Lepton identification overview

Kenta Uno (KEK)

Belle II physics week

November 2nd 2023

Self-introduction

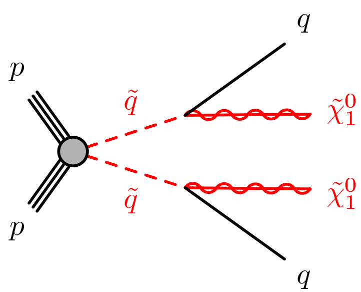
Kenta Uno

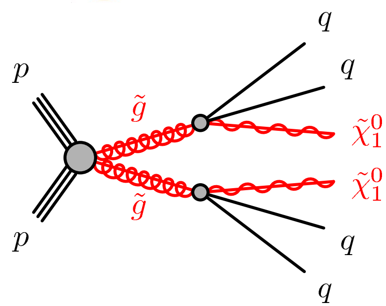
- 2015-2020: LHC-ATLAS experiment
    - Firmware development
    - SUSY search

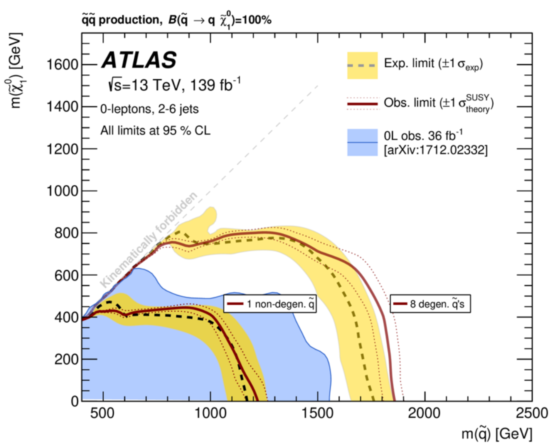
No Lepton in this analysis 😅

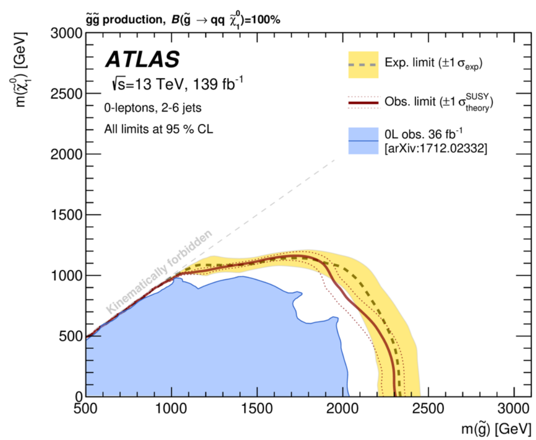
- 2020-2023: Belle II experiment
    - Tau physics analysis
    - ARICH operation
    - PID calibration

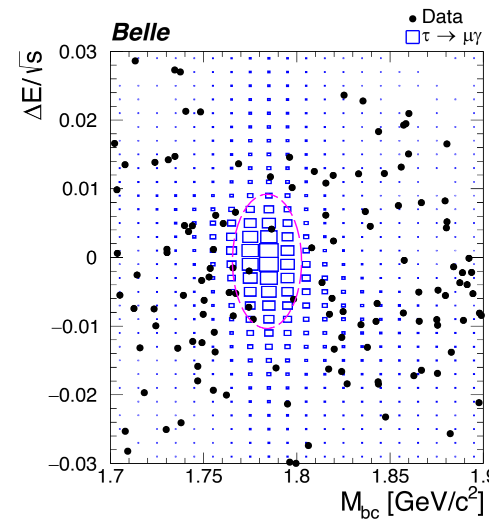
Learned lepton physics after joining Belle II ☺



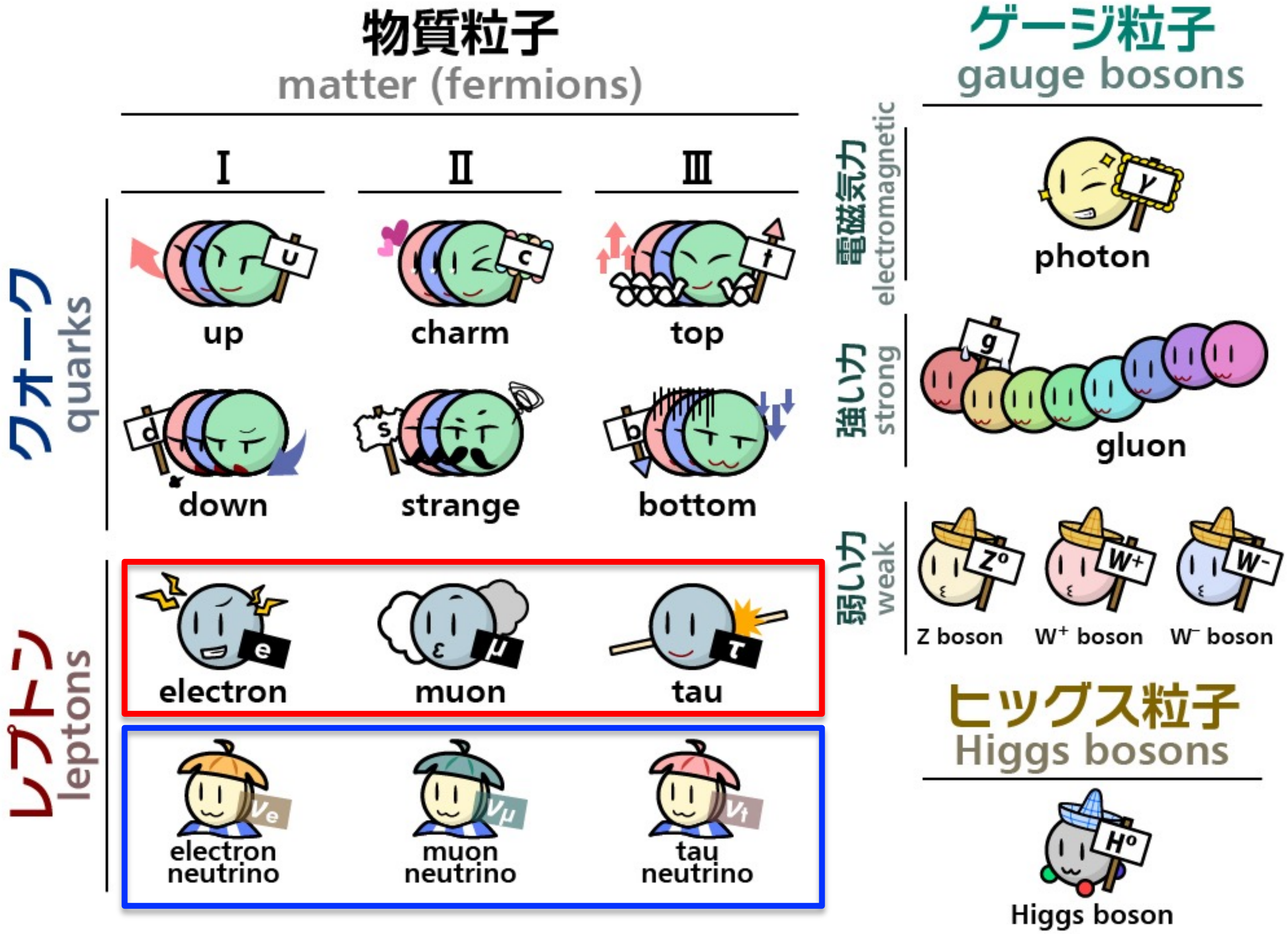








Introduction



Leptons:

- Charged Leptons: electron, muon, tau
- Neutral Leptons: electron neutrino, muon neutrino, tau neutrino

Introduction: Neutral Lepton

Neutrinos: ν_e, ν_μ, ν_τ

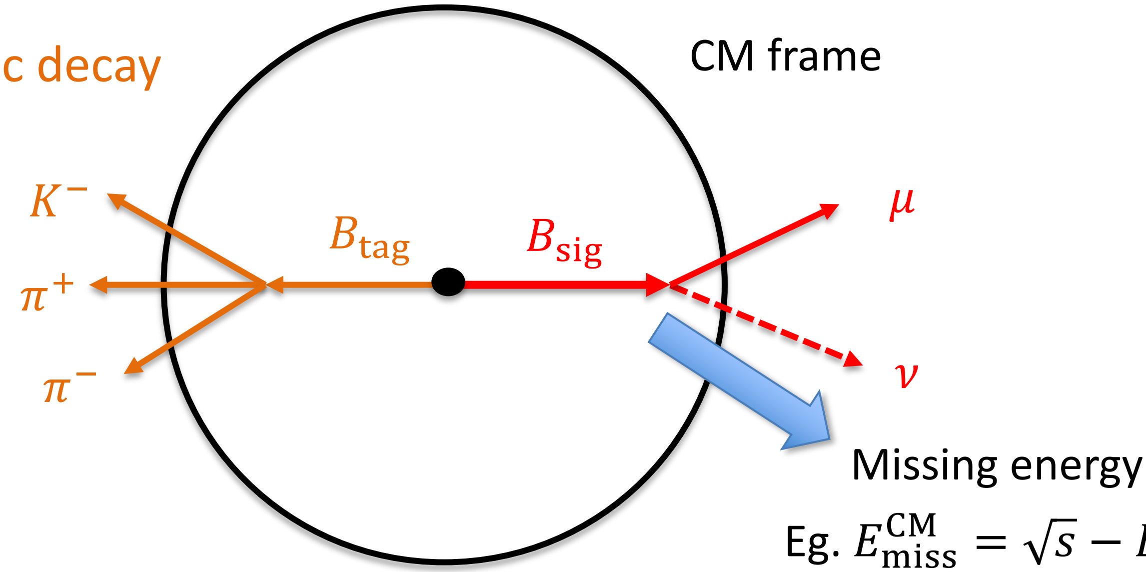
- It is not possible to detect them

→ The three-momentum and energy can be estimated

(if there is one neutrino in an event)

Eg. $ee \to \Upsilon(4S) \to B_{\text{sig}} B_{\text{tag}}$

- $B_{\text{sig}} \to \mu\nu$
- $B_{\text{tag}} \to$ hadronic decay



Eg. $E_{\text{miss}}^{\text{CM}} = \sqrt{s} - E_{\text{vis}}^{\text{CM}} = E_\nu^{\text{CM}}$

<u>Not detect neutrinos → No way to identify neutrinos</u>

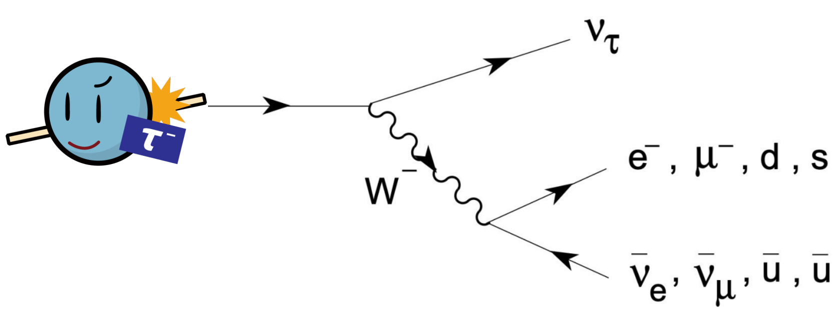
Introduction: Charged Lepton

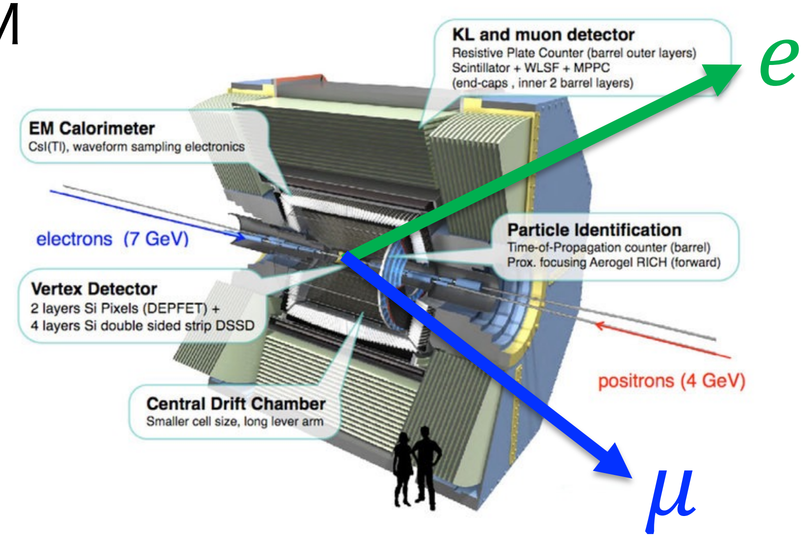
Electron (e)	Muon (μ)	Tau (τ)
• mass: 511 keV/c^2	• mass: 105 MeV/c^2	• mass: 1.78 GeV/c^2
• Lifetime: stable	• Lifetime: 2.2 μs	• Lifetime: 300 fs

Tau: Heaviest/Shortest lepton in the SM

- Many decays (Leptonic and Hadronic)
  - $\tau \to e\nu\nu, \mu\nu\nu, \pi\nu, \rho\nu, ..$

we cannot directly detect τ due to a short lifetime

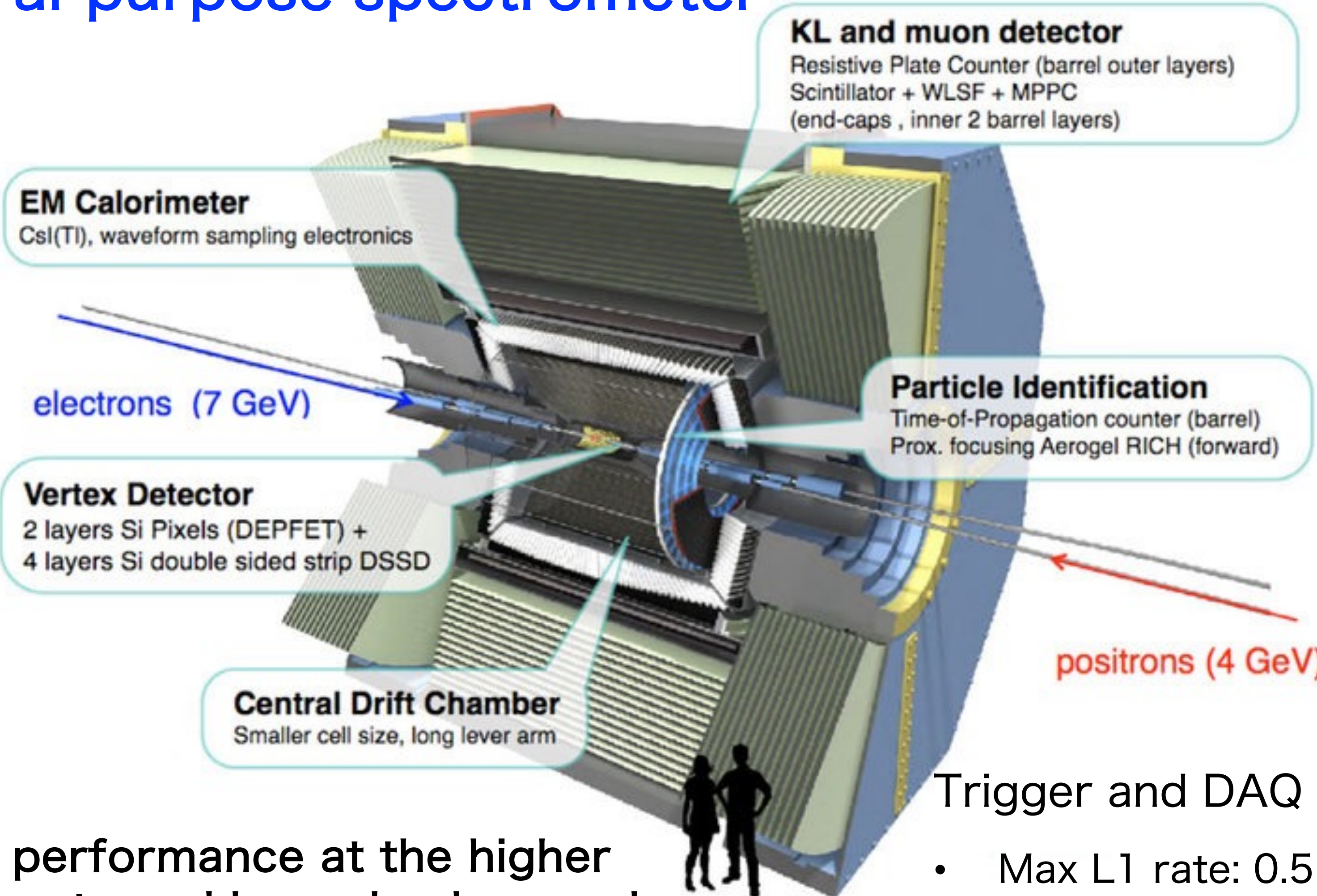




Lepton ID at Belle II: Identify "electron" and "muon"

Introduction: Belle II detector

General-purpose spectrometer

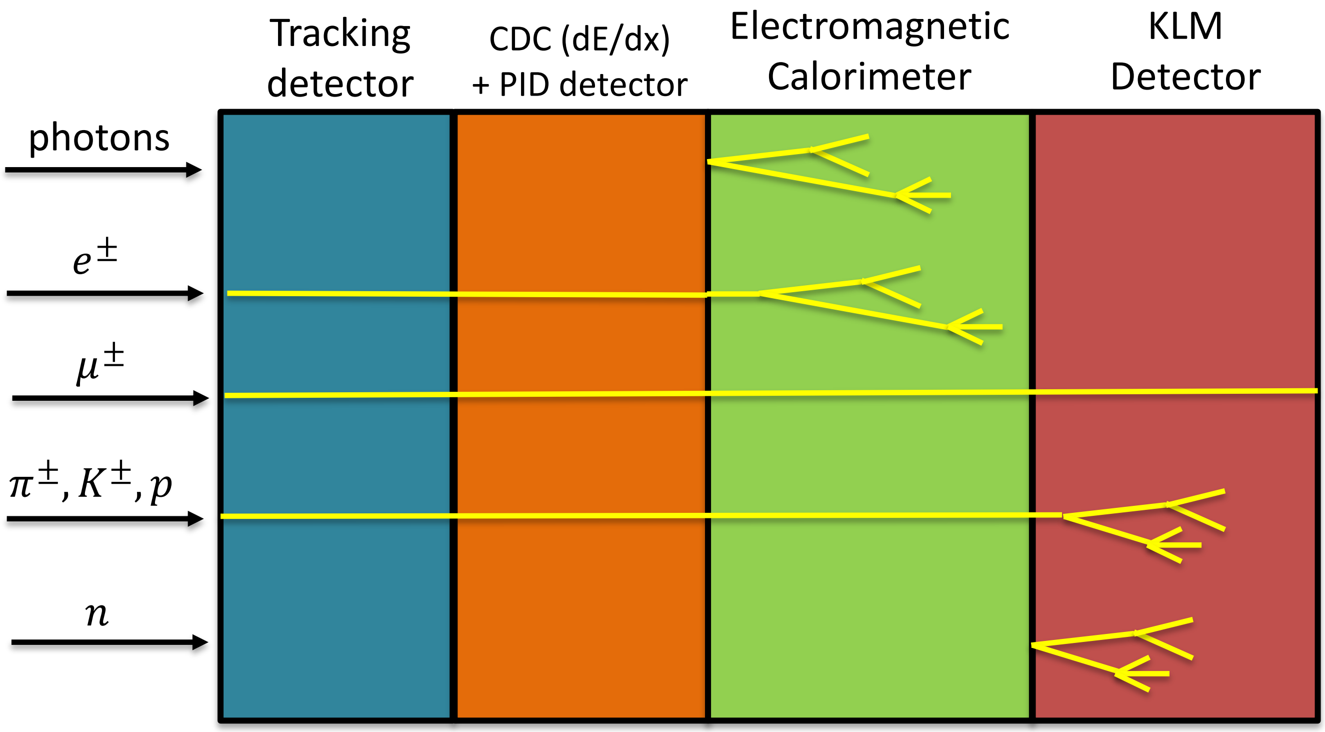


Trigger and DAQ

- Max L1 rate: 0.5 →30 kHz
- Pipeline readout

Better performance at the higher trigger rate and beam background

Introduction: Belle II detector



Combine the response from each sub-detector

- e vs π, K: CDC (dE/dx) and ECL mainly contribute
- μ vs π, K: CDC (dE/dx) and KLM mainly contribute

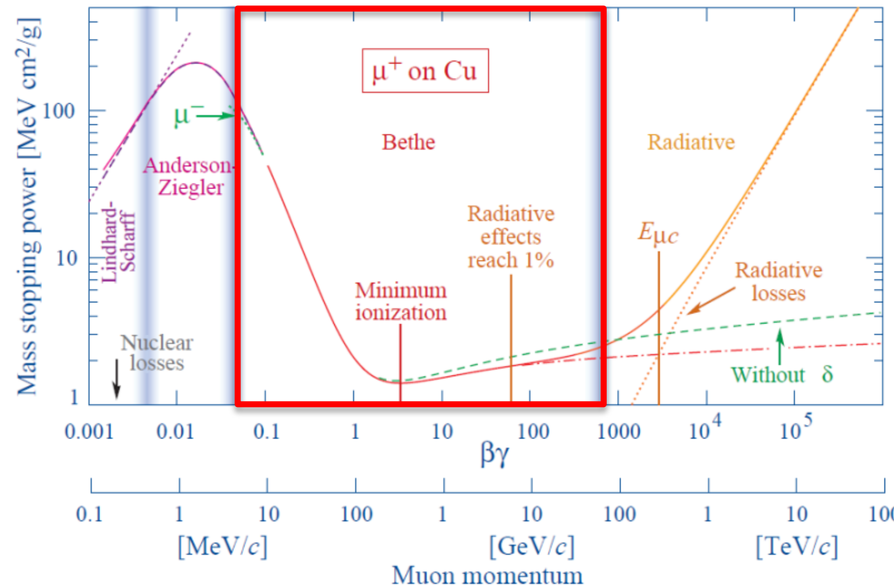
PID in CDC, ECL, KLM

PID in CDC: Energy loss

Bethe-Bloch equation

$$\left\langle -\frac{dE}{dx}\right\rangle = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_ec^2\beta^2\gamma^2W_{\max}}{I^2}-\beta^2-\frac{\delta(\beta\gamma)}{2}\right] \sim F(\beta)$$

W	energy transfer to an electron in a single collision	MeV
k	bremsstrahlung photon energy	MeV
z	charge number of incident particle	
Z	atomic number of absorber	
A	atomic mass of absorber	g mol^{-1}
K	$4\pi N_A r_e^2 m_e c^2$	0.307 075 MeV mol^{-1} cm^2
I	mean excitation energy	eV (*Nota bene!*)
$\delta(\beta\gamma)$	density effect correction to ionization energy loss	



Energy loss of heavy charged particles (except electrons)

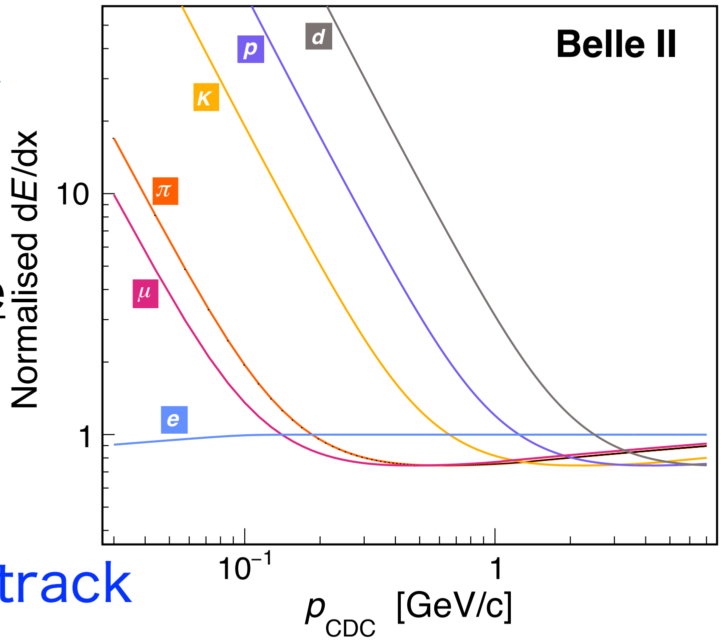
Energy loss generally depends on the velocity of the particle (β)

PID: dE/dx in CDC

$p = M\beta\gamma$ (M: mass of a particle)

- Different distribution for each particle

→ Useful for particle identification



dE/dx measurement in CDC on a given track

- Calculate a truncated mean from the signals from all sense wires
  - Discard lowest 5% and highest 25% of measurements
- Calculate a normalized deviation for each particle

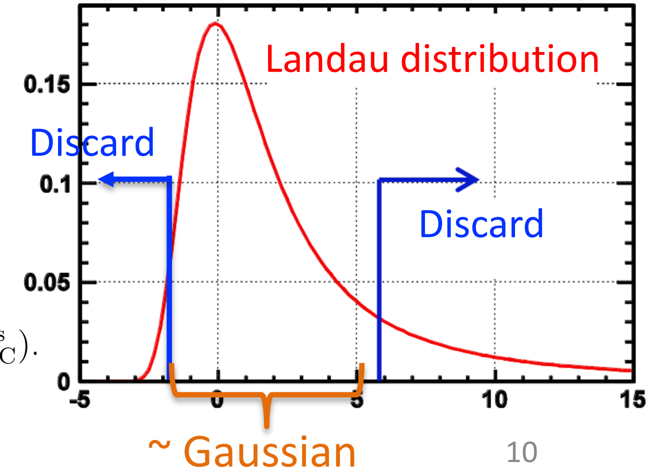
$$\chi_i = \frac{\Delta^i_{\mathrm{d}E/\mathrm{d}x}}{\sigma^i_{\mathrm{pred}}} = \frac{\boxed{\mathrm{d}E/\mathrm{d}x^i_{\mathrm{meas}}} - \boxed{\mathrm{d}E/\mathrm{d}x^i_{\mathrm{pred}}}}{\boxed{\sigma^i_{\mathrm{pred}}}}.$$

Measured truncated mean

Predicted truncated mean

Predicted resolution on $\Delta^i_{dE/dx}$

$$\sigma^i_{\mathrm{pred}}\left(\mathrm{d}E/\mathrm{d}x_{\mathrm{pred}}, \cos\theta, N^{\mathrm{hits}}_{\mathrm{CDC}}\right) = f^i_1(\mathrm{d}E/\mathrm{d}x_{\mathrm{pred}}) \times f^i_2(\cos\theta) \times f^i_3(N^{\mathrm{hits}}_{\mathrm{CDC}}).$$

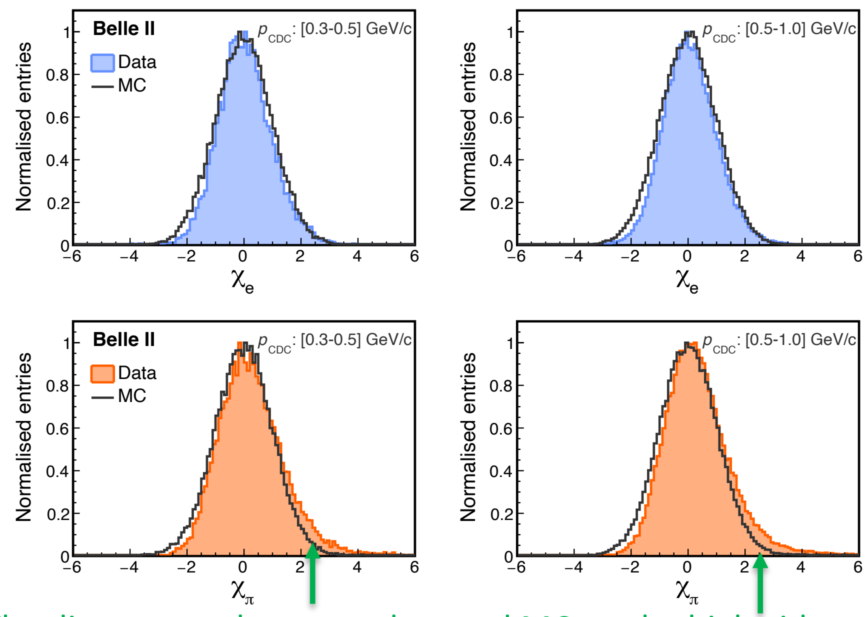


CDC likelihood

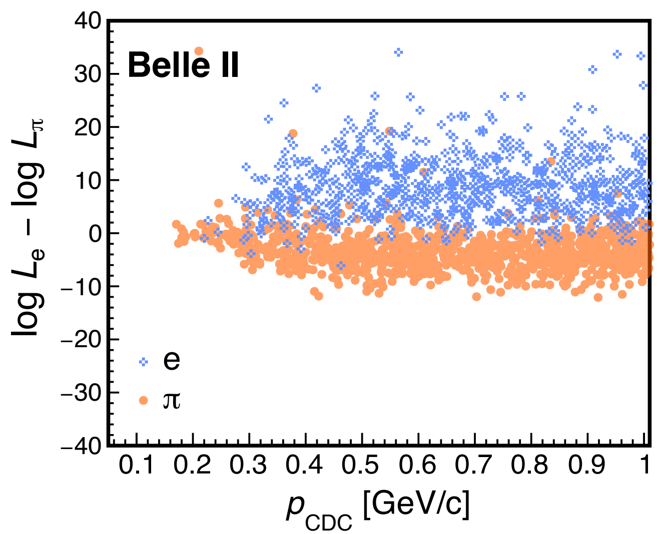
χ_i distribution ~ Gaussian

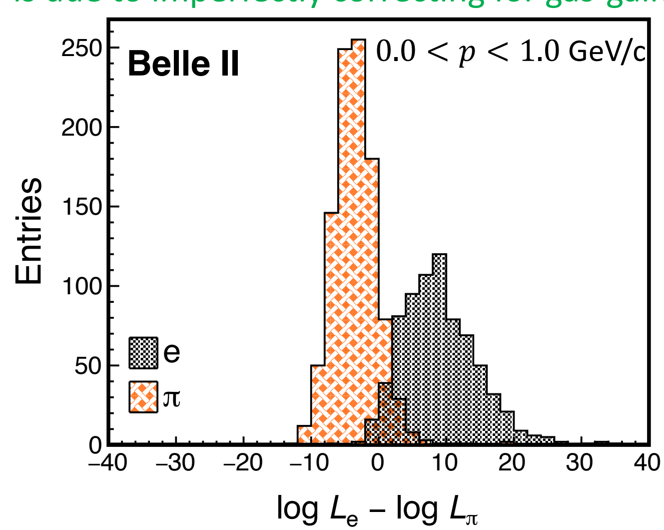
- Form log-likelihoods for the particle hypotheses from the χ_i

$$\log \mathcal{L}_i^{\mathrm{CDC}} = -\frac{\chi_i^2}{2},$$



The discrepancy between data and MC on the high-side tail is due to imperfectly correcting for gas-gain saturation



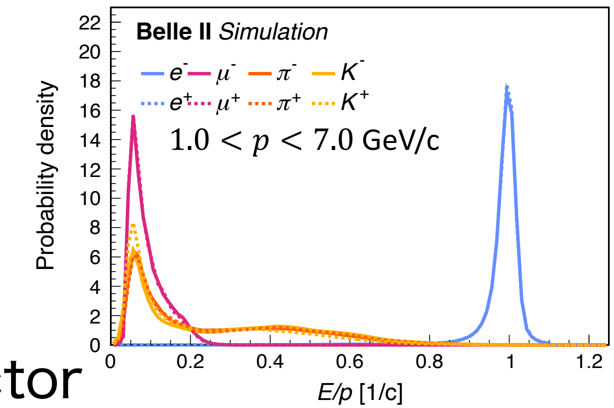


A good separation power between ℓ and π, K for low momentum

PID in ECL

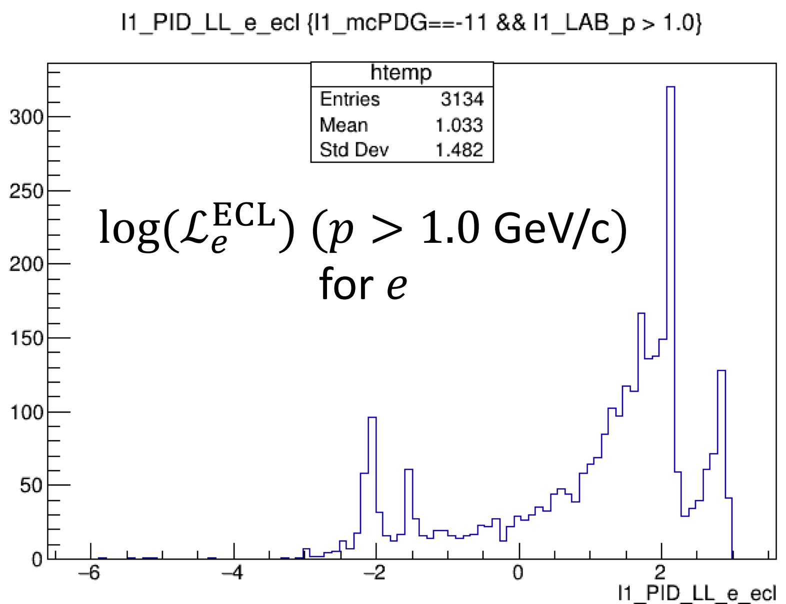
Use the ratio E/p for PID in the ECL

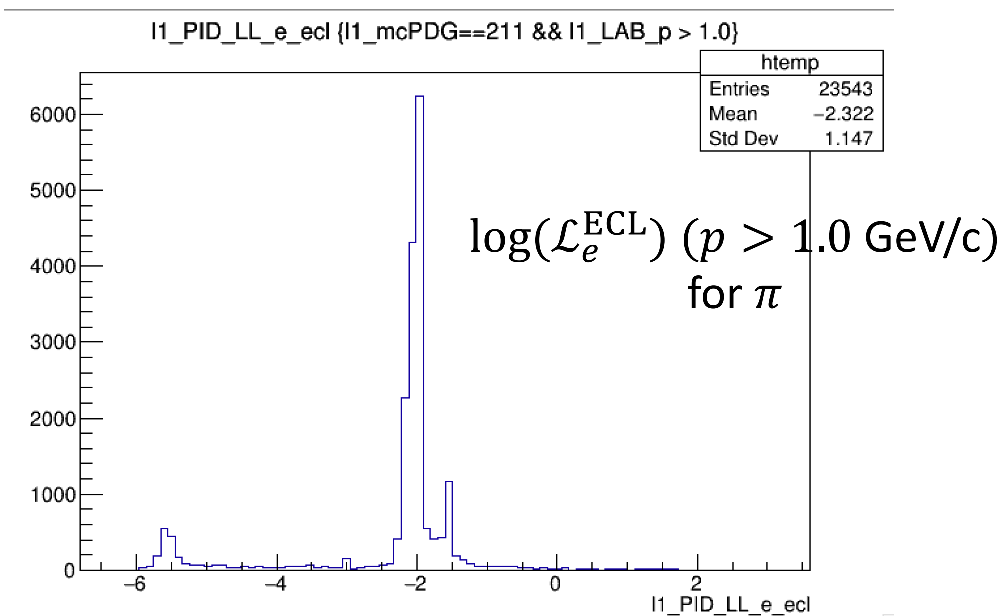
- Energy measured in the ECL
- Momentum reconstructed in tracking detector



→ Templates of the E/p distributions are constructed from single-particle simulations for each hypothesis

- Extract the likelihoods, $\mathcal{L}_i^{\mathrm{ECL}}$ based on the probability density function





PID in KLM

- Muon identification is simple: high penetration
  - No strong interaction → No interaction with hadrons.
  - No bremsstrahlung like electrons

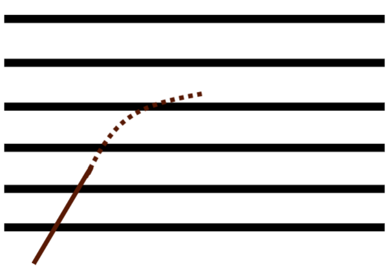
→ Difference in longitudinal penetration depth and transverse scattering

The KLM likelihoods per hypothesis i: $\mathcal{L}_i^{KLM} = \mathcal{L}_i^{KLM,L} \cdot \mathcal{L}_i^{KLM,T}$

The longitudinal components, $\mathcal{L}_i^{KLM,L}$

$$\mathcal{L}_{i,k}^{\mathrm{KLM},L} = \begin{cases} p_{i,k} & \text{if matched hit on } k\text{-th layer} \\ 1 - p_{i,k} \cdot \epsilon_k & \text{if no matched hit on } k\text{-th layer,} \end{cases}$$

$p_{i,k}$: Expected probability of finding a hit on k-th layer from simulation
ϵ_k: Measured detection efficiency of the layer.

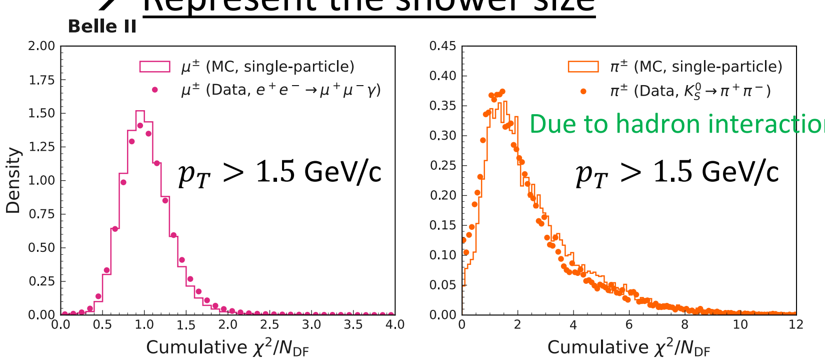


$$L = p_1 p_2 (1 - p_3 \varepsilon_3)\,(1 - p_4 \varepsilon_4)$$

The transverse components, $\mathcal{L}_i^{KLM,T}$

- Track fit quality χ^2 divided by NDF.
  - The track reconstructed in inner detectors is extrapolated to KLM

  → Represent the shower size



Combine info in all detectors

Each sub-detector provides a likelihood $\mathcal{L}_i^d$

- d: {SVD, CDC, TOP, ARICH, ECL, KLM}
- i: $\{e, \mu, \pi, K, p, d\}$

Eg. $\mathcal{L}_e^{\mathrm{ECL}}$: is related to probability(?) that the particle is an electron based on ECL info.

basf2 `pidLogLikelihoodValueExpert(pdgCode, detectorList)`

returns the log likelihood value of for a specific mass hypothesis and set of detectors.

Based on the likelihoods, we define "the global likelihood ratio"

$$p_i = \frac{\mathcal{L}_i}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}$$

$$\mathcal{L}_i = \prod_d^{d \in D} \mathcal{L}_i^d$$

we also define "the binary likelihood ratio"

$$p_{i,j} = \frac{\mathcal{L}_i}{\mathcal{L}_i + \mathcal{L}_j}$$

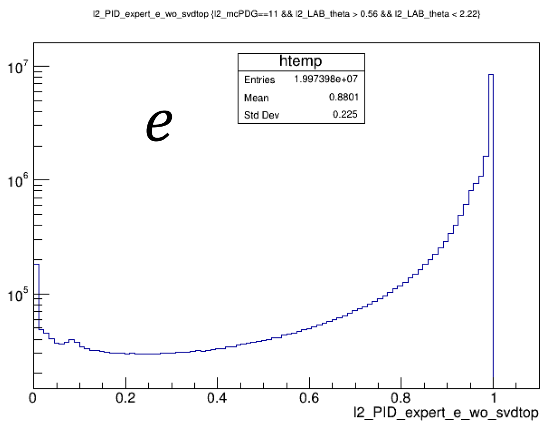
Global electron ID

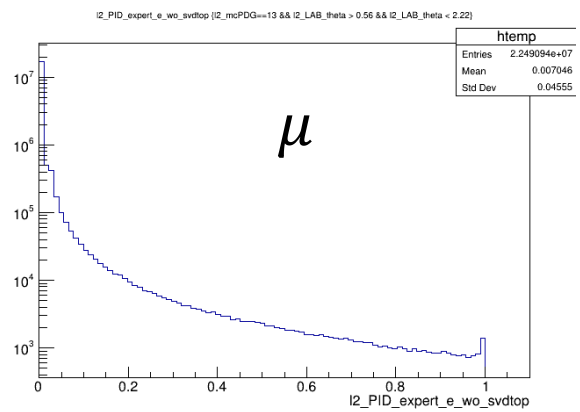
electronID_noSVD_noTOP:

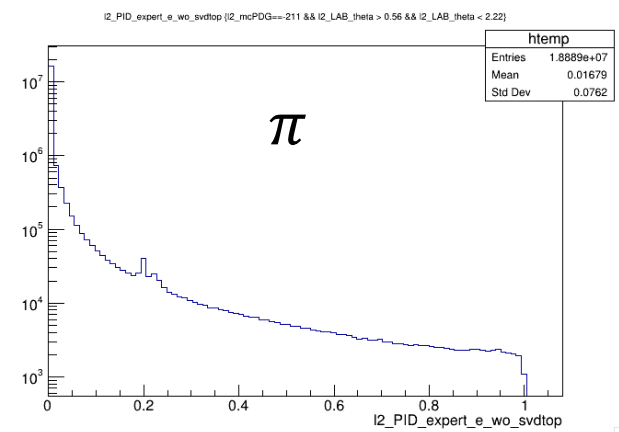
$$p_e = \frac{\mathcal{L}_e}{\mathcal{L}_e+\mathcal{L}_\mu+\mathcal{L}_\pi+\mathcal{L}_K+\mathcal{L}_p+\mathcal{L}_d}$$

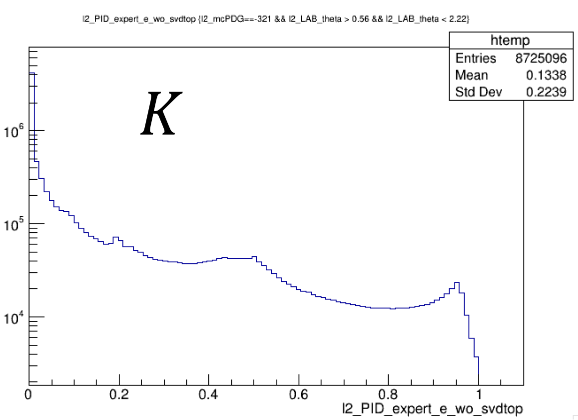
d: {CDC, ARICH, ECL, KLM}

$$\mathcal{L}_i = \prod_{d}^{d\in D} \mathcal{L}_i^d$$









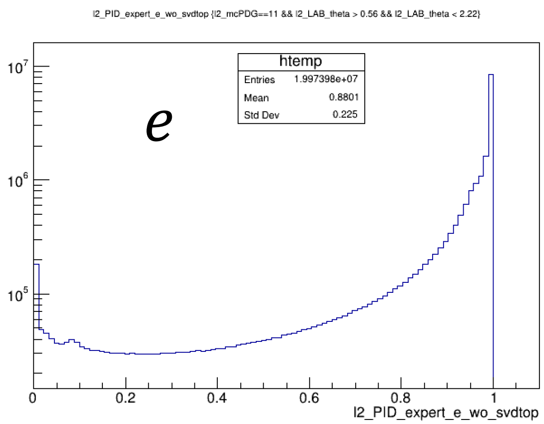
Global electron ID

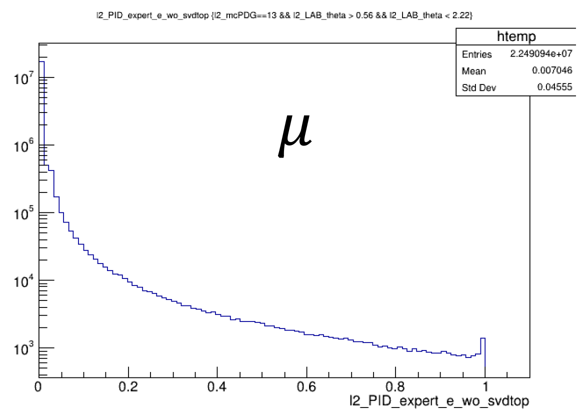
electronID_noSVD_noTOP:

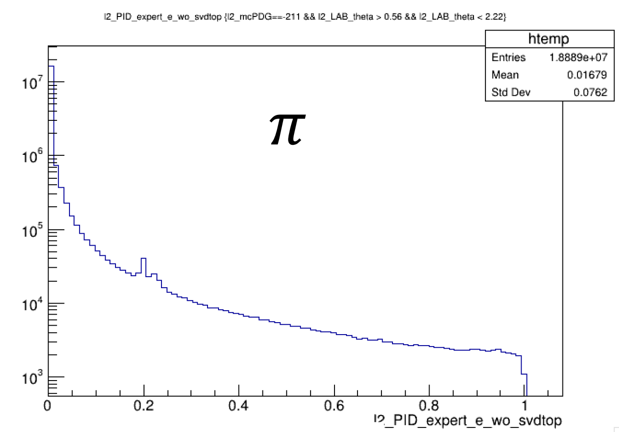
$$p_e = \frac{\mathcal{L}_e}{\mathcal{L}_e+\mathcal{L}_\mu+\mathcal{L}_\pi+\mathcal{L}_K+\mathcal{L}_p+\mathcal{L}_d}$$

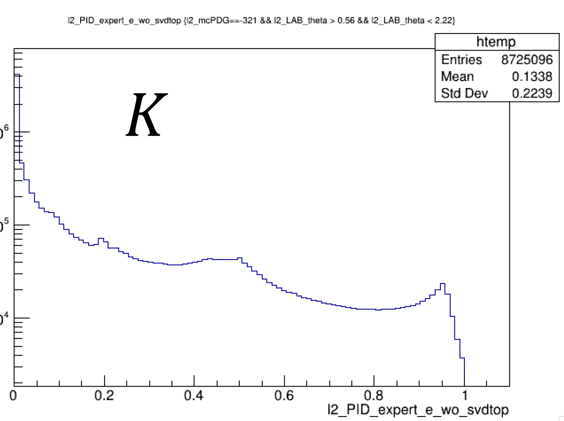
d: {CDC, ARICH, ECL, KLM}

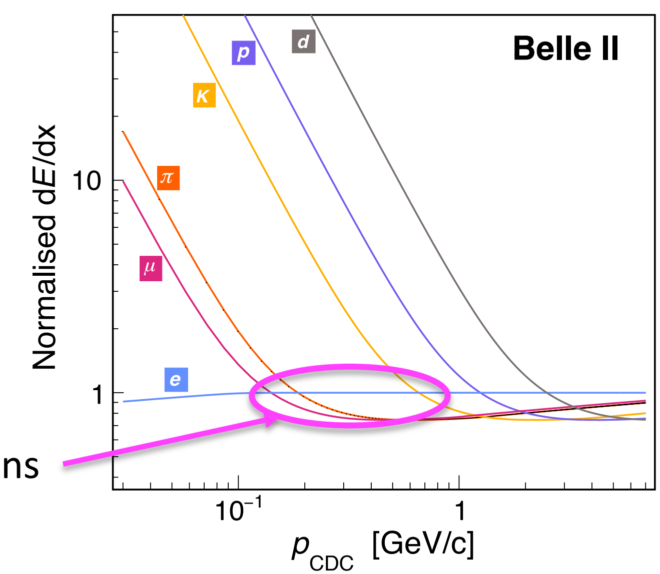
$$\mathcal{L}_i = \prod_{d}^{d\in D} \mathcal{L}_i^d$$











Difficult to distinguish electrons from pions/kaons

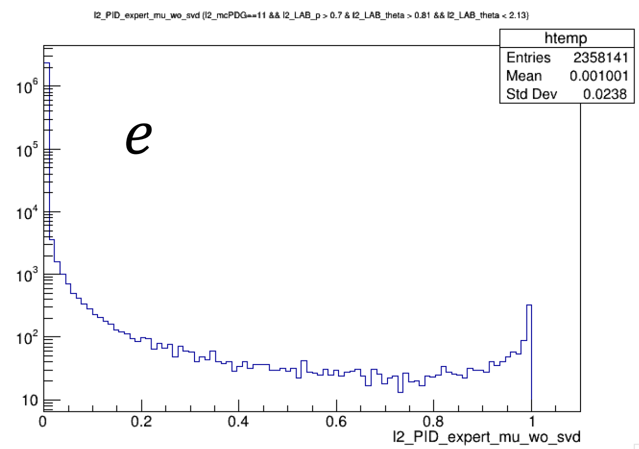
Global muon ID

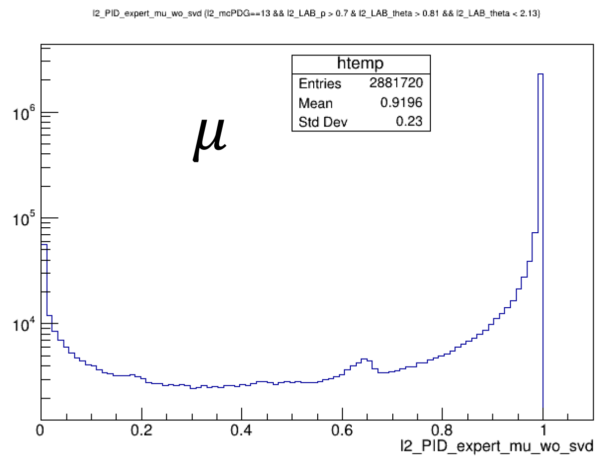
muonID_noSVD:

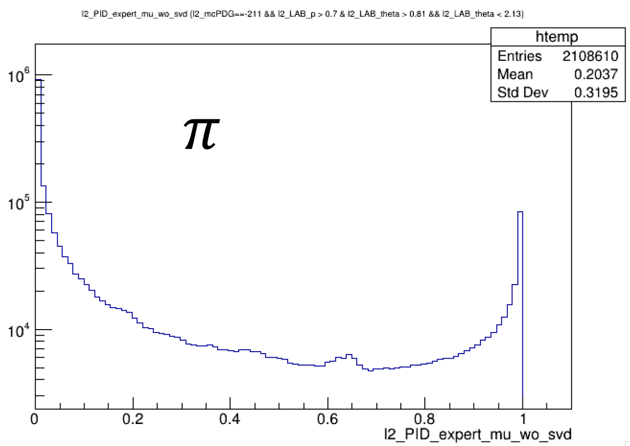
d: {CDC, TOP,ARICH, ECL, KLM}

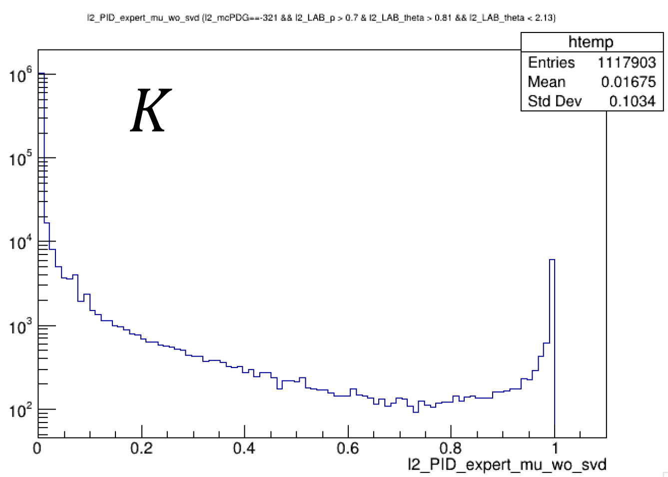
$$p_\mu = \frac{\mathcal{L}_\mu}{\mathcal{L}_e+\mathcal{L}_\mu+\mathcal{L}_\pi+\mathcal{L}_K+\mathcal{L}_p+\mathcal{L}_d}$$

$$\mathcal{L}_i = \prod_{d}^{d\in D} \mathcal{L}_i^d$$









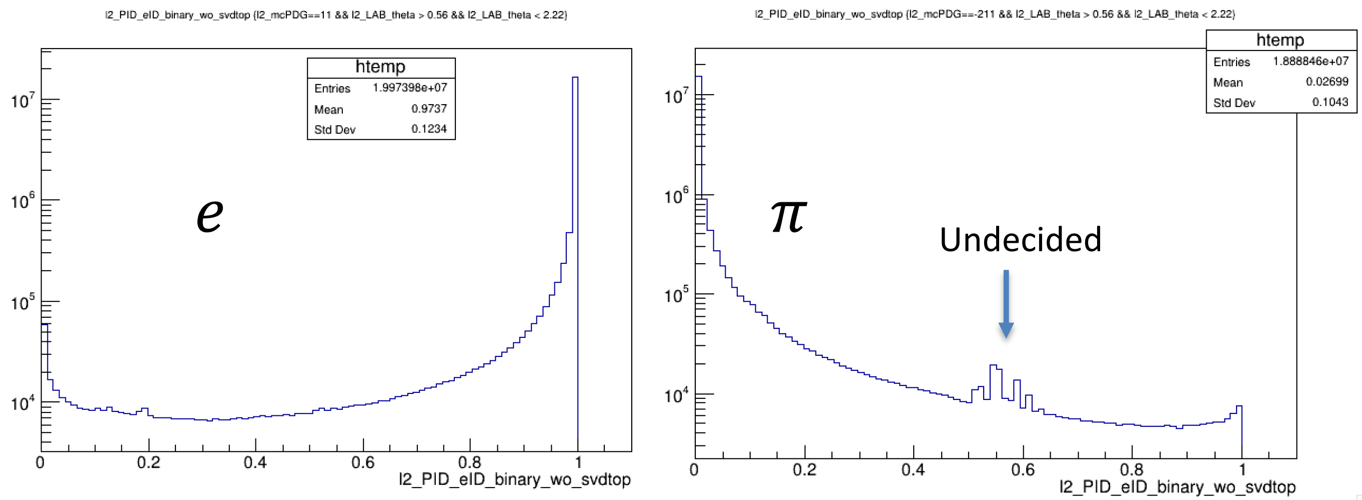
Binary lepton ID

Binary electronID_noSVD_noTOP:

$$p_e = \frac{\mathcal{L}_e}{\mathcal{L}_e + \mathcal{L}_\pi}$$

d: {CDC, ARICH, ECL, KLM}

$$\mathcal{L}_i = \prod_{d}^{d \in D} \mathcal{L}_i^d$$



Binary muonID_noSVD:

$$p_\mu = \frac{\mathcal{L}_\mu}{\mathcal{L}_\mu + \mathcal{L}_\pi}$$

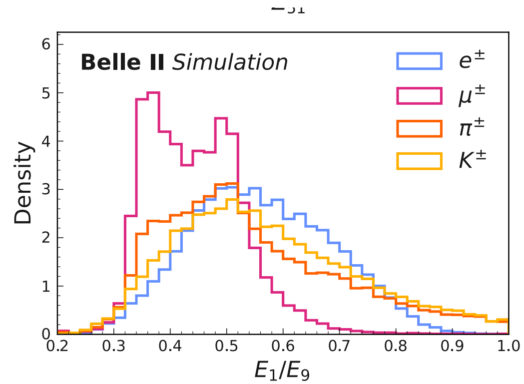
Multivariate approach

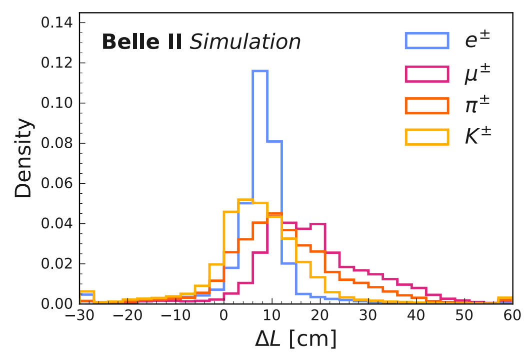
PID group provides BDT-based Lepton ID

- Use some ECL observables instead of ECL likelihood

→ Improve performance at $p < 1.0$ GeV/c by exploiting measurements that characterize ECL clusters

Variable	Description	Range
	Associated to the ECL clusters	
E/p [c]	Ratio of cluster energy over matching track momentum.	–
E_1/E_9	Ratio of the energy of the most energetic crystal over the energy sum of the 9 surrounding crystals.	–
E_9/E_{21}	Ratio of the energy sum of 9 crystals surrounding the one with maximum energy over the energy sum of the 25 surrounding crystals, minus 4 corners.	–
ΔL [cm]	Projection on the extrapolated track direction of the distance between the track entry point in the ECL and the cluster centroid [19].	–
LAT	Lateral moment of the electromagnetic shower [20].	–
$\|Z_{nm}\|$	2 Zernike moments $((n,m) \in \{(4,0),(5,1)\}$ calculated in a plane orthogonal to the shower direction [21].	–
	Associated to the non-ECL likelihoods (binary BDT)	
$\Delta \log \mathcal{L}(\ell/\pi)_{\mathrm{CDC}}$	Log-likelihood $\ell - \pi$ difference in the CDC	–
$\Delta \log \mathcal{L}(\ell/\pi)_{\mathrm{TOP}}$	Log-likelihood $\ell - \pi$ difference in the TOP	Barrel
$\Delta \log \mathcal{L}(\ell/\pi)_{\mathrm{ARICH}}$	Log-likelihood $\ell - \pi$ difference in the ARICH	FWD endcap
$\Delta \log \mathcal{L}(\mu/\pi)_{\mathrm{KLM}}$	Log-likelihood $\ell - \pi$ difference in the KLM	$p > 0.6$ GeV/c
	Associated to the non-ECL likelihoods (multiclass BDT)	
P_ℓ^{CDC}	Global normalised likelihood in the CDC.	–
P_ℓ^{TOP}	Global normalised likelihood in the TOP.	Barrel
P_ℓ^{ARICH}	Global normalised likelihood in the ARICH.	FWD endcap
P_ℓ^{KLM}	Global normalised likelihood in the KLM.	$p > 0.6$ GeV/c



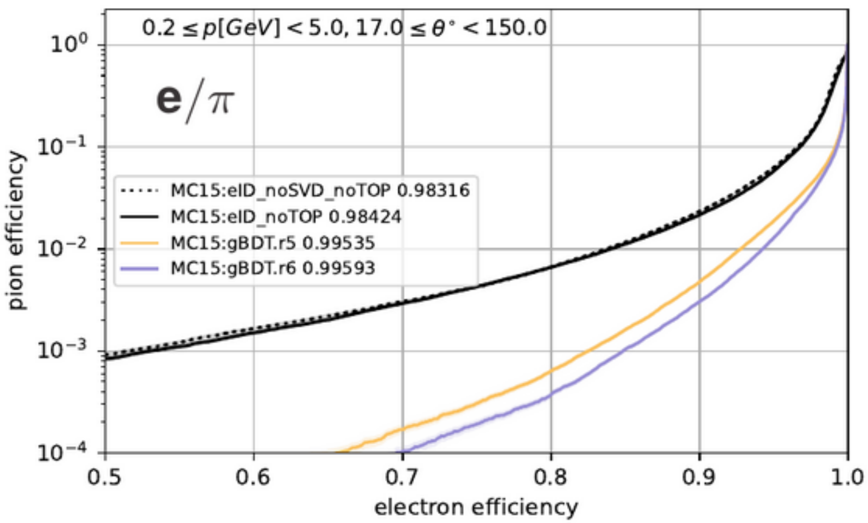


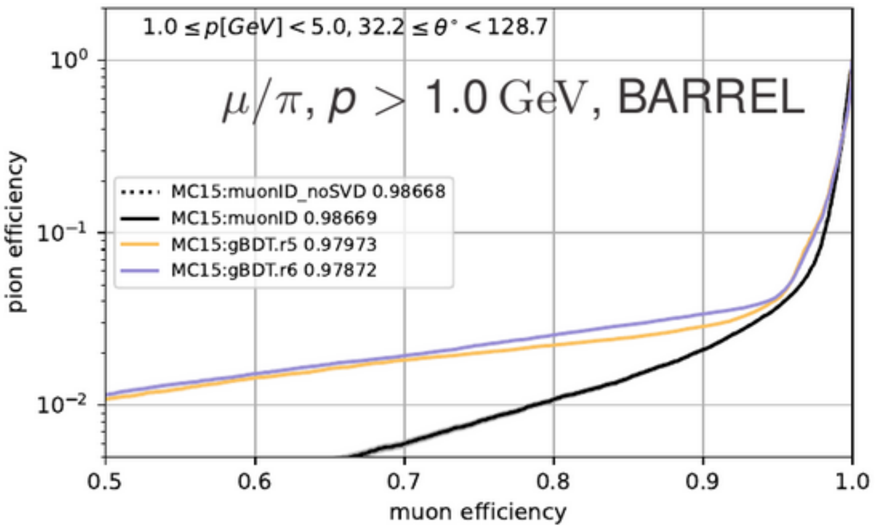
BDT LID: Performance

BDT Lepton ID gives better performance

- Electron ID: excellent improvement by utilizing ECL observables

→ Now, PID official recommendation is to use BDT-based electronID





LID efficiency calibration

LID efficiency correction

Lepton ID is an important item for high precision measurement

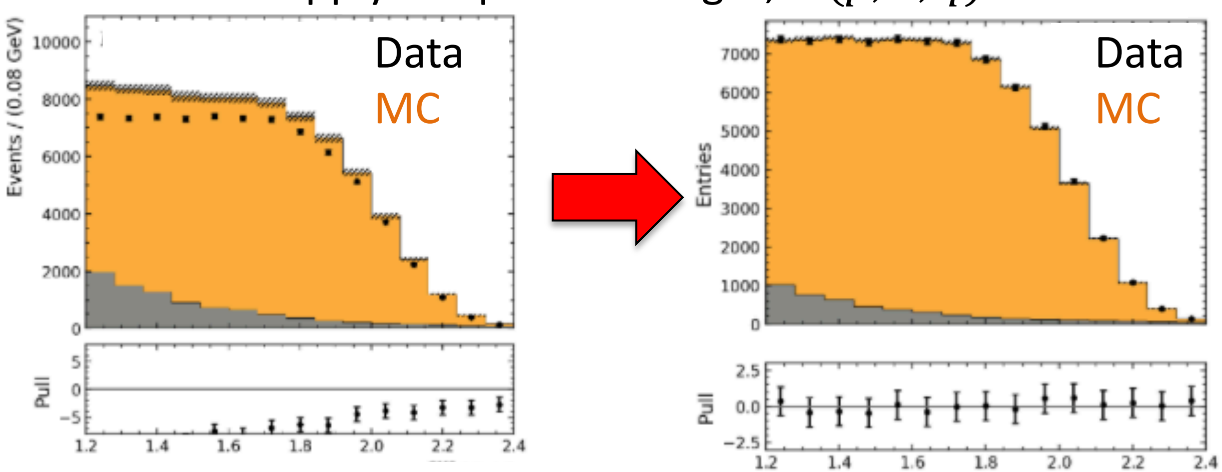
- Require good performance → less fake backgrounds ($\pi \to \ell, K \to \ell$)
- <u>**Require corrections (Data/MC) with small uncertainties**</u>
  - Reconstruction efficiency could be different from the data

    (MC simulation is not perfect..)

  → Essential to correct the difference for a precious measurement

Example.

Apply a Lepton ID weight, $w(p, \theta, q)$ to MC



Muon momentum [GeV]

Calibration channel

Evaluate LID efficiency/fake rate using several channels

Electron ID

- $ee \to (ee)ee$: momentum coverage is $p < 2.5$ GeV/c
- J/$\psi \to ee$: momentum coverage is $1.0 < p < 3.0$ GeV/c
- $ee \to ee(\gamma)$: momentum coverage is $1.0 < p$ GeV/c

Possible to cover full p range

Muon ID

- $ee \to (ee)\mu\mu$: momentum coverage is $p < 2.5$ GeV/c
- J/$\psi \to \mu\mu$: momentum coverage is $1.0 < p < 3.0$ GeV/c
- $ee \to \mu\mu$: momentum coverage is $1.0 < p$ GeV/c

Possible to cover full p range

$\boldsymbol{\pi \to \ell, K \to \ell}$ mis-ID

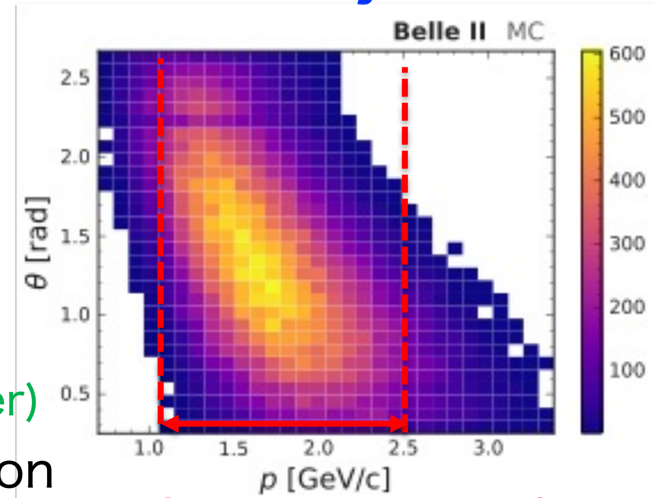
- $\mathrm{K_S^0} \to \pi\pi$: momentum coverage is $p < 1.0$ GeV/c
- $ee \to \tau(1p)\tau(3p)$: momentum coverage is $p < 1.0$ GeV/c
- $\mathrm{D}^{*+} \to (D^0 \to K^-\pi^+)\pi^+$ for $K \to \ell$ fake rates

J/ψ channel

Reconstruct J/ψ candidate from hadronic B decays

- Two tracks as originating from the IP
  - At least one track have a value of ℓID > 0.9
- $2.8 < M_{\ell^+\ell^-} < 3.3$ GeV/c^2
- Some requirements for bkg suppression

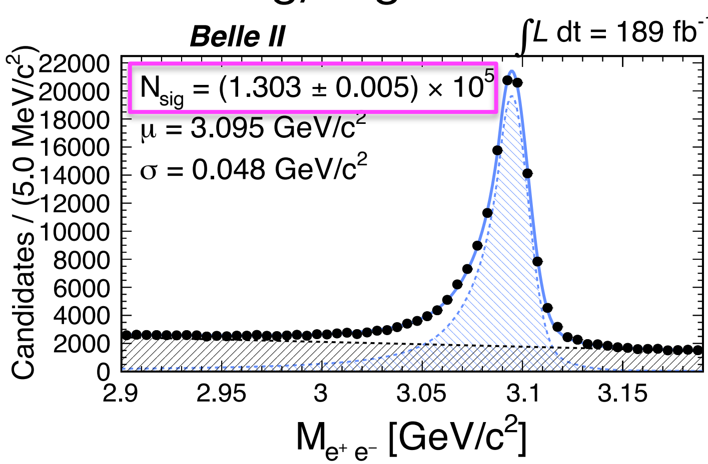
(eg. $R_2 < 0.4$, matching ECL cluster)

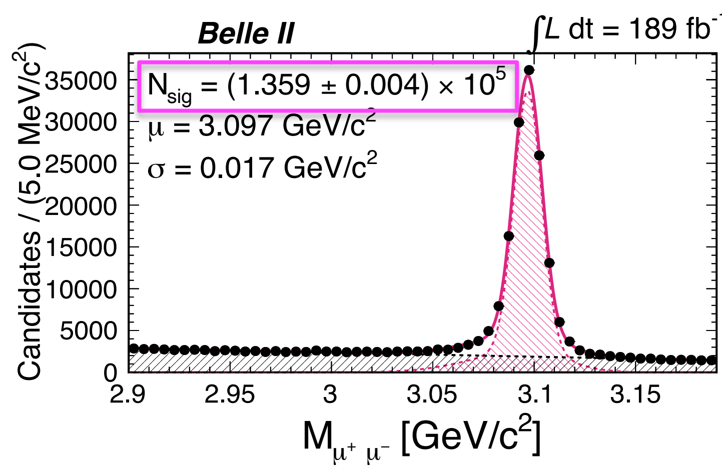


Coverage: $1.0 < p < 2.5$ GeV/c

→ Perform a binned likelihood fit to $M_{\ell^+\ell^-}$ distribution

Prepare suitable PDFs for sig/bkg





Extract the number of of J/ψ candidates based on the fit result

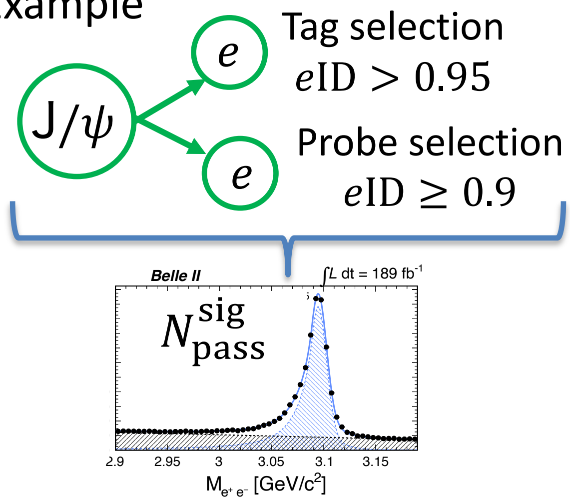
J/ψ channel

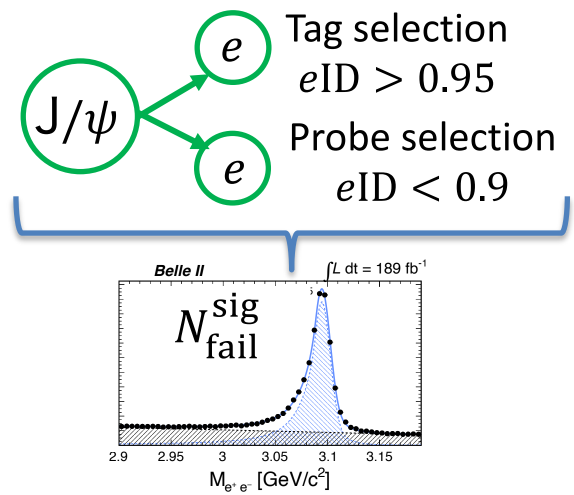
Tag-and-probe method is used to determine the LID efficiency

LID efficiency (ϵ)

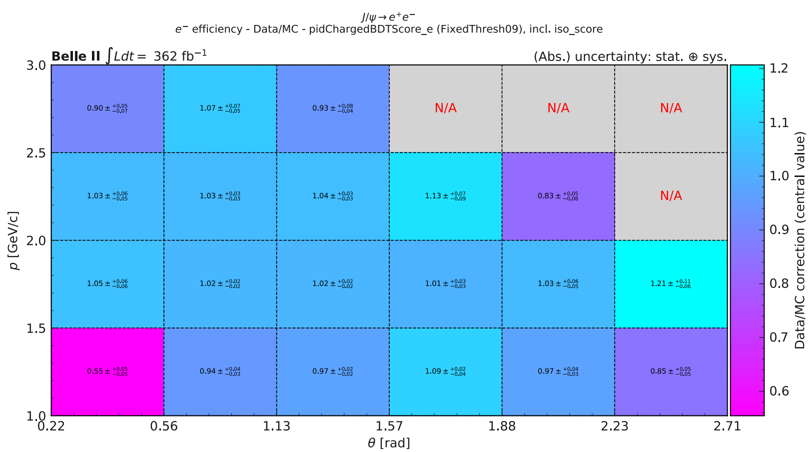
$$\varepsilon = \frac{N_{\text{pass}}^{\text{sig}}}{N_{\text{pass}}^{\text{sig}} + N_{\text{fail}}^{\text{sig}}},$$

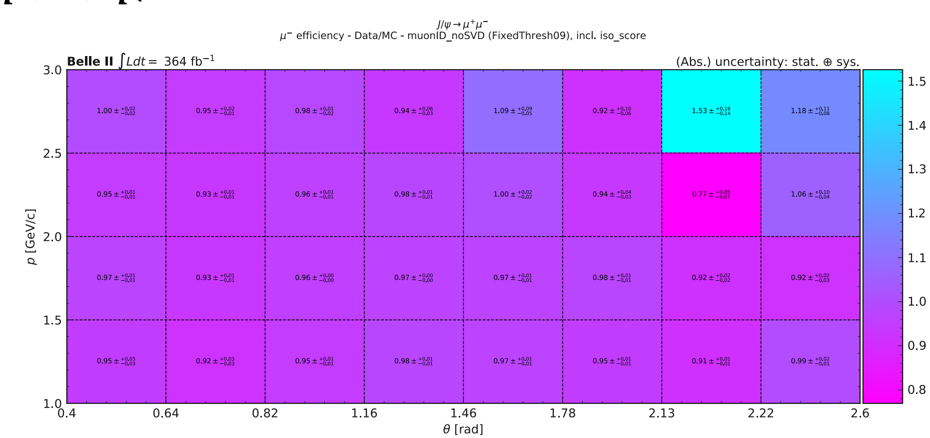
Example





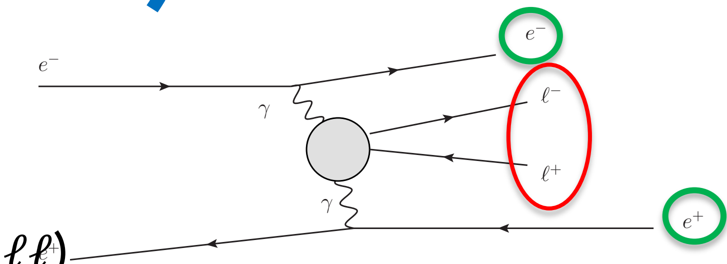
- Simultaneous fit is performed over the two "pass" and "fail" sets
- Calculate efficiencies for each (p, θ, q) bin





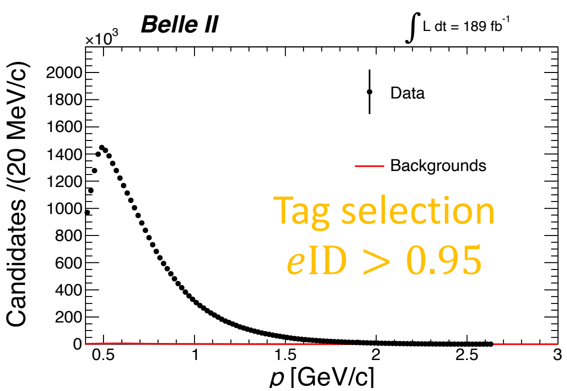
$ee \to (ee)\ell\ell$ (di-photon) channel

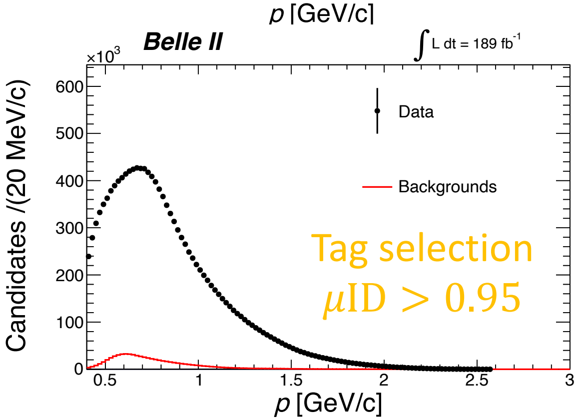
- Enough statistics ($p < 3.0$ GeV/c)
- $\ell^+\ell^-$ within the detector volume acceptance
- Suppress backgrounds ($ee \to qq, \tau\tau, ee\pi\pi, eeKK, \ell\ell$)
  - Eg. Total visible energy, p_T balance

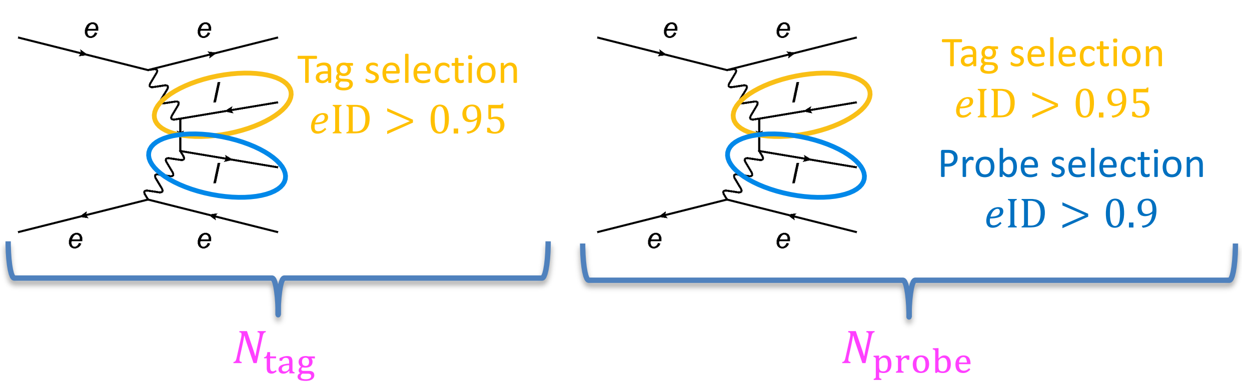


Escape into the beam pipe

Tag-and-probe method is used to determine the LID efficiency



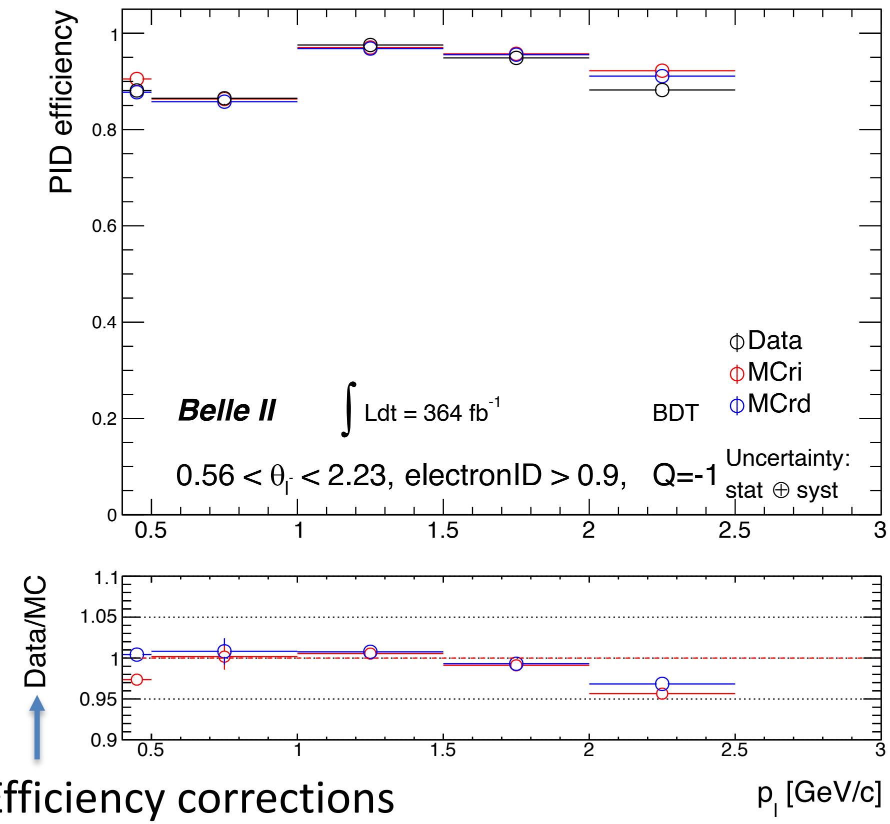


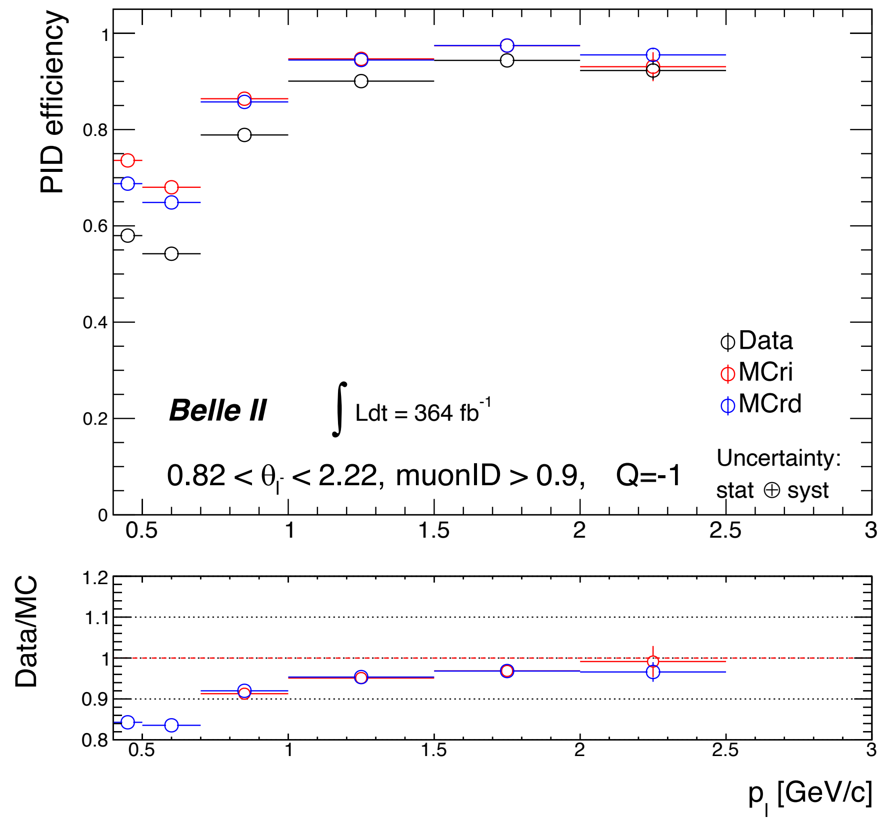


$$\varepsilon_{data} = \frac{N_\mathrm{probe} - N^\mathrm{bkg}_\mathrm{probe}}{N_\mathrm{tag} - N^\mathrm{bkg}_\mathrm{tag}} \qquad \varepsilon_{MC} = \frac{N^\ell_\mathrm{probe}}{N^\ell_\mathrm{tag}}$$

Bkg subtraction

$ee \to (ee)\ell\ell$ (di-photon) channel



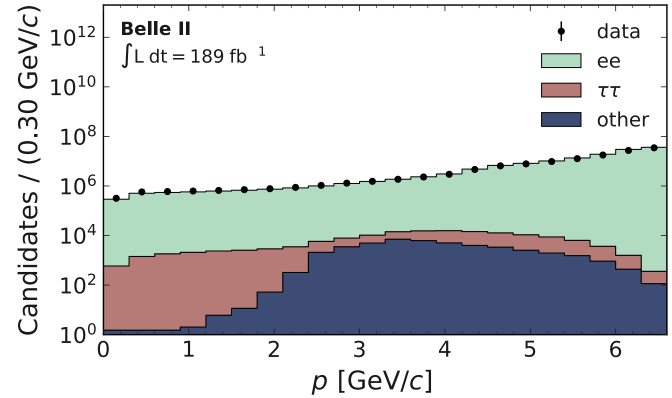


Efficiency corrections

- Electron ID: a good agreement between data and MC15rd
- Muon ID: Some discrepancy between data and MC15rd
  - Low momentum → CDC dE/dx calibration? Still under investigation

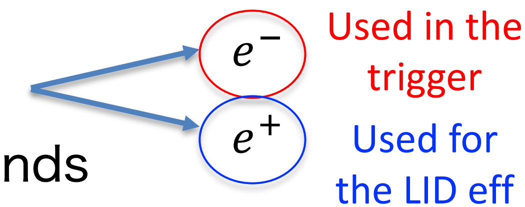
$ee \to ee(\gamma)$: Bhabha channel

- Large events across full momentum range
  - High momentum: $ee \to ee$
  - Low momentum: $ee \to ee\gamma$



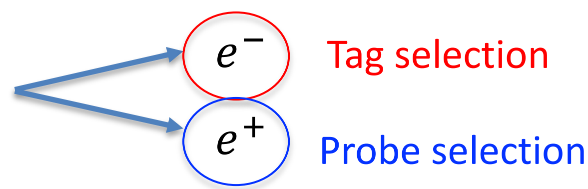
<u>Event selection</u>

- Use trigger requiring an ECL cluster (>2 GeV)
  - Probe the track, which is not used in the trigger
- Require the recoil mass M^2_{recoil} to suppress backgrounds



Tag-and-probe method

$$\epsilon_{data} = \frac{\rho_{tag+probe}\ N_{probe+tag}}{\rho_{tag}\ N_{tag}} \qquad \epsilon_{MC} = \frac{N_{probe+tag}}{N_{tag}}$$

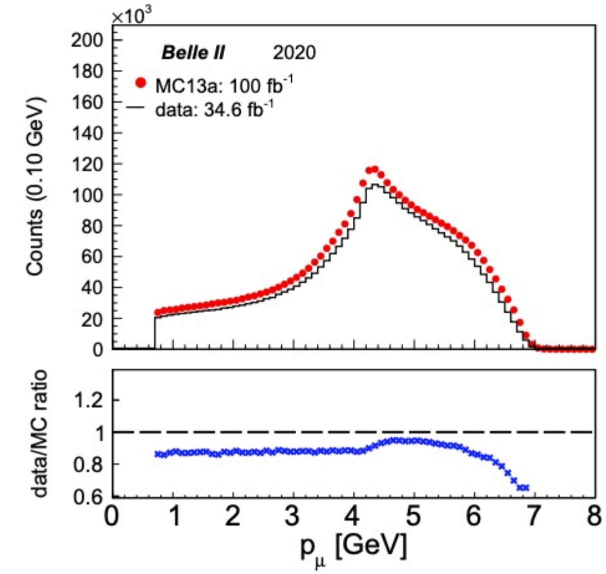


ρ: purities with $eeee, ee\mu\mu, \mu\mu, \tau\tau$ MC samples

In this B2GM, Phillip reported a recent progress in this channel.
Please check his slide [Link]

$ee \to \mu\mu\gamma$: di-muon channel

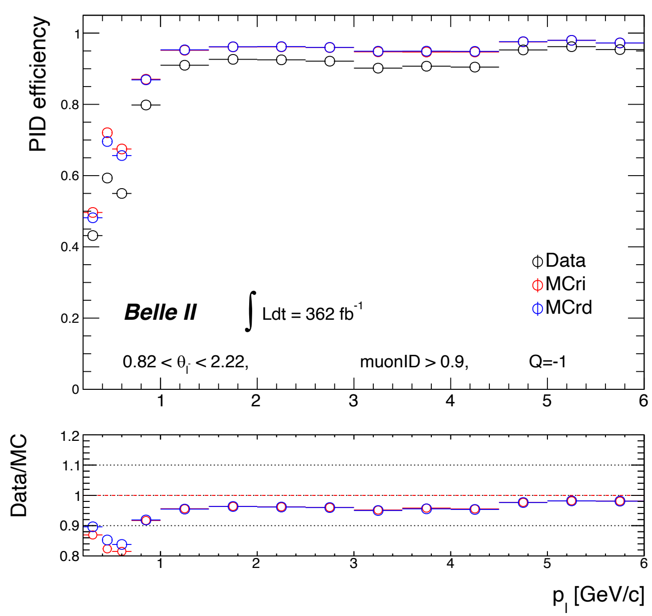
- Large events across full momentum range
- Clean environment
  - (Almost) no backgrounds

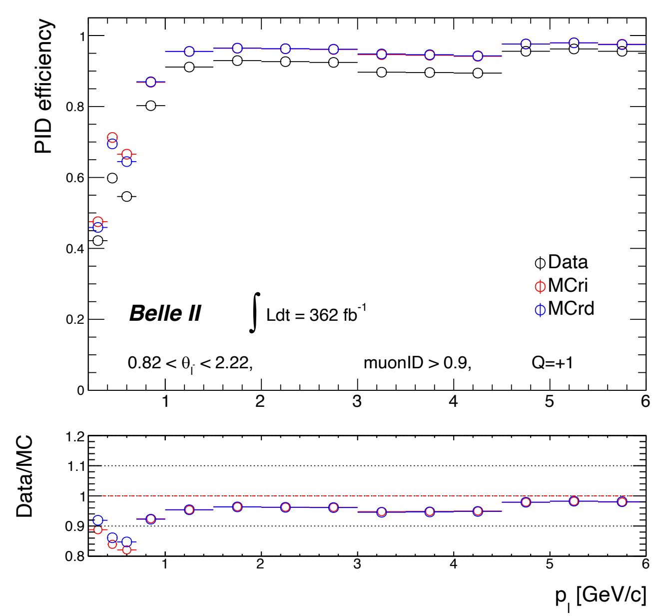


Tag-and-probe method

$$\varepsilon_{data} = \frac{N_{\rm probe} - N^{\rm bkg}_{\rm probe}}{N_{\rm tag} - N^{\rm bkg}_{\rm tag}} \qquad \varepsilon_{MC} = \frac{N^{\ell}_{\rm probe}}{N^{\ell}_{\rm tag}}$$

Bkg subtraction (eg. $ee \to \pi\pi\gamma, KK\gamma$)





$K_S \to \pi\pi$ channel

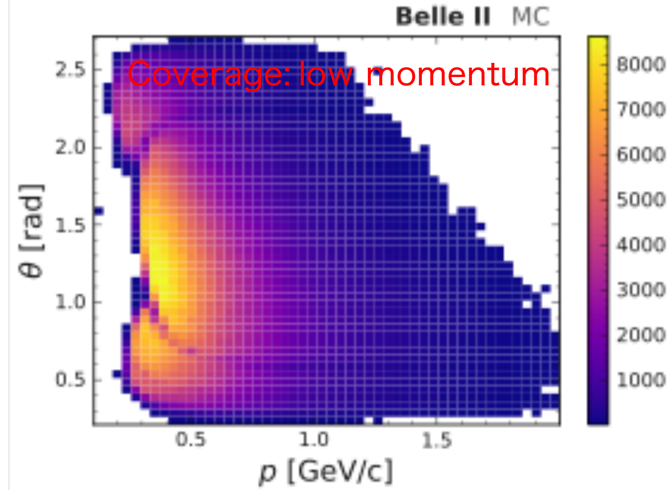
Methodology is similar as $J/\psi \to \ell\ell$

Inclusively reconstructed $K_S \to \pi\pi$ candidates are used

- Two tracks as originating from the IP
- $0.45 < M_{\pi^+\pi^-} < 0.55$ GeV/c^2
- Perform a vertex fit

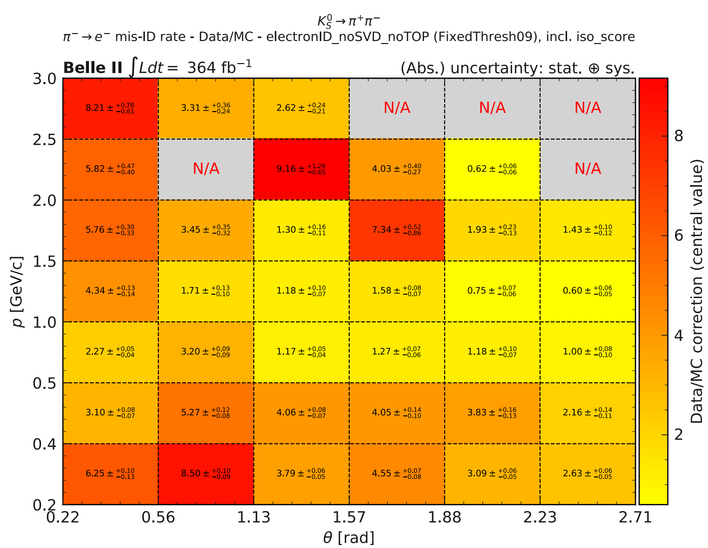
→ High signal purity (99.4%) of this channel

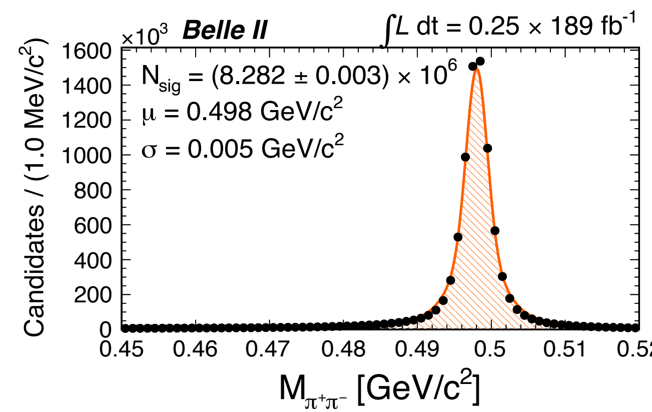
→ No need a tag selection



Misidentification rate: cut-and-count approach

- Compare N_{K_S} yields with and without ℓID cut

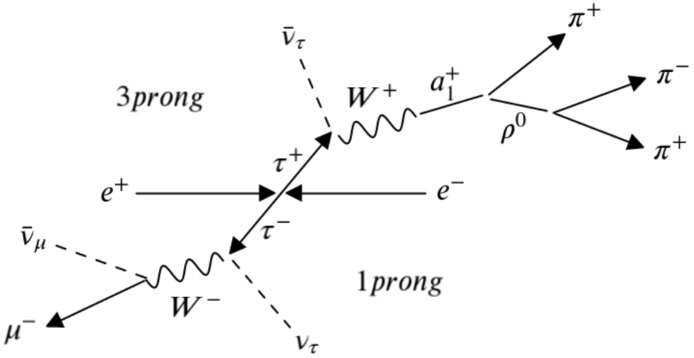




$ee \to \tau(1p)\tau(3p)$ channel

A clean sample of pions to measure pion misidentification probability

- $\tau \to \pi\pi\pi\nu$ (3-prong) and $\tau \to \ell\nu\nu$ (1-prong)
  - 98% of 3-prong tracks: pions (High purity)

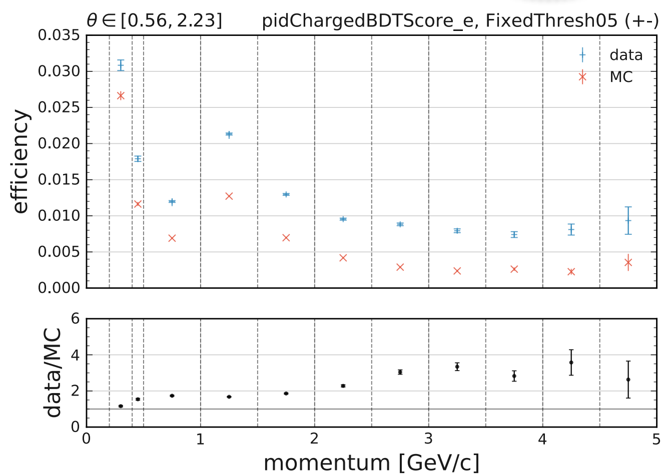


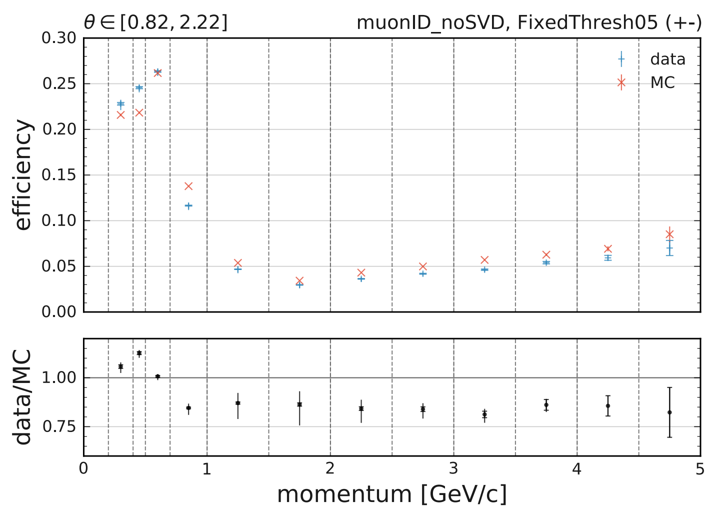
Event selection

- Take 1+3 charged track events
- Suppress backgrounds: Thrust, visible energy..

Bkg subtraction

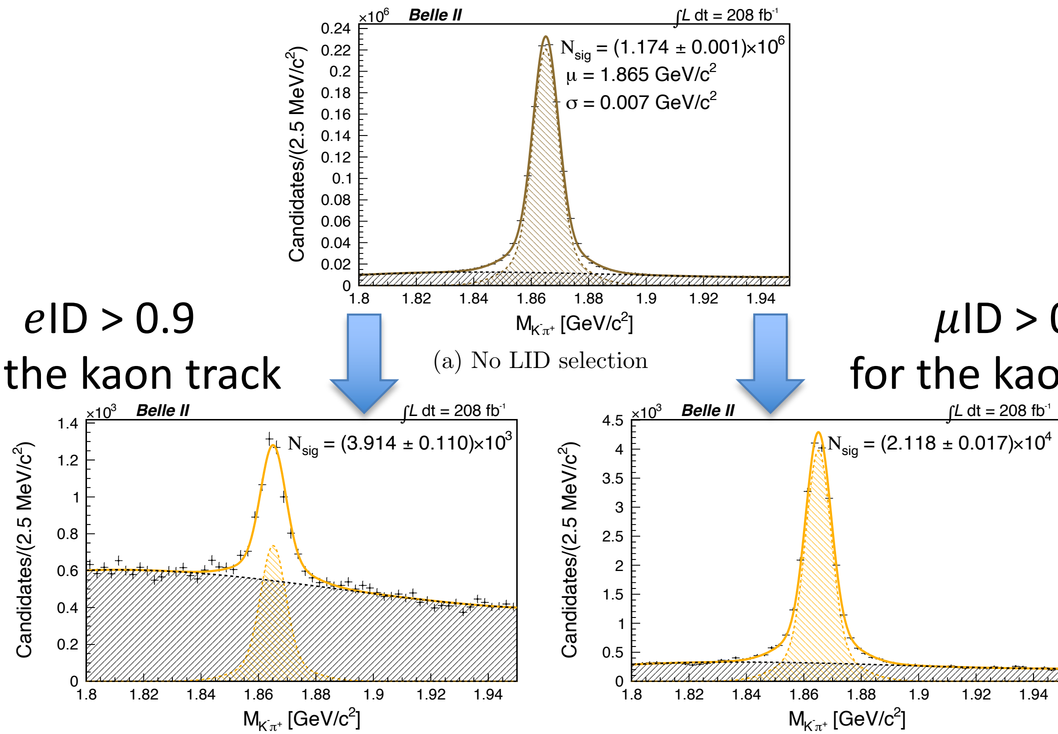
$$\varepsilon_{data} = \frac{N_{\mathrm{probe}} - N^{\mathrm{bkg}}_{\mathrm{probe}}}{N_{\mathrm{tag}} - N^{\mathrm{bkg}}_{\mathrm{tag}}} \qquad \varepsilon_{MC} = \frac{N^{\ell}_{\mathrm{probe}}}{N^{\ell}_{\mathrm{tag}}}$$





$D^{*+} \to D^0[\to K^-\pi^+]\pi^+$ channel

- Measure kaon-to-lepton misidentification probabilities
- D^{*+} mesons are produced in $e^+e^- \to c\bar{c}$ continuum events
  - Use all the data (on- and off-resonance data)

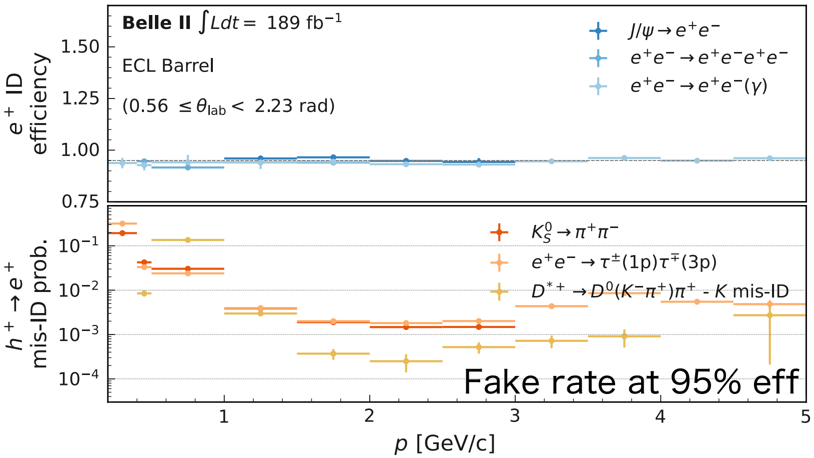


(a) No LID selection

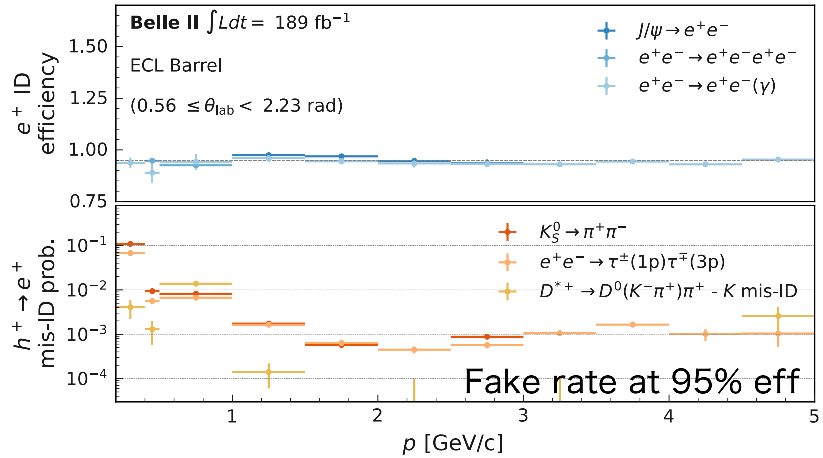
(b) $P_e > 0.9$ (c) $P_\mu > 0.9$

Figure 19: Fit to the M_{D^0} distribution in the $D^{*+} \to D^0[\to K^-\pi^+]\pi^+$ channel without any probe selection criteria (top), with $P_e > 0.9$ for the kaon track (bottom left) and with $P_\mu > 0.9$ for the kaon track (bottom right).

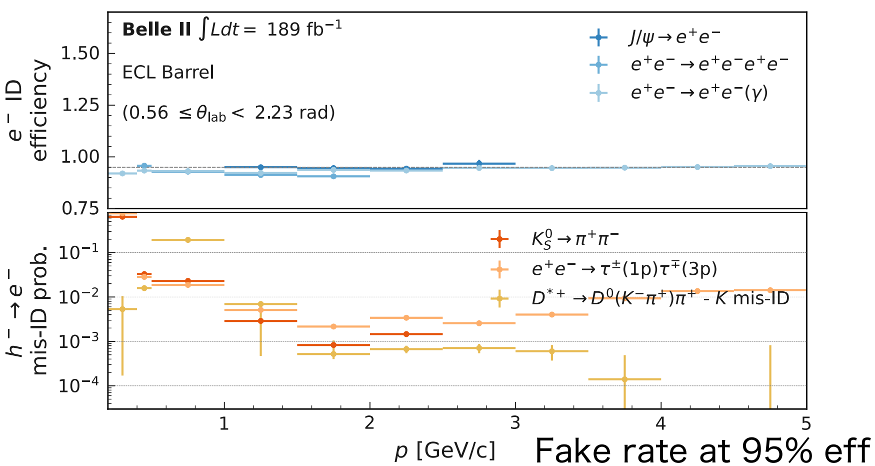
Results: electron ID



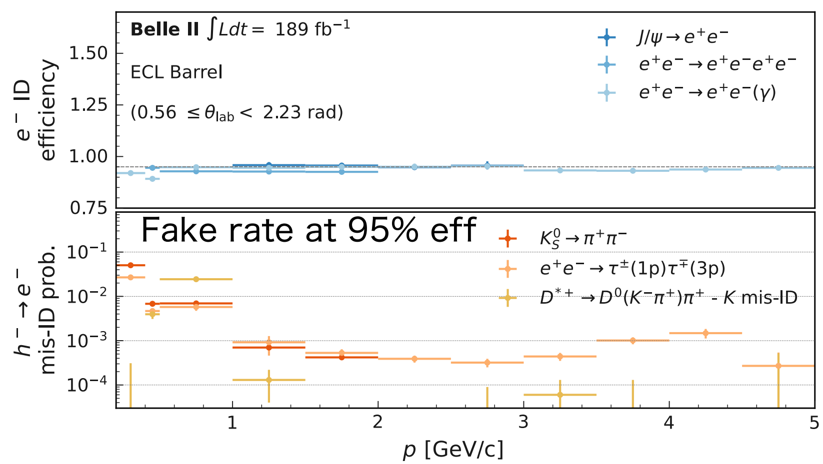
(a) e^+, likelihood-based LID.



(b) e^+, BDT-based LID.

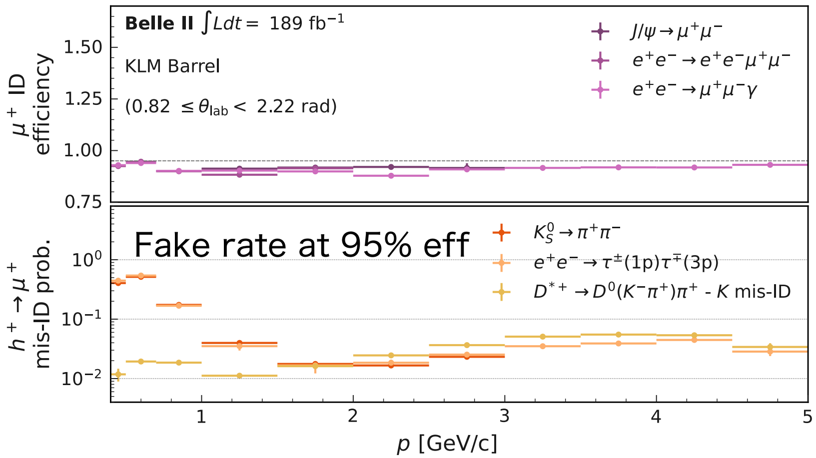


(c) e^-, likelihood-based LID.

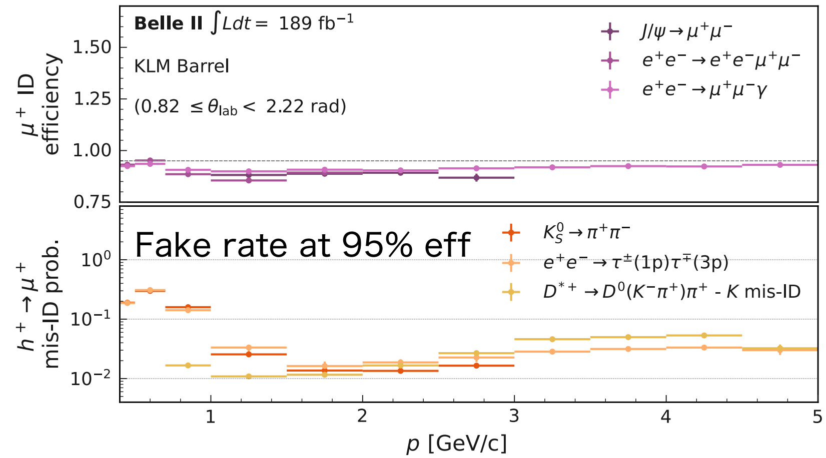


(d) e^-, BDT-based LID.

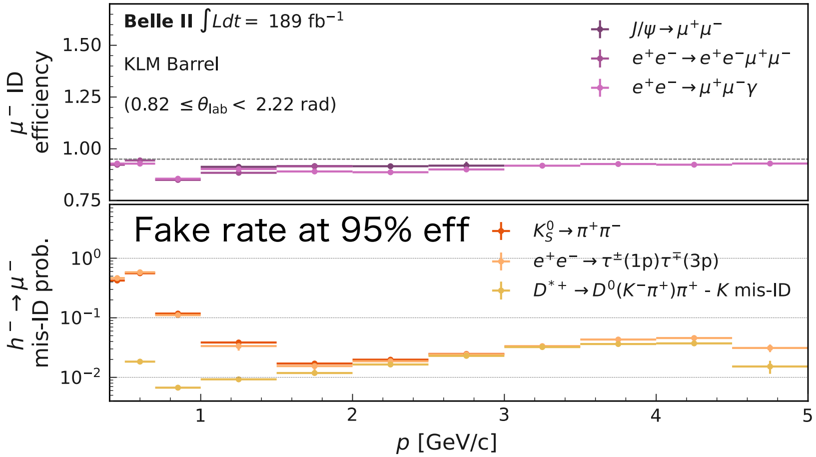
Results: muon ID



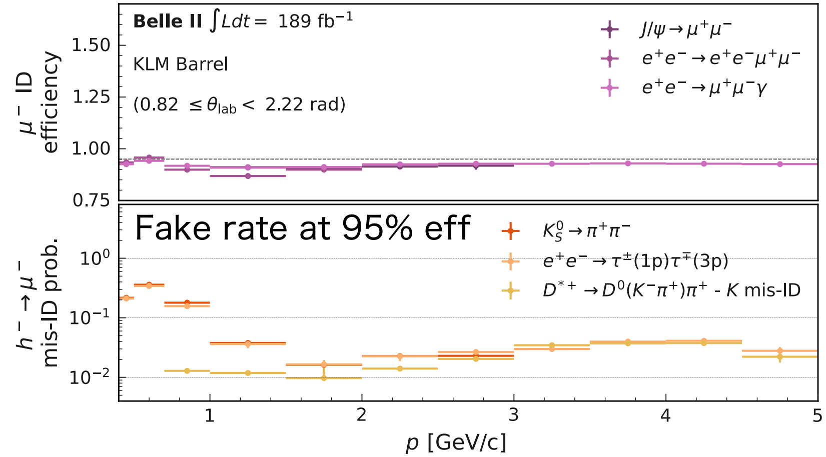
(a) μ^+, likelihood-based LID.



(b) μ^+, BDT-based LID.

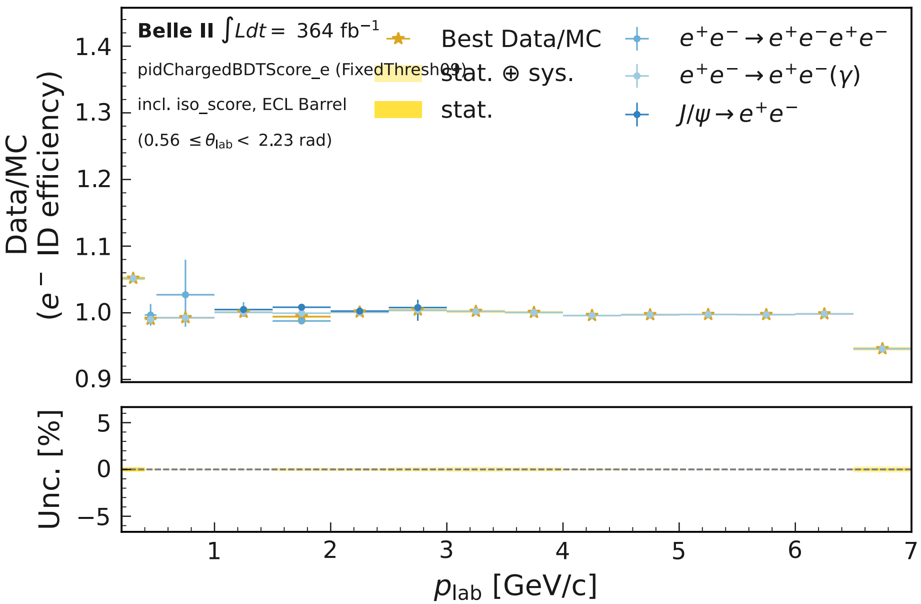


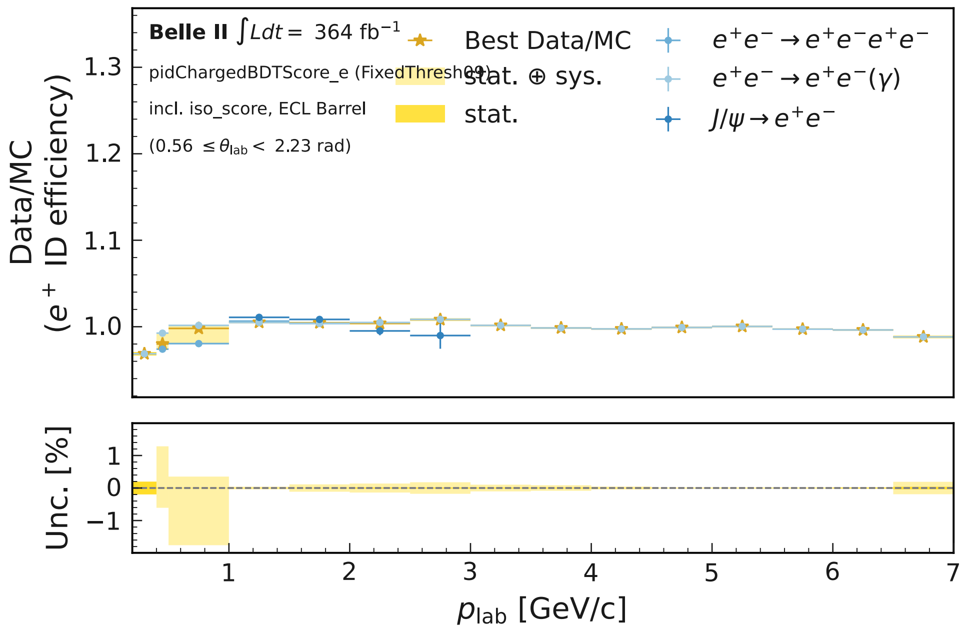
(c) μ^-, likelihood-based LID.



(d) μ^-, BDT-based LID.

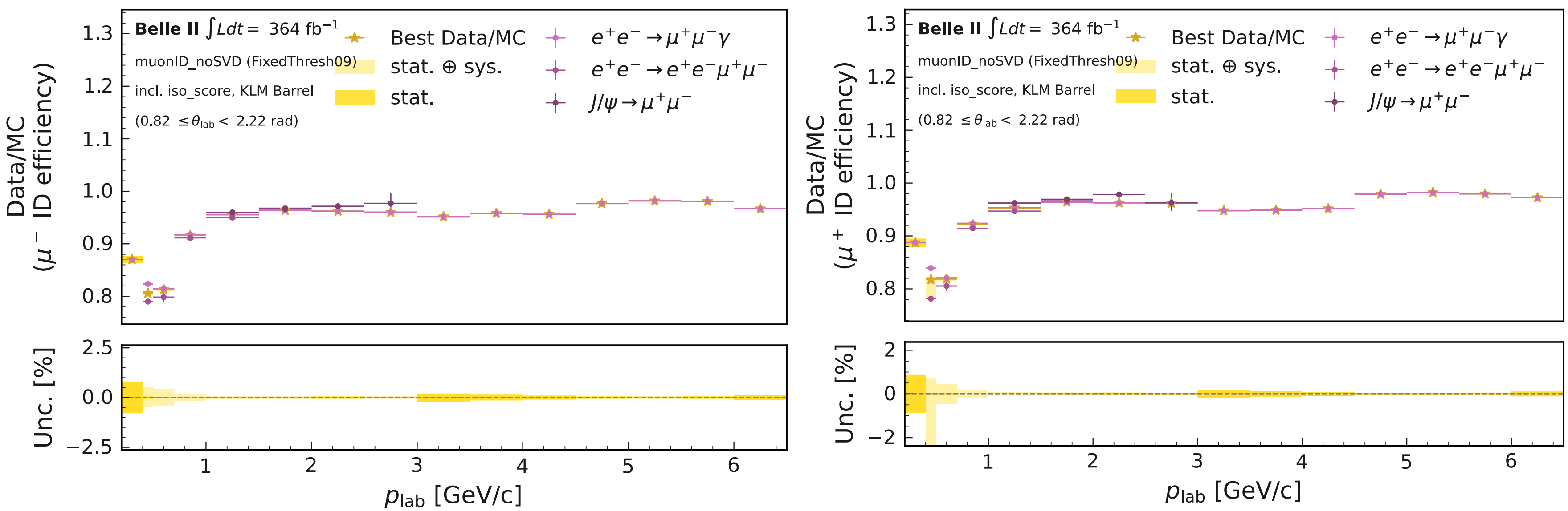
Combination: electron ID





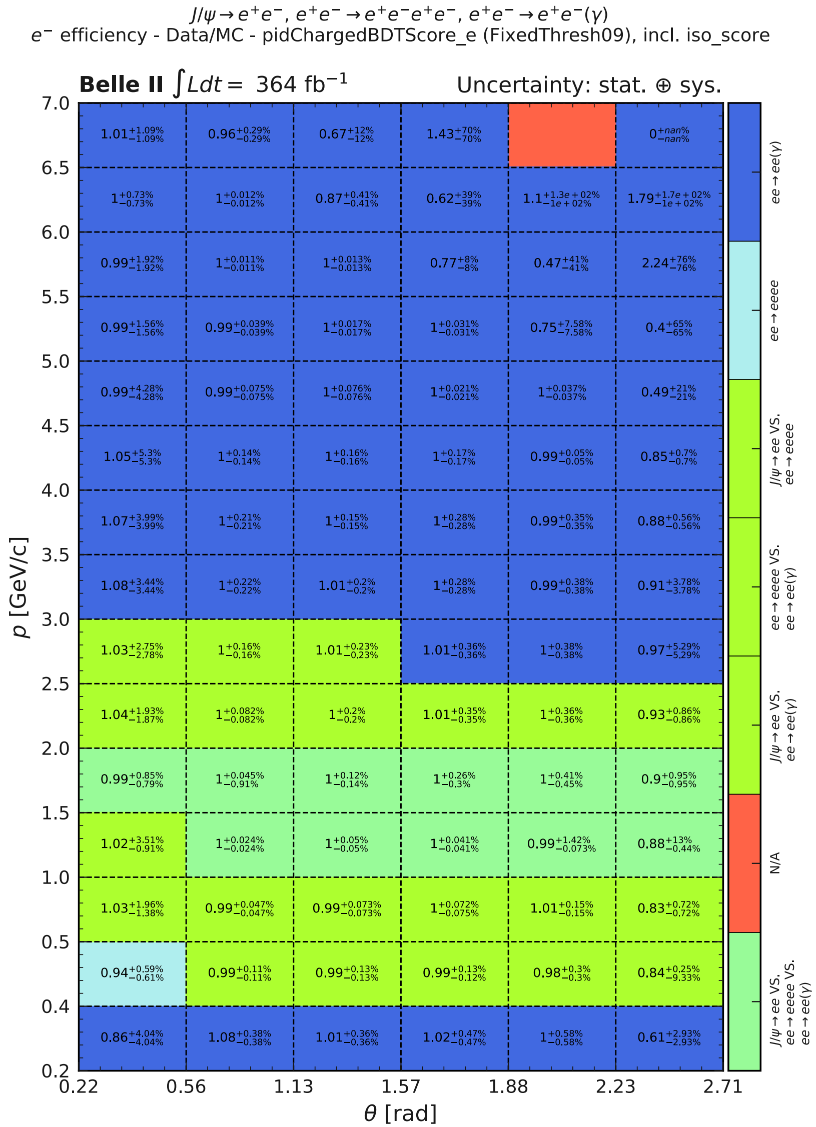
- Similar corrections in three calibration channels → High reliability

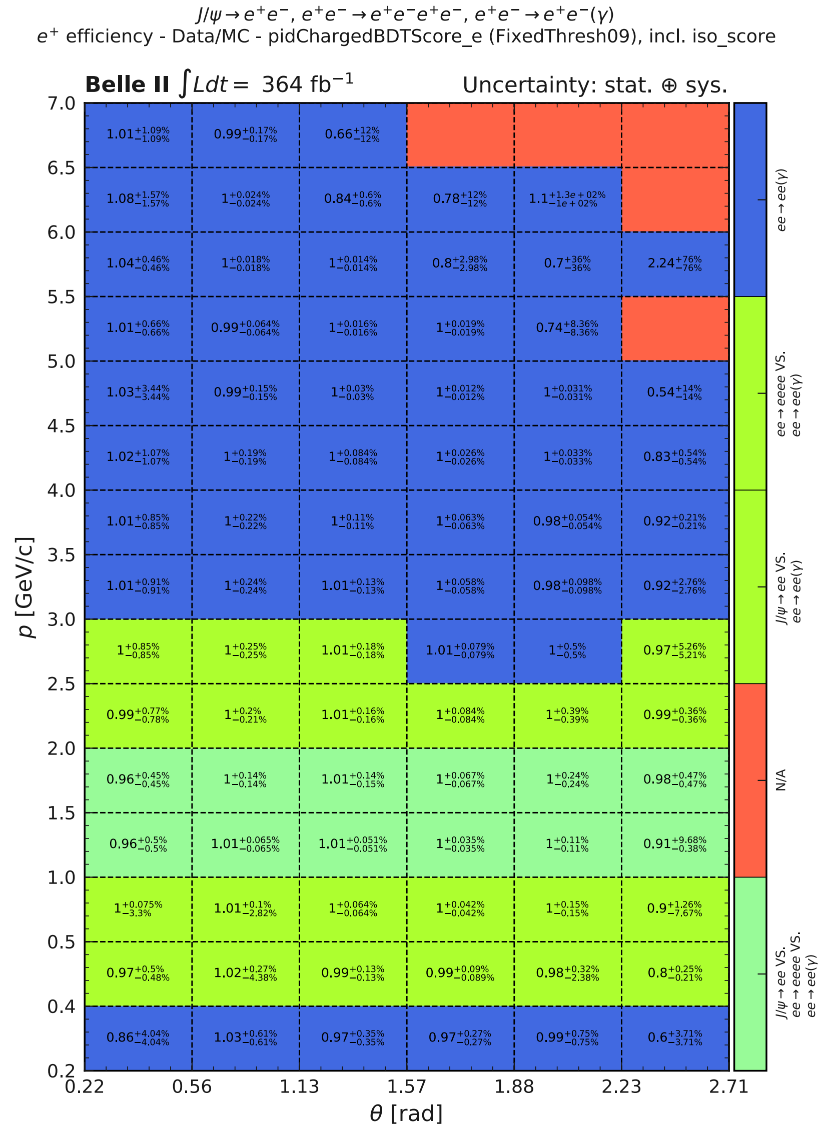
Combination: muonID



- Similar corrections in three calibration channels → High reliability

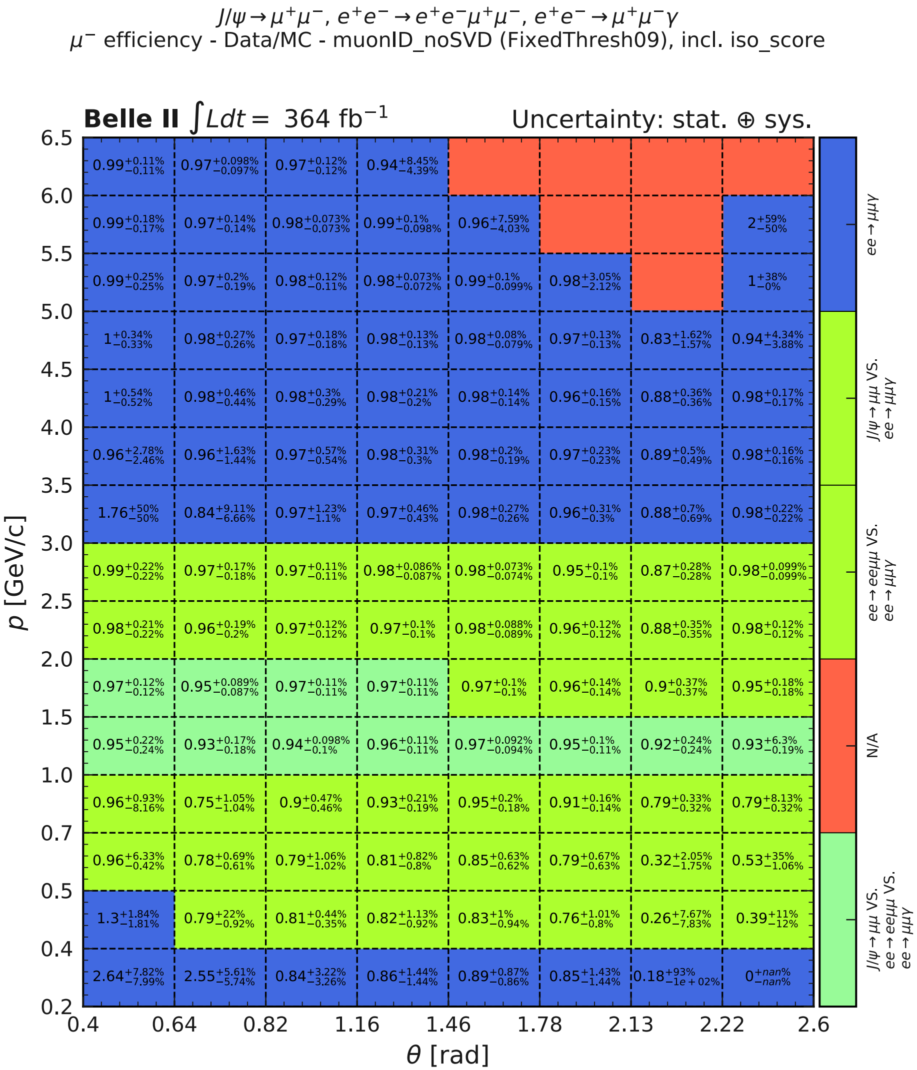
Combination: electron ID

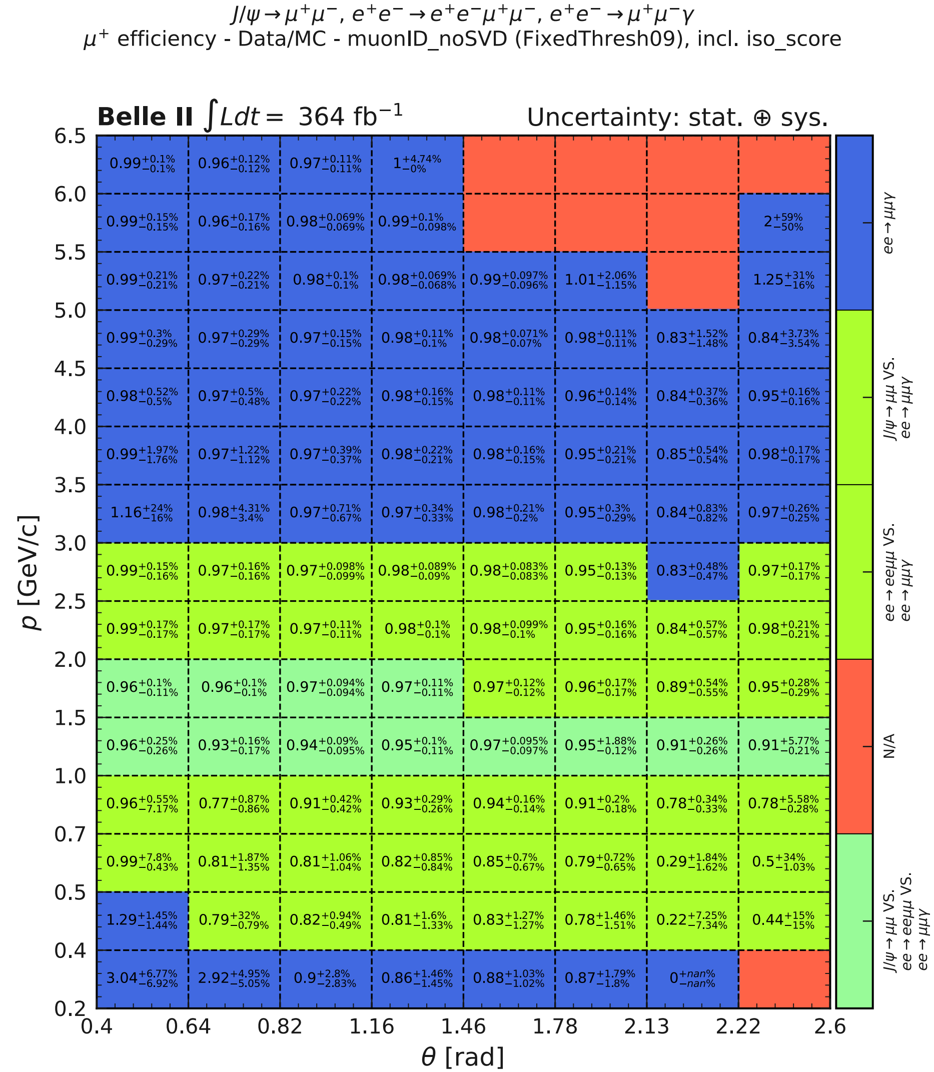




- Calculate corrections and uncertainties in bins of (p, θ, q)
  - Comparable to or better than those at Belle

Combination: muonID





- Calculate corrections and uncertainties in bins of (p, θ, q)
  - Comparable to or better than those at Belle

PID recommendation

Kenta Uno <kenta.uno@kek.jp> 6月26日(月) 16:52
To coll-members, physics, physics-performance, Kenta, Alessandro

Dear all,

We are very happy to announce new PID corrections for run-dependent MC (MC15rd).

Hadron ID correction, Lepton ID fake rate correction
Systematics Framework (SF) ntuples for D*, Lambda0 and KS samples are now available to the users.
You can estimate efficiency corrections depending on your PID criteria. This time, it is possible to estimate not only hadron ID but also fake rates (pion/kaon to lepton fake rates).
Please check the documentation page for the location of the ntuples.
https://syscorrfw.readthedocs.io/en/latest/
kekcc: /group/belle2/dataprod/Systematics/production/
naf: /nfs/dust/belle2/group/dataprod/Systematics/production/
You can also check Ale's slide at B2GM nagoya: https://indico.belle2.org/event/8946/contributions/61650/attachments/22724/33548/B2GM_PID_ale_0605.pdf

Lepton ID efficiency correction
The SF ntuples are not fully validated yet. Instead, we prepared usual csv tables on kekcc:
/group/belle2/users2022/unok/leptonid/combination/perf/PID/methods/LP2023/proc13prompt/MC15rd/v0
All details are collected here: https://confluence.desy.de/pages/viewpage.action?pageId=311564730
You can extract corrections from the csv files.

Let me remind you of the PID recommendation:
Electron ID: We recommend using the BDT-based electron ID because it provides much lower fake rates than likelihood-based electron ID for equivalent efficiency and smaller uncertainty on the data/MC corrections.
Muon ID: We recommend using likelihood-based muonID (muonID_noSVD).
Hadron ID: The default likelihood based variables (global or binary, depending on your needs) are recommended. The reweighted PID and DNN PID variables can be used, but they require some care: check that the performance is better that the default variables. Your feedback is very welcome.

Many thanks to Ale, Paolo, Paul, Philipp, Sviat for your effort.

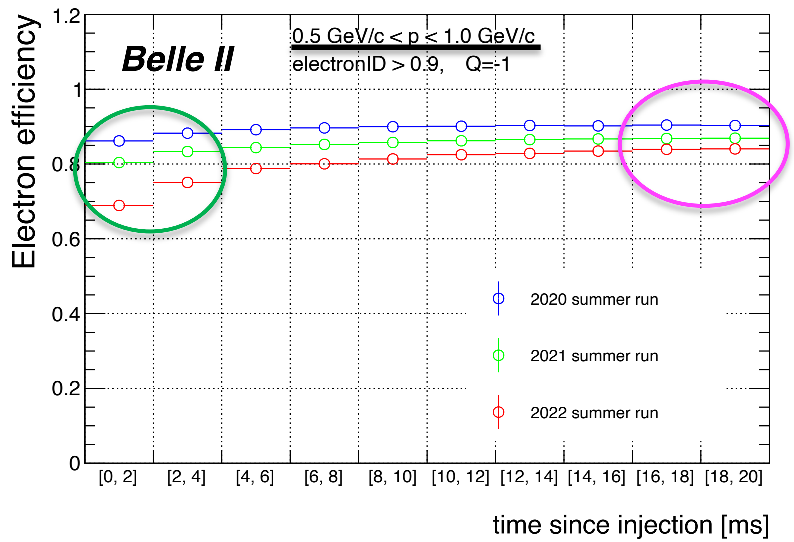
Some LID topics to keep in mind

(1) LID vs beam background

Observe a degradation of LID efficiency due to beam background

- 5-10% degradation of eID efficiency
- 10-15% degradation of eID efficiency

BDT-base eID



Beam bkg gives an impact on detectors

- CDC: gain drop (dE/dx)
- ECL: pedestal shift
- KLM: multi-strip-hit?

Expect larger beam background as the peak luminosity increases

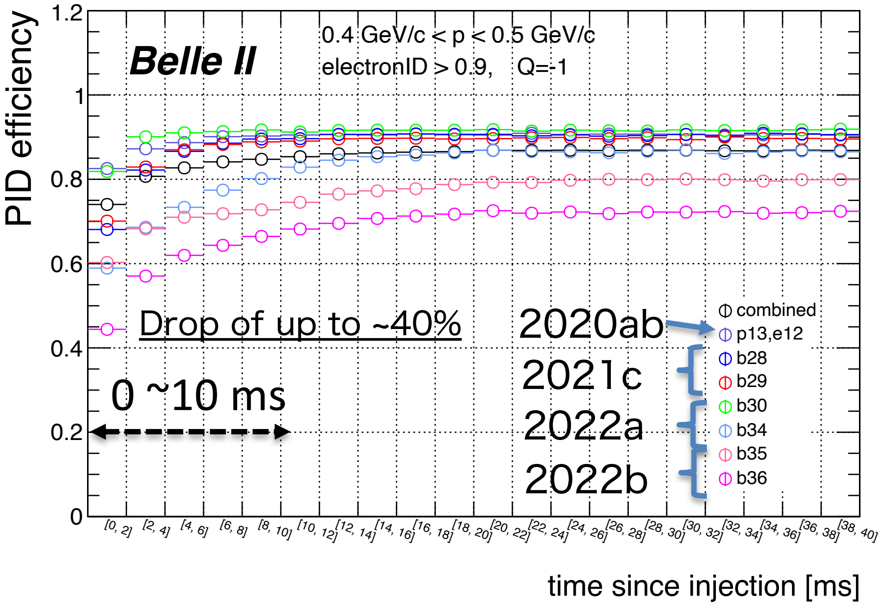
→ We cannot avoid such a degradation for the LID variables.

(1) Beam injection

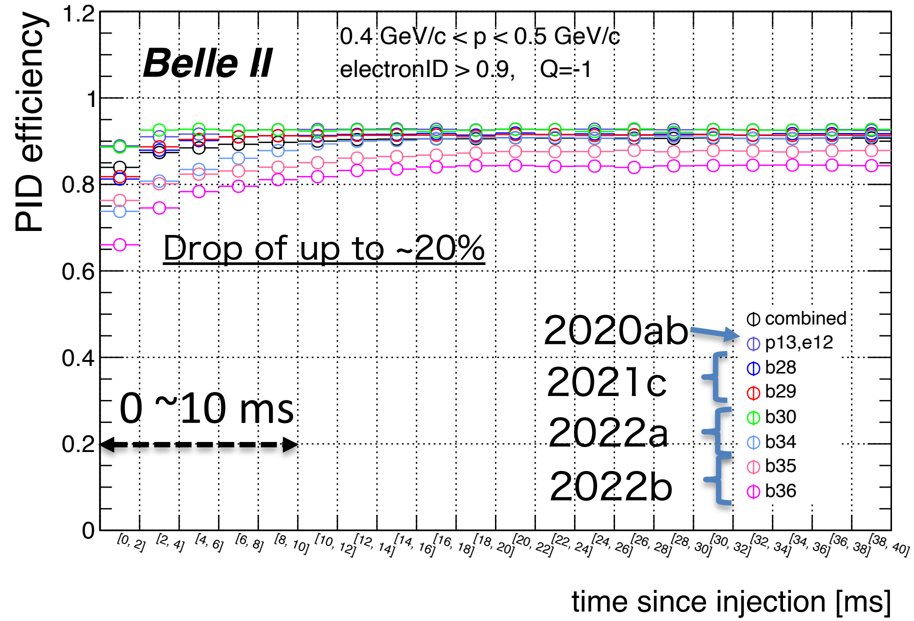
eID efficiency in $ee \to eeee$ channel

- $0.4 < p < 0.5$ GeV/c, ECL barrel ($\theta \in [0.56, 2.23]$)

Likelihood-base eID



BDT-base eID



- <u>Large efficiency drop due to the degradation of dE/dx info.</u>
  - CDC likelihood is affected by the injection backgrounds
- <u>BDT-base eID is more robust against the injection backgrounds</u>

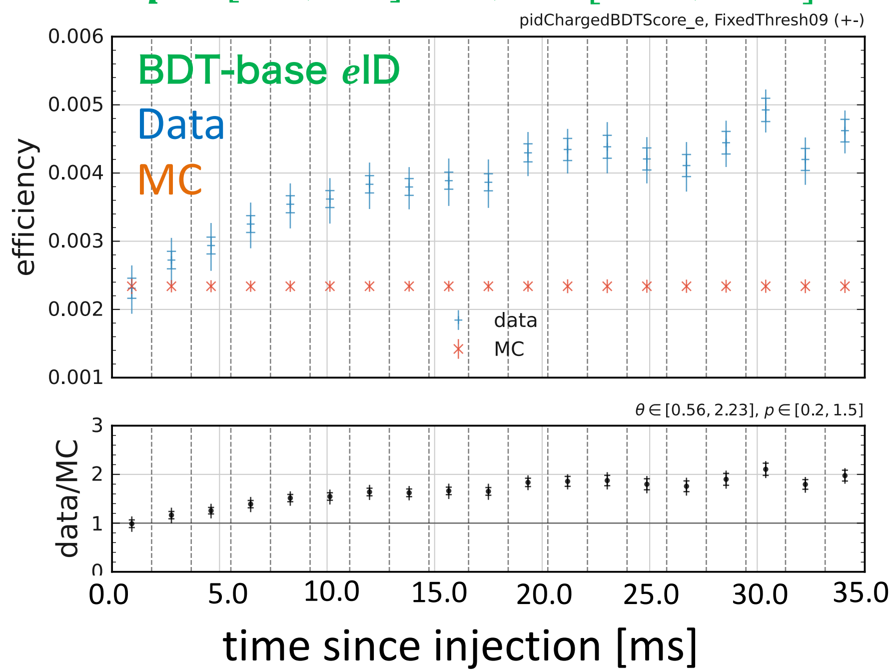
→ Utilize ECL information in BDT-base eID

(The effect of ECL variables is not large)

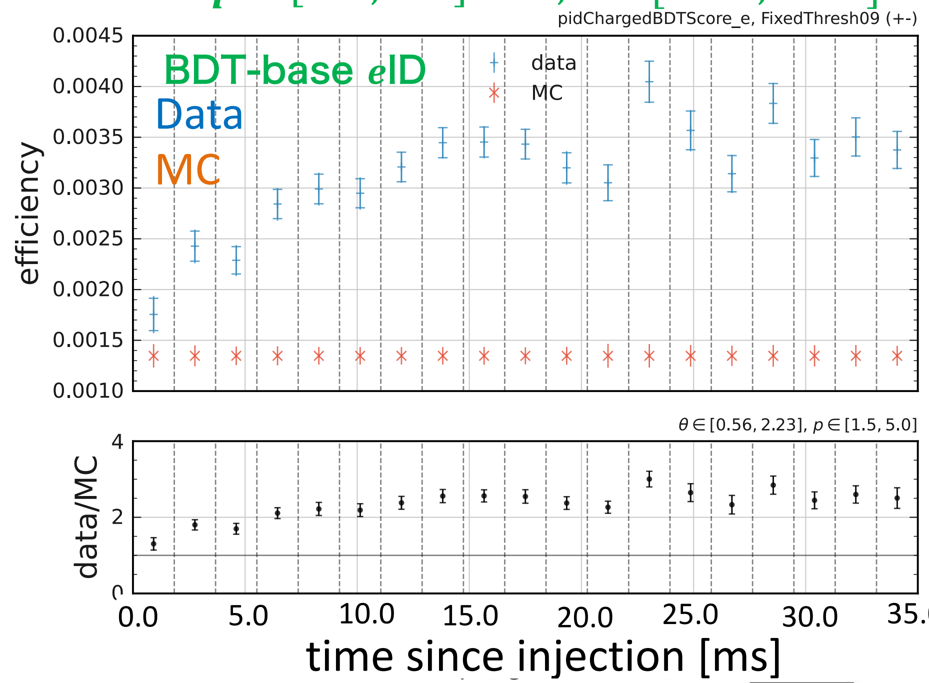
(1) $\pi \to e$ fake: beam injection

$\pi \to e$ fake ratio in $ee \to \tau\tau$ channel ($\tau \to \pi\pi\pi\nu$ decay)

$\boldsymbol{p \in [0.2, 1.5]}$ **GeV,** $\boldsymbol{\theta \in [0.56, 2.23]}$

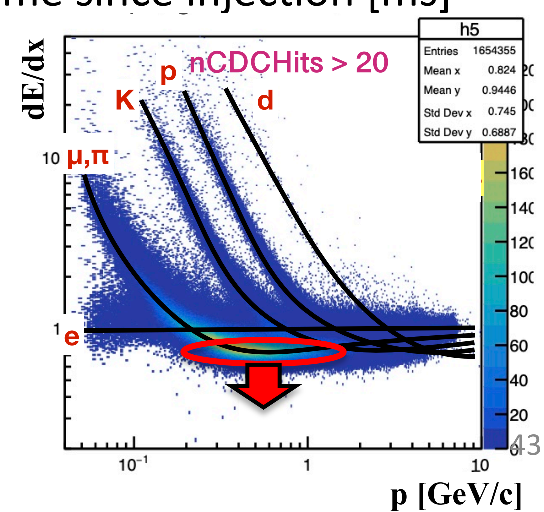


$\boldsymbol{p \in [1.5, 5.0]}$ **GeV,** $\boldsymbol{\theta \in [0.56, 2.23]}$



- <u>Small fake ratio ~10 ms after injections</u>
- Observe the shift to lower values of dE/dx

→ Pions are more unlikely to fake electrons

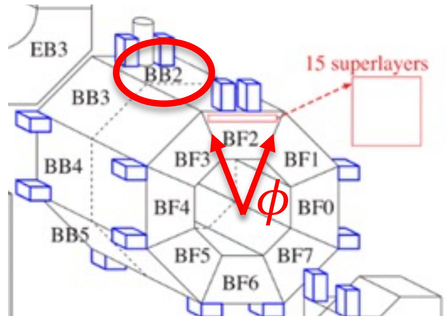


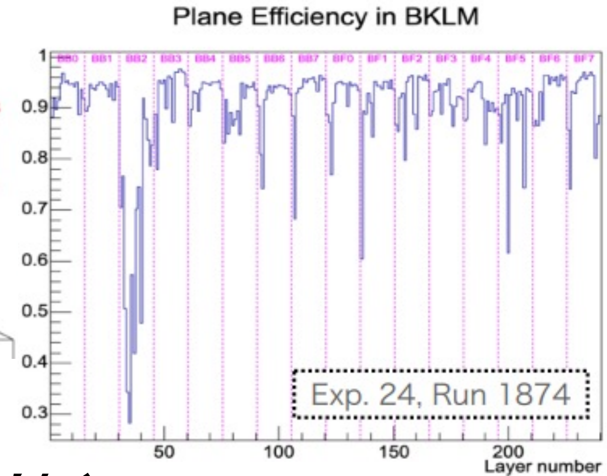
(2) μID effect due to KLM BB2

KLM BB2 issue

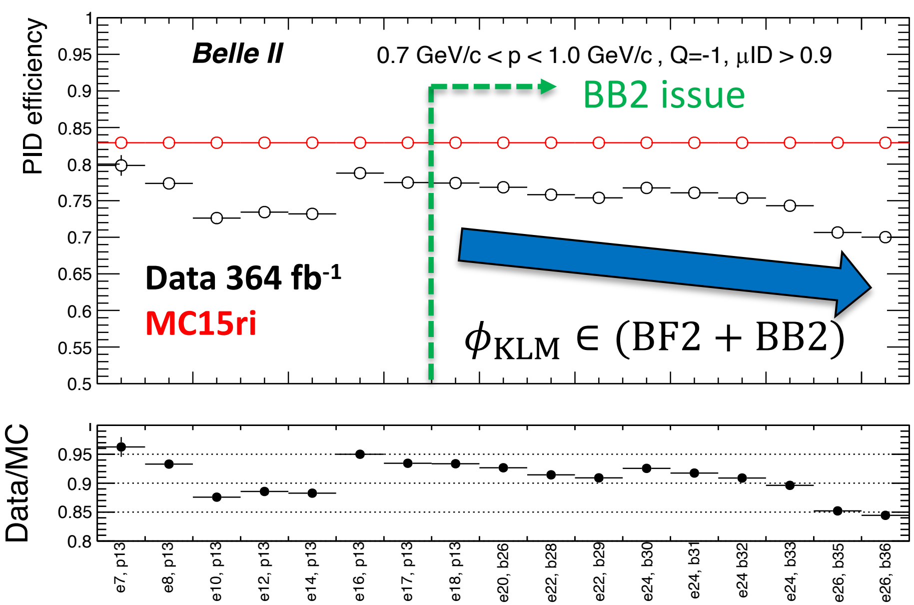
- Efficiency drop from e18
- Observed in only BB2
  - BB2: $-22.5° < \phi_{\mathrm{KLM}} < 22.5°$

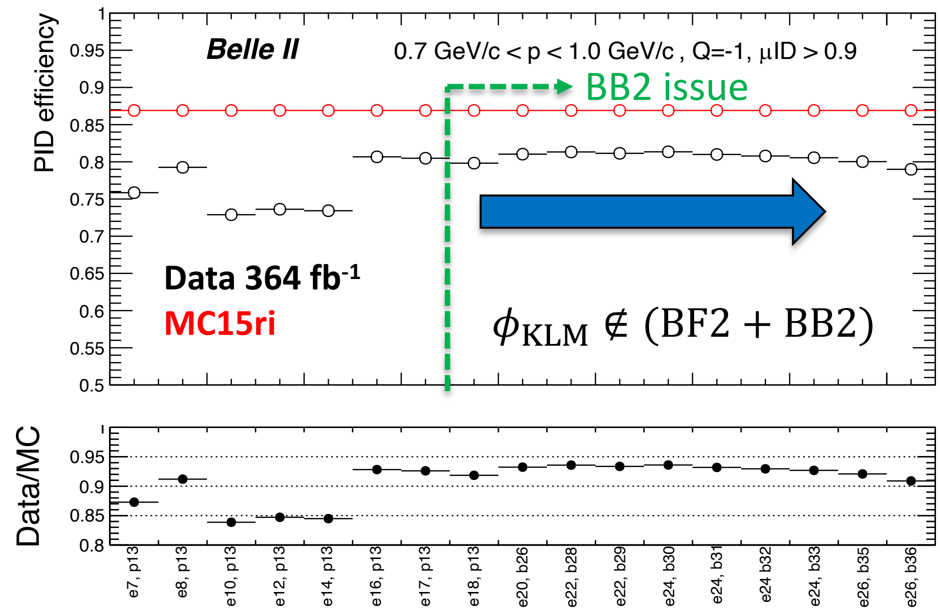
ϕ_{KLM}: Use helixExtPhi to probe a track associated with KLM





μID efficiency with $ee \to ee\mu\mu$ ($0.7 < p < 1.0$ GeV/c)

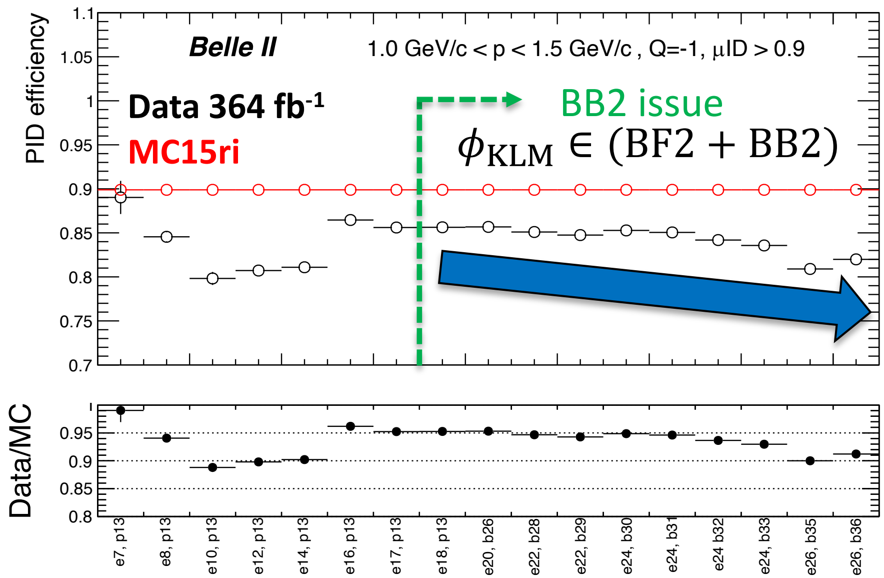




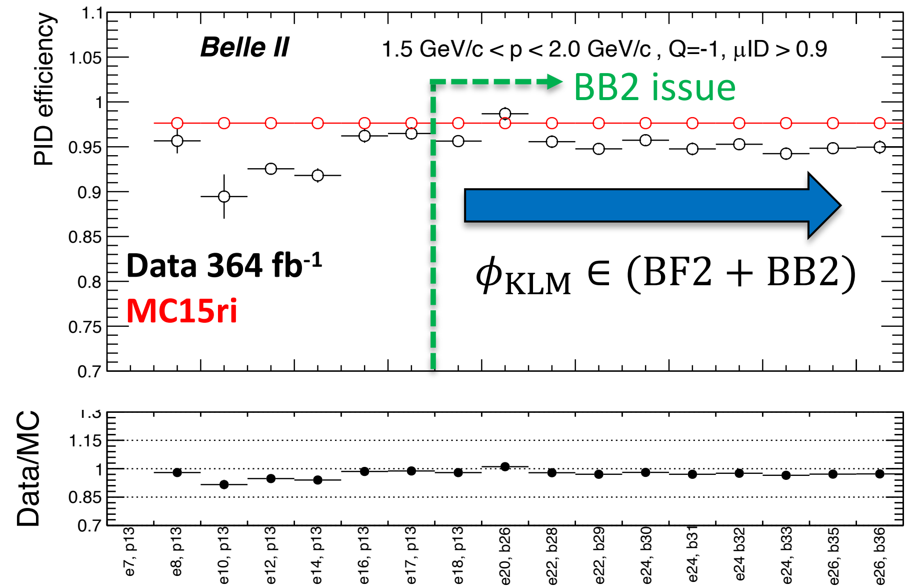
5 ~10% degradation of data efficiency in **the only BB2 region**

(2) Effect in high p region

$1.0 < p < 1.5$ GeV/c



$1.5 < p < 2.0$ GeV/c



(Almost) no impact on μID efficiency at $p > 1.5$ GeV/c

(3) Event multiplicity

Track isolation study

- PID effect due to particles in the vicinity of a candidate track

→ Compute the distance from the nearest particle (at cylindrical surfaces)

- Define "isolation score" based on the distances

→ Already implemented in basf2

`minET2ETIsoScoreAsWeightedAvg`(*referenceListName, useHighestProbMassForExt, detectorList*)

Returns a particle's isolation score s based on the weighted average:

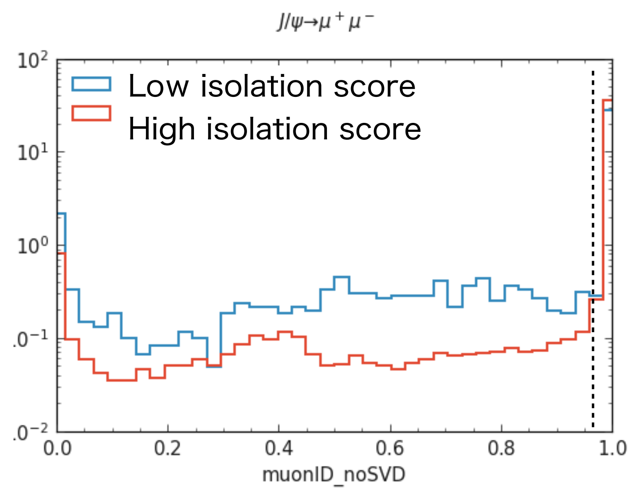
$$s = 1 - \frac{\sum_{\text{det}} \sum_{i}^{N_{\text{det}}^{\text{layers}}} w_{\text{det}} \cdot \frac{D_{\text{det}}^{\text{thresh}}}{d_i}}{\sum w_{\text{det}}},$$

where d_i is the distance to the closest neighbour at the i-th layer of the given detector (c.f., `minET2ETDist`), $N_{\text{det}}^{\text{layers}}$ is the number of layers of the detector, $D_{\text{det}}^{\text{thresh}}$ is a threshold length related to the detector's granularity defined in the `TrackIsoCalculator` module, and w_{det} are (negative) weights associated to the detector's impact on PID for this particle type, read from a CDB payload.

The score is normalised in [0, 1], where values closer to 1 indicates a well-isolated particle.

```
/**
 * Threshold values for the distance (in [cm]) to closest ext. helix to define isolated particles.
 * One for each detector layer.
 */
std::map<std::pair<std::string, int>, double> m_distThreshPerDetLayer = {
  { {Const::parseDetectors(Const::CDC), 0}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 1}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 2}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 3}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 4}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 5}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 6}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 7}, 5.0 },
  { {Const::parseDetectors(Const::CDC), 8}, 5.0 },
  { {Const::parseDetectors(Const::TOP), 0}, 22.0 },
  { {Const::parseDetectors(Const::ARICH), 0}, 10.0 },
  { {Const::parseDetectors(Const::ECL), 0}, 36.0 },
  { {Const::parseDetectors(Const::ECL), 1}, 36.0 },
  { {Const::parseDetectors(Const::KLM), 0}, 20.0 }
};
```

$D_{\text{det}}^{\text{thresh}}$



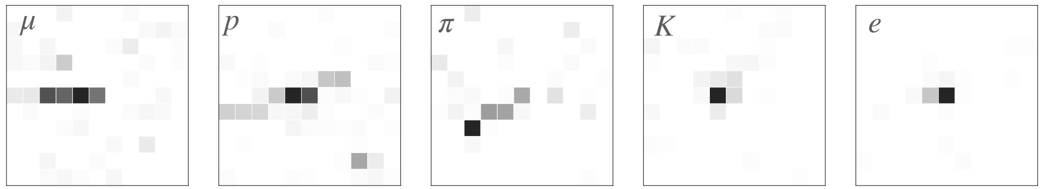
<u>Slightly improve the discrepancy b.t.w high- and low-multiplicity samples</u>

(4) PID with Convolutional NN

Anja Novosel's study

Exploit the specific patterns in the ECL crystals

- Utilize the spatial distribution of energy deposition in the ECL



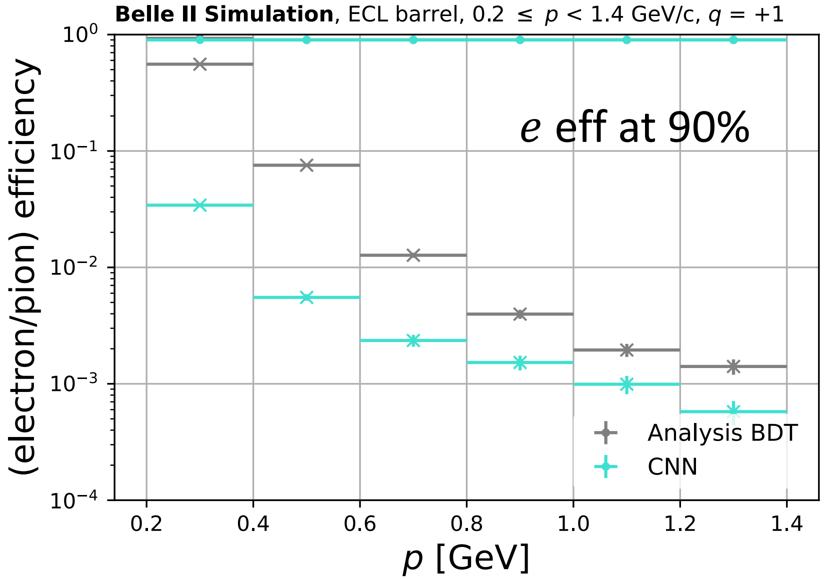
Examples of images for different particle species.

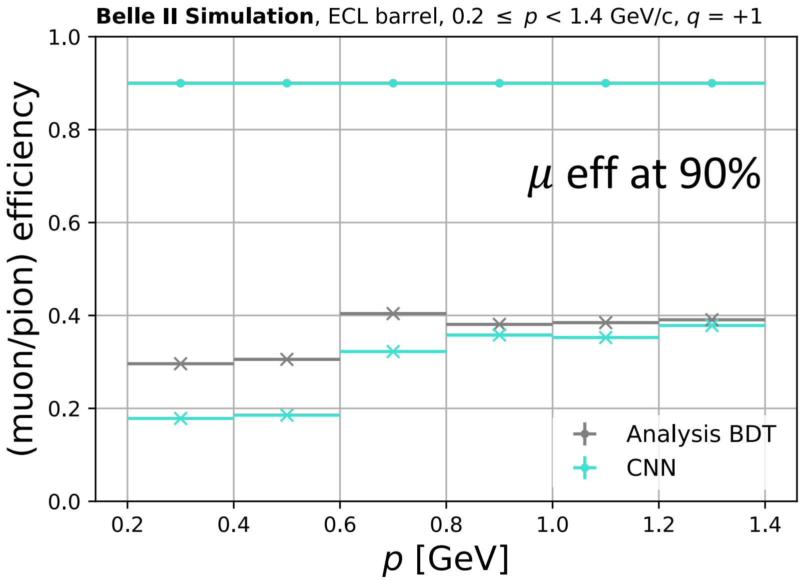
ECL-image: $N \times N$ neighboring crystals

Training sample

- Particle-gun (π, e, μ, K, p)

MC only study





$\pi \rightarrow e, \mu$ fake rate improves at low momentum

Utilize ECL information for low momentum☺

Summary

Sorry, there are many topics I couldn't cover

Please let me (+Ale, Stefan) know if you have questions

Lepton ID is an important item for several analysis

Lepton ID performance

- Identify "electron" and "muon" by exploiting all the detectors!
- Utilize Machine Learning technique to improve the performance

Lepton ID calibration

- Provide efficiency/fake rate corrections in bins of (p, θ, q)
  - A better data/MC agreement applying corrections in several analysis

Lepton ID development

- Many topics (eg. ML-based PID) are ongoing!

ありがとうございました!

BTW, this is my last presentation as PID convener.

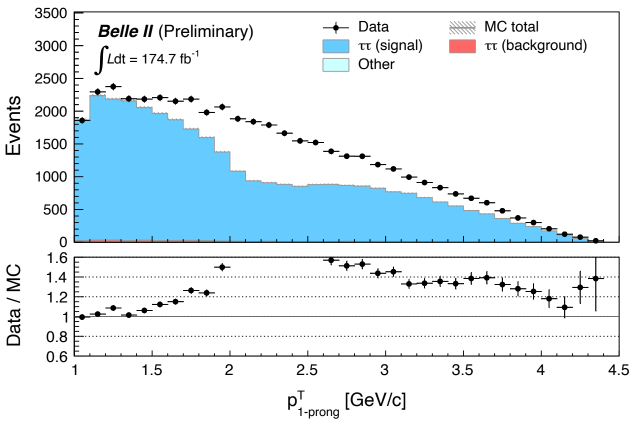
Thank you very much for your cooperation!

Backup

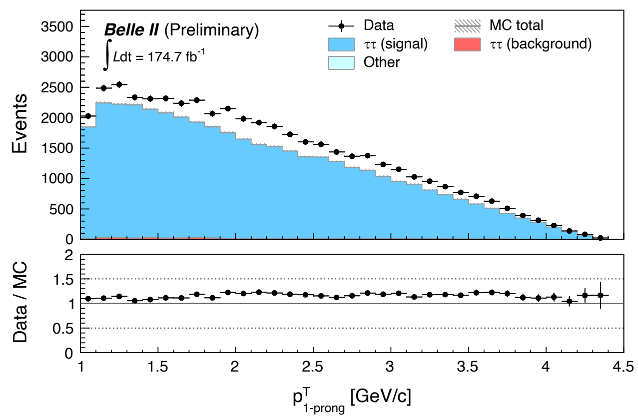
SVD likelihood issue

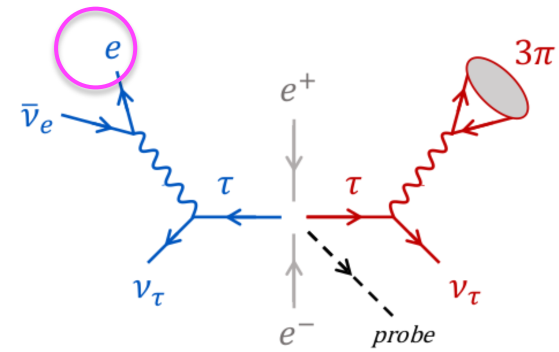
C.Lemettais, S.Xavior, A.Gaz

electronID_noTOP > 0.9



electronID_noSVD_noTOP > 0.9



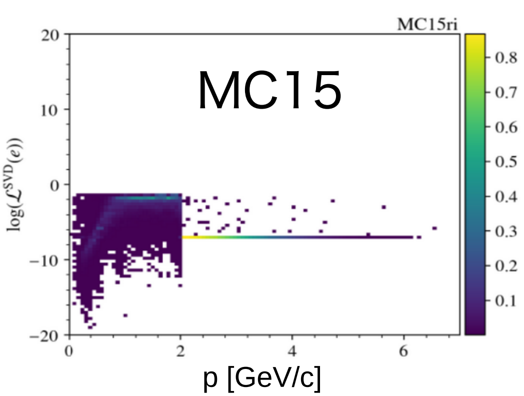


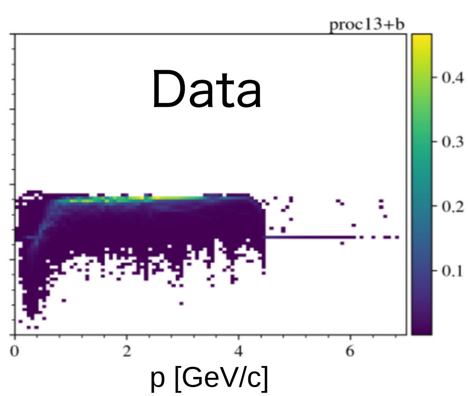
- $p > 2.0$ GeV/c: Distortion in the MC simulation

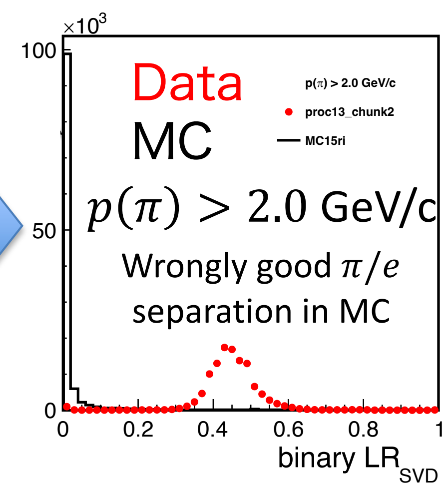
Problem in uploading the payloads for electron hypothesis

- The SVD likelihood is defined only up to 2 GeV/c in the MC

 (other particle hypotheses: up to ~ 5 GeV/c)







$$\frac{\mathcal{L}_e}{\mathcal{L}_e + \mathcal{L}_\pi}$$

PID effect due to SVD likelihood

This problem affects only the MC

Electron ID

- Biggest effects are expected

→ Exclude SVD (and TOP) info

$$p_e = \frac{\mathcal{L}_e}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}$$

$$p_\pi = \frac{\mathcal{L}_\pi}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}$$

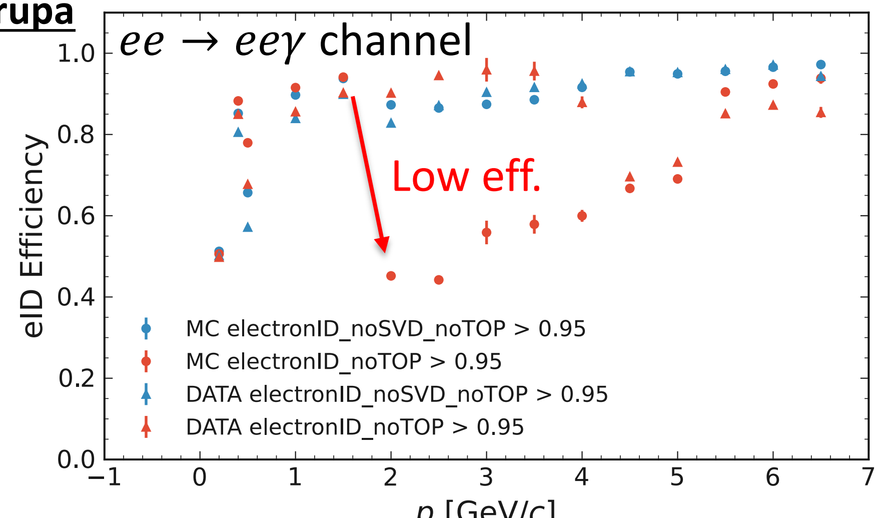
Hadron ID

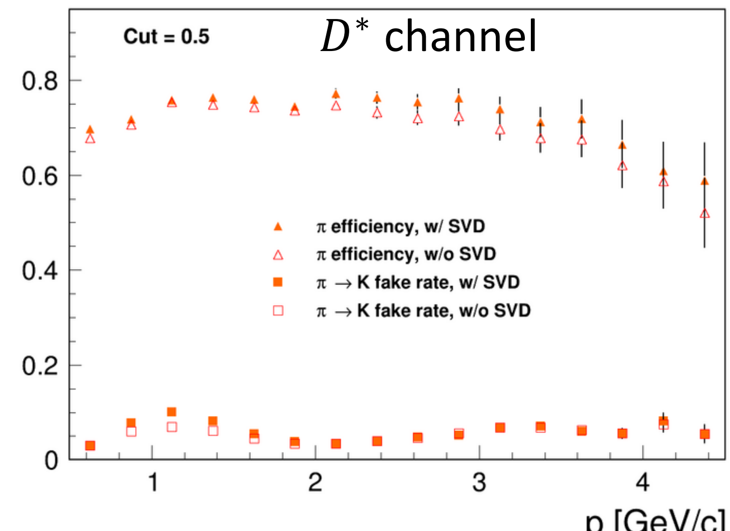
- Problem appears at the denominator

→ Small effects are expected, so keep SVD info

※ small effect for muon ID, but we exclude SVD info (to be consistent with eID)

J.Skorupa





A.Gaz

Discussion with SVD experts is ongoing