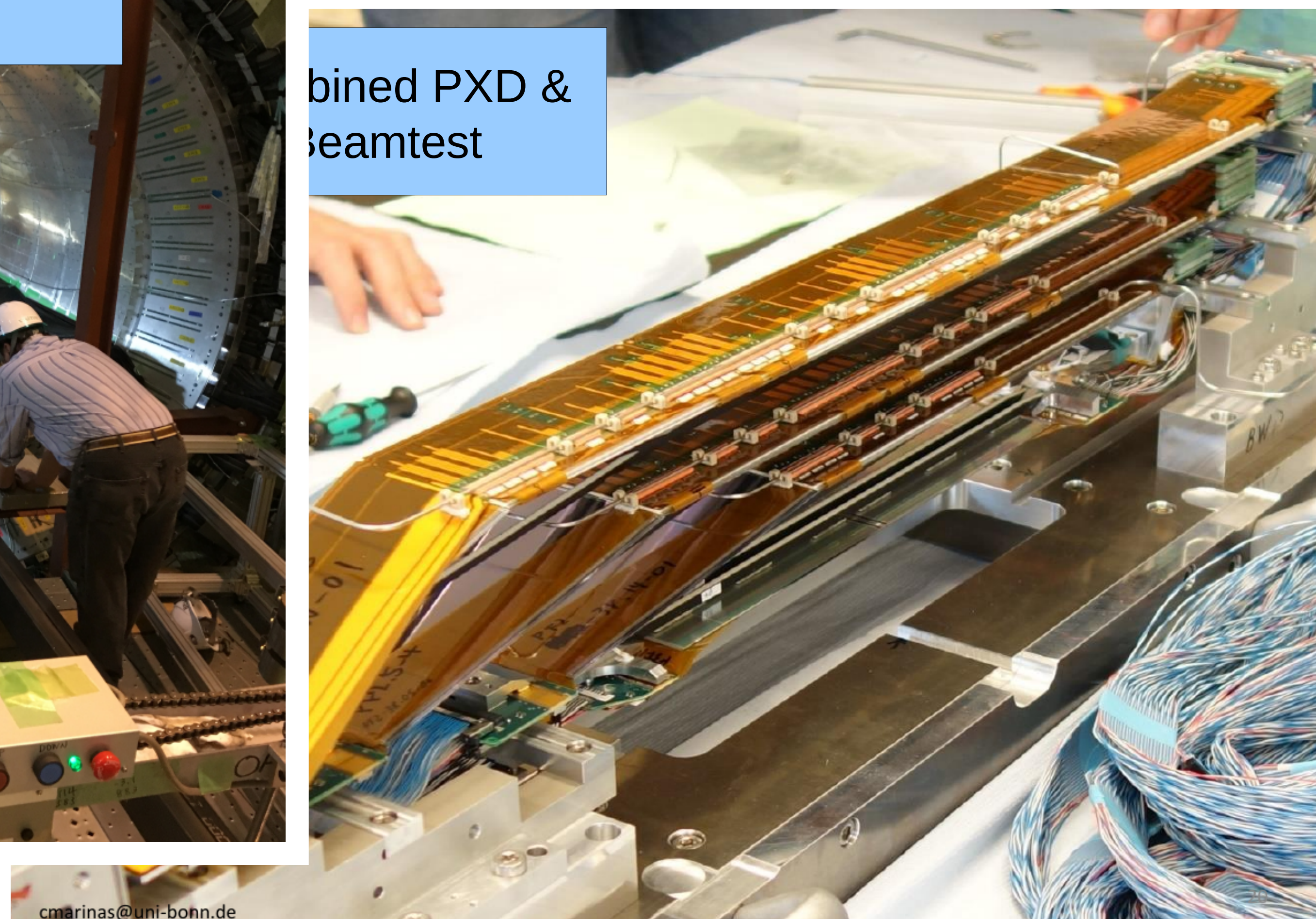


Belle II Detector - Recent Progress

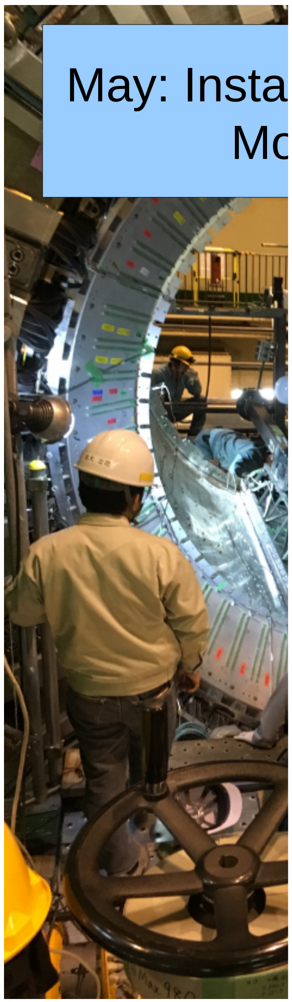


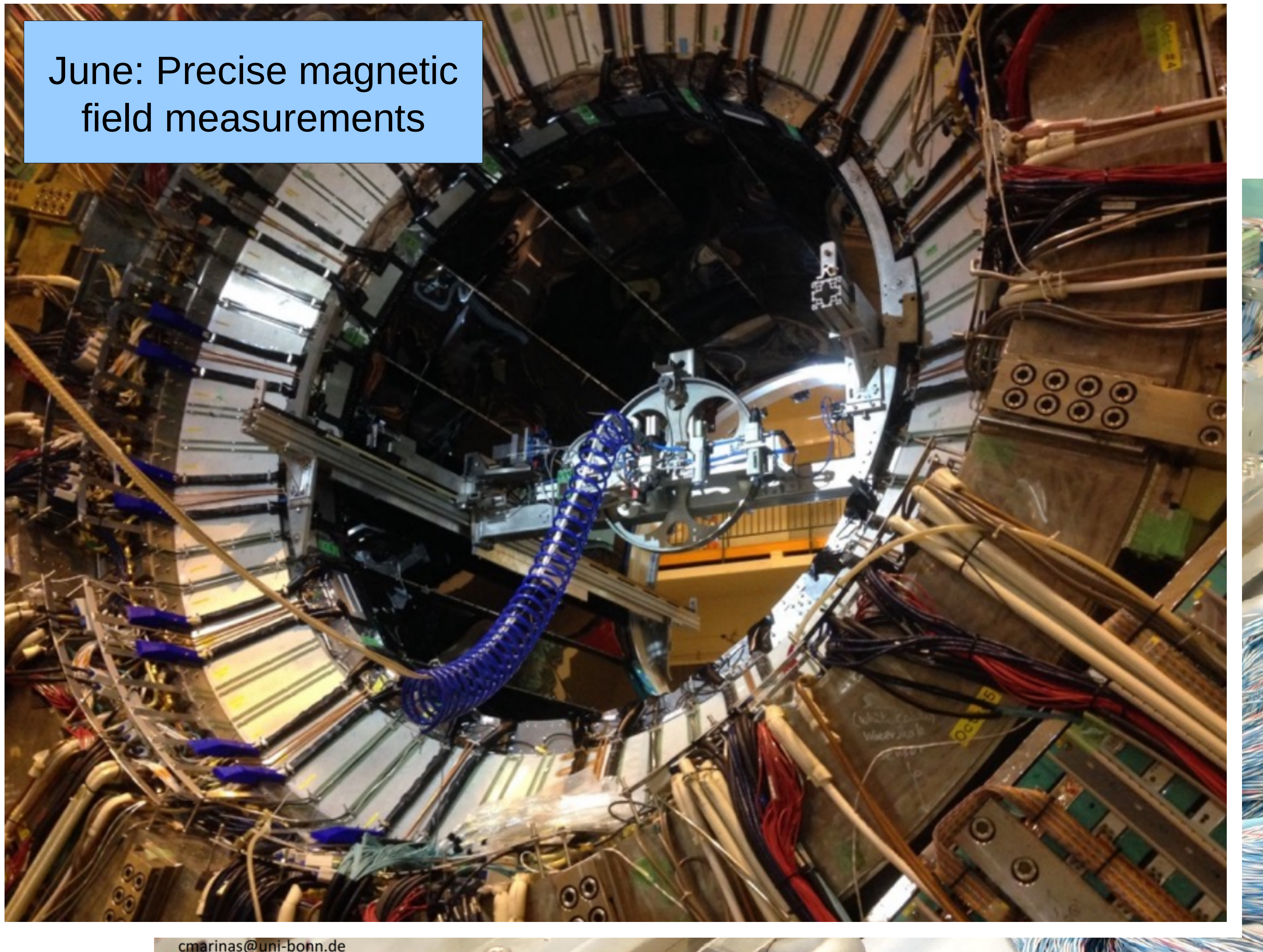
Belle II Detector - Recent Progress



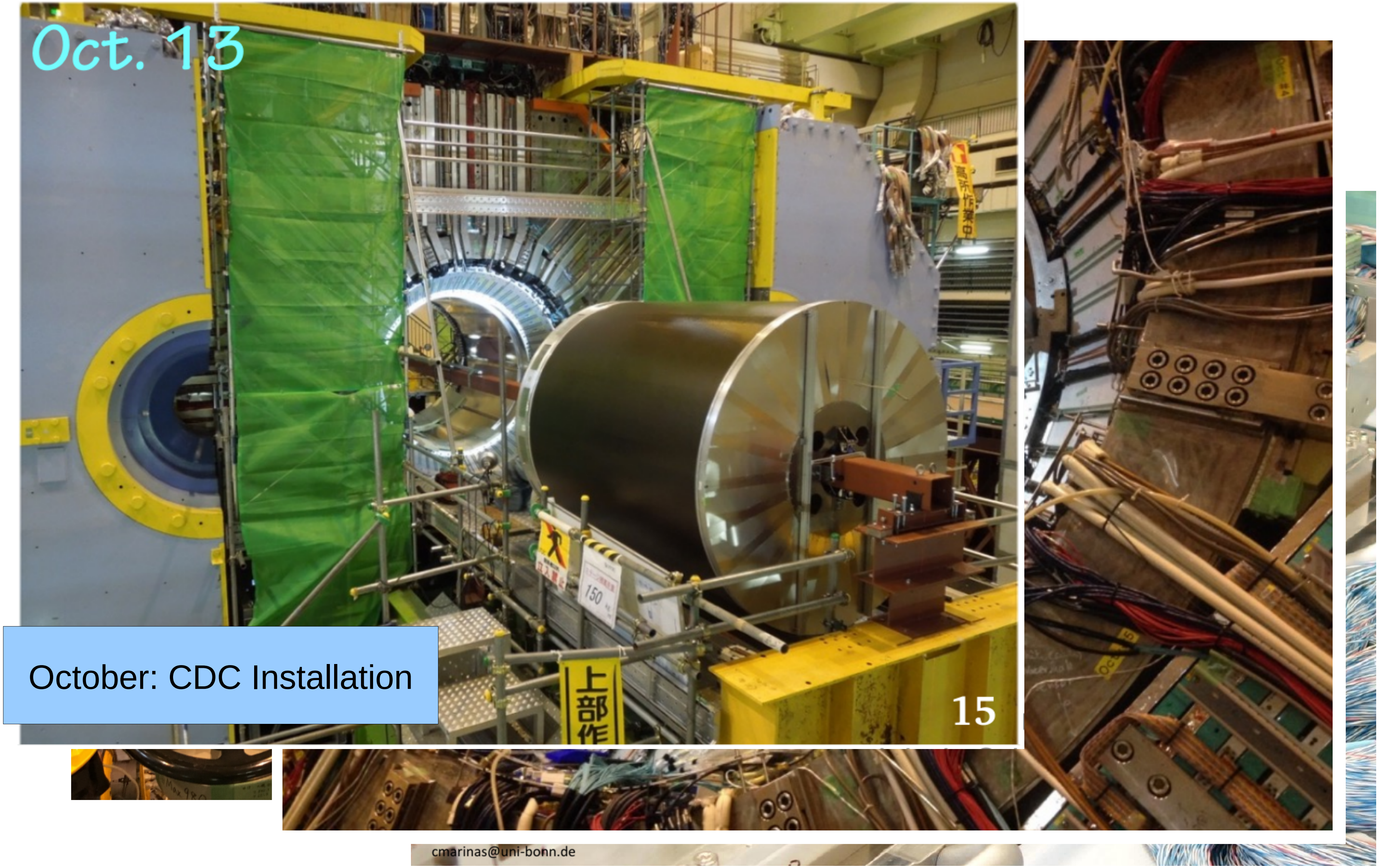


Belle II Detector - Recent Progress



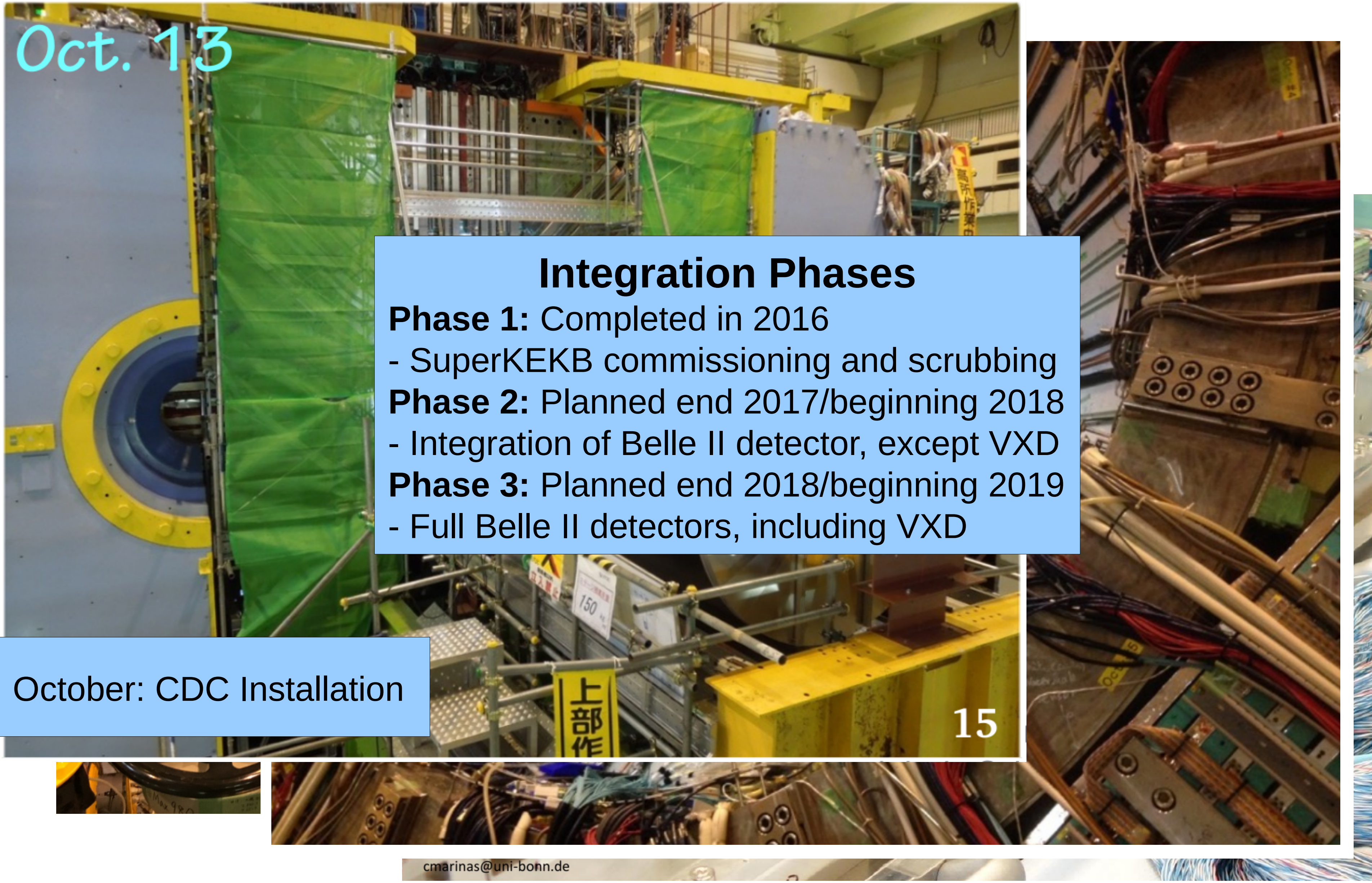


Belle II Detector - Recent Progress



October: CDC Installation

Belle II Detector - Recent Progress



Integration Phases

Phase 1: Completed in 2016
- SuperKEKB commissioning and scrubbing

Phase 2: Planned end 2017/beginning 2018
- Integration of Belle II detector, except VXD

Phase 3: Planned end 2018/beginning 2019
- Full Belle II detectors, including VXD

October: CDC Installation

Tracking Environment & Challenges

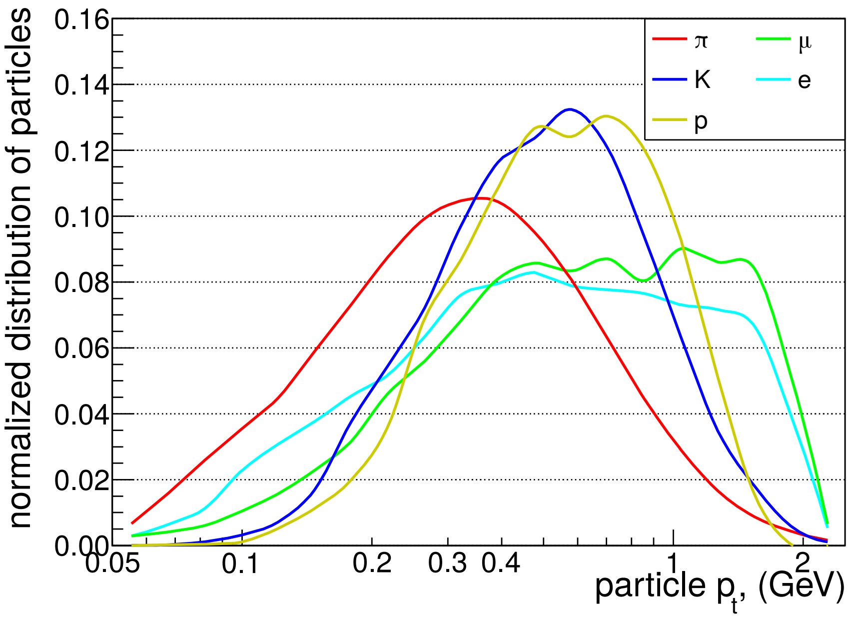
- The most widely used event categories for analysis are Y(4S) → BBar decays
- On average: 11 primary tracks / per event (but additional background hits)
- Here: almost all of the beam energy is converted into the B-Meson pair
- If all visible decay products of the two B-Meson decays can be reconstructed:
  - Very clean topolgies with well-known Kinematics
  - Search for rare decays by assigning all tracks to a particle candidate

Challenges for Tracking Hard & Software:

- Reconstruct **all** tracks, also down to very low-pt regions
- Cope with the huge contribution of background hits
- Low fake rate

Particle types visible in Tracking Detectors of typical Y(4S) event

Particle type	Average fraction
$\pi^{\pm}$	72.8%
$K^{\pm}$	14.9%
$e^{\pm}$	5.8%
$\mu^{\pm}$	4.7%
$p^{\pm}$	1.8%



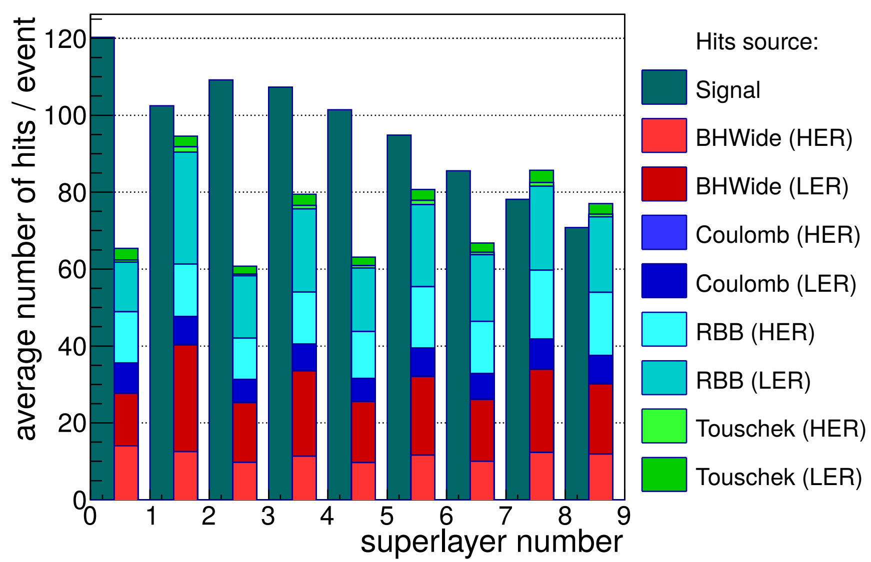
Beam-induced Background

- Compared to KEKB / Belle, the luminosity dependent background has increased by factors
- Space Points (SP) are combinations of u and v clusters on silicon strip detectors
  - In cases of more than 1 hit/sensor: "ghost hits" are created
- Luminosity dependent background:
  - Two-Photon processes (mainly in the VXD detectors)
  - Radiative Bhabha
- Beam-size dependent background
  - Touscheck Radiation
  - Beam-Gas scattering

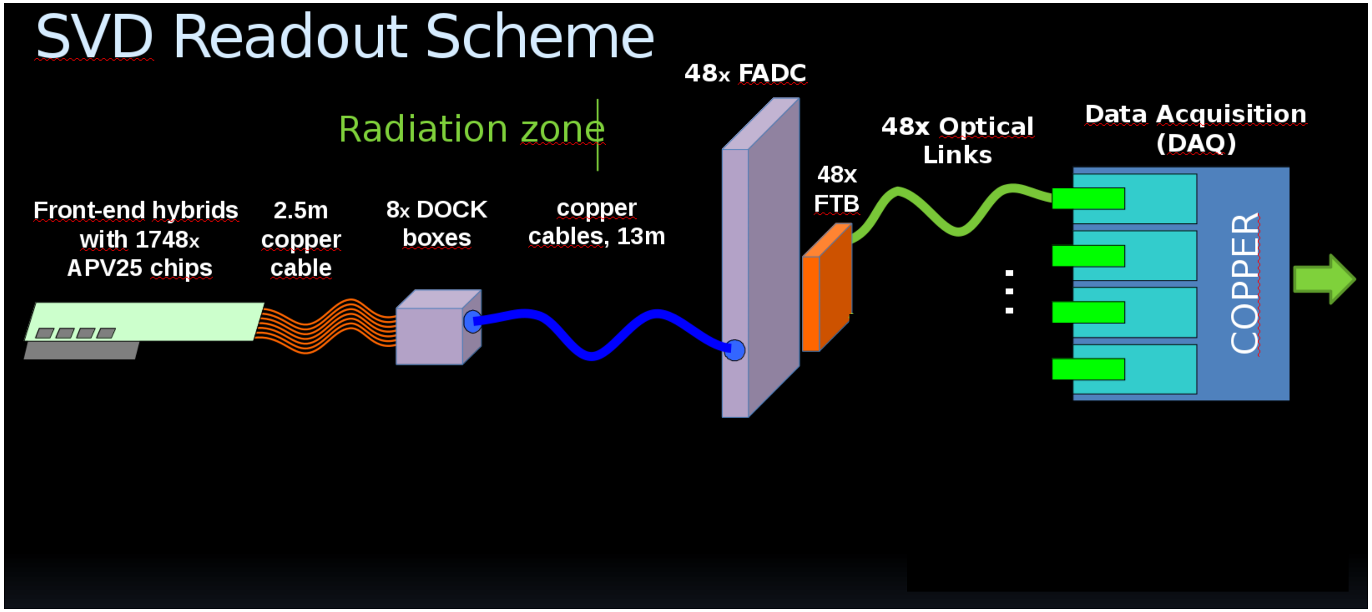
SVD Background Contribution

Case	$\Upsilon(4S)$-only	BG-only	$\Upsilon(4S)$ + BG
L3 strips u/v	49.2/36.7	260.0/121.7	308.1/158.0
L3 clusters u/v	11.8/11.8	39.0/37.9	50.3/49.3
L3 SPs	26.1	233.9	318.0
L4 strips u/v	39.4/29.1	120.3/61.2	159.1/90.1
L4 clusters u/v	12.7/12.6	29.9/26.7	42.5/39.2
L4 SPs	22.5	100.5	143.1
L5 strips u/v	37.3/28.5	122.7/67.2	160.1/95.8
L5 clusters u/v	12.3/12.1	35.0/30.5	47.3/42.7
L5 SPs	19.2	99.3	132.3
L6 strips u/v	38.3/28.6	134.6/76.8	172.9/105.4
L6 clusters u/v	12.4/12.2	42.1/36.3	54.4/48.5
L6 SPs	17.0	100.8	127.9
Average strips/layer u/v	164.3/122.8	159.4/81.7	200.1/112.3
Total clusters u/v	49.2/48.7	146.0/131.3	194.4/179.6
Total SPs	84.8	534.6	721.3
	Signal	Noise	

CDC Background Contribution



SVD Hit Timing Assignment



APV25 Readout Chip:

- Developed for CMS (70k units installed)
- Supports multi-peak readout mode to readout several (6) samples along signal shape

FADC

- Analog to digital converters
- Central FPGA is an Altera Stratix IV GX for further signal processing

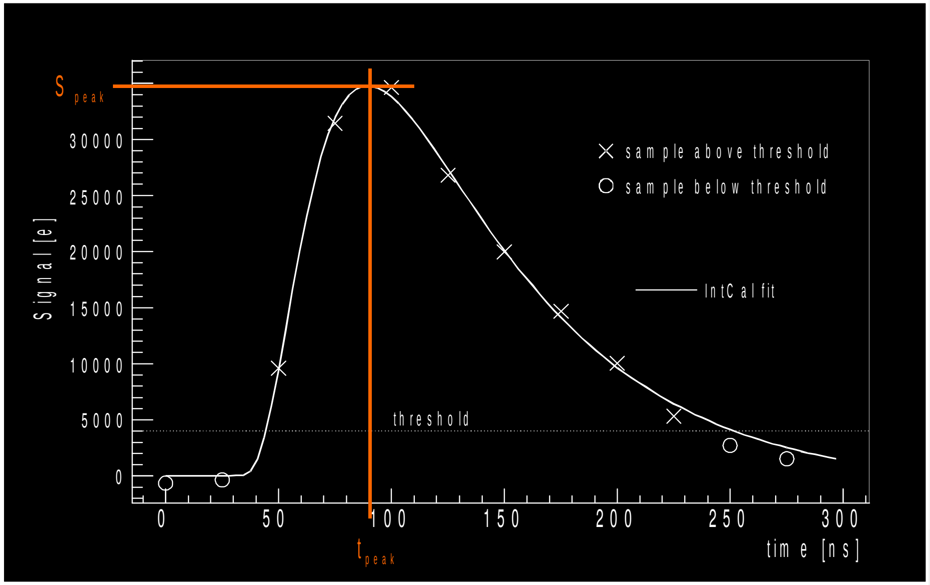
SVD Hit Timing Assignment

Readout Signal

- The APV25 has a CR-RC shaper which generates a non-zero signal with a certain duration (tunable, ~160 ns for one MIP)
- 6 samples around the maximum are recorded and read-out by the FADC
- Hit threshold an important tuning parameter:
  - Higher: better noise suppression
  - Lower: better efficiency

Fit of Hit timing

- Using a numerical fit: an accuracy of 2 ns RMS was achieved offline (not possible to implement on FPGA)
- New implementation using a neural network for the hit time fit: shows similar performance as the numeric fit
- Work ongoing to port this neural network method to the FDAC's FGPA

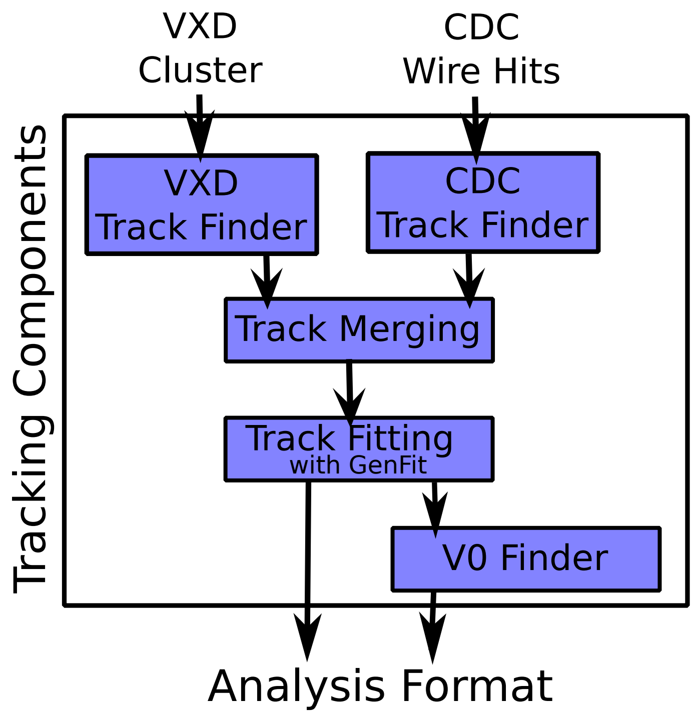


Hit timing information can be used to reject out-of-time background hits

Further studies with test beam and real beam data are necessary to understand the performance of the hit time determination

Belle II Modular Tracking

- The Software Framework of Belle II (`basf2`) allows to implement processing steps in a so-called Modules which are not coupled and exchange input and output data via a `DataStore`
- Due to the very different nature of the two major tracking systems (silicon based and drift chamber), two different track finders were implemented
- All stages of the Belle II Tracking software are implemented in independent framework modules which:
  - fits well with our (geographically) distributed development model
  - allows to re-arrange parts of the tracking software for different use cases (online / offline / beamtest scenarios)
- A common exchange format class (*RecoTrack*) is used to transfer pattern recognition and fit results between modules
- While the earlier Module (for example VXD Track Finder) are specifically developed for specific hardware, all downstream modules (like TrackFitting) can operate on the common *RecoTrack*

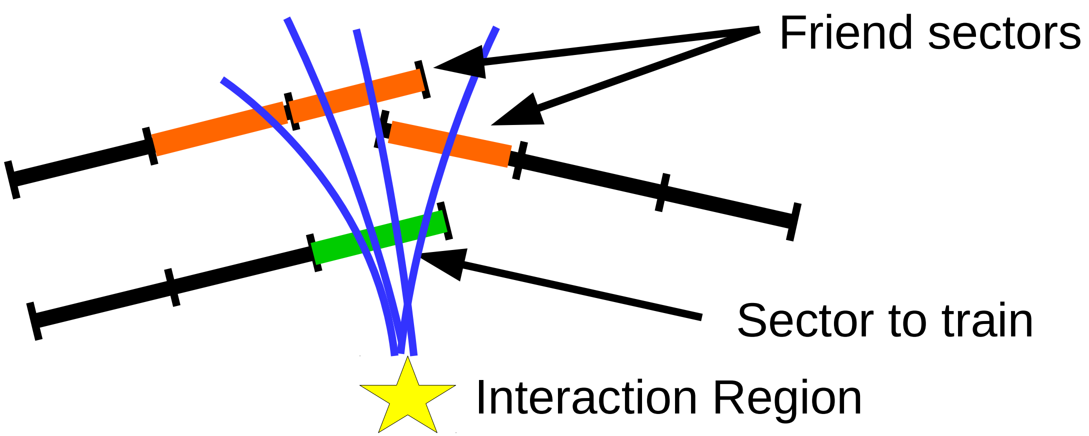


Belle II VXD Tracking

- The VXD Tracking has to operate in two scenarios:
  - High-Level-Trigger online reconstruction of 4 SVD layers
  - Offline reconstruction with 2 PXD and 4 SVD layers
- Very large combinatorics, due to huge number of background hits, especially in the innermost layers & ghost hits in the SVD

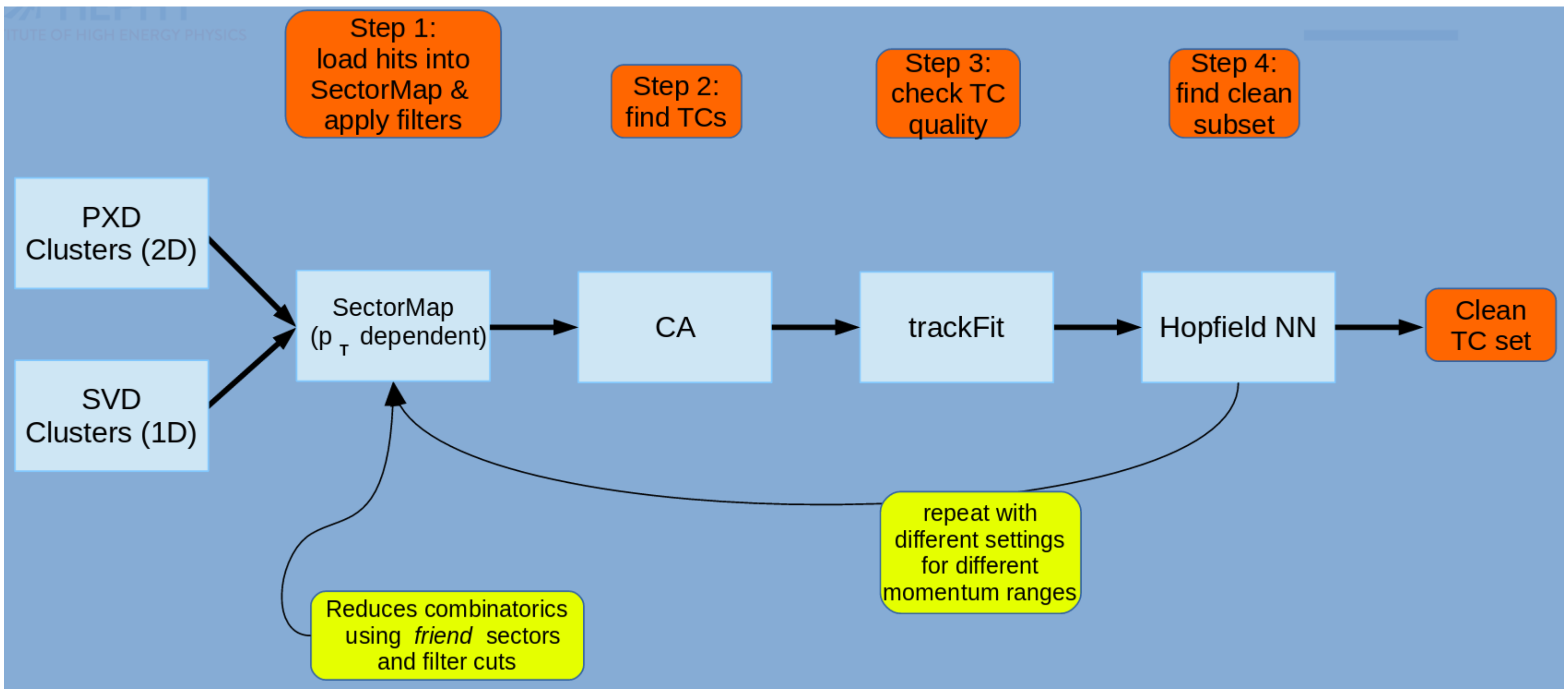
SectorMap – Concept to reduce combinatorics

- Each sensor is split into sectors (~10 per sensor)
- Using tracks from simulated Monte Carlo events, the possibility to hit the next sector is evaluated
- Connections between sectors are stored up to 4 sectors deep
- SectorMap forms an acyclic directed graph, couple of Megabyte to store to disk
- The SectorMap also stores cut values for two and three-hit filter combinations and is depending on different Pt values



Belle II VXD Tracking

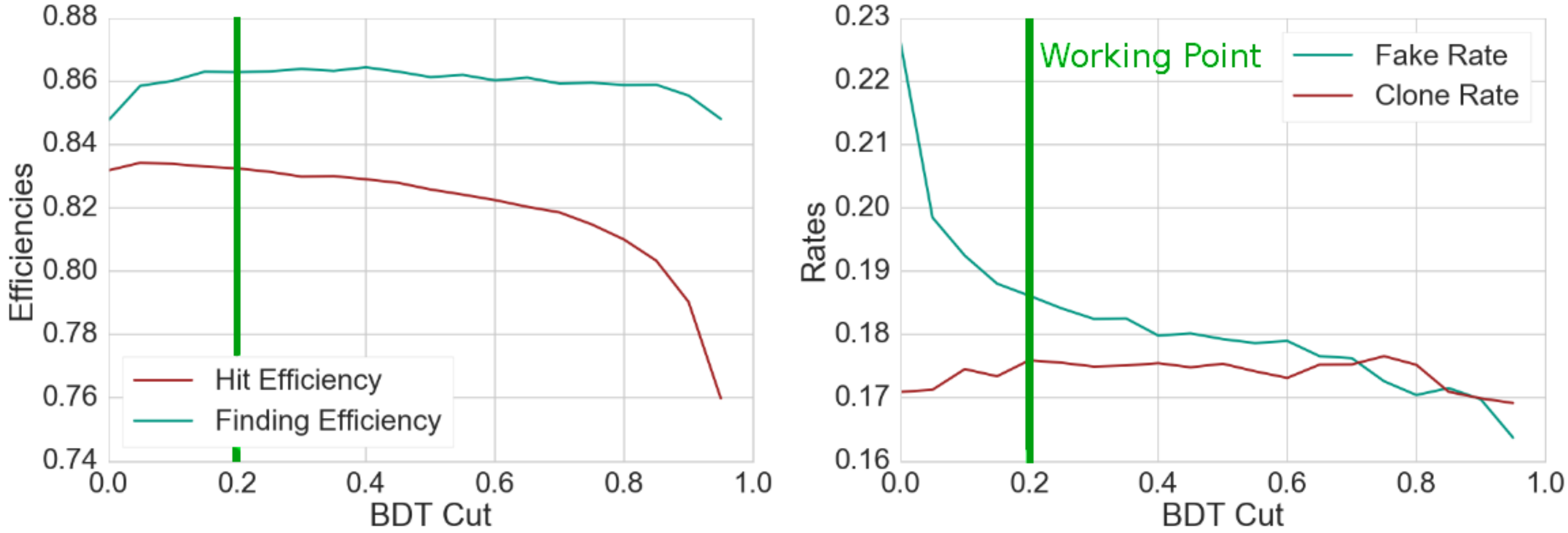
- A Cellular Automaton (CA) is used to explore the possible combinations of hits to form track candidates
- Quality of the these track candidates is evaluated with a fast fitting method
- A Hopfield neural network finds the best subset of available track candidates (in case of track with shared hits)



- **The VDX Track Finder has successfully been used in multiple PXD & SVD beam tests**
- We are currently working on a refactored version to allow for more modularity and flexibility: for example to support multiple track fit methods

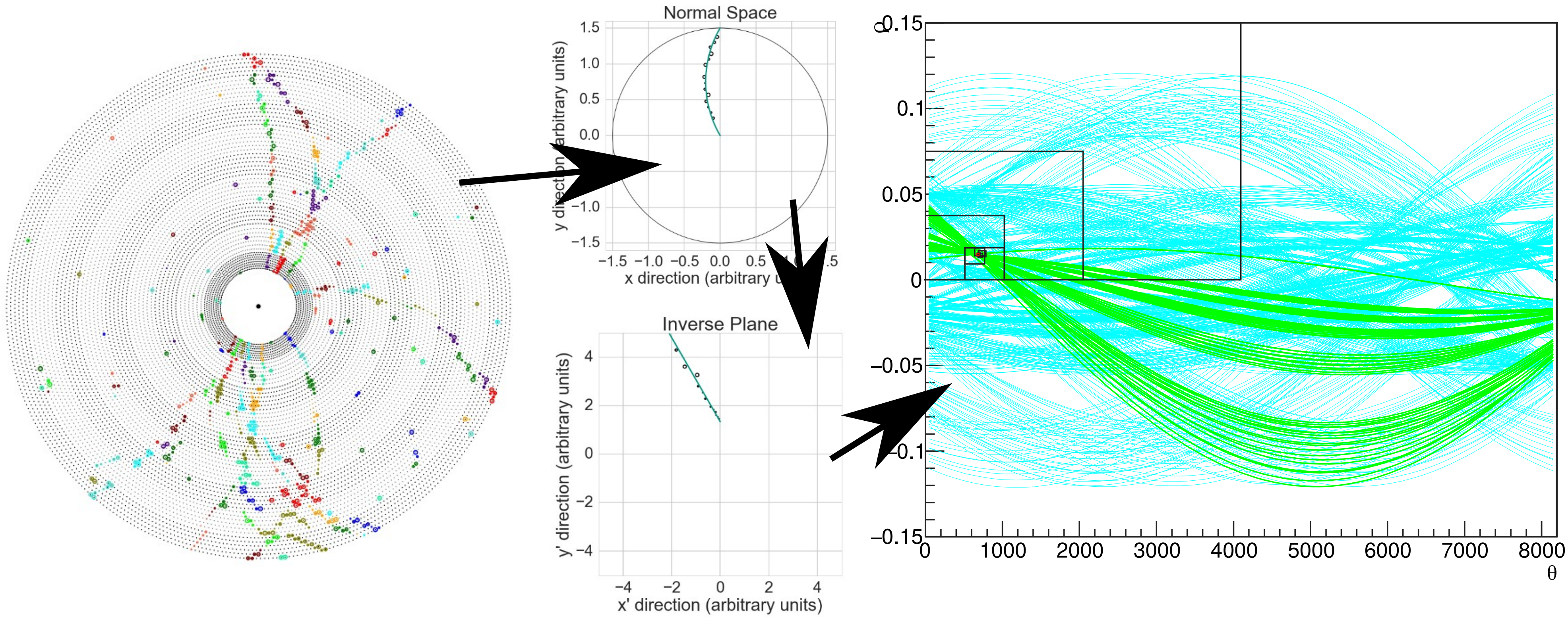
Belle II CDC Tracking: Background

- Around 40 percent of all CDC hits are results of machine-induced background
- Background hits have distinctly different properties:
  - Isolated
  - Don’t form larger clusters
- A boosted decision tree classifier is trained on: total and mean number of neighboring clusters, total and mean drift length, super-layer number



- The background filter is able to classify background hits reliably
- With the chosen cut value of 0.2, the finding efficiency and the fake rate can be improved
- Only very small loss in track hit efficiency observed

CDC Track Finding
Legendre-based Finder



- All axial CDC Hits are transformed to Legendre-space where each forms a sinusoidal curve
- Fast iterative quad tree algorithm is used to search for areas of high curve densities → parameters shared by one track
- Takes the drift length and the left/right ambiguity into account

Legendre-based Finder

- Multiple search iterations are performed to find various classes of tracks
- Numerous optimizations of the quad tree search algorithm:
  - Sliding bins are re-centered on the center of the hits for each quad tree search level → recovers hits otherwise lost to border effects
  - Monte-Carlo simulation was used to parametrize the optimal search depth depending on the track’s transverse momentum

    → Hit spread in Legendre-space of low-pt tracks is larger due to energy loss

    → Quad tree search depth depends on the Pt of the tracks
- Fast circle fit for track quality estimation and transformation of all unassigned hits using the point-of-closest-approach of each track

  → Allows to assign hits to tracks which are slightly displaced

- **Legendre-based search is a track finder targeting complete tracksoriginating from the interaction point**

Local Finder

- **Build segments from individual hits in each super layer**
  - A graph of hits in one CDC super-layer is created
  - Graph search algorithm: cellular automaton searches for connected entries (hits) in the graph which can belong to one track segment
  - All found track segments are fitted with a fast circle fit
- **Build tracks from connected segments**
  - A graph of segments is build
  - Neighboring segment pairs are connected in the graph
  - Loose feasibility cuts first, then judge compatibility by combined $\chi 2$ of the trajectories

Local Finder does not rely on any track origin: designed for displaced tracks and short tracks

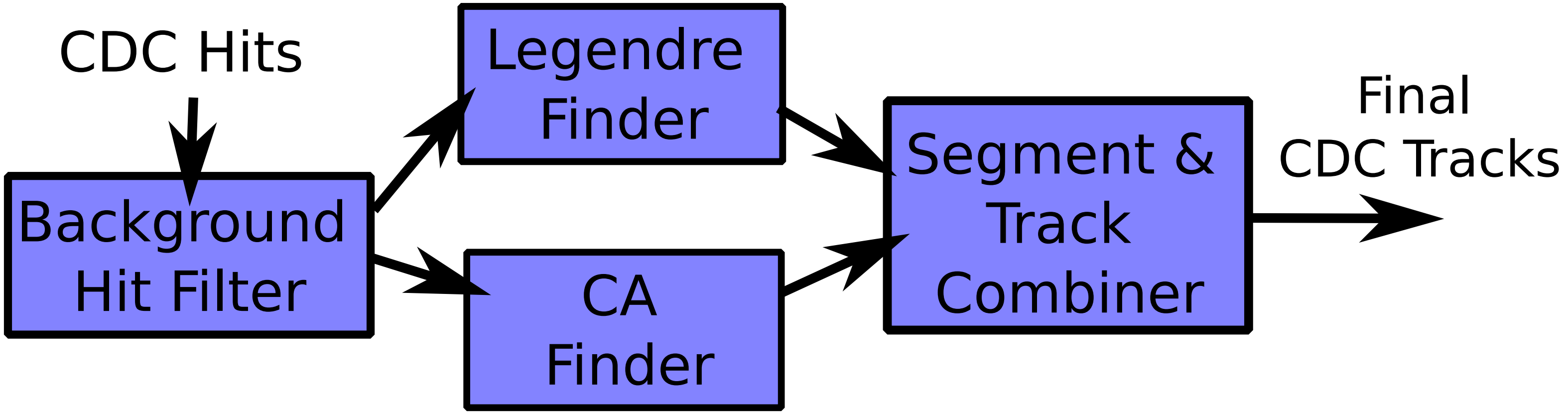
CDC Track Combination

To combine the tracks found by the global finder with the segments found by the local finder:

1) Combination of track and segments which share at least one hit
2) Removal of fake segments
3) Marking of segments which are not long enough to form a full track but are still valid and not fake
4) Matching of the remaining segments among themselves and with the tracks to find possible combinations
5) Filtering of track-segment combinations for fake tracks

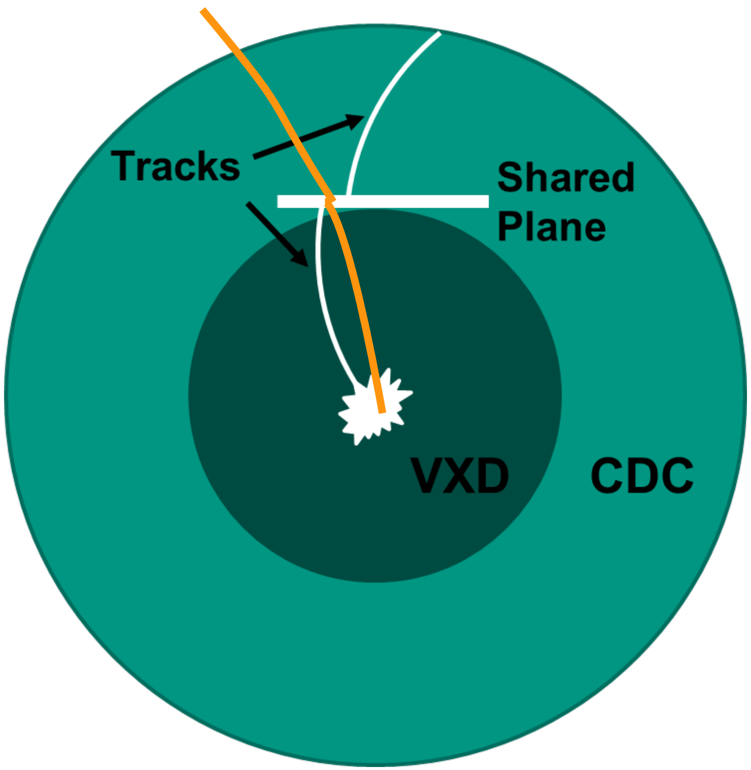
The filtering process uses boosted decision trees to find the optimal selection.

This combination and filter sequence proved to give the best results in terms of finding efficiency, hit efficiency and purity of the final tracks



VXD & CDC Track Merging

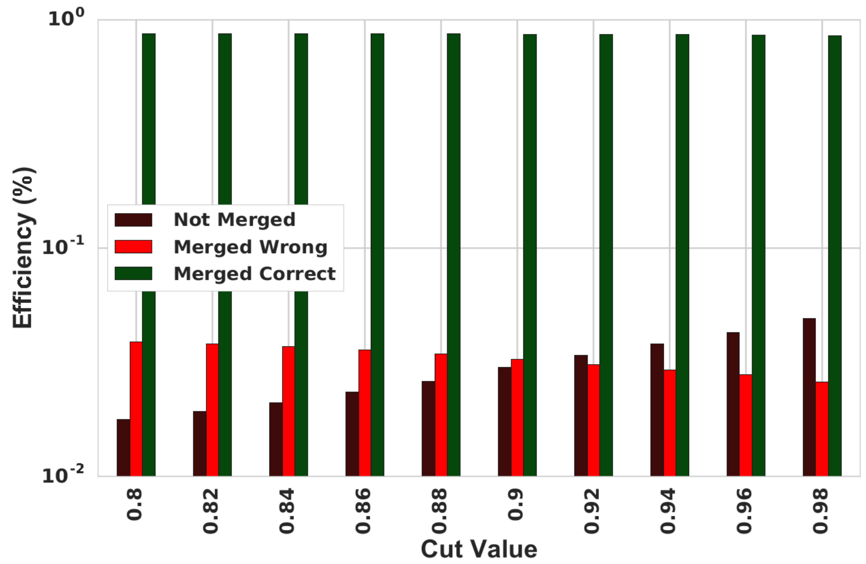
- Tracks found in the VXD and CDC detectors need to be combined to form one consecutive track
- The merging algorithm must handle these cases correctly
  - Merge the correct VXD & CDC track parts, if both were found
  - Don’t merge VXD & CDC track part , if VXD or CDC track part was not found
  - Don’t merge unrelated VXD & CDC track parts



BDT-based Track Merging

- A novel method using boosted decision trees (BDT) has been developed
- The BDT classifier is trained on valid combinations of CDC and VXD tracks
- The input track parameters are the estimations of the track finders, which are more coarse then a Kalman fit procedure but sufficient for the BDT
- Among the BDT input variables are:
  - Transverse Momentum of Tracks
  - Charge Difference
  - Theta & Phi Angle (and absolute differences)
  - Vertex position of both tracks
- For each combination, a classifier output from 0 to 1 is returned
- Combinations are created based on their classifier output
- **Advantages:**
  - Better merging efficiency than distance-based
  - Very fast runtime: the BDT-Combiner can operate without a track fit of the individual VXD and CDC tracks
- **Planned Improvements:** global optimization of track combinations in one event using a Hopfield network

Merging Efficiency as a function of the BDT cut value



GenFit package

- Belle II plays an active role in the experiment-independent fitting package GenFit

    Now on GitHub: https://github.com/GenFit/GenFit
- User community, not only in HEP: Panda, SHIP
- Lightweight interface for material, magnetic field and measurements to integrate GenFit in basf2
- Both VXD and CDC hits are fitted as one track via GenFit
    - We can also fit the T_0 of each event using GenFit and CDC drift times
- GenFit offers multiple fit variants: Kalman Fit or Deterministic Annealing Fit to reject outliers
- We use the DAF fit as our default for final tracks

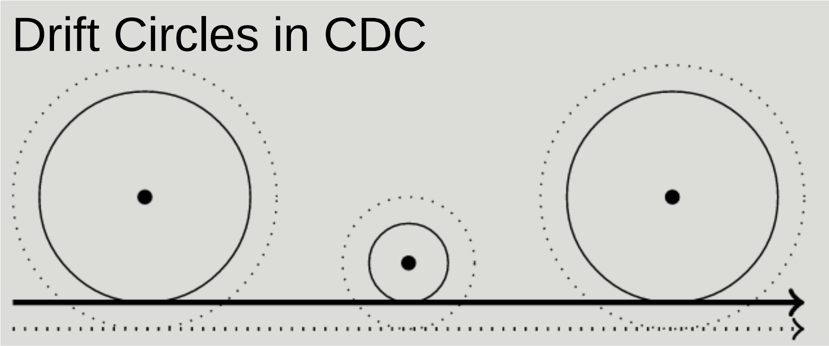
CDC Track Finders and GenFit successfully used throughout this year's cosmics data taking with the CDC !

Fitting T0 with GenFit

- As a drift chamber is a device which actually measures time (and position is inferred), it can also be used to fit the T0 time of one event
- The time is added as one additional parameter to fit (in addition to the track's parameters)
- Level-1 trigger provides a timing resolution of ~10 ns (bunch x every 2ns)
  - The TOP detector has much better timing resolution, IF a track is found there
- If we know T0 to ~2ns, we can assign the collision event to a specific bunch crossing

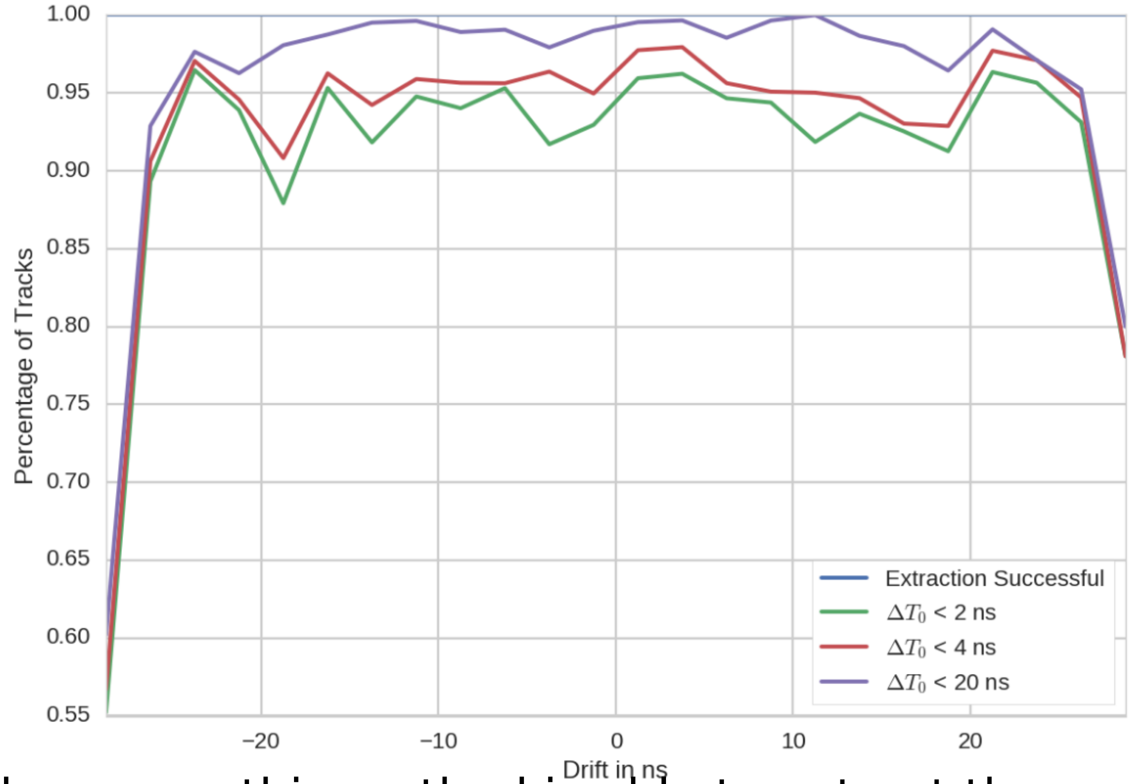
Procedure

- A "good track" (high pt) is picked and the trigger timing is used for the initial fit
- The difficulty is to achive a T0 fit which converges:
  - If T0 is very off, the drift circles (and there CDC "hit positions") are very wrong
- Iterative grid search approach with this track:
  - 4 T0 seeds are created
  - 2 iterative fits in each grid cell
  - continue with best result
  - If fit converged use this as t0



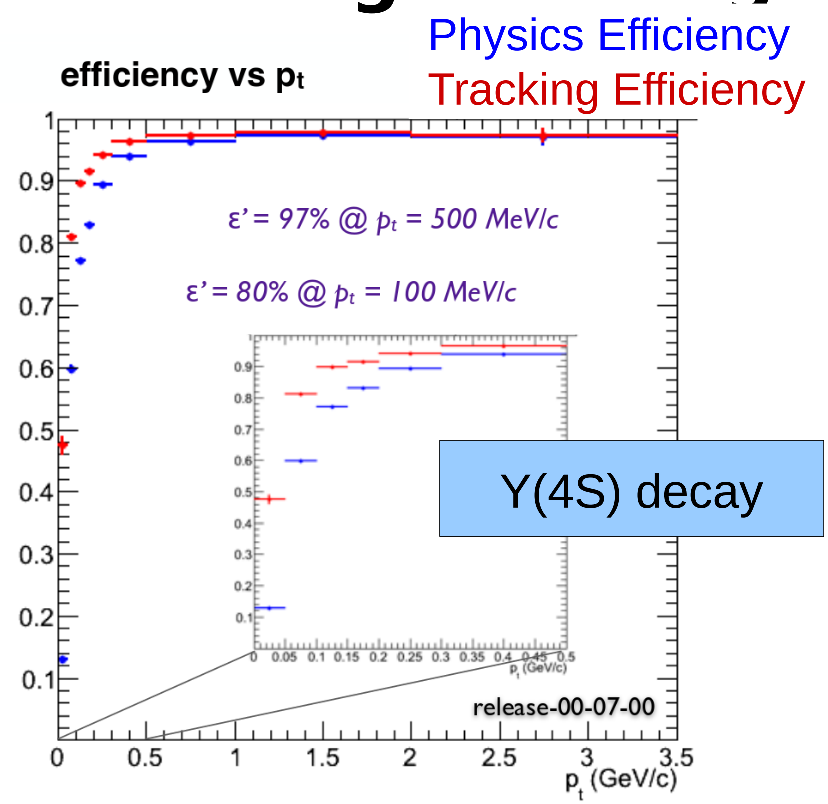
Fitting T0 with GenFit

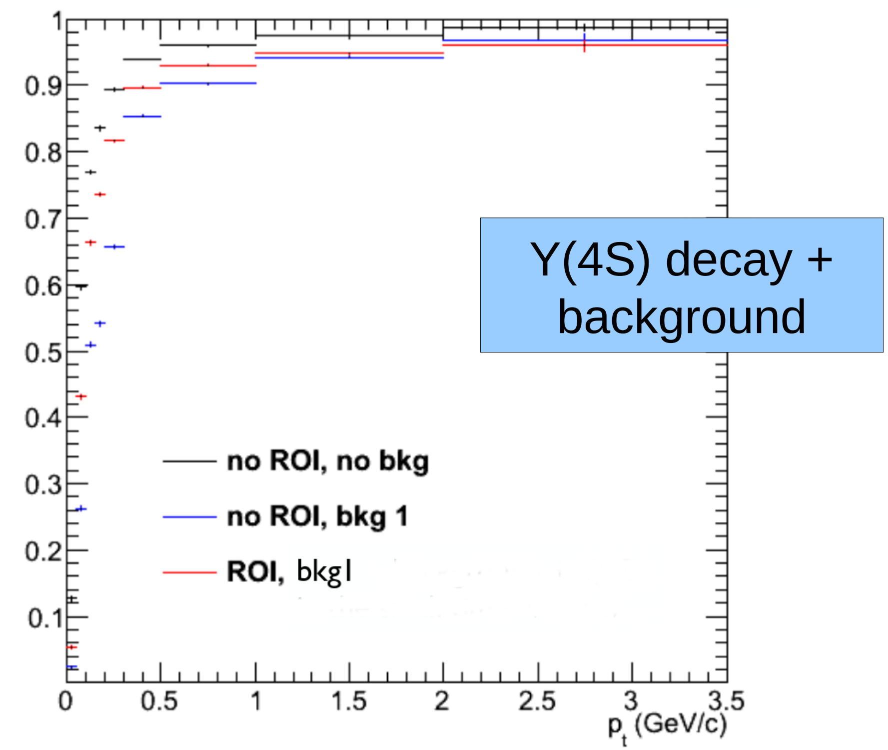
- Method has been tested on Monte Carlo events where the T0 time for simulation has been artificially shifted (resulting in inaccurate drift times)



- In~94% of the cases, this method is able to extract the correct T0 with an offset of less than 2ns for initial T0 times as large as +- 25ns
- This method will be used in an upcoming beamtest CDC cosmics beamtest
- Application of this method in the fully integrated Belle II will be discussed as soon as the precision of Level 1 and TOP trigger timing is available

Physics Performance of
Track Finding & Fitting





- Physics Efficiency: efficiency normalized to the initial generator particles
- Tracking Efficiency: factored out detector acceptance & inefficiencies
- With realistic background estimations:
  - As expected: decreased efficiencies
  - Region-of-Interest (ROI) selection on the PXD improves the performance significantly
- Studies are ongoing to further reduce the impact of background (for example SVD hit timing)
- We are looking forward to measured background samples from Belle II Phase 2 to test and optimize our background rejection techniques

Conclusion

- The Belle II Pattern Recognition Software supports and integrates two quite distinct detectors:
  - Silicon-based PXD & SVD
  - Central Drift Chamber
- Modular structure of the track finding and fitting algorithms allows for easy replacement and improvement of existing code
- Major components of the pattern recognition chain have been developed and validated in test beam and cosmics setups
- Current focus of tracking development is to further test and validate the existing algorithms with test beam and cosmics data

Thank you for your attention !