



dE/dx Estimator Optimization for Anomalous Ionization Searches

Michael John Noamesi¹ Emery Nibigira²

¹Gettysburg College

²University of Tennessee Knoxville



THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

Introduction

- Supersymmetric and other BSM models predict **Heavy Stable Charged Particles (HSCPs)** that move at sub-relativistic speeds $\beta = \frac{v}{c} < 1$ and deposit unusually large **energy loss per unit length (dE/dx)** in multiple layers of the silicon detector.
- HSCPs can traverse the whole detector or decay far from the interaction point, so they sit in the broader class of **Long-Lived Particles (LLPs)**.
- A particle's dE/dx follows a Landau-like asymmetric distribution, so robust track-level estimators (not a simple mean) are essential to suppress high-tail outliers and extract a reasonable **most probable value (MPV)**.
- However, existing CMS estimators (e.g. Harmonic-2) leave a residual bias in reconstructed HSCP mass due to heavy-tailed per-hit energy-loss distributions.

dEdx and Estimators

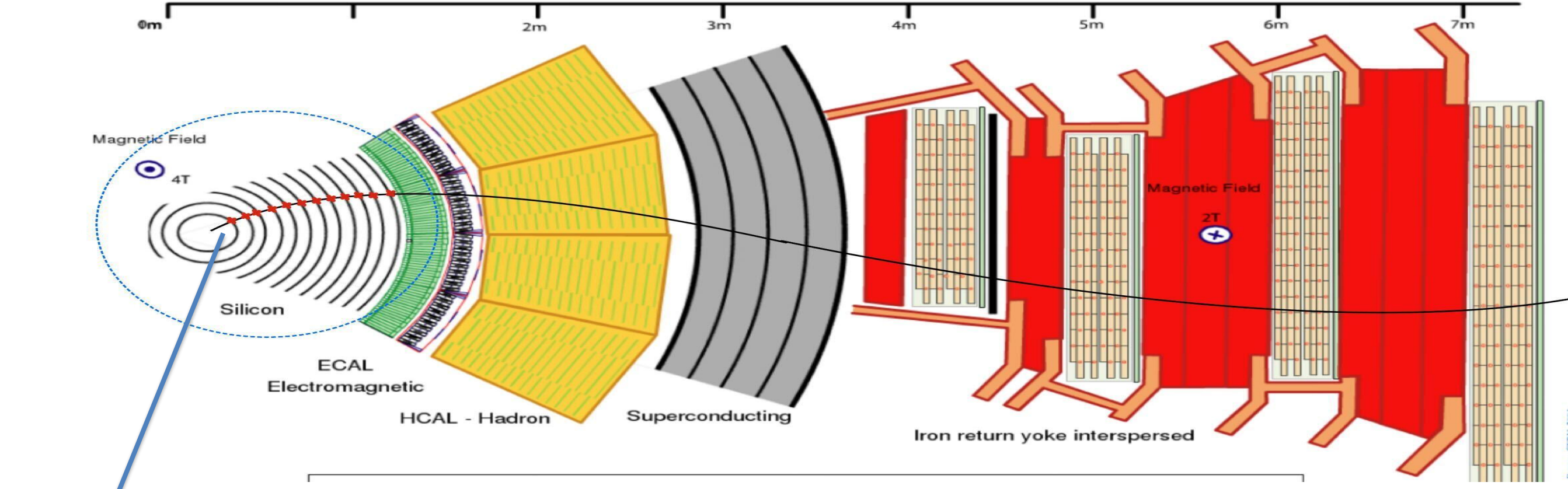


Figure 1. Transverse cut-out of CMS with an HSCP track

Estimators translate a track's collection of hit-level energy-loss measurements into a single, most-probable dE/dx value used to reconstruct mass. Yet, it fares poorly because the underlying hit distribution is Landau-skewed: a handful of abnormally large deposits ("pathological" outliers) pull the mean upward and blur the reconstructed mass peak. To suppress those tails, CMS adopts the **Harmonic-2 estimator (I_h)**, which weights each hit by $\frac{1}{(\frac{dE}{dx})^2}$ and thus down-weights extreme

$$I_h = \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{dE}{dx} \right)_i^{-2} \right]^{-\frac{1}{2}}$$

values. Even so, I_h leaves a residual high-tail bias and may not capture the true dE/dx MPV, motivating our exploration of alternative approaches, especially for anomalous ionization signatures.

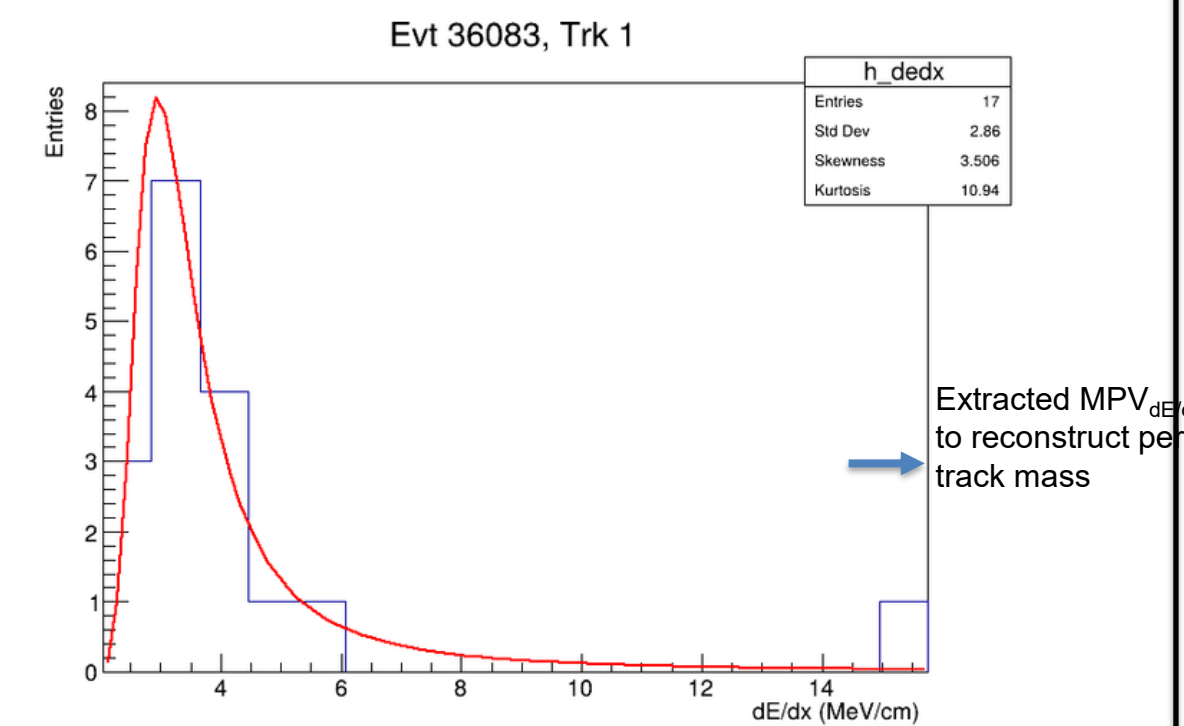


Figure 2. Typical distribution of a particle's dE/dx through the strip detector (hit level)

Bethe-Bloch function and mass reconstruction

dE/dx of charged particles traversing a material depends on $\beta\gamma$ as described by the Bethe-Bloch function. [fig. 3]

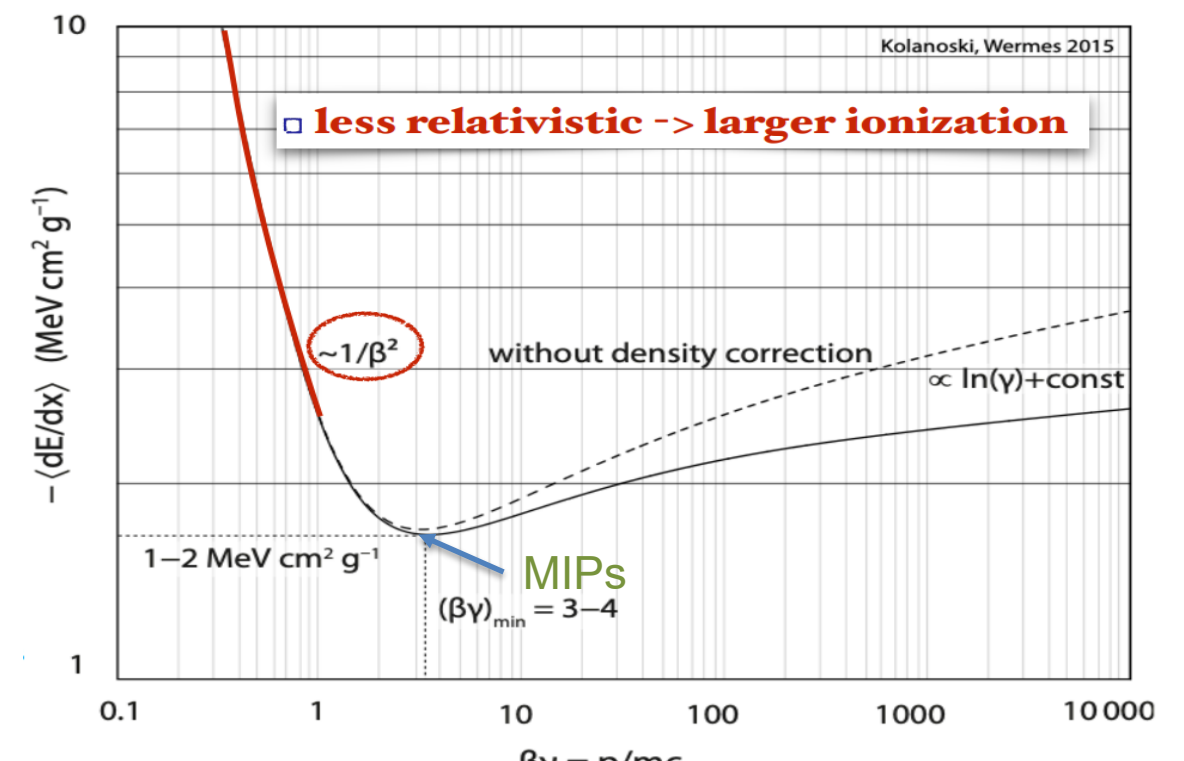


Figure 3. Bethe-Bloch curve with HSCP signal region in red

- in figure 4 marks the signal region for HSCPs on the bethe-bloch, but CMS design inclination is towards **Minimum Ionizing particles (MIPs)**, and may fail to detect less relativistic particles accurately.
- Mass is reconstructed by linearizing the first order Bethe-Bloch parameterization. Using the track's estimated dE/dx MPV and its momentum p , we solve:

$$\frac{dE}{dx} \approx f\left(\frac{1}{\beta}\right) \approx f\left(\frac{m^2}{p^2}\right)$$

$$\frac{dE}{dx} = K \frac{m^2}{p^2} + C$$

.....K and C are extracted from a Bethe-Bloch fit to the proton reference band in data [fig. 4]

$$mass_{rec} = p \sqrt{\frac{\frac{dE}{dx} - C}{K}}$$

.....extrapolated to any mass

- dE/dx** is our **direct mass handle**, sharpening the estimator tightens the HSCPs mass peak and boosts discovery reach.

Data and Methods

Dataset.

2018 HSCP AOD(Analysis Object Data)1800 GeV gluino signal simulation sample ($\approx 136k$ tracks HSCP based filtered to $\approx 15k$ tracks for mass reconstruction) –no background.

Event filters.

- HLT_Mu50 trigger • IsoTrack $p_T > 55$ GeV • $|\eta| < 1$ • $\chi^2/\text{dof} < 5$
- IsoTrack_isHighPurityTrack • IsoTrack_dxy $< 0.02\text{cm}$ • IsoTrack_dz $< 0.1\text{cm}$
- IsoTrack_fractionOfValidHits > 0.8

Baseline estimator.

- Harmonic-2 (I_h).
- f-truncated mean ($f = 10 - 40\%$)

Alternative estimators prototyped.

- Landau-MPV per-track fit with adaptive **Freedman-Diaconis (FD)** binning; 90 % convergence in 0.44 ms/track.
- Per-track adaptive truncator (gaussianity checker).

Early Diagnostics

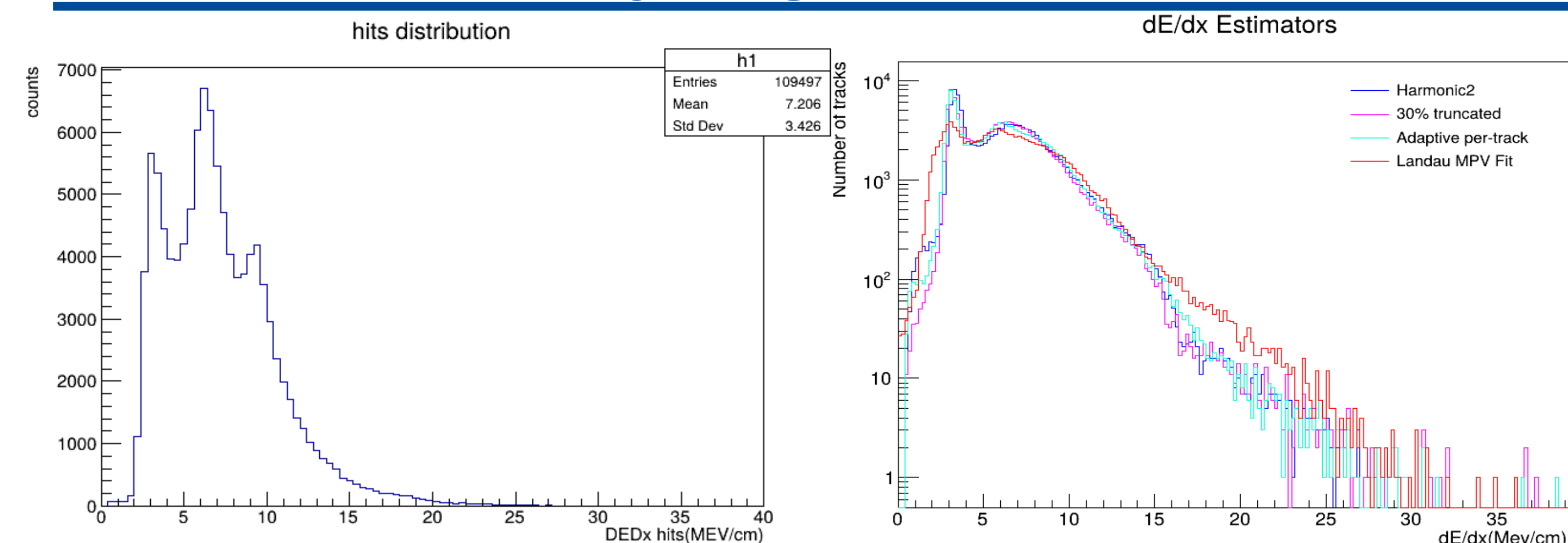


Figure 5. Sample Hit-level dE/dx spectrum — raw Landau-like distribution that every estimator must tame.

- Landau-skewed with 109 k entries, modal (dE/dx) ≈ 7.2 MeV cm^{-1} and $\sigma \approx 3.4$ MeV cm^{-1} . The extended right tail is precisely what biases simple averages and motivates robust estimators.

This **asymmetry and long right-hand tail** originate from fundamental stochastic and detector-level effects:

- Landau fluctuations:** Most ionizations events yield small energy losses, but rare collisions with atomic electrons create **δ-rays (knock-on electrons)** that carry away large energy, dominating the non-Gaussian tail.
- Detector artefacts:** Finite charge-collection efficiency, electronic noise, and imperfect path-length corrections broaden the low-energy side, while cluster splitting or merged hits enhance the upper tail.
- Geometry variations:** Tracks traversing sensors at shallow angles deposit energy over longer paths, raising measured dE/dx and widening the spread.

Figure 6. Comparison of dE/dx Estimators for IsoTracks (log y-scale)

The distributions compare several per-track energy-loss estimators used on the CMS data: **Harmonic-2 (I_h)**, **30 % truncated mean**, and two prototype **adaptive per-track truncator estimator** and a **Landau-MPV fit estimator**.

- Harmonic-2 (I_h)** — the CMS Run-2 baseline — applies inverse-square weighting that down-weights large strip hits, reducing σ/μ by $\approx 15\%$ relative to the arithmetic mean while retaining a long high-ionization tail (residual outliers > 20 MeV cm^{-1}).
- Landau-MPV fit** performs a per-track Landau fit; its fitted core is $\approx 5\%$ narrower than the arithmetic mean yet still broader than a 30 % truncation. About 5 % of tracks fail or yield near-zero fits, suggesting that refitting or seed tuning could further tighten the distribution.
- 30 % truncated mean** hard-cuts the upper 30 % of hit charges, shrinking σ/μ by $\approx 20\%$ and almost eliminating > 20 MeV cm^{-1} entries. The trade-off is partial loss of legitimate highly-ionizing HSCP signals.

- Adaptive per-track estimator** dynamically trims or reweights hit tails per track, seeking an optimal Gaussian-like core while preserving genuine high-ionization responses.

In the overlay, all estimators converge near the main MIP peak ($\sim 3\text{--}4$ MeV cm^{-1}) but diverge in the non-Gaussian tail region.

Prototype Estimators

Landau-MPV Fit Estimator

Rather than reducing the Landau-skewed hit spectrum with a weight-based average, we use a **Minuit regressor to fit a Landau function to each track's dE/dx histogram and take the fit's most-probable value (MPV)** as the estimator. Because the Landau model describes the particle's collisional energy loss through ionization, the method should yield an **unbiased central value—once pathological outlier hits are tamed**.

Adaptive bins for stable χ^2 .

With $> O(10^4\text{--}10^6)$ —tracks, a fixed binning scheme would either over-smooth fine structure or leave vast swaths of empty bins, inflating the χ^2 that drives the Minuit fit. We therefore histogram **each track with the Freedman-Diaconis rule**

$$\text{bin width} = \frac{2(q_3 - q_1)}{\sqrt[5]{n}}$$

which minimizes the asymptotic **Integrated Mean-Squared Error (IMSE)** between the histogram and its probability density function.

Intuitive seeding $\rightarrow$ 90 % convergence.

MPV guess = modal-bin centre, amplitude = modal-bin height, $\sigma_{guess} \approx 0.3 \times \text{RMS}$. With these seeds, the Landau fit converges for $\approx 89\%$ in the test set.

Residual pathologies.

- $\sim 0.2\%$ return negative MPVs—each has at least one pathological dE/dx hit wildly deviating from the modal distribution of per track dE/dx hits; likely detector artefacts. *Interesting how the estimator catches them.*
- $\approx 5\%$ of **successful fits** give $\sigma \rightarrow 0$, dominated by tracks with ≤ 4 **strip clusters**. These cases inflate σ/μ and need a rescue strategy (re-seed or fall back to Harmonic-2).

Adaptive Truncation Estimator

For another approach to suppress pathological dE/dx hits, we developed a **variable-truncation (VT)** estimator that adaptively removes non-Gaussian tails on a per-track basis. Each track's hit distribution is iteratively analyzed to determine the optimal truncation fraction that minimizes shape asymmetry.

Algorithm workflow.

- Input:** vector of per-track hit dE/dx values.
- Initialization:** compute skewness and kurtosis; define candidate truncation levels (e.g. 10–40 %).
- Adaptive loop:** for each truncation level, iteratively exclude the highest hits and recompute Gaussian fit parameters.
- Convergence rule:** select the truncation level that minimizes $|\text{skew}| + |\text{excess kurtosis}|$ (i.e. most Gaussian).
- Output:** truncated-mean estimator $\langle dE/dx \rangle$ optimized for that track.

Preliminary mass reconstruction outlook

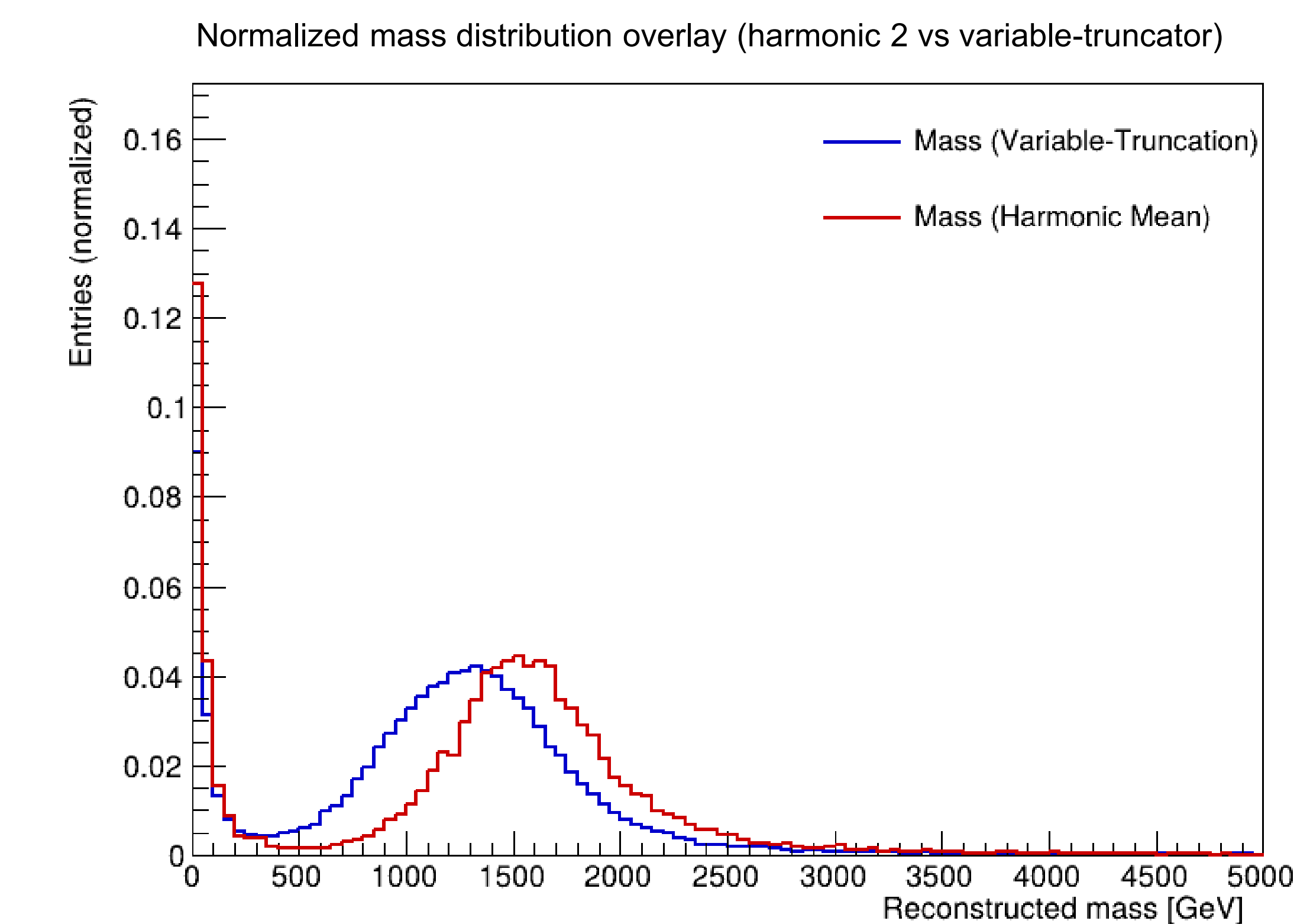


Figure 7. **Normalized comparison of reconstructed masses for Harmonic-2 (red) and Variable-Truncation (blue).**

The adaptive method shifts the peak slightly lower and tightens the width relative to other estimators, indicating improved bias suppression and stability in mass estimation.

Even with inconsistent track counts and an incomplete filtered input sample for the variable truncator, the resulting distribution is **notably more symmetric**, suggesting that the estimator's adaptive trimming inherently stabilizes the reconstructed mass against outlier contamination and imperfect preprocessing.

Work to do

- K and C constants are unique to each estimator; prepping new AOD data to feed into pipeline to extract K and C constants for each estimator.
- Implement all mass reconstructions using consistent filters and matched event selections across estimators to ensure unbiased performance comparisons and compare resolution as success criteria.
- Investigate Landau-MPV fit anomalous hits detection as basis for per-track hit masking.
- Single hit masking: train and develop an Isolation Forest anomaly detection algorithm to classify outliers.

Acknowledgements

This work was supported by the DOE grant RENEW-HEP: U.S. CMS SPRINT - "A Scholars Program for Research Internship" at Tougaloo College, MS (DE-SC0023681), Brown University (DE-SC0023651), RI, University of Puerto Rico-Mayaguez, PR (DE-SC0023680), and University of Wisconsin – Madison, WI (DE-SC0023643) and U.S. CMS Operations at the Large Hadron Collider (NSF-2121686).

