A Unified Geometric Framework for Understanding Collider Event Manifolds

Based on Work in Collaboration with:

Tianji Cai[1,2]

arXiv:2609.xxxxx

[1] School of Physical Science and Engineering, Tongji University, Shanghai 200092, China

[2] State Key Laboratory of Autonomous Intelligent Unmanned Systems, MOE Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China

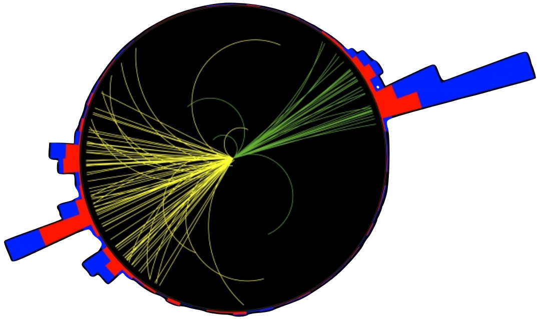
Hancheng Li

Rutgers University

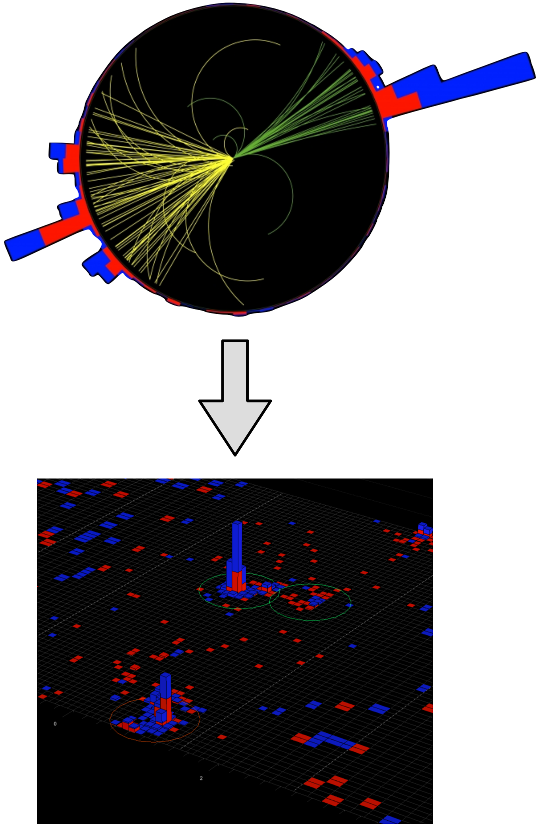
hancheng.li@rutgers.edu

Introduction to Collider Metrics

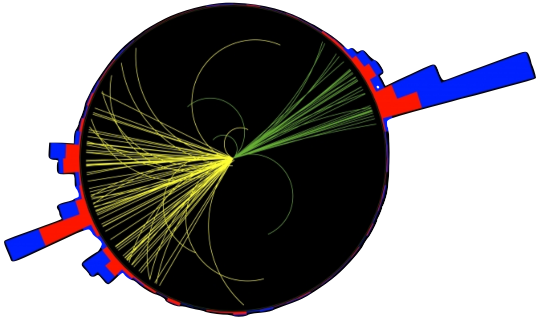
Introduction to Collider Metrics

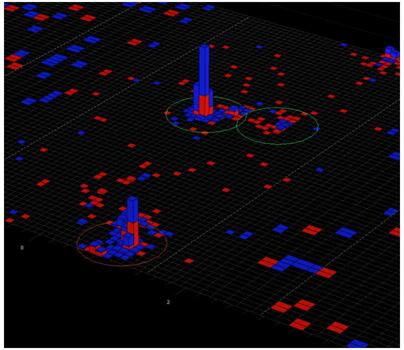


Introduction to Collider Metrics



Introduction to Collider Metrics





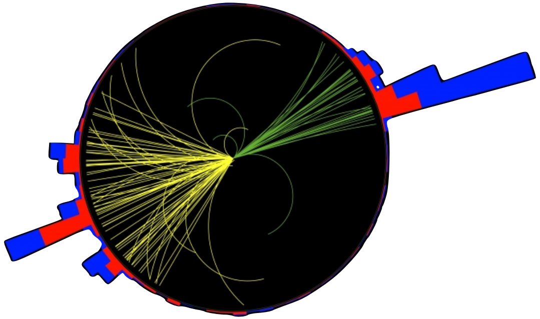
- Wasserstein Distances:

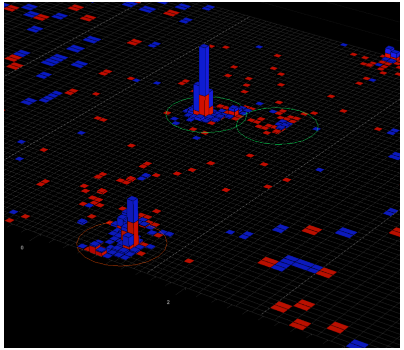
$$W_2^2(\mathscr{E}, \mathscr{E}') = \min_{\gamma \in \Pi(\mathscr{E}, \mathscr{E}')} \sum_{i,j} \gamma_{ij} \left((\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2 \right)$$

- A rigorous definition of distance between collider events.
- First introduced by [1] to particle physics

[1]: Komiske et al., *Phys. Rev. Lett.* **123**, 041801 (2019). arXiv:1902.02346

Introduction to Collider Metrics

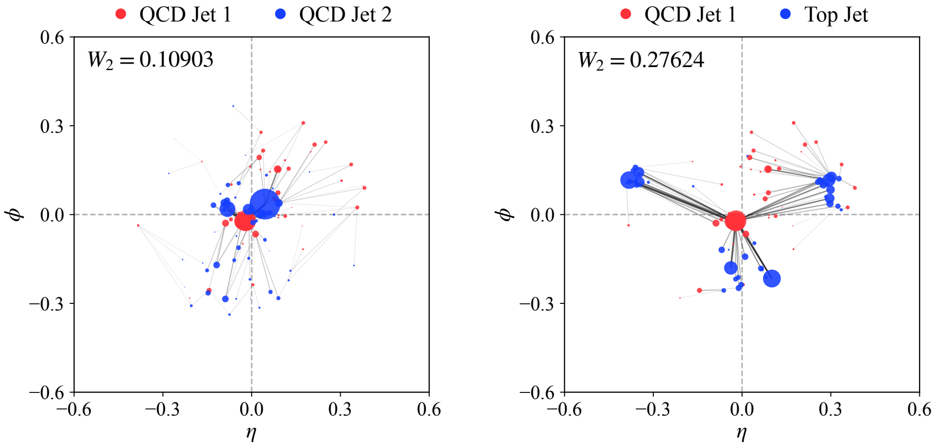




- Wasserstein Distances:

$$W_2^2(\mathcal{E},\mathcal{E}') = \min_{\gamma\in\Pi(\mathcal{E},\mathcal{E}')} \sum_{i,j} \gamma_{ij}\left((\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2\right)$$

- A rigorous definition of distance between collider events.
- First introduced by [1] to particle physics



[1]: Komiske et al., *Phys. Rev. Lett.* **123**, 041801 (2019). arXiv:1902.02346

Zoo of Collider Metrics

- **Spectral Wasserstein Distance**[1]
- Compress the energy distribution into a 1D *Spectral Function*
- Spectral 2-Wasserstein Distance (SW_2) = W_2 of the spectral functions
- **Phase Space Distance**[2]
- Represent an event as its dLIPS_N
- The Phase Space Distance (PS) = the distance between two dLIPS_N

- **And many more…**

[1]: A. J. Larkoski & J. Thaler, "A spectral metric for collider geometry," **JHEP 08 (2023) 107**, arXiv:2305.03751.
[2]: T. Cai, J. Cheng, N. Craig, G. Koszegi & A. J. Larkoski, "The phase space distance between collider events," **JHEP 09 (2024) 054**, arXiv:2405.16698. [Springer Nature Link]

What Should We Do With The Zoo of Metrics?

What Should We Do With The Zoo of Metrics?

Theoretically,

- Are the geometric structures of the manifolds* induced by these metrics the same?

    1. Geometry of a single jet type manifold

    2. Geometry of the decision boundary manifold

3. How can you unify the metrics?

*: Distributions of events equipped with metrics

What Should We Do With The Zoo of Metrics?

Theoretically,

- Are the geometric structures of the manifolds* induced by these metrics the same?

    1. Geometry of a single jet type manifold

    2. Geometry of the decision boundary manifold

3. How can you unify the metrics?

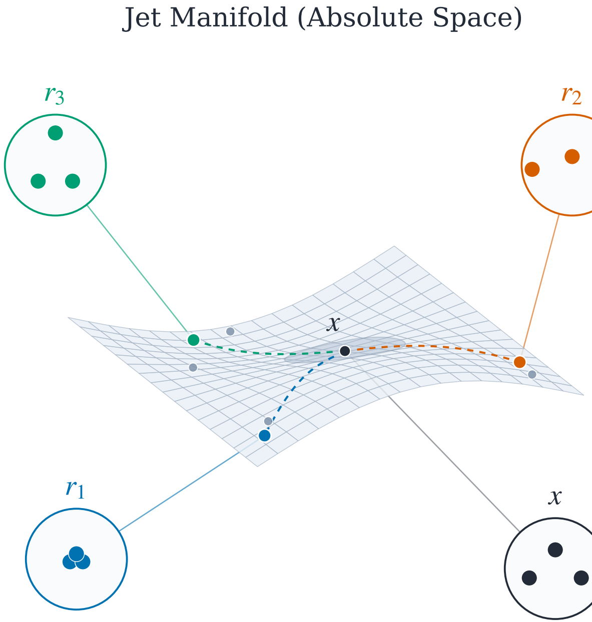
*: Distributions of events equipped with metrics

We need to study the geometry of these manifolds in a systematic way and on equal-footing!

Multi-Reference Relative Representation (M3R)

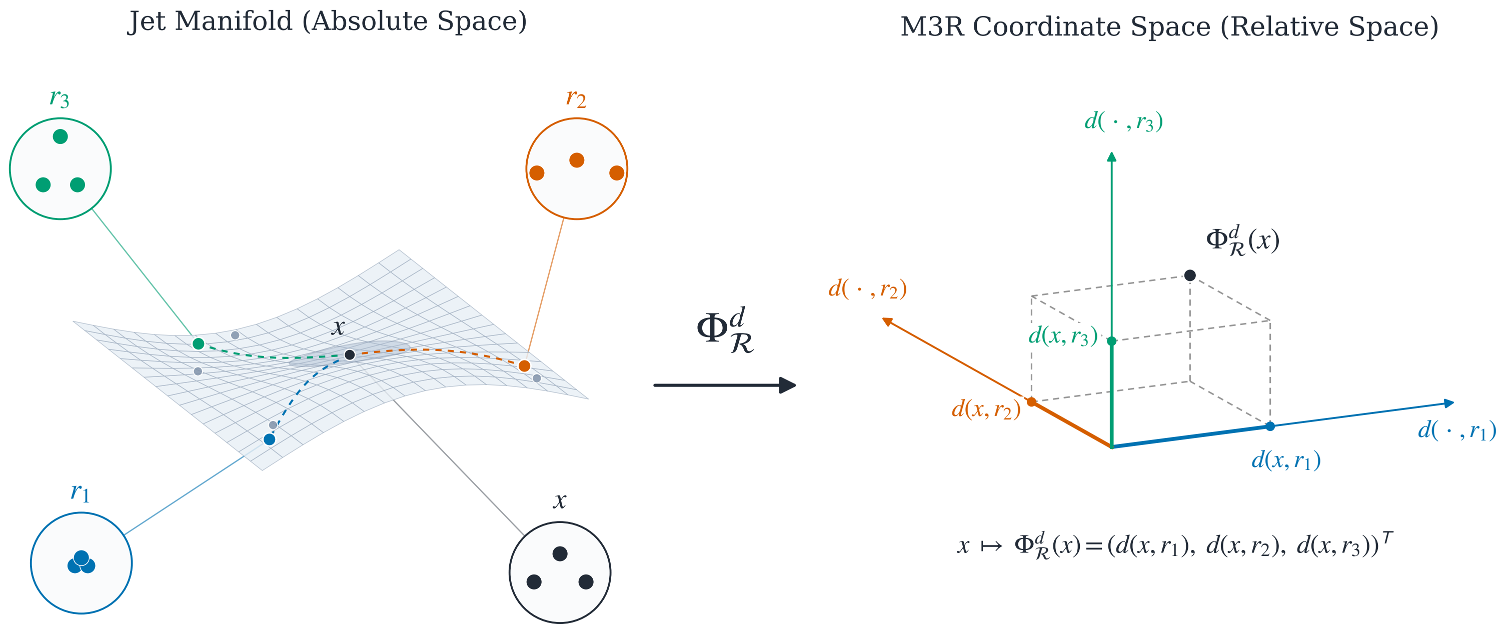
Moschella et al., *ICLR* (2023). arXiv:2209.15430

Multi-Reference Relative Representation (M3R)



Moschella et al., *ICLR* (2023). [arXiv:2209.15430](https://arxiv.org/abs/2209.15430)

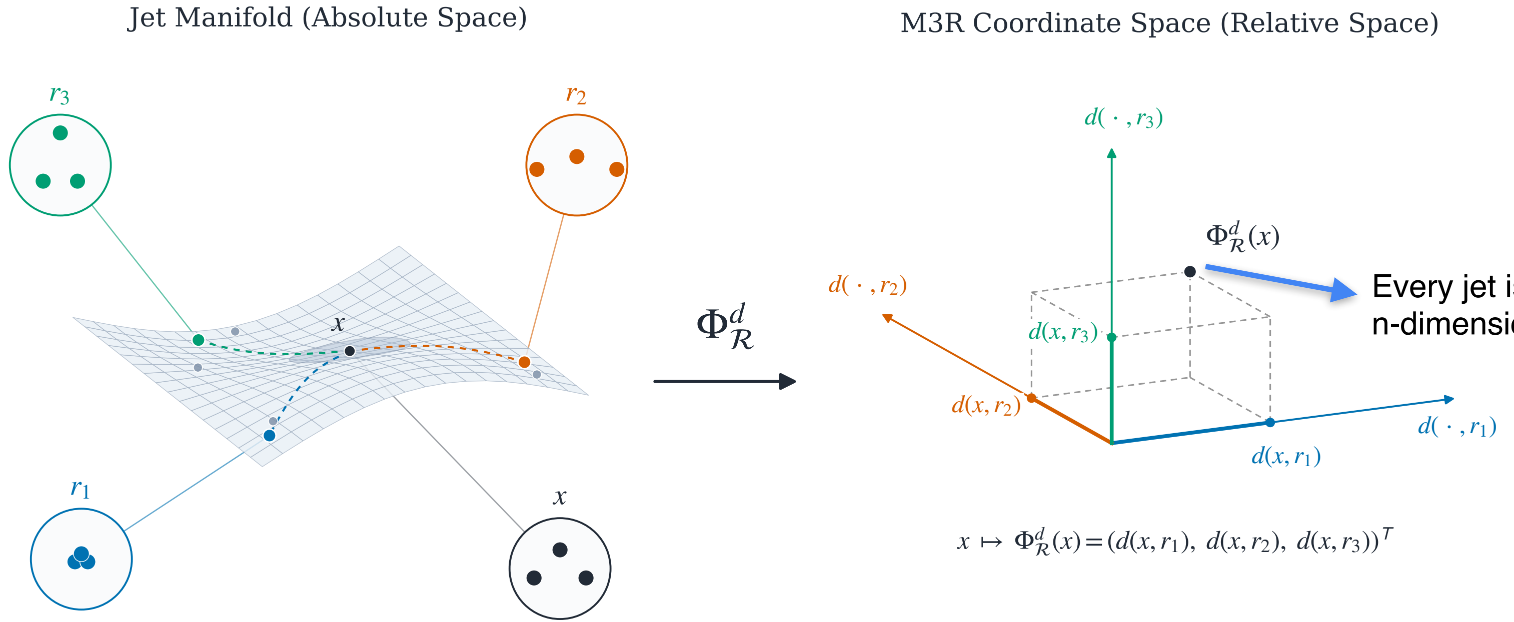
Multi-Reference Relative Representation (M3R)



$$x \mapsto \Phi^d_{\mathcal{R}}(x) = (d(x, r_1),\ d(x, r_2),\ d(x, r_3))^T$$

Moschella et al., *ICLR* (2023). [arXiv:2209.15430]

Multi-Reference Relative Representation (M3R)



Every jet is a point in n-dimensional space

Distance between two events in absolute space → Euclidean distance in relative space

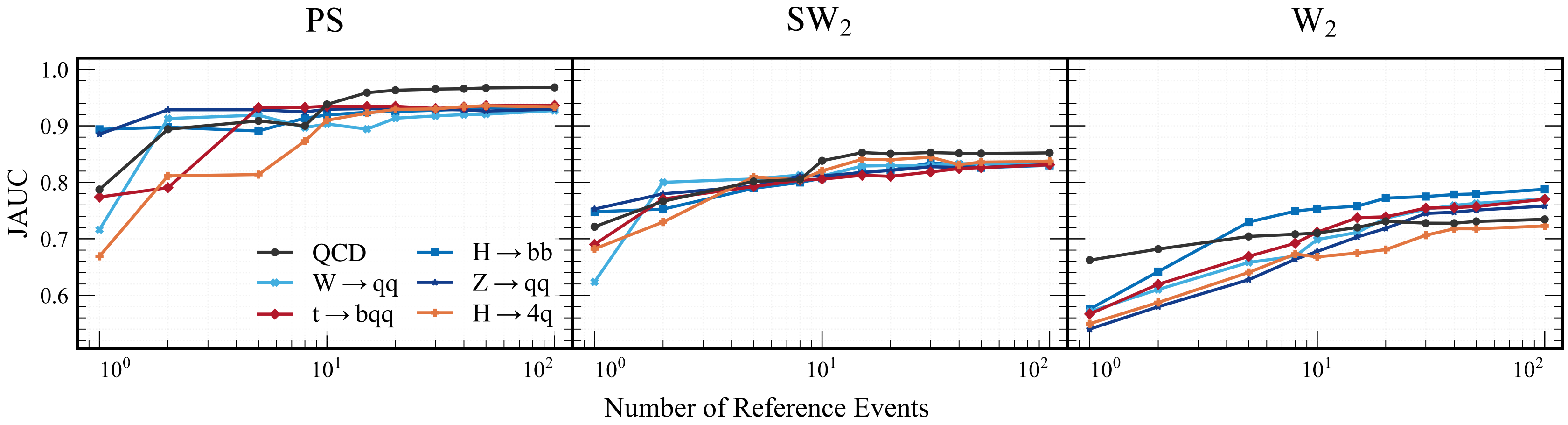
Moschella et al., *ICLR* (2023). [arXiv:2209.15430]

Geometry of a Single Jet Type Manifold

Geometry of The Decision Boundary Manifold

Unifying The Metrics

Single Jet Type Manifolds

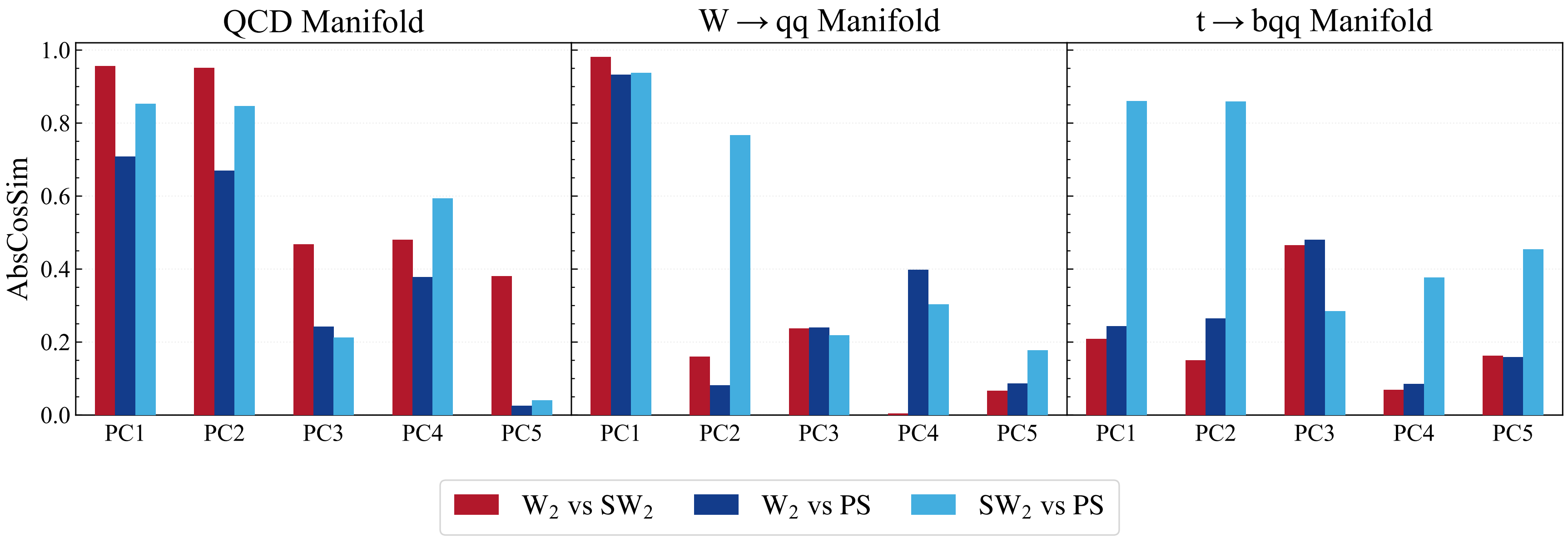


- PS is clearly the easiest for M3R to reconstruct, followed by SW$_2$, then W$_2$

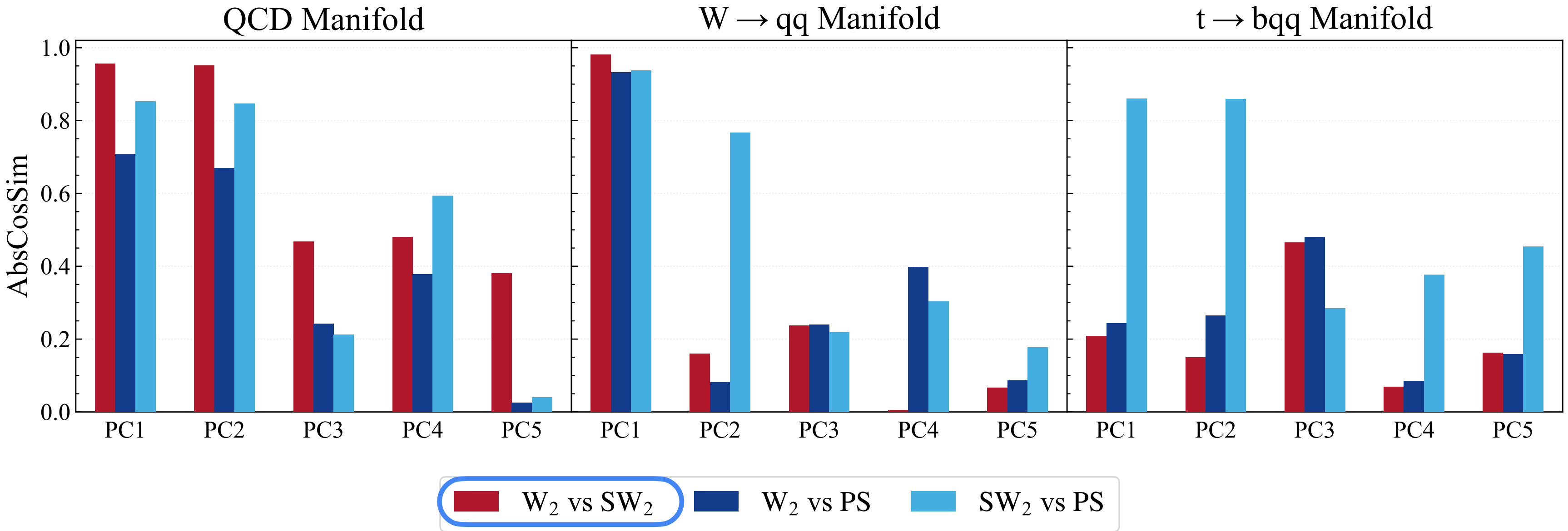
- Is this consistent with the physics expectation?

The `JetClass` Dataset is used for this study

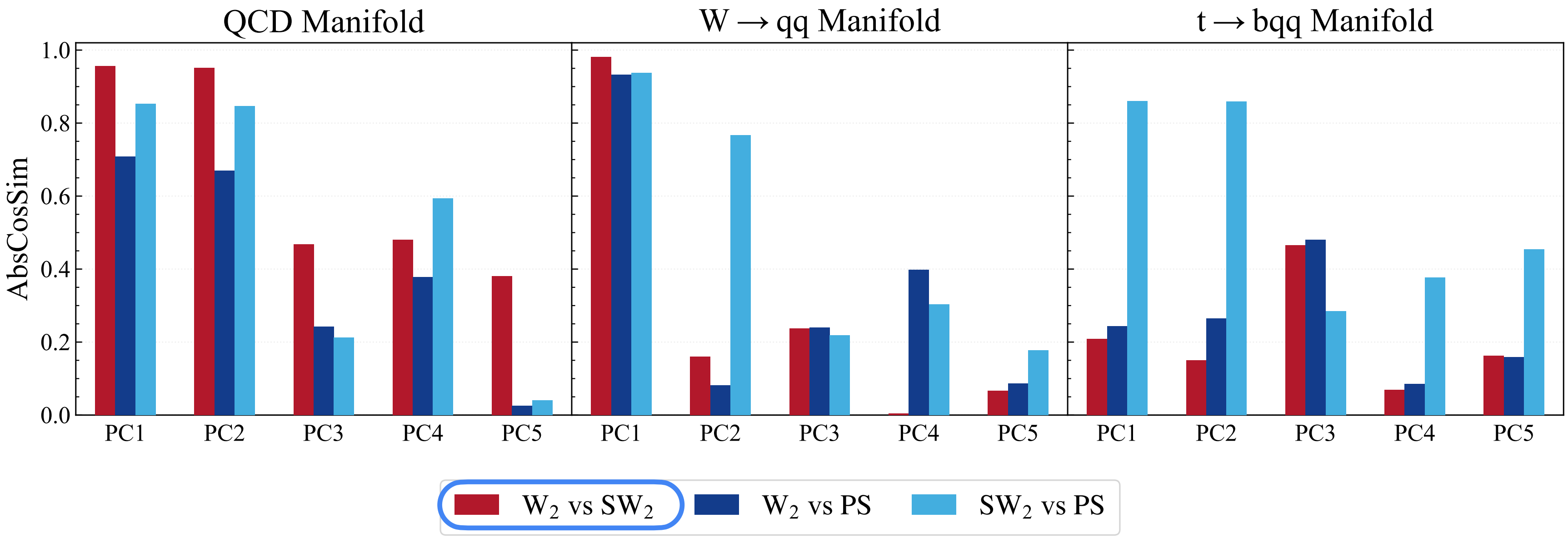
Manifold Alignment



Manifold Alignment



Manifold Alignment



Numerically confirming the theoretical calculations and in [1] and [2], extending them to the non-perturbative* regime

[1]: A. J. Larkoski & J. Thaler, "A spectral metric for collider geometry," **JHEP 08 (2023) 107**, arXiv:2305.03751.
[2]: R. Gambhir, A. J. Larkoski & J. Thaler, "SPECTER: Efficient evaluation of the spectral EMD," **JHEP 12 (2024) 219**, arXiv:2410.05379.

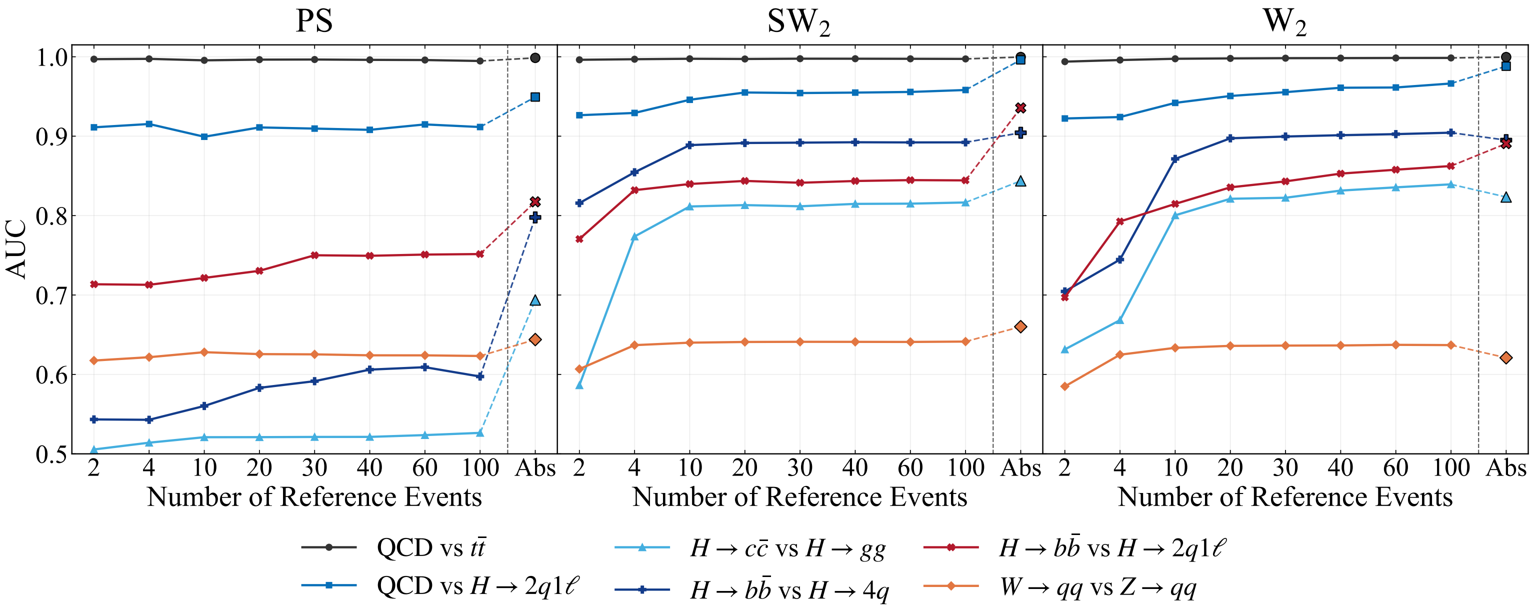
*: Simulation

Geometry of a Single Jet Type Manifold

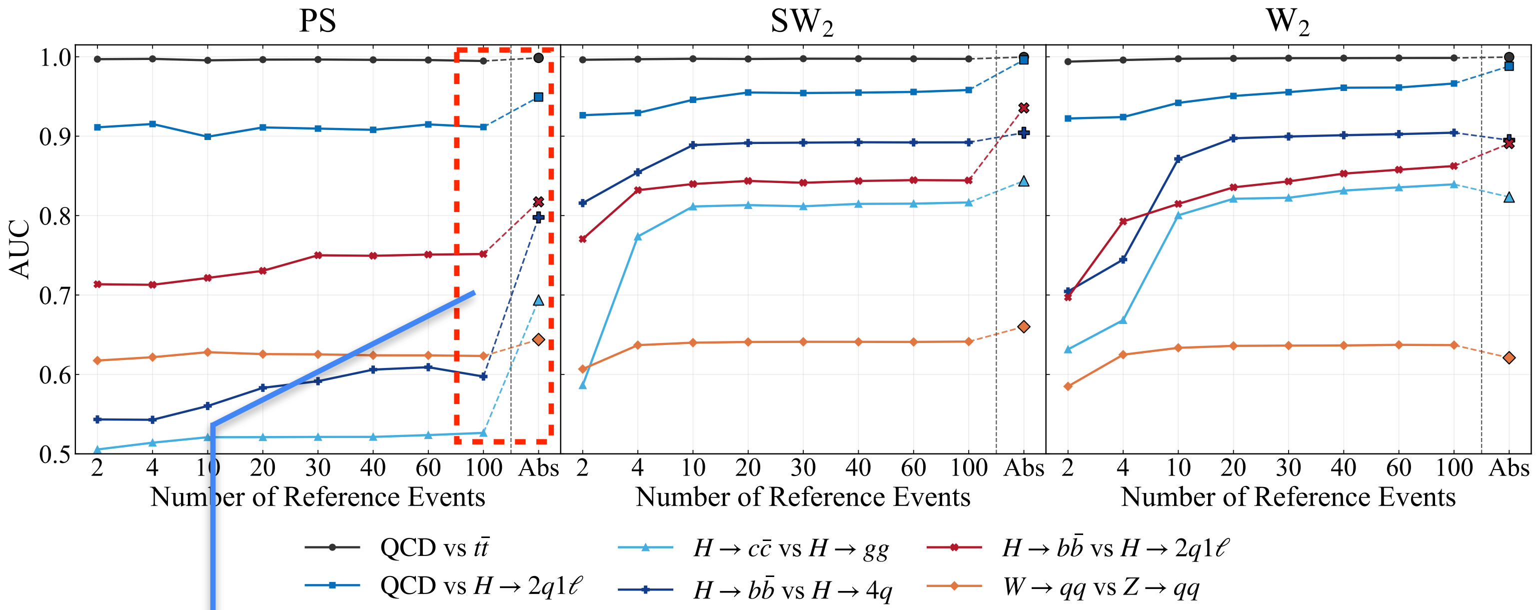
Geometry of The Decision Boundary Manifold

Unifying The Metrics

Tagging Performances



Tagging Performances



What's happening for the PS distance?

Why?

- If PS is easier to reconstruct; shouldn't we expect tagging performance to be similar to absolute space?

- If you think about it more carefully…
  - PS is the easiest to reconstruct for a single jet type ↔ PS doesn't capture the detailed physics well
  - PS doesn't capture the detailed physics well → Decision boundary is more complex → Harder for M3R to reconstruct → Worst tagging performance compared to the absolute space

Geometry of a Single Jet Type Manifold

Geometry of The Decision Boundary Manifold

Unifying The Metrics

Unified Representation

Unified Representation

$$\Phi^{d}_{\mathscr{R}}(x) = \big(d(x, r_1), \ldots, d(x, r_m)\big)$$

Unified Representation

$$\Phi^{d}_{\mathscr{R}}(x) = \big(d(x, r_1), \ldots, d(x, r_m)\big)$$

$$\Downarrow$$

$$\Phi^{\mathrm{Uni}}_{\mathscr{R}}(x) = (d_1(x, r_1), \ldots, d_1(x, r_m), \ldots, d_n(x, r_1), \ldots, d_n(x, r_m))$$

Unified Representation

$$\Phi^{d}_{\mathscr{R}}(x) = \big(d(x, r_1), \ldots, d(x, r_m)\big)$$

$$\Phi^{\mathrm{Uni}}_{\mathscr{R}}(x) = (d_1(x, r_1), \ldots, d_1(x, r_m), \ldots, d_n(x, r_1), \ldots, d_n(x, r_m))$$

Same Reference Set

Arbitrary Metrics/Similarities

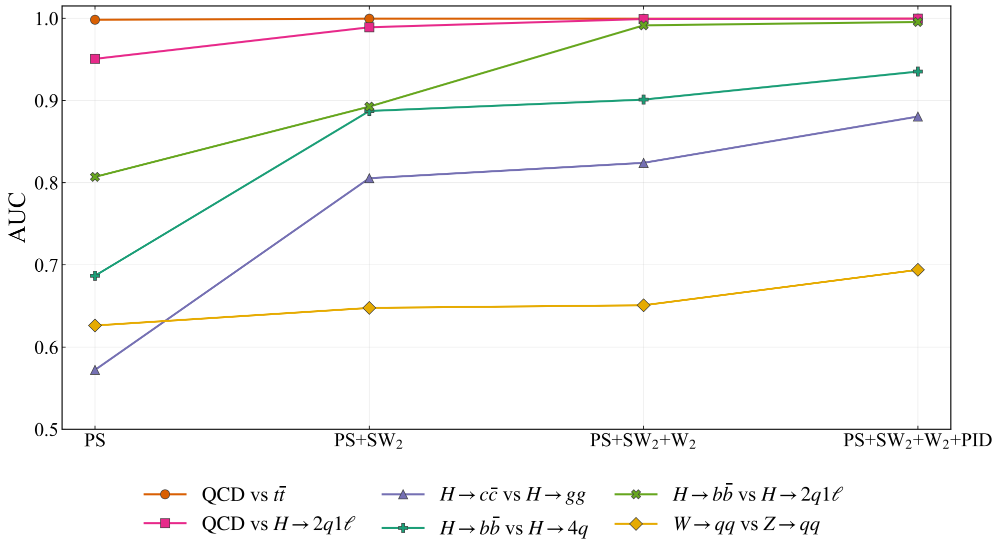
Unified Representation

$$\Phi^{d}_{\mathcal{R}}(x) = \big(d(x, r_1), \ldots, d(x, r_m)\big)$$

$$\Phi^{\mathrm{Uni}}_{\mathcal{R}}(x) = (d_1(x, r_1), \ldots, d_1(x, r_m), \ldots, d_n(x, r_1), \ldots, d_n(x, r_m))$$

Same Reference Set

Arbitrary Metrics/Similarities



Takeaways and Outlook

- M3R provides insights to the different of physics captured by different metrics
  - Easier reconstruction but poor tagging performance of the PS metric
  - Numerical confirmation of theoretical calculations comparing SW_2 and W_2
- Gives a straightforward way to unify the metrics
- (Greatly reduces the computational complexity of pairwise distance matrix computation)

More excitingly:

- The M3R feature vector, defined *per-event,* opens the door to the realm of feature engineering and **deep learning**

Backup Slides

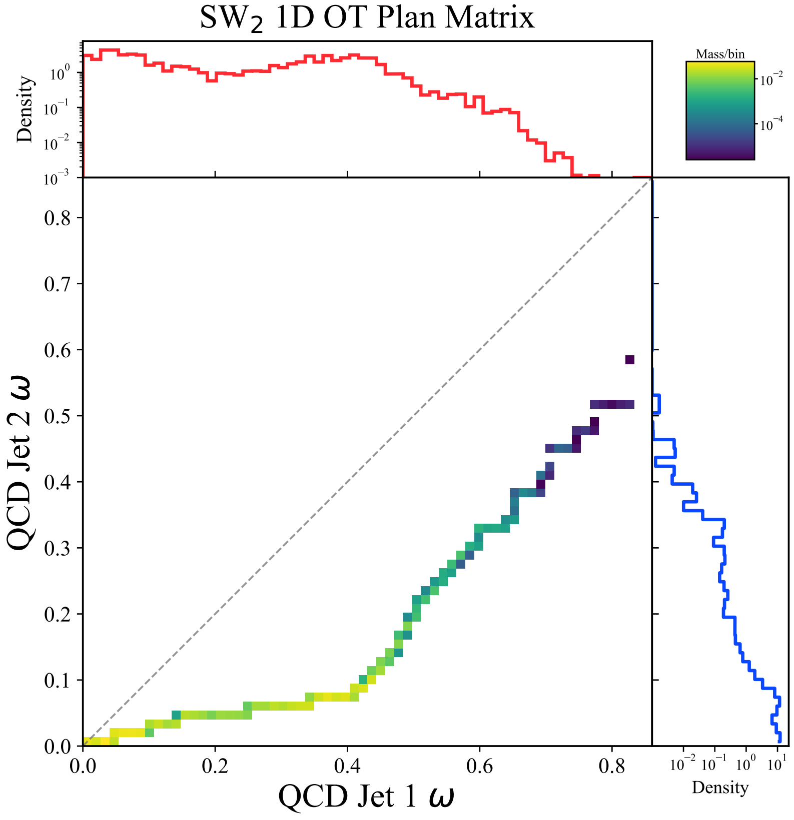
Properties of M3R

- $d(x, y)$ is a metric $\rightarrow \|\Phi^d_{\mathscr{R}}(x) - \Phi^d_{\mathscr{R}}(y)\|_2$ is a metric*
- $\dfrac{1}{\sqrt{n}}\|\Phi^d_{\mathscr{R}}(x) - \Phi^d_{\mathscr{R}}(y)\|_2 \leq d(x, y)$
- $d(x, y)$ is IRC-safe $\rightarrow \|\Phi^d_{\mathscr{R}}(x) - \Phi^d_{\mathscr{R}}(y)\|_2$ is IRC-safe

M3R is a faithful reconstruction of the absolute space

*: Almost always

Spectral Wasserstein Distance



Faithful M3R Reconstruction

$$\Phi(x) = \{d(x,r)\}_{r\in\mathcal{X}}$$

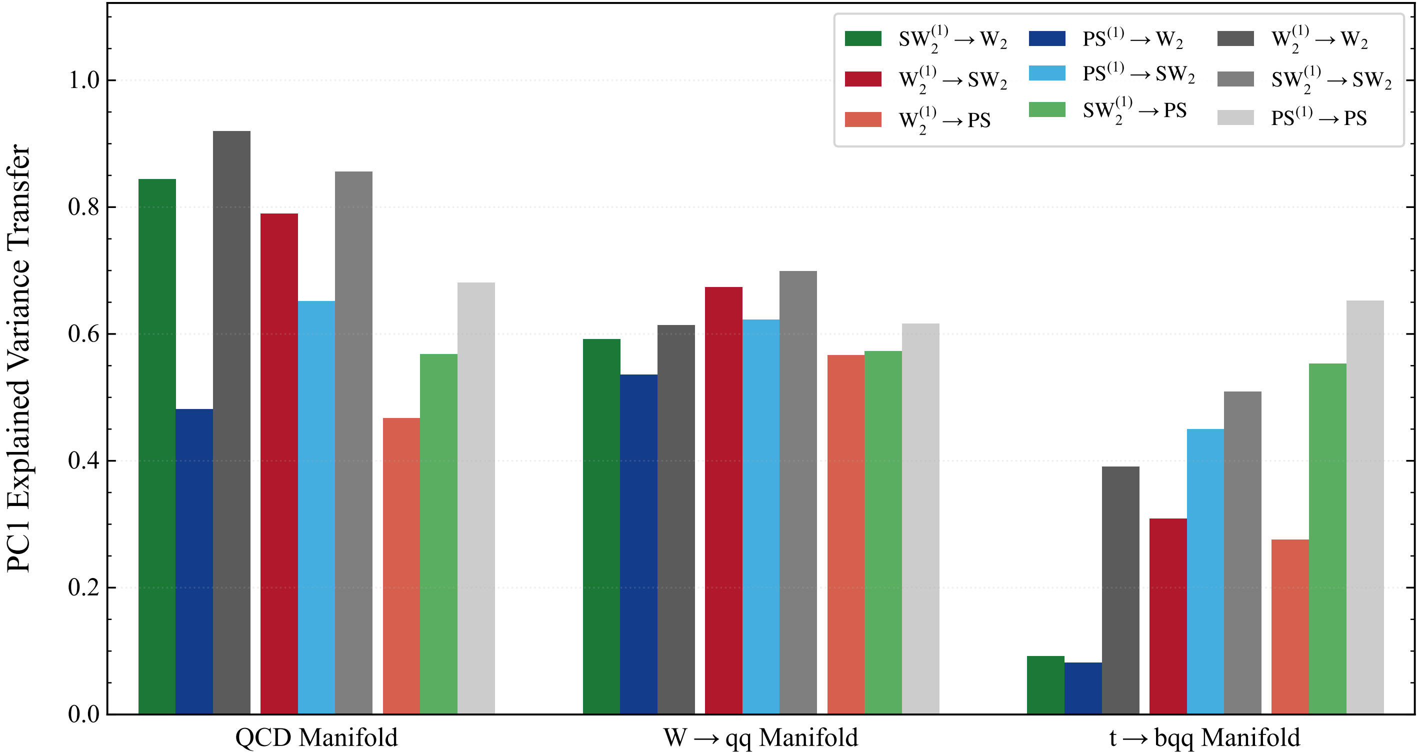
$$\|\Phi(x) - \Phi(y)\|_\infty = \sup_{r\in\mathcal{X}} |\, d(x,r) - d(y,r)\,|$$

Reverse triangle inequality: $\quad |\, d(x,r) - d(y,r)\,| \leq d(x,y)$

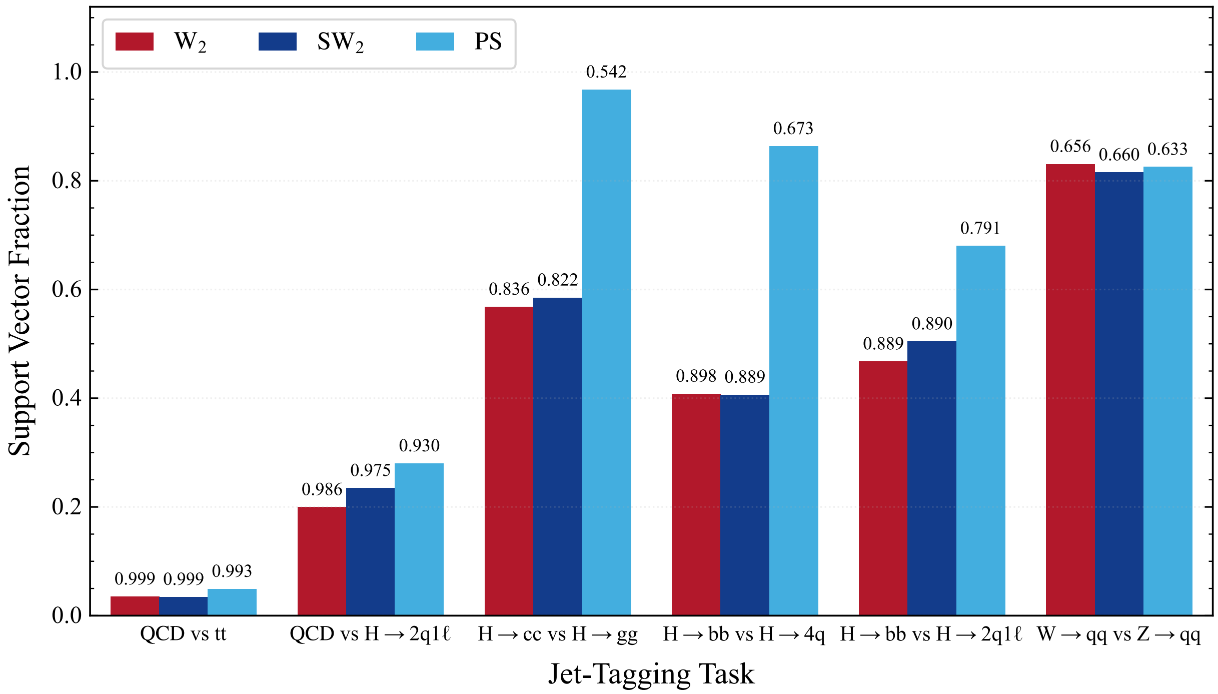
Choose $r = x$: $\quad |\, d(x,x) - d(y,x)\,| = d(x,y)\,.$

$$\|\Phi(x) - \Phi(y)\|_\infty = d(x,y)$$

Manifold Alignment



Geometry of Inter-Manifold Decision Boundaries



Jaccard AUC (JAUC)

$$J_k(x) = \frac{\left| \mathcal{N}_k^{\text{abs}}(x) \cap \mathcal{N}_k^{\text{rel}}(x) \right|}{\left| \mathcal{N}_k^{\text{abs}}(x) \cup \mathcal{N}_k^{\text{rel}}(x) \right|}$$

$$\text{JAUC} = \int_{q_{\min}}^{q_{\max}} \left\langle J_{qN}(x) \right\rangle_x dq, \qquad q = \frac{k}{N}.$$

Computation Efficiency of M3R

Usually pairwise distance calculation:

$\mathcal{O}(n^2)$ expensive metric distance calculations, where n is the number of events

M3R pairwise distance calculation:

$\mathcal{O}(mn)$ expensive distance calculations, where m is the # of reference events, $m \ll n$ and $\mathcal{O}(n^2)$ extremely cheap euclidean distance calculations

$\rightarrow \mathcal{O}(mn) \approx \mathcal{O}(n)$ expensive distance calculations

PID Similarities

PID Similarities

PID Similarities:

$$\boldsymbol{\delta}_{\text{pid}}(J, J') = \left| \mathbf{f}_{\text{pid}}(J) - \mathbf{f}_{\text{pid}}(J') \right|$$

$$\mathbf{f}_{\text{pid}}(J) = \left(N_e, N_\mu, N_\gamma, N_{\text{ch}}, N_{\text{neu}}, Q_J \right)$$