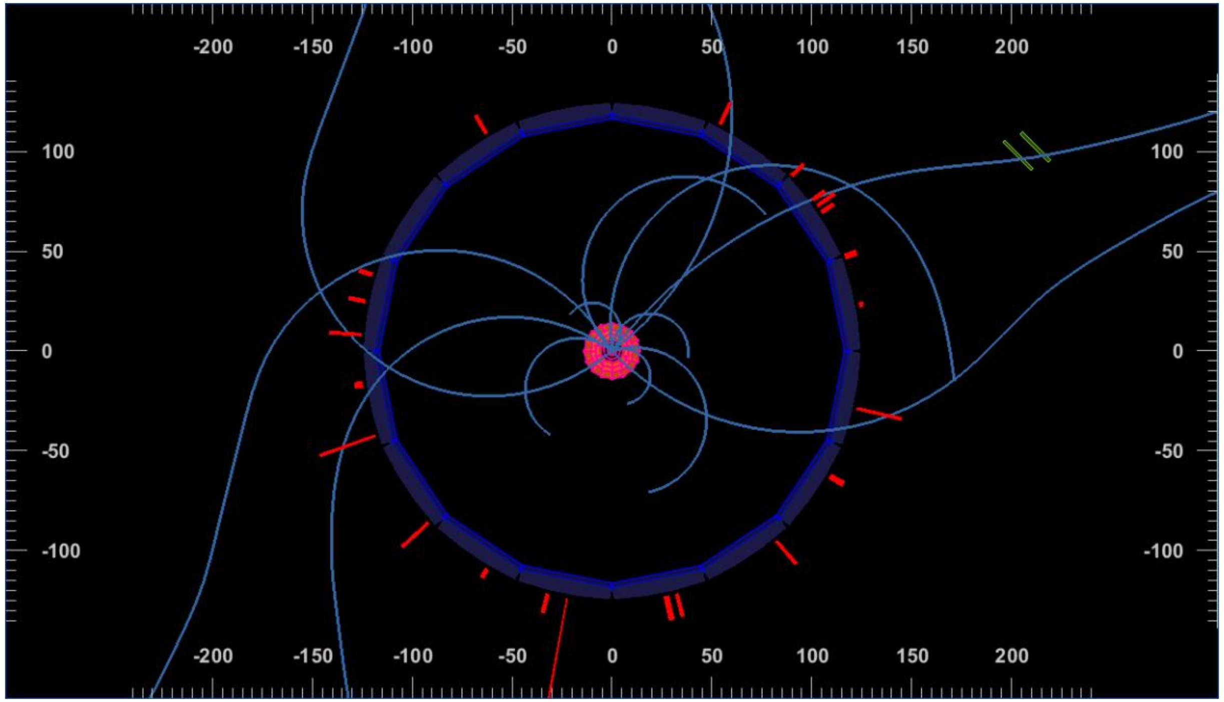
Reconstruction of Full Decays using Transformers and Hyperbolic Embedding at Belle II

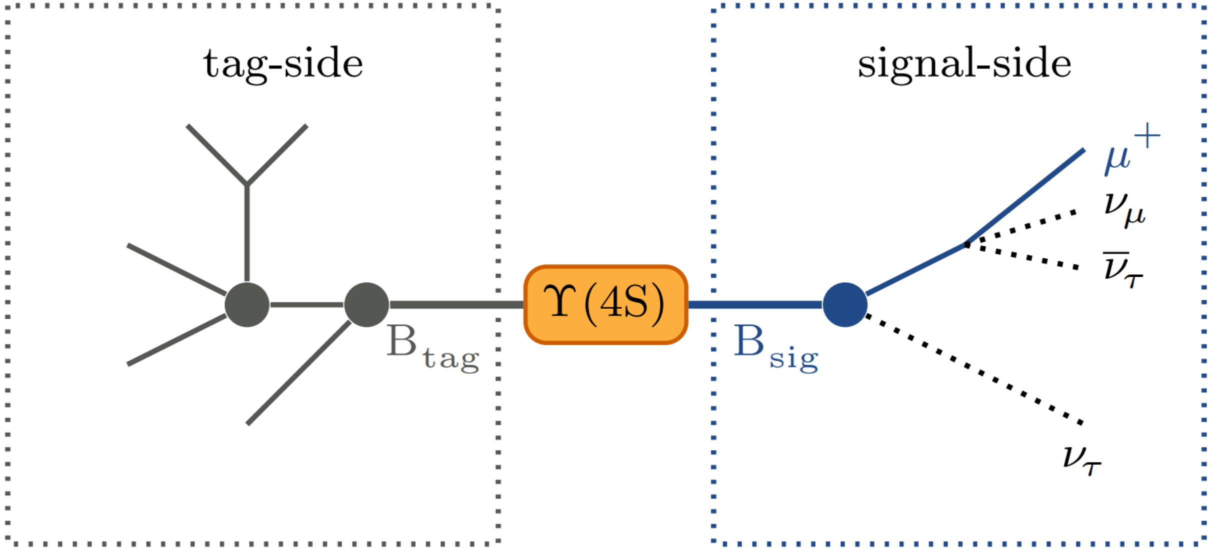
Boyang Yu, Nikolai Hartmann, Thomas Kuhr
Ludwig-Maximilians-Universität München

CHEP, Oct 22nd, 2024

Boyang.Yu@physik.uni-muenchen.de

Reconstruction of full decays

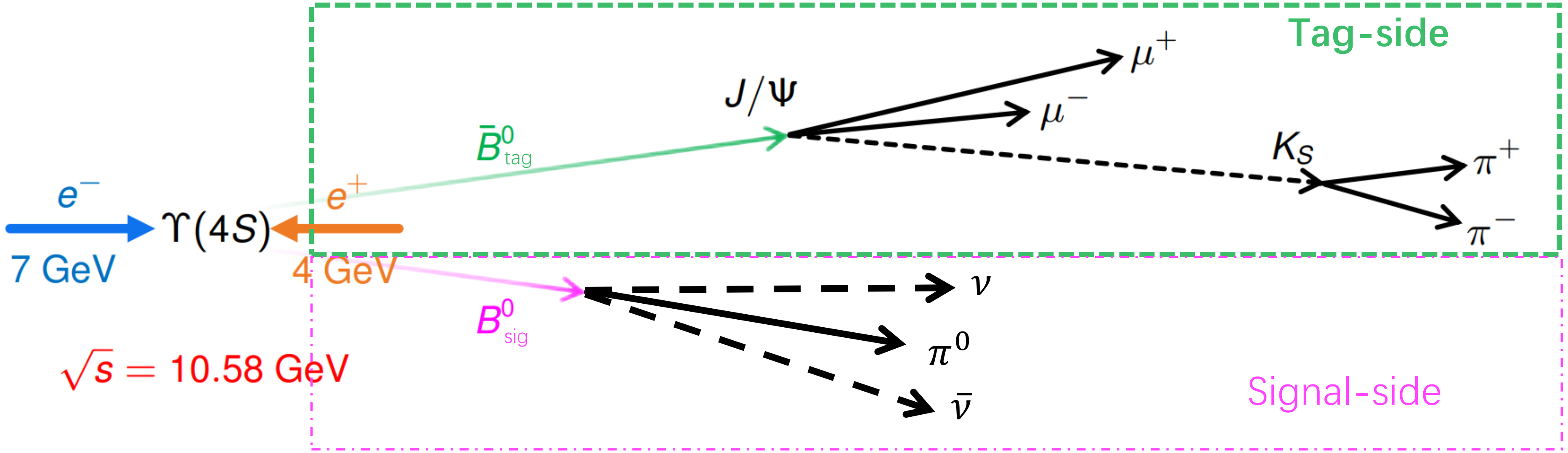




Detector information → Decay information

Tagging Method

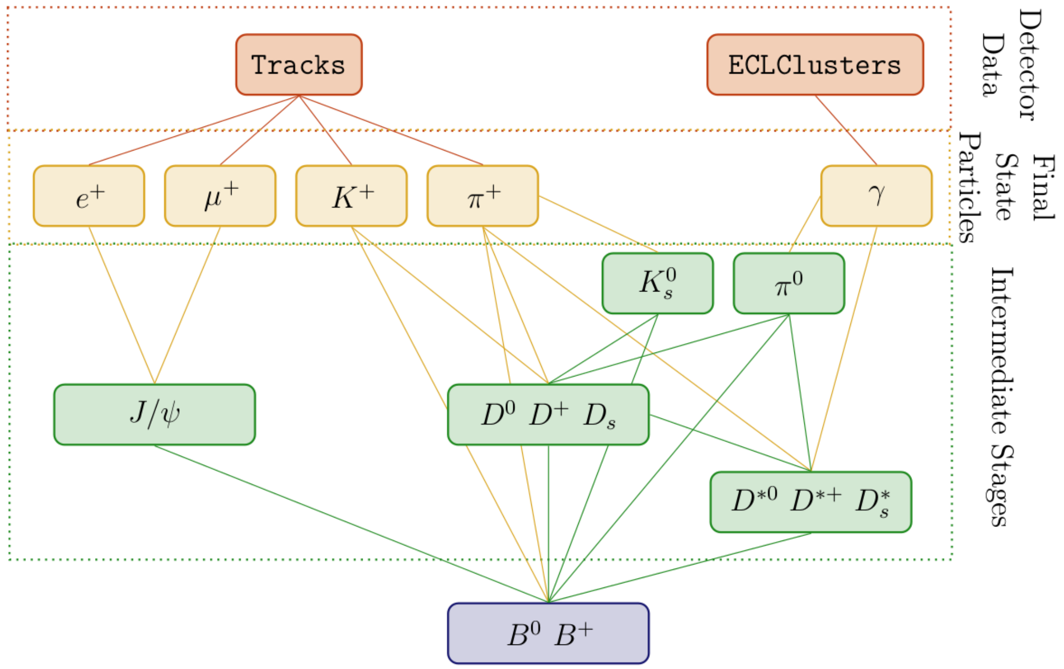
- Fully reconstruct one B meson, as tag-side
- Reconstruct the other side as signal under constraints on kinematics and flavour



Reconstruction of full decays

Metrics

- Purity: $\frac{\#\text{Correctly reconstructed}}{\#\text{Reconstructed}}$
- Efficiency: $\frac{\#\text{Reconstructed}}{\#\text{Total}}$



State-of-the-art @ Belle II: Full Event Interpretation(FEI)

- Binary classification of each individual mother-daughter combination using $\mathcal{O}(10^3)$ boosted decision trees (BDTs)
  -> High accuracy classification = high purity
- Hierarchical reconstruction of the whole decay tree in several stages

Limitations:

- Hard-coded decay channels for each particle
- Hard-coded particle types at each stage

-> Low efficiency: $\mathcal{O}(1\%)$

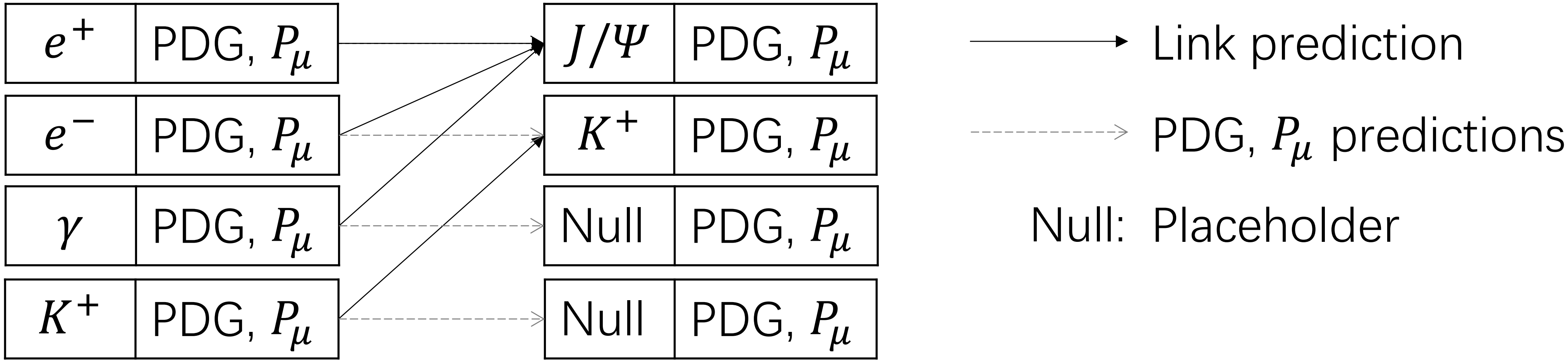
Improvement

For high efficiency:

- No restrictions on decay channels
  -> Predictions of particles instead of fixed combinations
  -> PDG (Particle type) + P_μ (Four momentum) + Link predictions

Example:

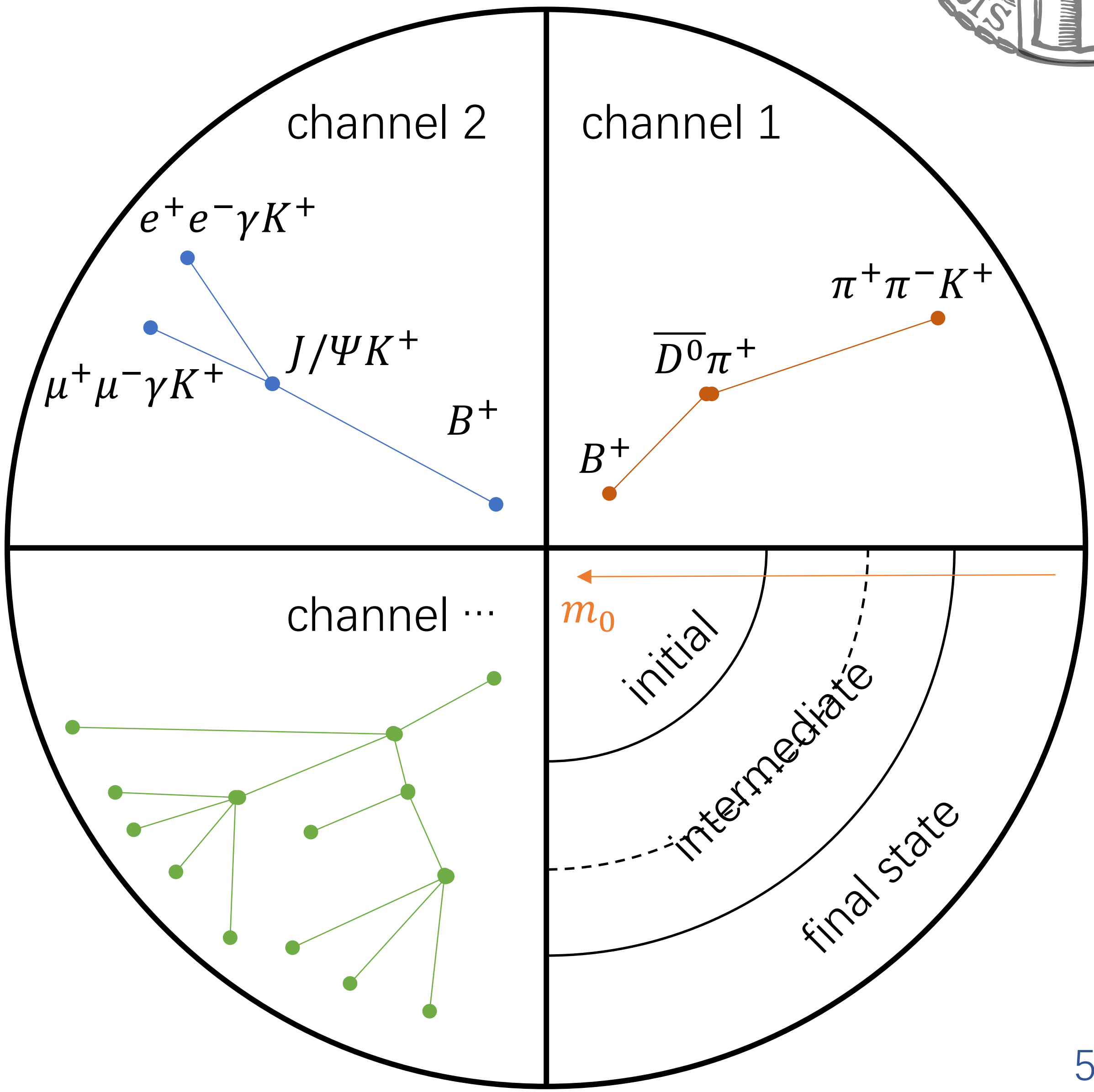
Given final state particles (including particle information): $e^+e^-\gamma K^+$



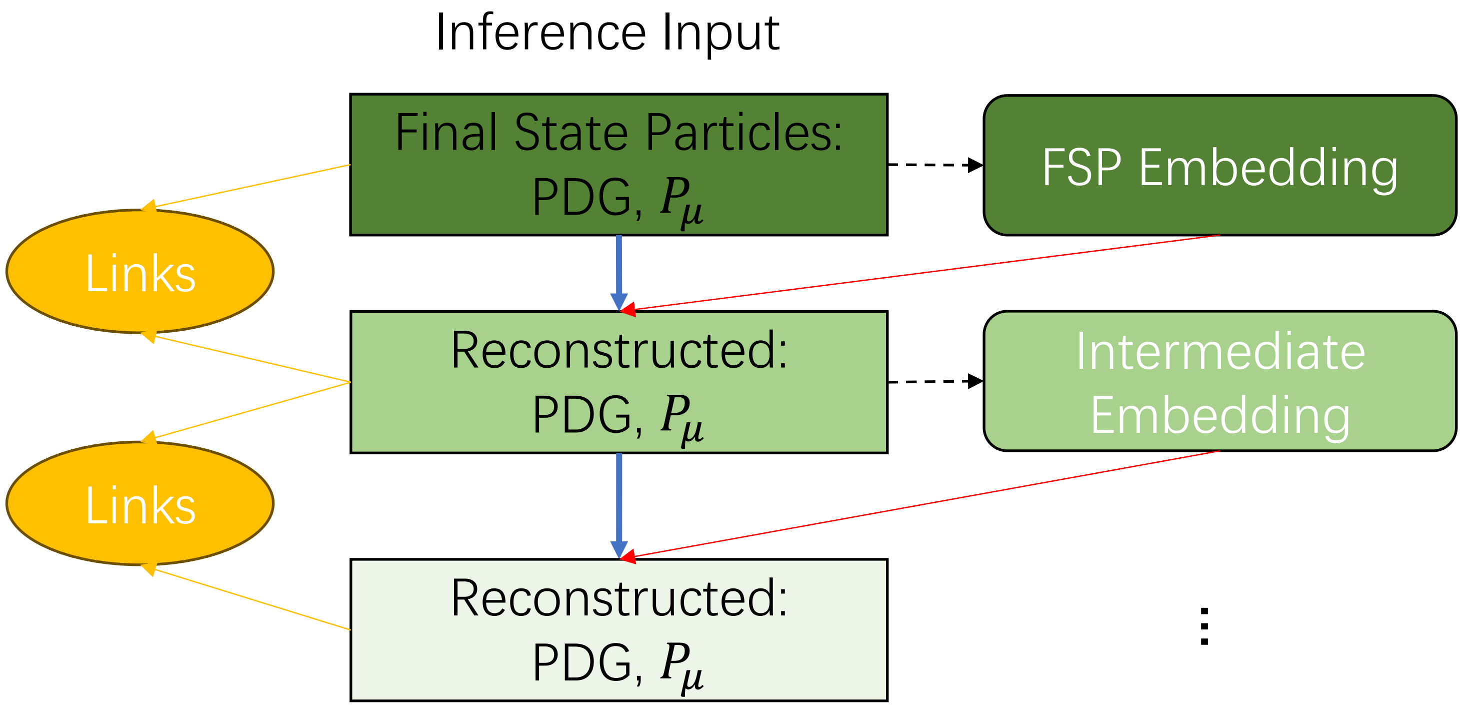
Improvement

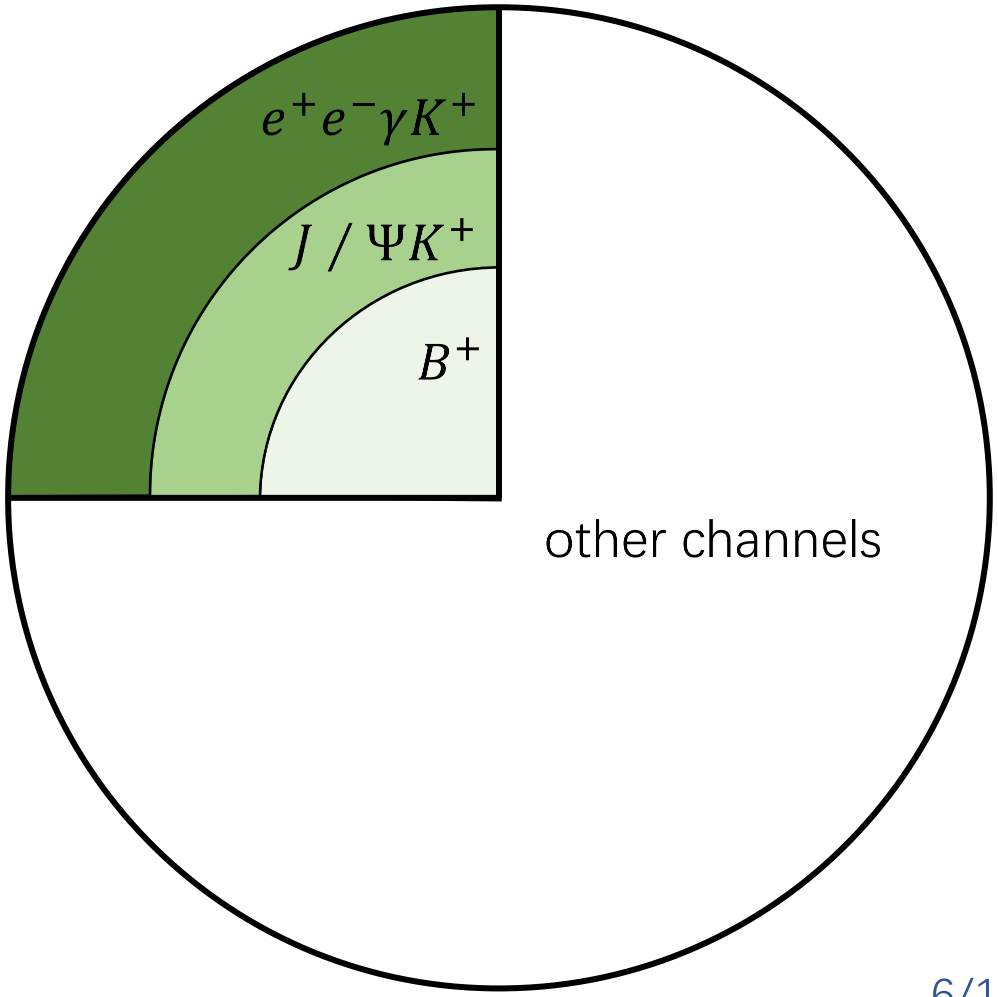
For high efficiency:

- No restrictions on available particle types at each stage
  -> Relaxed definition of stages/levels
- Continuous representation of the decay information in an embedding space to maintain high purity
  - Hyperbolic space for high representative power on hierarchical task
- Arbitrary number of channels
  - Decay pattern: dictionary of daughter PDGs of initial particle
  - Continuous channel comparison: cosine similarity of decay patterns



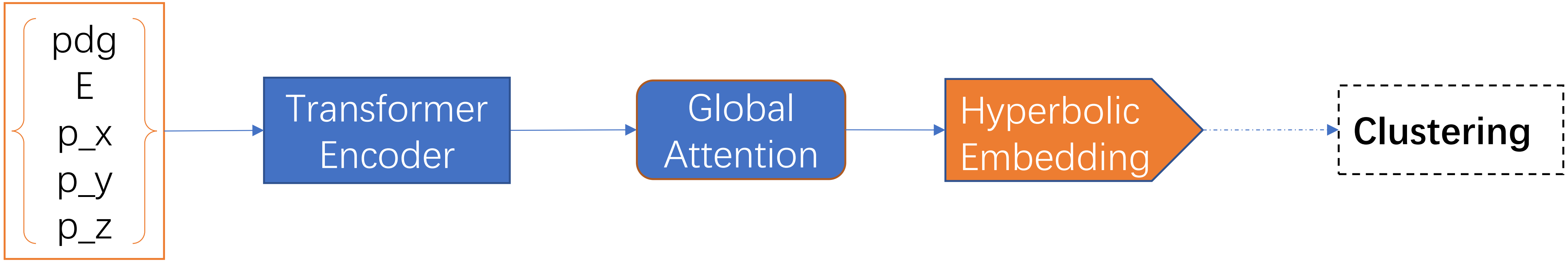
Hierarchical Reconstruction with Hyperbolic Embedding





- Training: on the whole decay tree
- Inference: start from Final State Particles (FSPs)
  - Repeat until the initial particle is found
  - Signal probability of whole reconstruction summarized from link predictions

Embedding

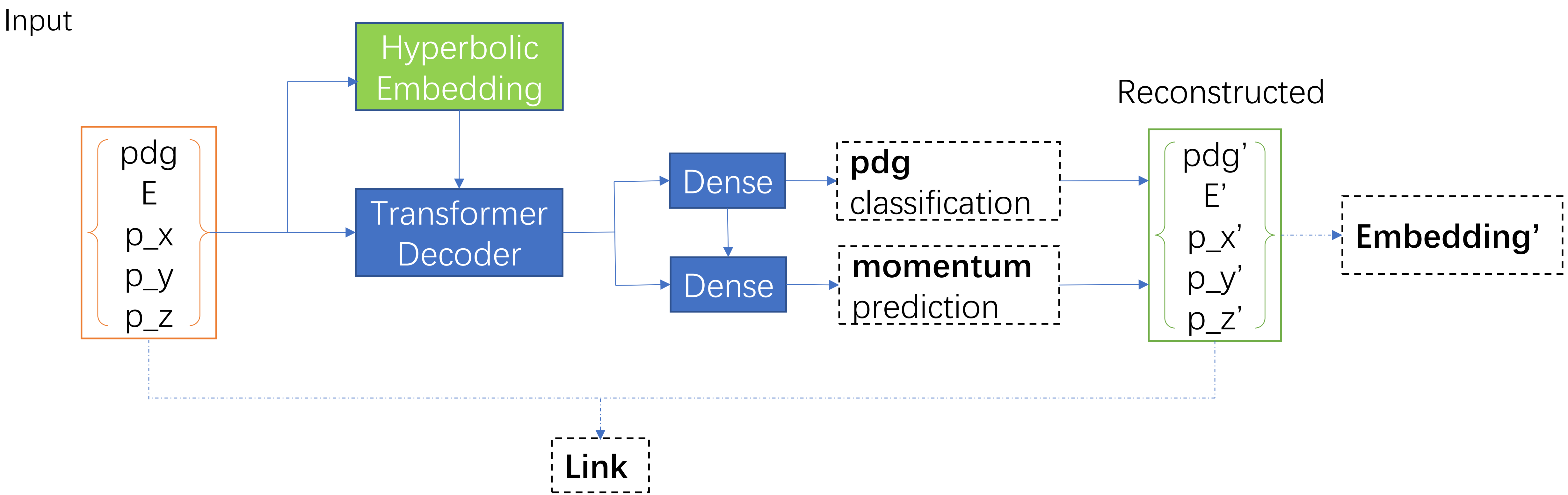


Clustering with combined loss:

- Continuous regularization loss:
    - Variance, Invariance, Covariance (VIC, arXiv:2105.04906) of embeddings
    - Similarity of decay patterns instead of positive-negative labels
    - Hyperbolic distance instead of Euclidean distance
- Radius loss
    - Mean square error between hyperbolic radius and $r_{\text{goal}} = a\sqrt{1 - m_i^*} + b$

      where $m_i^* = \frac{\max(m)@\text{level}_i}{m_{\text{initial}}}$, and a, b hyperparameters

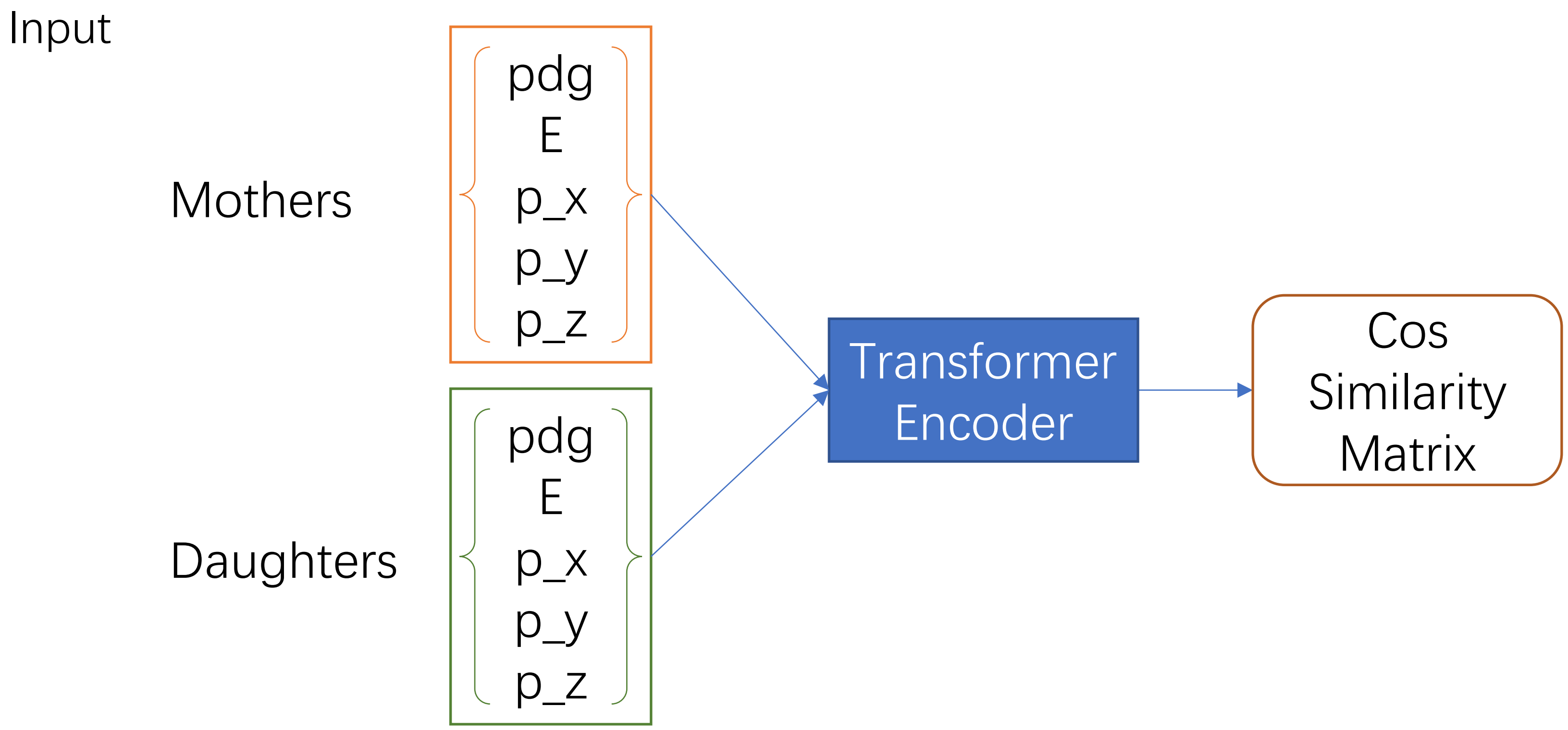
Reconstruction



Multi-task-losses:

- PDG classification: cross entropy
- Momentum prediction: mean absolute error
- Embedding loss: hyperbolic distance of embedding prediction with predicted particles
- Link loss: cross entropy of link prediction with predicted particles

Link Prediction



Link Prediction Loss: binary cross entropy of predicted similarity matrix

Full Reconstruction

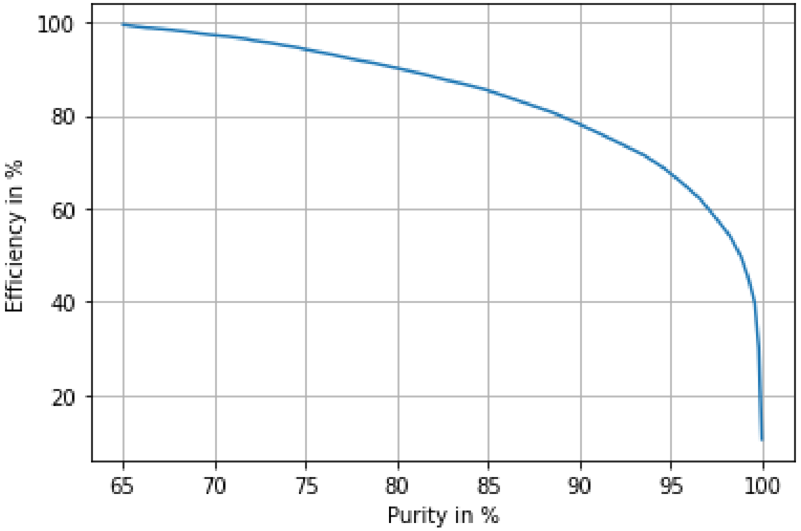
GraFEI dataset (DOI 10.5281/zenodo.6983257)

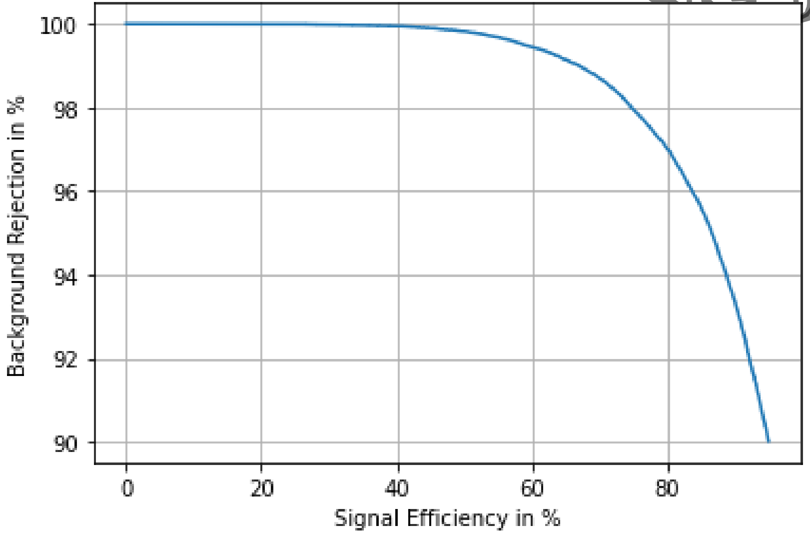
- Phase space decay following momentum conservation
- Simplification compared to generic dataset for analysis
  - Only 200 decay channels
  - No particle charge
  - No missing particles
  - No unused particles
  - Balanced decay channels

Full Reconstruction

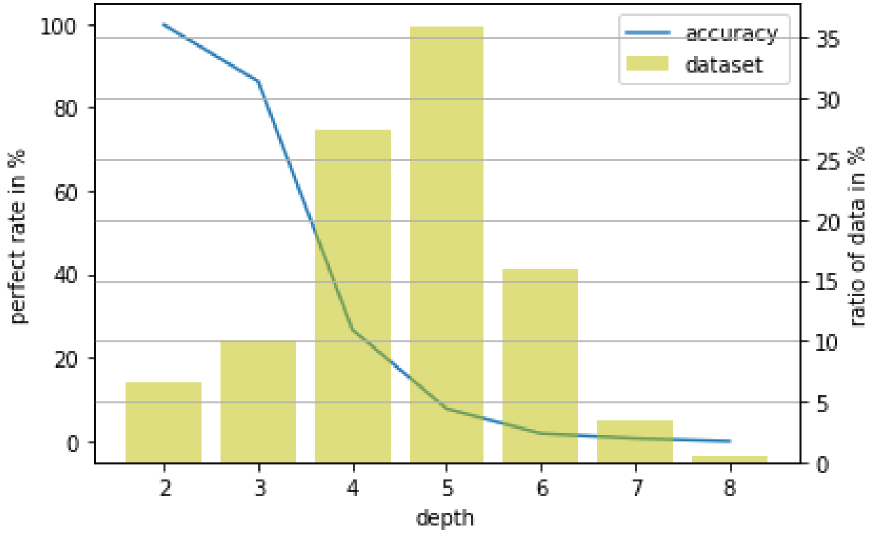
Performance on GraFEI dataset

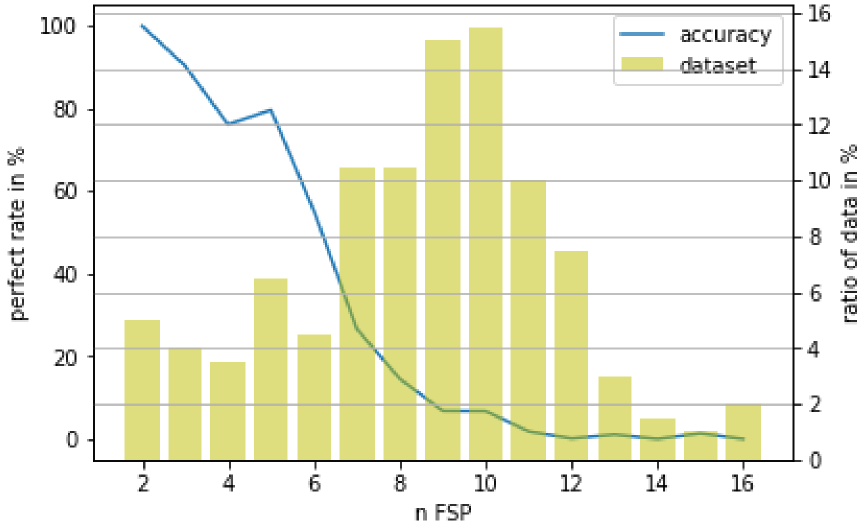
- Full reconstruction succeeded on all decays
- Efficiency is controlled by varying threshold for signal probability





Perfect reconstruction under 100% efficiency





Summary

Conclusion:

- HyperTagging showed reasonable performance on GraFEI dataset

On going:

- Improving the performance of HyperTagging with particle level embedding and reconstruction with GPT structure
- Training and evaluation of HyperTagging on generic dataset
- Implementation of HyperTagging in searching for $B \to \pi\nu\bar{\nu}$ analysis at Belle II

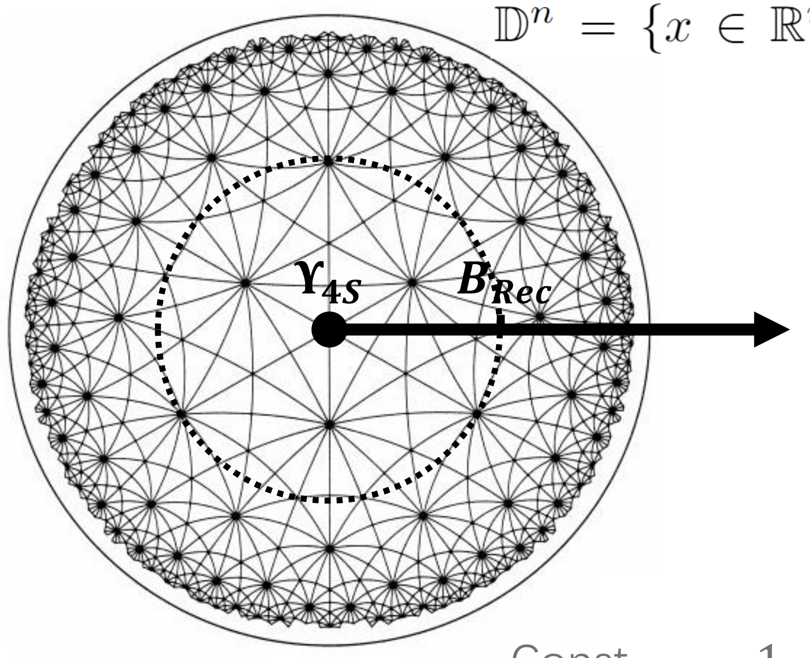
Thank You for your Attention

Boyang.Yu@physik.uni-muenchen.de

Backup

Hyperbolic Space (2D example – Poincare disc)

$$\mathbb{D}^n = \{x \in \mathbb{R}^n : c\|x\|^2 < 1, c \geq 0\}$$



Const. $\sim 1 - Radius^2$

$$E_{\text{ROE}} = E_{\Upsilon(4S)} - E_{\text{Reconstructed}}$$

$$\rightarrow r \sim \sqrt{E_{\text{ROE}}}$$

Properties:

- The size of an object with distance d to the centre: $1 - d^2$
  -> Embedded events will never reach the boundary
  -> Effective space near the boundary is infinite
- Volume of the space scales exponentially with radius
  -> Comparable to tree-structured data (decay relations)

Metrics:

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c\|\mathbf{y}\|^2)\mathbf{x} + (1 - c\|\mathbf{x}\|^2)\mathbf{y}}{1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c^2\|\mathbf{x}\|^2\|\mathbf{y}\|^2}$$

- Hyperbolic distance

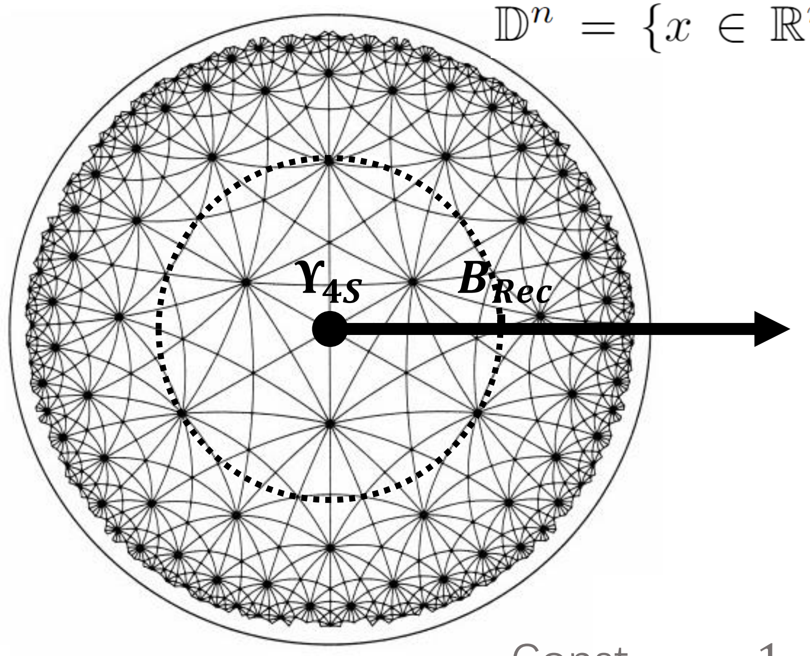
$$D_{hyp}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} \text{arctanh}(\sqrt{c}\| - \mathbf{x} \oplus_c \mathbf{y}\|)$$

- Hyperbolic angle/cosine similarity (the same as euclidical)

$$D_{cos}(\mathbf{z}_i, \mathbf{z}_j) = \left\| \frac{\mathbf{z}_i}{\|\mathbf{z}_i\|_2} - \frac{\mathbf{z}_j}{\|\mathbf{z}_j\|_2} \right\|_2^2 = 2 - 2\frac{\langle \mathbf{z}_i, \mathbf{z}_j \rangle}{\|\mathbf{z}_i\|_2 \cdot \|\mathbf{z}_j\|_2}$$

Hyperbolic Space (2D example – Poincare disc)

$$\mathbb{D}^n = \{x \in \mathbb{R}^n : c\|x\|^2 < 1, c \geq 0\}$$



Const. $\sim 1 - Radius^2$

$$E_{\text{ROE}} = E_{\Upsilon(4S)} - E_{\text{Reconstructed}}$$

$$\rightarrow r \sim \sqrt{E_{\text{ROE}}}$$

Properties:

- The size of an object with distance d to the center ~ $1 - d^2$
  -> Embedded events will never reach the boundary
  -> Effective space near the boundary is infinite
- Volume of the space scales exponentially with radius
  -> Comparable to tree-structured data (decay relations)

Ideal Embedding:

- Center: Singularity containing all full reconstructions of $\Upsilon(4S)$
  -> Empty rest of event (ROE)
- Bulk points: Partially reconstructed decays
- Points near boundary: Starting points of reconstructions
  -> The less reconstructed, the smaller branching ratio
  (taking less place in embedded space)
  -> Enable all possible decays

Hyperbolic metrics

- Hyperbolic distance

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c\|\mathbf{y}\|^2)\mathbf{x} + (1 - c\|\mathbf{x}\|^2)\mathbf{y}}{1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c^2\|\mathbf{x}\|^2\|\mathbf{y}\|^2}$$

$$D_{hyp}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} \operatorname{arctanh}(\sqrt{c}\| - \mathbf{x} \oplus_c \mathbf{y}\|)$$

- Hyperbolic angle/cosine similarity (the same as euclidical)

$$D_{cos}(\mathbf{z}_i, \mathbf{z}_j) = \left\| \frac{\mathbf{z}_i}{\|\mathbf{z}_i\|_2} - \frac{\mathbf{z}_j}{\|\mathbf{z}_j\|_2} \right\|_2^2 = 2 - 2\frac{\langle \mathbf{z}_i, \mathbf{z}_j \rangle}{\|\mathbf{z}_i\|_2 \cdot \|\mathbf{z}_j\|_2}$$

- Cross entropy losses w.r.t the two metrics for positive pairs (i, j)

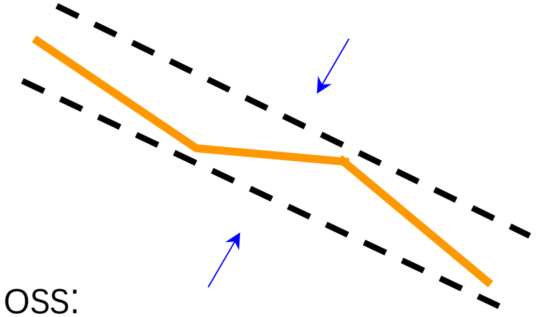
$$l_{i,j} = -\log \frac{\exp\left(-D(\mathbf{z}_i, \mathbf{z}_j)/\tau\right)}{\sum_{k=1, k\neq i}^{K} \exp\left(-D(\mathbf{z}_i, \mathbf{z}_k)/\tau\right)}$$

- Cross entropy losses for general cases
  (no well defined decay channels to specify "positive pairs")

$$l_{i,j} = -\delta_{c_i, c_j} \log \frac{\exp\left(-D(\mathbf{z}_i, \mathbf{z}_j)\right)}{\sum_k \exp\left(-D(\mathbf{z}_i, \mathbf{z}_k)\right)} \to -\log \frac{M_{i,j} \exp\left(-D(\mathbf{z}_i, \mathbf{z}_j)\right)}{\sum_k M'_{i,k} \exp\left(-D(\mathbf{z}_i, \mathbf{z}_k)\right)}$$

Contrast losses

- Intra loss:
  Align the samples from the same decay event



- Inter loss:
  Cluster/Separate the samples according to their decay channels