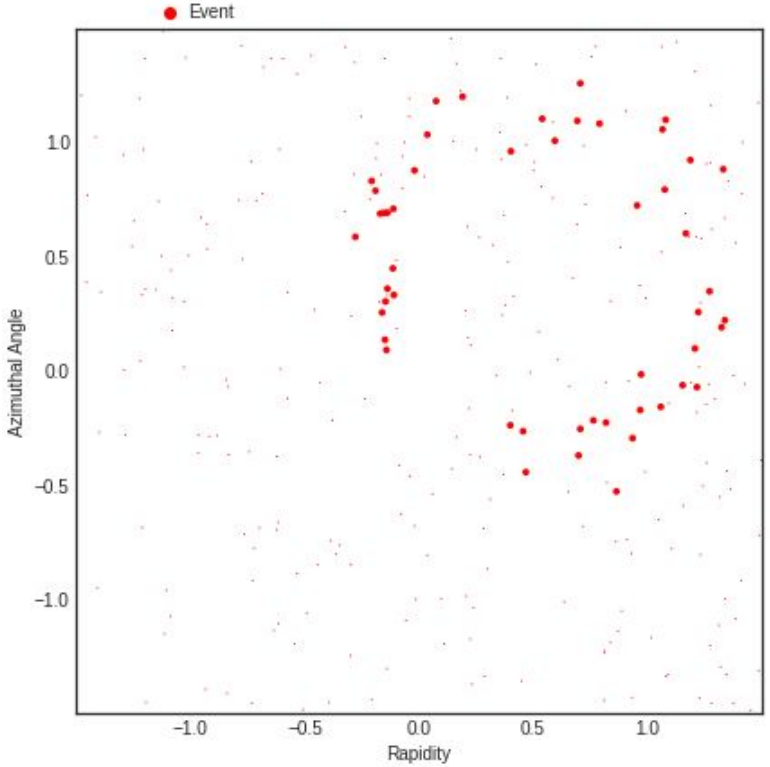
Can you hear the shape of a jet?

Rikab Gambhir
With Akshunna S. Dogra (AI), Demba Ba (AI), & Jesse Thaler (AI MIT)

Fundamental Question: What shape is this?

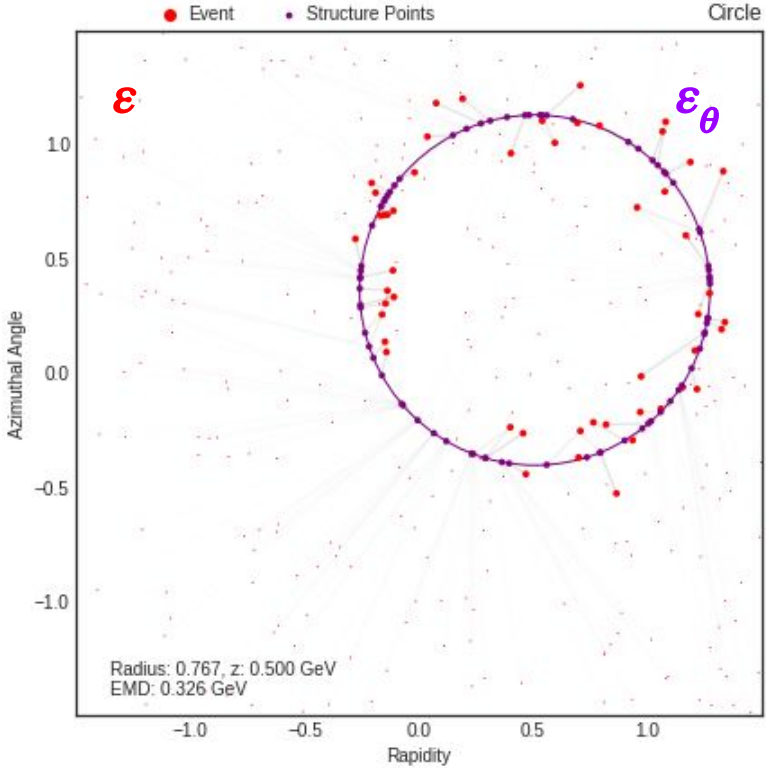


Pictured: (Fake) event that you might have measured at the LHC

Red dots are detector hits on a patch of the LHC cylinder, weighted by energy

Goal: Construct an observable $\mathcal{O}$ that generically answers this question!

Fundamental Question: What shape is this?



Using the **SHAPER** framework …

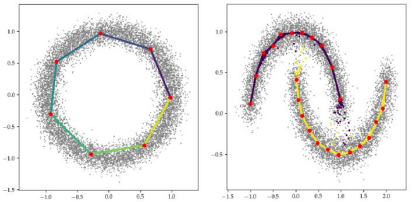
$$\mathcal{O}_{\mathcal{M}}(\mathcal{E}) = \min_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

$$\theta = \operatorname*{argmin}_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

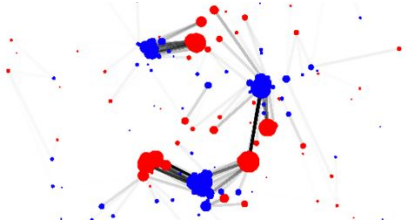
Circle with radius 0.767, center (0.50, 0.36) and a “circle-ness” value of 0.32.

Yes, you *CAN* hear the shape of a jet!

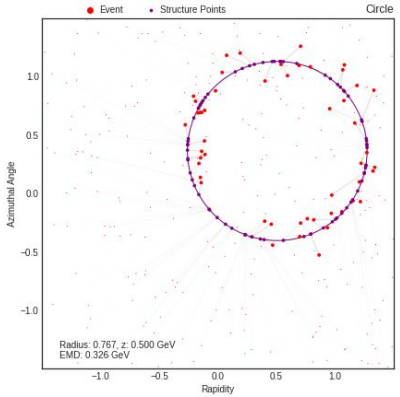
NSF AI Institute for Artificial Intelligence & Fundamental Interactions



Piecewise-Linear Manifold Approximation with K-Deep Simplices (KDS, 2012.02134)



Well-Defined Metric on Particle Collisions using Energy Mover's Distance (EMD, 2004.04159)



SHAPER: Learning the Shape of Collider Events

$$\mathcal{O}_{\mathcal{M}}(\mathcal{E}) = \min_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

$$\theta = \operatorname*{argmin}_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

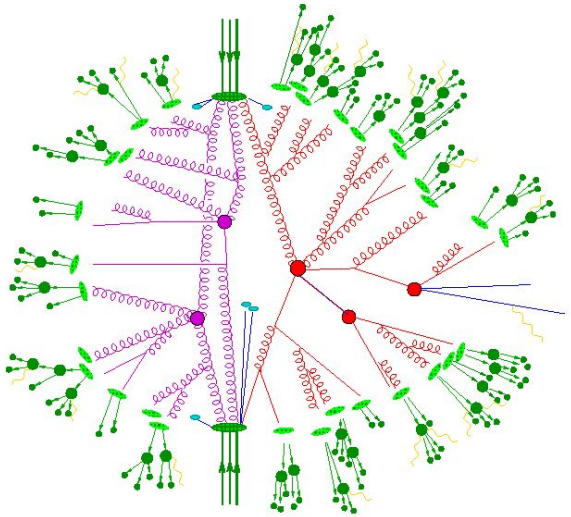
Framework for defining and calculating useful observables for collider physics!

Rikab Gambhir

Energy Flows

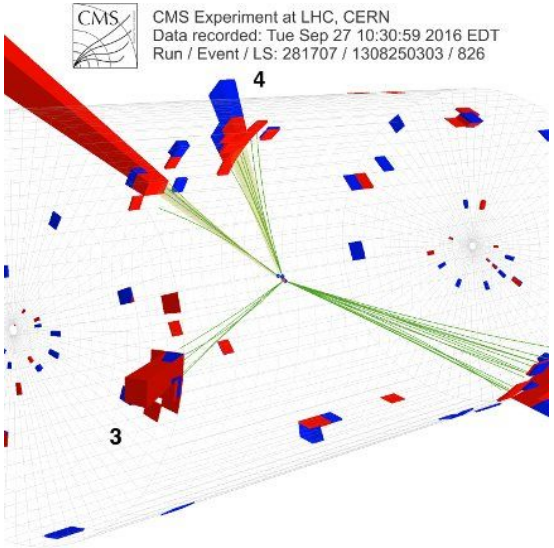
Robust Observables

[Enrico Bothmann et. al., 1905.09127;
CMS, 1810.10069]



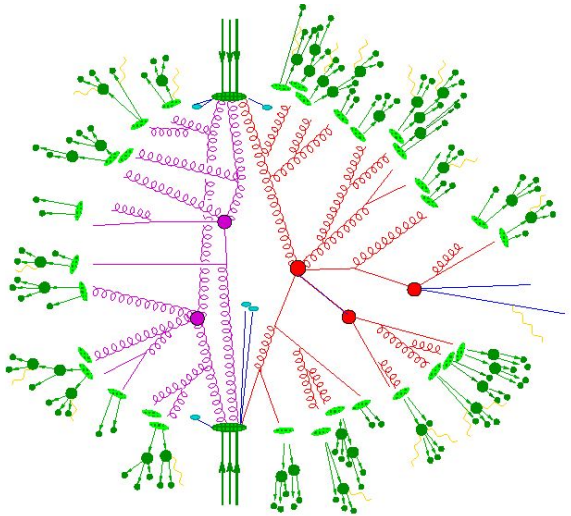
Theory

Experiment



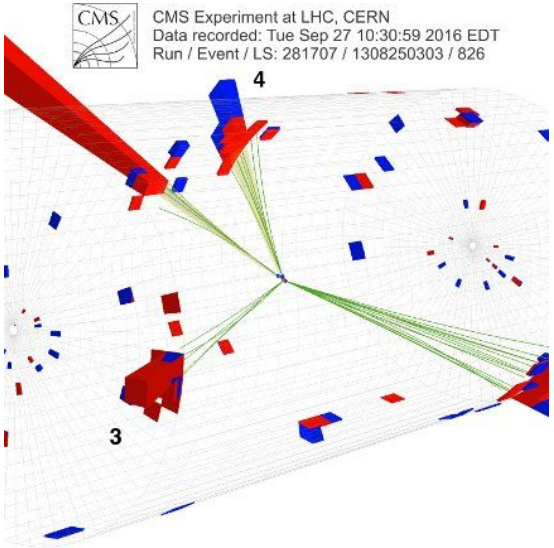
Robust Observables

[Enrico Bothmann et. al., 1905.09127;
CMS, 1810.10069]



Theory

Experiment

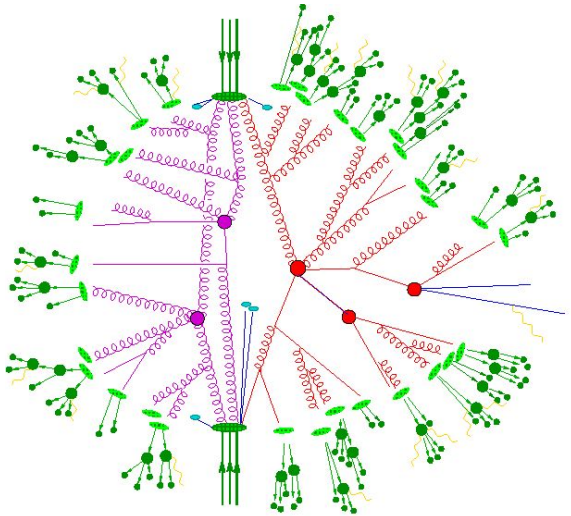


Perturbativity?
Hadronization?
Parton Shower Model?

Finite Calorimeter Resolution Effects?
Different Resolution between Detectors?

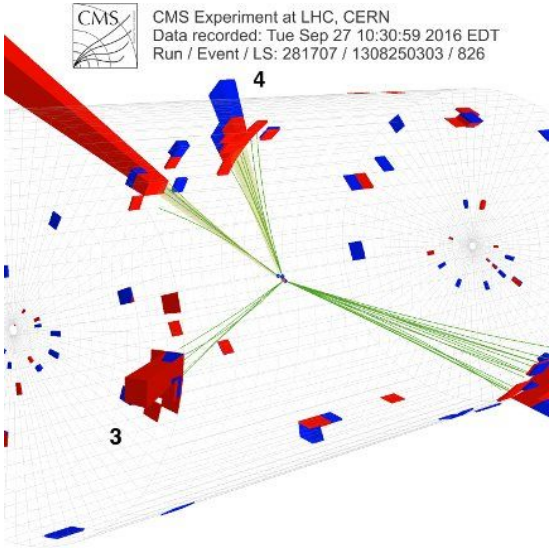
Robust Observables

[Enrico Bothmann et. al., 1905.09127;
CMS, 1810.10069]



Theory

Experiment



Perturbativity?
Hadronization?
Parton Shower Model?

Energy Flow!

Finite Calorimeter Resolution Effects?
Different Resolution between Detectors?

The Energy Flow

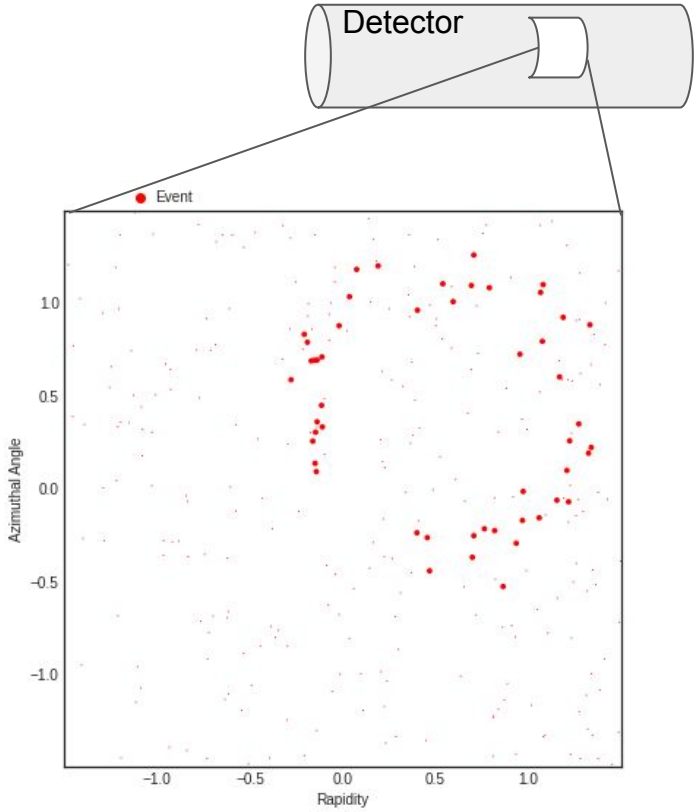
The Infrared and Collinear (IRC) safe information about a state is contained within its **Energy Flow**:

Detector Coordinate (η, ϕ)

$$\mathcal{E}(\vec{y}) = \sum_i z_i \delta(\vec{y} - \vec{y}_i)$$

Energy Fraction E_j / E_{Tot}

Can be either **real** or **idealized**.

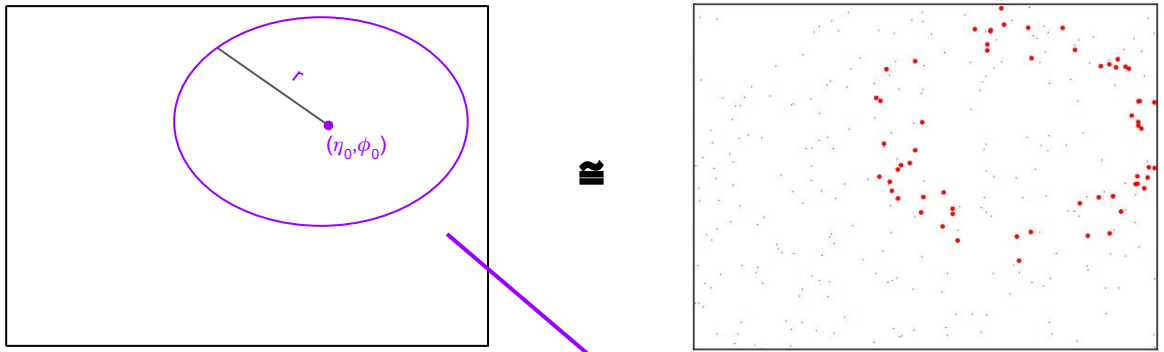


This plot *is* the energy flow for an event

Shapes and the Wasserstein Metric

Shapes and the Energy Flows

Translate our question about shapes into energy flows!



≅

Idealized Energy Flow

Real Energy Flow

What parameterized energy flow, $\boldsymbol{\varepsilon_\theta}$, *best* matches the observed energy flow $\boldsymbol{\varepsilon}$?

[D. Ba, A. Dogra, R. Gambhir, A. Tasissa, J. Thaler, In progress]

Shape Observables

Minimize over the manifold of parameterized energy flows to determine shape!

$$\mathcal{O}_{\mathcal{M}}(\mathcal{E}) = \min_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

$$\theta = \operatorname*{argmin}_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

Learns the "shapiness" $\mathcal{O}$ and the optimal shape parameters θ

Observables $\Leftrightarrow$ Manifold of Parameterized Flows correspondence!

[P. Komiske, E. Metodiev, J. Thaler, 1902.02346;
see also T. Cai, J. Cheng, K. Craig, N. Craig, 2111.03670;
see also C. Zhang, Y. Cai, G. Lin, C. Shen, 2003.06777;
see also L. Hou, C. Yu, D. Samaras, 1611.05916;
see also M. Arjovsky, S. Chintala, L. Bottou, 1701.07875]

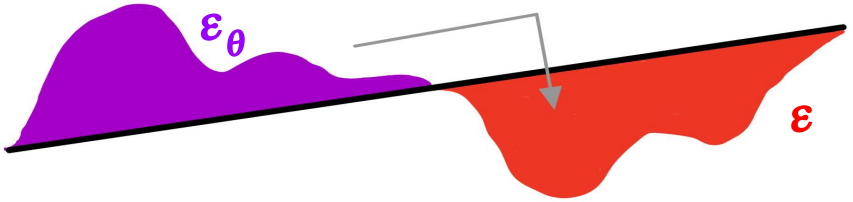
The Wasserstein Metric

There is a natural metric on probability distributions, the **Wasserstein Metric**

$$\mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta) = \sum_{i,j} f_{ij} \frac{|\vec{y}_i - \vec{y}'_j|}{R} + \left|E_i - E'_j\right|$$

$$\text{where} \sum_j f_{ij} = E_i, \sum_i f_{ij} = E'_j, \sum_{i,j} f_{ij} = \min(E_i, E'_j)$$

Also known as the Earth/Energy Mover's Distance (EMD) - geometric structure on events!



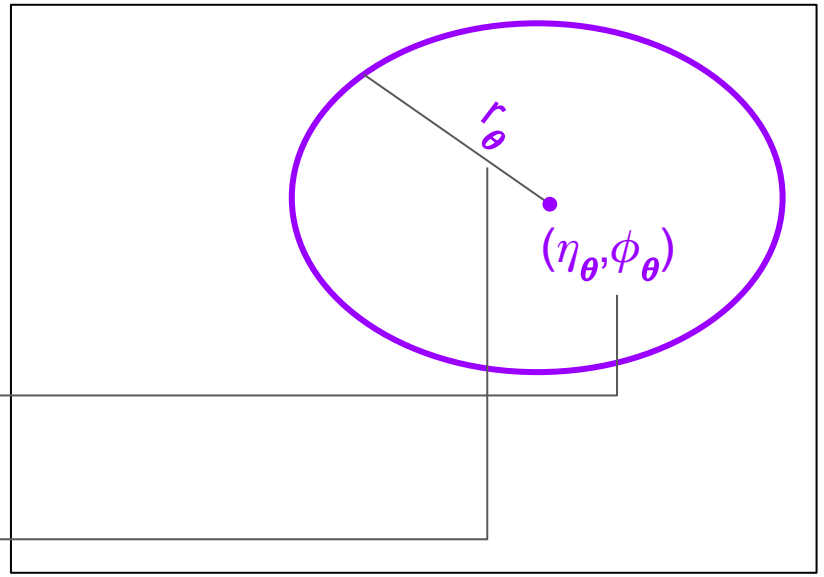
EMD = Work done to move "dirt" optimally

Defining Shapes

Define a *shape* as any parameterized energy flow

For example, for a circle with parameterized radius and center*

$$\mathcal{E}(\vec{y}) = \begin{cases} \frac{1}{2\pi r_\theta} & |\vec{y} - \vec{y}_\theta| = r_\theta \\ 0 & |\vec{y} - \vec{y}_\theta| \neq r_\theta \end{cases}$$



Our parameterized circle written as an energy flow

*Uniform prior by choice for simplicity. In principle, we can pick any parameterized normalized distribution.

Observable ⇔ Manifolds

[P. Komiske, E. Metodiev, and J. Thaler, 2004.04159;
J. Thaler, and K. Van Tilburg, 1011.2268;
I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, 1004.2489.;
S. Brandt, C. Peyrou, R. Sosnowski and A. Wroblewski, PRL 12 (1964) 57-61;
C. Cesarotti, and J. Thaler, 2004.06125]

Many existing observables have this form!

- *N*-subjettines ⇔ Manifold of *N*-point events
- *N*-jettiness ⇔ Manifold of *N*-point events with floating energy
- Thrust ⇔ Manifold of back-to-back point events
- Event Isotropy ⇔ Uniform distribution
- … and more!

All of the form “How much like **[shape]** does my **event** look like?”

We generalize this to build more observables!

The SHAPER Framework

[D. Ba, A. Dogra, R. Gambhir, A. Tasissa, J. Thaler, In progress]

SHAPER

Shape-**H**unting **A**lgorithm using **P**arameterized **E**nergy **R**econstruction

Framework for defining and building IRC-safe observables using parameterized objects

Easy to define new observables by specifying parameterization, or by combining shapes

Returns EMD distance and optimal shape parameters

$$\min_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

SHAPER

Sampling Energy Flows
Gradient Descent
Linear Programming
Simplex Projections

Loss and Shape ($\mathcal{O}$, θ)

[J. Feydy, tel.archives-ouvertes.fr/tel-02945979;
B. Charlier, J. Feydy, J. Alexis Glaunès F. D. Collin, G. Durif, JMLR:v22:20-275;
J. Feydy, T. Séjourné, F. X. Vialard, S. Amari, A. Trouvé, G. Peyré, 1810.08278]

Estimating Wasserstein

Sinkhorn Divergence: A strictly convex approximation to EMD! Dual potential formalism*:

$$\mathrm{EMD}_\epsilon(\mathcal{E}, \mathcal{E}') = \max_{f,g:\mathcal{X}\to\mathbb{R}} \left[\sum_{i\in\mathcal{E}} z_i f(x_i) + \sum_{j\in\mathcal{E}'} z'_j g(y_j) - \epsilon \sum_{ij\in\mathcal{E},\mathcal{E}'} z_i z'_j \left(e^{\frac{1}{\epsilon}(f(x_i)+g(y_j)-\theta_{ij})} - 1 \right) \right]$$

distance matrix

Can take gradients with respect to the entire event!

$$\nabla_{z_i} \mathrm{EMD}_\epsilon = f(x_i)$$
$$\nabla_{x_i} \mathrm{EMD}_\epsilon = z_i \nabla_{x_i} f(x_i)$$

Algorithm 3.4: Symmetric Sinkhorn algorithm, with debiasing

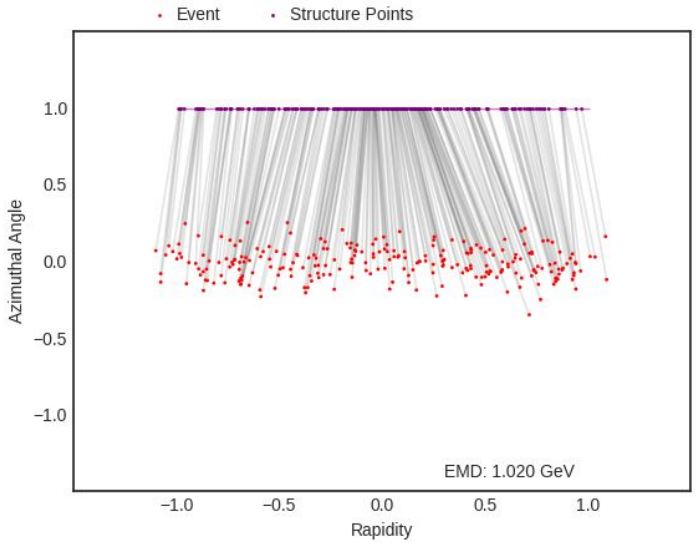
Parameters: Cost function $\mathbf{C} : (x_i, y_j) \in \mathcal{X} \times \mathcal{X} \mapsto \mathbf{C}(x_i, y_j) \in \mathbb{R}$,
Temperature $\varepsilon > 0$.
Input: Positive measures $\alpha = \sum_{i=1}^{N} \alpha_i \delta_{x_i}$ and $\beta = \sum_{j=1}^{M} \beta_j \delta_{y_j}$ with the same mass.

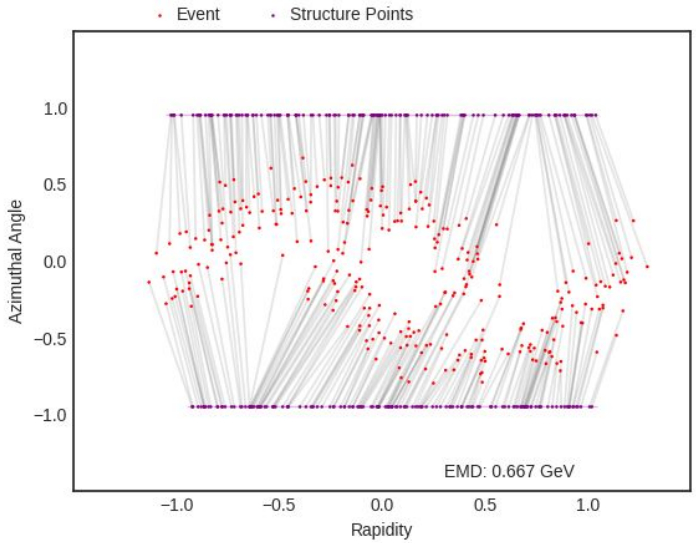
$f_i^{\beta\to\alpha}, g_j^{\alpha\to\beta}, f_i^{\alpha\leftrightarrow\alpha}, g_j^{\beta\leftrightarrow\beta} \leftarrow \mathbf{0}_{\mathbb{R}^N}, \mathbf{0}_{\mathbb{R}^M}, \mathbf{0}_{\mathbb{R}^N}, \mathbf{0}_{\mathbb{R}^M}$ ▷ Dual vectors.
repeat ▷ The four lines below are executed **simultaneously**.
$f_i^{\beta\to\alpha} \leftarrow \frac{1}{2} f_i^{\beta\to\alpha} + \frac{1}{2} \min_{y\sim\beta,\varepsilon} \left[\mathbf{C}(x_i, y) - g^{\alpha\to\beta}(y) \right]$, ▷ $\alpha \leftarrow \beta$
$g_j^{\alpha\to\beta} \leftarrow \frac{1}{2} g_j^{\alpha\to\beta} + \frac{1}{2} \min_{x\sim\alpha,\varepsilon} \left[\mathbf{C}(x, y_j) - f^{\beta\to\alpha}(x) \right]$, ▷ $\beta \leftarrow \alpha$
$f_i^{\alpha\leftrightarrow\alpha} \leftarrow \frac{1}{2} f_i^{\alpha\leftrightarrow\alpha} + \frac{1}{2} \min_{x\sim\alpha,\varepsilon} \left[\mathbf{C}(x_i, x) - f^{\alpha\leftrightarrow\alpha}(x) \right]$, ▷ $\alpha \leftarrow \alpha$
$g_j^{\beta\leftrightarrow\beta} \leftarrow \frac{1}{2} g_j^{\beta\leftrightarrow\beta} + \frac{1}{2} \min_{y\sim\beta,\varepsilon} \left[\mathbf{C}(y, y_j) - g^{\beta\leftrightarrow\beta}(y) \right]$. ▷ $\beta \leftarrow \beta$
until convergence up to a set tolerance. ▷ Monitor the updates on the potentials.
return $f_i^{\beta\to\alpha} - f_i^{\alpha\leftrightarrow\alpha}$, $g_j^{\alpha\to\beta} - g_j^{\beta\leftrightarrow\beta}$ ▷ Debiased dual potentials $F(x_i)$ and $G(y_j)$.

Implemented using the KerOps Python Package!

*Needs to be debiased, not shown here for simplicity

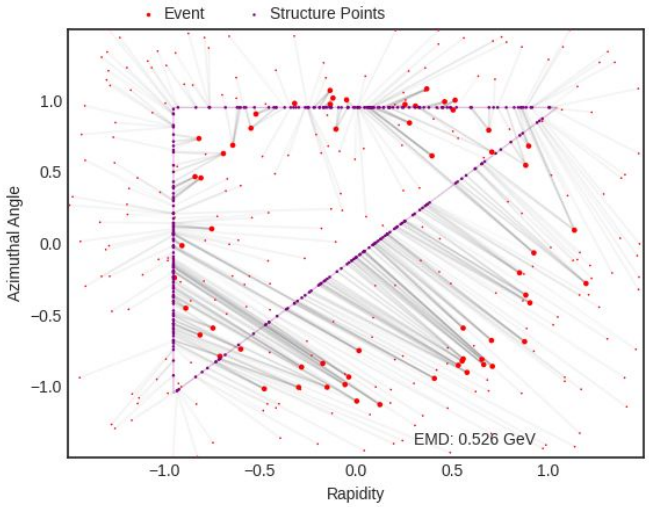
Fun Animations

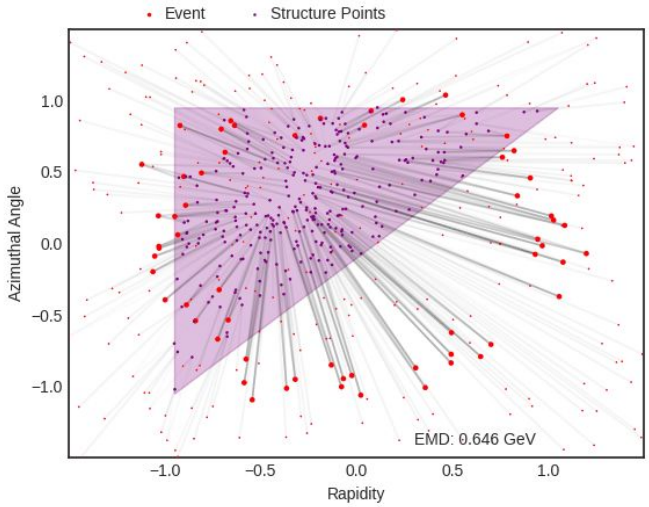




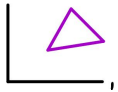
Red: Event Y
Purple: Shape ε_θ with structure points a_i
Grey: Matrix x_{ij} connecting y's and a_i's

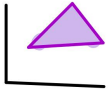
Fun Animations Cont'd





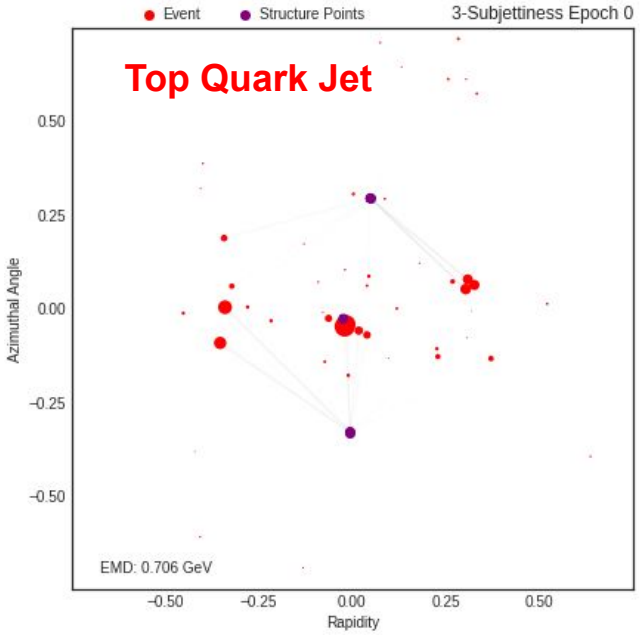
Red: Event *Y*
Purple: Shape ε_θ with structure points a_i
Grey: Matrix f_{ij} connecting y's and a_i's

Left: ε_θ =  EMD = 0.245

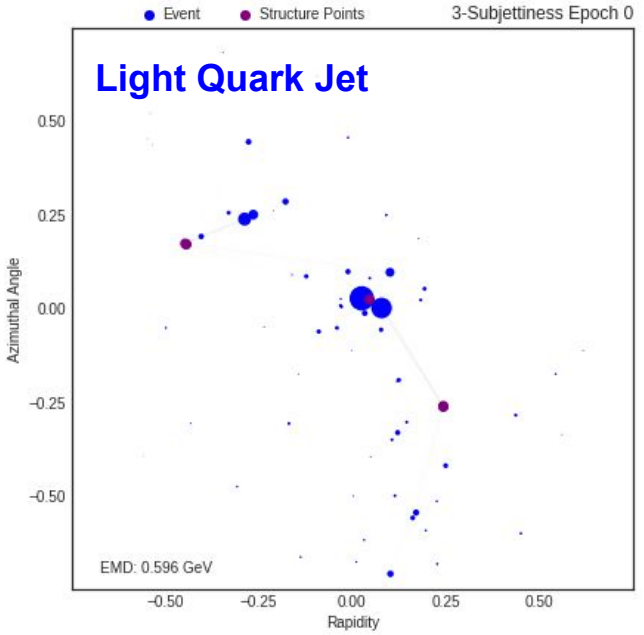
Right: ε_θ =  EMD = 0.279

N-Subjettiness

Easy to compute classic jet observables!



$\tau_3 = 0.214$



$\tau_3 = 0.343$

[see also L., B. Nachman, A. Schwartzman, C. Stansbury, 1509.02216;
see also B. Nachman, P. Nef, A. Schwartzman, M. Swiatlowski, C. Wanotayaroj, 1407.2922;
see also M. Cacciari, G. Salam, 0707.1378]

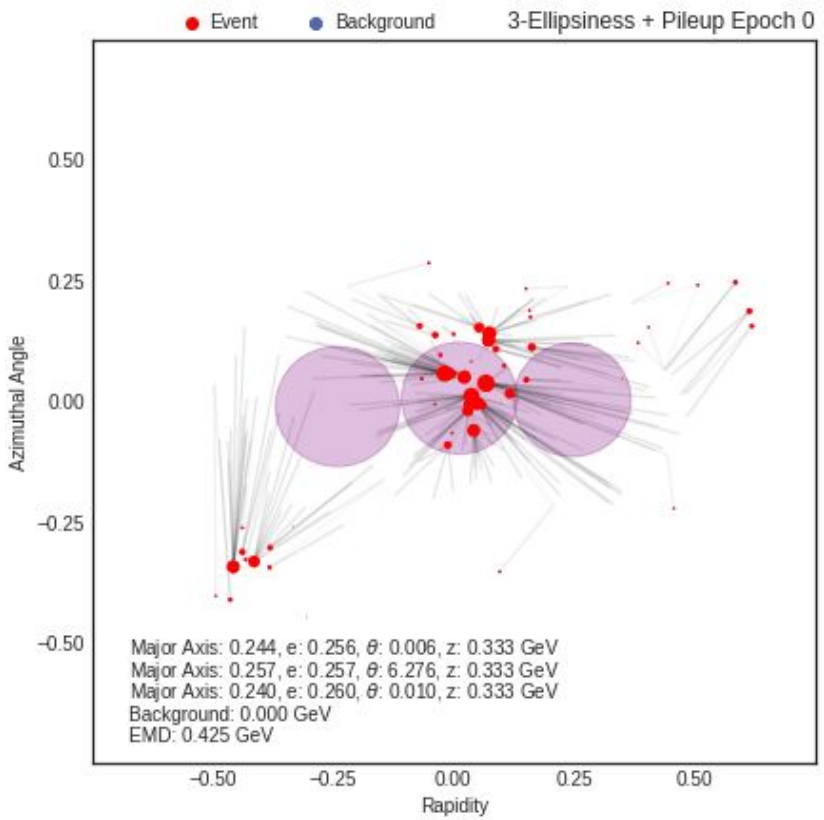
New IRC-Safe Observables

The **SHAPER** framework makes it easy to invent new jet observables!

e.g. ***N*-Ellipsiness+Pileup** as a jet algorithm.

- Learn jet centers
- Dynamic jet radii (no *R* hyperparameter)
- Dynamic eccentricities and angles
- Dynamic jet energies
- Uniform Pileup Subtraction
- Learned parameters for discrimination

Can design custom specialized jet algorithms to learn jet substructure!



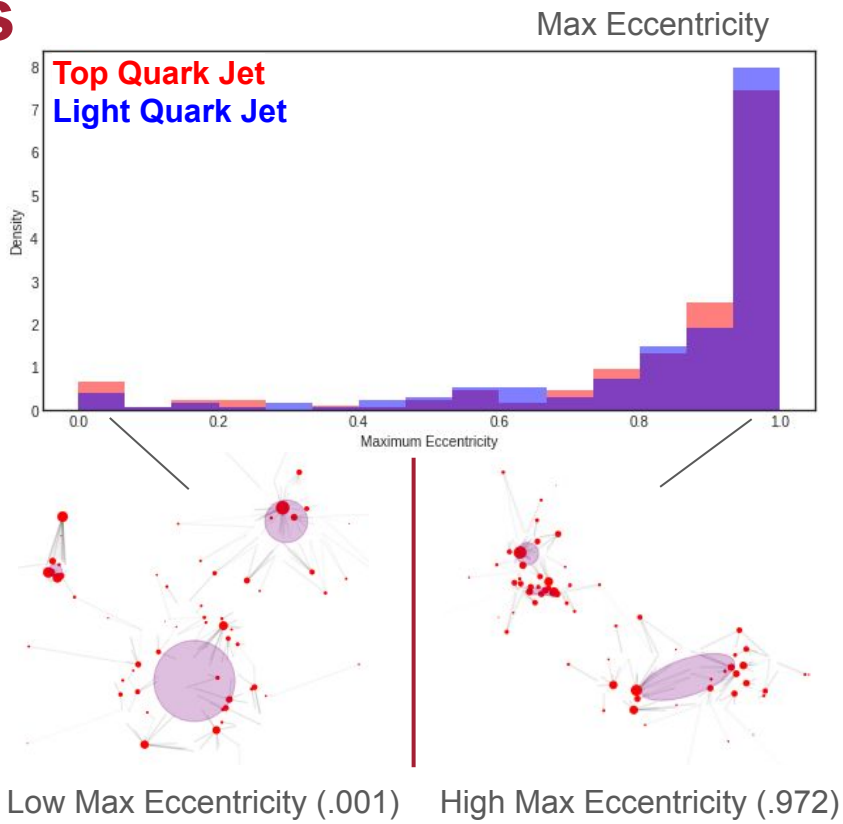
New IRC-Safe Observables

The **SHAPER** framework makes it easy to invent new jet observables!

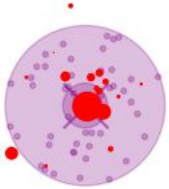
e.g. ***N*-Ellipsiness+Pileup** as a jet algorithm.

- Learn jet centers
- Dynamic jet radii (no *R* hyperparameter)
- Dynamic **eccentricities** and angles
- Dynamic jet energies
- Uniform Pileup Subtraction
- Learned parameters for discrimination

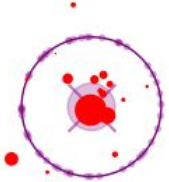
Can design custom specialized jet algorithms to learn jet substructure!



Observables on CMS OpenData

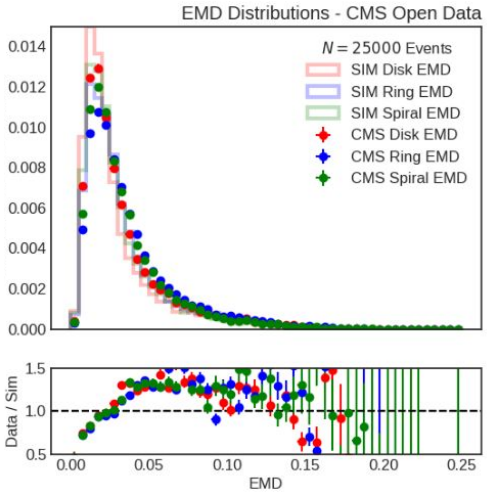


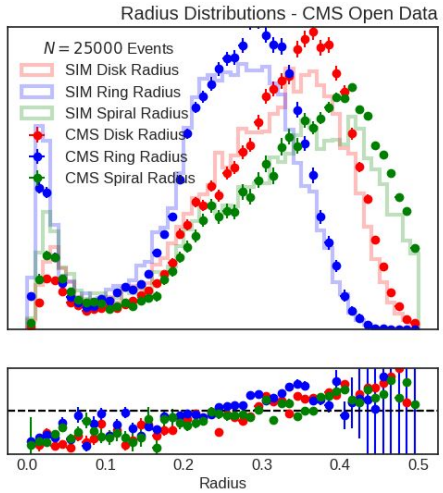
Disk + δ-function

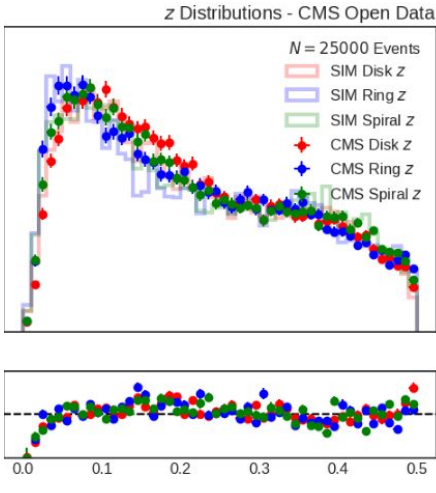


Ring + δ-function

Spiral + δ-function







Probing **collinear** (δ-function) and **soft** (shape) structure!

Outlook

SHAPER, a machine learning framework for calculating robust observables for collider physics based on IRC-safety and Wasserstein geometry!

Playground for defining and building custom observables and jet algorithms!

$$\mathcal{O}_{\mathcal{M}}(\mathcal{E}) = \min_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

$$\theta = \operatorname*{argmin}_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

Outlook

SHAPER, a machine learning framework for calculating robust observables for collider physics based on IRC-safety and Wasserstein geometry!

Playground for defining and building custom observables and jet algorithms!

$$\mathcal{O}_{\mathcal{M}}(\mathcal{E}) = \min_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

$$\theta = \operatorname*{argmin}_{\mathcal{E}'_\theta \in \mathcal{M}} \mathrm{EMD}(\mathcal{E}, \mathcal{E}'_\theta)$$

Using **SHAPER**, you ***CAN*** hear the shape of a jet!

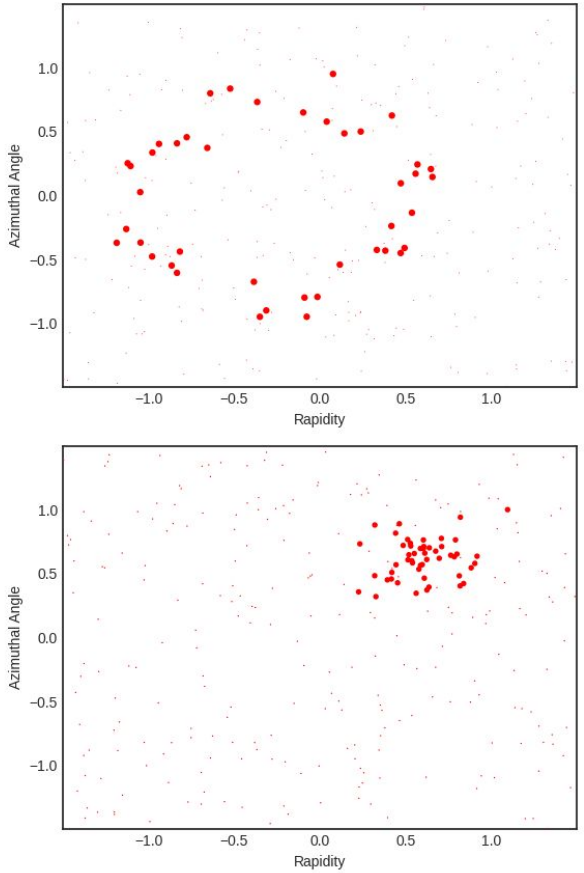
More questions? Email me at rikab@mit.edu

Appendices

Toy Analysis

Define the following two types of events:

- Type 1: Ring-Like
  - Pois(50) “signal” particles, uniformly making up 70%-90% of event energy
  - Arranged in a ring with radius $r \sim N(0.75, 0.25)$, width = 0.1
  - Pois(250) “background” particles making up remaining energy
- Type 2: Disc-Like
  - Pois(50) “signal” particles, uniformly making up 70%-90% of event energy
  - Arranged in a disc with radius $r \sim N(0.50, 0.25)$
  - Pois(250) “background” particles making up remaining energy



Observables

In analogy with *N-jettiness*, define the *A-shapeliness* of an event as the value of the loss when evaluated on the shape *A*.

For this toy analysis, define:

- Shape A_1: Filled-in triangle, parameterized by its vertices
- Shape A_2: Boundary of a triangle, parameterized by its vertices (∂A1)

Both shapes have $R = 0.25$

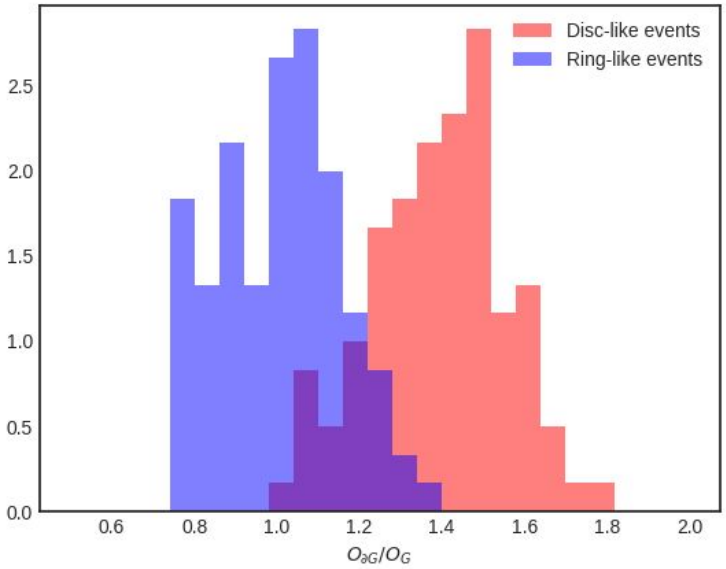
The ratio of the shapeliness values should be a proxy for if the event is ring-like or disc-like

Technical Aspects

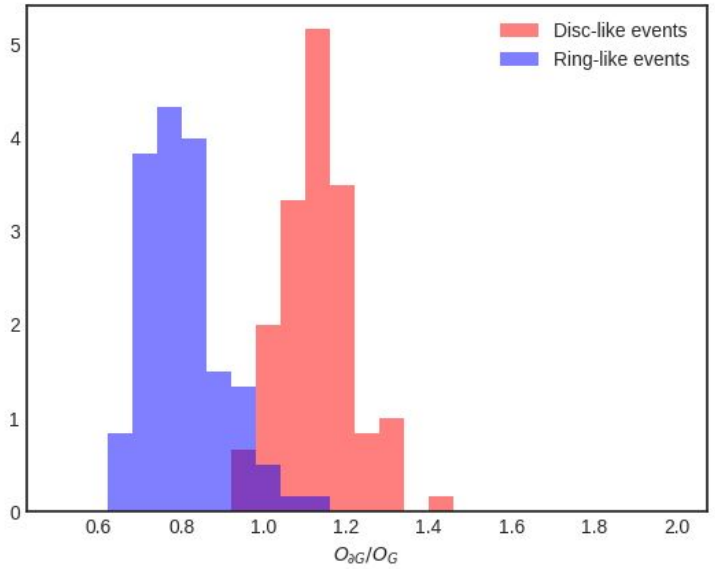
- 125 sample points defining shapes
- Adam optimizer, lr = 0.05
- 125 epochs with *z0* frozen, then 125 epochs with *z0* unfrozen
  - Early stopping if loss has not improved after 10 epochs
- Triangles initialized at (-1, -1), (-1, 1), (1,1)

With these settings, each observable takes about 2-3 seconds per event.

Results

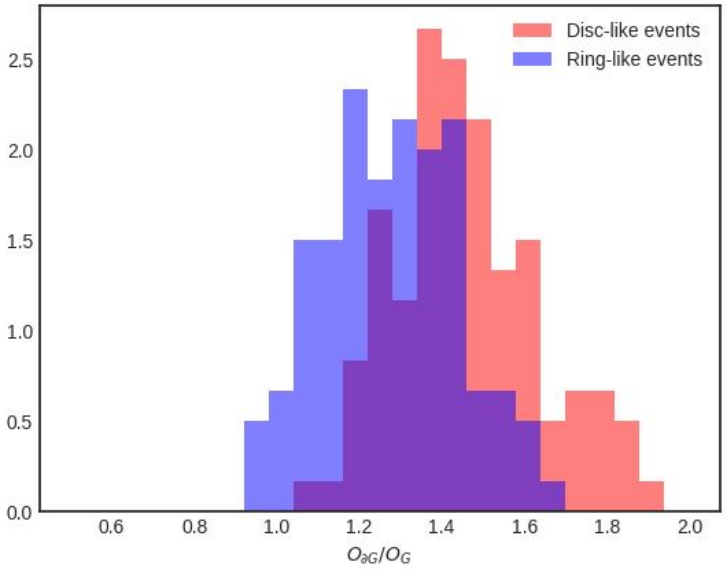


Without unclustered term

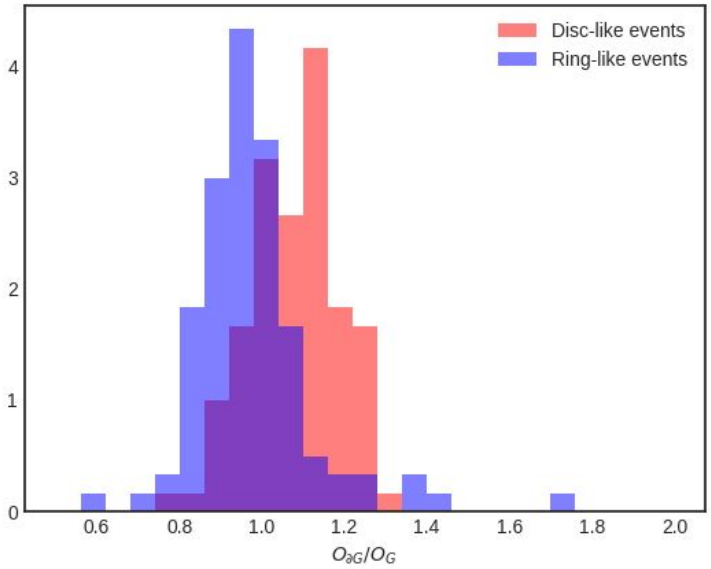


With unclustered term

Results - High Background (40%-60%)



Without unclustered term



With unclustered term

IRC Safety

Infrared Safety: An observable is unchanged under a soft emission

Collinear Safety: An observable is unchanged under a collinear splitting

[Pranay Tankala, Abiy Tasissa, James M. Murphy, Demba Ba, 2012.02134]

Implementation

To practically determine this minimum …

1. Initialize *A* by random sampling the energy flow
2. Freeze *A*. Calculate the Wasserstein Metric loss *L*, and the corresponding transport matrix f_{iy}
3. Freeze *f*. Calculate the gradients of *L* with respect to *A* [ignoring dependencies on *f*]
4. Gradient update *A*
5. Freeze *f* and *A*. Gradient update weights *z* by numerical derivative
6. Repeat 2-5 until convergence.