LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Machine Learning at Belle II

James Kahn

Ludwig Maximilians Universität München
DFG cluster of excellence
"Origin and Structure of the Universe"

April 2018

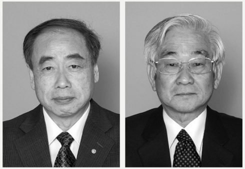
Bundesministerium
für Bildung
und Forschung

Outline

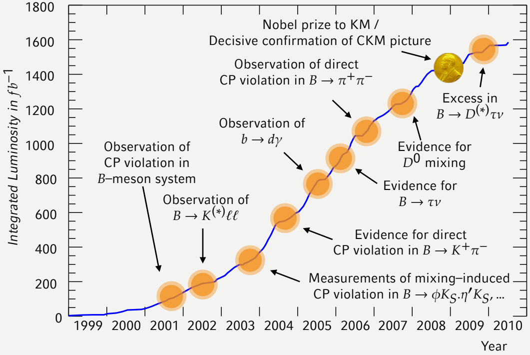
- The Belle II experiment
  - History
  - B-factories
  - Expectations
- ML in analysis
  - Continuum suppression
  - Full event reconstruction
- ML in hardware
  - Z-vertex trigger
- Challenges

The Belle II Experiment

- Collaboration formed in 2009 following success of Belle experiment:
  - Confirmed Kobayashi–Maskawa–mechanism (Nobel prize 2008).
  - UT parameters, heavy flavour spectroscopy, CPV, rare B decays, etc.



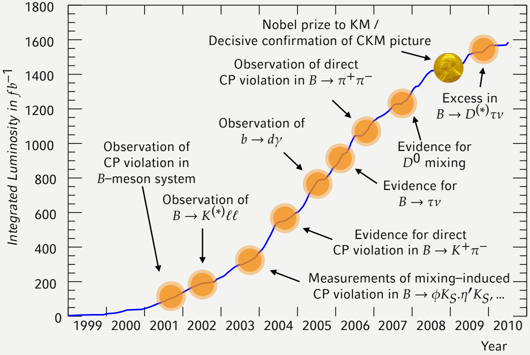
Makoto Kobayashi Toshihide Maskawa



The Belle II Experiment

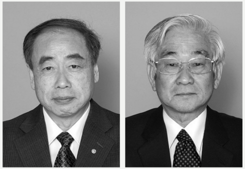
- Collaboration formed in 2009 following success of Belle experiment:
  - Confirmed Kobayashi–Maskawa–mechanism (Nobel prize 2008).
  - UT parameters, heavy flavour spectroscopy, CPV, rare B decays, etc.
- New physics searches (Sources of CPV, (semi–)leptonic decay, LFV, etc.)

Makoto Kobayashi Toshihide Maskawa

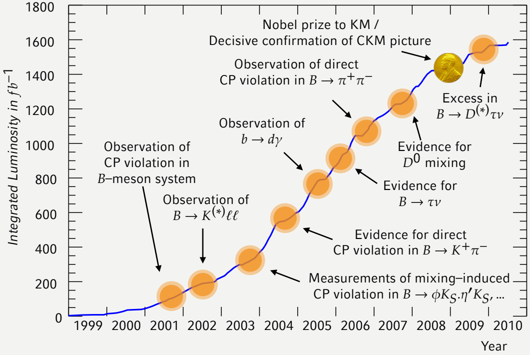


The Belle II Experiment

- Collaboration formed in 2009 following success of Belle experiment:
  - Confirmed Kobayashi–Maskawa–mechanism (Nobel prize 2008).
  - UT parameters, heavy flavour spectroscopy, CPV, rare B decays, etc.
- New physics searches (Sources of CPV, (semi–)leptonic decay, LFV, etc.)
- Still tensions in SM and big unanswered questions



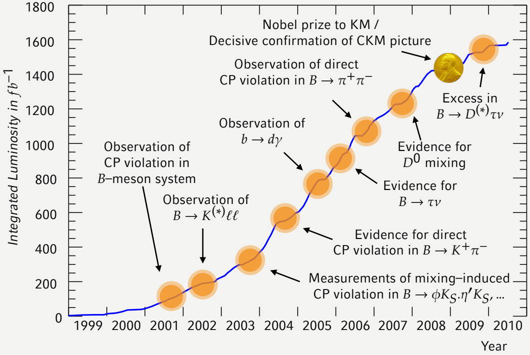
Makoto Kobayashi Toshihide Maskawa



The Belle II Experiment

- Collaboration formed in 2009 following success of Belle experiment:
  - Confirmed Kobayashi–Maskawa–mechanism (Nobel prize 2008).
  - UT parameters, heavy flavour spectroscopy, CPV, rare B decays, etc.
- New physics searches (Sources of CPV, (semi–)leptonic decay, LFV, etc.)
- Still tensions in SM and big unanswered questions
- Current standing:
  - 700+ Members, 106 institutes, 25 countries
  - First calibration data: 2018
    First analysis data: 2019

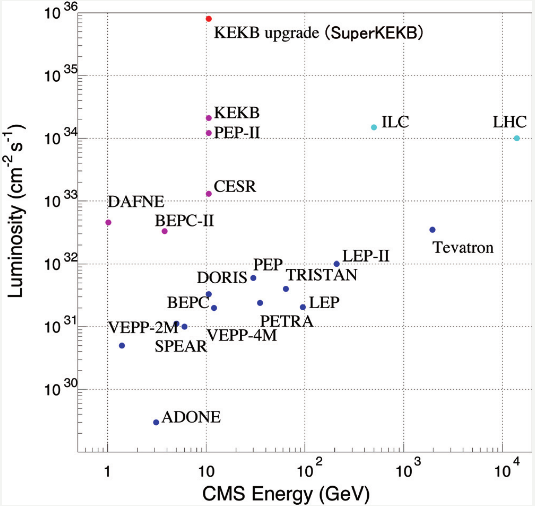
Makoto Kobayashi Toshihide Maskawa



B-Factory

Two approaches for HEP:

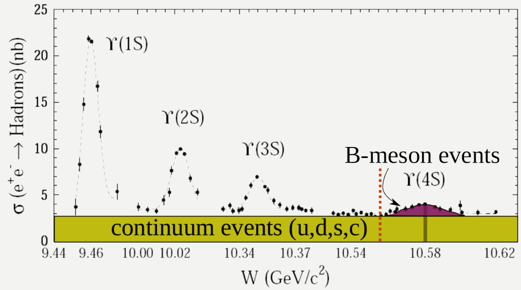
1. Energy frontier (direct search):
   - Powerful enough to produce new particles directly. (LHC)
2. Precision frontier (indirect):
   - Focus on specific energy range. (Belle II)



B-Factory

Two approaches for HEP:

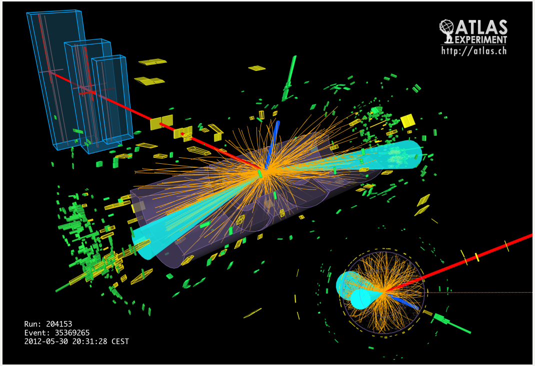
1. Energy frontier (direct search):
   - Powerful enough to produce new particles directly. (LHC)
2. Precision frontier (indirect):
   - Focus on specific energy range. (Belle II)



B-Factory

Two approaches for HEP:

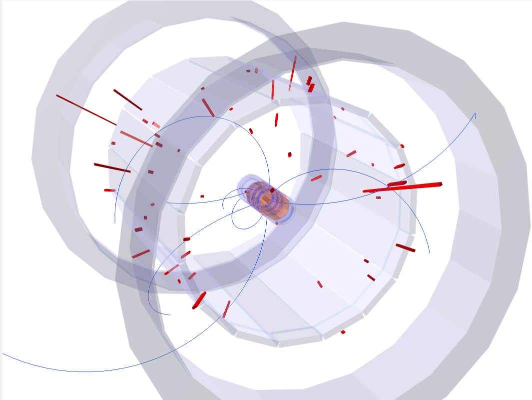
1. Energy frontier (direct search):
   - Powerful enough to produce new particles directly. (LHC)
2. Precision frontier (indirect):
   - Focus on specific energy range. (Belle II)



B-Factory

Two approaches for HEP:

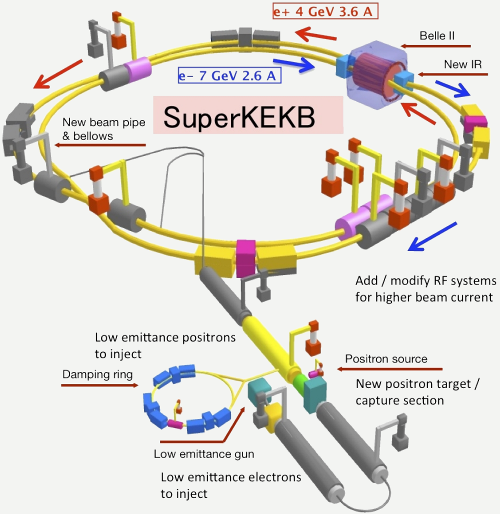
1. Energy frontier (direct search):
   - Powerful enough to produce new particles directly. (LHC)
2. Precision frontier (indirect):
   - Focus on specific energy range. (Belle II)



Unique advantages to B-factory:

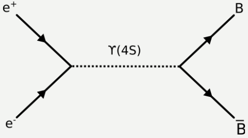
- Clean environment from e^+e^- - no partons
- Precise knowledge of initial energy
- Ability to directly measure branching fractions
- Missing energy searches

SuperKEKB

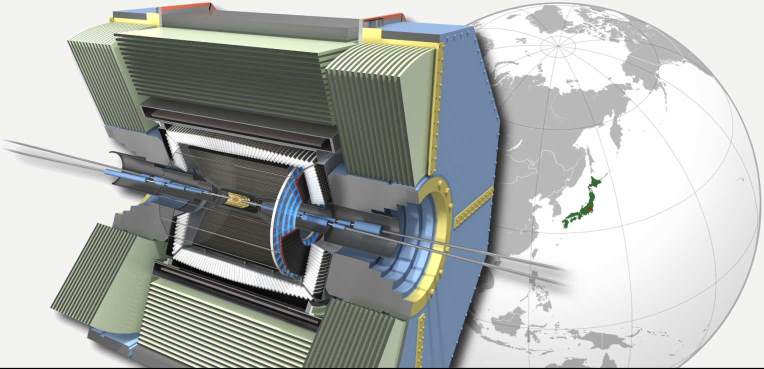

e+ 4 GeV 3.6 A
Belle II
New IR
e− 7 GeV 2.6 A
SuperKEKB
New beam pipe
& bellows
Add / modify RF systems
for higher beam current
Low emittance positrons
to inject
Positron source
Damping ring
New positron target /
capture section
Low emittance gun
Low emittance electrons
to inject

The Belle II Experiment

Asymmetric e^+e^- experiment mainly at the $\Upsilon(4S)$ resonance (10.58 GeV)

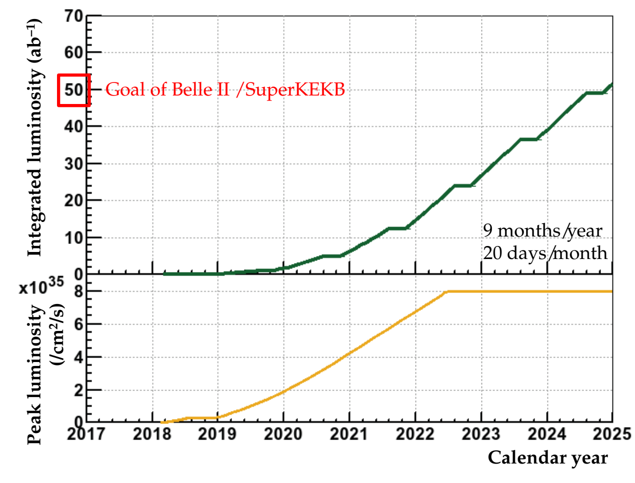


Focus on B, charm and τ physics



	KEKB/Belle	SuperKEKB/Belle II
operation	1999–2010	2018–
peak luminosity	$2.11 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$	$8 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$
integrated luminosity	$1\,\mathrm{ab}^{-1}$ (772 million $B\bar{B}$ pairs)	$50\,\mathrm{ab}^{-1}$

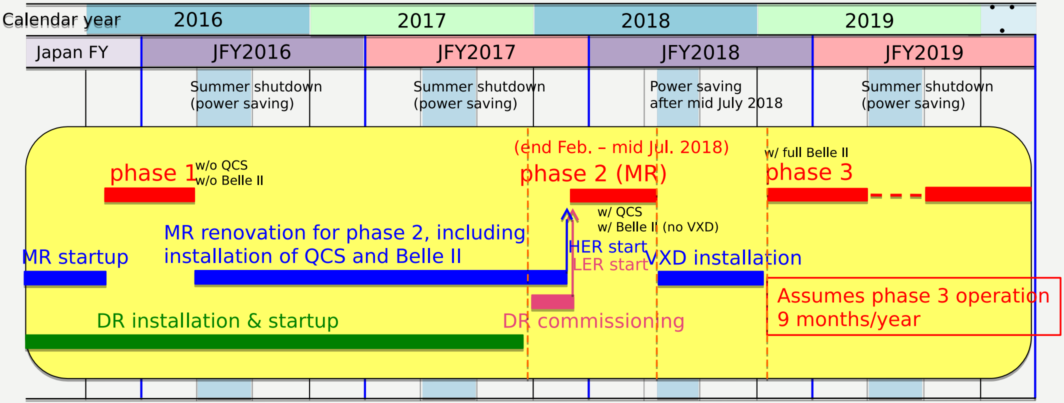
Timeline



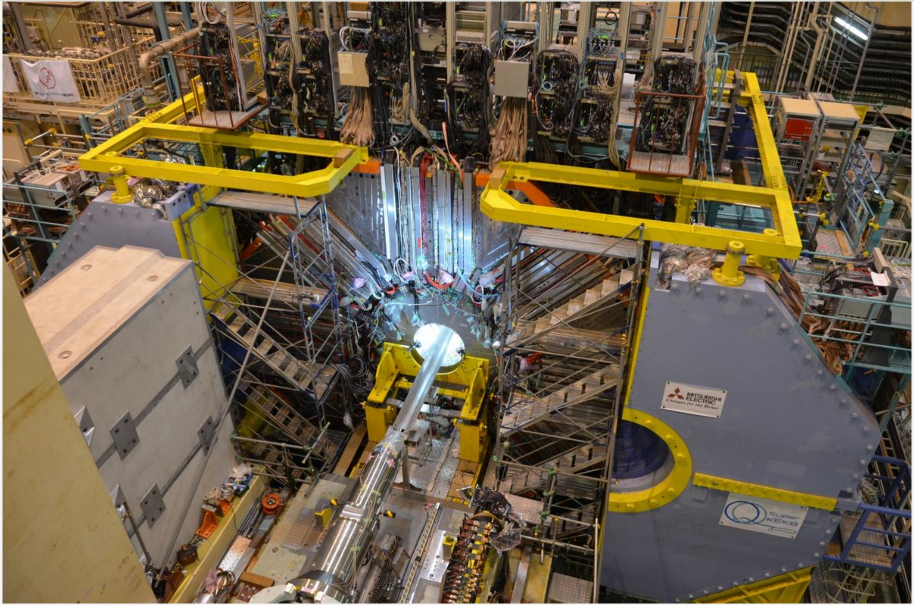
- Peak instantaneous luminosity: $8 \times 10^{35} cm^{-2} s^{-1}$ (Belle: $2.11 \times 10^{34} cm^{-2} s^{-1}$)
- Total integrated luminosity: $50 ab^{-1}$ (Belle: $1 ab^{-1}$)

Process	$\sigma[nb]$	No. events [$\times 10^9$]
$B\bar{B}$	1.1	55
$q\bar{q}$	2.52	185.45
$\tau^+\tau^-$	0.92	45.95

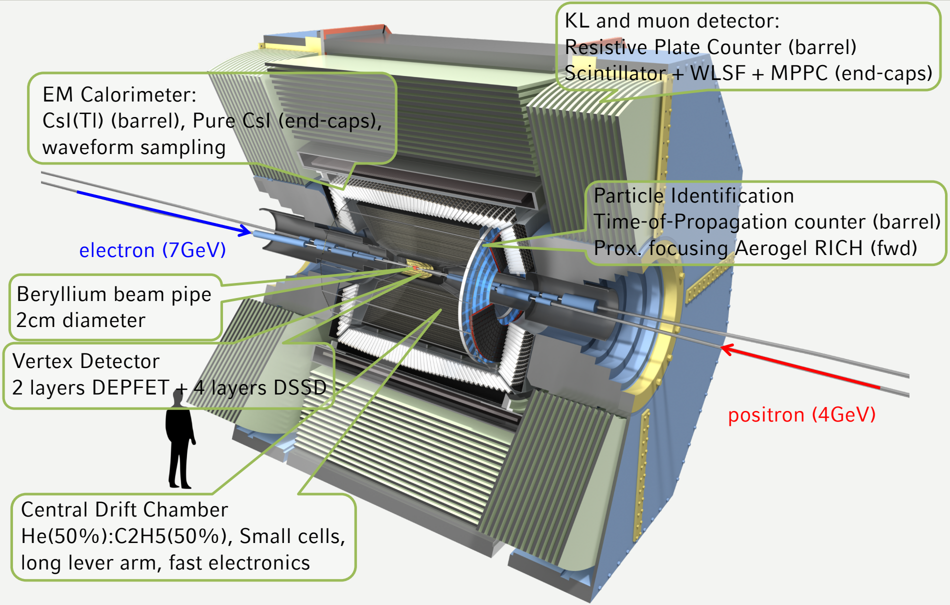
Timeline


Calendar year
2016
2017
2018
2019
Japan FY
JFY2016
JFY2017
JFY2018
JFY2019
Summer shutdown
(power saving)
Summer shutdown
(power saving)
Power saving
after mid July 2018
Summer shutdown
(power saving)
(end Feb. – mid Jul. 2018)
w/ full Belle II
phase 1
w/o QCS
w/o Belle II
phase 2 (MR)
phase 3
w/ QCS
w/ Belle II (no VXD)
MR renovation for phase 2, including
installation of QCS and Belle II
MR startup
HER start
LER start
VXD installation
Assumes phase 3 operation
9 months/year
DR installation & startup
DR commissioning

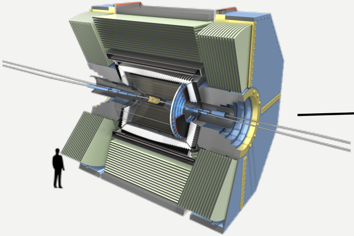
Belle II Detector



Belle II Detector



Data Flow



DAQ

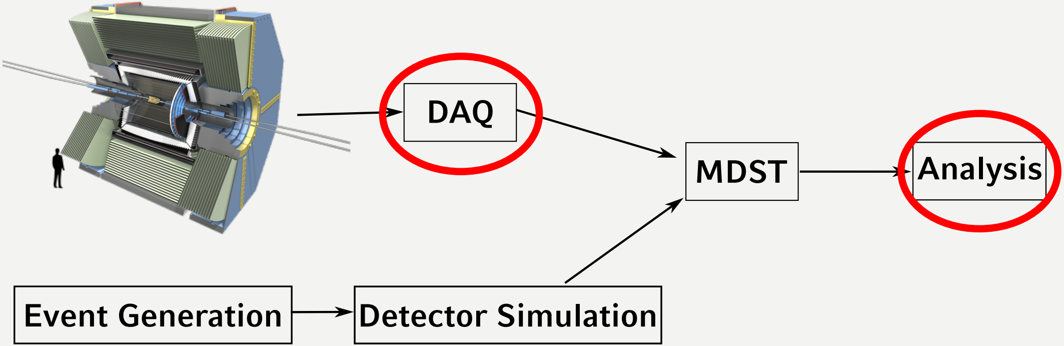
MDST

Analysis

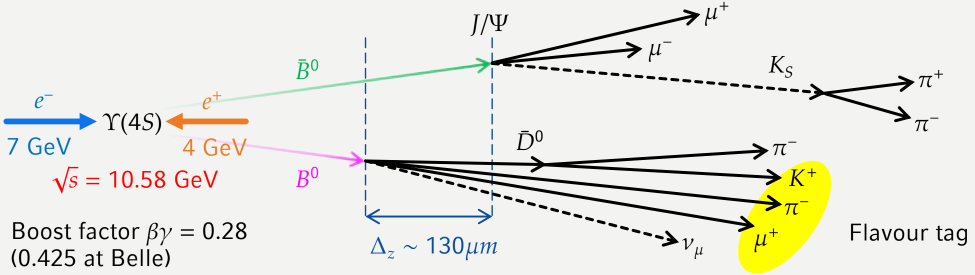
Event Generation

Detector Simulation

Data Flow



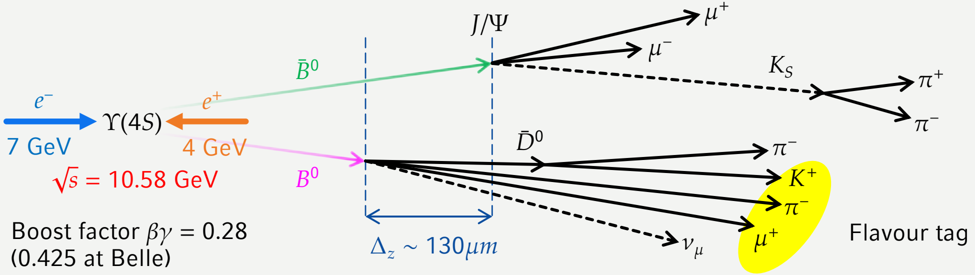
Machine Learning in Analysis



Traditional analysis procedure:

1. Reconstruct signal side
2. Cut-based selection
3. Continuum Suppression
4. Tag side reconstruction
5. Fitting

Machine Learning in Analysis



Boost factor $\beta\gamma = 0.28$
(0.425 at Belle)

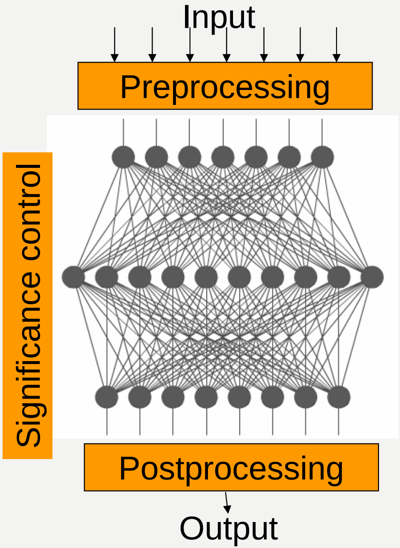
Traditional analysis procedure:

1. Reconstruct signal side
2. Cut-based selection
3. **Continuum Suppression**
4. **Tag side reconstruction**
5. Fitting

Machine Learning in Analysis

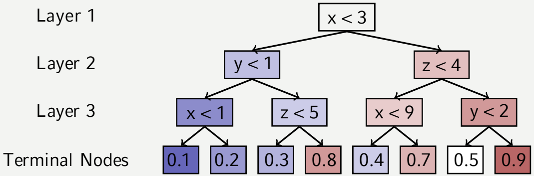
Standard toolset available to users:

- Neurobayes
  - Developed by phi-t at KIT
  - First use of neural networks
  - Deprecated



Standard toolset available to users:

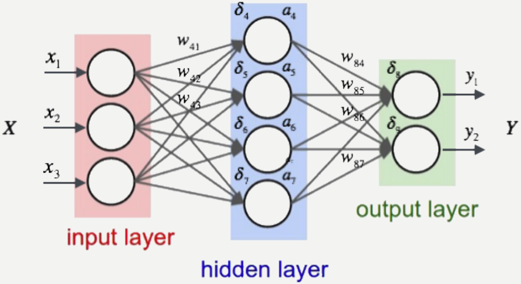
- Neurobayes
  - Developed by phi-t at KIT
  - First use of neural networks
  - Deprecated
- FastBDT
  - Modified boosted decision tree
  - Robust against overfitting



Machine Learning in Analysis

Standard toolset available to users:

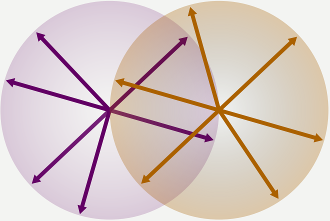
- Neurobayes
  - Developed by phi-t at KIT
  - First use of neural networks
  - Deprecated
- FastBDT
  - Modified boosted decision tree
  - Robust against overfitting
- Tensorflow/Keras
  - Newly implemented in Belle II
  - Integrated within software framework
  - Requires no extra dataset preparation



Continuum Suppression

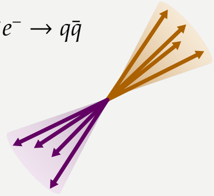
What is Continuum?

$e^+e^- \to \Upsilon(4S) \to B\bar{B}$

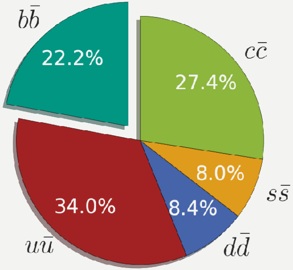


Spherical

$e^+e^- \to q\bar{q}$



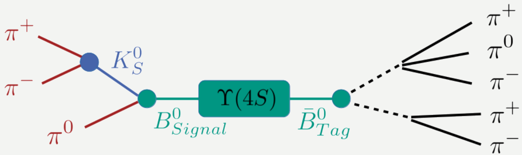
Jet-like



- Continuum biggest background in many analyses
- Utilise kinematics to suppress
- Specialised tool and variables developed

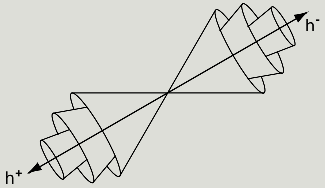
Continuum Suppression

Established Method



1. Reconstruct B^0_{Signal} candidate
2. Collect remaining event
3. Construct high level variables
4. Belle: Neurobayes
   Belle II: FastBDT

CleoCones



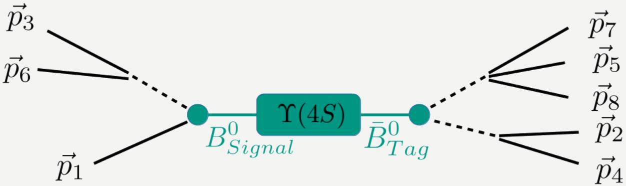
KSFW Variables

- Fox Wolfram moments
- Calculated on primary daughters

Thrust axis

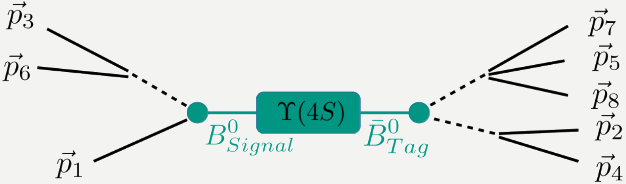
Compare B_{Signal} to B_{Tag} flight directions

New Approach



- Use direct detector information
  - Tracks in drift chamber
  - Energy deposits in ECL
- Information on 10 highest momenta:
  - Momentum vector
  - Particle ID
  - Uncertainties
- 200 low level variables

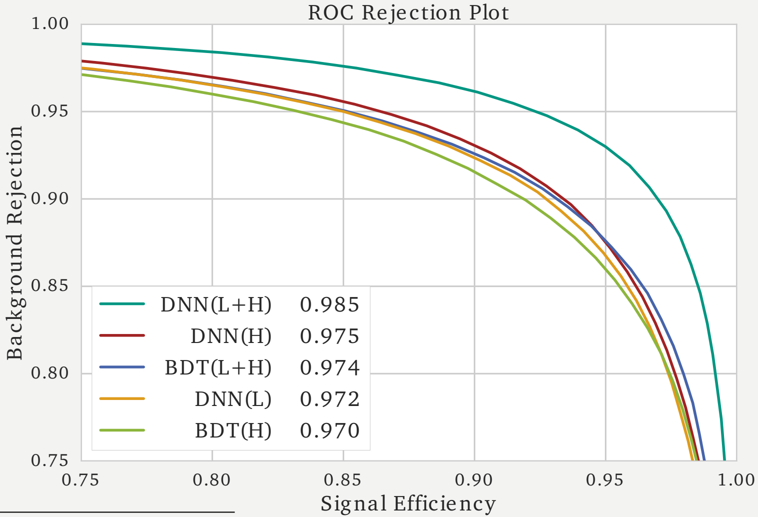
New Approach



- Use direct detector information
  - Tracks in drift chamber
  - Energy deposits in ECL
- Information on 10 highest momenta:
  - Momentum vector
  - Particle ID
  - Uncertainties
- 200 low level variables

- Train on $1M$ MC events:
  - 50% $B \to K^0_S \pi^0$
  - 50% continuum
- 10% for validation
- 3 different set of variables for training:
  - H: High Level variables from the established method
  - L: Low Level variables from the new approach
  - L+H: combination of the 2 sets
- 4 hidden layers: 180, 120, 60, 30

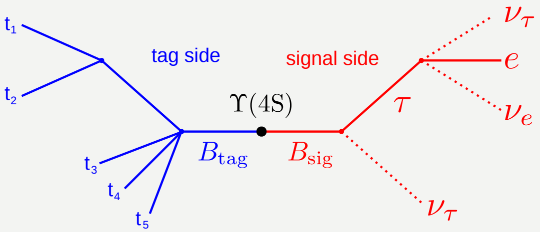
Results



[1] D. Weyland, "Continuum Suppression with Deep Learning techniques for the Belle II Experiment", MA thesis (KIT, Karlsruhe, ETP, 2017-11-02)

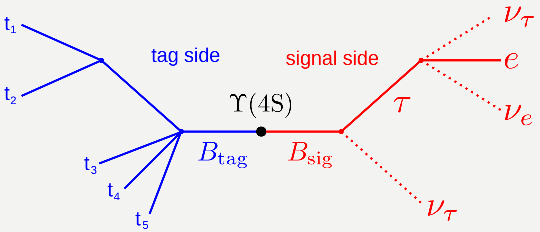
Full Reconstruction (BELLE)

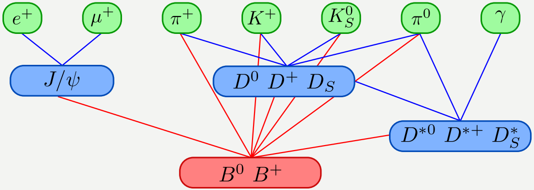
- Unique ability of B-factories
- Necessary for missing energy searches
  - $B \to K\nu\nu$
  - $B \to D^{**}\ell\nu$
  - $B \to \tau\nu$



Full Reconstruction (BELLE)

- Unique ability of B-factories
- Necessary for missing energy searches
  - $B \to K\nu\nu$
  - $B \to D^{**}\ell\nu$
  - $B \to \tau\nu$
- Hierarchical reconstruction (Neurobayes)





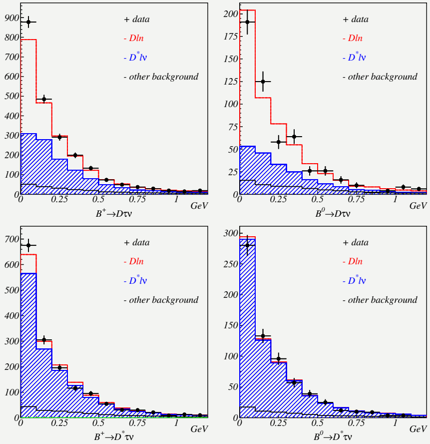
Full Reconstruction (BELLE)

- Unique ability of B-factories
- Necessary for missing energy searches
  - $B \to K\nu\nu$
  - $B \to D^{**}\ell\nu$
  - $B \to \tau\nu$
- Hierarchical reconstruction (Neurobayes)
- Used in $R(D^*)$ measurement hadronic tag:

$$\mathscr{R}(D^{(*)}) = \frac{\mathscr{B}(\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau)}{\mathscr{B}(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell)}$$

- Fit to remaining energy in detector:

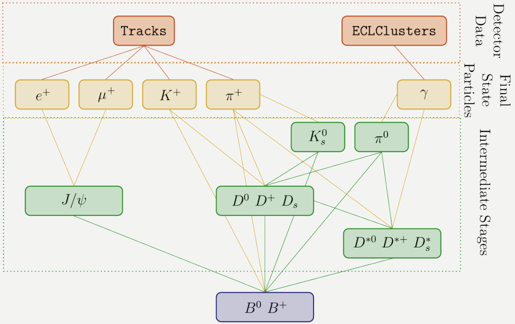
$$E_{ECL} = E_{tot} - E_{rec}$$



[1] I. Adachi et al., “Measurement of B —> D(*) tau nu using full reconstruction tags”, in Proceedings, 24th International Symposium on Lepton-Photon Interactions at High Energy (LP09): Hamburg, Germany, August 17-22, 2009 (2009)

Full Event Interpretation

- New in Belle II
- Same hierarchical reconstruction method

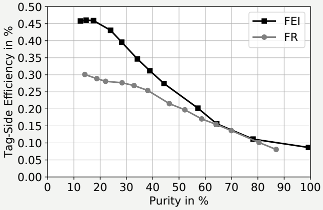


[1] T. Keck, *The Full Event Interpretation*, personal communication

Full Event Interpretation

- New in Belle II
- Same hierarchical reconstruction method
- Use FastBDT, added many more decay modes, improved efficiency

Figure: $B+$



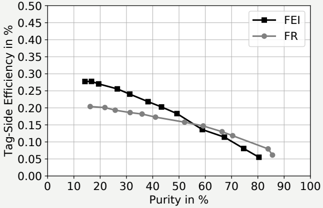


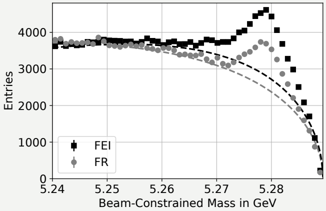
Figure: $B0$

[1] T. Keck, *The Full Event Interpretation*, personal communication

Full Event Interpretation

- New in Belle II
- Same hierarchical reconstruction method
- Use FastBDT, added many more decay modes, improved efficiency

Figure: $B+$



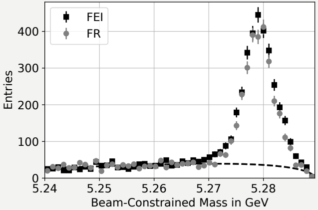
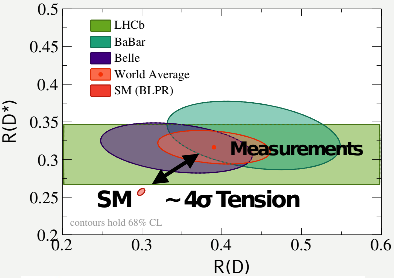


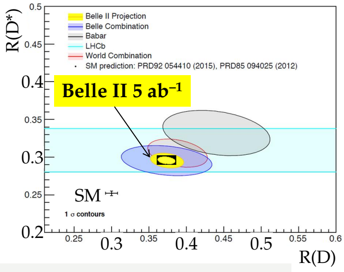
Figure: $B0$

[1] T. Keck, *The Full Event Interpretation*, personal communication

Full Event Interpretation

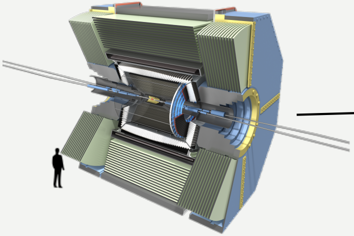
- New in Belle II
- Same hierarchical reconstruction method
- Use FastBDT, added many more decay modes, improved efficiency
- Repeat previous measurements with improved results





[1] T. Keck, *The Full Event Interpretation*, personal communication

ML in Hardware



DAQ

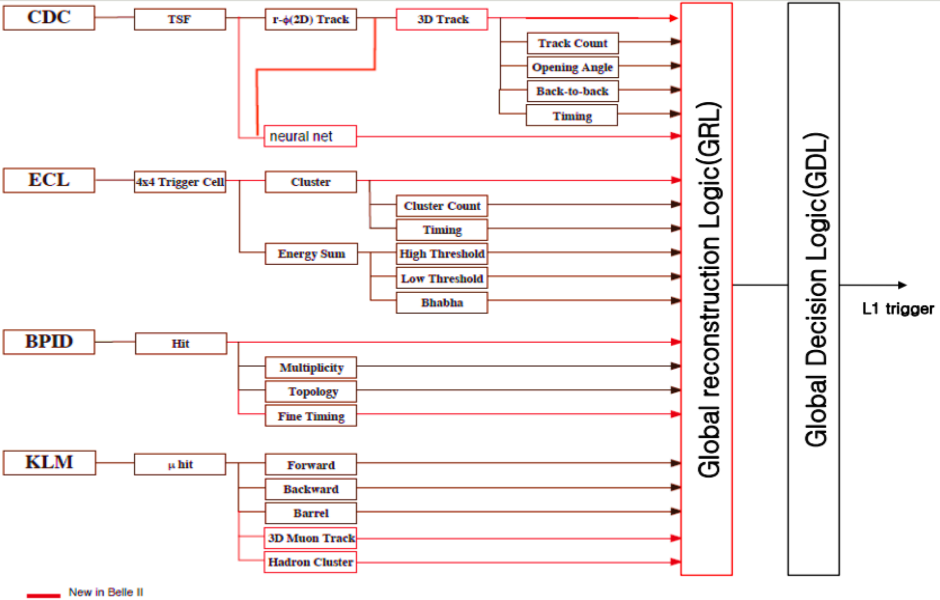
MDST

Analysis

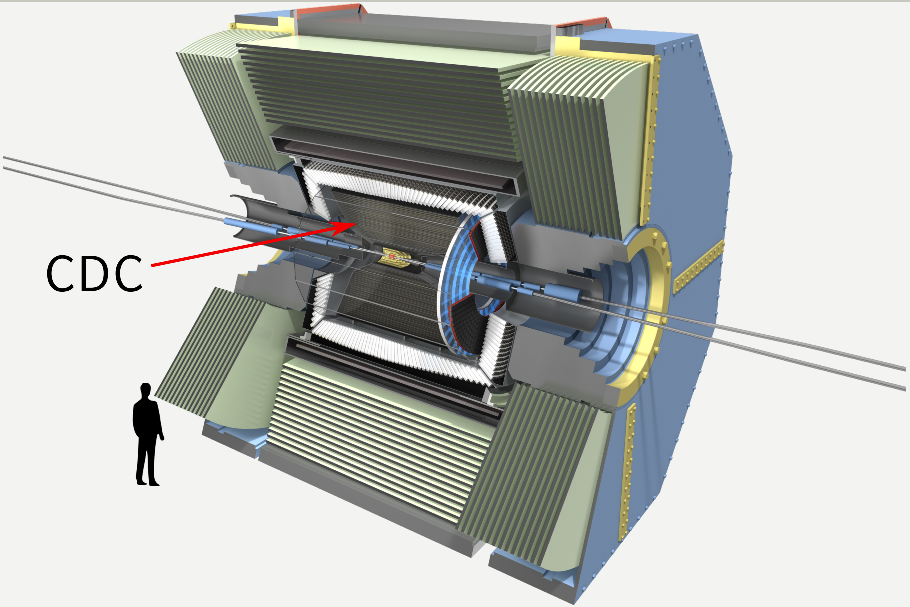
Event Generation

Detector Simulation

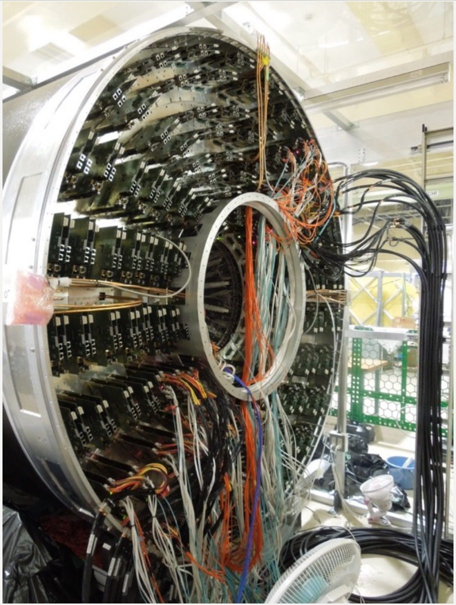
ML in Hardware



Z-Vertex Trigger



Z-Vertex Trigger



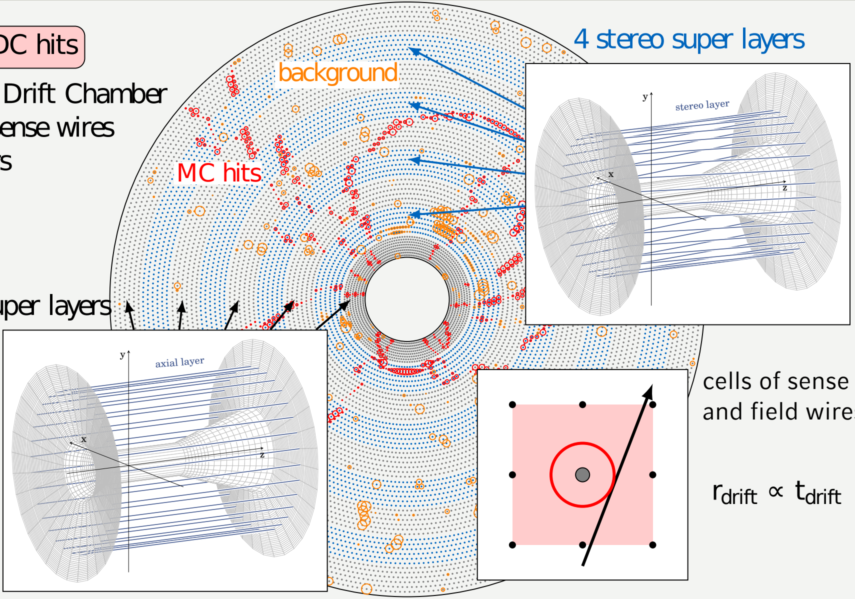
Z-Vertex Trigger

input: CDC hits

Central Drift Chamber
14336 sense wires
56 layers

4 stereo super layers

5 axial super layers

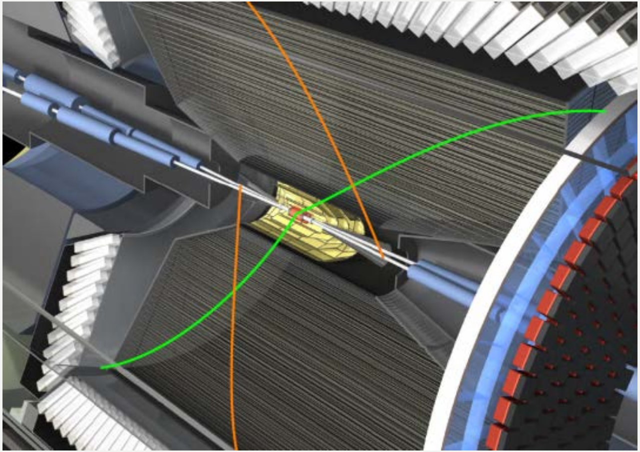


cells of sense wires and field wires

$r_{\mathrm{drift}} \propto t_{\mathrm{drift}}$

Z-Vertex Trigger

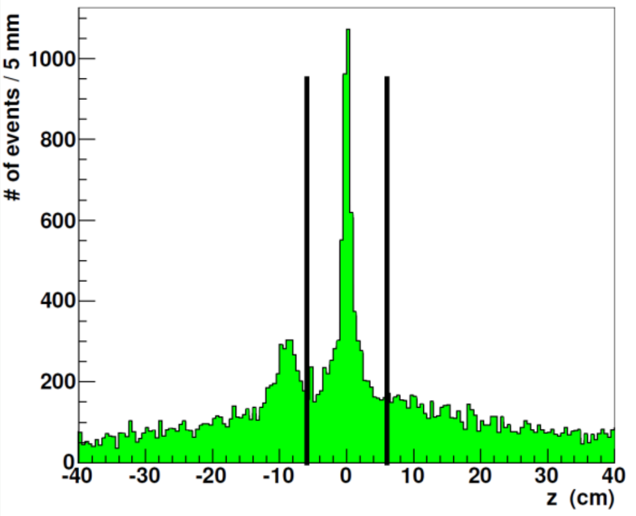
- Need to suppress non-event backgrounds:
  - Intra-beam interactions (Touscheck scattering)
  - Scattering from residual beampipe gas



Z-Vertex Trigger

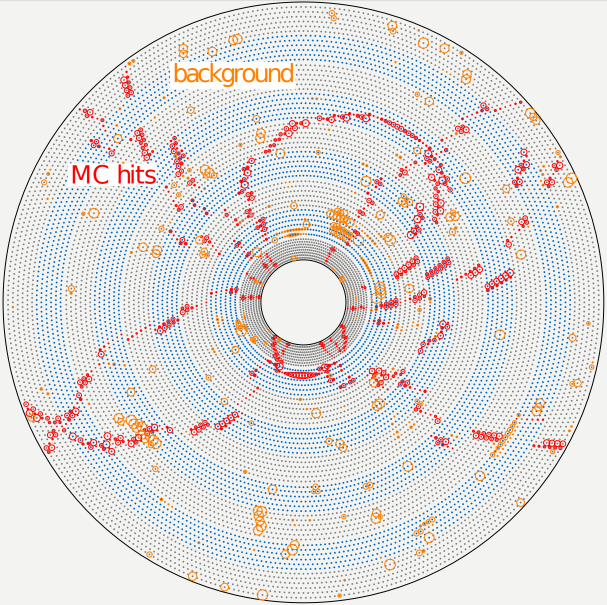
- Need to suppress non-event backgrounds:
  - Intra-beam interactions (Touscheck scattering)
  - Scattering from residual beampipe gas

- Want to discriminate between real events and machine background



Z-Vertex Trigger

- Need to suppress non-event backgrounds:
  - Intra-beam interactions (Touscheck scattering)
  - Scattering from residual beampipe gas
- Want to discriminate between real events and machine background
- First step: Track segment finder



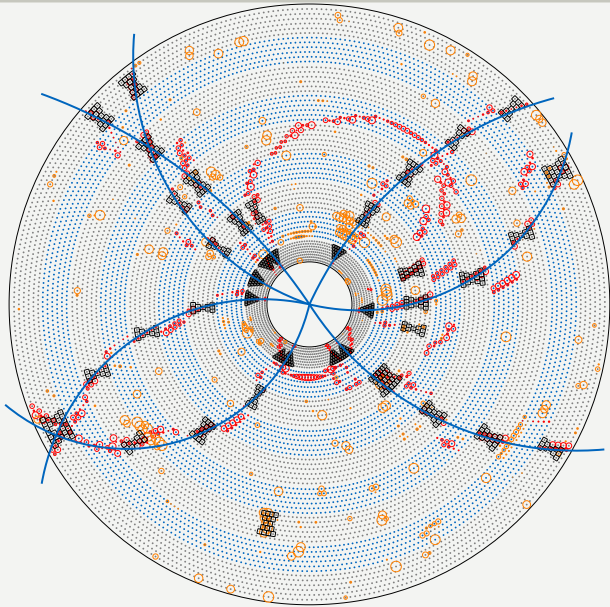
Z-Vertex Trigger

- Need to suppress non-event backgrounds:
  - Intra-beam interactions (Touscheck scattering)
  - Scattering from residual beampipe gas

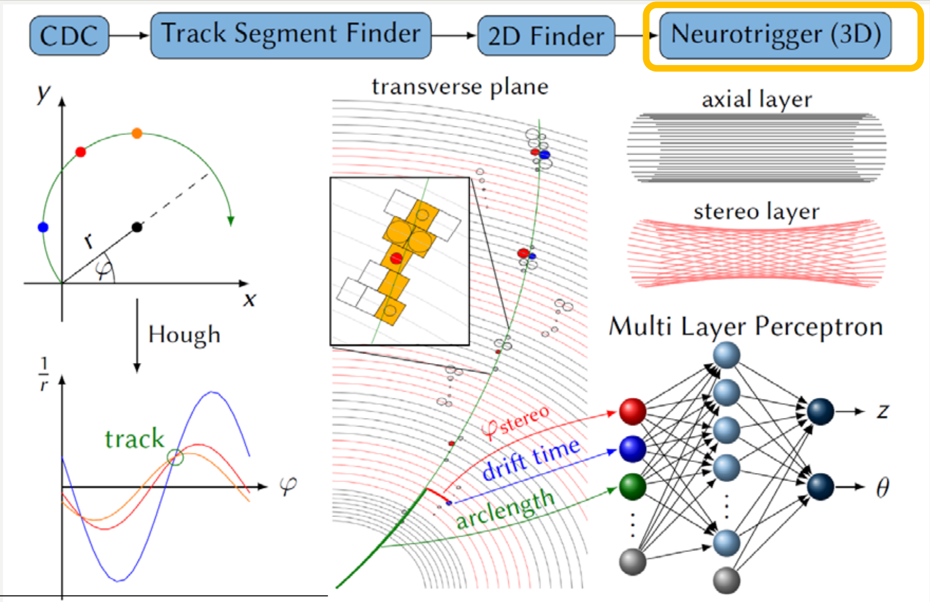
- Want to discriminate between real events and machine background

- First step: Track segment finder

- Second step: 2D track finder

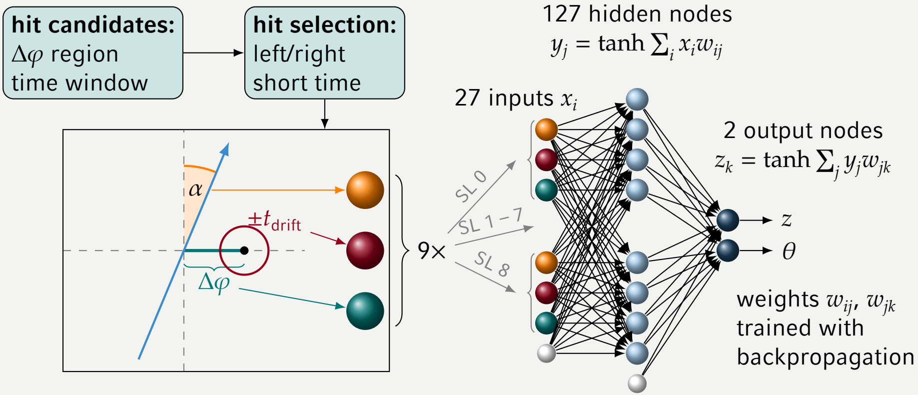


Z-Vertex Trigger



[1]S. Pohl, *The Belle II z-Vertex Trigger*, sneuhaus@mpp.mpg.de

Z-Vertex Trigger

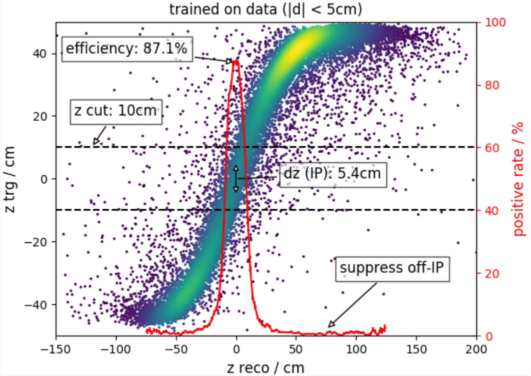


- Crossing angle α: track curvature
- Missing axial hit: default inputs (0,0,0)
- Missing stereo hit: expert network

[1] S. Pohl, *The Belle II z-Vertex Trigger*, *sneuhaus@mpp.mpg.de*

Z-Vertex Trigger

- Currently trained on MC
- Single hidden layer feed-forward network
- First tests performed on cosmic ray backgrounds
- Achieved resolution of $\mathcal{O}(1cm)$
- Mid 2018 calibration data collected



Challenges

- Strong understanding of systematics essential in HEP
- Begins with accurately representative simulations

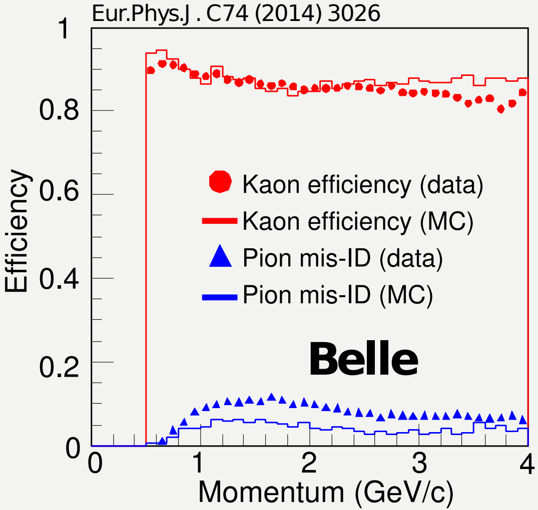


Figure: Kaon identification, $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$

Challenges

- Strong understanding of systematics essential in HEP
- Begins with accurately representative simulations
- First step is recalibration according to detector response

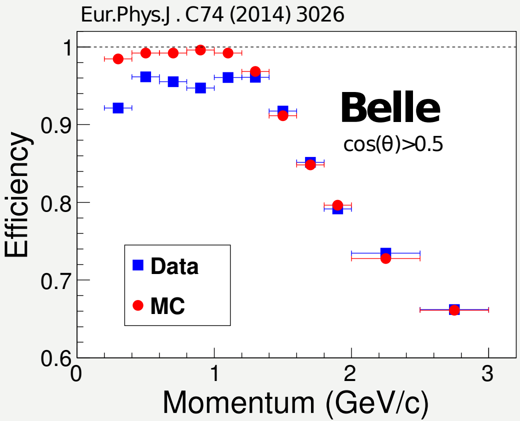


Figure: Proton identification, $\Lambda \to p\pi^-$

Challenges

- Strong understanding of systematics essential in HEP
- Begins with accurately representative simulations
- First step is recalibration according to detector response
- Currently no collision data → phase 2

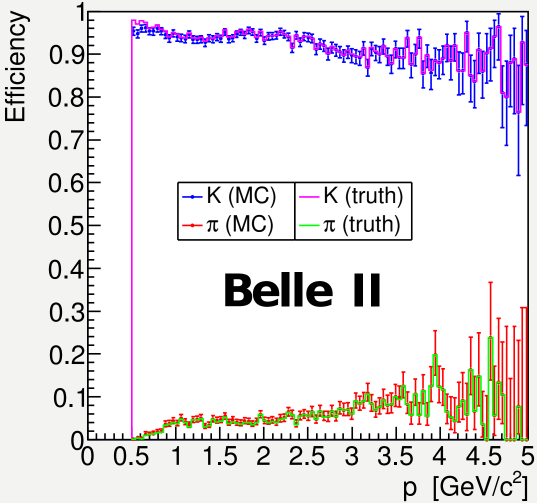
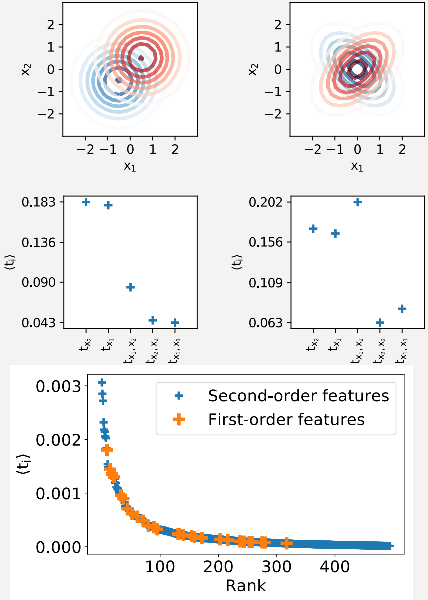


Figure: Kaon identification

Challenges

- Strong understanding of systematics essential in HEP
- Begins with accurately representative simulations
- First step is recalibration according to detector response
- Currently no collision data → phase 2
- First attempt at systematics estimation submitted (Preprint: arXiv:1803.08782) (`github.com/stwunsch/tensorflow_derivative`)

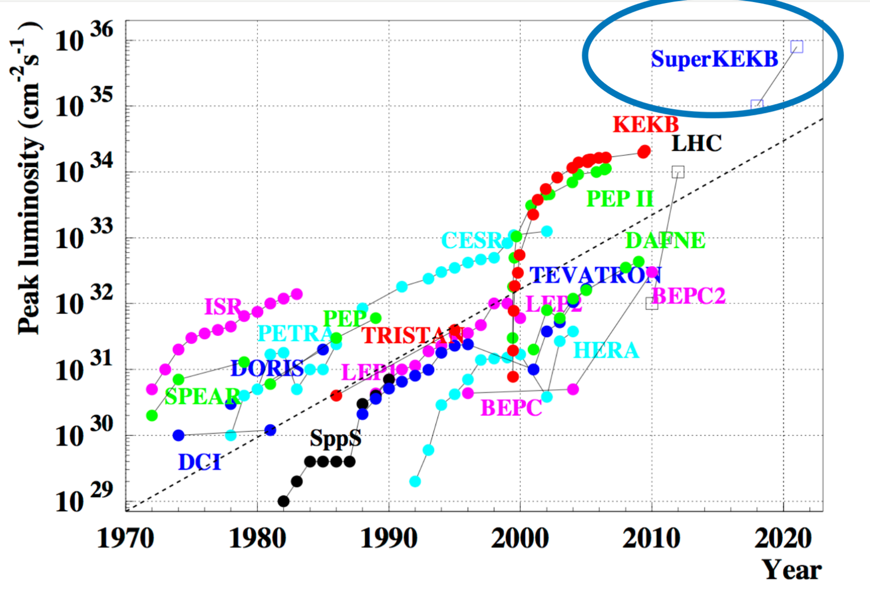


Summary

- Belle II beginning operation this year - calibration
- Data taking to begin in 2019
- Machine learning implemented in collaboration-wide tools
- BDTs already used
- Neural networks beginning to be implemented throughout
- Investigation into systematics propagation ongoing

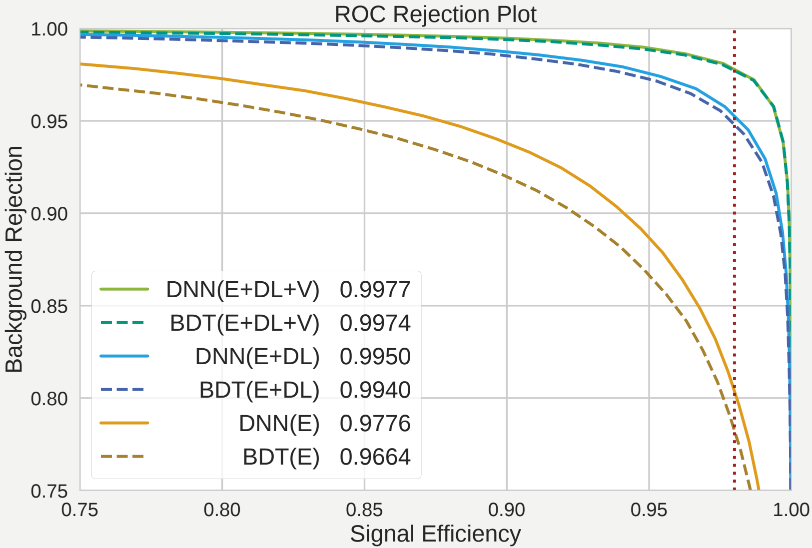
BACKUP

B-Factory



Peak luminosity ($cm^{-2}s^{-1}$)
10^{36}
10^{35}
10^{34}
10^{33}
10^{32}
10^{31}
10^{30}
10^{29}
1970
1980
1990
2000
2010
2020
Year
SuperKEKB
KEKB
LHC
PEP II
DAFNE
CESR
TEVATRON
BEPC2
LEP2
ISR
PEP
PETRA
TRISTAN
HERA
DORIS
LEP1
SPEAR
BEPC
SppS
DCI

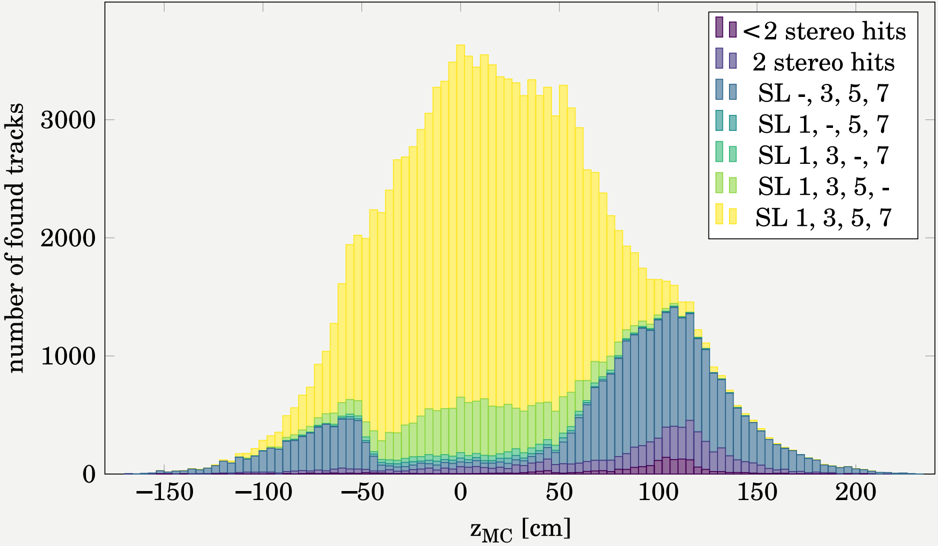
Results


ROC Rejection Plot
Background Rejection
Signal Efficiency
1.00
0.95
0.90
0.85
0.80
0.75
0.75
0.80
0.85
0.90
0.95
1.00
DNN(E+DL+V) 0.9977
BDT(E+DL+V) 0.9974
DNN(E+DL) 0.9950
BDT(E+DL) 0.9940
DNN(E) 0.9776
BDT(E) 0.9664

Continuum Suppression

- Introduced by CLEO collaboration in 1996
- Cones distributed in steps of $10\,\mathrm{deg}$
- Measure momentum flow into concentric areas around thrust axis

Z-Vertex Trigger



Z-Vertex Trigger