Hyperbolic Machine Learning for Jet Physics

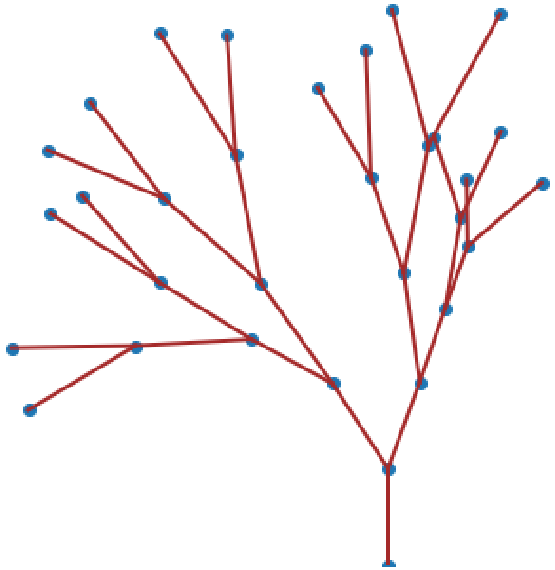
Nate Woodward (MIT 25’), Jeff Krupa, Sang Eon Park, Philip Harris

nswood@mit.edu

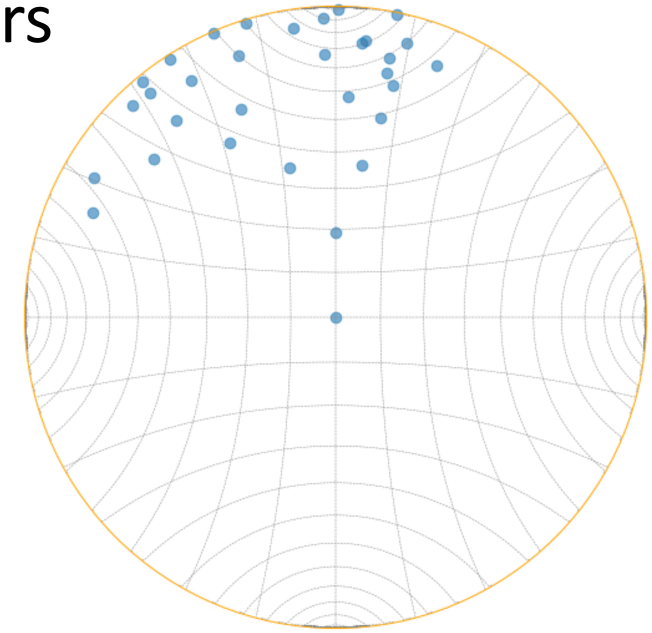
The NSF Institute for Artificial Intelligence and Fundamental Interactions

Motivation

- Optimal representations of jets are key for maximizing performance
- New jet representation fruitful for jet analysis (ParticleNet Qu et al. 1902.08570)
- Non-Euclidean geometries provide a powerful method to better represent physics data & reduce parameters



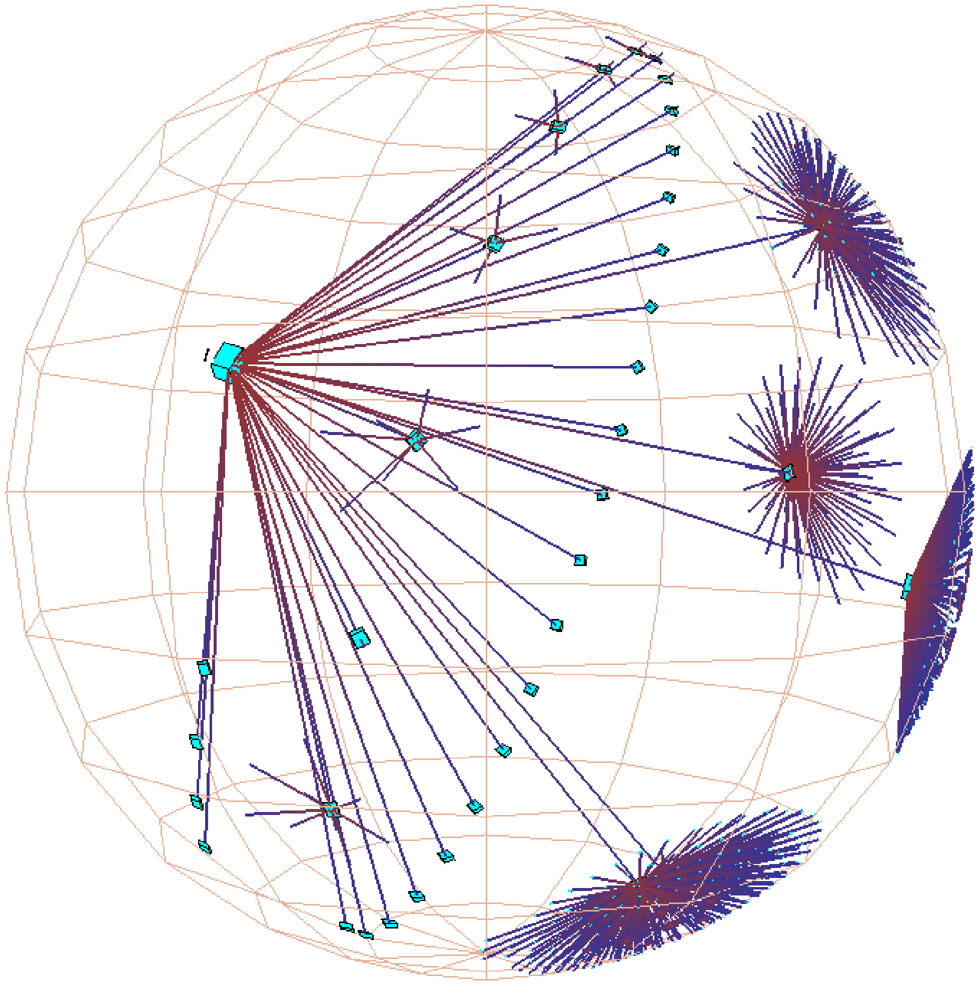
Graph Representation



Non-Euclidean Representation

Overview

- Hyperbolic geometries naturally represent inherent hierarchies
  - Exponential growth in surface area with distance
- Showering & hadronization exhibits hierarchical structure [Park et al. 2208.05484]
- Investigate optimal representation geometry
  - Quark & gluon jets exhibit varying jet curvature
- Need non-Euclidean aspect to fully represent jets?



Munzner, Tamara. ``Exploring Large Graphs in 3D Hyperbolic Space''

Goals

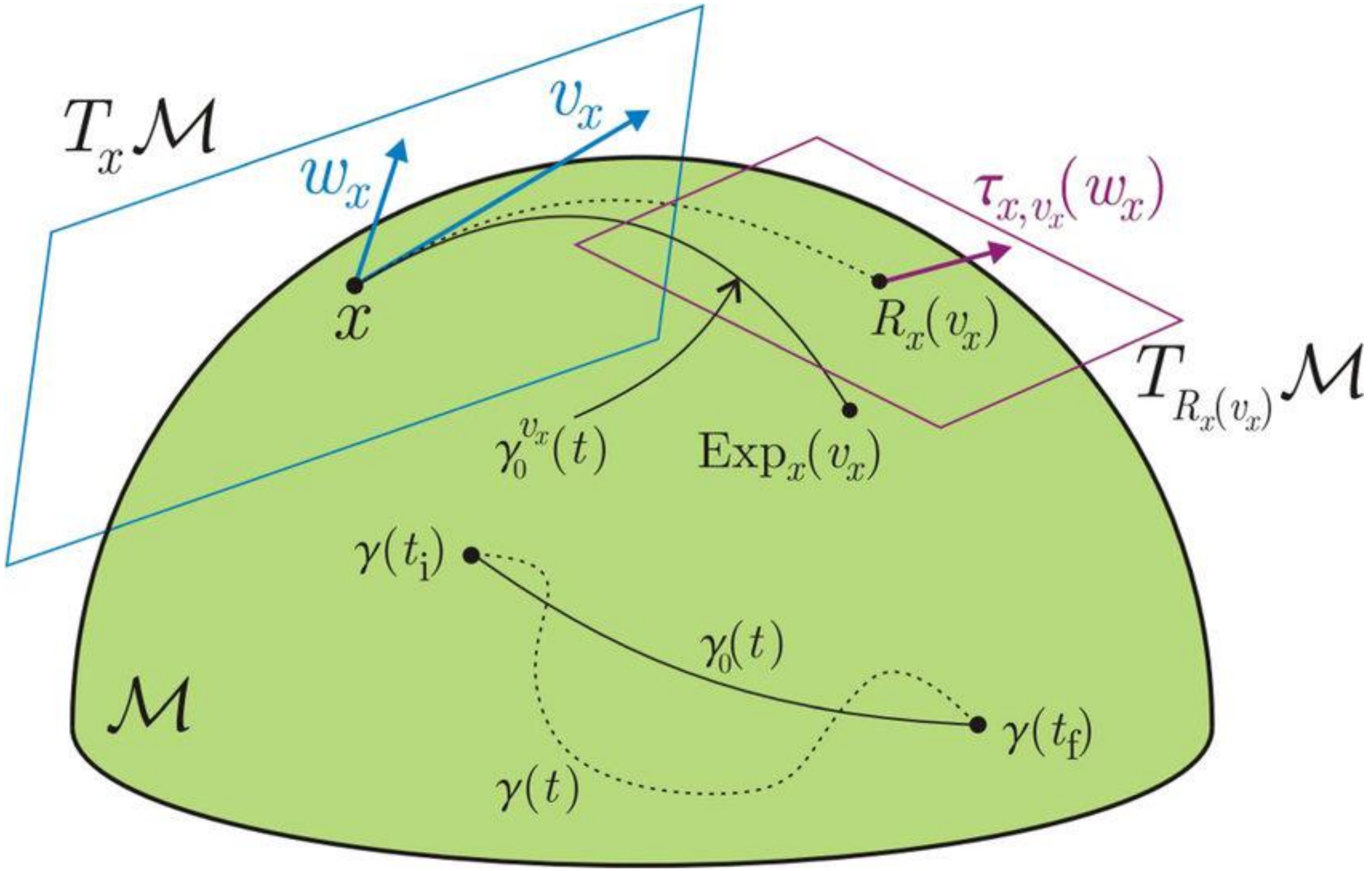
Demonstrate **equivalence** or **superiority** of geometric representations

Develop **low-parameter** models for jet classification in non-Euclidean geometries

Mathematical Prerequisites

Riemannian Manifolds

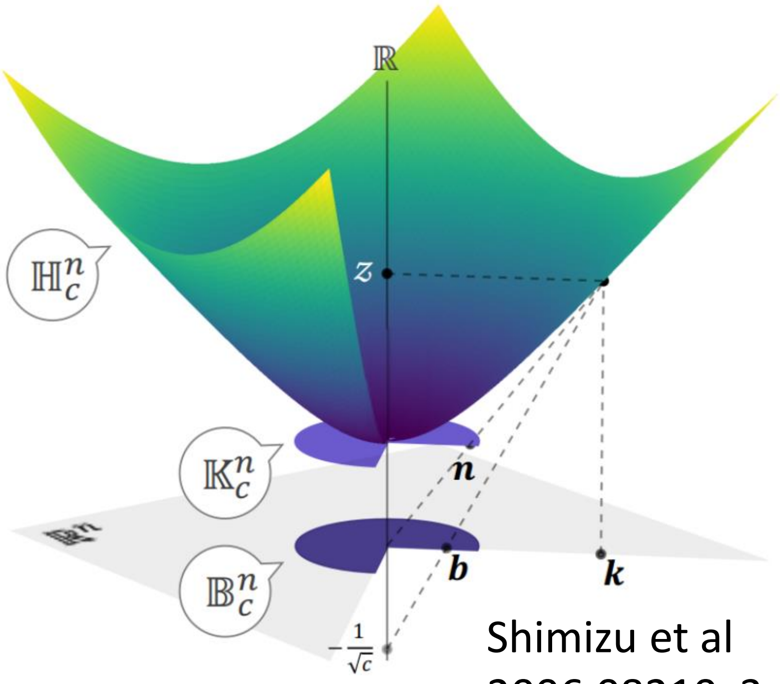
- Riemannian manifold M is a **real**, **smooth manifold** with a **positive definite inner product** on the tangent space $T_x M$ at each point.
- $Exp_x(v_x): T_x M \rightarrow M$
- $Log_x(z): M \rightarrow T_x M$



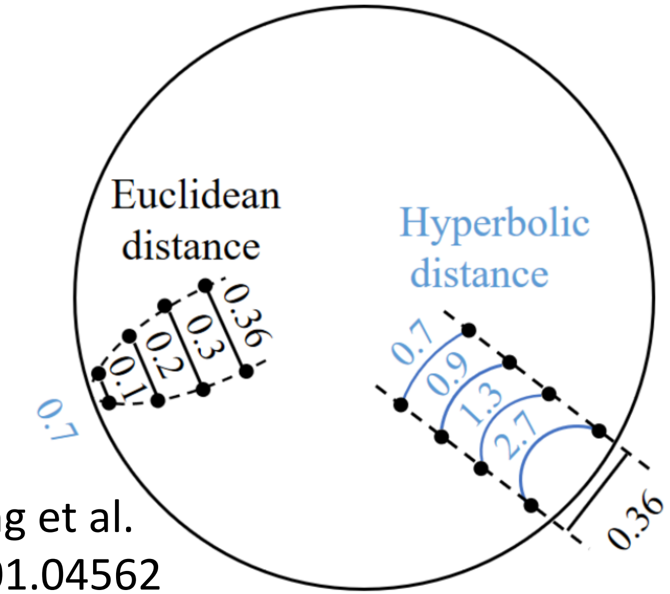
Luchikov et al. 2007.01287

Hyperbolic Geometry

- 5 Isometric representations: we choose the **Poincaré ball**
- Gyrovector spaces **replaces vector space** through Möbius transformations
- Employ the follow geometries:
  - Flat: $\boldsymbol{E^n}$
  - Hyp: $\boldsymbol{B_c^n}$
  - Simple Product: $\boldsymbol{E^{\frac{n}{2}} \times B_c^{\frac{n}{2}}}$
  - QG Product: $\boldsymbol{E^{\frac{n}{3}} \times B_{c_{quark}}^{\frac{n}{3}} \times B_{c_{gluon}}^{\frac{n}{3}}}$



Shimizu et al
2006.08210v3



Peng et al.
2101.04562

Gromov$-\delta$ Curvature

A metric space (X, d) is δ $-$hyperbolic iff for all $x, y, u, v \in X$

$$\big(d(x,y) + d(u,v)\big) - \big(d(x,u) + d(y,v)\big) \leq 2\delta$$

- Define space by min δ
- δ $-$hyperbolicity metric notion of **tree-likeness** [Adcock et al. 2013]
- **δ computable from jet data**

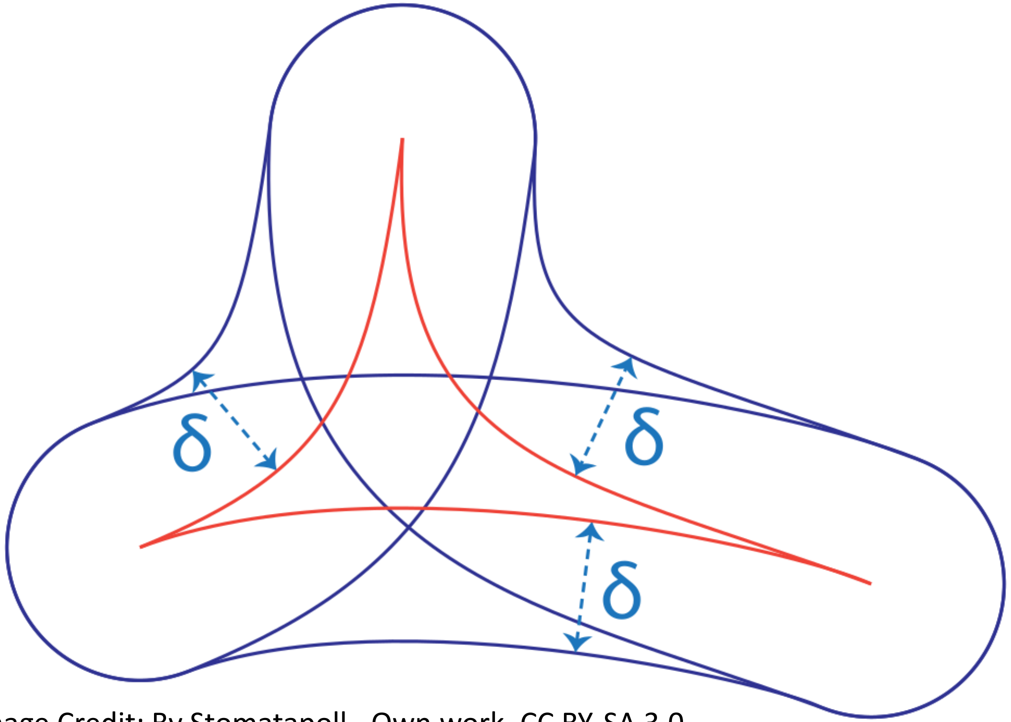


Image Credit: By Stomatapoll - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=22898943

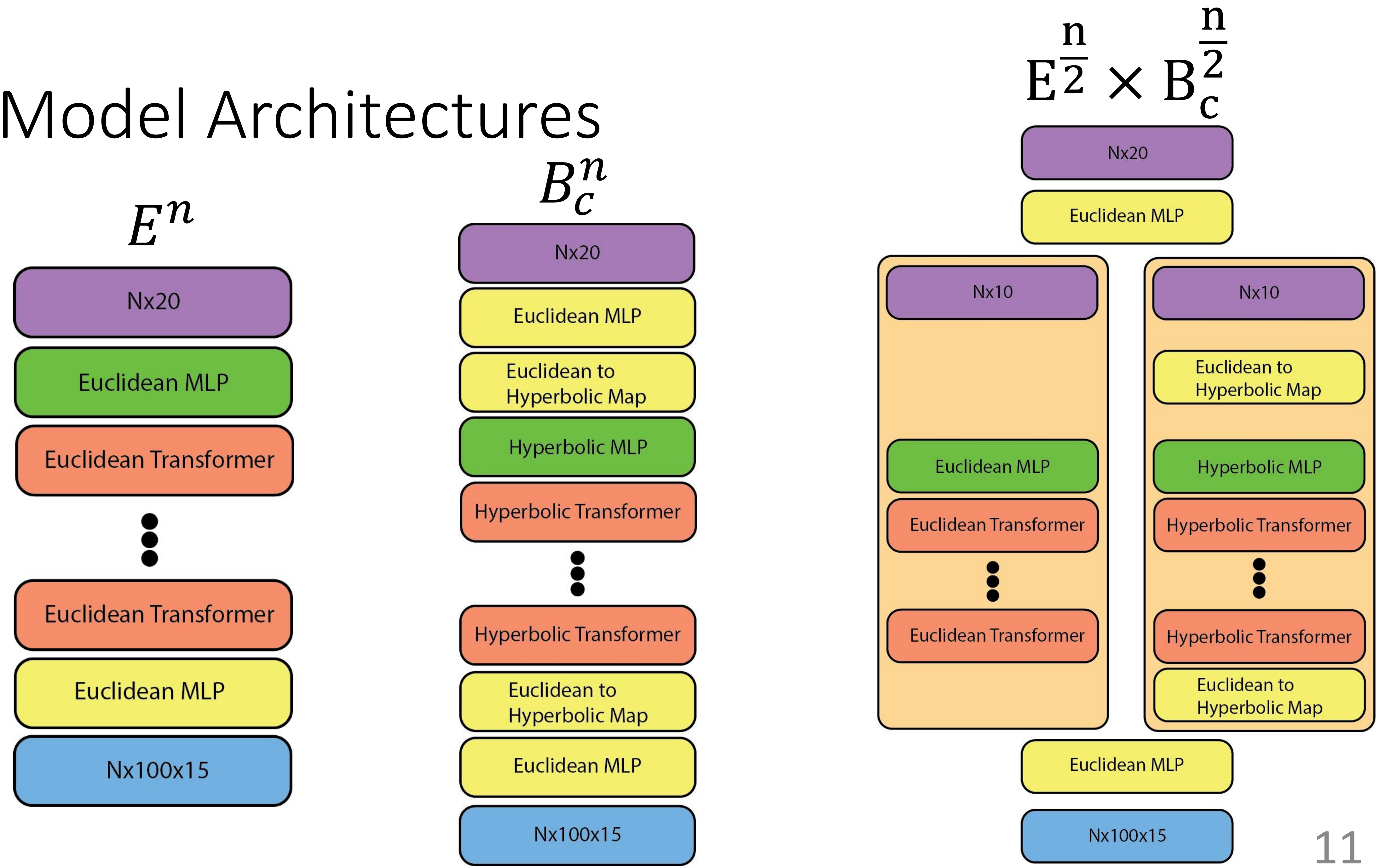
Riemannian Analogs for Euclidean ML

Euclidean	Riemannian Analog
Fully connected layer	**Gyrovector** matrix formulation or robust formulation from [Shimizu et al. 2006.08210v3]
Batch normalization	**Differentiable Fréchet** batch norm by [Lou et al. 2003.00335]
Attention aggregation	Hyperbolic distance metric & **Einstein weighted midpoint**
Optimization algorithms	Manifold aware versions of Adam, SGD, Lion

*Implementation specifics in *Additional Slides*

Model Architectures

Model Architectures



Physics Results

Quark Gluon Dataset Overview

- Public Quark-Gluon Jet Class dataset from Qu et al.
- 2 million training, 200k validation
- Zero-padded highest 100 pT particles per jet
- Particle features:
  - Particle pT/Jet pT
  - px,py,pz
  - η, ϕ
  - $\Delta\eta, \Delta\phi$
  - Charge
  - Particle hot encoding

Curvature Calculation for Quark & Gluon Jets

<u>Initial geometry:</u> B_c and $E \times B_c$, c is calculate from entire Q-G dataset

- $c = 1.139$

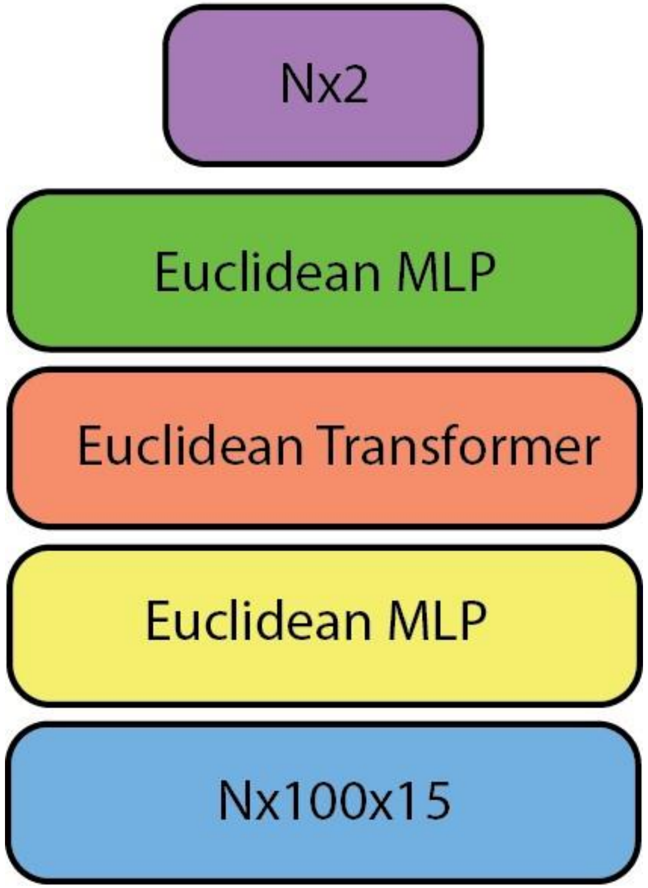
<u>Q-G Curvatures:</u> $E \times B_{c_{quark}} \times B_{c_{gluon}}$, calculate c separately

- $c_{quark} = 4.925$
- $c_{gluon} = 1.086$

Different physics can manifest through different curvature

Initial Tests

- Bottleneck models to output 2D representation of jet
  - **Only possible** 2D geometries are $\boldsymbol{E^2}$ and $\boldsymbol{B_c^2}$
- Minimize model architecture
- Scan over broad range of total model parameters



Example model

Results 2D Bottlenecking

Params	Flat ROC	Hyp ROC	Flat SE @ 5%	Hyp SE @ 5%
0.9k	0.8711	**0.8769**	0.4499	**0.4521**
1.4k	0.8823	**0.8871**	0.4660	**0.4800**
3.1k	**0.8889**	**0.8880**	**0.4874**	**0.4886**
9.4k	0.8821	**0.8928**	0.4604	**0.4905**
33k	**0.8941**	**0.8947**	0.5002	**0.5095**

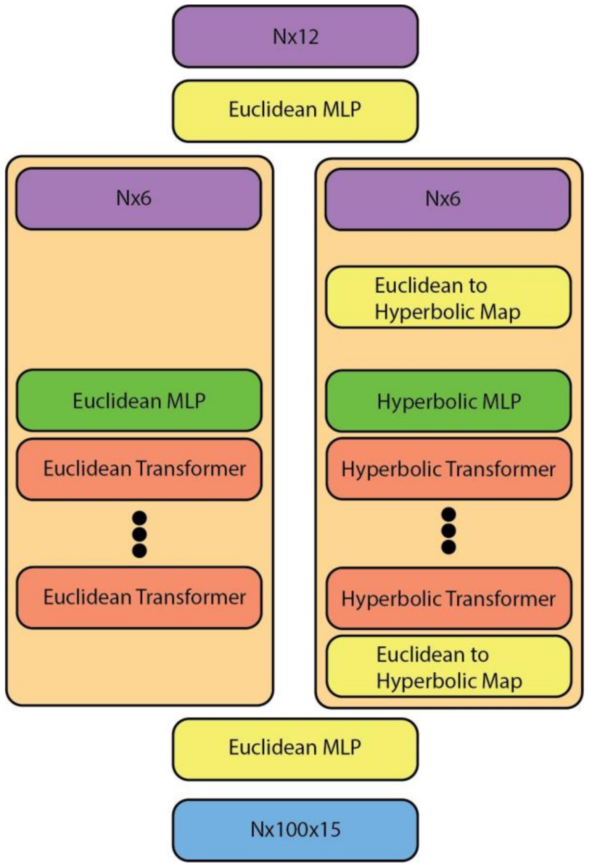
Product Space Geometric Learning

Inspired by work from Gu et al. 2019

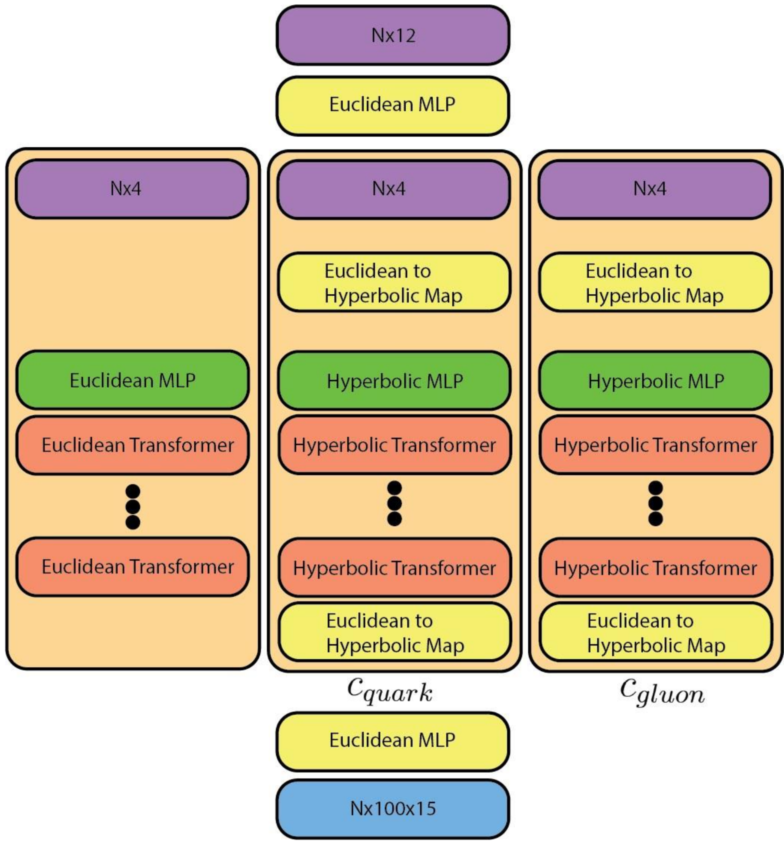
Product Space models

- Hyperbolic geometries succeed on small param models & low dimensions
- Now we explore physically motivated product space representations in 12D:
  - Simple Product: $\boldsymbol{E}^{\frac{n}{2}} \times \boldsymbol{B}_{c}^{\frac{n}{2}}$
  - QG Product:

$$\boldsymbol{E}^{\frac{n}{3}} \times \boldsymbol{B}_{c_{quark}}^{\frac{n}{3}} \times \boldsymbol{B}_{c_{gluon}}^{\frac{n}{3}}$$



$$\boldsymbol{E}^{\frac{n}{2}} \times \boldsymbol{B}_{c}^{\frac{n}{2}}$$



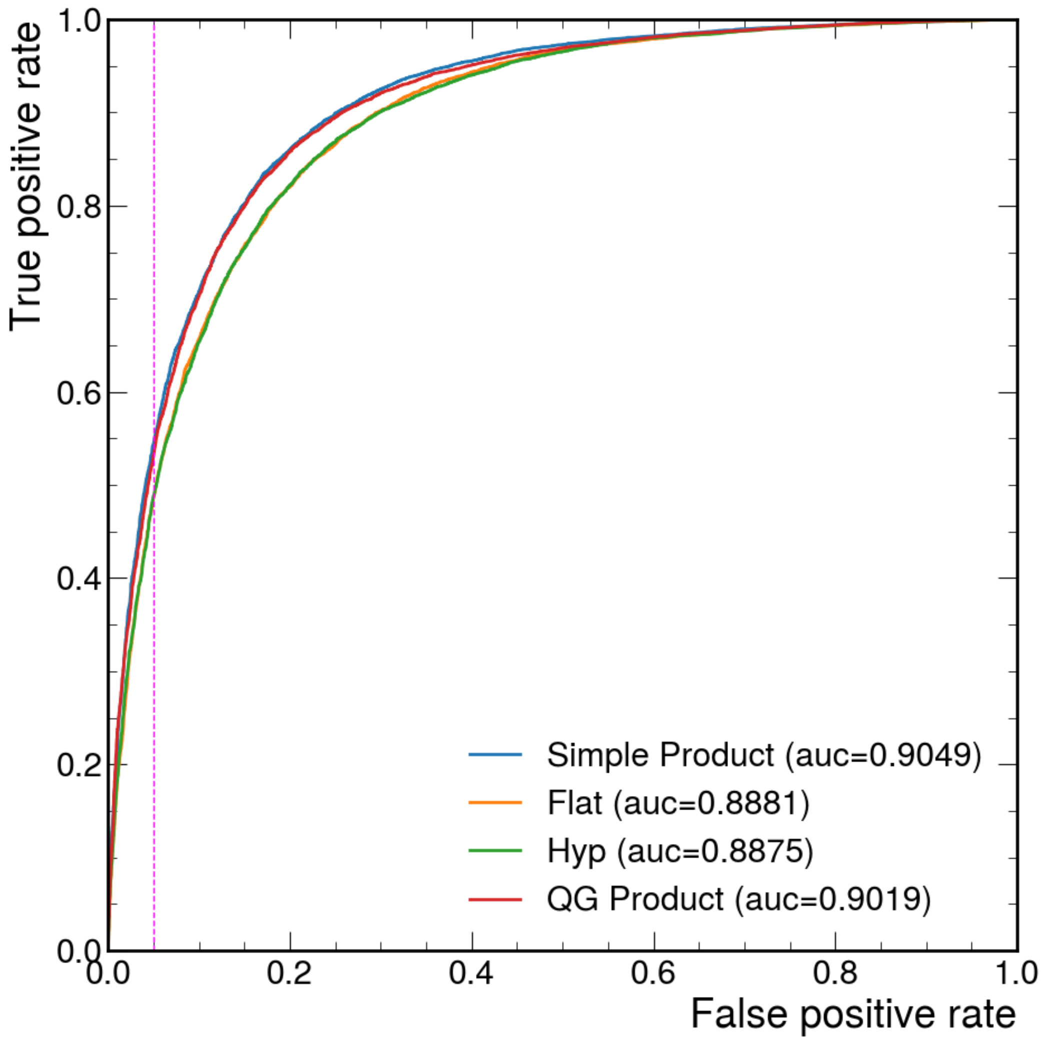
$$\boldsymbol{E}^{\frac{n}{3}} \times \boldsymbol{B}_{c_{quark}}^{\frac{n}{3}} \times \boldsymbol{B}_{c_{gluon}}^{\frac{n}{3}}$$

12D Jet Tagging: Quark vs. Gluon

<u>Signal Efficiency at 5%</u>

- **Simple Product: 0.4969**
- QG Product: 0.4906
- Hyp: 0.4849
- Flat: 0.4660

- **Product space models outperform** other geometries at QG tagging
- Combines **convenience of Euclidean** geometries with **expressive power of hyperbolic** geometry.



Conclusion

Conclusion

Results

- Hyperbolic models **equal/better** in small param & embedding dim
- **Product space models extend these improvements** to larger models & representations

Takeaway

- Non-Euclidean representations are an **efficient** avenue to improve jet physics analyses

Next Steps

- **Full comparison of non-Euclidean** models to popular jet tagging approaches

Additional Slides

Details: Machine Learning by Manifold Preserving Operations

Fully connected (FC) layers & Activation

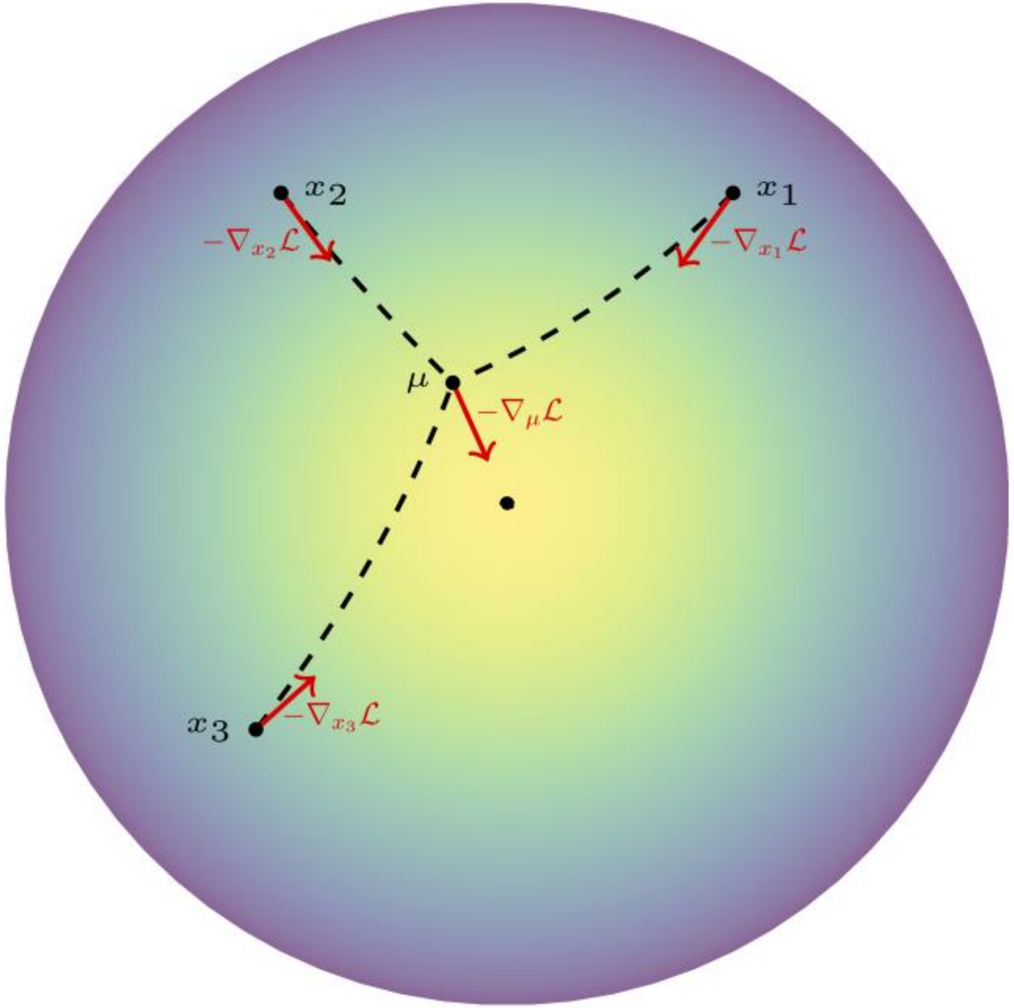
- Naive formulation: reformat FC propagation through gyrovector operations:

$$y = Wx \oplus_c b$$

- Where $W = [v_1, v_2, \dots, v_n]$ for $v_i \in B_c^m, x \in B_c^n, b \in B_c^m$
- We employ more robust formulations from the literature [Shimizu et al. 2006.08210v3]
- We utilize ReLU and LeakyReLU as they are norm reducing manifold preserving activation functions [Liu et al. 1910.12892]

Batch normalization

- The Fréchet mean generalizes centroids to metric spaces
- Work by [Lou et al. 2003.00335] developed differentiable Fréchet mean for Riemannian Batch Norm
- No analytic solution, but iterable



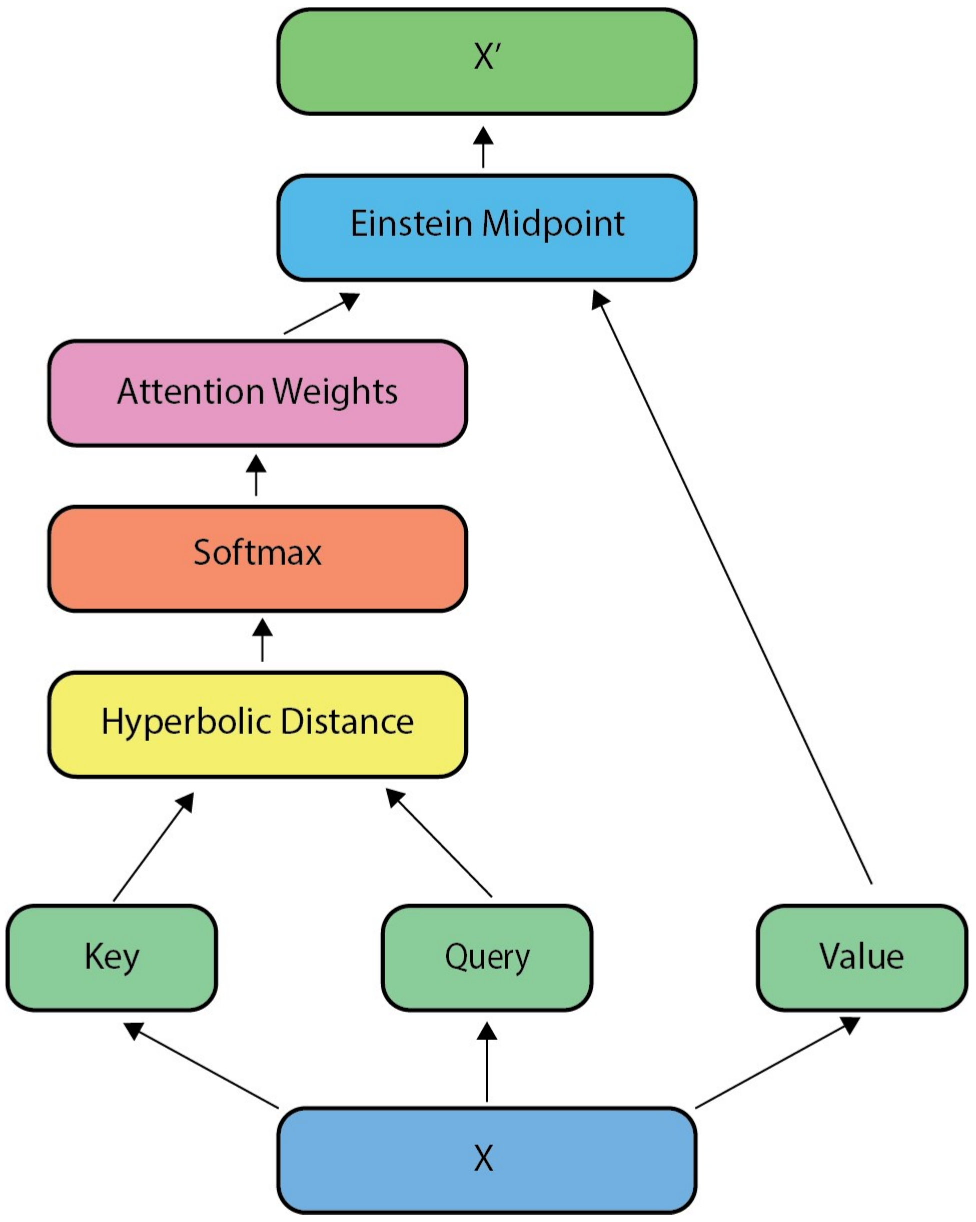
Lou et al. 2003.00335

Attention

- Attention weights calculated through hyperbolic distance metric
- Attention aggregation through weighted Einstein midpoint [Gulcehre et al. 1805.09786v1]
- $v_i \in B_c^m, w \in \mathbb{R}$, and γ is the conformal factor

$$m_{Ein}(v_1, \ldots, v_h; w_1, \ldots, w_h)$$

$$= \frac{1}{2} \oplus \sum_{k=1}^{h} p_k \gamma_k \boldsymbol{v}_k \Big/ \sum_{k=1}^{h} p_k (1 - \gamma_k)$$

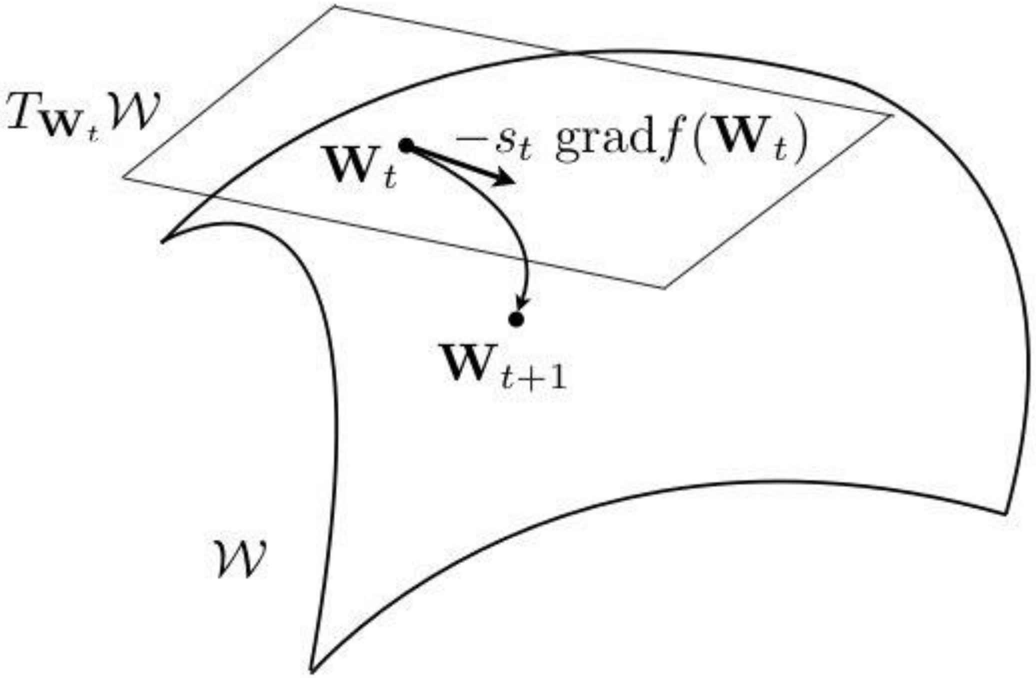


Optimization algorithms

- Adjustments for manifold optimization:

1. Adjust gradient magnitude for manifold
2. Update using parallel transport

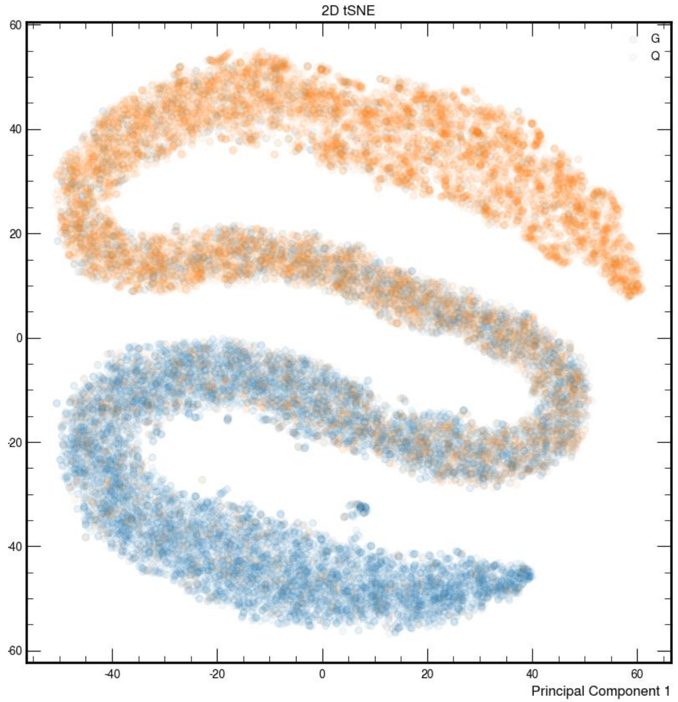
- Available algorithms: RiemannianAdam , RiemannianSGD RiemannianLion



Meyer et al. 1006.1288

Jet Contrastive Embedding: TSNE Quark vs Gluon

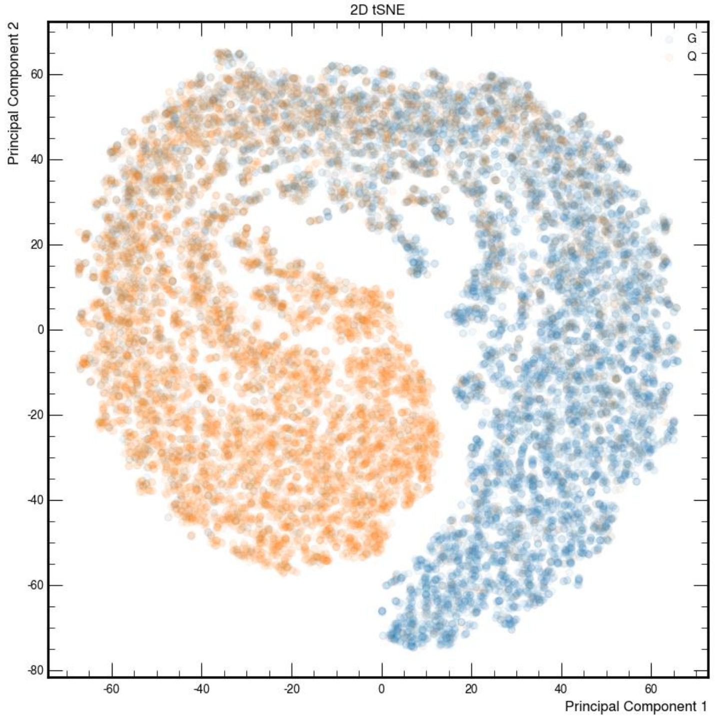
Flat



SIMCLR Loss:

5.3569

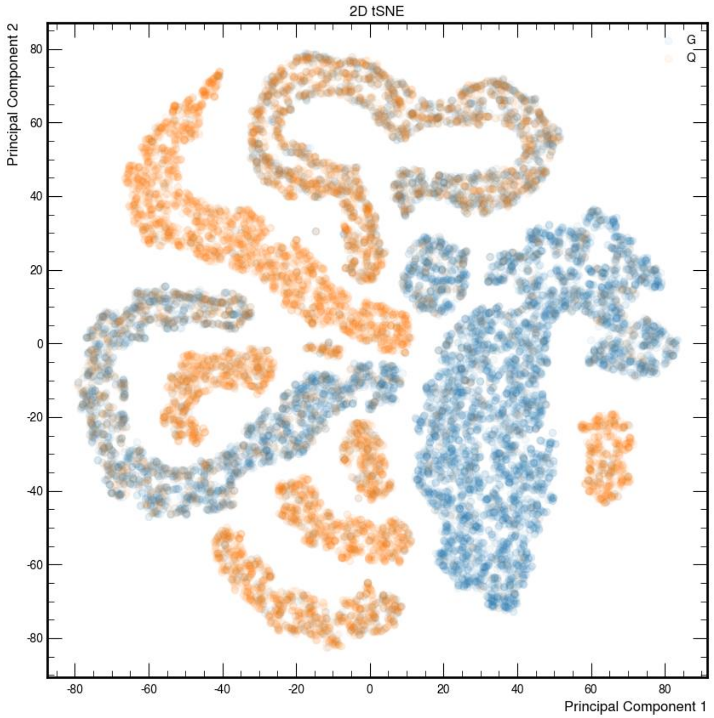
Hyperbolic



SIMCLR Loss:

5.3652

Mixed



SIMCLR Loss:

5.3554

Large Model Parameters

Architecture

- Transformers: 3
- Attention Heads: 8
- Head-size: 64
- Embedding-Dim: 160

Total Parameters

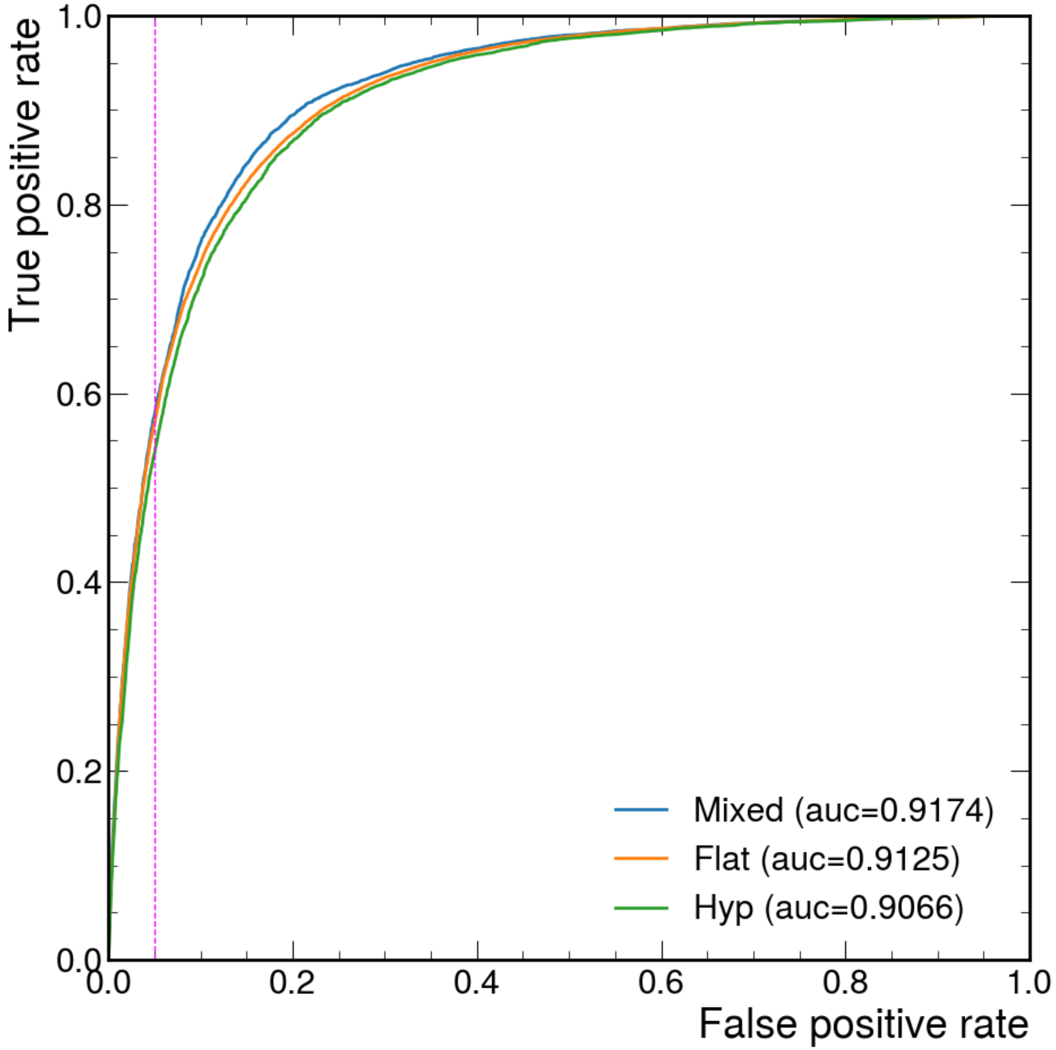
- Mixed: 341486 parameters
- Flat: 419342 parameters
- Hyp: 419330 parameters

Jet Tagging: Quark vs. Gluon

<u>Signal Efficiency at 5%</u>

- **Mixed: 0.5759**
- Flat: 0.5673
- Hyp: 0.5361

- Superior performance of mixed curvature model over traditional transformer
- Increased performance with **20% fewer** model parameters



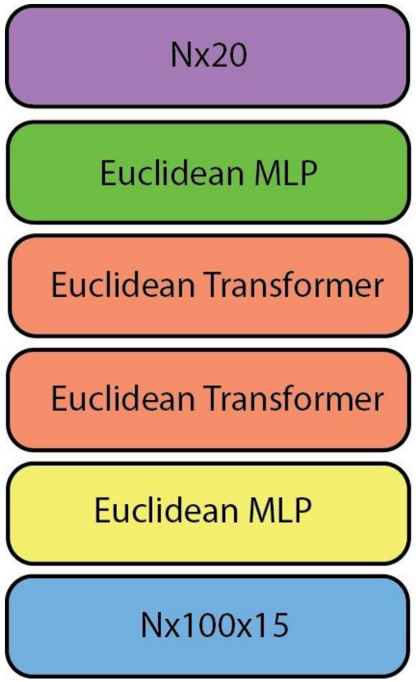
Small Models

(SUPER SMALL 1k params, embed 2d, 8d, 16d)

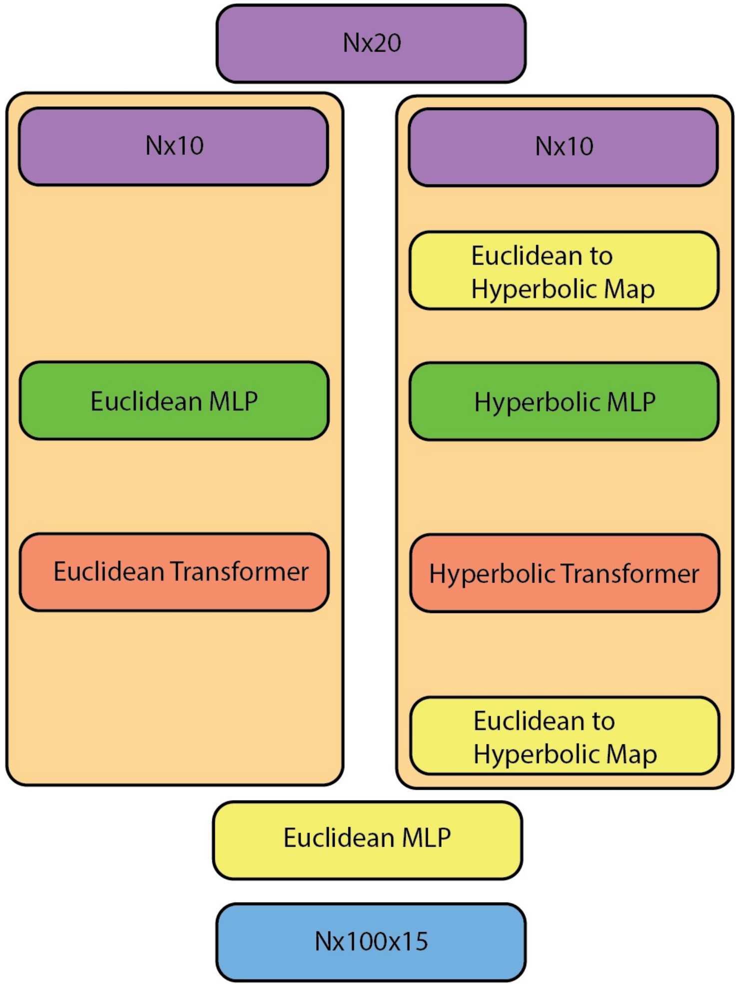
Model Architectures

- Limit model to 2 transformers
- What is optimal model architecture?
- Architectures:
  - 10% previous parameters
  - Euclidean: 31k params
  - Product Space: 38k params

Euclidean Model



Product Space Model

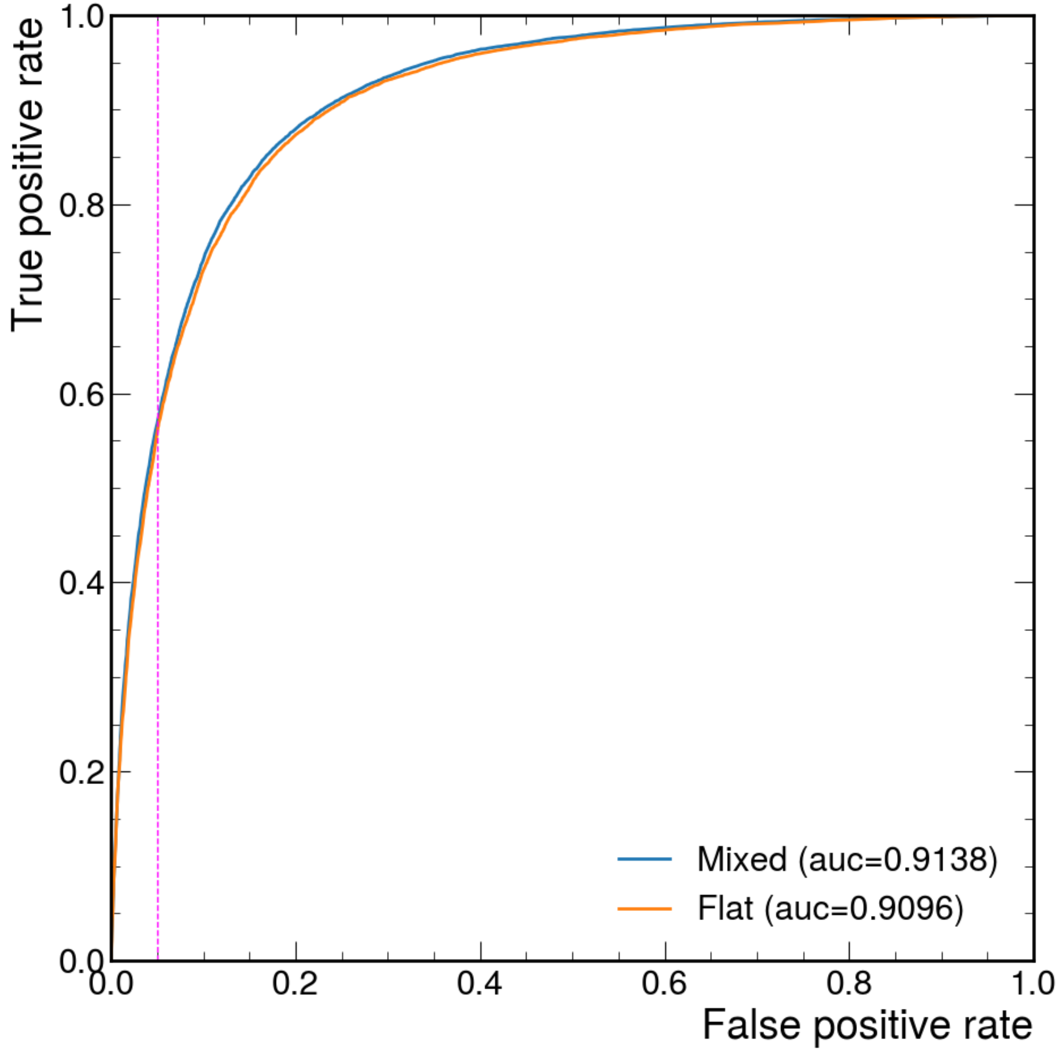


Small Jet Tagging: Quark vs. Gluon

ROC at 5%

- **Mixed: 0.5682**
- Flat: 0.5588

- Superior on Quark-Gluon tagging
- Non-Euclidean representations provide superior jet representation



Expanded Product Space Models

Expanded Mixed Models Jet Tagging: Quark vs. Gluon

<u>ROC at 5%</u>

- **Mixed: 0.5750**
- Flat: 0.5475

- Product space model superior to sequential Euclidean
- Non-Euclidean representations more easily extract jet information