

Reconstruction of Ξ^- hyperons
from simulated Au+Au collisions
at 10 GeV/nucleon in the CBM
spectrometer using ATKFPF and
XGBoost

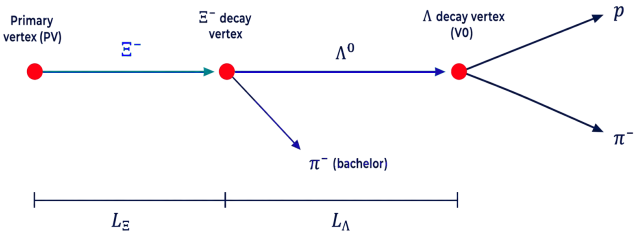
Illya Yurchanka

University of Warsaw, Faculty of Physics

September 14, 2026

Properties and decay topology of the Ξ^-

- Quark content: dss .
- Mass: $1321.71 \pm 0.07 \text{ MeV}/c^2$.
- Mean lifetime: $163.9 \pm 1.5 \text{ ps}$ [1].
- Reconstructed decay: $\Xi^- \rightarrow \Lambda \pi^-$, followed by $\Lambda \rightarrow p \pi^-$.



[1] Particle Data Group, P. A. Zyla et al. ‘Review of Particle Physics’. In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020), p. 083C01. doi: 10.1093/ptep/ptaa104

Simulated dataset

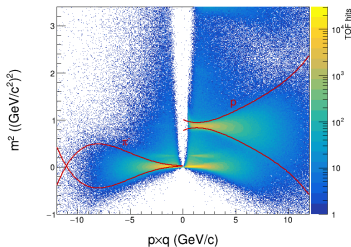
- Simulated Au+Au collision sample for CBM studies.
- Beam kinetic energy: 10 GeV/nucleon.
- Events produced with a PHQMD.
- Detector response propagated with GEANT3.
- Production campaign: July 2025.
- Data stored in .analysistree.root format.
- 5000 runs: 1000 training runs and 4000 reconstruction runs.

Location:

```
/lustre/cbm/prod/mc/phqmd52_winn/jul25p1_v18.8.2_jan24p5/  
geant3/auau/pbeam12agev/mbias/small_clusters/day-1/AT/*
```

Simulation sample and Ξ^- reconstruction

- ATKFPARTICLEFINDER [2] with Kalman-filter fits and default PID.
- Loose topological selections before XGBoost.

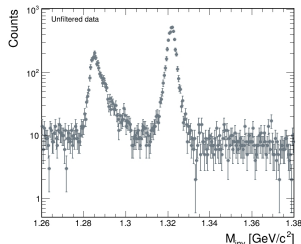
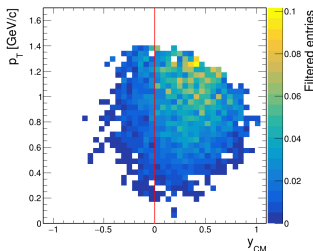


```
{ "ntuple_pdg": 3122, "ntuple_name": "Lambda", "daughters": ["Pi", "P*"] },
{ "ntuple_pdg": 3312, "ntuple_name": "Xi",
  "daughters": ["Pi", "PiLambda", "PLambda"],
  "v0_name": "Lambda" },
],
"reconstruction_list": [3312, 3122],
"store_tree_pdg": [3312, 3122],
"cuts": {
  "chi_primary_track": 1.0,
  "max_distance_between_particles": 2.0,
  "chi_primary_2d": 3.0,
  "chi2_2d": 25.0,
  "ldl_2d": 0.0,
  "ldl_track_particle_manybody": 5.0,
  "chi2_topo_track_particle_manybody": 5.0,
  "chi2_geo_track_particle_manybody": 6.0
},
"pid_cuts": {
  "track_length_min": 500.0,
  "track_length_max": 1400.0,
  "track_to_f_time_min": 16.0,
  "track_to_f_time_max": 62.0,
  "sp": [
    [0.056908, -0.0470572, 0.0216465, -0.0021016, 8.50396e-05],
    [0.00943075, -0.00635429, 0.00998695, -0.0011527, 7.77811e-05],
  ]
}
```

ATKFPF cut configuration.

[2] Yingjie Zhou. *ATKFPARTICLEFINDER*. <https://git.cbm.gsi.de/z.yingjie/afkf>. Software repository

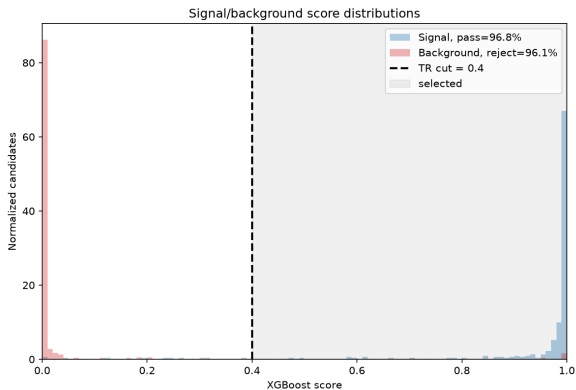
Candidate distributions after topological selection



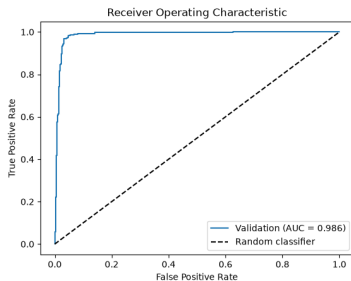
$$y_{\text{CM}} = y_{\text{LAB}} - y_{\text{NN}} \quad (1)$$

- Detector acceptance gives more candidates in the positive-rapidity range.
- Large combinatorial background remains after topological selections at low masses.

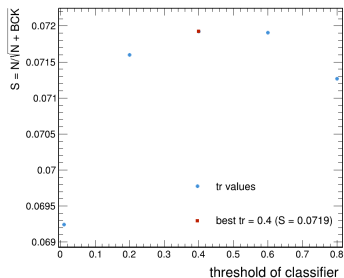
XGBoost discrimination



XGBoost discrimination and threshold optimization



ROC curve

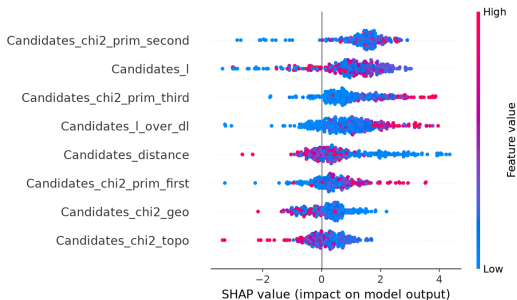


Signal significance scan

- XGBoost [3]: eight input features; 500 trees; maximum depth 3.
- Optimal score threshold: 0.4, chosen by maximizing signal significance.

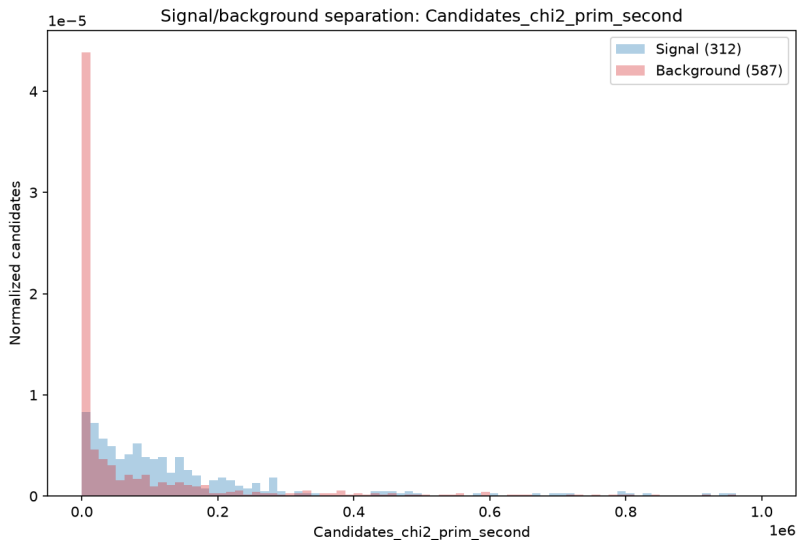
[3] Tianqi Chen and Carlos Guestrin. 'XGBoost: A Scalable Tree Boosting System'. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785–794. doi: 10.1145/2939672.2939785

SHAP feature attribution for XGBoost

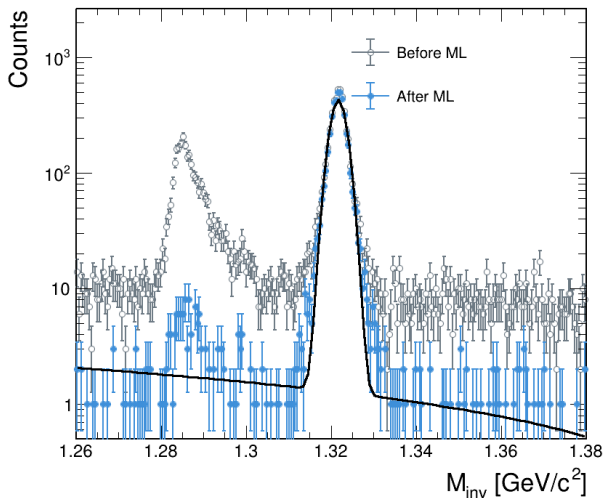


- Primary-vertex χ^2 of pion tracks ranks highest.
- Decay-length information also contributes strongly.

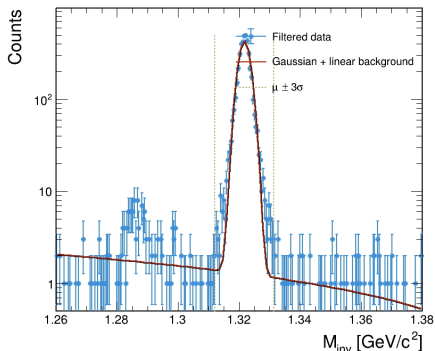
Distribution of χ^2_{prim} for the pion from Λ



Comparison before and after ML



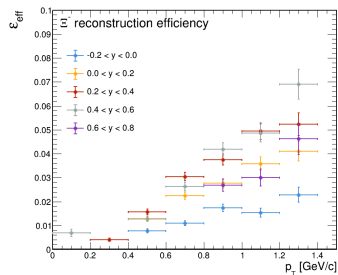
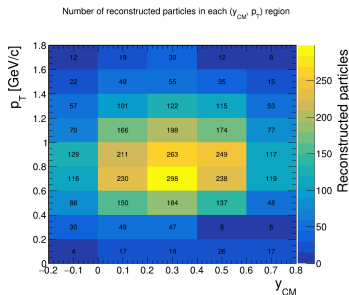
Invariant-mass fit and signal extraction



- Fit model:
Gaussian signal + linear background.
- Signal window:
 $\mu \pm 3\sigma$.
- Only sectors with significance > 2 are accepted.
- Yield: $S = T - B$.

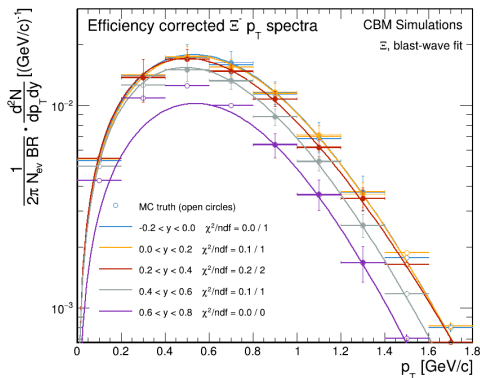
- Reconstructed mean: $\mu = 1.32171(3) \text{ GeV}/c^2$.
- PDG mass: $m_{\Xi^-} = 1.32171 \text{ GeV}/c^2$ [1].

Two-dimensional reconstruction efficiency



- Selected candidate counts depend on p_T and rapidity.
- $\epsilon = N_{\text{reco}}^{\text{true}} / N_{\text{MC}}^{\text{true}}$ in each bin.
- The correction accounts for acceptance and reconstruction losses.

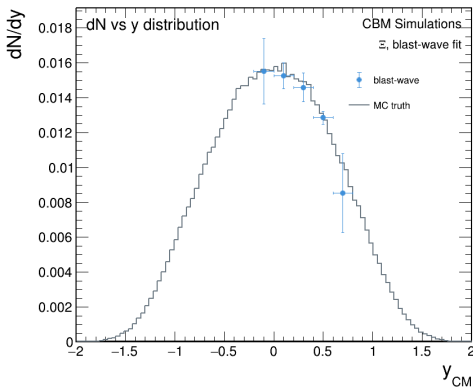
Efficiency-corrected transverse-momentum spectra



Blast-Wave model: Schnedermann, Sollfrank and Heinz, Phys. Rev. C 48, 2462 (1993) [4].

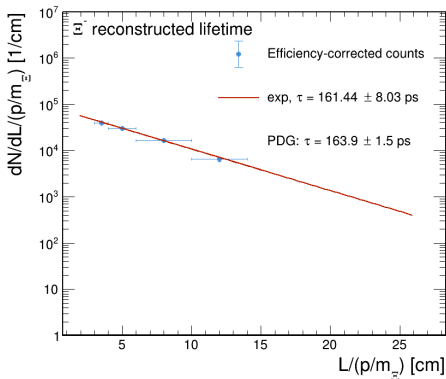
- Spectra are divided by efficiency and normalized by event count and branching ratio.
- Separate spectra are measured in rapidity intervals.
- Blast-Wave fits [4] extrapolate outside the measured p_T range.

Rapidity density of reconstructed Ξ^- hyperons



- Measured spectra are integrated using the trapezoidal rule.
- Blast-Wave integrals provide the unmeasured p_T tails.
- Reconstructed dN/dy agrees with MC truth within the statistical uncertainties.

Ξ^- lifetime fit



$$x = \frac{L}{\beta\gamma},$$

$$N(x) = N_0 \exp\left(-\frac{x}{c\tau}\right)$$

- Fit variable: decay length corrected for the Lorentz boost.
- Corrected yields are fitted in $L/(\beta\gamma)$ bins.
- Extracted lifetime is consistent with the PDG value [1].

[1] Particle Data Group, P. A. Zyla et al. 'Review of Particle Physics'. In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020), p. 083C01. doi: 10.1093/ptep/ptaa104

Conclusions and outlook

Conclusions

- ATKFParticleFinder and XGBoost can be used for Ξ^- reconstruction.
- Efficiency-corrected dN/dy agrees with MC truth within the statistical uncertainties.
- The fitted lifetime agrees with the PDG value.

Outlook

- Increase simulation statistics.
- Evaluate systematic uncertainties.
- With increased statistics, test finer binning.

References

[1] Particle Data Group, P. A. Zyla et al. 'Review of Particle Physics'. In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020), p. 083C01. doi: 10.1093/ptep/ptaa104.

[2] Yingjie Zhou. *ATKFPARTICLEFINDER*. <https://git.cbm.gsi.de/z.yingjie/afkf>. Software repository.

[3] Tianqi Chen and Carlos Guestrin. 'XGBoost: A Scalable Tree Boosting System'. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785-794. doi: 10.1145/2939672.2939785.

[4] Ekkard Schnedermann, Josef Sollfrank and Ulrich Heinz. 'Thermal phenomenology of hadrons from 200A GeV S+S collisions'. In: *Physical Review C* 48 (1993), pp. 2462-2475. doi: 10.1103/PhysRevC.48.2462.

[5] Haowen Deng et al. 'Ensemble learning for the early prediction of neonatal jaundice with genetic features'. In: *BMC Medical Informatics and Decision Making* 21, 338 (Dec. 2021). doi: 10.1186/s12911-021-01701-9.



Backup Slides

Fit functions: Blast-Wave

Blast-Wave model [4]

$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = A m_T \int_0^1 r dr I_0\left(\frac{p_T \sinh \rho(r)}{T}\right) K_1\left(\frac{m_T \cosh \rho(r)}{T}\right)$$

- $m_T = \sqrt{p_T^2 + m_0^2}$: transverse mass, with rest mass m_0 (natural units $c = k_B = 1$).
- T : kinetic freeze-out temperature, when elastic scattering ends.
- r : radius divided by the source radius, $0 \leq r \leq 1$.
 $\rho(r) = \tanh^{-1} \beta_T(r)$: transverse flow rapidity. A common profile is $\beta_T(r) = \beta_s r^{n_{\text{flow}}}$, with surface speed β_s and profile exponent n_{flow} .
- I_0 , K_1 : modified Bessel functions of the first kind (order 0) and second kind (order 1), from integration over emission azimuth and longitudinal source rapidity.

Fit functions: Lévy-Tsallis

$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = A \left(1 + \frac{m_T - m_0}{nT} \right)^{-n}$$

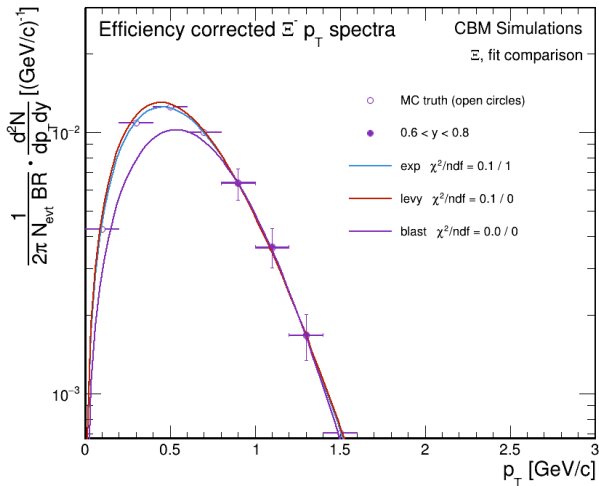
- $m_T - m_0$: transverse kinetic energy.
- $T > 0$: effective slope scale with energy units. Its temperature interpretation depends on the model.
- n : dimensionless exponent controlling the tail. Smaller n gives a heavier tail. Require $n > 2$ for a finite p_T -integrated yield over $0 \leq p_T < \infty$.
- As $n \rightarrow \infty$, the shape approaches $e^{-(m_T - m_0)/T}$. At sufficiently high p_T , it falls as p_T^{-n} . This n differs from the Blast-Wave flow-profile exponent.

Fit functions: m_T exponential

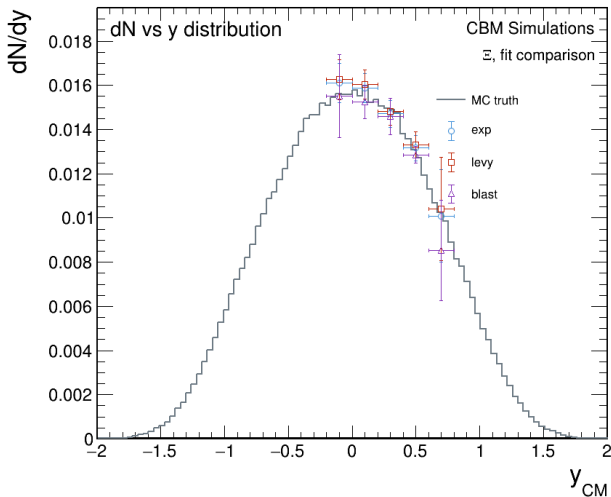
$$\frac{1}{2\pi p_T} \frac{d^2 N}{dp_T dy} = A e^{-m_T/T}$$

- $T > 0$: inverse-slope parameter with energy units. Larger T gives a flatter spectrum.
- Thermal Boltzmann suppression motivates the exponential shape. Collective flow can affect the fitted slope, so T alone does not determine the freeze-out temperature.
- Using $e^{-(m_T - m_0)/T}$ gives the same shape for fixed m_0 when the normalization is free.

p_T reconstruction comparison

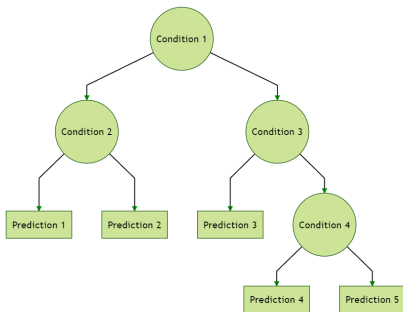


Rapidity distribution comparison



Decision trees for candidate selection

- Inputs: candidate features $\mathbf{x}$, such as decay geometry and track-fit quality. Training uses signal/background labels.
- Each condition $x_j < a$ splits candidates into two branches. Training learns the feature x_j and threshold a .
- A leaf assigns the final prediction or score. Different branches can use different cuts.
- Training chooses splits that improve the classification objective.

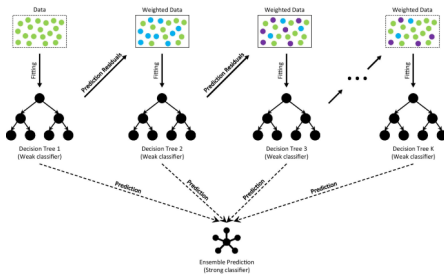


Source: TechyBuddy: ML algorithms, pros and cons

Boosted decision trees (BDT)

$$F_M(\mathbf{x}) = F_0 + \eta \sum_{m=1}^M f_m(\mathbf{x}), \quad s(\mathbf{x}) = \frac{1}{1 + e^{-F_M(\mathbf{x})}}$$

- Each new tree uses loss gradients to correct the ensemble.
- f_m : tree score, M : tree count, F_0 : initial score, η : learning rate.
- The binary logistic sigmoid maps F_M to $s \in [0, 1]$. Select candidates with $s > t$.



Source: Deng et al. (2021), BMC Med. Inform. Decis. Mak. 21, 338.

XGBoost: regularized gradient boosting

Extreme Gradient Boosting: efficient boosted trees using first and second derivatives of the loss.



$$\mathcal{L} = \sum_i \ell(z_i, F(\mathbf{x}_i)) + \sum_m \Omega(f_m), \quad \Omega(f) = \gamma L + \frac{\lambda}{2} \sum_{j=1}^L w_j^2$$

- z_i : true class label, ℓ : prediction loss (e.g. binary logistic loss), Ω : complexity penalty.
- L : number of leaves, w_j : leaf scores. γ penalizes extra leaves and λ suppresses large scores.
- Tree depth and random sampling of candidates or features control complexity. Early stopping uses validation performance to choose the number of boosting rounds.

Chen and Guestrin (2016), XGBoost: A Scalable Tree Boosting System. KDD, pp. 785-794.

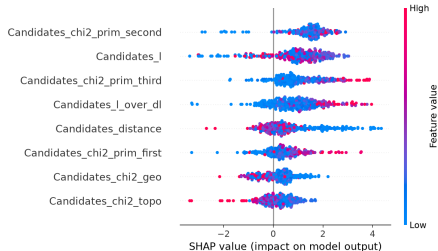
Logo: Wikimedia Commons.

SHAP: interpreting the classifier

SHapley Additive exPlanations

$$F(\mathbf{x}) = \phi_0 + \sum_{j=1}^d \phi_j$$

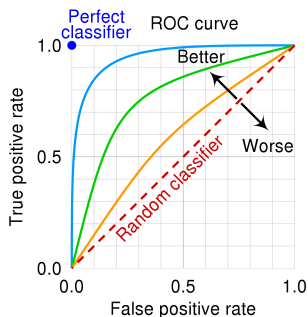
- ϕ_0 : baseline. ϕ_j : contribution of feature j (d in total).
- Positive: favours signal. Negative: favours background. Raw XGBoost output uses log-odds.
- Each dot: one candidate. Colour: feature value. Mean $|\phi_j|$ ranks importance.



ROC curves and selection thresholds

$$\text{TPR}(t) = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR}(t) = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

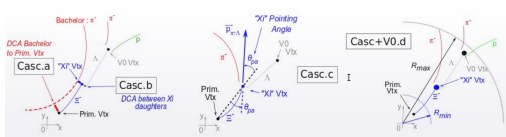
- **ROC**: signal efficiency (TPR) versus background acceptance (FPR), varying threshold t .
- TP / FN: accepted / rejected signal. FP / TN: accepted / rejected background. Rejection is $1 - \text{FPR}$.
- Upper left is ideal. AUC is 1 for perfect ranking, 0.5 for random ranking.



Source: Insight Media Group, ROC curve.

Method reference: `scikit-learn`, ROC metrics.

Ξ^- reconstruction



Source: Fizyka zderzeń jąder atomowych, Zakład Fizyki Jądrowej, Wydział Fizyki UW, Krzysztof Piasecki.

Topological cut parameters

- `chi_primary_track = 1.0`: daughter-track compatibility with the primary vertex.
- `max_distance_between_particles = 2.0`: maximum DCA between decay products.
- `chi_primary_2d = 3.0`: 2D candidate compatibility with the primary vertex.
- `chi2_2d = 25.0`: 2D secondary-vertex fit quality.
- `ldl_2d = 0.0`: transverse decay-length significance, $L/\Delta L$.
- `ldl_track_particle_manybody = 5.0`: decay-length significance for track-particle vertices.
- `chi2_topo_track_particle_manybody = 5.0`: topological fit quality.
- `chi2_geo_track_particle_manybody = 6.0`: geometrical vertex-fit quality.