Contribution ID: **192**   Type: **Poster**

GAN-based Particle Identification over Large Hadron Collider beauty Run III Data

The performance of Particle Identification (PID) in the LHCb experiment is critical for numerous physics analyses. Classifiers, derived from detector likelihoods under various particle mass hypotheses, are trained to tag particles using calibration samples that involve information from the Ring Imaging Cherenkov (RICH) detectors, calorimeters, and muon identification chambers. However, these control channels often differ significantly in feature distributions from the physics channels under study. This mismatch limits the precision with which PID response can be predicted in analyses, particularly in statistically limited datasets like the beam-gas ones collected at LHCb with the System for Measuring Overlap with Gas (SMOG).

In this work, we propose a novel deep generative strategy to learn multidimensional PID distributions from real calibration data using a GAN-based architecture (PIDGAN). A GAN-based architecture enables the generalization across multiple calibration channels, effectively learning high-dimensional PID responses conditioned on experimental features. Our method opens a path towards improved PID calibration with scalable, data-driven models that capture correlations and non-linear effects in PID variables more comprehensively, offers an alternative approach to PID studies for physics analyses, and illustrates a broader strategy for generative modeling of real-world, high-dimensional sensor data.

Significance

This presentation explores novel techniques to enhance GANs for modeling low-statistics regions in high-energy physics data, with a focus on improving fidelity in the tails of physical distributions. By incorporating strategies such as targeted noise injection, we address key challenges in rare event generation. Unlike many approaches that rely on simulated samples, our method is trained exclusively on real detector data, which allows us to directly model the true underlying distributions without sim-to-real domain shifts. These methods go beyond status reporting by providing actionable improvements to generative model training and evaluation in the context of realistic detector conditions.

This work also represents an important incremental step within a broader effort to develop fast, accurate calibration tools for particle identification at LHCb. It contributes to the long-term goal of integrating machine learning-based generative models into the high-throughput analysis pipelines of current and future LHC runs.

References

While I (Josef) was not involved at the time, this work aims to extend what was started by part of our group here: https://arxiv.org/abs/2110.10259 and address its limitations.

Experiment context, if any

Real-Time Analysis at the LHCb, PID-calibration

Author: RUZICKA GONZALEZ, Josef (Costa Rica Center for High Technology)

Co-authors: MENESES ROJAS, Esteban (Consejo Nacional de Rectores (CONARE) (CR)); ANDERLINI, Lucio (Universita e INFN, Firenze (IT)); BARBETTI, Matteo (INFN CNAF); MARIANI, Saverio (CERN); ARGUEDAS CUENDIS, Sergio (Consejo Nacional de Rectores (CONARE) (CR))

Presenters: RUZICKA GONZALEZ, Josef (Costa Rica Center for High Technology); MARIANI, Saverio (CERN); ARGUEDAS CUENDIS, Sergio (Consejo Nacional de Rectores (CONARE) (CR))

Session Classification: Poster session with coffee break

Track Classification: Track 2: Data Analysis - Algorithms and Tools

Author: RUZICKA GONZALEZ, Josef (Costa Rica Center for High Technology)

Co-authors: MENESES ROJAS, Esteban (Consejo Nacional de Rectores (CONARE) (CR)); ANDERLINI, Lucio (Universita e INFN, Firenze (IT)); BARBETTI, Matteo (INFN CNAF); MARIANI, Saverio (CERN); ARGUEDAS CUENDIS, Sergio (Consejo Nacional de Rectores (CONARE) (CR))

Presenters: RUZICKA GONZALEZ, Josef (Costa Rica Center for High Technology); MARIANI, Saverio (CERN); ARGUEDAS CUENDIS, Sergio (Consejo Nacional de Rectores (CONARE) (CR))

Session Classification: Poster session with coffee break