Summary: Software, Performance, and Computing

Conveners: Lea Reuter, Nikolai Hartmann, Eldar Ganiev

October 2, 2024, Belle II Germany meeting

Session Overview

09:00
- **PubDB, the new analysis database for Belle II** — *David Koch*
  Seminar room 4a, Building 1B — 09:00 - 09:20
- **b2luigi - bringing batch to luigi** — *Alexander Heidelbach*
  Seminar room 4a, Building 1B — 09:20 - 09:40
- **ECL clustering for high beam backgrounds** — *Torben Ferber*
  Seminar room 4a, Building 1B — 09:40 - 10:00

10:00
- **Improving cluster reconstruction in the ECL using GNNs** — *Jonas Eppelt*
  Seminar room 4a, Building 1B — 10:00 - 10:20
- **coffee break**
  DESY, Notkestrasse 85, 22607 Hamburg — 10:30 - 11:00

11:00
- **News of the Belle II Collaborative Services and Tools** — *Andreas Gellrich et al.*
  Seminar room 4a, Building 1B — 11:00 - 11:30
- **The DESY Analysis Facility** — *Christian Voss et al.*
  Seminar room 4a, Building 1B — 11:30 - 12:00

12:00
- **Visit to the DESY Computer Center** — *Thomas Hartmann et al.*
  Seminar room 4a, Building 1B — 12:00 - 12:20

14:00
- **PXD background generation using generative models** — *Fabio Novissimo*
  Seminar room 3, Building 1B — 14:00 - 14:20
- **Using machine learning to distinguish signal and background SVDSpacePoints** — *Christian Wessel*
  Seminar room 3, Building 1B — 14:20 - 14:40
- **CDC MVA-based background filters** — *Yulan Fan*
  Seminar room 3, Building 1B — 14:40 - 15:00

15:00
- **Graph Neural Networks based Tracking** — *Lea Reuter*
  Seminar room 3, Building 1B — 15:00 - 15:20
- **Track Momentum Scaling** — *Raimundo Hoppe Elsholz*
  Seminar room 3, Building 1B — 15:20 - 15:40

PubDB, the new analysis database for Belle II

David Koch

Motivation

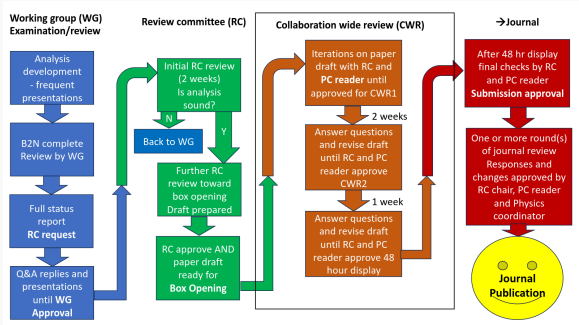
1. The software that runs Belle II's **document server** (docs.belle2.org) is outdated and needs replacement.
2. There is no streamlined way to follow the process of creating a paper.
   - there are well-defined policies regarding the process of creating a Belle II paper, ie. when and how to form a review committee, how the review process should take place, who is responsible for what, ...
   - no centrally organized mechanism which **implements the policy**: all e-mail based
   - no **database** that keeps track of the status of ongoing analyses

→ PubDB

PubDB, the new analysis database for Belle II

David Koch

Demo



The following is a series of screenshots that demonstrate how PubDB guides (fictional) users along this flowchart.

https://authors.belle2.org//deliver.pl?F=1

Timeline

October
- finish beta testing
- implement critical fixes / features
- prepare migration from Invenio to PubDB

November
- migrate
- test migration

December
- collaboration switches to PubDB
- 2.12. put Invenio in read-only mode

Timeline

- 28.2. turn off Invenio

The Analysis Code

Institute of Experimental Particle Physics

Without b2luigi

```
1_gamma                  feature_ana_paper
2_gamma                  losses.py
9_1_22_3.csv             models.py
9_1_22_4.csv             print_crystal_data.py
9_2_1_56.csv             __pycache__
9_2_1_57.csv             rebatch.py
dataset.py               run_1_barrel.py
feature_ana              run_2_barrel.py
feature_analysis.ipynb   run_2_full.py
```

```
run_2_nominal.py         utils_gnn.py
run_34_early.py          utils_train_inf.py
run_34_nominal.py        zaper_run_1_early.py
run_feature_ana_0b.py    zaper_run_1_nominal.py
run_feature_ana_0.py     zaper_run_feature_ana_early.py
run_feature_ana_1b.py    zaper_run_feature_ana_nominal.py
run_feature_ana_1.py
run_feature_ana_2.py
test.py
```

With b2luigi

- Workflow Management Systems…
  - … enable the user to structure the workflow by defining dependencies
  - … build the dependency graph and schedule the execution

b2luigi - bringing batch to luigi

Alexander Heidelbach

luigi in a Nutshell

- **Building a pipeline**
  - Dependency Resolution
  - Workflow Management
  - Visualisation
  - Handling Failures
  - Command Line Integration
  - …
- https://github.com/spotify/luigi

Spotify®

"Hello World" in luigi:

```
class MyTask(luigi.Task):
    parameter = luigi.Parameter()

    def run(self):
        do_smth(self.parameter)

    def output(self):
        return Target("some/file")

    def requires(self):
        yield OtherTask()
```

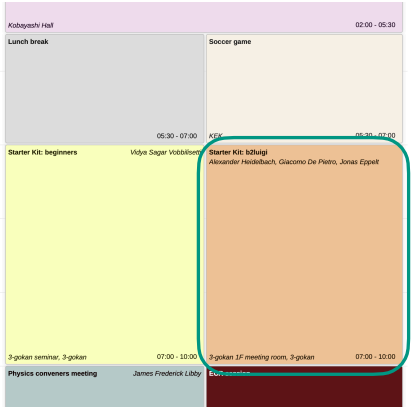
Example: HTCondor

```python
class MyTask(b2luigi.Task):
    parameter = b2luigi.IntParameter()
    batch_system = "htcondor"

    @property
    def executable(self):
        return ["$MY_PYTHON"]

    def output(self):
        yield self.add_to_output("test.txt")

    def run(self):
        with open(self.get_output_file_name("test.txt"), "w") as f:
            f.write(f"Test {self.parameter}")

class Wrapper(b2luigi.WrapperTask):
    def requires(self):
        for i in range(10):
            yield MyTask(parameter=i)

if __name__ == "__main__":
    b2luigi.set_setting("env_script", "setup.sh")

    b2luigi.set_setting("result_dir", "results")

    b2luigi.process(Wrapper(), batch=True, workers=100)
```

- Script executed on the batch job side
- Location needs to be accessible on the batch job side
- Wrapper tasks need no output
- Definition of the batch system
- Executable for this specific task
  - E.g. environment variable set in setup.sh

Advertisement

- **b2luigi workshop at the next B2GM**
  - Friday 11th October at 3pm
  - 3-gokan 1F meeting room
  - **In person only!**
- **Topics**:
  - (b2)luigi basics
  - basf2 analysis from simulation to ntuple
  - Sending b2luigi tasks to batch systems
  - Offline analysis hints

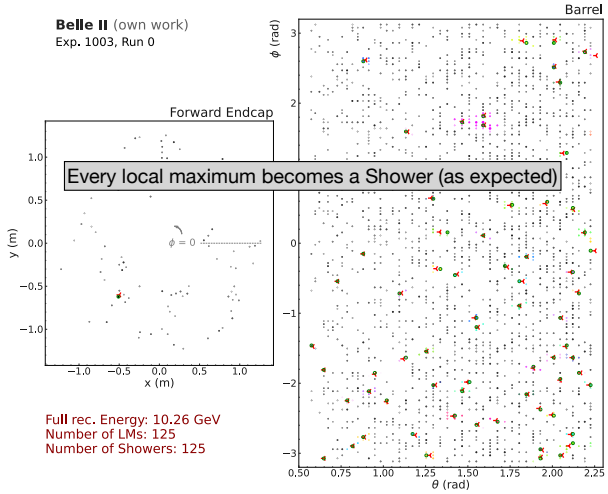


ECL clustering for high beam backgrounds

Torben Ferber

ECL clustering event display exp1003 backgrounds

plot by Isabel Haide

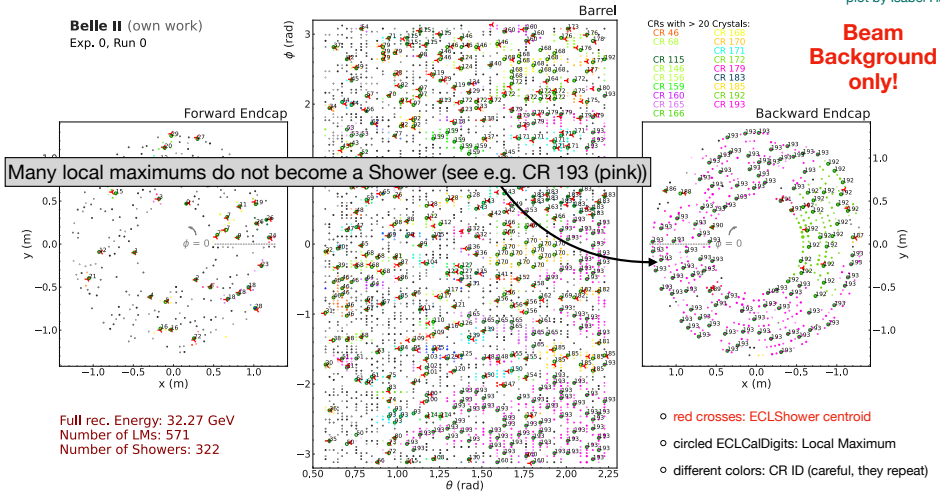


ECL clustering for high beam backgrounds

Torben Ferber

ECL clustering event display exp0 backgrounds

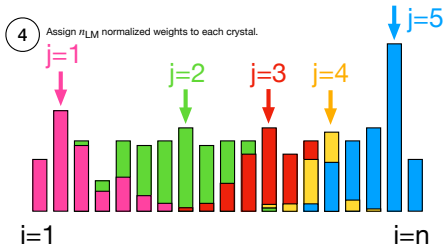
plot by Isabel Haide



ECL clustering for high beam backgrounds

Torben Ferber

ECL clustering: Splitter N1 for $n_{\mathbf{LM}} > 1$



These conditions remove a lot of potential shower candidates if there are too many energy depositions of similar magnitude because the reweighting becomes unstable.

The solution for this to limit the size of the CR (see Jonas talk for state-of-the art LM classification).

Repeat the weighting procedure from step 3, but use the updated centroid positions and ECLCalDigit weights.

1) Calculate the mean centroid shift of all LMs. If it is below 1mm, exit the procedure.

2) Calculate the weighted energy per crystal in the shower candidates. If this value is maximal in an **ECLCalDigit that is not identical to the LM, mark this shower candidate for deletion**.

3) If the highest weighted energy is **below a threshold $E_{\mathbf{min}}$, mark this shower candidate for deletion**.

4) If any condition 2) or 3) is fulfilled, remove those shower candidates*.

*This can mean that we remove all shower candidates and leave an empty CR, which can be seen in the event display on slide 7 in CR193.

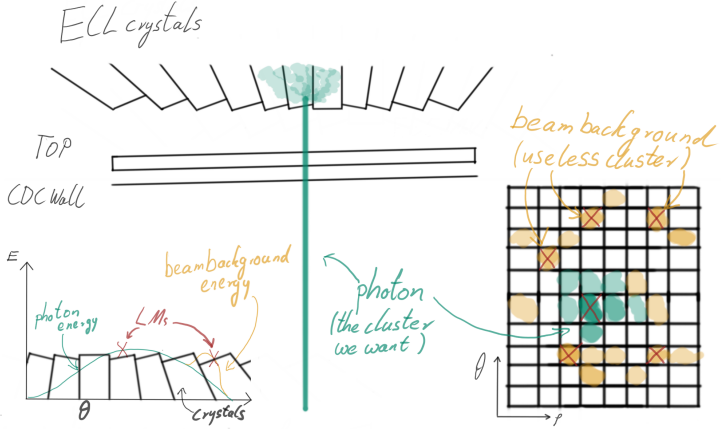
Conclusion

- The background conditions, especially close to injection, are extreme, **analysts need to filter events** that have a large number of out-of-time crystals **if they are sensitive to inefficiencies** (we do this in the single photon analysis for example)
- Post-reconstruction MVAs do not help for this challenge
- ECL will implement the following changes to reconstruction:
  - **Increase seed and growth thresholds to 20 MeV and apply a timing cut of 200 ns to seed crystals (via module parameters, no code changes needed)**
  - future: limit the weight distance of crystals possibly belonging to a LM to 5 CsI Moliere radii (~20cm)
  - future: allow LM shifts by one crystal (relax condition 3 on page 11)
  - medium future: LM classifier (Jonas), GNN reconstruction (F. Wemmer et al., "Photon Reconstruction in the Belle II Calorimeter Using Graph Neural Networks", https://link.springer.com/article/10.1007/s41781-023-00105-w)

Improving cluster reconstruction in the ECL using GNNs

Jonas Eppelt

Challenges from beam background

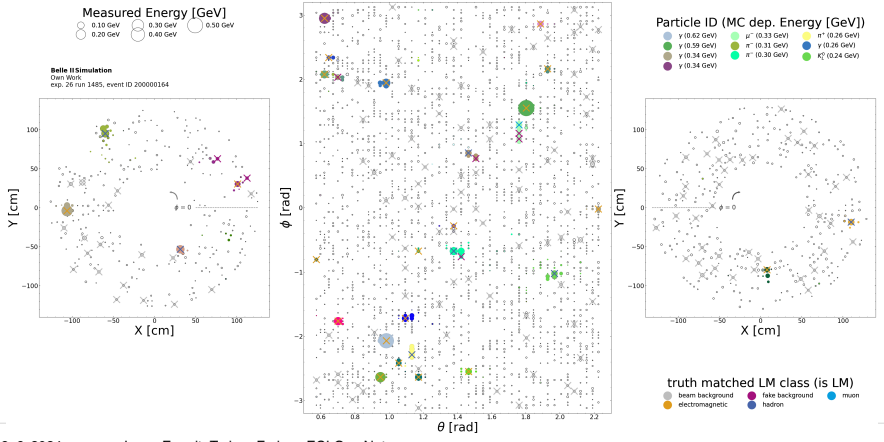


Improving cluster reconstruction in the ECL using GNNs

Jonas Eppelt

Karlsruher Institut für Technologie

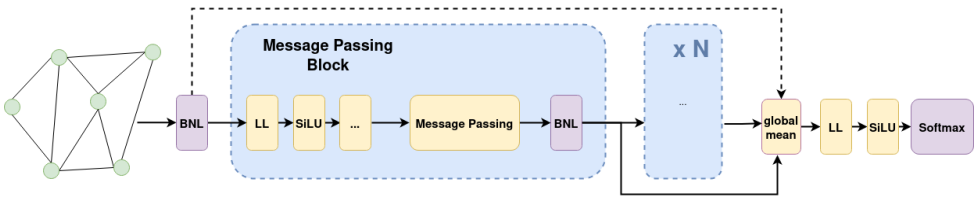
Examplary B-event with target LMs



Improving cluster reconstruction in the ECL using GNNs

Jonas Eppelt

The Seedclassifier model

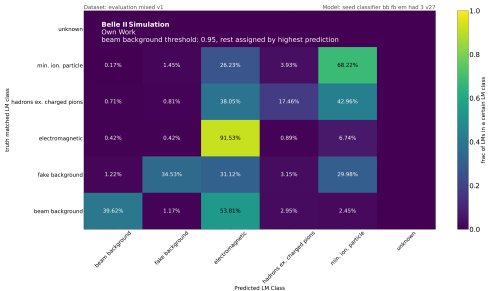


- Message passing blocks collect information about the neighborhood.
- A Focal Loss (weighted cross-entropy) accounts for a large imbalance towards background samples.
- The training is done on Particle Gun events enriched with low energetic particles.

Improving cluster reconstruction in the ECL using GNNs

Jonas Eppelt

Summary and outlook



Summary

- Using GNNs, we can reduce the number of beam background LMs.
- Further, we can identify electromagnetic LMs.

Outlook

- Seperating hadrons and muons is more challenging. I plan to use different training targets: *minimal ionizing particle* and *other hadrons*.
- Using these labels, I plan to build clusters from them.

News of the Belle II Collaborative Services and Tools

Andreas Gellrich

Collaborative Services and Tools (B2CS)

Infrastructure

- DESY IT Infrastructure
  - User registry (**accounts**) + **MFA**
  - *ZMS* (**web** pages) (www.belle2.org)
  - XWIKI (**wiki**)
  - *Sympa* (**mailing lists**)
  - *Indico* (**agenda**)
  - *GitLab* (**repo**, **issues**)

- XEN-infrastructure for Belle II
  - 4 x [HV (20 cores), SSD (447GiB), SAS (237GiB)]; NetApp (800GiB) (**recently renewed for 36k EUR**)
  - ~35 VMs; systems (OS, cert, storage) provided and maintained by DESY (AG)
  - authors, b2mms, calibration, chat, docs, evdisp, dqm, elog, mirabelle, pxd, rundb, questions, shift, software …
  - Buildbot (EL7/8/9, Ubuntu 20/22/(24)); DB squids

Collaborative Services and Tools (B2CS) cont'd

The Belle II wiki – a migration story

- In 2015 Belle II decided to install collaborative services from scratch
- During 2016 the TWIKI contents was migrated to *confluence* in a (*painful*) semi-automatic procedure

- Confluence suffered from a number of security incidents with (*painful*) service interruptions
- ATLASSIAN finally changed the support model from on-premise to cloud with enormous costs for DESY

- As a successor XWIKI was identified in 2022
- **ATLASSIAN** tools reached end-of-life (**EOL**) in Feb 2024
- The XWIKI company helped DESY IT to build and adjust a migration tools

- **Finally the Belle II confluence space (biggest in DESY) was migrated in Sep 2024**

MFA @ DESY

Will be mandatory for all services at some point

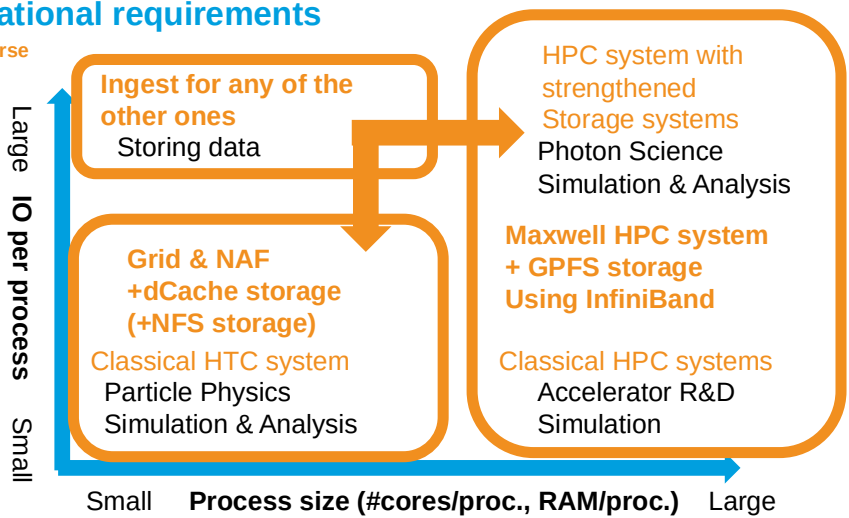
- **779** out of 1211 active members **have** a MFA token
- **430** still **without** a token
  - can not use **X-Wiki** & other services with MFA
  - can not create a SSH-tunnel
  - can not use self-service password reset
  - **please spread the word**

The DESY Analysis Facility

Yves Kemp

Computational requirements

Very very coarse

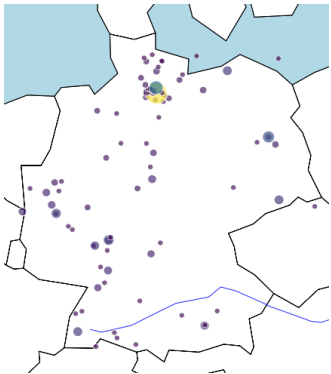


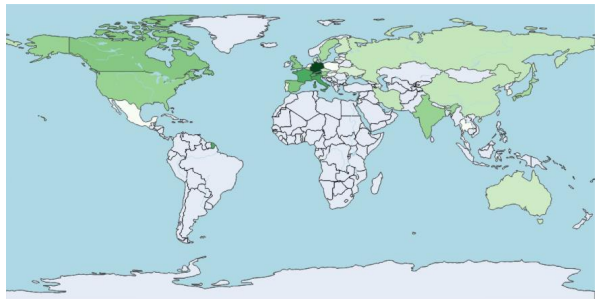
The DESY Analysis Facility

Yves Kemp

… and where they come from

logins during two weeks in October 2023





Only NAF & Maxwell logins are accounted for (no Grid submission)

… mostly from academia (universities and institues)

… some commercial users

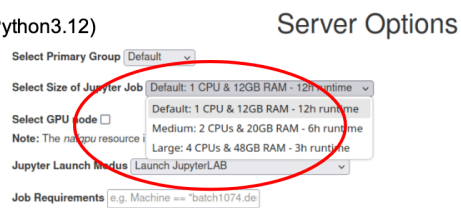
The DESY Analysis Facility

Yves Kemp

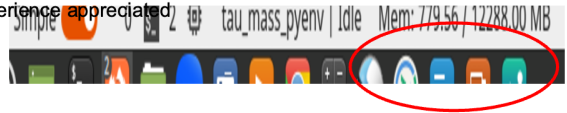
JUPYTER Notebooks

Next generation

- JHUB & notebooks upgraded (JupyterHub version 5.0.0, Python3.12)
- New notebook classes:
  - Default: 1 CPU / 12 GB RAM / 12h runtime
  - Medium: 2 CPUs / 20 GB RAM / 6h runtime
  - Large: 4 CPUs / 48 GB RAM / 3h runtime



- Default notebooks run on all pool nodes in reserved slots (similar setup to old pool)
- Medium & large notebooks run on 2 dedicated servers
- Feedback about new sizing and user experience appreciated
- RAM taxometer now in place



- Suggestion: Have a ‘show-us-your-notebook’ session later this year in order to connect notebook users over VO/batchsystem borders and discuss further experiences and needs

The DESY Analysis Facility

Yves Kemp

NAF intro school

FH Sustainable Computing Workshop

Oct 7 – 8, 2024
Europe/Berlin timezone

Enter your search term

- Overview
- Timetable
- Contribution List
- My Conference
  - My Contributions
- Registration
- Participant List

Contact:

- eleanor.jones@desy.de
- juliette.alimena@desy.de
- ben.brueers@desy.de
- nils.gillwald@desy.de

The FH Sustainability Forum and FH IT experts present a 2-day workshop on Computing. In this workshop, participants will learn to write efficient code, use batch computing in a sustainable way, and more. Best practices will be taught by experts. This workshop is targeted towards students and postdocs, but open to anyone who feels like they would benefit from refreshing their memory or learning current best practices. Participants will learn how to develop more efficient code and how to maximize resources and minimize waste, with hands-on examples. These examples will be demonstrated with local computing clusters and are tailored to the needs of the FH division.

The fourth installment of this workshop will occur on October 7 and 8, 2024, in person on the Hamburg campus (with some zoom available). This fourth installment is targeted towards beginners. We plan to offer this workshop at regular intervals in the future, so that incoming students and postdocs can profit. The next workshop will target more advanced users.

Coffee breaks will be provided free of charge. Lunch is not included.

Zoom connection:
https://cern.zoom.us/j/62785538071?pwd=dXJFZndHK0g4eVNaelArSjVjaUFQZz09

Mattermost channel: https://chat.desy.de/desy/channels/fh-sustainable-computing-workshop

Live notes for last minute updates:

Gitlab repository for the workshop: https://gitlab.desy.de/fh-sustainability-forum/sustainable-coding-tutorial

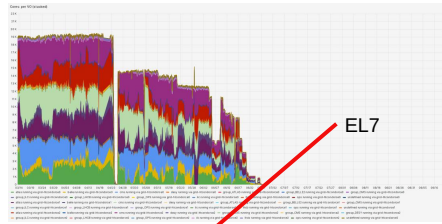
https://indico.desy.de/event/43906/

Grid @ DESY-HH

- HTCondor Cluster

  HTC Clusters at DESY-HH

  - ~25k Cores, ~402 kHS23
    (mix HS06 mapped and HS23 natively measured)
  - Started now to cull older, inefficient nodes
  - Recycled old nodes into dedicated preprod cluster
- Rolling migration to Condor23 & CondorCE6 on RHEL9
  - Reinstalled last remaining set of EL7 nodes with EOL
  - EL9 migration worked reasonably
    - Some minor remaining elements still have to migrated (py2.7, cgroups v1,... monitoring scripts etc.) :-/
- Getting non-token ready groups onto the cluster was/is a PITA
  - SSL authz working, but expertise on the user/group site limited



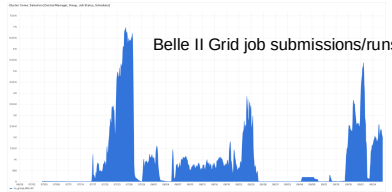
The DESY Analysis Facility

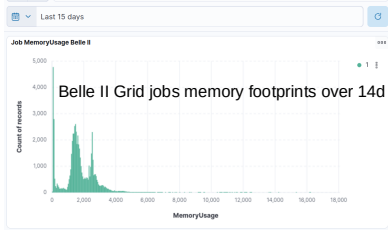
Thomas Hartmann

Status Belle II

Belle II Grid Production

- CondorCE authz via legacy SSL user/robot DN
  - Status token workflows on factory side?
- Utilization of grid cores below nominal share
  - Local or upstream issue?
  - Steady queue pressure desirable
- Memory footprints seem to leak beyond 2GB/core
  - Plans for mcores or more dynamic core & mem scheduling?
  - VOCard still valid?



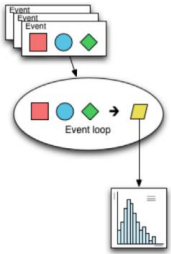


Columnar Analysis

Illustrations by Nick Smith

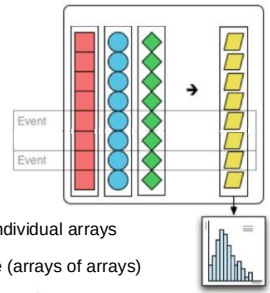
A Different Paradigm for Data Analysis – But is it really?

Event Loop



- Typically Object based data ➔ Particle with associated track
- Read data from event into local variables
- Calculate and store derived variables
- Rinse and repeat for all entries
- Explicit outer loop

Columnar Structure



- Load data into individual arrays
- Nested structure (arrays of arrays)
- Perform calculation on all elements in array
- Implicit inner loop
- Data structure in TTree and Pandas dataframe
- **Advantageous with modern CPU architectures**

Columnar Analysis Tool-Kits

Scale-out Option Similar to Pandas

- Single Node Jupyter Notebooks are limited in resources and Python by default not clever at data import
- Ease of use of Pandas et al. further escalate problem
- Investigate scale-out tools, most commonly used: Apache Spark and Python Dask

Apache Spark

- Written in Java; industry tool
- Different syntax than Pandas
- Not Python bound → easier to deploy/maintain
- Extensive data-caching in memory possible
- Easy to plug external data sources and data types (e.g. PostgreSQL import, ROOT data-types)
- Limited support for HEP data formats

Python Dask

- Pure pythonic; developed in HPC community
- Very similar syntax than Pandas
- Pure Python → tedious to deploy/maintain (versioning)
- Extensive data-caching in memory possible
- Support for HEP data formats (through uproot/awkward arrays)
- Supports only numpy compatible data-types

- Community tool-boxes build around these tools → Coffea/Casa
- ROOT recent addition RDataframe can be connected to Dask/Spark setups (CERN Swan)

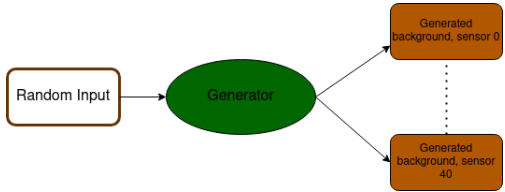
Generating pixels with GAN

Previous approach:

- GAN conditioned on sensor number with a transformer-based relational reasoning module to reproduce the correlations between sensors(IEA-GAN).

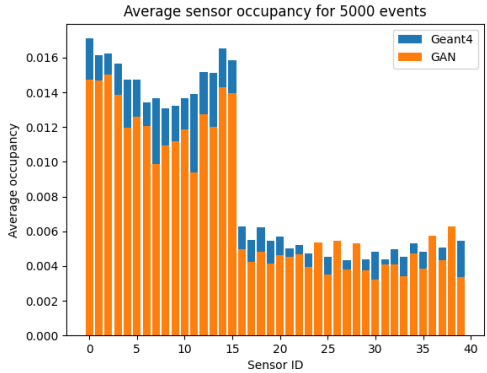
New approach: generate the background using a GAN without conditioning on the sensor number.

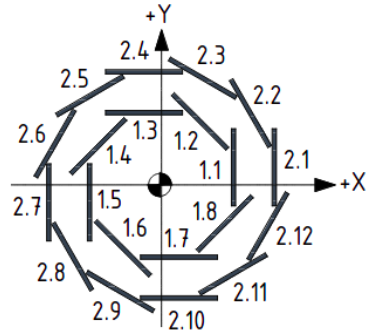
- Generate instances of background for all sensors at once.
- Wasserstein GAN with CNN layers used in the Generator and Discriminator.



Evaluation - Occupancy per sensor

The model seems to reproduce quite well the sensor occupancy, aside from some minor details probably due to some fluctuations in the weights of the model.

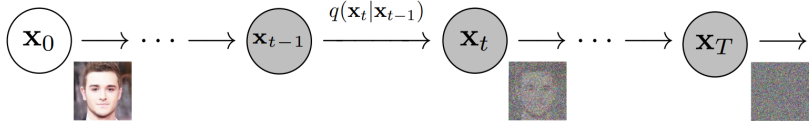




PXD background generation using generative models

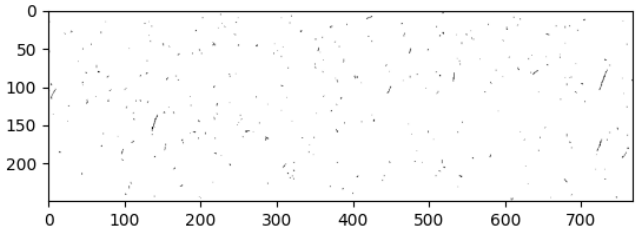
Fabio Novissimo

Diffusion model: forward process



$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbb{I}) \qquad \{\beta_t \in (0,1)\}_{t=1}^{T}$$

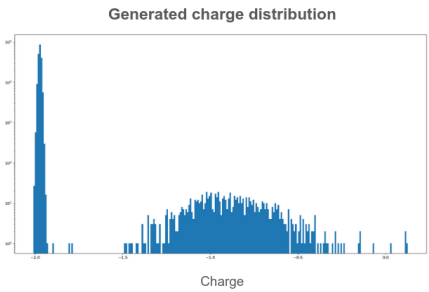
Preliminary results: background

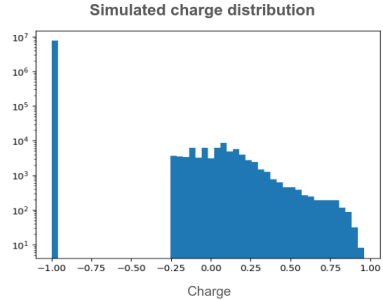


PXD background generation using generative models

Fabio Novissimo

Preliminary results: charge distribution





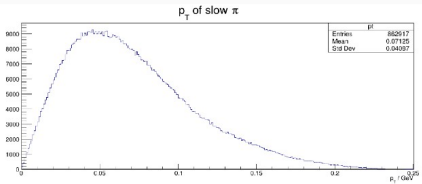
Using machine learning to distinguish signal and background SVDSpacePoints

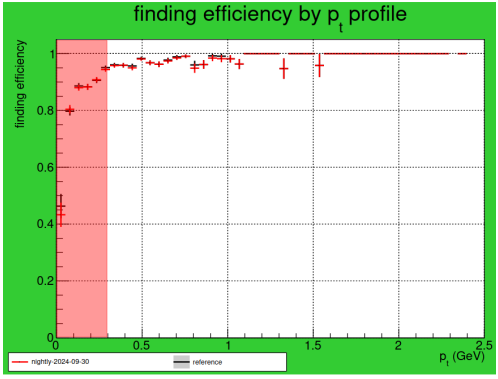
Christian Wessel

The issue

Why?

- We lose a lot of tracks in the (very) low p_T region
- Many analyses suffer from this, for instance and in particular B –> D^{*+} [–> D^0 π_s^+]X decays, where the D^{*+} can't be reconstructed if the resulting slow pion is missing
- But not limited to these processes
- Can we improve low p_T tracking?
- I think we can!





Using machine learning to distinguish signal and background SVDSpacePoints

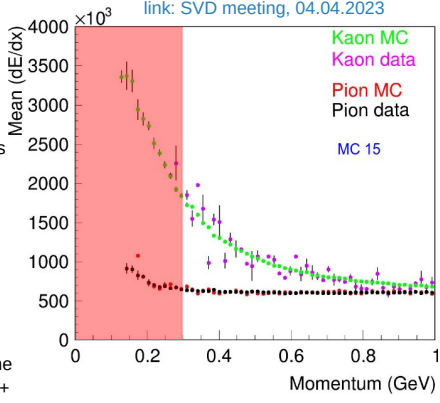
Christian Wessel

The idea

How?

- Slow pions, but also other low p_T particles, should have higher than average energy loss in SVD
- Can we use these information to identify slow pion SVDClusters / SVDSpacePoints?
- Train MVA (FBDT) with SpacePoints in B –> D^{*+} X events
- Only ~4 SpacePoints per event are from slow pion, but O(900) SpacePoints are from background particles (nominal BG, excluding SpacePoints from other signal particles)
  - Train on all slow pion SpacePoints
  - Only use a small fraction of background (0.7%) SpacePoints to obtain a balanced training data set
- Information used from clusters (u + v): (seed) charge, time + time error, size, signal-to-noise ratio, local coordinates + errors, layer, ladder, sensor

link: SVD meeting, 04.04.2023



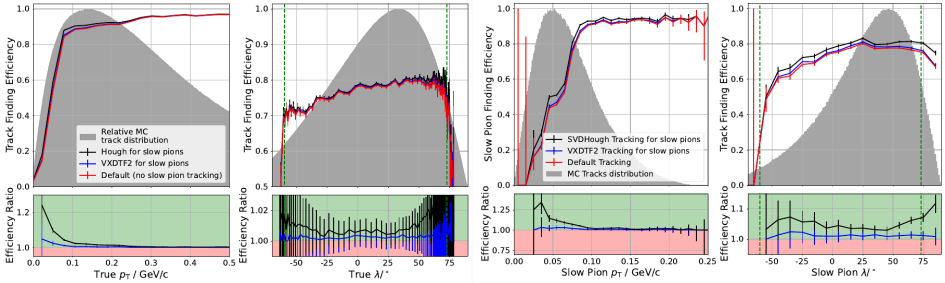
Using machine learning to distinguish signal and background SVDSpacePoints

Christian Wessel

Slow Pion tracking

Find slow pions in SVD and PXD

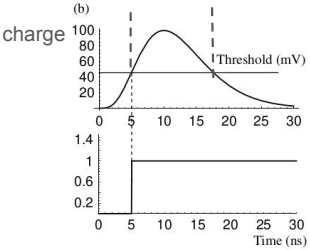
- Proven to work for SVD-only tracking using VXDTF2 and HoughTracking
- For now full Hough Space used, will extend this to avoid high pt region –> focus on low p_T particles only
- Significant improvement in general (up to +0.9% on average, left) and slow pion (up to +3.7% on average, right) track finding efficiency
  - Below p_T of ~70 MeV still significant room for improvement –> need SVD+PXD tracking



CDC MVA-based background filters

Yulan Fan

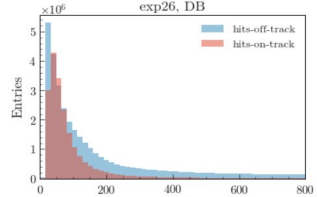
Motivation



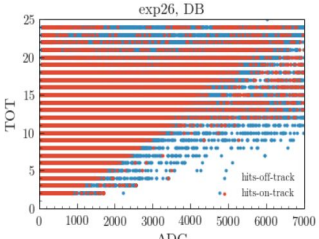
$$\mathrm{ADC} = \int \frac{dc}{dt} dt$$

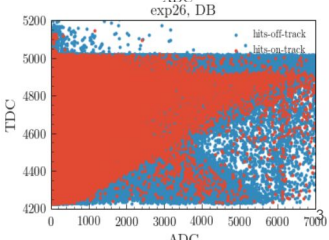
Time over threshold (TOT)

Database(DB) as reference:
ADC>=15, TOT>=2, ADC/TOT>=3 (Inner super layers)
ADC>=18, TOT>=2, ADC/TOT>=3 (Outer super layers)



Sizable amount of hits-off-tracks & different correlation patterns for hits on tracks and hits off tracks for ADC, TOT and TDC, MVA can be used to distinguish them





CDC MVA-based background filters

Yulan Fan

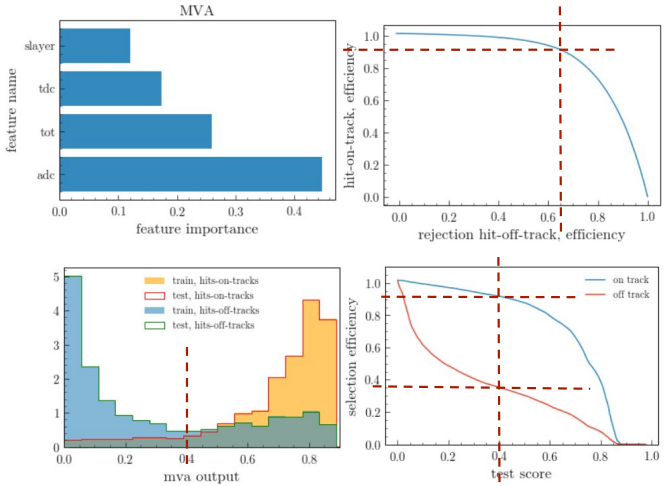
MVA: training preparation and performance

All:
ADC>=1, TOT>=1, ADC/TOT>=0
(payload generated in wireHitRequirement)

Training Variables:
ADC, TOT, TDC and inner/outer super layer

Hyper-parameter:
optimized with Optuna

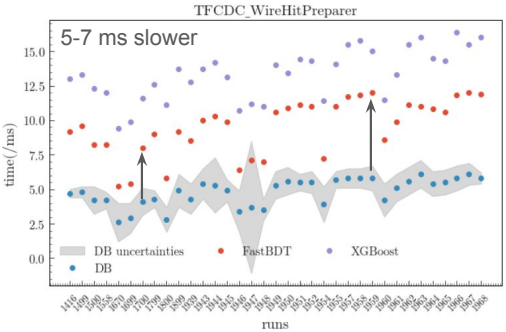
Training sample:
1.2 million exp26 raw data

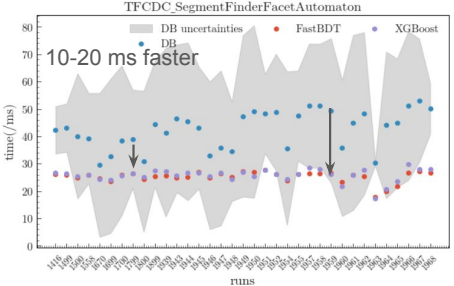


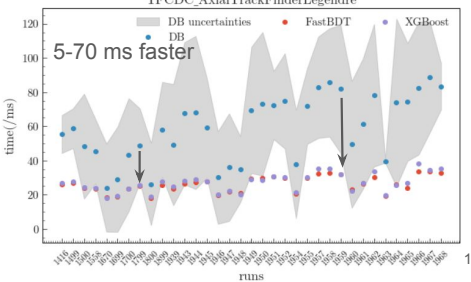
CDC MVA-based background filters

Yulan Fan

Time consumption







The main CDC tracking algorithms becomes almost twice faster !

CDC MVA-based background filters

Yulan Fan

Standard validation: rel8 nominal background, cdc standalone

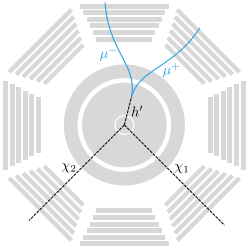
CDC standalone (%)	DB	hadron MVA	dimu MVA	MC0 MVA
finding_charge_efficiency	**69.1**	**78.5**	**79.6**	**80.8**
finding_efficiency	**71.3**	**80.9**	**81.7**	**83.2**
charge_efficiency	**96.9**	**97.1**	**97.5**	**97.1**
charge_asymmetry	**2.31**	**1.60**	**0.68**	**0.84**
fake_rate	**3.39**	**3.52**	**2.88**	**2.93**
clone_rate	**0.65**	**0.64**	**0.58**	**0.53**
hit_efficiency	**80.6**	**82.4**	**81.0**	**85.7**

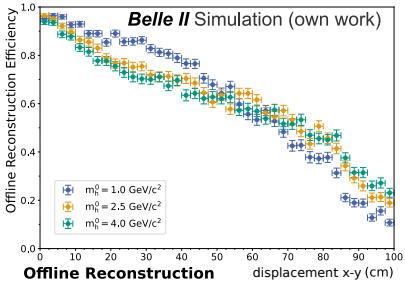
Graph Neural Networks based Tracking

Lea Reuter

Motivation - Displaced Vertices

Displaced vertices important signature in searches for new physics (for example [1] [2] and arXiv:2012.08595, arXiv:202.03452, arXiv:1911.03176)





Efficiency decreases depending on displacement (K_S^0, Λ^0, Dark Sector searches)

[1] Search for a long-lived spin-0 mediator in $b \to s$ transitions at the Belle II experiment (arXiv:2306.02830)

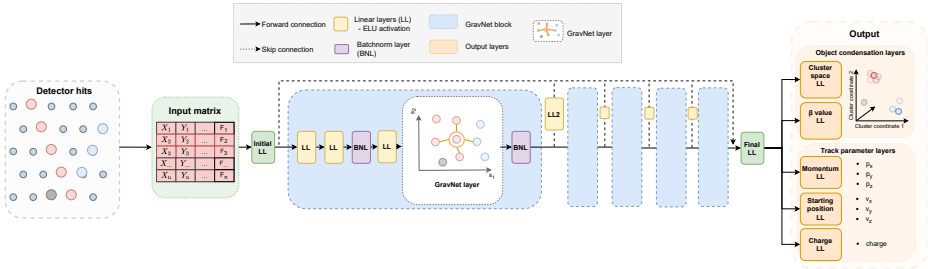
[2] Search for Inelastic Dark Matter produced in association with a Dark Higgs (BELLE2-NOTE-PH-2024-019)

Graph Neural Networks based Tracking

Lea Reuter

Approach

- Find track parameters: momentum, starting position and charge
- Computing resource and time constraint → Graph Neural Networks
- Find unknown number of tracks → Object Condensation (arXiv:2002.03605)



Inputs:

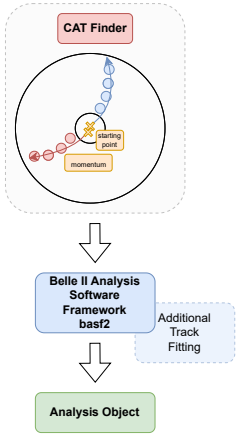
- x and y wire position
- tdc and adc signal information
- layer, superlayer, and clayer

Adjustable Parameters result in a total of 797,812 trainable parameters (3MB weightfiles)

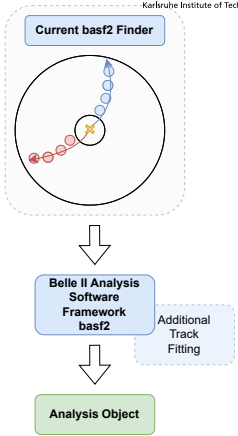
Graph Neural Networks based Tracking

Lea Reuter

New Track Finder: CDC AI Track (CAT) Finder



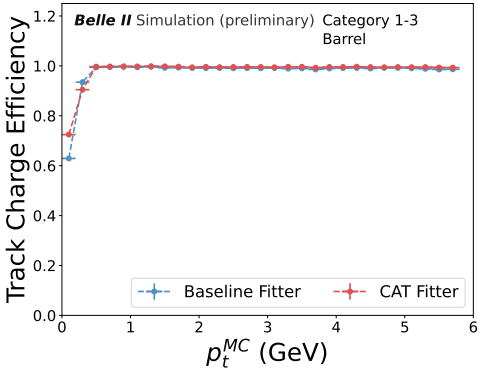
Fair comparison for track finding algorithms on analysis level for CDC only



Graph Neural Networks based Tracking

Lea Reuter

Performance on prompt particles for CDC only



1-12 μ^+ and μ^- per event,
run dependent beam-backgrounds from experiment 26,
barrel acceptance ($35.4° < \theta < 123°$)

integrated	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$
Baseline	$94.15^{+0.06}_{-0.06}$	$1.83^{+0.04}_{-0.04}$	$0.9^{+0.03}_{-0.03}$
CAT	$95.11^{+0.06}_{-0.06}$	$1.25^{+0.03}_{-0.03}$	$0.34^{+0.02}_{-0.02}$

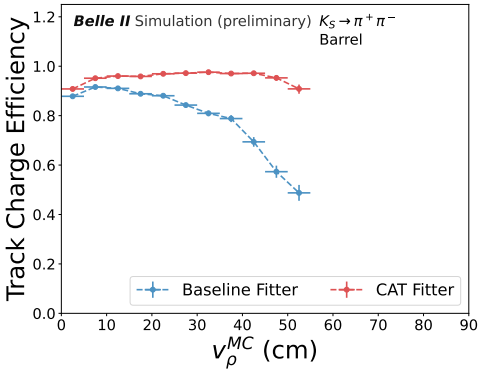
$$\varepsilon_{\text{track}} = \frac{n_{\text{tracks}}(\text{matched to particle and charge})}{n_{\text{simulated}}(\geq 1\text{matched hit})}$$

Extensive studies shown in paper draft (BELLE2-NOTE-TE-2024-008)

Graph Neural Networks based Tracking

Lea Reuter

Performance on displaced particles for CDC only



Particle gun $K_s^0 \to \pi^+\pi^-$, $p_t(K_s^0) = [0.05, 3]$GeV
run dependent beam-backgrounds from experiment 26,
barrel acceptance

integrated	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$
Baseline	$88.1^{+0.1}_{-0.1}$	$5.39^{+0.07}_{-0.07}$	$0.59^{+0.02}_{-0.03}$
CAT	$92.99^{+0.08}_{-0.08}$	$5.13^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$

$$\varepsilon_{\text{track}} = \frac{n_{\text{tracks}}(\text{matched to particle and charge})}{n_{\text{simulated}}(\geq 1\text{matched hit})}$$

Extensive studies shown in paper draft (BELLE2-NOTE-TE-2024-008)

Graph Neural Networks based Tracking

Lea Reuter

Current Status: Implementation in Full Reconstruction

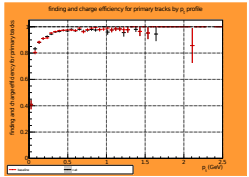
Karlsruhe Institute of Technology

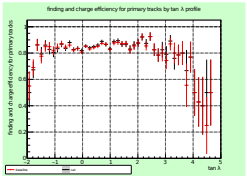
b2validation, `release-08-01-08`, exp. 1003, Full Tracking Validation **including VXD**

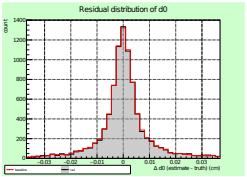
`add_reconstruction(path, useCAT=True/False)`

Extrapolation to the inner tracking detectors work fine, compatible results

CAT Finder vs **Baseline**







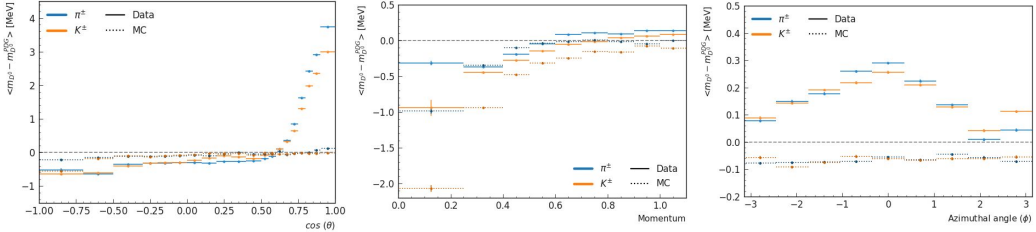
Tag	finding charge efficiency	charge efficiency	charge asymmetry	fake rate	clone rate
Baseline	92.94	98.97	1.47	3.16	3.48
CAT	92.95	99.65	1.30	3.76	5.21

Track Momentum Scaling

Raimundo Hoppe Elsholz

Summary of the problem

- A disparity between the reconstructed mass of particles and the PDG nominal values was found.
  - Polar angle (θ) — Corrected
    - ➡ Attributed to a misestimation of the detector magnetic field
      - ○ Sample dependent (MC / Data)
  - Momentum — Corrected
    - ➡ Attributed to an insufficient energy loss correction due to detector material
      - ○ Sample and Particle type dependent
  - Azimuthal angle (φ) — Not Corrected
    - ➡ Reason still unknown
      - ○ Sample, Charge and decay topology dependent

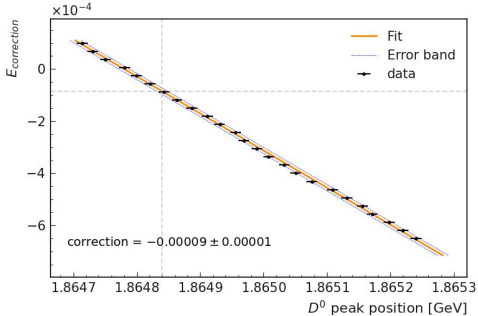


Track Momentum Scaling

Raimundo Hoppe Elsholz

Estimating the Corrections

- Correction payloads were developed for the cos(θ) and momentum dependent shifts.
- The corrections for both were estimated in the same manner using $D^0 \rightarrow K^{\pm}\pi^{\mp}$ events.
- Correction estimation:
  1. The data is separated into bins [cos(θ) or p].
  2. A list of 'N' guess corrections is generated around the "identity correction".
  3. In each bin the data is replicated 'N' times and each guess correction is applied to one of this copies.
  4. The mass peak position resulting from the fit in each replica is plotted against the guess correction that was used (scan plot on the right).
  5. A 1D polynomial is fitted to the resulting scan plot to extract the correction value corresponding to the PDG D^0 mass for each bin.
- Due to their nature, the corrections are performed differently:



The cos(θ) corrections:

$$P_{new} = P_{old} \bullet corr$$

Energy Loss correction:

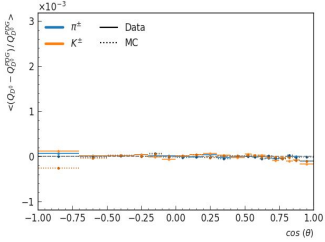
$$E_{new} = E_{old} + corr$$

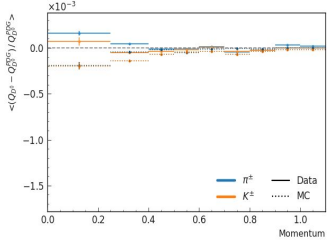
Track Momentum Scaling

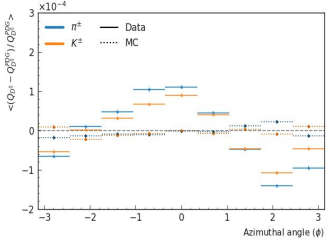
Raimundo Hoppe Elsholz

Summary

- Correction Payloads are available. Currently waiting for approval on the Note.
- Some small mass peak shifts still remain in low momentum MC events.
- Azimuthal angle shifts are still present and their cause is unknown.
- The corrections have been validated for Release 8 using chunk 1. No problems were found (Slide on backup)





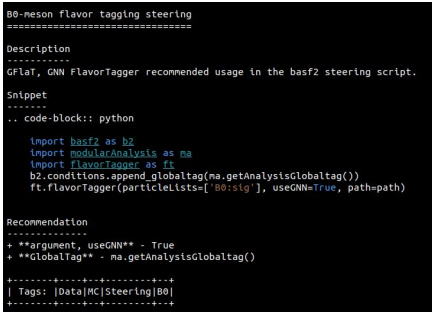


slide from Sviatoslav Bilokin

Links: Offline Reweighting, Correction table repository

New analysis tools

- `b2help-recommendation`
  - Provides up-to-date information about analysis from command line
  - Includes code snippets
  - Sorted by MC campaigns: MC15, MC16
- Offline reweighting or SysVar
  - Tool to add weights to ntuples
  - Supports PID and FEI weight tables
  - New version will be released in winter light releases
- Correction table repository
  - Contains recommended weight tables that can be applied offline
  - Sorted by campaign and MCrd or MCri
  - Please contact conveners to add your weight tables that can be of general interest!



Thanks to all speakers for the great talks and all participants for the fruitful discussions!