Charged Particle Identification Overview

Stefan Wallner
(swallner@mpp.mpg.de)

Max Planck Institute for Physics

Belle II Physics Week
October 18, 2024

Introduction

- Particle of various species are produced at Belle II and need to be distinguished
- For example, τ can decay to electrons, muons, pions, and kaons; which can be separated only by experimentally identifying the species of the particles
  - ➥ Requires charged particle identification (PID)
- Test lepton-flavor universality in $\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau$ decays

$R_\mu = \frac{\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau(\gamma))}{\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau(\gamma))}$

  - Most precise test of $\mu - e$ universality in τ decays from a single measurement at Belle II [JHEP 08 (2024) 205]
  - Consistent with Standard Model expectation
  - Strongly relies on PID

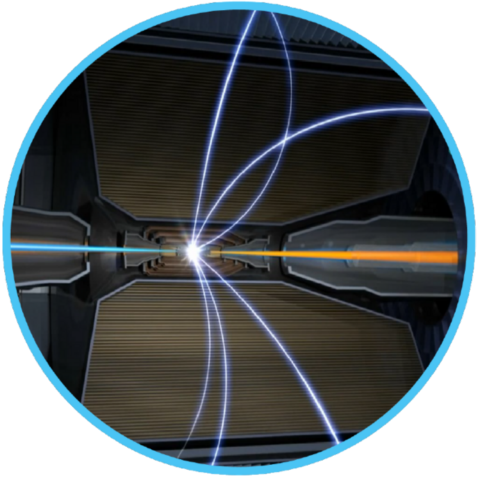
Introduction

- Particle of various species are produced at Belle II and need to be distinguished
- For example, τ can decay to electrons, muons, pions, and kaons; which can be separated only by experimentally identifying the species of the particles
  - ➥ Requires charged particle identification (PID)
- Test lepton-flavor universality in $\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau$ decays

  $R_\mu = \frac{\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau(\gamma))}{\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau(\gamma))}$
  - Most precise test of $\mu - e$ universality in τ decays from a single measurement at Belle II [JHEP 08 (2024) 205]
  - Consistent with Standard Model expectation
  - Strongly relies on PID

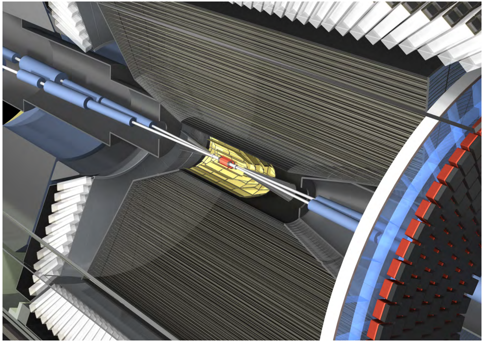
Introduction

- Tracks of charged particles measured in tracking detectors (PXD, SVD, CDC)
  - Measurement of track position and momentum
- Six species of charged particles that are “stable” within the Belle II detector
  - $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $\overset{(-)}{p}$, $\overset{(-)}{d}$
- Requires additional experimental measurement to identify the species of the track
  - ➥ PID
- Measure quantity that differs for the six particle species
  - Mass
  - Type of interaction
- Translate to a classification variable $\mathcal{L}_h$ representing how likely it is that the particle is of species h



Introduction

- Tracks of charged particles measured in tracking detectors (PXD, SVD, CDC)
  - Measurement of track position and momentum
- Six species of charged particles that are “stable” within the Belle II detector
  - $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $\overset{(-)}{p}$, $\overset{(-)}{d}$
- Requires additional experimental measurement to identify the species of the track
  - ➡ PID
- Measure quantity that differs for the six particle species
  - Mass
  - Type of interaction
- Translate to a classification variable $\mathcal{L}_h$ representing how likely it is that the particle is of species h



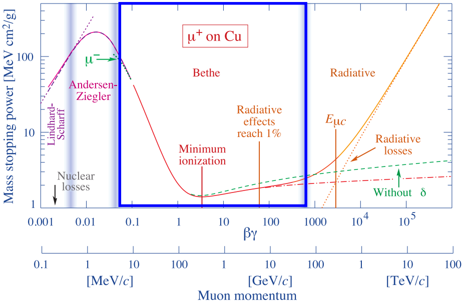
Outline

Physics Principles and PID Detectors

$$\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2 m_e c^2 b^2 \gamma^2 W_{\max}(\beta\gamma)}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right]$$

Physics Principles

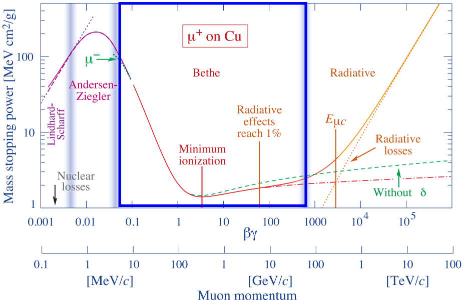
- Electronic energy loss of charged particles (except $e^{\pm}$) described by Bethe-Bloch equation
- Energy loss depends only on velocity of particle (also for $e^{\pm}$) and medium properties
- For given measured momentum, the energy loss is different for different particle masses
  - ➥ Identification of particle species
- Crossing points where energy loss is similar for different particle species
  - ➥ PID via energy loss works only in certain momentum regions



$$\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2 m_e c^2 b^2 \gamma^2 W_{\max}(\beta\gamma)}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right]$$

Physics Principles

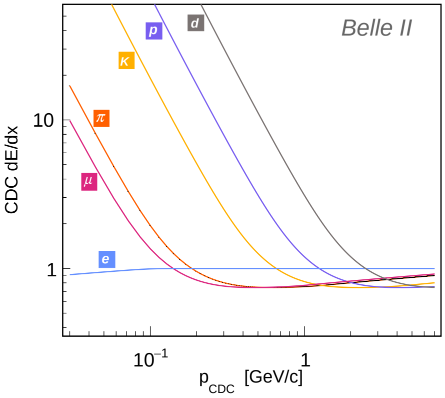
- Electronic energy loss of charged particles (except $e^{\pm}$) described by Bethe-Bloch equation
- Energy loss depends only on velocity of particle (also for $e^{\pm}$) and medium properties
- For given measured momentum, the energy loss is different for different particle masses
  - ➥ Identification of particle species
- Crossing points where energy loss is similar for different particle species
  - ➥ PID via energy loss works only in certain momentum regions



$$\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2 m_e c^2 b^2 \gamma^2 W_{\max}(\beta\gamma)}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] = F(p; m)$$

Physics Principles

- Electronic energy loss of charged particles (except $e^{\pm}$) described by Bethe-Bloch equation
- Energy loss depends only on velocity of particle (also for $e^{\pm}$) and medium properties
- For given measured momentum, the energy loss is different for different particle masses
  - ➥ Identification of particle species
- Crossing points where energy loss is similar for different particle species
  - ➥ PID via energy loss works only in certain momentum regions

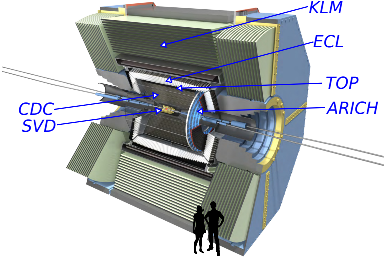


CDC and SVD PID

- Energy loss measured by ionization in CDC and SVD (and PXD)
- Formulate a likelihood for the particle-species hypotheses h

$$\log \mathcal{L}_h^{\mathrm{CDC}} = -\frac{\chi_h^2}{2} = -\frac{[\mathrm{d}E/\mathrm{d}x_{\mathrm{meas.}} - \mathrm{d}E/\mathrm{d}x^h_{\mathrm{pred.}}]^2}{2[\sigma^h_{\mathrm{pred.}}]^2}$$

  - Using calibration data to determine $\mathrm{d}E/\mathrm{d}x^h_{\mathrm{pred.}}$ for each particle species h
- For example, good e/π separation for $p \gtrsim 0.3\,\mathrm{GeV}/c$

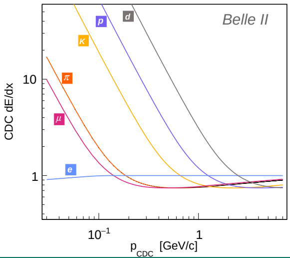


CDC and SVD PID

- Energy loss measured by ionization in CDC and SVD (and PXD)
- Formulate a likelihood for the particle-species hypotheses h

$$\log \mathcal{L}_h^{\mathrm{CDC}} = -\frac{\chi_h^2}{2} = -\frac{[\mathrm{d}E/\mathrm{d}x_{\mathrm{meas.}} - \mathrm{d}E/\mathrm{d}x_{\mathrm{pred.}}^h]^2}{2[\sigma_{\mathrm{pred.}}^h]^2}$$

  - Using calibration data to determine $\mathrm{d}E/\mathrm{d}x_{\mathrm{pred.}}^h$ for each particle species h
- For example, good e/π separation for $p \gtrsim 0.3\,\mathrm{GeV}/c$

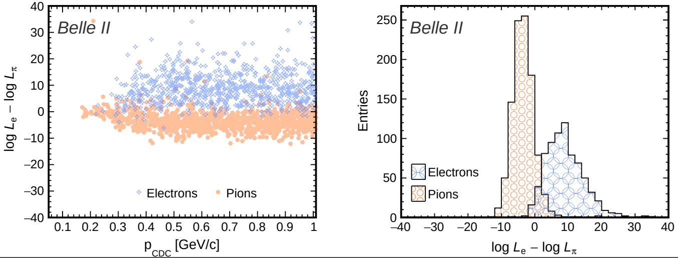


CDC and SVD PID

- Energy loss measured by ionization in CDC and SVD (and PXD)
- Formulate a likelihood for the particle-species hypotheses h

$$\log \mathcal{L}_h^{\mathrm{CDC}} = -\frac{\chi_h^2}{2} = -\frac{[\mathrm{d}E/\mathrm{d}x_{\mathrm{meas.}} - \mathrm{d}E/\mathrm{d}x^h_{\mathrm{pred.}}]^2}{2[\sigma^h_{\mathrm{pred.}}]^2}$$

  - Using calibration data to determine $\mathrm{d}E/\mathrm{d}x^h_{\mathrm{pred.}}$ for each particle species h
- For example, good e/π separation for $p \gtrsim 0.3\,\mathrm{GeV}/c$



Physics Principles

- Particles traversing a medium faster than the speed of light in the medium emit Cherenkov light
- Cherenkov light emitted on a cone with opening angle

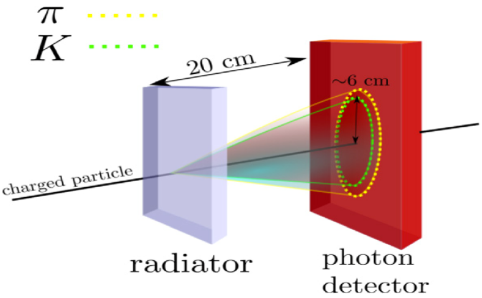
$$\cos\theta_\mathrm{C} = \frac{1}{n\beta}$$

- For given measured momentum, the Cherenkov angle is different for different particle masses
  - ➥ Identification of particle species
- Minimal momentum to produce Cherenkov light

$$p_\mathrm{th.} = \frac{m}{\sqrt{n^2 - 1}}$$

- Also number of Cherenkov photons sensitive to mass

$$\frac{\mathrm{d}^2 N}{\mathrm{d}E\mathrm{d}x} = \frac{\alpha z^2}{\hbar c}\sin^2\theta_\mathrm{C}$$



Physics Principles

- Particles traversing a medium faster than the speed of light in the medium emit Cherenkov light
- Cherenkov light emitted on a cone with opening angle

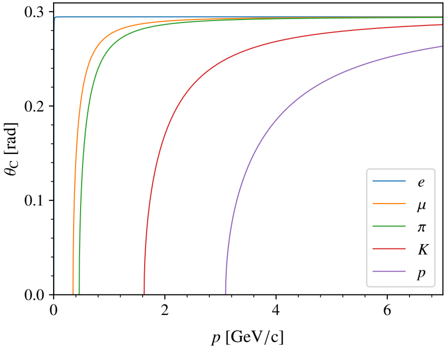
$$\cos\theta_{\mathrm{C}} = \frac{1}{n\beta} = G(p; m)$$

- For given measured momentum, the Cherenkov angle is different for different particle masses
  - ➥ Identification of particle species
- Minimal momentum to produce Cherenkov light

$$p_{\mathrm{th.}} = \frac{m}{\sqrt{n^2 - 1}}$$

- Also number of Cherenkov photons sensitive to mass

$$\frac{\mathrm{d}^2 N}{\mathrm{d}E\mathrm{d}x} = \frac{\alpha z^2}{\hbar c}\sin^2\theta_{\mathrm{C}}$$

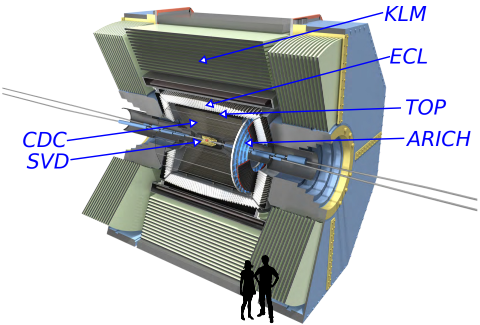


ARICH PID

- Covers forward region
- Cherenkov photons produced in silica aerogel radiator
- Measured by hybrid avalanche photo detectors
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ARICH}}$ for particle-species hypothesis h taking into account the probability for each individual pixel to be hit or not hit

TOP PID

- Covers barrel region
- Cherenkov photons produced in quartz transported via internal reflection and detected at the end
- Time of propagation depends on Cherenkov angle and position where photons leave the bar
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{TOP}}$

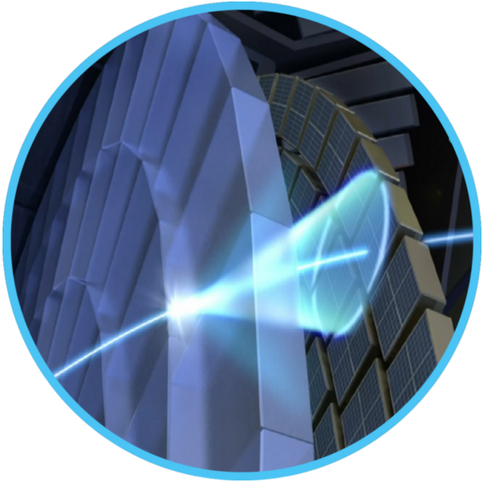


ARICH PID

- Covers forward region
- Cherenkov photons produced in silica aerogel radiator
- Measured by hybrid avalanche photo detectors
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ARICH}}$ for particle-species hypothesis h taking into account the probability for each individual pixel to be hit or not hit

TOP PID

- Covers barrel region
- Cherenkov photons produced in quartz transported via internal reflection and detected at the end
- Time of propagation depends on Cherenkov angle and position where photons leave the bar
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{TOP}}$

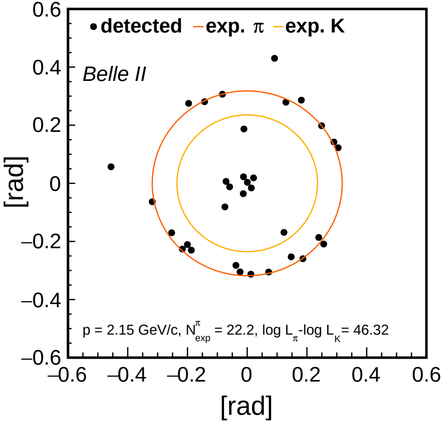


ARICH PID

- Covers forward region
- Cherenkov photons produced in silica aerogel radiator
- Measured by hybrid avalanche photo detectors
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ARICH}}$ for particle-species hypothesis h taking into account the probability for each individual pixel to be hit or not hit

TOP PID

- Covers barrel region
- Cherenkov photons produced in quartz transported via internal reflection and detected at the end
- Time of propagation depends on Cherenkov angle and position where photons leave the bar
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{TOP}}$

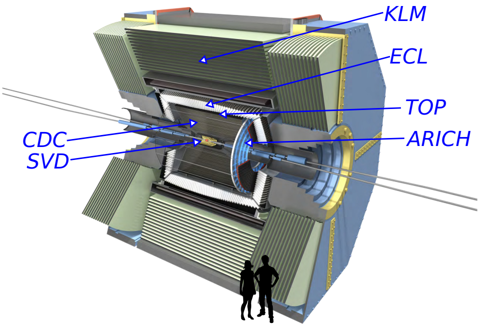


ARICH PID

- Covers forward region
- Cherenkov photons produced in silica aerogel radiator
- Measured by hybrid avalanche photo detectors
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ARICH}}$ for particle-species hypothesis h taking into account the probability for each individual pixel to be hit or not hit

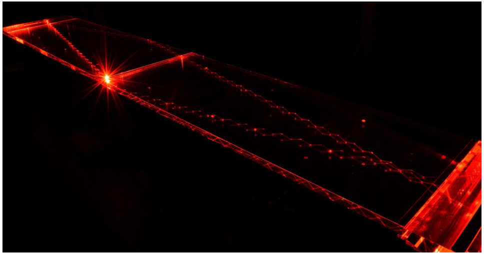
TOP PID

- Covers barrel region
- Cherenkov photons produced in quartz transported via internal reflection and detected at the end
- Time of propagation depends on Cherenkov angle and position where photons leave the bar
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{TOP}}$



ARICH PID

- Covers forward region
- Cherenkov photons produced in silica aerogel radiator
- Measured by hybrid avalanche photo detectors
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ARICH}}$ for particle-species hypothesis h taking into account the probability for each individual pixel to be hit or not hit

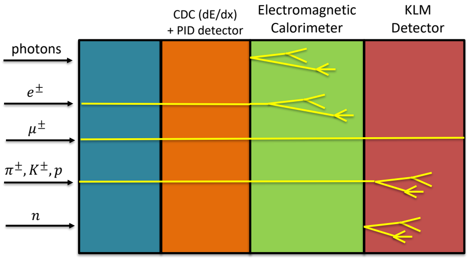


TOP PID

- Covers barrel region
- Cherenkov photons produced in quartz transported via internal reflection and detected at the end
- Time of propagation depends on Cherenkov angle and position where photons leave the bar
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{TOP}}$

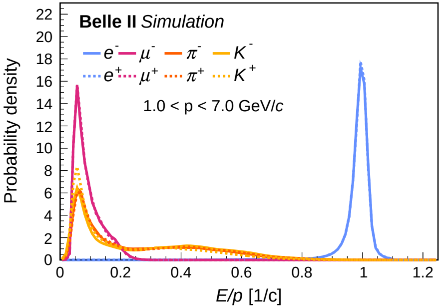
Physics Principles

- Electron energy fully absorbed in ECAL
- Ratio of in ECL deposited energy E and measured track momentum p is $E/p = 1$ for electrons
- Other species leave only fraction of their energy in ECL
- Depends on track momentum
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ECAL}}$ based on the expected energy deposition in ECAL



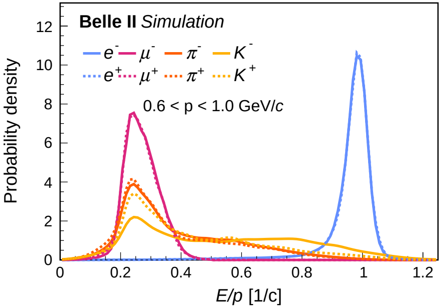
Physics Principles

- Electron energy fully absorbed in ECAL
- Ratio of in ECL deposited energy E and measured track momentum p is $E/p = 1$ for electrons
- Other species leave only fraction of their energy in ECL
- Depends on track momentum
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ECAL}}$ based on the expected energy deposition in ECAL



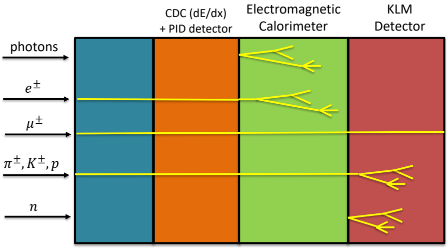
Physics Principles

- Electron energy fully absorbed in ECAL
- Ratio of in ECL deposited energy E and measured track momentum p is $E/p = 1$ for electrons
- Other species leave only fraction of their energy in ECL
- Depends on track momentum
- Formulate likelihood $\log \mathcal{L}_h^{\mathrm{ECAL}}$ based on the expected energy deposition in ECAL



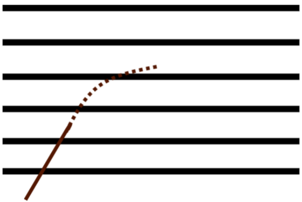
Physics Principles

- Muons have a large penetration depth fully traversing the KLM
  - Bremsstrahlung suppressed by $\frac{1}{m_\mu^2}$ with respect to electrons
  - No strong interaction
- In the KLM, muons have different
  - longitudinal penetration depth
  - transverse scattering
- Both used to formulate likelihood $\log \mathcal{L}_h^{\mathrm{KLM}}$ comparing the extrapolated track from inner detectors to KLM hits



Physics Principles

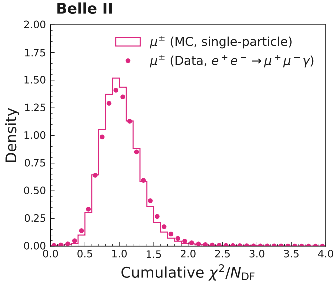
- Muons have a large penetration depth fully traversing the KLM
  - Bremsstrahlung suppressed by $\frac{1}{m_\mu^2}$ with respect to electrons
  - No strong interaction
- In the KLM, muons have different
  - longitudinal penetration depth
  - transverse scattering
- Both used to formulate likelihood $\log \mathcal{L}_h^{\mathrm{KLM}}$ comparing the extrapolated track from inner detectors to KLM hits

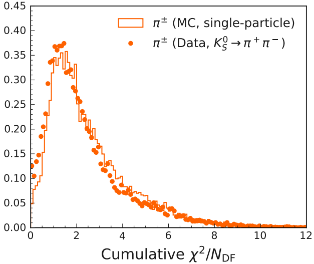


$$L = p_1 p_2 (1 - p_3 \varepsilon_3)\,(1 - p_4 \varepsilon_4)$$

Physics Principles

- Muons have a large penetration depth fully traversing the KLM
  - Bremsstrahlung suppressed by $\frac{1}{m_\mu^2}$ with respect to electrons
  - No strong interaction
- In the KLM, muons have different
  - longitudinal penetration depth
  - transverse scattering
- Both used to formulate likelihood $\log \mathcal{L}_h^{\mathrm{KLM}}$ comparing the extrapolated track from inner detectors to KLM hits





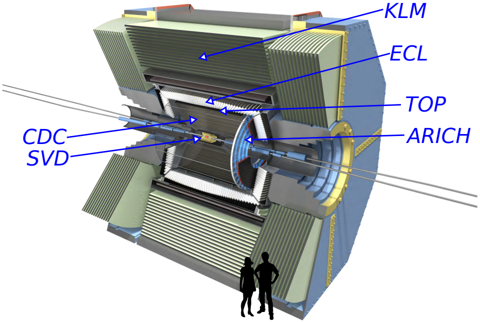
Physics Principles

- Muons have a large penetration depth fully traversing the KLM
  - Bremsstrahlung suppressed by $\frac{1}{m_\mu^2}$ with respect to electrons
  - No strong interaction
- In the KLM, muons have different
  - longitudinal penetration depth
  - transverse scattering
- Both used to formulate likelihood $\log \mathcal{L}_h^{\mathrm{KLM}}$ comparing the extrapolated track from inner detectors to KLM hits

Physics Principles and PID Detectors

Belle II PID

- In total 6 subdetectors that yield PID information
  - Different coverage of detector regions
  - Different coverage (separation power) of momentum regions
- Each provides likelihood for all 6 hypotheses
  - ➥ In total 36 likelihoods

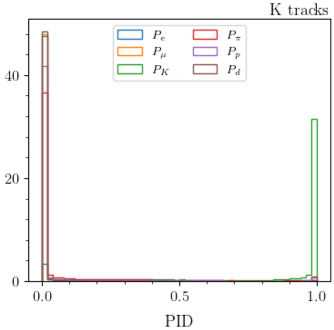


Using PID for Physics Analysis

Global PID

- Combine local detector likelihoods $\mathcal{L}_h^d$ to the global PID probability $\mathcal{P}_h$
- Assuming detector likelihoods are independent

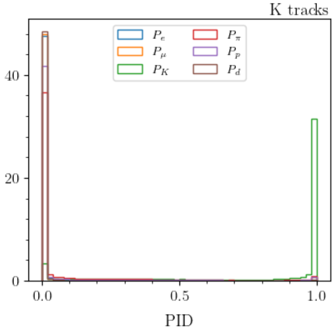
$$\mathcal{P}_h = \frac{\prod_d \mathcal{L}_h^d}{\sum_{h'} \prod_d \mathcal{L}_{h'}^d}$$

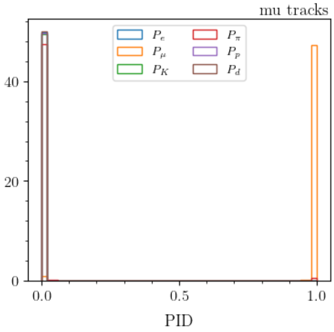


Global PID

- Combine local detector likelihoods $\mathcal{L}_h^d$ to the global PID probability $\mathcal{P}_h$
- Assuming detector likelihoods are independent

$$\mathcal{P}_h = \frac{\prod_d \mathcal{L}_h^d}{\sum_{h'} \prod_d \mathcal{L}_{h'}^d}$$

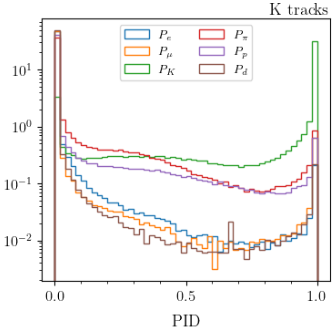


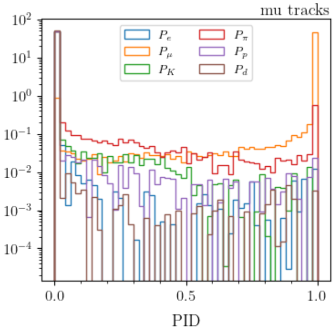


Global PID

- Combine local detector likelihoods $\mathcal{L}_h^d$ to the global PID probability $\mathcal{P}_h$
- Assuming detector likelihoods are independent

$$\mathcal{P}_h = \frac{\prod_d \mathcal{L}_h^d}{\sum_{h'} \prod_d \mathcal{L}_{h'}^d}$$





Global PID

- Accessible in basf2 via `ID` variables
  - `electronID` $= \mathcal{P}_e$; `muonID` $= P_\mu$; `pionID` $= P_\pi$; `kaonID` $= P_K$; `protonID` $= P_p$; `deuteronID` $= P_d$
- Sometimes, subdetectors need to be excluded from PID for better performance, e.g.

$$\text{muonID\_noSVD} = \frac{\prod_{d \notin \{\text{SVD}\}} \mathcal{L}_h^d}{\sum_{h'} \prod_{d \notin \{\text{SVD}\}} \mathcal{L}_{h'}^d}$$

Global PID

- Accessible in basf2 via `ID` variables
  - `electronID` $= \mathcal{P}_e$; `muonID` $= P_\mu$; `pionID` $= P_\pi$; `kaonID` $= P_K$; `protonID` $= P_p$; `deuteronID` $= P_d$
- Sometimes, subdetectors need to be excluded from PID for better performance, e.g.

$$\mathrm{muonID\_noSVD} = \frac{\prod_{d \notin \{\mathrm{SVD}\}} \mathcal{L}_h^d}{\sum_{h'} \prod_{d \notin \{\mathrm{SVD}\}} \mathcal{L}_{h'}^d}$$

Binary PID

- If only a certain subset of species needs to be separated by PID in physics analysis, normalize PID probability to only this subset
- If pions need to be separated only from kaons use binary π/K PID probability

$$P_{\pi/K} = \frac{\prod_d \mathcal{L}_\pi^d}{\prod_d \mathcal{L}_\pi^d + \prod_d \mathcal{L}_K^d} = \frac{P_\pi}{P_\pi + P_K}$$

- In basf2, this reads `pionID/(pionID+kaonID)`

- Simple combination of detector likelihoods, $\prod_d \mathcal{L}_d^h$ is imperfect
  - Ignores correlations among detector likelihoods; does not use full information; approximations in likelihoods
  - ➥ Train MVA method on simulated data to yield better PID variables

LeptonID BDT

- Use CDC, TOP, ARICH, KLM likelihoods
- Use ECL E/p and cluster shape observables
- Improve performance for electron ID $p < 1\,\mathrm{GeV}/c$

PID Neural Network for K/π separation

- Use all likelihoods from all 6 subdetectors
- Use measured track momentum and charge
- Improve performance for K/π separation for low fake rates

- Simple combination of detector likelihoods, $\prod_d \mathcal{L}_d^h$ is imperfect
  - Ignores correlations among detector likelihoods; does not use full information; approximations in likelihoods
  - ➥ Train MVA method on simulated data to yield better PID variables

LeptonID BDT

- Use CDC, TOP, ARICH, KLM likelihoods
- Use ECL E/p and cluster shape observables
- Improve performance for electron ID $p < 1\,\mathrm{GeV}/c$

PID Neural Network for K/π separation

- Use all likelihoods from all 6 subdetectors
- Use measured track momentum and charge
- Improve performance for K/π separation for low fake rates

- Simple combination of detector likelihoods, $\prod_d \mathcal{L}_d^h$ is imperfect
  - Ignores correlations among detector likelihoods; does not use full information; approximations in likelihoods
  - ➥ Train MVA method on simulated data to yield better PID variables

LeptonID BDT

- Use CDC, TOP, ARICH, KLM likelihoods
- Use ECL E/p and cluster shape observables
- Improve performance for electron ID $p < 1\,\mathrm{GeV}/c$

PID Neural Network for K/π separation

- Use all likelihoods from all 6 subdetectors
- Use measured track momentum and charge
- Improve performance for K/π separation for low fake rates

- PID setting depends on the data production
- PID recommendations can be found on PID Wiki page

Recommendations for Release06 data (MC15ri, MC15rd, proc13+prompt)

	Global	Binary (ℓ/π or K/π)
electron	electronID_noSVD_noTOP, pidChargedBDTScore_e	binaryElectronID_noSVD_noTOP_pi, pidPairChargedBDTScore_e_pi
muon	muonID_noSVD, pidChargedBDTScore_mu	binaryMuonID_noSVD_pi, pidPairChargedBDTScore_mu_pi
pion	pionID	pionIDNN
kaon	kaonID	kaonIDNN
proton	protonID	

Using PID for Physics Analysis

PID Recommendations

- PID setting depends on the data production
- PID recommendations can be found on PID Wiki page

Recommendations for Release06 data (MC15ri, MC15rd, proc13+prompt)

	Global	Binary (ℓ/π or K/π)
electron	electronID_noSVD_noTOP, pidChargedBDTScore_e	binaryElectronID_noSVD_noTOP_pi, pidPairChargedBDTScore_e_pi
muon	muonID_noSVD, pidChargedBDTScore_mu	binaryMuonID_noSVD_pi, pidPairChargedBDTScore_mu_pi
pion	pionID	pionIDNN
kaon	kaonID	kaonIDNN
proton	protonID	

PID Performance

- Correctly estimating PID effects is crucial for physics analysis
- Study the PID performance in real data and simulation
  - Efficiency to identify particle of species s: $P(s \to s)$
  - Fake rate to wrongly identify particle of species s as particle-species hypothesis h: $P(s \to h)$
- Requires sample of tracks where species is known without detector PID information
  - ➥ Use known decays of particles, where the species of the daughter particles is know for the dominant decay mode and where all other decay modes are strongly suppressed

- Correctly estimating PID effects is crucial for physics analysis
- Study the PID performance in real data and simulation
  - Efficiency to identify particle of species s: $P(s \to s)$
  - Fake rate to wrongly identify particle of species s as particle-species hypothesis h: $P(s \to h)$
- Requires sample of tracks where species is known without detector PID information
  - ➥ Use known decays of particles, where the species of the daughter particles is know for the dominant decay mode and where all other decay modes are strongly suppressed

D^* sample of kaons and pions

$$D^{*,+} \longrightarrow D^0 \, \pi^+_{slow}$$
$$\hookrightarrow K^- \pi^+$$

- Two-body D^0 decays
  - The negative decay product is almost always a K^-
  - The positive decay product is almost always a π^-
- Analogously for $\bar{D}^0$ decays
- Select $\overset{(-)}{D}{}^0$ signal and distinguish D^0 from $\bar{D}^0$ in $D^{*,\pm} \to \overset{(-)}{D}{}^0 \pi^\pm_{\mathrm{slow}}$ decays
  - Reconstructed D^0 masses separates signal from background
    - Statistical background subtraction using sPlot technique
  - Covers large kinematic region

D^* sample of kaons and pions

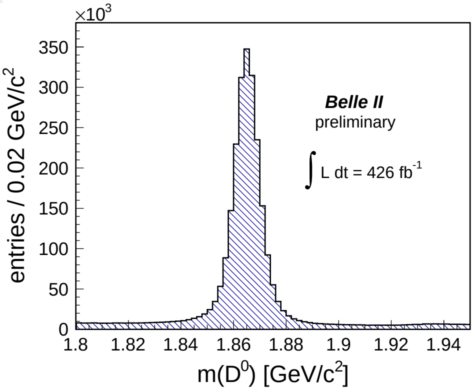
$$D^{*,+} \longrightarrow D^0\,\pi^+_{slow}$$
$$\hookrightarrow K^- \pi^+$$

- Two-body D^0 decays
  - The negative decay product is almost always a K^-
  - The positive decay product is almost always a π^-
- Analogously for $\bar{D}^0$ decays
- Select $\overset{(-)}{D}{}^0$ signal and distinguish D^0 from $\bar{D}^0$ in $D^{*,\pm} \to \overset{(-)}{D}{}^0\pi^\pm_{\mathrm{slow}}$ decays
  - Reconstructed D^0 masses separates signal from background
    - Statistical background subtraction using sPlot technique
  - Covers large kinematic region

D^* sample of kaons and pions

$$D^{*,+} \longrightarrow D^0\,\pi^+_{slow}$$
$$\quad\quad\quad\quad \hookrightarrow K^-\pi^+$$

- Two-body D^0 decays
  - The negative decay product is almost always a K^-
  - The positive decay product is almost always a π^-
- Analogously for $\bar{D}^0$ decays
- Select $\overset{(-)}{D}{}^0$ signal and distinguish D^0 from $\bar{D}^0$ in $D^{*,\pm} \rightarrow \overset{(-)}{D}{}^0\pi^\pm_{\mathrm{slow}}$ decays
  - Reconstructed D^0 masses separates signal from background
    - Statistical background subtraction using sPlot technique
  - Covers large kinematic region



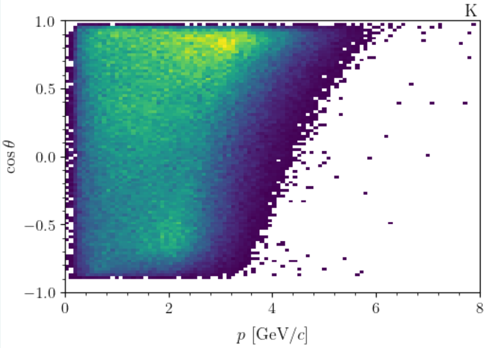
PID Performance

Performance Samples

D^* sample of kaons and pions

$$D^{*,+} \longrightarrow D^0 \, \pi^+_{slow}$$
$$\quad\quad\quad\quad \hookrightarrow K^- \pi^+$$

- Two-body D^0 decays
  - The negative decay product is almost always a K^-
  - The positive decay product is almost always a π^-
- Analogously for $\bar{D}^0$ decays
- Select $\overset{(-)}{D}{}^0$ signal and distinguish D^0 from $\bar{D}^0$ in $D^{*,\pm} \to \overset{(-)}{D}{}^0 \pi^{\pm}_{\text{slow}}$ decays
  - Reconstructed D^0 masses separates signal from background
    - Statistical background subtraction using sPlot technique
  - Covers large kinematic region



K_S^0 sample for pions

$$K_S^0 \to \pi^- \pi^+$$

- K_S^0 mainly decay to pions
- Covering mainly low-momentum region

Λ sample for pions and protons

$$\Lambda \to p\pi^-$$

- Λ decays mainly to pions and kaons
- Separate proton from pion by kinematics (Armenteros plot)

τ sample for pions

$$\tau^- \to \pi^- \pi^- \pi^+$$

- Identify two same-charge pions
- Used for pion to lepton fake rate

K^0_S sample for pions

$$K^0_S \to \pi^- \pi^+$$

- K^0_S mainly decay to pions
- Covering mainly low-momentum region

Λ sample for pions and protons

$$\Lambda \to p \pi^-$$

- Λ decays mainly to pions and kaons
- Separate proton from pion by kinematics (Armenteros plot)

τ sample for pions

$$\tau^- \to \pi^- \pi^- \pi^+$$

- Identify two same-charge pions
- Used for pion to lepton fake rate

J/ψ sample for electrons and muons

$$J/\psi \to \ell^- \ell^+$$

Four-Lepton sample for electrons and muons

$$e^- e^+ \to e^- e^+ \ell^- \ell^+$$

Two-Lepton sample for electrons and muons

$$e^- e^+ \to \ell^- \ell^+ (\gamma)$$

Efficiency

- Efficiency is the fraction of true particles of species s that pass a certain PID cut $P_s > t$ where t is the PID threshold

$$P(s \to s) = \frac{\#_s(P_s > t)}{\#_s(\mathrm{all})}$$

- For example, the kaon efficiency for a kaonID cut of 0.6 is

$$P(K \to K) = \frac{\#_K(P_K > 0.6)}{\#_K(\mathrm{all})}$$

Fake rate

- The fake rate or misidentification rate is the fraction of true particles of species s that pass a certain PID cut $P_h > t$ for hypothesis h

$$P(s \to h) = \frac{\#_s(P_h > t)}{\#_s(\mathrm{all})}$$

- For example, the pion to kaon fake rate is for a kaonID cut of 0.6 is

$$P(\pi \to K) = \frac{\#_\pi(P_K > 0.6)}{\#_\pi(\mathrm{all})}$$

Efficiency

- Efficiency is the fraction of true particles of species s that pass a certain PID cut $P_s > t$ where t is the PID threshold

$$P(s \to s) = \frac{\#_s(P_s > t)}{\#_s(\mathrm{all})}$$

- For example, the kaon efficiency for a kaonID cut of 0.6 is

$$P(K \to K) = \frac{\#_K(P_K > 0.6)}{\#_K(\mathrm{all})}$$

Fake rate

- The fake rate or misidentification rate is the fraction of true particles of species s that pass a certain PID cut $P_h > t$ for hypothesis h

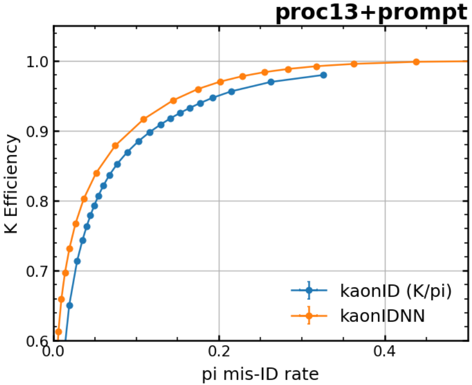
$$P(s \to h) = \frac{\#_s(P_h > t)}{\#_s(\mathrm{all})}$$

- For example, the pion to kaon fake rate is for a kaonID cut of 0.6 is

$$P(\pi \to K) = \frac{\#_\pi(P_K > 0.6)}{\#_\pi(\mathrm{all})}$$

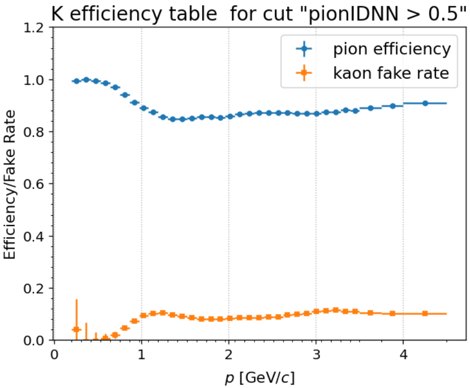
K/π separation

- Overall good K/π separation
- Improved separation using PIDNN
- Increase in efficiency for $p \lesssim 1\,\mathrm{GeV}/c$
  - Due to $\mathrm{d}E/\mathrm{d}x$ measurement in CDC
- Good separation in barrel and forward region in $\cos\theta \gtrsim -0.5$
- PID performance is a function of $(p, \cos\theta, q)$



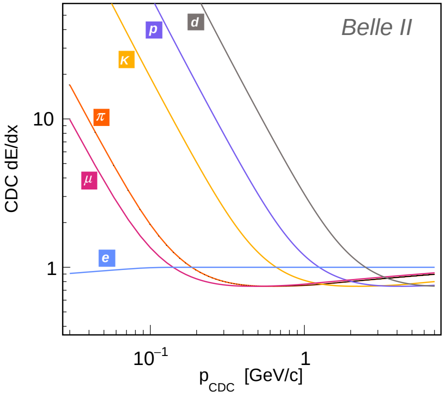
K/π separation

- Overall good K/π separation
- Improved separation using PIDNN
- Increase in efficiency for $p \lesssim 1\,\mathrm{GeV}/c$
  - Due to $\mathrm{d}E/\mathrm{d}x$ measurement in CDC
- Good separation in barrel and forward region in $\cos\theta \gtrsim -0.5$
- PID performance is a function of $(p, \cos\theta, q)$



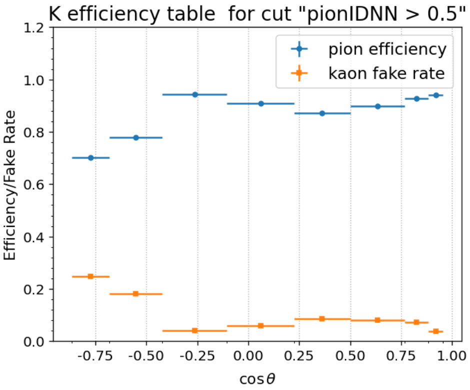
K/π separation

- Overall good K/π separation
- Improved separation using PIDNN
- Increase in efficiency for $p \lesssim 1\,\mathrm{GeV}/c$
  - Due to $\mathrm{d}E/\mathrm{d}x$ measurement in CDC
- Good separation in barrel and forward region in $\cos\theta \gtrsim -0.5$
- PID performance is a function of $(p, \cos\theta, q)$



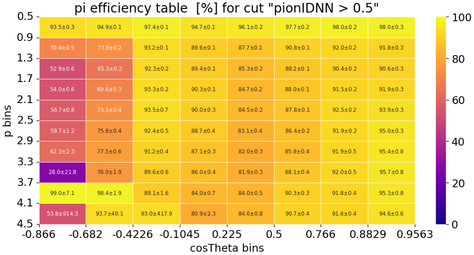
K/π separation

- Overall good K/π separation
- Improved separation using PIDNN
- Increase in efficiency for $p \lesssim 1\,\mathrm{GeV}/c$
  - Due to $\mathrm{d}E/\mathrm{d}x$ measurement in CDC
- Good separation in barrel and forward region in $\cos\theta \gtrsim -0.5$
- PID performance is a function of $(p, \cos\theta, q)$



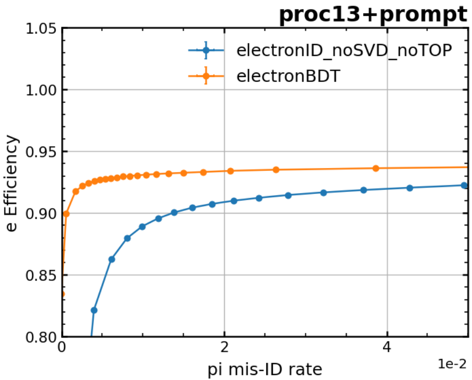
K/π separation

- Overall good K/π separation
- Improved separation using PIDNN
- Increase in efficiency for $p \lesssim 1\,\mathrm{GeV}/c$
  - Due to $\mathrm{d}E/\mathrm{d}x$ measurement in CDC
- Good separation in barrel and forward region in $\cos\theta \gtrsim -0.5$
- PID performance is a function of $(p, \cos\theta, q)$



ℓ/π separation

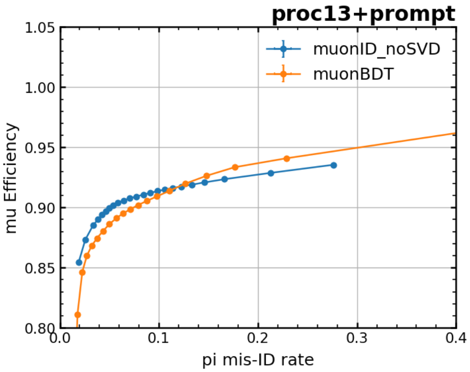
- Very good e/π separation due to ECL
  - Good electron efficiency ($\approx 95\,\%$)
  - Low pion fake rate ($\lesssim 1\,\%$)
- Very good μ/π separation
  - For $p \gtrsim 1\,\mathrm{GeV}/c$
  - For $p \lesssim 1\,\mathrm{GeV}/c$, particle does not reach KLM
    - ⇒ Worse μ/π separation as $m_\mu \approx m_\pi$



PID Performance
Lepton ID Performance

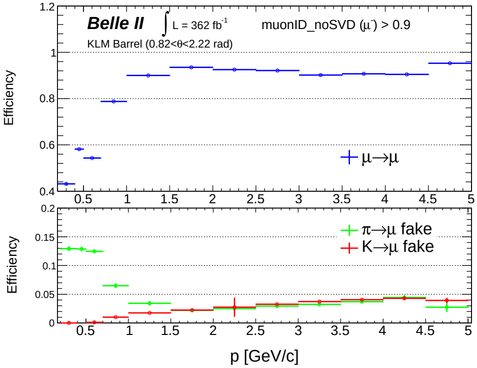
ℓ/π separation

- Very good e/π separation due to ECL
  - Good electron efficiency ($\approx 95\,\%$)
  - Low pion fake rate ($\lesssim 1\,\%$)
- Very good μ/π separation
  - For $p \gtrsim 1\,\mathrm{GeV}/c$
  - For $p \lesssim 1\,\mathrm{GeV}/c$, particle does not reach KLM
    - ➥ Worse μ/π separation as $m_\mu \approx m_\pi$



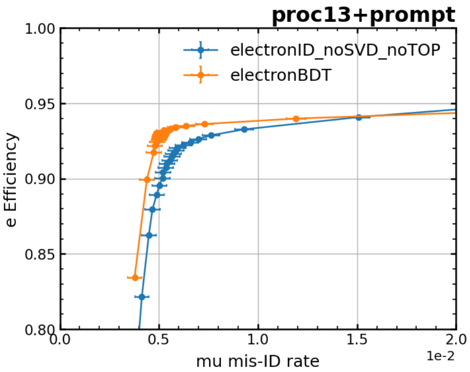
ℓ/π separation

- Very good e/π separation due to ECL
  - Good electron efficiency ($\approx 95\,\%$)
  - Low pion fake rate ($\lesssim 1\,\%$)
- Very good μ/π separation
  - For $p \gtrsim 1\,\mathrm{GeV}/c$
  - For $p \lesssim 1\,\mathrm{GeV}/c$, particle does not reach KLM
    - ➥ Worse μ/π separation as $m_\mu \approx m_\pi$



e/μ separation

- Good e/μ separation
  - Good electron efficiency ($\approx 95\,\%$)
  - Low muon fake rate ($\lesssim 1\,\%$)



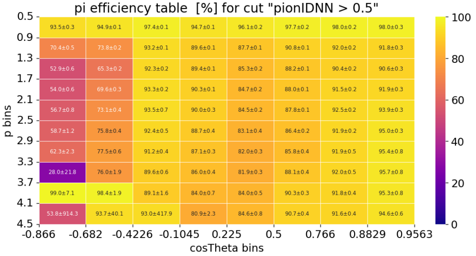
Correcting for PID Effects

Correcting for PID Effects

- Not all particle tracks can be identified
  - ➥ Not all decays are reconstructed and selected
  - ➥ Number of measured decays smaller than actual number of decays
  - ➥ Acceptance / efficiency
- Acceptance is non-uniform in phase-space of the particle
  - ➥ Causes deformation of measured distribution
- Acceptance correction done using detector Monte Carlo (MC) simulation of signal process
  - ➥ Requires detector simulation to match real detector performance

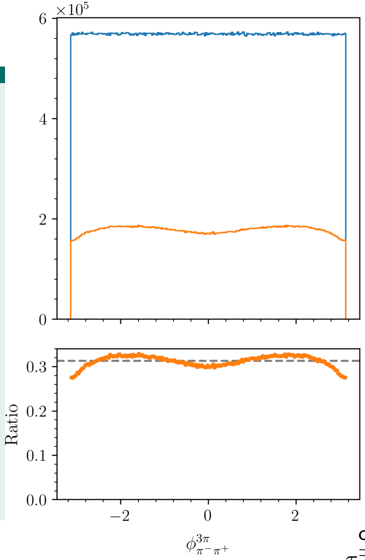
Correcting for PID Effects

- Not all particle tracks can be identified
  - ➥ Not all decays are reconstructed and selected
  - ➥ Number of measured decays smaller than actual number of decays
  - ➥ Acceptance / efficiency
- Acceptance is non-uniform in phase-space of the particle
  - ➥ Causes deformation of measured distribution
- Acceptance correction done using detector Monte Carlo (MC) simulation of signal process
  - ➥ Requires detector simulation to match real detector performance



Correcting for PID Effects

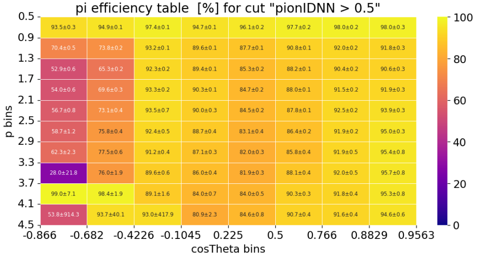
- Not all particle tracks can be identified
  - ➥ Not all decays are reconstructed and selected
  - ➥ Number of measured decays smaller than actual number of decays
  - ➥ Acceptance / efficiency
- Acceptance is non-uniform in phase-space of the particle
  - ➥ Causes deformation of measured distribution
- Acceptance correction done using detector Monte Carlo (MC) simulation of signal process
  - ➥ Requires detector simulation to match real detector performance



Angular distribution of all and selected $\tau^\mp \to \pi^\mp\pi^\mp\pi^\pm \overset{(-)}{\nu_\tau}$

Correcting for PID Effects

- Not all particle tracks can be identified
  - ➥ Not all decays are reconstructed and selected
  - ➥ Number of measured decays smaller than actual number of decays
  - ➥ Acceptance / efficiency
- Acceptance is non-uniform in phase-space of the particle
  - ➥ Causes deformation of measured distribution
- Acceptance correction done using detector Monte Carlo (MC) simulation of signal process
  - ➥ Requires detector simulation to match real detector performance

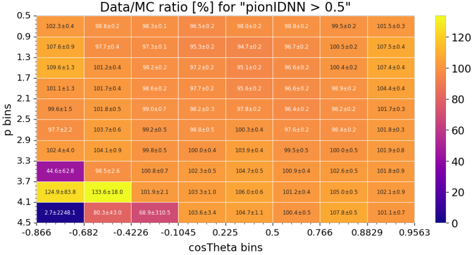


Angular distribution of all and selected $\tau^{\mp} \to \pi^{\mp}\pi^{\mp}\pi^{\pm} \overset{(-)}{\nu_\tau}$

Correcting for PID Effects

Real-Data / Simulation Agreement

- Compare PID performance from detector MC simulation with real-data using performance samples
- Overall fair agreement on few %-level
- Agreement depends on $(p, \cos\theta, q)$



Real-Data / Simulation Correction

- Simulation needs to be corrected for real-data/simulation disagreement
- Extract correction factor for each identified particle as a function of $(p, \cos\theta, p)$ from performance samples of real and simulated data

Offline reweighting and SysVar package

- Lepton ID: Correction tables at fixed working points (PID cut thresholds) available
- Hadron ID: Correction tables can be extracted from the Systematic Corrections Framework
  - At analyst-defined working point
  - Analyst can define additional track selection criteria that may affect PID performance
- The correction tables can be applied to simulated data using the SysVar package (see basf2 documentation)

Real-Data / Simulation Correction

- Simulation needs to be corrected for real-data/simulation disagreement
- Extract correction factor for each identified particle as a function of $(p, \cos\theta, p)$ from performance samples of real and simulated data

Offline reweighting and SysVar package

- Lepton ID: Correction tables at fixed working points (PID cut thresholds) available
- Hadron ID: Correction tables can be extracted from the Systematic Corrections Framework
  - At analyst-defined working point
  - Analyst can define additional track selection criteria that may affect PID performance
- The correction tables can be applied to simulated data using the SysVar package (see basf2 documentation)

Real-Data / Simulation Correction

- Simulation needs to be corrected for real-data/simulation disagreement
- Extract correction factor for each identified particle as a function of $(p, \cos\theta, p)$ from performance samples of real and simulated data

Offline reweighting and SysVar package

- Lepton ID: Correction tables at fixed working points (PID cut thresholds) available
- Hadron ID: Correction tables can be extracted from the Systematic Corrections Framework
  - At analyst-defined working point
  - Analyst can define additional track selection criteria that may affect PID performance
- The correction tables can be applied to simulated data using the SysVar package (see basf2 documentation)

Real-Data / Simulation Correction

- Simulation needs to be corrected for real-data/simulation disagreement
- Extract correction factor for each identified particle as a function of $(p, \cos\theta, p)$ from performance samples of real and simulated data

Offline reweighting and SysVar package

- Lepton ID: Correction tables at fixed working points (PID cut thresholds) available
- Hadron ID: Correction tables can be extracted from the Systematic Corrections Framework
  - At analyst-defined working point
  - Analyst can define additional track selection criteria that may affect PID performance
- The correction tables can be applied to simulated data using the SysVar package (see basf2 documentation)

Considerations for analysts

- Lepton ID samples only for fixed PID working points
  - ➥ You need to tune your selection to the available working points
- Additional track selection criteria, e.g. cut on nCDCHits, may affect the PID performance
  - ➥ Lepton ID correction tables available only for no additional track criteria
  - ➥ Hadron ID correction tables must be tuned to your additional track criteria
    - Systematic Corrections Framework per default applies cut on nCDCHits>20
- Recommended PID variables, cuts, correction tables, ... differ between releases (data productions)
  - Check PID recommendations

Considerations for analysts

- Lepton ID samples only for fixed PID working points
  - ➥ You need to tune your selection to the available working points
- Additional track selection criteria, e.g. cut on nCDCHits, may affect the PID performance
  - ➥ Lepton ID correction tables available only for no additional track criteria
  - ➥ Hadron ID correction tables must be tuned to your additional track criteria
    - Systematic Corrections Framework per default applies cut on nCDCHits>20
- Recommended PID variables, cuts, correction tables, ... differ between releases (data productions)
  - Check PID recommendations

Considerations for analysts

- Lepton ID samples only for fixed PID working points
  - ➡ You need to tune your selection to the available working points
- Additional track selection criteria, e.g. cut on nCDCHits, may affect the PID performance
  - ➡ Lepton ID correction tables available only for no additional track criteria
  - ➡ Hadron ID correction tables must be tuned to your additional track criteria
    - Systematic Corrections Framework per default applies cut on nCDCHits>20
- Recommended PID variables, cuts, correction tables, ... differ between releases (data productions)
  - Check PID recommendations

Systematic Uncertainties

- Real-data/simulation corrections have statistical and systematic uncertainty
  - Finite performance-sample size
  - Uncertainties from background subtraction (sPlot method, background modeling, ...)
  - ...
- Uncertainties provided with correction tables and propagated via the SysVar package
- Dominating systematic uncertainty for some analyses (e.g. lepton-flavor universality in τ decays)
- Many improvements possible
  - Take into account correlations in lepton ID systematic uncertainties
  - Improve background subtraction
  - ....

Systematic Uncertainties

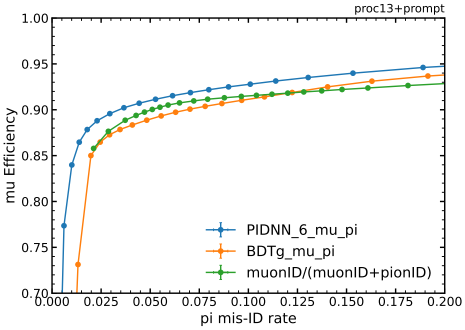
- Real-data/simulation corrections have statistical and systematic uncertainty
  - Finite performance-sample size
  - Uncertainties from background subtraction (sPlot method, background modeling, ...)
  - ...
- Uncertainties provided with correction tables and propagated via the SysVar package
- Dominating systematic uncertainty for some analyses (e.g. lepton-flavor universality in τ decays)
- Many improvements possible
  - Take into account correlations in lepton ID systematic uncertainties
  - Improve background subtraction
  - ....

Outlook

Outlook

Release08 data (proc16/MC16) and beyond

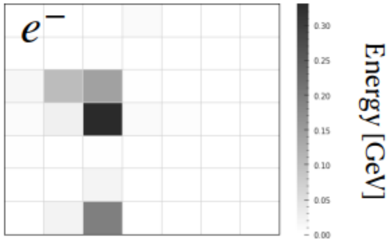
- Neural network PID will be extended and to all species
  - Can be used already now for release 06 (proc13/MC15) data
- Also lepton ID corrections will be provided via the Systematic Corrections Framework
- Improvements in detector likelihoods (algorithm and calibration)
  - Convolutional neural network for ECL reconstruction
  - KLM reconstruction using neural networks
  - Improved CDC/SVD calibrations taking into account time after injection
  - ...

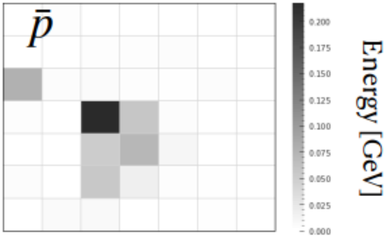


Outlook

Release08 data (proc16/MC16) and beyond

- Neural network PID will be extended and to all species
  - Can be used already now for release 06 (proc13/MC15) data
- Also lepton ID corrections will be provided via the Systematic Corrections Framework
- Improvements in detector likelihoods (algorithm and calibration)
  - Convolutional neural network for ECL reconstruction
  - KLM reconstruction using neural networks
  - Improved CDC/SVD calibrations taking into account time after injection
  - ...





Summary

Summary

- PID information from various subdetectors covering different kinematic regimes
- Various PID variables available in basf2 (check recommendation)
- Real-data/simulation corrections provided by PID group (check matching your analysis)
- If you have any questions, reach out to us
- You can contribute to improving our PID: Contact us for a service task

Links / References

- Wiki of PID performance group
- PID Recommendations
- List of service tasks
- PID mailing list: physics-performance-pid
- PID performance-group meetings on Thursday
- PID conveners: Alessandro Gaz, Stefan Wallner

Backup

Outline