Data Analysis Tools in Belle II

JENNIFER2
Computing Workshop

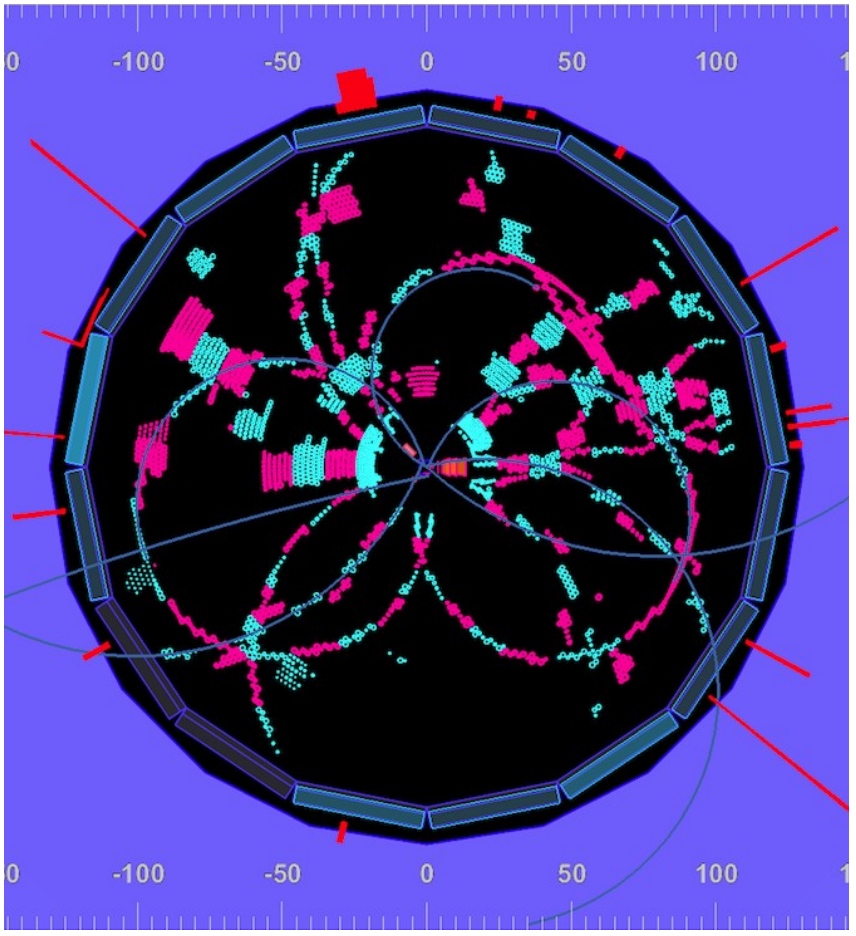
Thomas Kuhr

12.12.2019

Belle II Events

- e^+e^- collisions at $\sqrt{s} \approx 10.6$ GeV
- Design luminosity 8×10^{35} $cm^{-2}s^{-1}$ (40x KEKB)

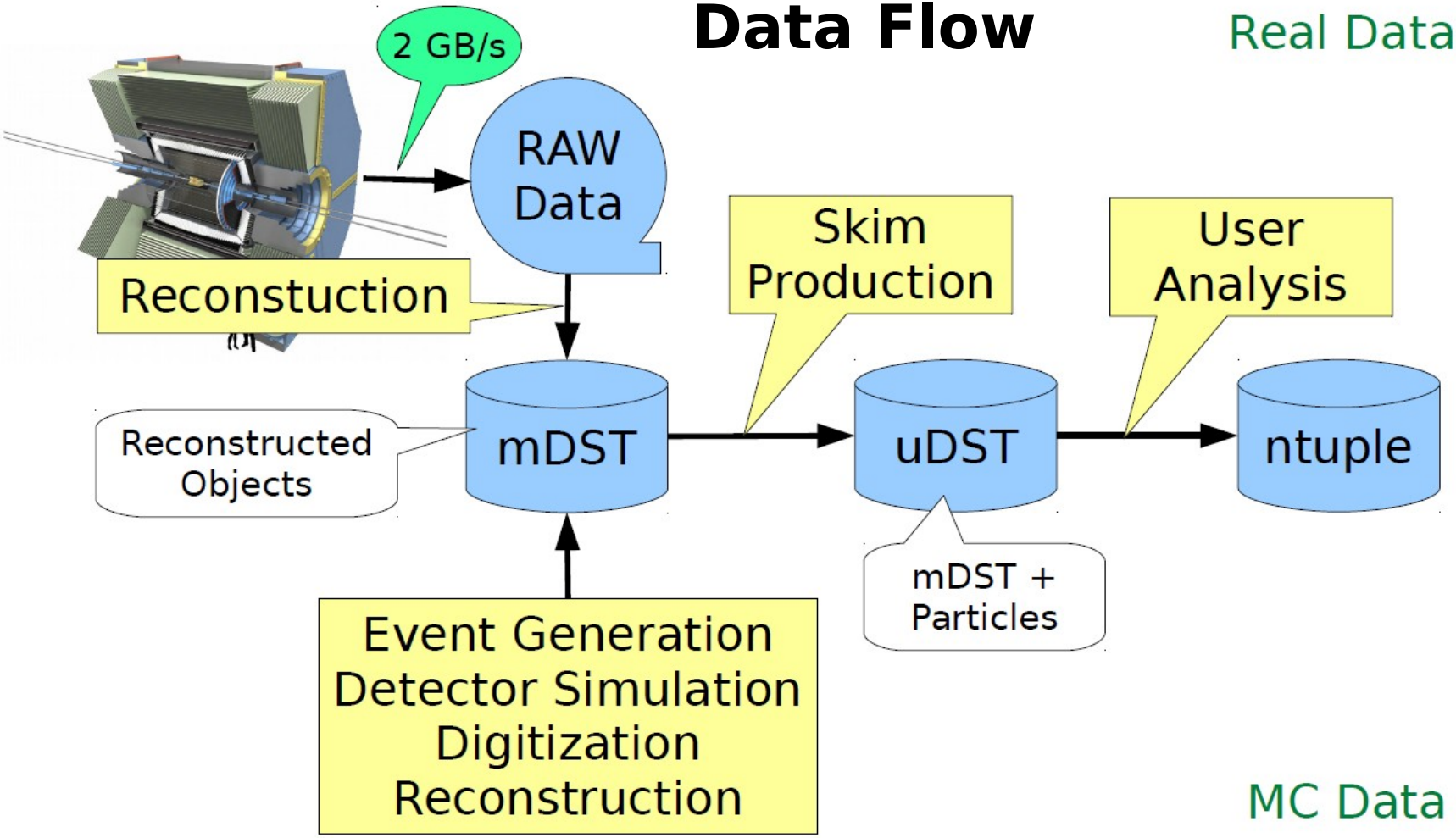
Type of event	Rate [kHz]
$\Upsilon(4S) \rightarrow B\bar{B}$	0.8
$q\bar{q}$	3.0
$\tau\,\bar{\tau}$	0.7
$e^+e^-(\gamma)$	240
$e^+e^-e^+e^-$	32



- O(10) tracks per Y(4S) event

Data Flow

Real Data



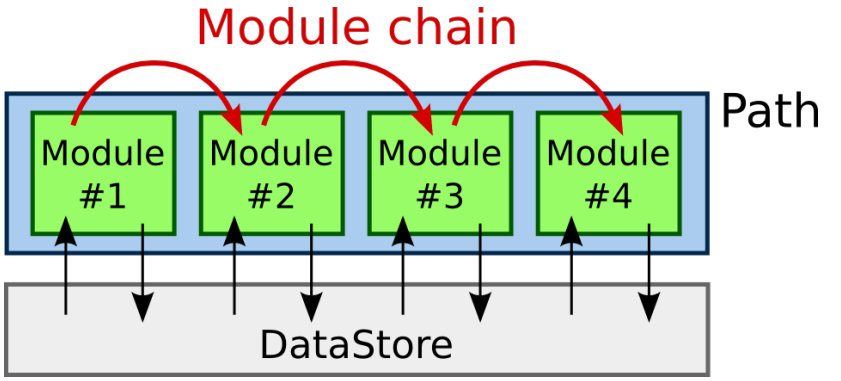
MC Data

Data Analysis Workflow

- Signal MC, Data/MC skims
- Reconstruction, (pre-)selection
- Tools like vertex fits, flavor tagging, etc.
- Ntuples
- Selection optimization
- Corrections / calibrations
- Fit
- Systematics
- Result

Belle II Analysis Software Framework (basf2)

- Used online and offline
- Dynamic loading of modules
- Data exchange via DataStore
- Relations
- Root I/O
- Parallel processing
- Steering via python
  → meta-frameworks




```
StoreArray<Track> tracks;
for (const Track& track: tracks) {
  const PIDLikelihood* pid =
    track->getRelated<PIDLikelihood>();
}
```

Modular Analysis

```
import modularAnalysis as ma
import variables.collections as vc

ma.inputMdst(...)

# create Ks -> pi+ pi- list from V0
# keep only candidates with 0.4 < M(pipi) < 0.6 GeV
ma.fillParticleList('K_S0:pipi', '0.4 < M < 0.6', path=...)

# reconstruct J/psi -> mu+ mu- decay
# keep only candidates with 3.0 < M(mumu) < 3.2 GeV
ma.reconstructDecay('J/psi:mumu -> mu+:loose mu-:loose', '3.0 < M < 3.2', ...)

# reconstruct B0 -> J/psi Ks decay
# keep only candidates with 5.2 < M(J/PsiKs) < 5.4 GeV
ma.reconstructDecay('B0:jspiks -> J/psi:mumu K_S0:pipi', '5.2 < M < 5.4', ...)

# perform B0 kinematic vertex fit using only the mu+ mu-
# keep candidates only passing C.L. value of the fit > 0.0 (no cut)
vertex.vertexRave('B0:jspiks', 0.0, 'B0 -> [J/psi -> ^mu+ ^mu-] K_S0', ...)

# build the rest of the event associated to the B0
ma.buildRestOfEvent('B0:jspiks', ...)

# perform MC matching (MC truth association)
ma.matchMCTruth('B0:jspiks', ...)

# determine B flavor
flavorTagger.flavorTagger('B0:jspiks', ...)

# create and fill flat Ntuple with MCTruth and kinematic information

ma.variablesToNtuple('B0:jspiks', vc.mc_truth + vc.detae_mbc, ...)
```

- ➢ Analysis on steering file level using decay strings
- ✔ Particle reconstruction and selection
- ✔ MC matching
- ✔ Vertex fits
- ✔ Flavor tagging
- ✔ Continuum suppression

Variable Manager

- Particle properties
- Functions,
  e.g. daughter, matchedMC
- Event properties
- Aliases
- Few hundred variables
- Output to:
  - Ntuple (candidate based)
  - Event based tree
  - Histogram
  - HDF5

Basf2 Variable Index

A | B | C | D | E | F | G | H | I | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

A

abs
acoplanarityAngle
ancestorHasWhichFlavor
angleToClosestInList
angleToMostB2BInList
aplanarity
ArmenterosDaughter1Qt
ArmenterosDaughter2Qt
ArmenterosLongitudinalMomentumAsymmetry
atcPIDBelle
averageValueInList
azimuthalAngleInDecayPlane

B

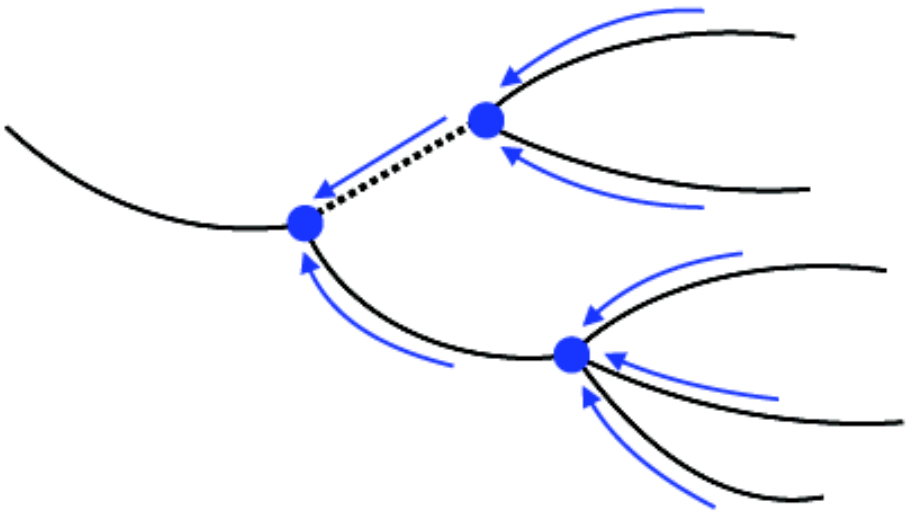
B0mcErrors
b2bClusterPhi
b2bClusterTheta
b2bPhi
b2bTheta
backwardHemisphereEnergy
backwardHemisphereMass
backwardHemisphereMomentum
backwardHemisphereX
backwardHemisphereY
backwardHemisphereZ
beamE
beamPx
beamPy
beamPz
belleECLEnergy
bssMassDifference
BtagToWBosonVariables

C

charge
chargeTimesKaonLiklihood

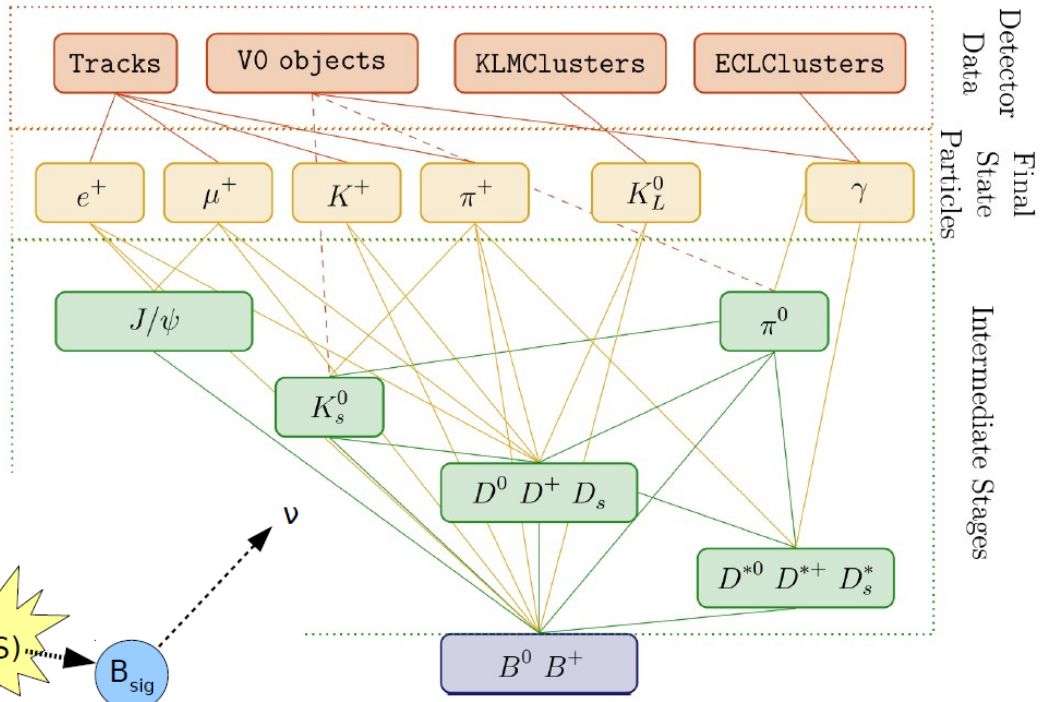
Vertex Fitting Tools

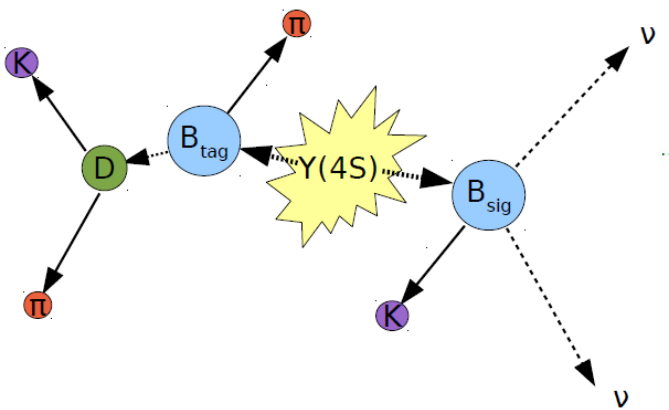
- KFit ported from Belle
- RAVE (no support by developers any more)
- TreeFitter
  - Adaption of decay chain fitter used by BaBar and LHCb
  - Speed up due to use of EIGEN library



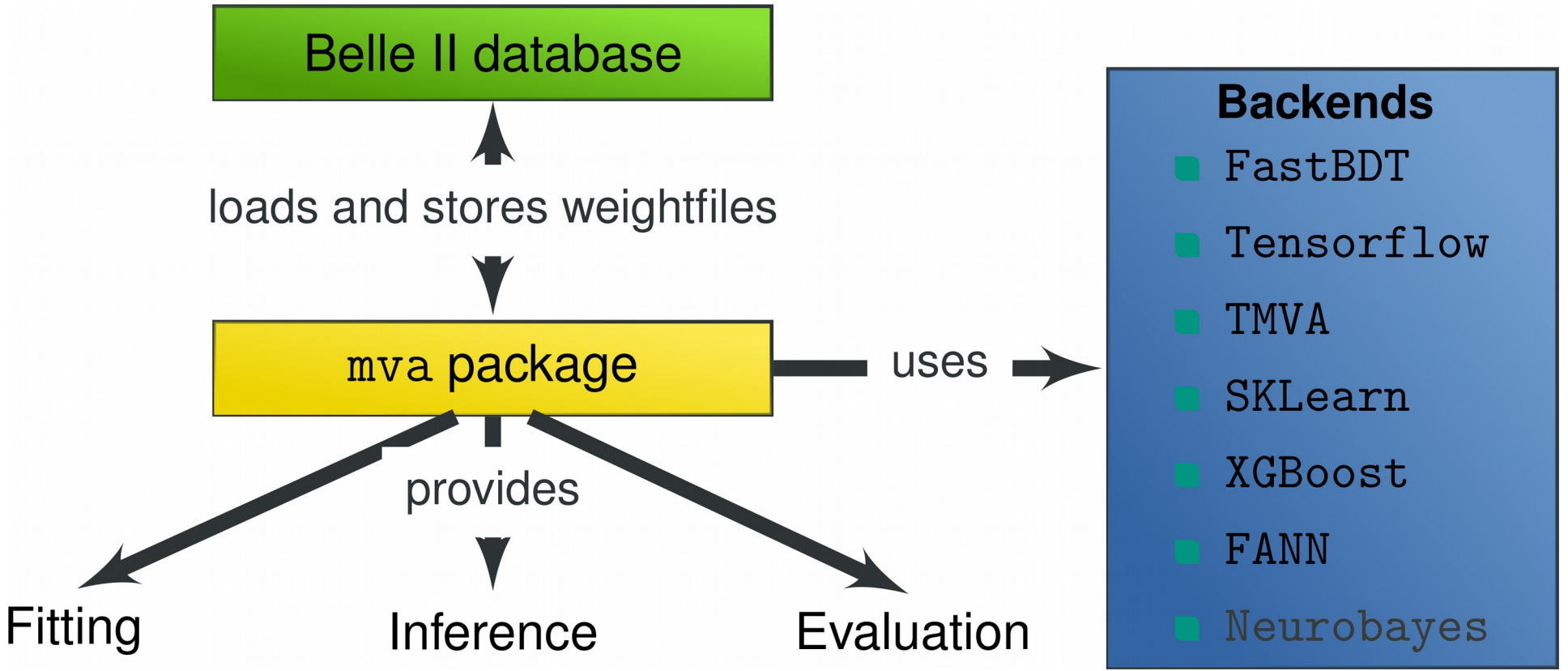
Full Event Interpretation (FEI)

- Full reconstruction of one B meson
- Momentum and charge of signal B meson known
- All remaining particles belong to signal B meson
- Reconstruction of decays with neutrinos

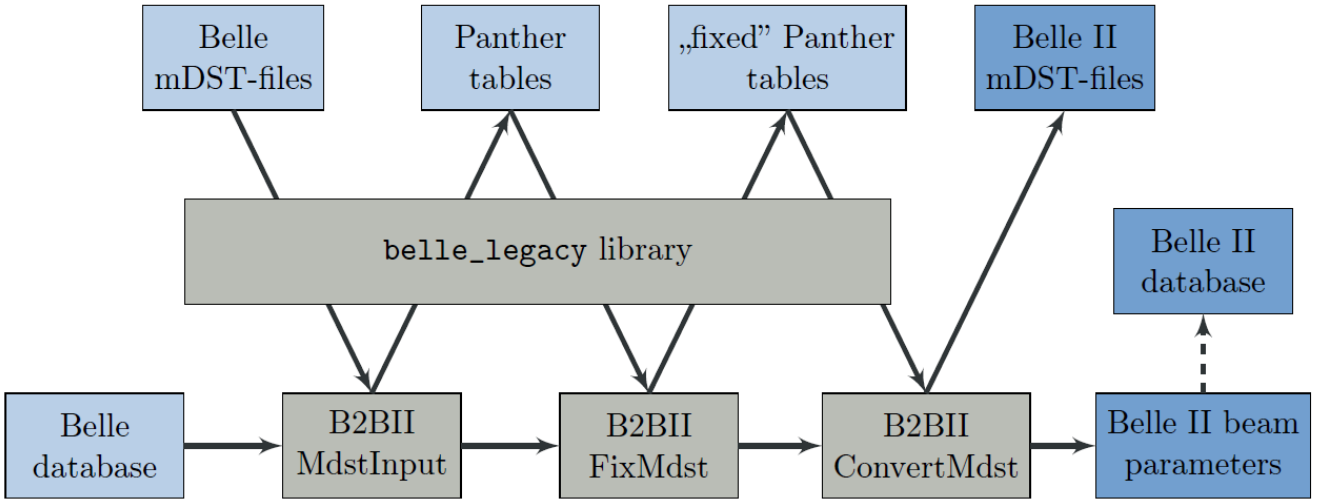


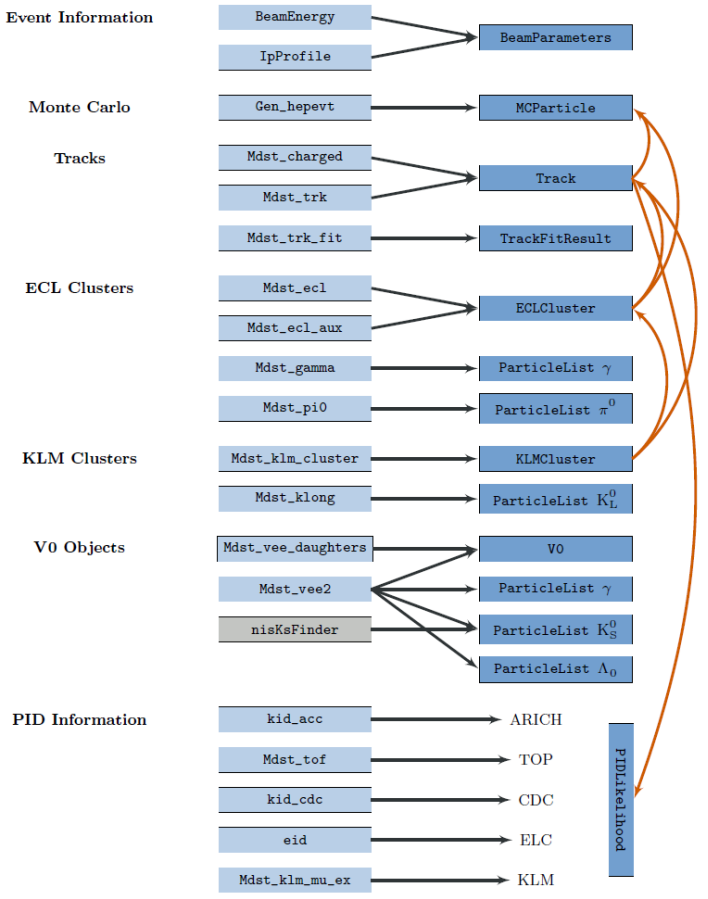


mva package

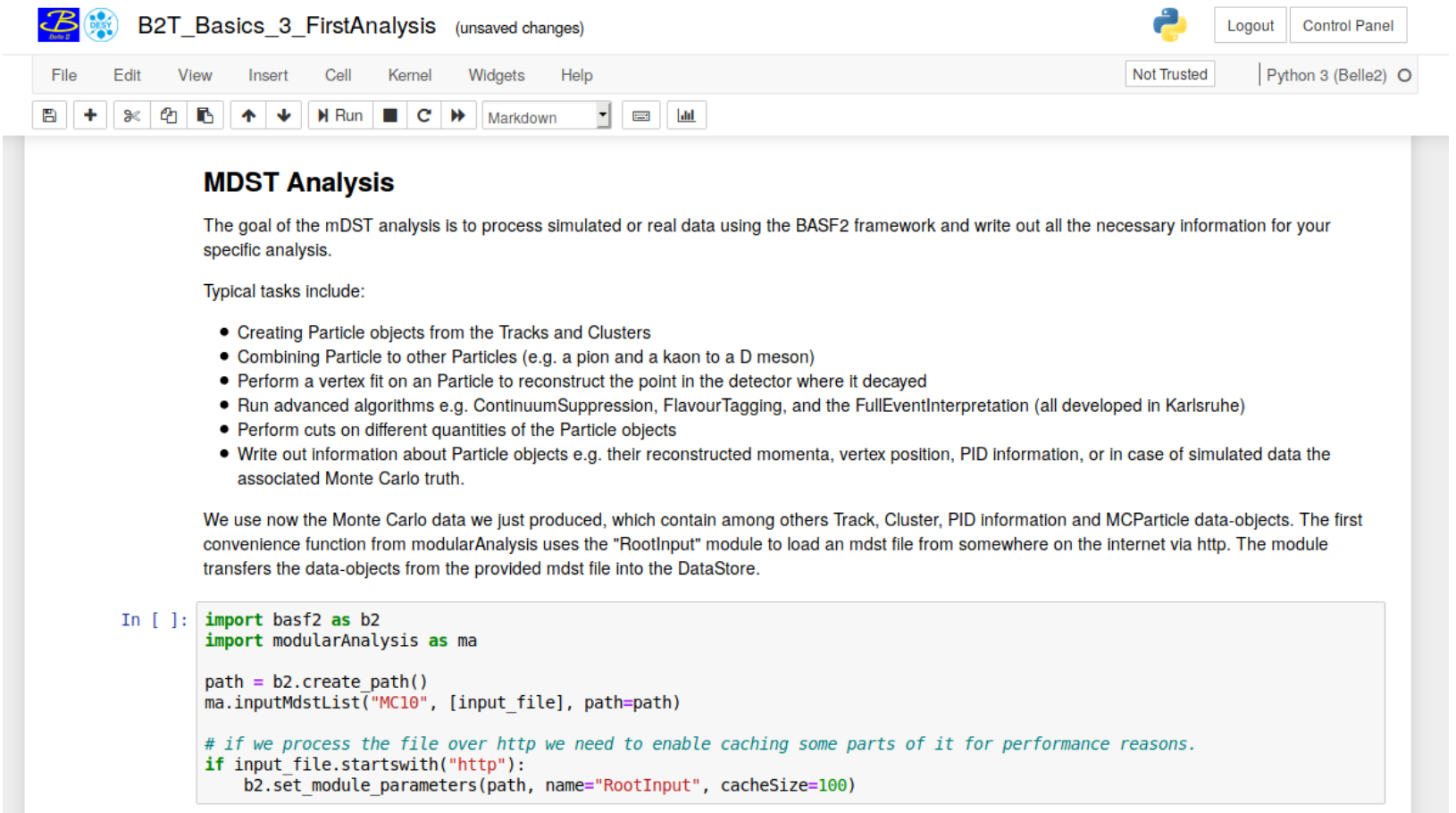

Belle II database
loads and stores weightfiles
mva package
uses
provides
Fitting
Inference
Evaluation
Backends
FastBDT
Tensorflow
TMVA
SKLearn
XGBoost
FANN
Neurobayes

Belle mdst input in basf2 (b2bii)





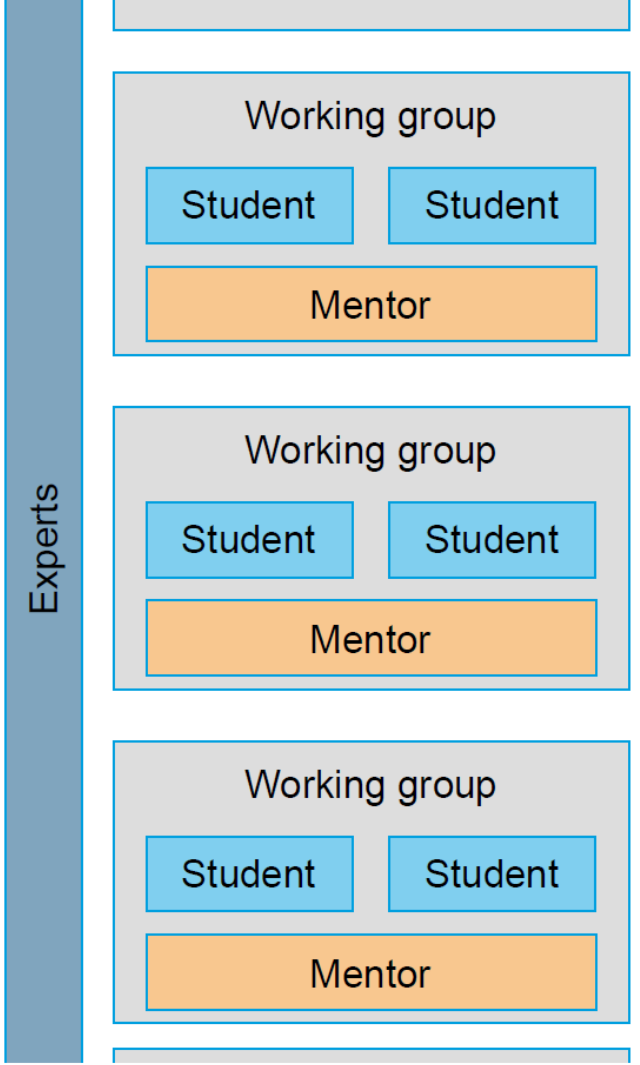
Jupyter Notebooks



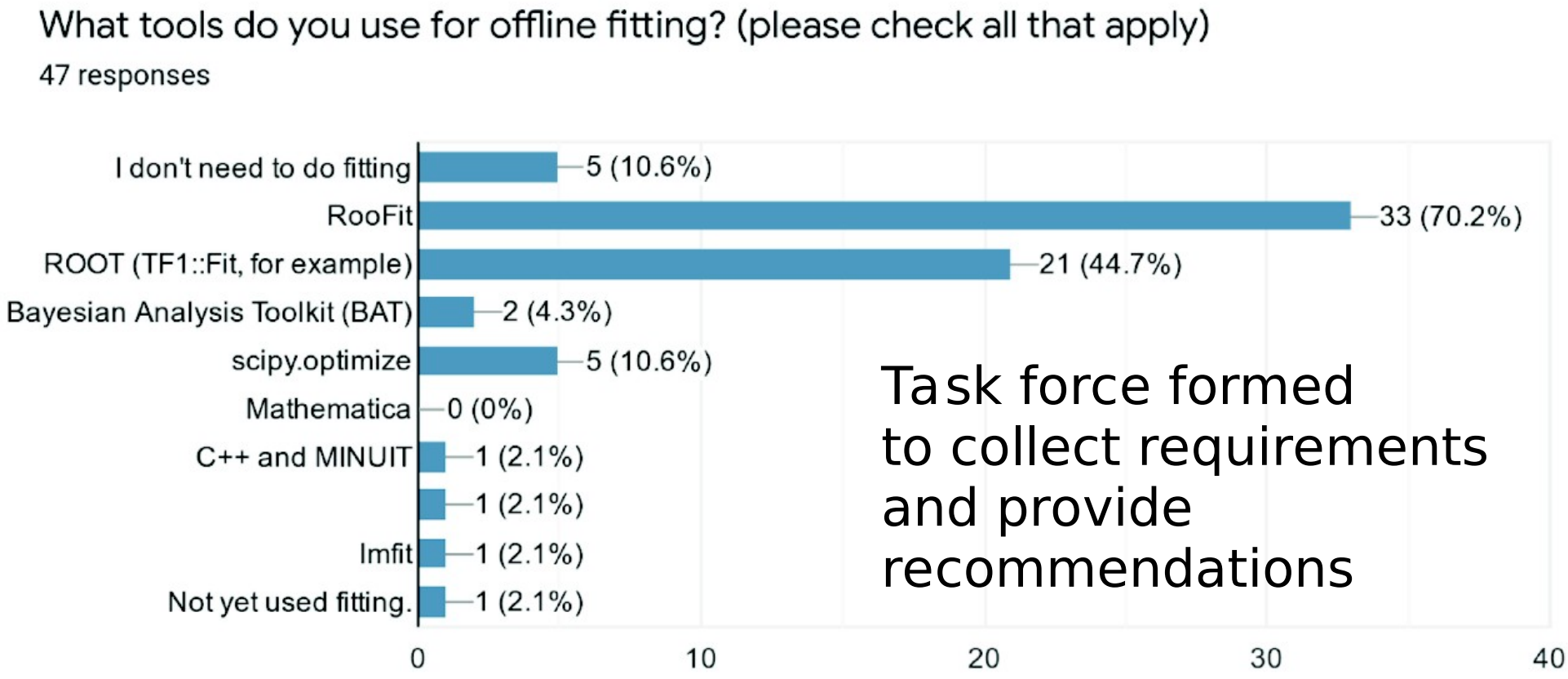
StarterKit

1st B2StarterKIt workshop, October 2017

	Day 1	Day 2	Day 3
Morning	Introduction to the software	“Expert” lectures	
Afternoon	Hands-on tutorial	Work in groups	



Fitting Tools



Task force formed to collect requirements and provide recommendations

Further Topics Under Discussion

- Systematics framework
- Analysis preservation
- ...

Summary

- Modular analysis and variable manager
  have simplified and standardized many analysis steps
- High level (B factory specific) analysis tools
  are implemented and well integrated
- Trends in ntuple-level analysis:
  - Python and industry standard tools
  - Browser based (jupyter notebooks)
  - Machine learning
- Some unresolved issues (systematics, preservation)