Charged PID performance overview

Marco Milesi (on behalf of the Hadron ID, Lepton ID sub-groups)

School of Physics - The University of Melbourne

BPAC Review, Summer 2021
Electron ID performance - Overview

- $e/\pi$ separation dominated by ECL ($\rightarrow E/p$ only in likelihood-based PID), w/ CDC ($\rightarrow dE/dx$) also providing a contribution.

- New in release-05: CDC $dE/dx$ vs. $\beta\gamma$ predictions estimated from calibrated control samples in data.

- TOP provides some separation only at low momenta. Unfortunately, TOP(e) PDFs are affected by a bug in release-05 w/ $\delta$-ray photon yield modelling, so we remove them temporarily from the likelihood-based PID definition. This is going to be fixed in release-06 conditions DB for autumn processing.
Electron ID performance - Algo improvements

- Including several ECL shower shapes + PSD info in a BDT greatly improves $e/\pi, e/K$ separation, especially at low momenta.

- Factor x10 reduction in $\pi, K \rightarrow e$ fake rates for $p < 0.6$ GeV/c @ 90% efficiency.

- BDT-based electron eff./fake rates are now being calibrated in data as well, and are recommended for usage in analysis.

Selected ECL observables combined w/ (non-ECL) $\mathcal{L}_e/\sum_i \mathcal{L}_i$ in BDT, low $p$, barrel region

- Factor x10 reduction in fake rates for $\pi, K \rightarrow e$ with $p < 0.6$ GeV/c.
Muon ID performance - Overview & algo improvements

- $\mu/\pi$ separation dominated by KLM for $p > 600$ GeV/c (acceptance limit). ECL and TOP can improve low $p$ separation, especially when combined in a BDT w/ multiple calorimetric observables.

- Promising results from using a CNN trained on $\mu/\pi$ ECL cell images at reconstruction stage, aiming to be integrated in basf2 by the end of 2021.
Electron and muon ID performance - MC calibration w/ data

- Preliminary calibration of $e$, $\mu$ efficiency in $J/\psi \rightarrow \ell^+\ell^-$ candidates show stability between rel-4 and rel-5.
- Data and MC are in good agreement overall.
- In the plots, efficiency is set to be 90% uniformly across $(p, \theta)$ bins.

- Efficiency calibration is performed also in low-multiplicity channels ($e^+e^-(\gamma), \mu^+\mu^-\gamma, e^+e^-\ell^+\ell^-$).
Electron and muon ID performance - MC calibration w/ data

- Preliminary calibration of $\pi \rightarrow e$, $\pi \rightarrow \mu$ fake rate in $K^0_S \rightarrow \pi^+\pi^-$ candidates show clear improvement in rel-5 wrt rel-4 for electrons, stable performance for muons.

- Owing to improved CDC PDF modelling in simulation.

- Detailed study in $e^+e^- \rightarrow \tau^\pm(1p)\tau^\mp(3p)$ shows dependence of $e$ fake rates upon event activity ("isolation"). Currently testing a parametrisation based on distance to closest track at a detector surface to reduce systematics, dominated by differences between event topologies.
$D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$ process, kinematically tagged by OS $K, \pi +$ slow pion.

- SVD not yet available in proc12, MC14. ARICH limited to tracks w/ $p > 0.55$ GeV/c
- MC14 performs better than data overall (but opposite for CDC only), disagreement driven by TOP.

- TOP: no major news since March BPAC → performance and data/MC difference not yet understood.
- Performance -stable across data taking periods, except for CDC $dE/dx$, which degrades in most recent “buckets” ($\rightarrow$ NB prompt processing not fully re-calibrated)

$K/\pi$ performance - MC calibration w/ data
$\Lambda^0 \rightarrow p^+\pi^-$ process.

- SVD not yet available in proc12, MC14.

- MC14 performs better than data overall (but opposite for CDC only), disagreement driven by TOP.
CDC $dE/dx$ status for hadron identification

- As reported in March BPAC, progress made in improving data/MC agreement.

- Some remaining discrepancies are still observed (details in backup):

  - $\pi \rightarrow K$ fake rate above 1 GeV/C, where MC overestimates data
    - Likely due to bias in the predicted $dE/dx$ vs. $\beta\gamma$ mean/resolution for MC.
  - $K$ efficiency dropping near $\cos \theta = 0$
    - Possibly due to imperfections in the hadron saturation corrections.
• CDC → asymmetries observed for tight PID requirements, with opposite signs in data and MC.

• $dE/dx$ has no mechanism to produce asymmetry. Effect is inherited from tracking, e.g. asymmetric distribution in number of hits, which has opposite sign compared to data.

• In data, additional asymmetry for tight requirements may result from imperfect 1D cell correction. Plan to move to 2D cell correction for next processing.

• TOP → asymmetries observed for kaons and protons

• Likely dominated by material interactions. Difference in data vs. MC trend needs to be investigated.

• Asymmetries increased after inclusion of $\delta$-ray component in PDFs.
Customising & automating the PID calibration workflow

- Choosing an arbitrary cut on PID (e.g., electronID > 0.9) clearly sub-optimal → one size doesn’t fit all!

- Providing corrections for target efficiency / background rejection is better, but still might not be the best choice for your analysis...

- The new “Systematic Framework” aims at overcoming this issue.

  1. PP defines the calibration channels \((\Lambda^0 \rightarrow p^+\pi^-, D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+)\)

  2. NTuples for different modes (and possibly their combination) are centrally produced and distributed to analysers → include sWeights to obtain pure “signal” histogram templates.

  3. Users produce ROC curves and automatically obtain tables w/ efficiencies and corrections, for any choice of PID (binary, trinary, global, BDT...).

  4. Also obtains statistical and systematic uncertainties.

- Being implemented as a b2luiji pipeline, it provides full automation, which will be ever more critical given how often updated corrections are needed.

- Lifts large amount of technical (boring) work from PP experts.
Backup
Electron ID performance - BDT inputs

- Input variables for the BDT are reasonably well modelled in MC (checked on $J/\psi \rightarrow e^+e^-$ candidates, sidebands bkg. sub.).
Input variables for the BDT are reasonably well modelled in MC (checked on $J/\psi \rightarrow \mu^+\mu^-$ candidates, sidebands bkg. sub.).
BPAC: CDC dE/dx calibration

Issue#1: costheta=0 dip in kaon efficiency (in data only)
- Electrons based calibration (“cosine correction”) using Bhabha electrons
  - electrons are at rise relativistic rise
- Hadron based calibration (“hadron-sat correction”) using (P/k/pi/mu) samples (rel. to electrons)
  - current version works well overall but some imperfections

In Data
- The dip in efficiency at cosθ near 0 is most likely due to imperfections in the saturation corrections.
  - dE/dx means are not flat vs. cos theta for min-I pions
    - decreases separation power hence affects kaon efficiency
- Work in progress to improve current imperfections
  - weighted bins, unfolding costh, detailed parameters studies

In MC
- Current track-level CDC dE/dx MC is not detailed enough to reproduce this effect
  - hits level MC in future (ongoing but a long project)
BPAC: CDC $dE/dx$ calibration

**Issue #2: K/Pi charge asymmetry in Data and MC**

**In MC**
- Current track-level CDC $dE/dx$ MC has no mechanism to produce an asymmetry in the $dE/dx$ portion
  - This means current/any MC asymmetry must be "inherited" from tracking, such as a difference in number of hits for each charge.

  ![Graph showing resolution predictions as a function of nhits (layer hits used in dE/dx truncation)](image)

  **Our resolution predictions use track hits ($dE/dx$ internal variable but correlated with nCDChits)**

**In Data**
- New 1D Cell corrections with release/05 removes most of such effects
  - remaining cal. imperfections seem causing issues for higher PID cuts
  - further improvements are expected with
    - 2D cell corrections in future (e.g. proc13 or in 2021c data)

![Graph showing Bella II preliminary CDC only MC14a vs CDC only proc13+B](image)
**BPAC: CDC dE/dx calibration**

**Issue #3: Prompt vs proc12 vs MC**

3a) Underperforming MC
- MC is underperformed due to higher ($\pi \to K$) fake rates for $p > 1\text{GeV}/c$
  - CDC dE/dx MC is produced using data based hadron calibrations
  - Higher $\pi$ fake is probably due to the biases of hadron cal. (mean/reso. predictions)
    - reducing $K/\pi$ band separations in MC
- Chunk2 Data/related-MC might show better results (a different approach is used in hadron cal)

3b) Lower performance in prompt
- Low performance w.r.t. major processing is due to hadron calibration
  - Hadron calibration is not re-calibrated in prompt due to complicated design and time-taking process
  - A stable version from s-proc1 (release/05 + exp12) is currently used in recent prompt buckets (>=16)
    - Bucket16 (exp14) shows a very good compatibility
    - Buckets (17→20) : A gain degradation and change is cosine dip is observed in these buckets and that affects overall compatibly with hadron (saturation) corrections.
    - Bucket19: additional poor performance is probably due to other reasons

**Solutions?**
- We will try to perform hadron calibration using bucket20/21 data and make it available for bucket22 (or with 23). Work in process!
$K/\pi$ performance - release 4 VS. release 5

- $K^0_S \rightarrow \pi^+\pi^-$ process, probes pions w/ $p < 3$ GeV/c. Study $\pi \rightarrow K$ mis-identification probability (among others).

- Data and MC agree better in release-05, owing to improved CDC calibration and updated CDC MC $dE/dx$ PDFs.

- $\pi \rightarrow K$ rate sizably reduced in both data and MC. Mostly thanks to overhauled PDF model to include $\delta$-rays.

New rel-5 PDF models “fast” $\delta$-ray emission ($\rightarrow dE/dx$ ($\beta$) -dependent effect, hence dependent on particle’s mass at given $p$)
Investigating TOP electron response in release 5

- Study on a MC14 $\tau\tau$ sample.
- $\Delta \text{Log} \mathcal{L}_{e/\pi} = \text{Log} \mathcal{L}(e) - \text{Log} \mathcal{L}(\pi)$ should always be $> 50\%$ for truth matched electrons (i.e., PID no worse than random guessing).
- Seems not the case in release 5. Possible issue w/ $\delta$-ray modelling for (primary) electrons.

- Observed large discrepancy in “expected vs. simulated” difference of $e, \pi$ photon yield in release 5.
- Expected photon yield was off in rel-4, but the relative $e, \pi$ difference was correctly modelled.