Software & Computing

Track Summary

Thomas Lück (LMU), Michel Villanueva (DESY)

Belle II Germany Meeting
Sep 19 - 21, 2022

Session Overview

Thank you to all speakers!!

Basf2 and related topics	*Giacomo De Pietro*
B132, LMU	09:20 - 09:40

Validation of basf2 releases	*Patrick Ecker*
B132, LMU	09:40 - 10:00

Introduction to RDataFrames	*David Koch*
B132, LMU	10:00 - 10:20

Coffee break	
B132, LMU	10:20 - 11:00

Tracking overview	*Christian Wessel*
B132, LMU	11:00 - 11:20

GNN-based Track - and Vertex - finding	*Lea Reuter*
B132, LMU	11:20 - 11:40

Improving ECL Clustering on Trigger Level with Object Condensation	*Isabel Haide*
B132, LMU	11:40 - 12:00

Clustering Energy Depositions in the ECL using graph neural networks	*Florian Wemmer*
B132, LMU	12:00 - 12:20

Representation of decay relations in hyperbolic space	*Boyang Yu*
B132, LMU	14:00 - 14:20

Systematics framework	*Sviatoslav Bilokin*
B132, LMU	14:20 - 14:40

Workflow management systems for Belle II	*Caspar Schmitt*
B132, LMU	14:40 - 15:00

Distributed computing at Belle II	*Michel Hernandez Villanueva*
B132, LMU	15:00 - 15:20

Coffee break	
B132, LMU	15:20 - 16:00

Computing development projects at KIT	*Matthias Schnepf et al.*
B132, LMU	16:00 - 16:20

NAF at DESY	*Thomas Hartmann*
B132, LMU	16:20 - 16:40

Storage infrastructure at DESY	*Christian Voss*
B132, LMU	16:40 - 17:00

Belle II

The software group & basf2 & other stuffs   Giacomo De Pietro

Developing, validating and deploying basf2

- **release-07** tagged in July:
  - features:
    - new externals
    - new track refinement step to reduce charge misreconstruction
    - improved TOP reconstruction
    - improved ECL likelihoods
  - release-07 validation still ongoing (got delayed)

- **release-08:**
  - first post LS1 basf2 release
  - wishlist:
    - upgrade Geant4 version
    - reduce memory usage and mdst size
    - get rid of deprecated ROOT classes
    - improve analysis toolkits (e.g. FEI; flavor tagger)
    - further improve charge reconstruction

Collaborative tools
- during LS1 migration from Atlassian to GitLab: no more Stash, Bamboo, Jira
- Confluence will stay (for a while …)
- migration of Jira issues and repositories automized
- transition from Bamboo to GitLab is ongoing

Belle II

The software group & basf2 & other stuffs

Giacomo De Pietro

Developing, valid...

- **release**
  - feat...
    - ...
    - ...
    - improved TOP reconstruction
    - ...
    - ...oing

- **release-08:**
  - first post LS1 basf2 release
  - wishlist:
    - ...ion
    - ...e and
    - ...ROOT
    - ...kits (e.g.
    - ...e

- ...
  - ...n to GitLab: no
  - ...)
  - ...ories automized
  - ...s ongoing

Belle II

4

Analysis Validation in Belle II

Patrick Ecker et al.

Old Approach

- over 59 modes, involving 10 WG liaisons
- slow and work intensive

Current Status of Analysis Validation

Produced centrally from the list of modes provided by WGs

Checked by WG liasons

Before: done locally by WG liaisons
Now: run centrally

Display results

mDST (validation MC modes)

Valid. steering 1 → Val histo 1 → Valid. fit 1
Valid. steering 2 → Val histo 2 → Valid. fit 2
...
Valid. steering n → Val histo 1 → Valid. fit n
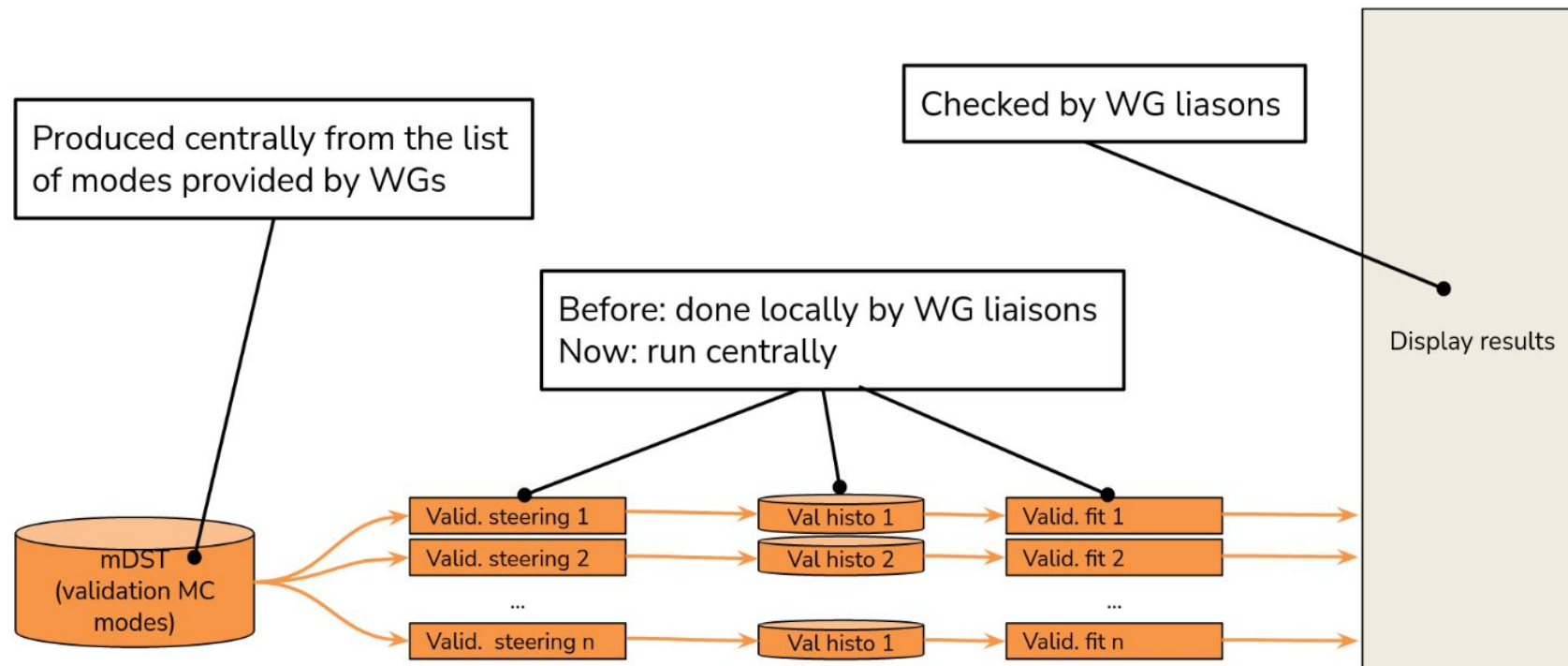
Analysis Validation in Belle II

Patrick Ecker et al.

Planned new Approach

- produce only 6 validation modes
- more done centrally
- modular framework: easy to add new validation
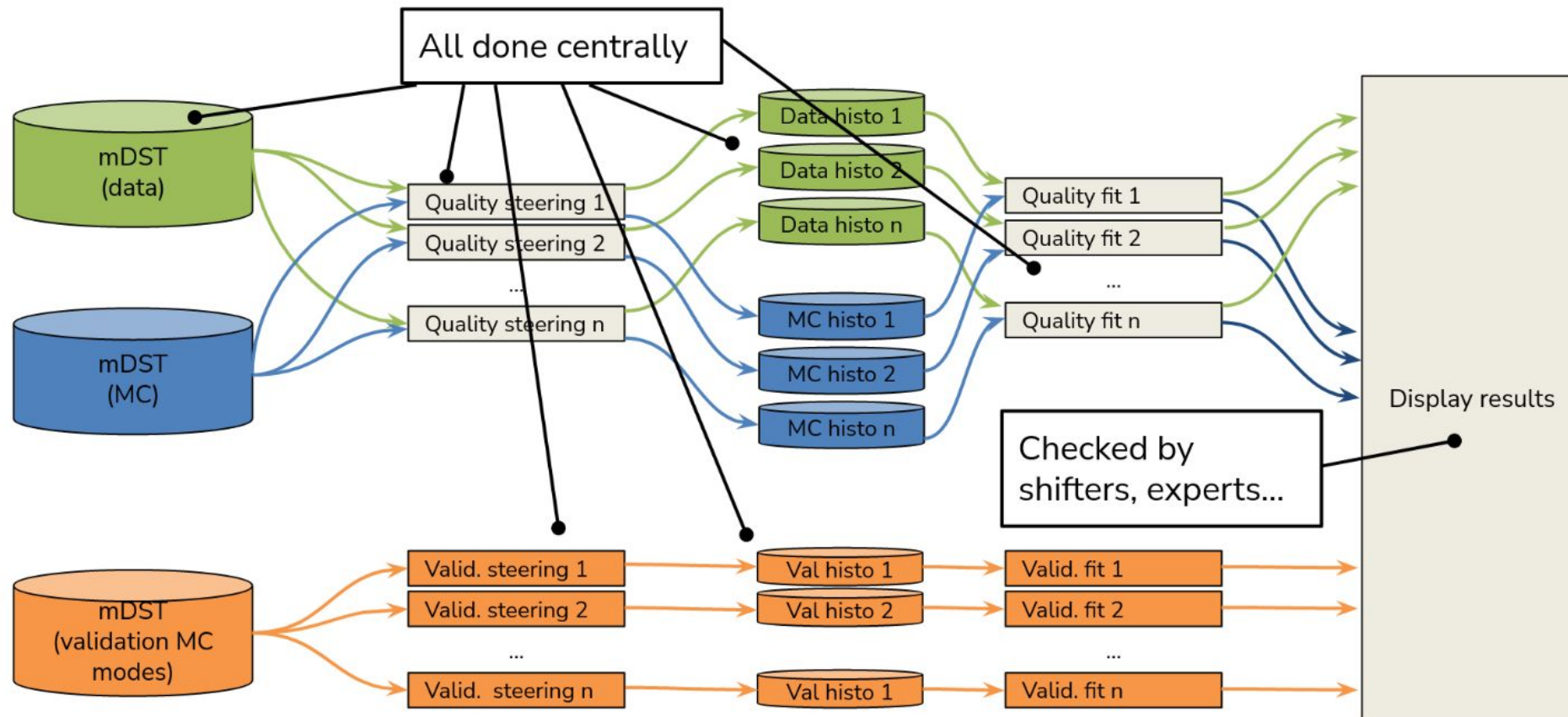
Introduction to RDataFrames

David Koch

And a short peek into uproot

- RDataFrame ROOTs answer to pythons root_pandas

- if you want or need to use ROOT in your analysis but dont want to miss pythonian style editing of your data you now can do so

- features:

  - easy to use

  - scalable: still experimental feature of supporting distributed running on clusters

  - interoperable with python ecosystem: you can convert RDataFrame into Numpy arrays

  - automated inspection:

    - create workflows or cut flows for inspection

  - automated systematic variations (still experimental)
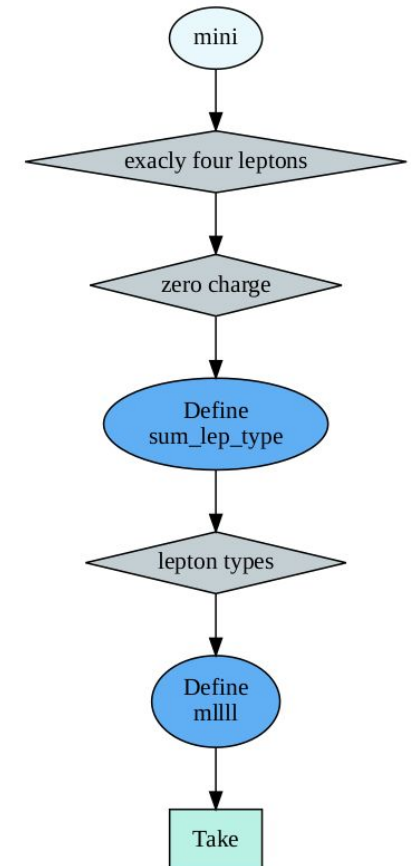
Belle II

Issues in Tracking/Performance

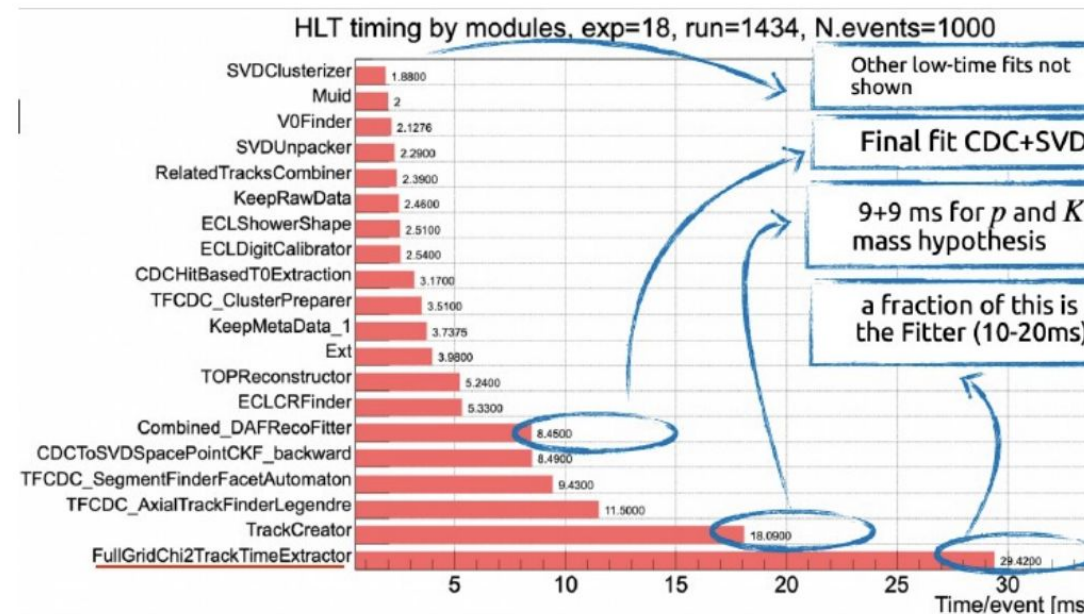### Tracking in Belle II
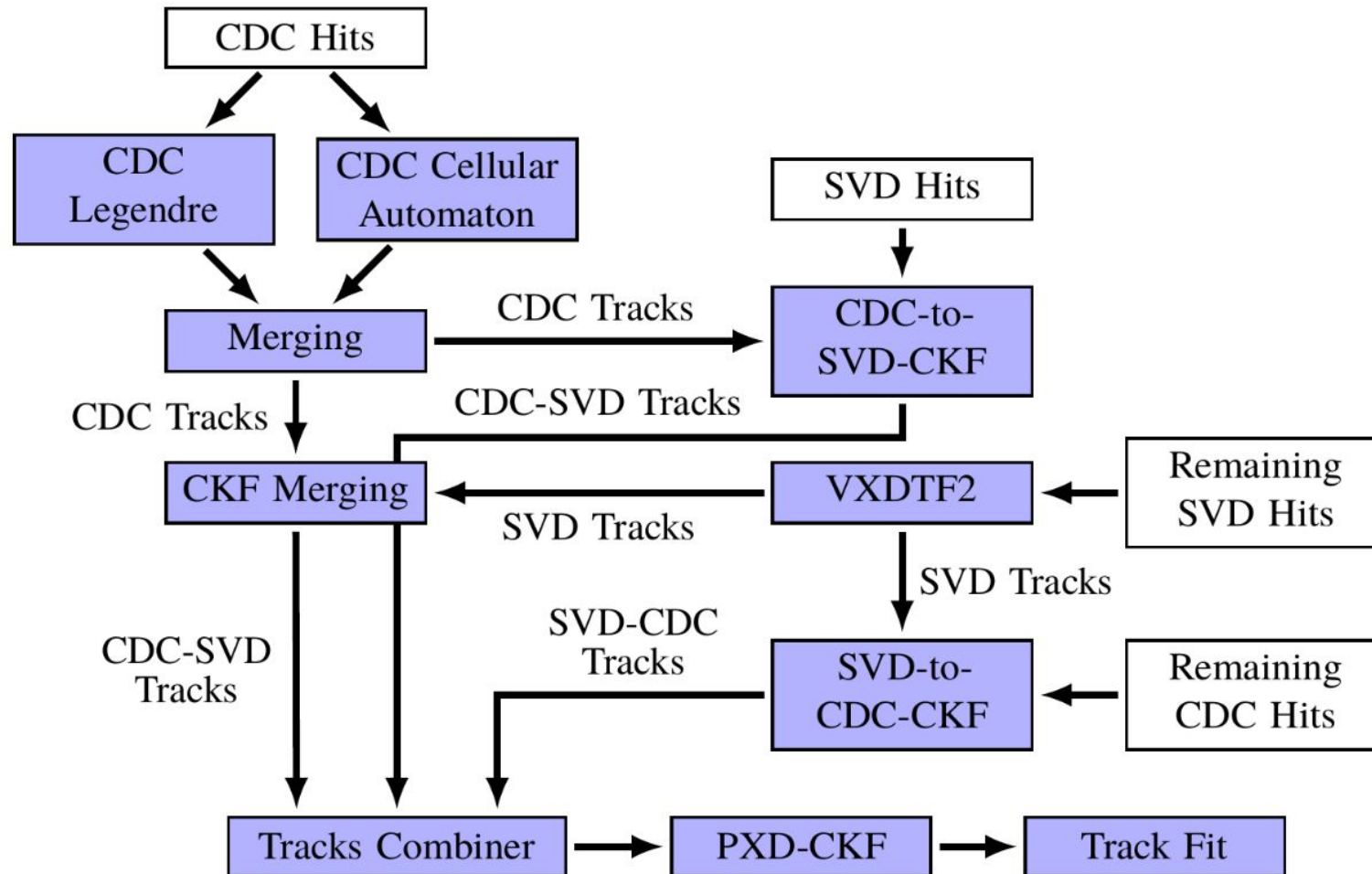
Tracking Overview

Christian Wessel

Issues in Tracking/Performance

- "Basically running stable".

- Issues in tracking/Performance:

  - **Fake** tracks (random combination of hits) and **clone** tracks (found multiple times)
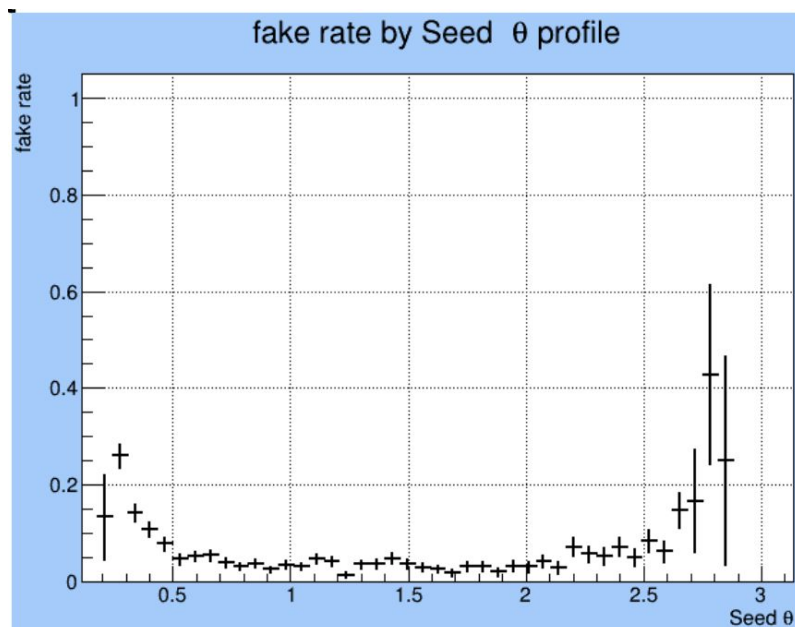
- Long execution time.

  - Several track **finding** algorithms take a long time per event.
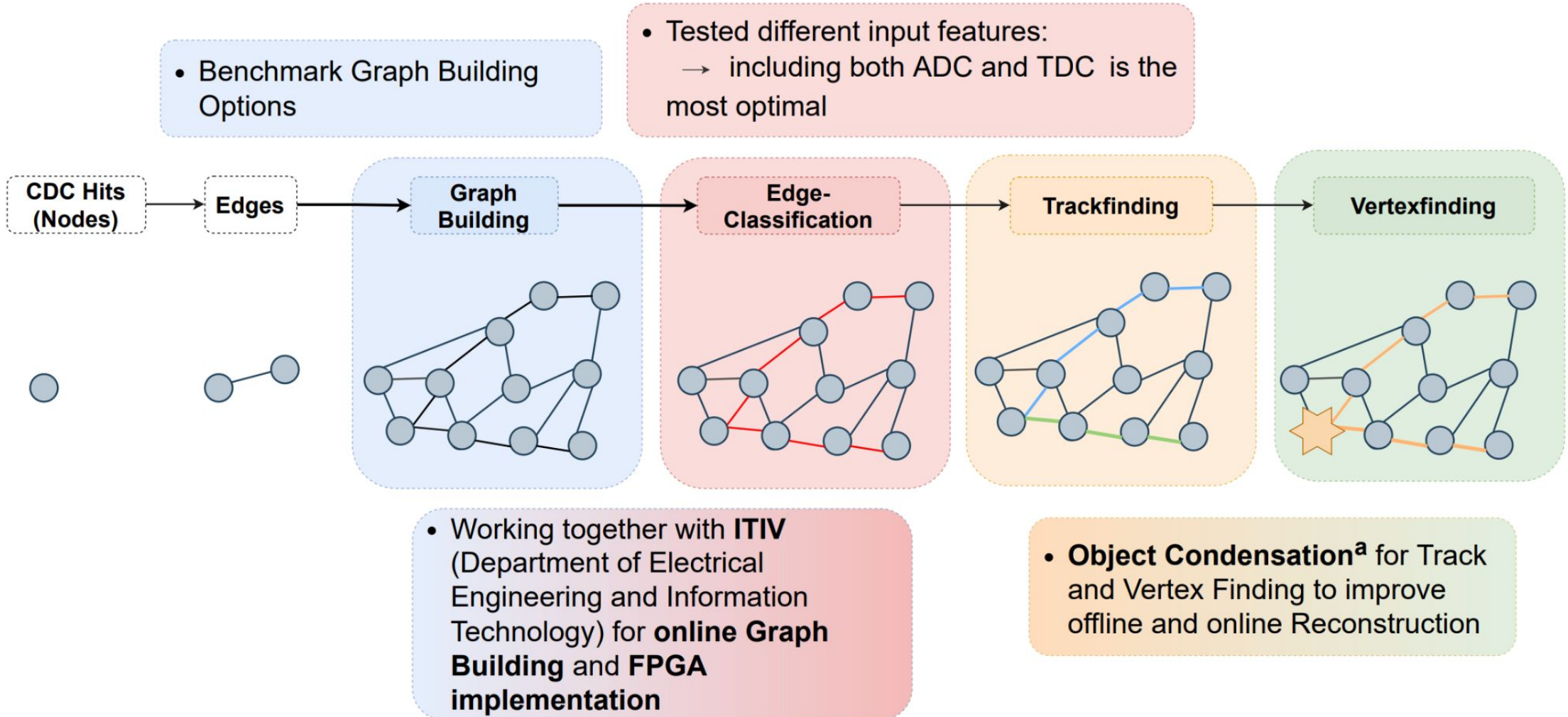
  - Track **fitting** is slow.

- **Call for help!** If you are interested in Tracking contact Thomas and Christian.

GNN-based Track and Vertex Finding

Lea Reuter et al.

Searches for displaced vertices

- Develop vertex finding using Graph Neural Networks.

- Benchmark Graph Building Options

- Tested different input features:
  → including both ADC and TDC is the most optimal

CDC Hits (Nodes) → Edges → Graph Building → Edge-Classification → Trackfinding → Vertexfinding

- Working together with **ITIV** (Department of Electrical Engineering and Information Technology) for **online Graph Building** and **FPGA implementation**

- **Object Condensation**[a] for Track and Vertex Finding to improve offline and online Reconstruction

Improving ECL Clustering on Trigger Level with Object Condensation

Isabel Haide et al.

During online reconstruction for L1 trigger decisions

- Object condensation (OC) for multi-particle reconstruction.

    - Objects and their defining properties are condensed into one representative condensation point.
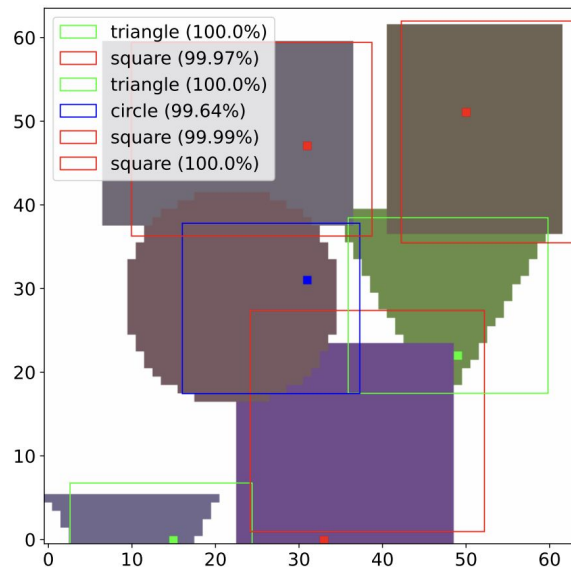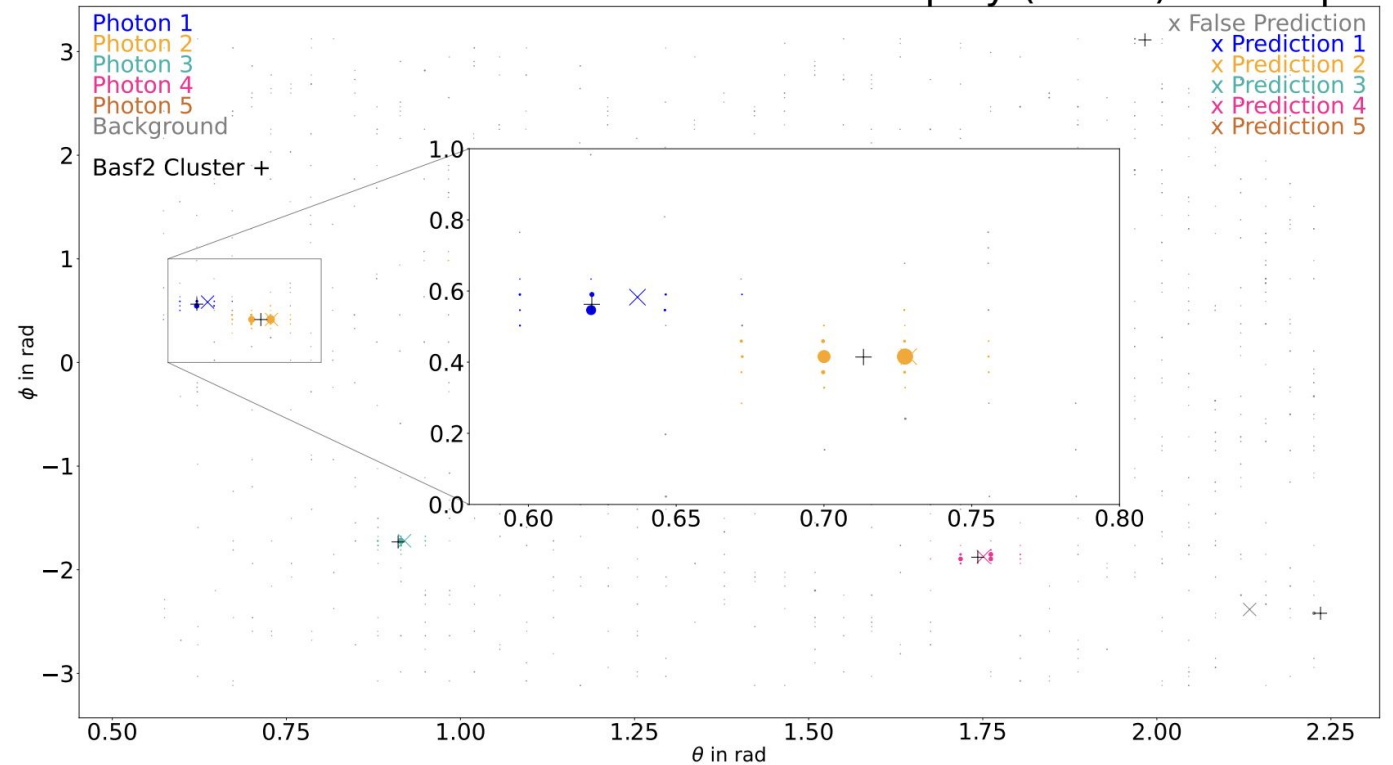
Improving ECL Clustering on Trigger Level with Object Condensation

Isabel Haide et al.

During online reconstruction for L1 trigger decisions

- Energy resolution and efficiency of basf2 still outperforms OC, especially for low-energy photons.

    - Improvements through tuned network architectures and optimized trainings.
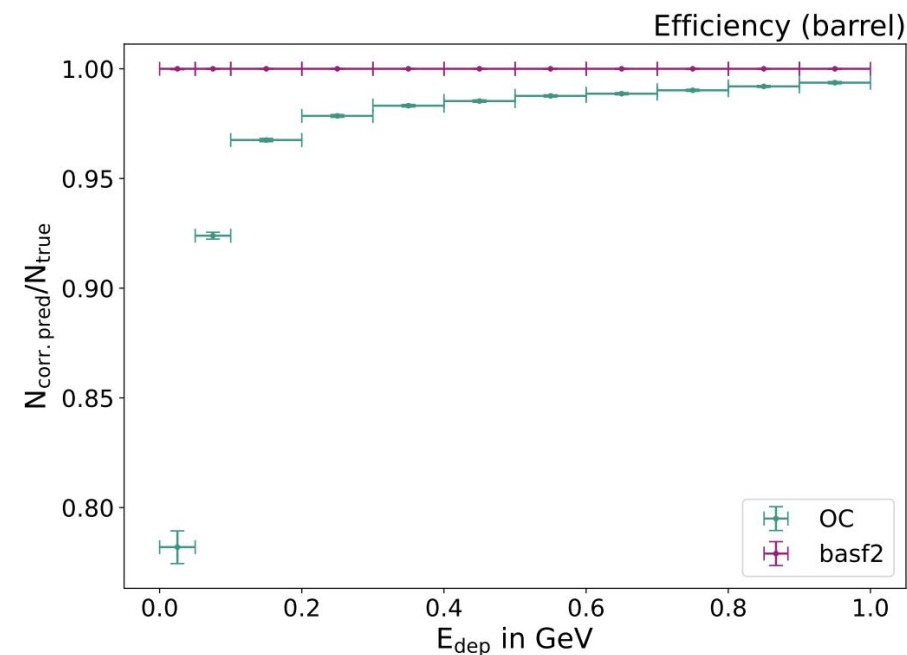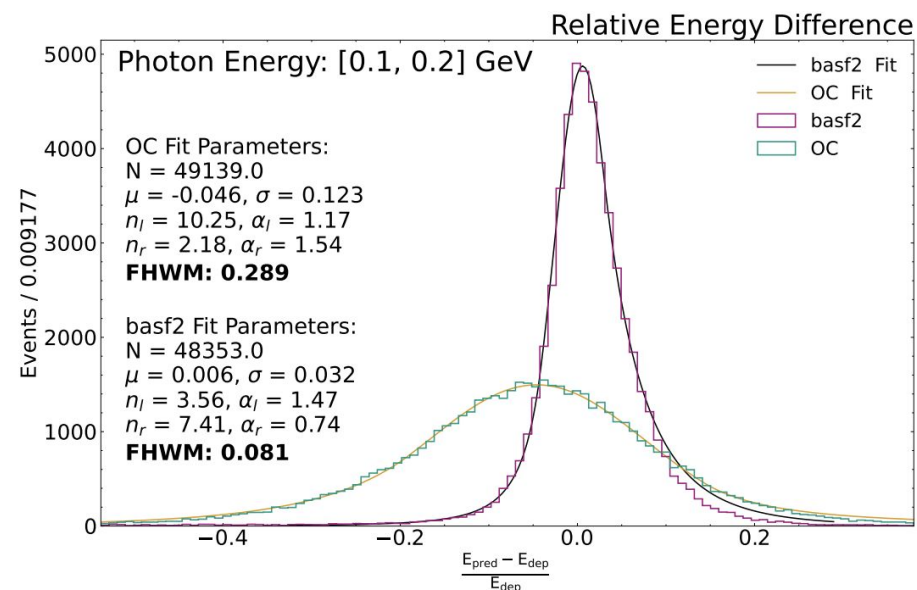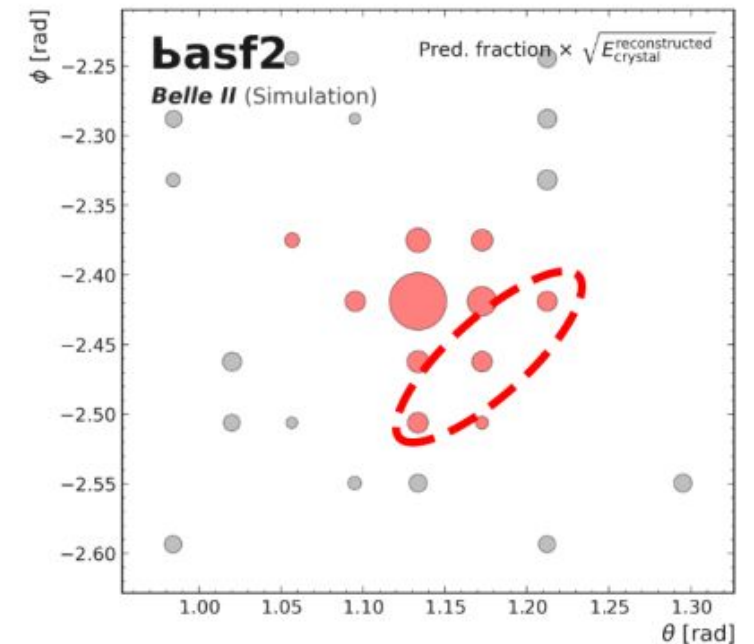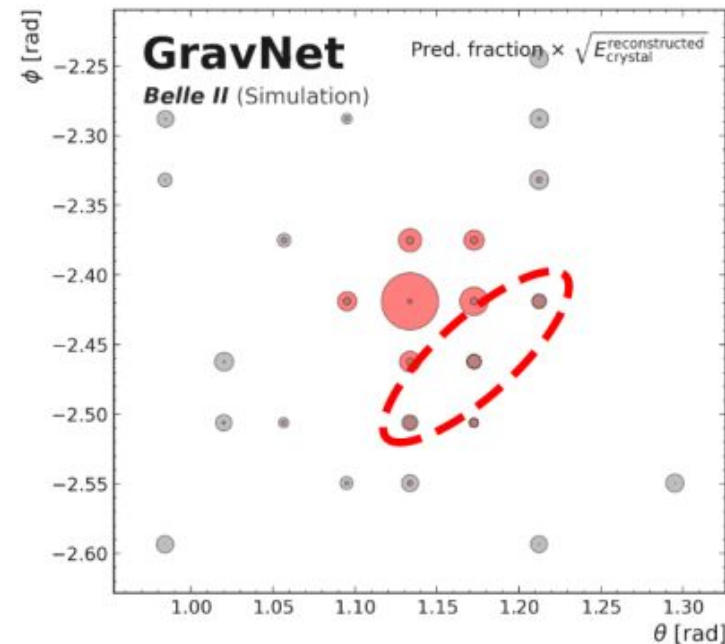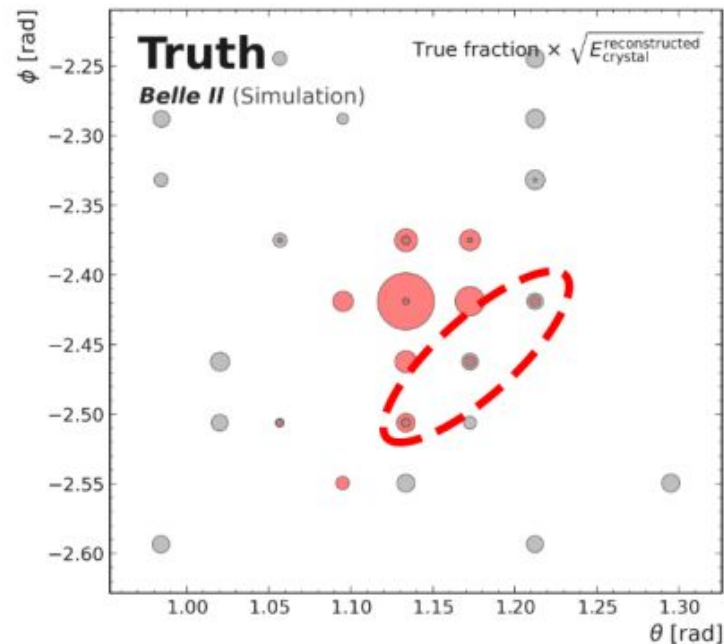
Clustering Energy Depositions in the ECL using GNNs

Soft clustering including background

Florian Wemmer et al.

- using Graph Neural Network on 9x9 cluster of crystals

- train on features: Energy, time, Pulse shape, crystal coordinates

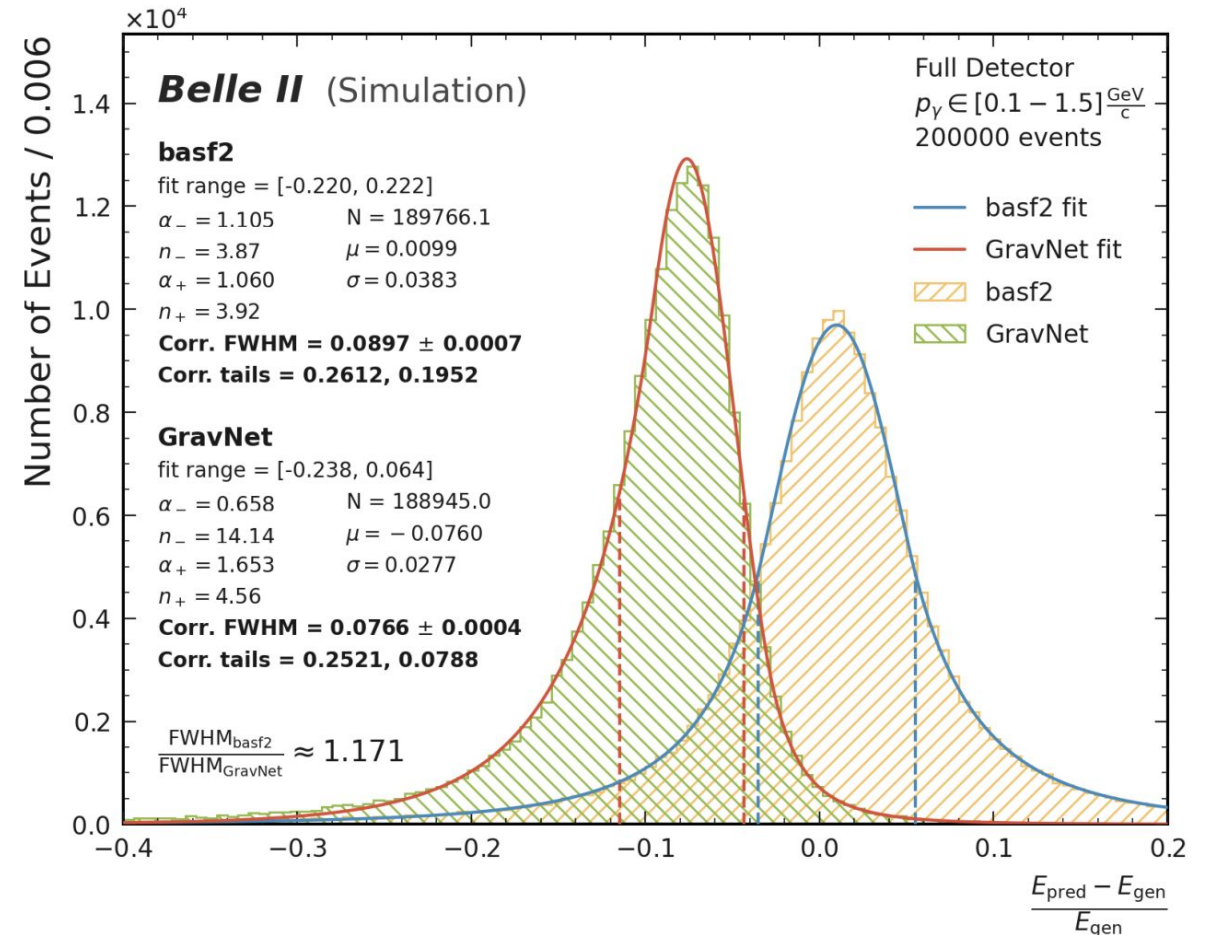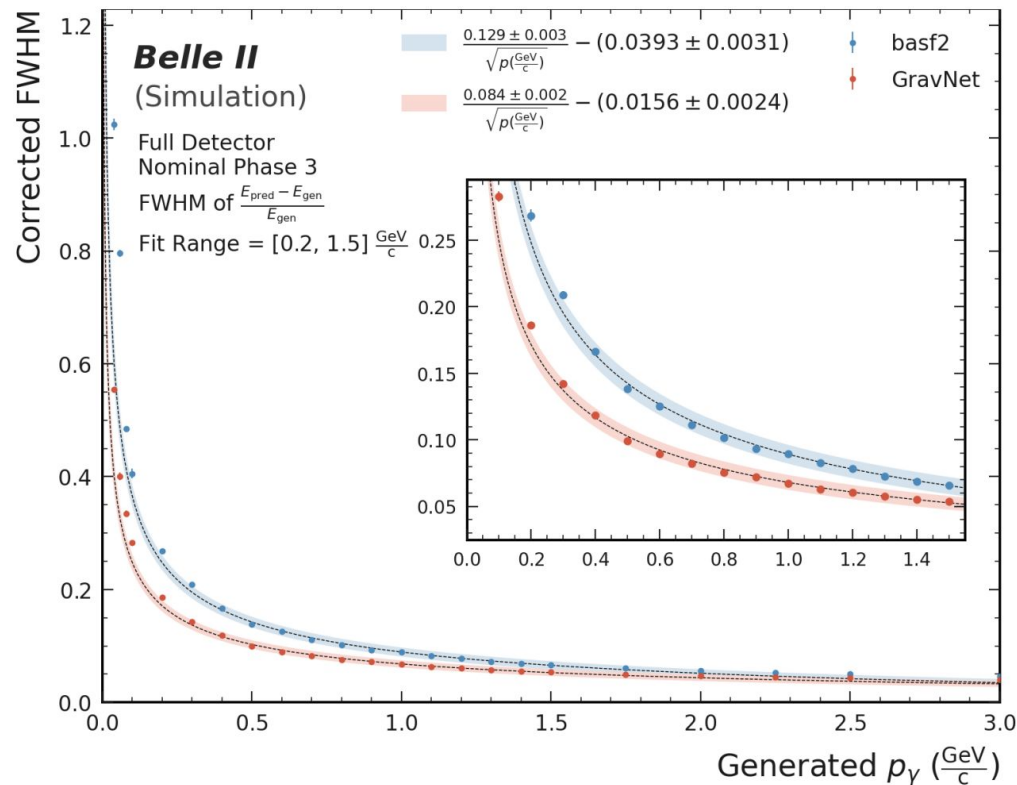- good discrimination between signal and background cluster

Clustering Energy Depositions in the ECL using GNNs

Soft clustering including background

Florian Wemmer et al.

- 17% improvement in resolution over basf2 algorithm
- next steps:
  - investigate feasibility for basf2
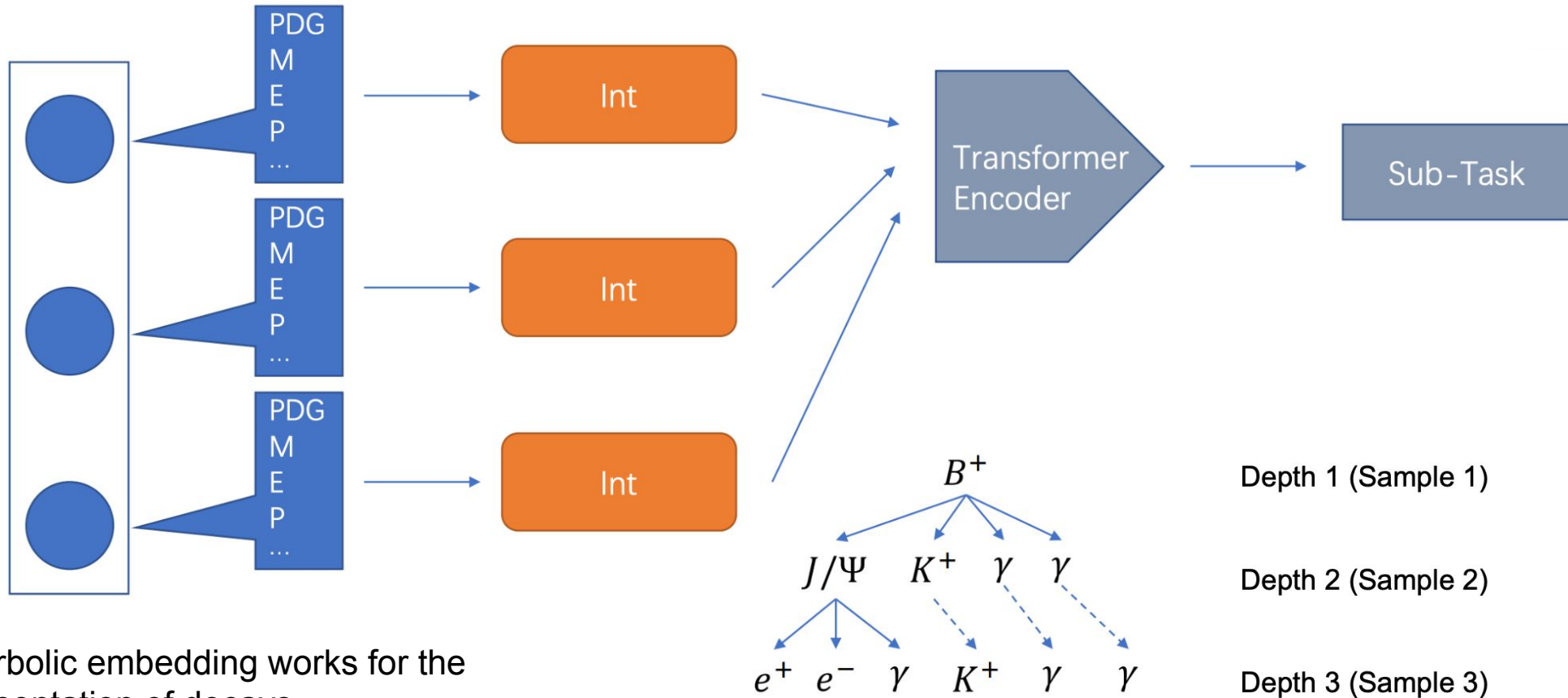 - tests on data
  - technical paper

Representation of Decay Relations in Hyperbolic Space

Prediction of decay channels from final-state particles

Boyang Yu

- Proof of concept: prediction of decay channels from final state particles for toy MC.

Depth 1 (Sample 1)

Depth 2 (Sample 2)

Depth 3 (Sample 3)

- Hyperbolic embedding works for the representation of decays

Each particle carries 12 features (**Bold** for reconstruction part)
PDG, mass, charge, energy, production time, x, y, z, **px**, **py**, **pz**, nDaughters

Belle II

15

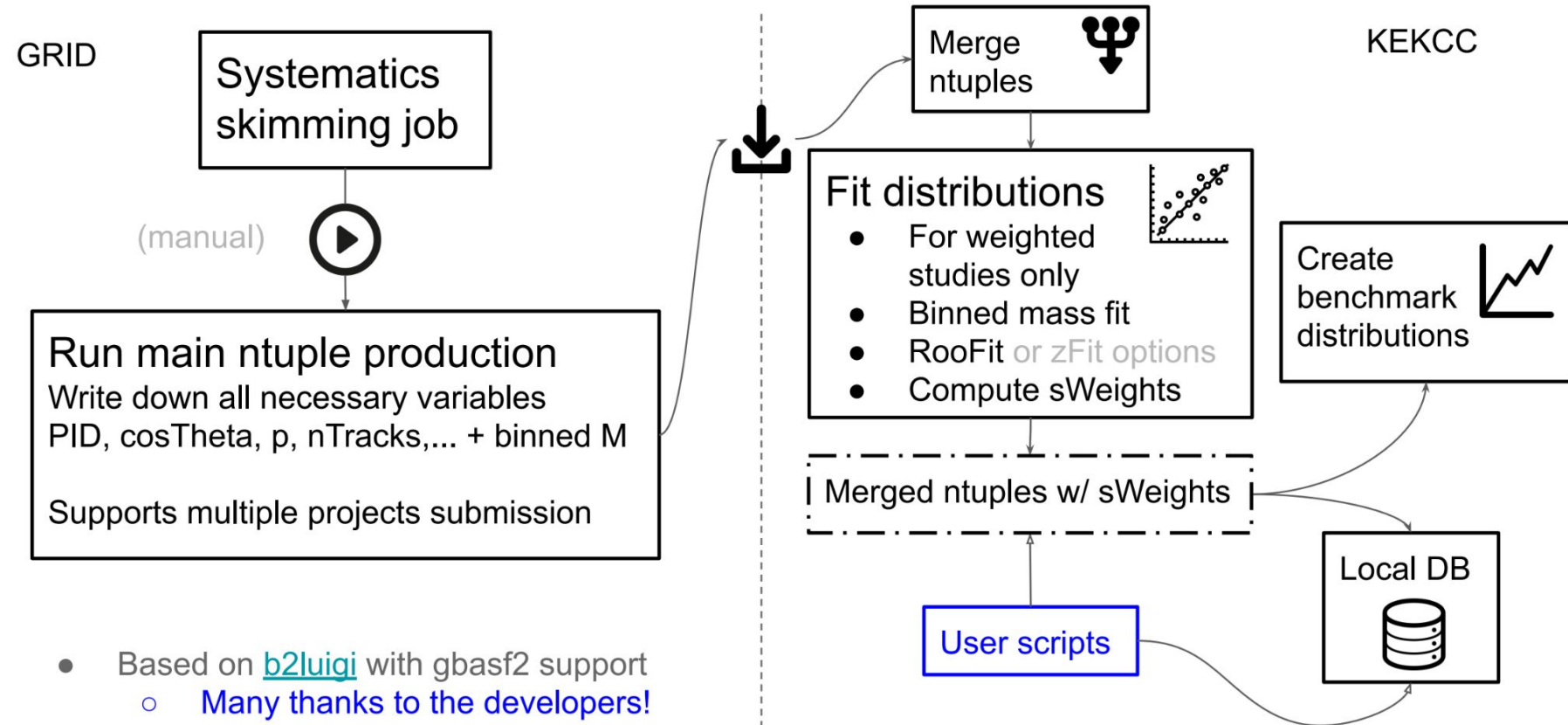# Systematics Framework

Sviatoslav Bilokin

Unify and automate computation of systematic uncertainties

- The framework works with the modes:

$$D^{*+} \rightarrow \left[D^0 \rightarrow K^- \ \pi^+ \right] \pi^+$$
$$\Lambda^0 \rightarrow p\pi^-$$
$$K_S \rightarrow \pi^+\pi^-$$
$$\left[\tau \rightarrow 3\pi\nu \right] \left[\tau \rightarrow l\nu\nu \right]$$

- Fit mass distributions for each weighted performance mode and compute sWeights and signal-like histograms.

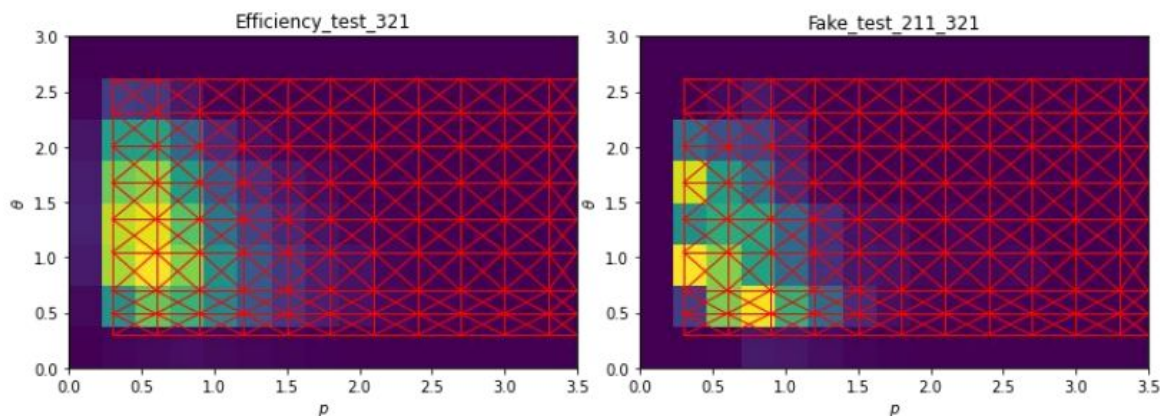- Ntuple production workflow for Weighted models:

- Based on b2luigi with gbasf2 support
  - Many thanks to the developers!

Systematics Framework

Sviatoslav Bilokin

New features available

- **Interactiion with PIDVar**
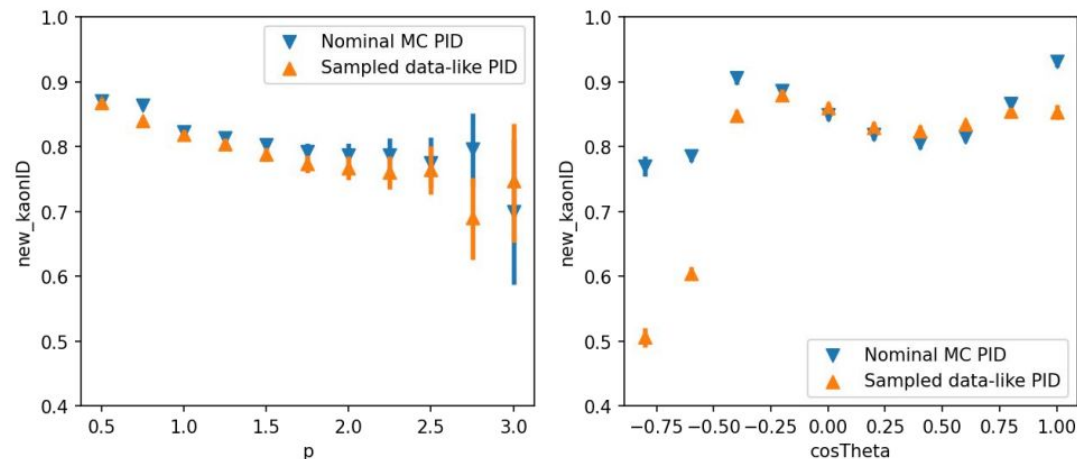 - Produce the weights in a notebook and immediately apply them to MC ntuple.

- **Multidimensional PID Fit**
  - Fit PID as function of (p,cosθ,...)

- **PID resampling**
 - ROOT files from the PID KDE fit are used to sample the data-like PID on MC.

- ✓ Added a possibility to integrate other types of studies
- ✓ Processed proc12 + all buckets up to 25th + s-proc's
- ✓ Introduced PID KDE Fit in LHCb fashion
- ➢ Integration of Lepton ID modes is in progress
- ☒ Create meta-variables in basf2 fashion
- ☒ Duplicate dataset to other servers, e.g. BNL or DESY
- ☒ Integration with B2Production framework
- ☒ Integration with basf2 and b2conditiondb

Workflow Management in HEP

Single executable, full analysis automatized and portable

Caspar Schmitt
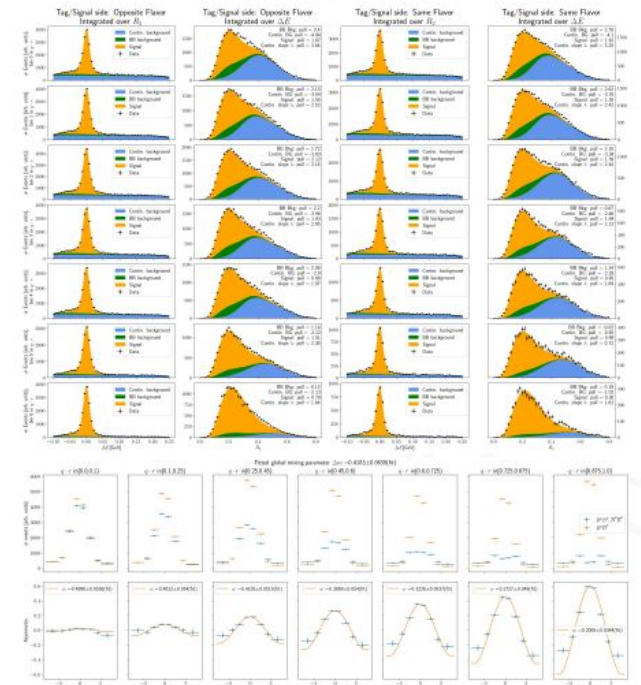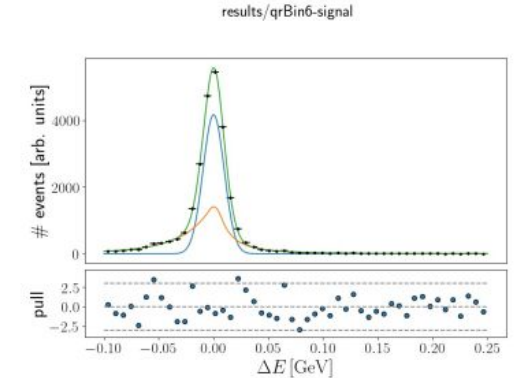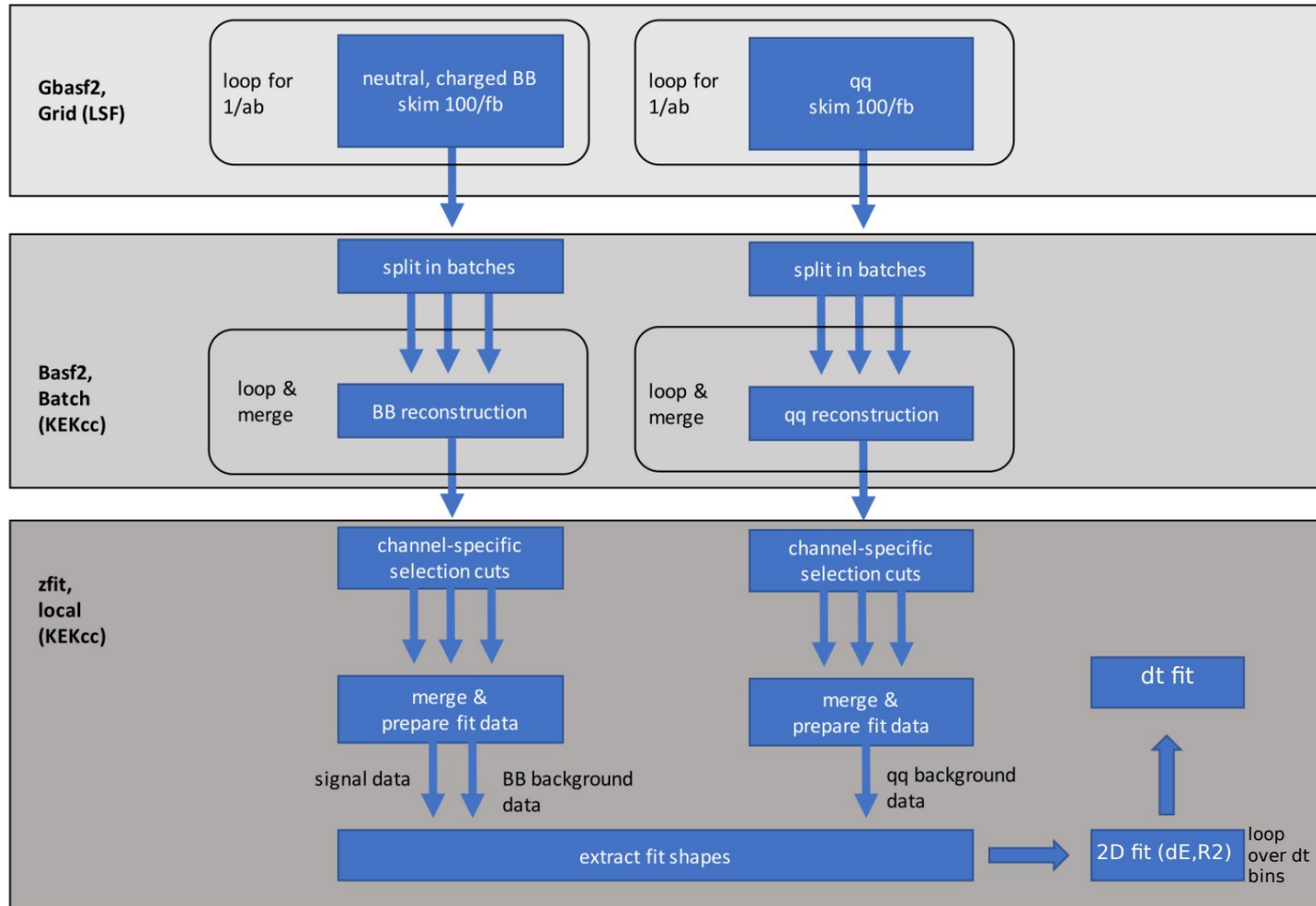
Workflow Management in HEP

Caspar Schmitt

Comparison of Selected Workflow Management Systems

Single applications:

Luigi

- ▶ developed by Spotify, open sourced
- ▶ Python syntax
- ▶ target based
- ▶ <u>integrated</u> analysis code and WF logic
- ▶ focus on <u>dynamic DAG visualization</u> and remote execution support
- ▶ automatised job steering on LSF and KEKcc...

Snakemake

- ▶ developed at Uni Duisburg-Essen
- ▶ Custom Python-based syntax
- ▶ target based
- ▶ Analysis code and WF logic <u>factorize</u>
- ▶ focus on <u>environment management</u> and remote execution support
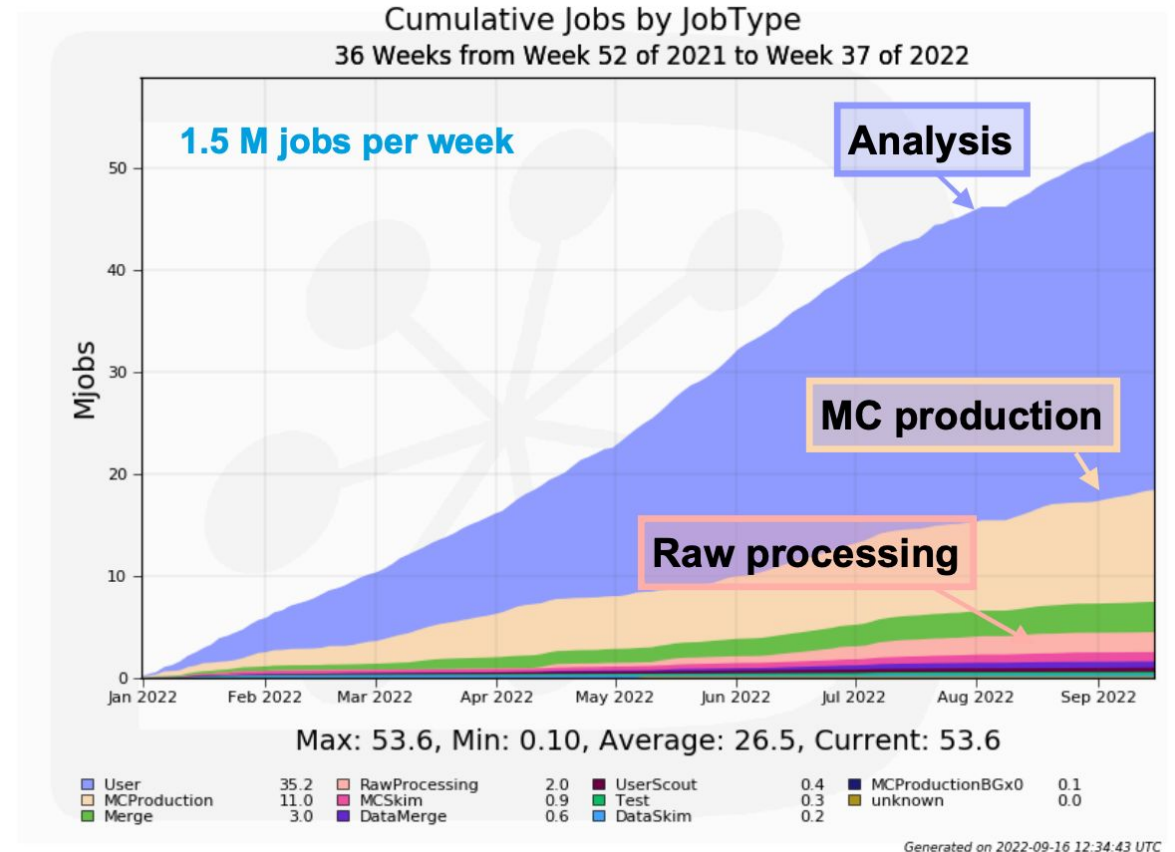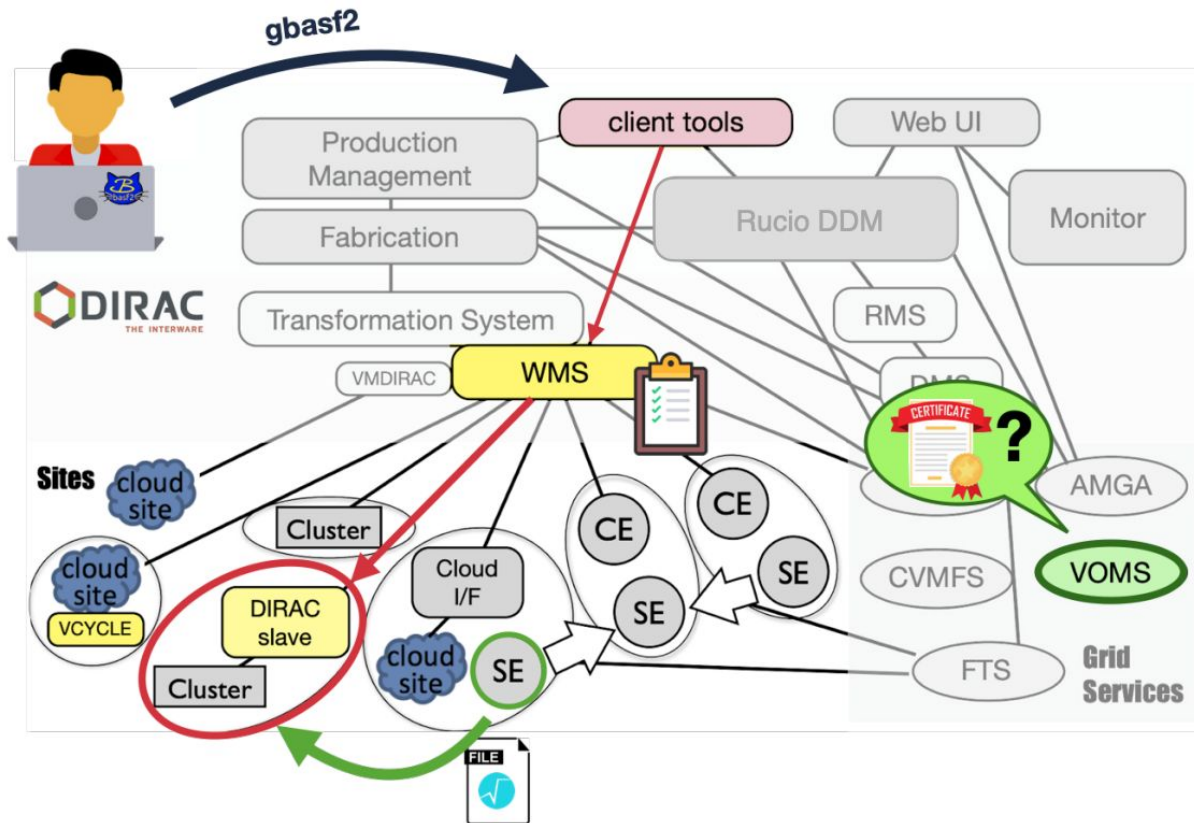- ▶ automatised job steering on LSF and KEKcc...

Server-based: e.g. **Reproducible Analysis Platform** developed by CERN.

Distributed Computing at Belle II

Michel Villanueva

Analysis on the grid

- DIRAC pilots allow to integrate heterogeneous computing resources with a single interface (gbasf2).
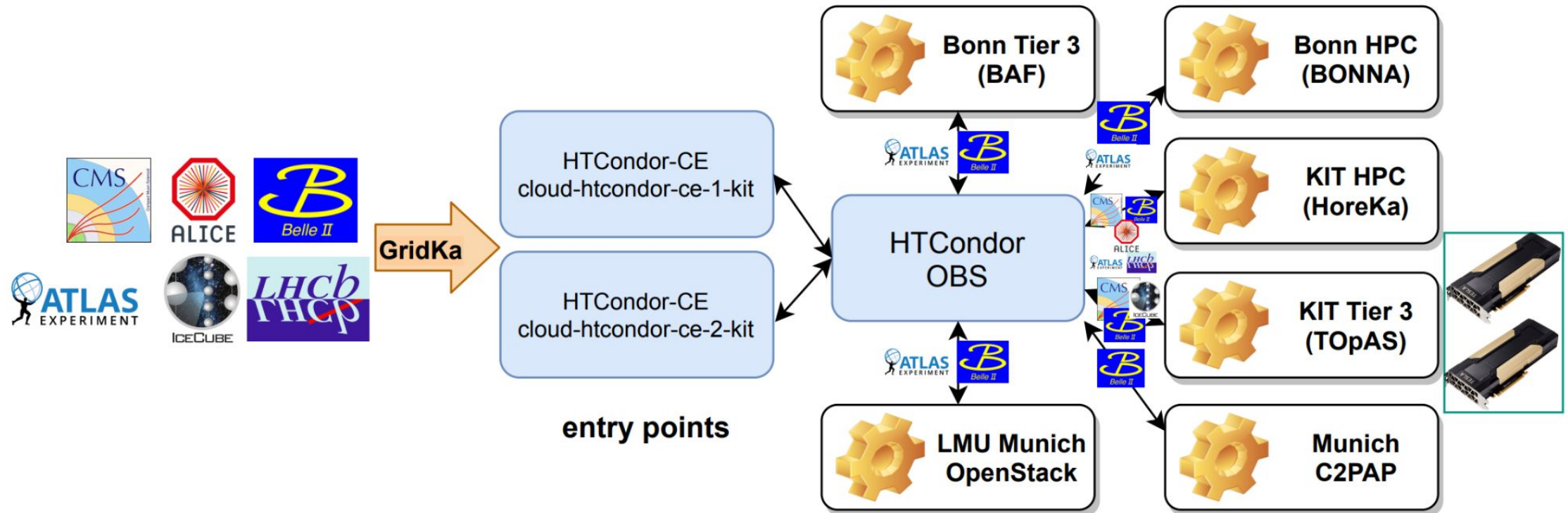

- Priorities on development during LS1 presented.

Belle II Grid Computing Developments at KIT

Opportunistic Resources for Belle II

Moritz Bauer et al.

- Dynamically providing up to 17k cores, 2k of those to Belle II.
  - Always looking for sites interested in joining!

Belle II Grid Computing Developments at KIT

## Caching with XRootD   Moritz Bauer et al.

- Why caching?
  - Some sites and some datasets are in higher demand than others.
  - Sometimes sites fail while SE is still functional, leading to waiting jobs at remaining sites.

- Implemented caching setup using xRootD.
  - XRootD is not supported on all sites but HTTPS is almost universal.
  - Trying to come up with solutions for an HTTP(S) redirector plugin.

- Monitoring setup also available:

NAF@DESY

Thomas Hartmann
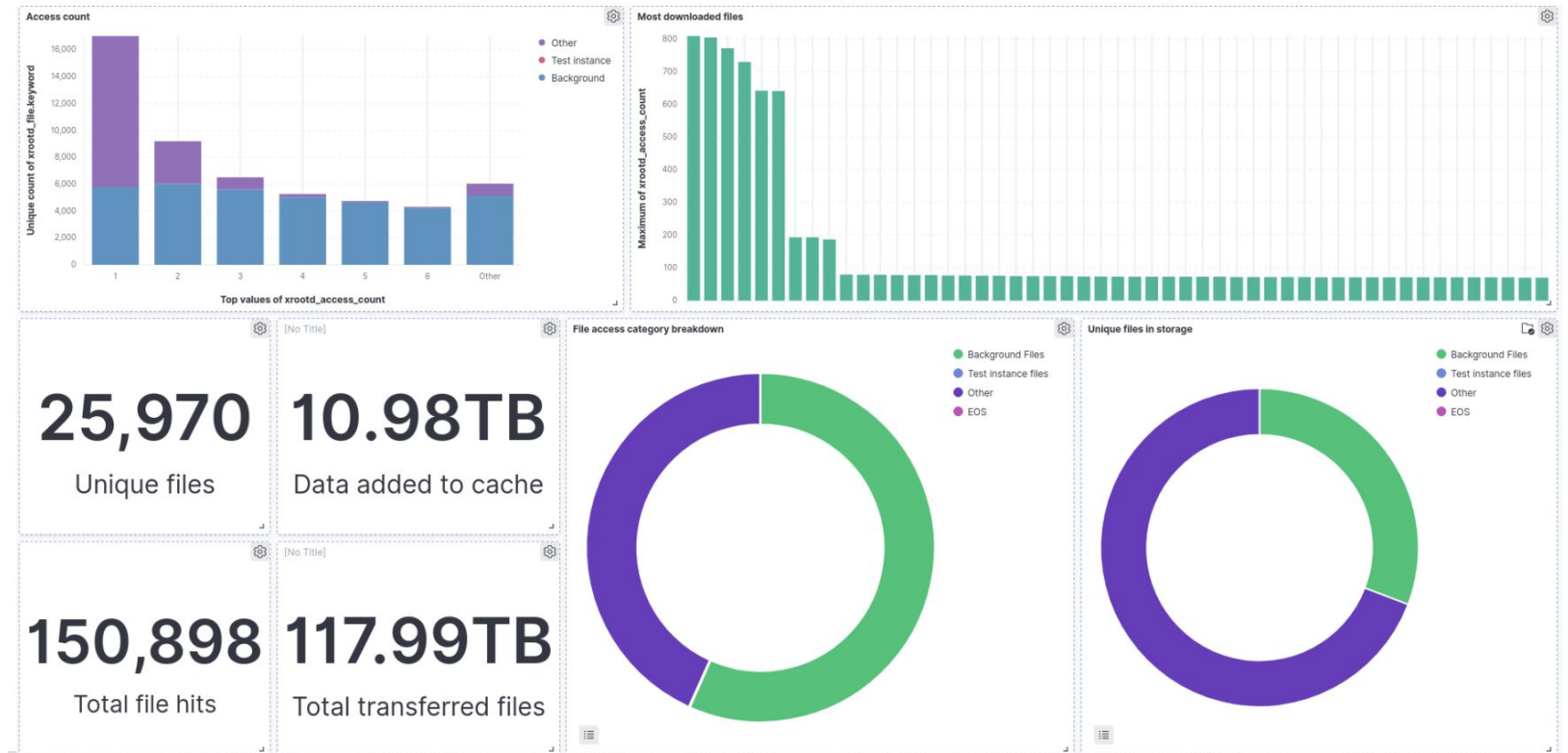
Complementing the Grid

High Throughput Computing

High Performance Computing

NAF

- same foundation as the Grid
  - HTCondor
  - HW in conjunction with the Grid
  - NAF nodes with user workload optimization

- Grid: a few dedicated prod roles & workflows

- NAF: broad spectrum of users and groups

~1600 users

Belle II

23

NAF@DESY

Thomas Hartmann
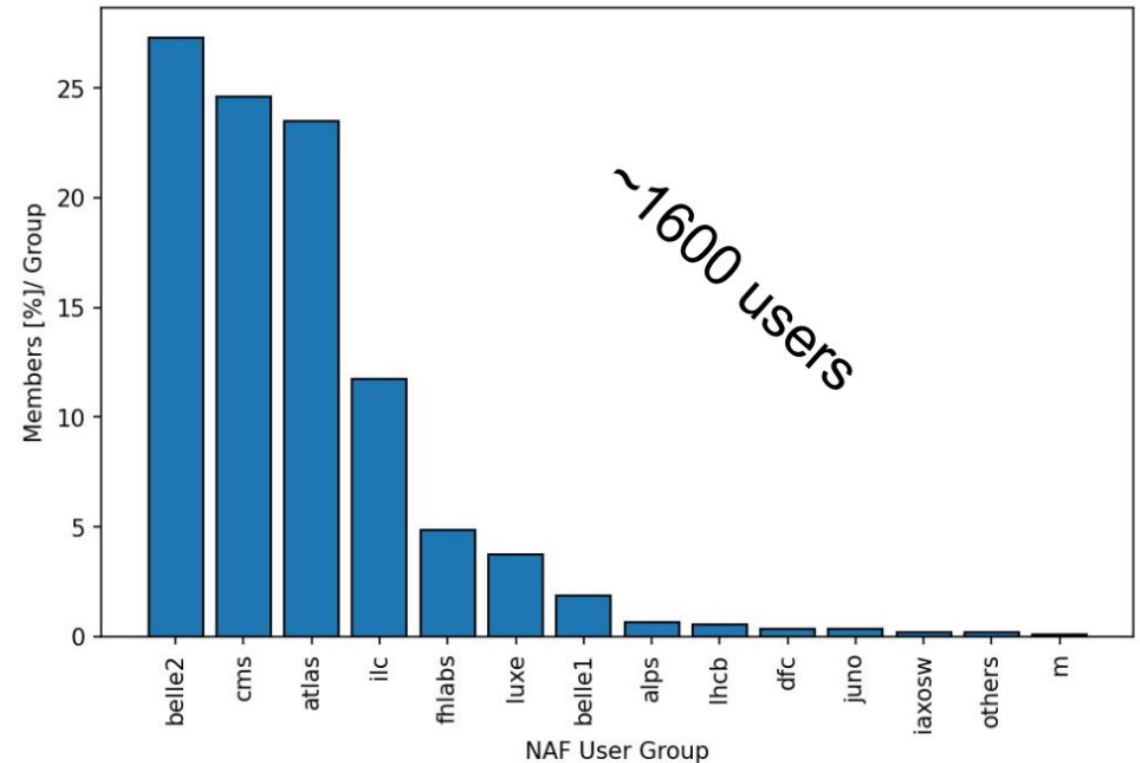
Access

- User access via:

  - SSH: classic shell access.

  - FastX: browser based X11.

  - Jupyter: Interactive Python Notebooks.

- Notebook jobs start transparently in the background.
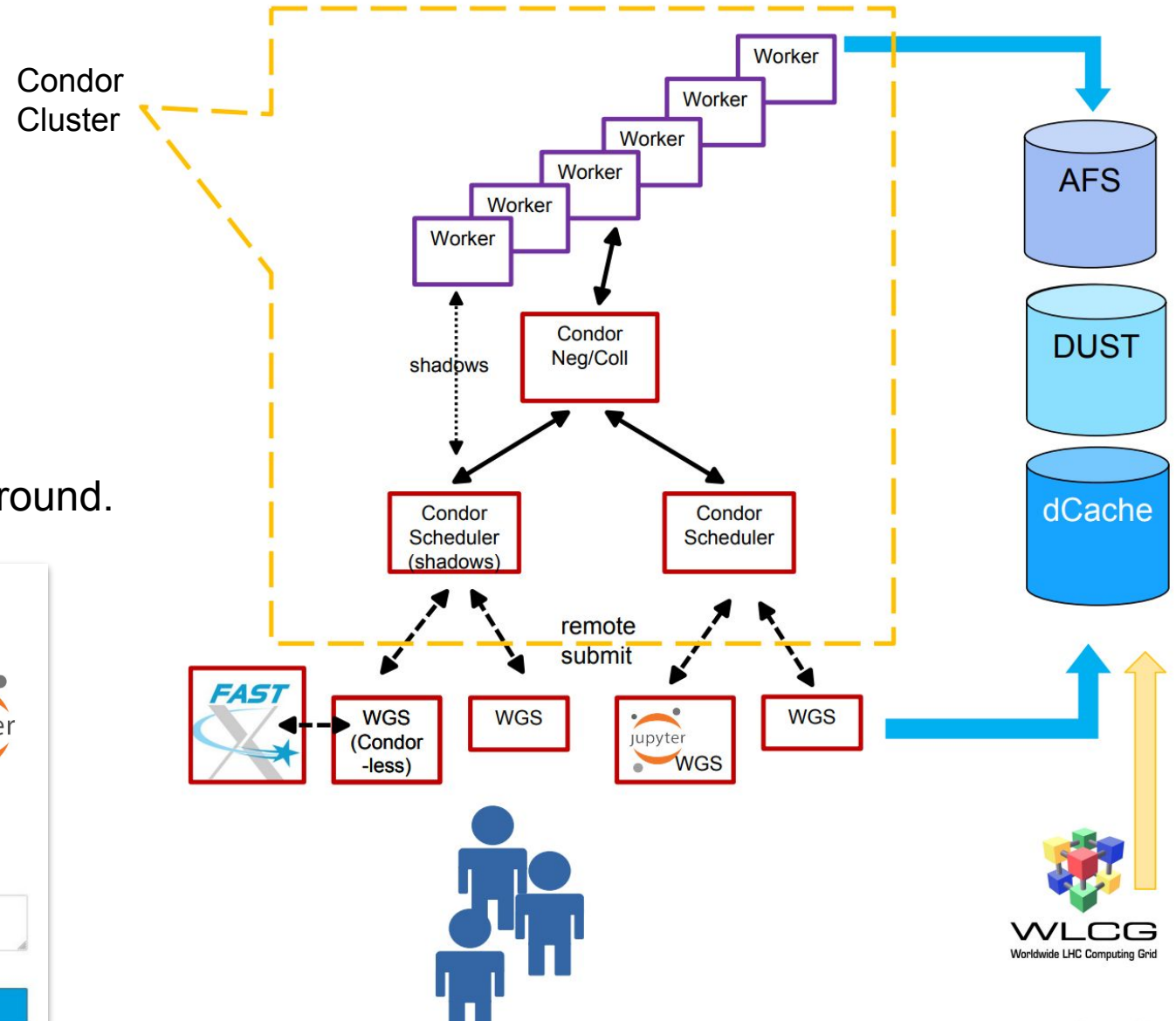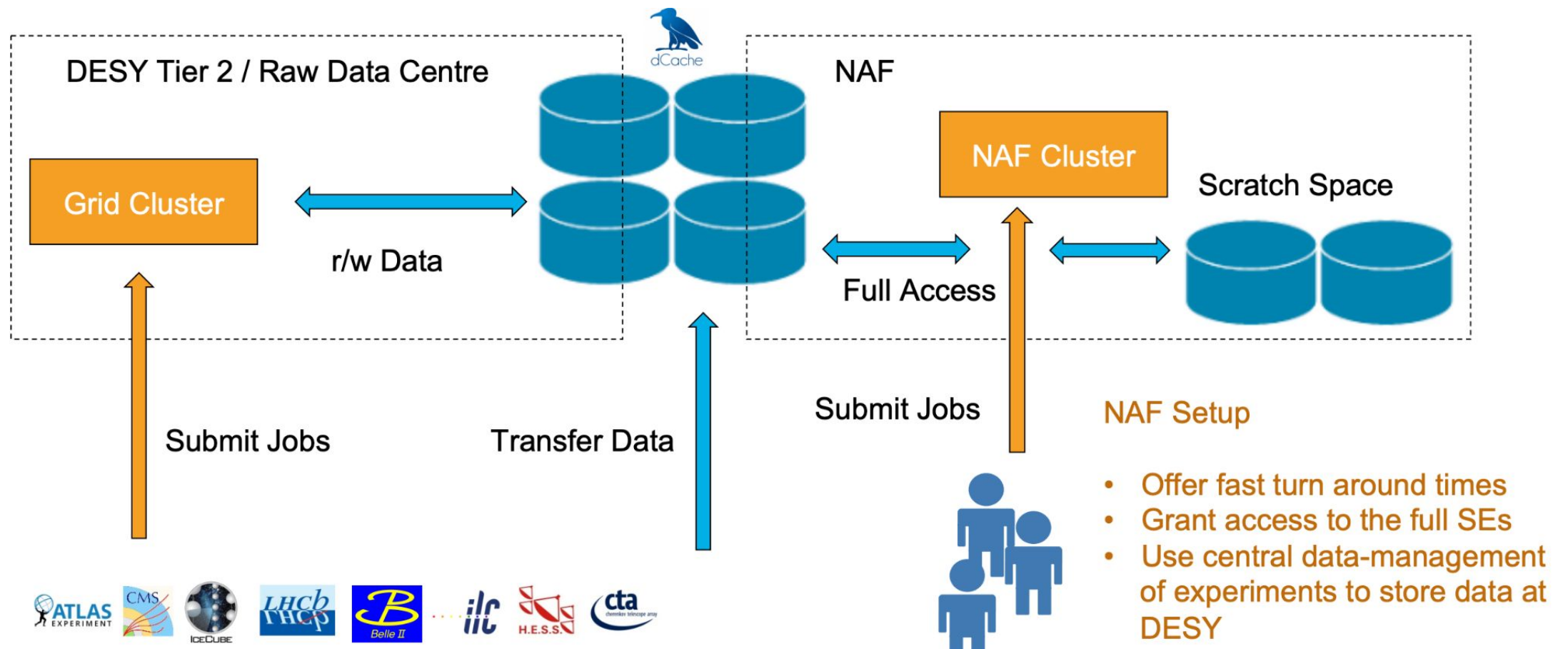
Storage infrastructure at DESY

Mass-Storage for Belle II in Grid and NAF   Christian Voss

- dCache storage offers a uniform namespace **available via NAF and Grid** (DESY and remote).
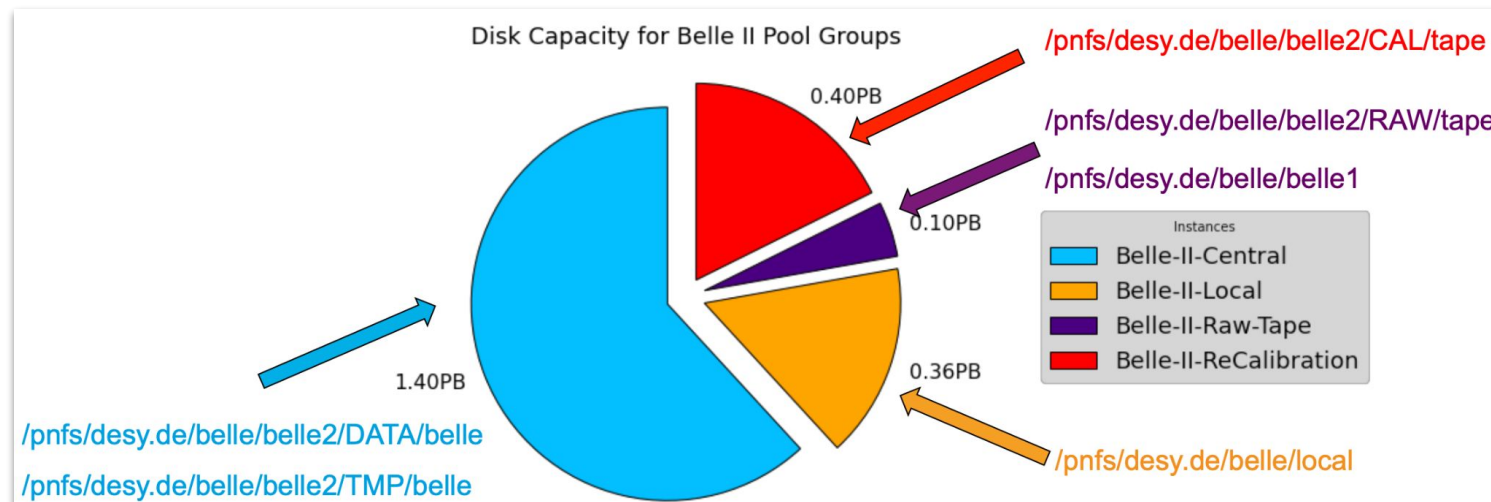
Storage infrastructure at DESY

Disk and Tape system
Christian Voss

- Pool groups are accessible via paths in the namespace.

- Access to directories with all supported protocols in Grid and NAF.

Disk Capacity for Belle II Pool Groups

/pnfs/desy.de/belle/belle2/CAL/tape
/pnfs/desy.de/belle/belle2/RAW/tape
/pnfs/desy.de/belle/belle1
/pnfs/desy.de/belle/local
/pnfs/desy.de/belle/belle2/DATA/belle
/pnfs/desy.de/belle/belle2/TMP/belle

0.40PB
0.10PB
0.36PB
1.40PB

Instances
Belle-II-Central
Belle-II-Local
Belle-II-Raw-Tape
Belle-II-ReCalibration

- Tape still in heavy use.
  - High investment but very limited running costs.
  - If used correctly: good streaming performance.

- 1 PB of usage by Belle II.
  - Workflows for accessing Raw and Calibration data are different.
  - They may be reviewed to improve performance.

Belle II
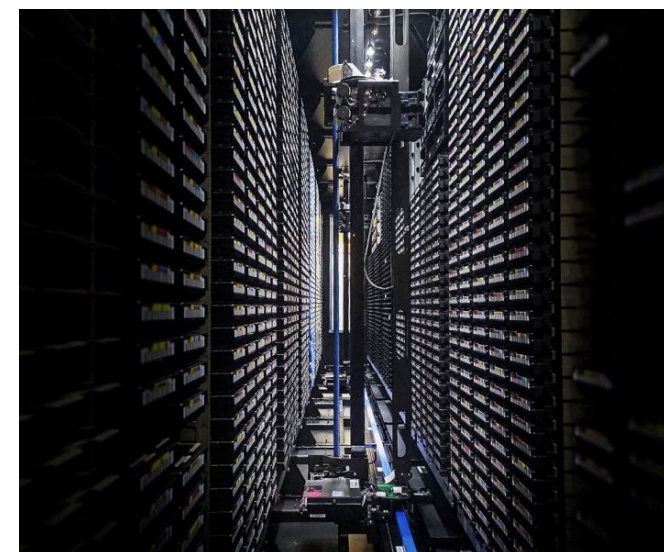
Calls for help!

■ Flaming needs:

 ■ New convenors for the Timing and Event T0 group
  ■ Both Giulia and Marko would like to step down 🔥
 ■ A new convenor for the Documentation and training group
  ■ Kilian left the collaboration :(🔥🔥🔥
 ■ Note: we already contacted few persons, but we received only "no"s

- Computers are not so smart. Sometimes, they fail.
  - **"Sometimes" x Huge Resources = "Often"**
  - The computing system need 24 hour x 7 day care.

- Please help us as a Data Production Shifter. You can book at shift.belle2.org
  - A nice manual is already prepared.
  -
- If you have some experience as data production shifter, please become an Expert Shifter.
  - The Expert Shifter training course is open.

Belle II

Thank you