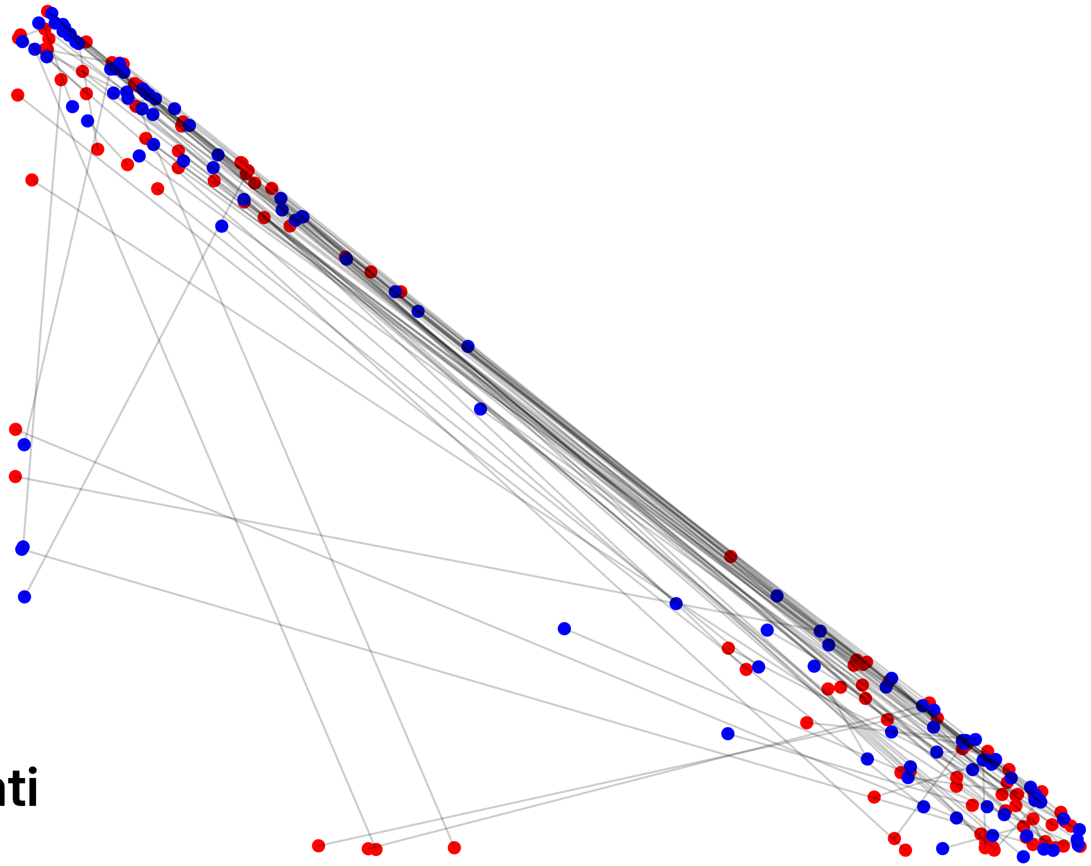
Optimal transport and CP violation: The Dalitz Mover's Distance



Tony Menzo
PhD candidate, University of Cincinnati

In collaboration with Adam Davis, Ahmed Youssef, and Jure Zupan

Based on 2301.13211

Standard lore

Given a field X

- The charge operator conjugates all internal quantum numbers e.g.

$$C[X^{\pm}] = X^{\mp}$$

- The parity operator inverts 'handedness' ($\vec{x} \to -\vec{x}$), for fermionic $X^{\pm}$

$$P[X^{\pm}_{L}] = X^{\pm}_{R}$$

* Chiral couplings tend to cause CP violation

CP in weak meson decays

$$-\mathcal{L}_W = \frac{g}{\sqrt{2}} \overline{u_{L,i}} \gamma^\mu \underbrace{(V_{\mathrm{CKM}})_{ij}}_{\text{Contains a single CP-violating phase}} d_{L,j} W_\mu^+ + h.c.$$

CP violation has been observed (at 5σ+) in ~30 different hadron decay modes involving s, c, and b quarks.

Direct CP violation

Requires at least two interferring amplitudes:

$$A(X \to \cdots) = a_1 e^{i\phi_1} e^{i\delta_1} + a_2 e^{i\phi_2} e^{i\delta_2}$$

$$A(\bar{X} \to \overline{\cdots}) = a_1 e^{-i\phi_1} e^{i\delta_1} + a_2 e^{-i\phi_2} e^{i\delta_2}$$

$$\downarrow$$

$$\frac{\Gamma(X) - \Gamma(\bar{X})}{\Gamma(X) + \Gamma(\bar{X})} = \frac{-a_1 a_2 \sin(\phi_1 - \phi_2) \sin(\delta_1 - \delta_2)}{|a_1|^2 + |a_2|^2 + 2 a_1 a_2 \cos(\phi_1 - \phi_2) \cos(\delta_1 - \delta_2)}$$

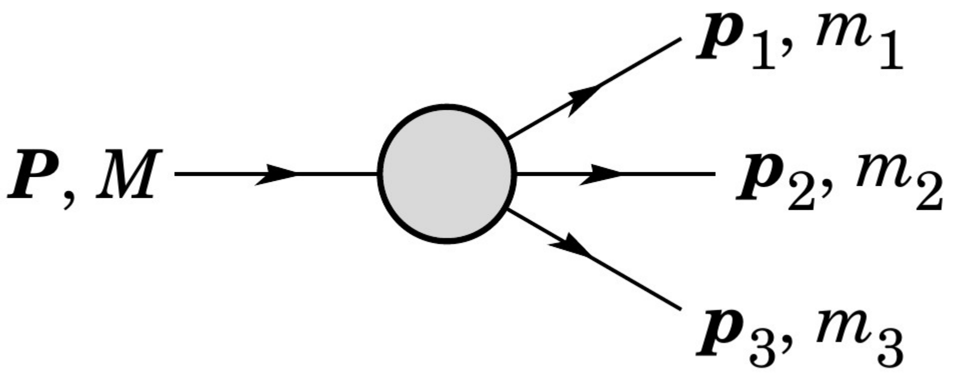
Neccesary condition for direct CPV: $\frac{\mathcal{A}(X \to \cdots)}{\mathcal{A}(\bar{X} \to \overline{\cdots})} \neq 1$

3-body kinematics

12 degrees of freedom
- 10 constraints

2 independent variables

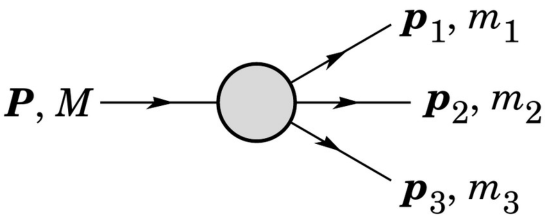
Phase space-dependent CPV



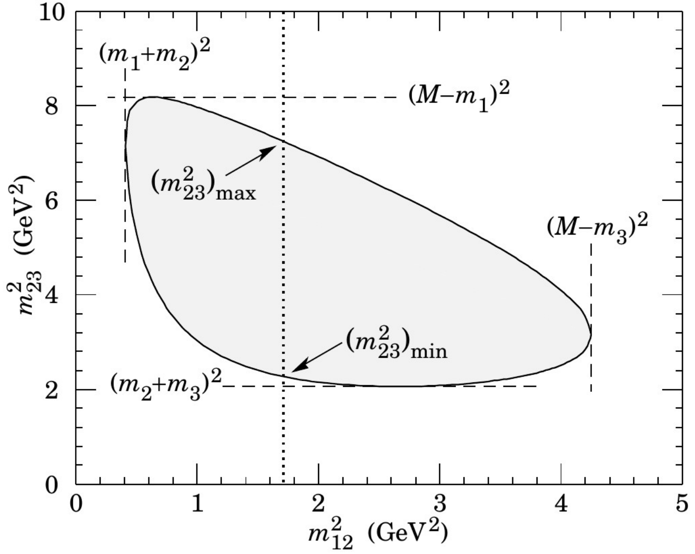
** Generically for N-body decays there are 3N – 7 degrees of freedom

Dalitz analysis

Consider a 3-body decay:



$$m_{ij}^2 = (p_i + p_j)^2$$

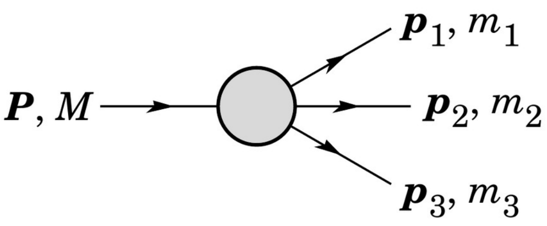


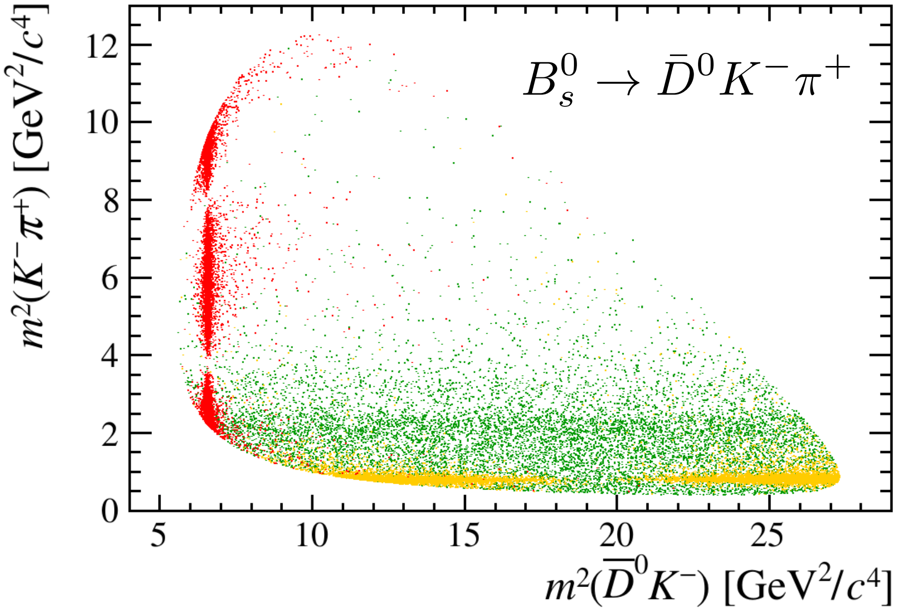
What to remember:

$$d\Gamma \propto |\mathcal{M}|^2 dm_{12}^2 dm_{23}^2$$

Dalitz analysis

Consider a 3-body decay:





What to remember:

$$d\Gamma \propto |\mathcal{M}|^2 dm_{12}^2 dm_{23}^2$$

Direct CPV manifests as local density asymmetries between conjugate Dalitz plots

Is there CPV?

1. Binned test statistic

$$A_{\mathrm{CP}} = \frac{N - \bar{N}}{N + \bar{N}}$$

2. Model the conjugate amplitudes

3. Unbinned test statistic

$$T = \sum_{i,j>i}^{N} \frac{\psi_{ij}}{N(N-1)} + \sum_{i,j>i}^{\bar{N}} \frac{\psi_{ij}}{\bar{N}(\bar{N}-1)} - \sum_{i,j}^{N,\bar{N}} \frac{\psi_{ij}}{N\bar{N}}, \qquad \psi_{ij} \equiv \psi(d_{ij};\sigma) = e^{-d_{ij}^2/2\sigma^2}$$

Is there CPV?

1. Binned test statistic

$$A_{\mathrm{CP}} = \frac{N - \bar{N}}{N + \bar{N}}$$

- Binning washes out asymmetry

2. Model the conjugate amplitudes

3. Unbinned test statistic

$$T = \sum_{i,j>i}^{N} \frac{\psi_{ij}}{N(N-1)} + \sum_{i,j>i}^{\bar{N}} \frac{\psi_{ij}}{\bar{N}(\bar{N}-1)} - \sum_{i,j}^{N,\bar{N}} \frac{\psi_{ij}}{N\bar{N}}, \qquad \psi_{ij} \equiv \psi(d_{ij};\sigma) = e^{-d_{ij}^2/2\sigma^2}$$

Is there CPV?

1. Binned test statistic

$$A_{\mathrm{CP}} = \frac{N - \bar{N}}{N + \bar{N}}$$

- Binning washes out asymmetry

2. Model the conjugate amplitudes

- Complex + many fit parameters, model dependent

3. Unbinned test statistic

$$T = \sum_{i,j>i}^{N} \frac{\psi_{ij}}{N(N-1)} + \sum_{i,j>i}^{\bar{N}} \frac{\psi_{ij}}{\bar{N}(\bar{N}-1)} - \sum_{i,j}^{N,\bar{N}} \frac{\psi_{ij}}{N\bar{N}}, \qquad \psi_{ij} \equiv \psi(d_{ij};\sigma) = e^{-d_{ij}^2/2\sigma^2}$$

Is there CPV?

1. Binned test statistic

$$A_{\mathrm{CP}} = \frac{N - \bar{N}}{N + \bar{N}}$$

– Binning washes out asymmetry

2. Model the conjugate amplitudes

– Complex + many fit parameters, model dependent

3. Unbinned test statistic

Widely used in experimental searches

$$T = \sum_{i,j>i}^{N} \frac{\psi_{ij}}{N(N-1)} + \sum_{i,j>i}^{\bar{N}} \frac{\psi_{ij}}{\bar{N}(\bar{N}-1)} - \sum_{i,j}^{N,\bar{N}} \frac{\psi_{ij}}{N\bar{N}},$$

$$\psi_{ij} \equiv \psi(d_{ij};\sigma) = e^{-d_{ij}^2/2\sigma^2}$$

What do we need?

- Discrete, 2D test statistic
- Sensitive to density asymmetries

What do we need?

- Discrete, 2D test statistic
- Sensitive to density asymmetries

⟶ Statistical distances

Statistical distance

Input two statistical samples and output scalar value.

Larger values = dissimilar distributions

Smaller values = similar distributions

Given samples x, y and z from distribution X:

$$d(x, y) \geq 0$$

$$d(x, y) = 0$$ if and only if x = y

$$d(x, y) = d(y, x)$$

$$d(x, y) \leq d(x, y) + d(y, z)$$

Examples:

Total variation distance

Wasserstein or Earth Mover's distance (EMD)

Hellinger's distance

Statistical distance

Input two statistical samples and output scalar value.

Larger values = dissimilar distributions

Smaller values = similar distributions

Given samples x, y and z from distribution X:

Divergence $d(x, y) \geq 0$ $d(x, y) = 0$ if and only if x = y

Metric $d(x, y) = d(y, x)$ $d(x, y) \leq d(x, y) + d(y, z)$

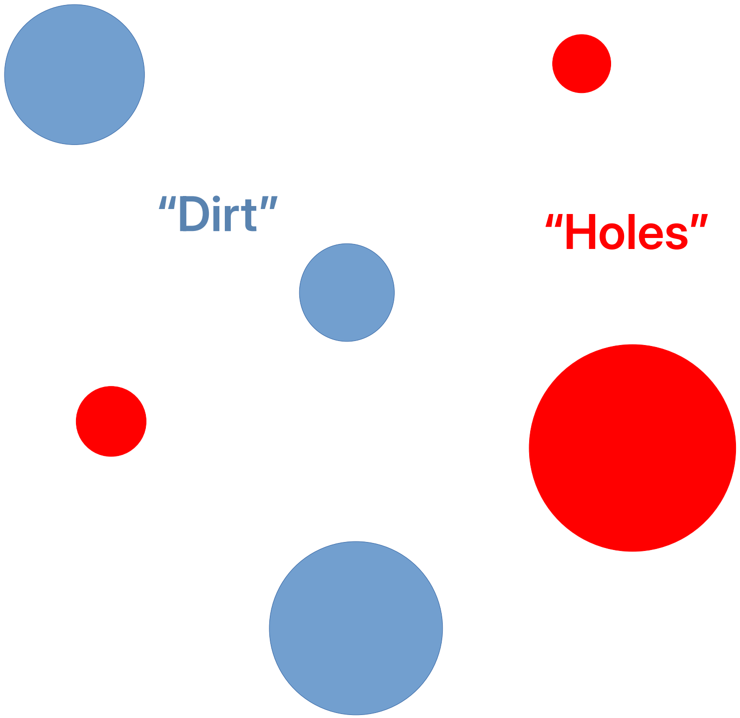
Examples:

Total variation distance

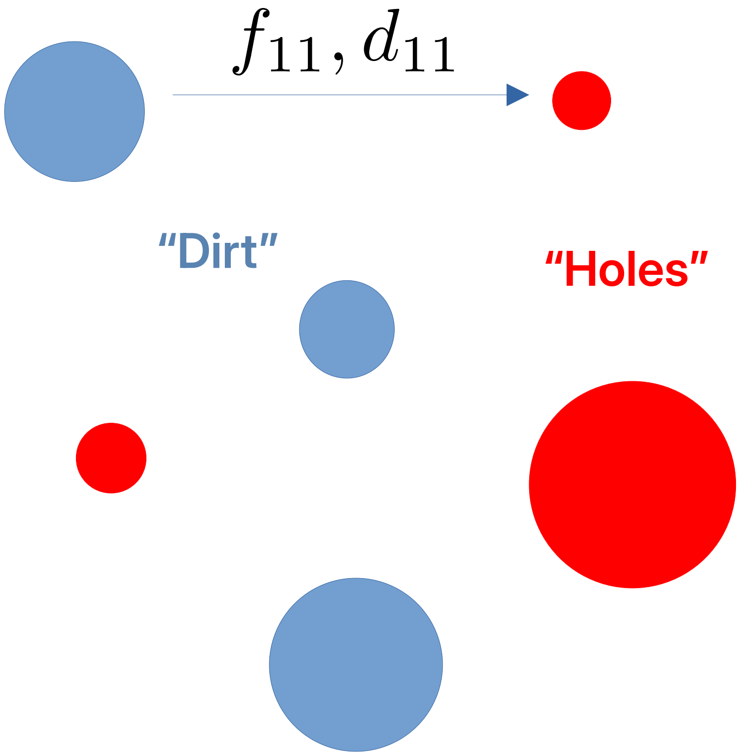
Wasserstein or Earth Mover's distance (EMD)

Hellinger's distance

Wasserstein or Earth mover's distance



Wasserstein or Earth mover's distance



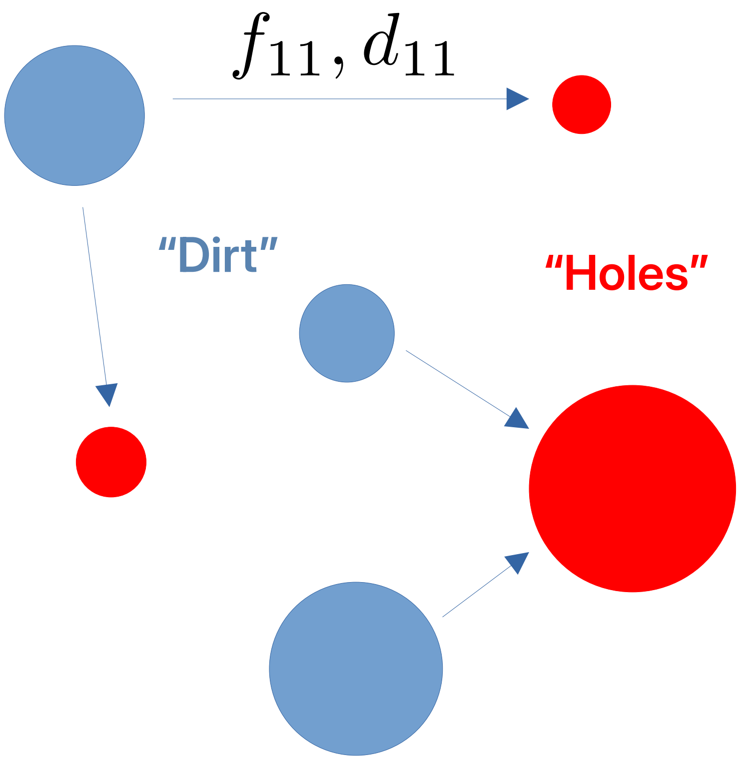
Each pile has some associated "work" required to transport from a given pile to a hole.

Given distance and transport matrices, the total cost of a particular "plan" is given by

$$\text{Total cost} = \sum_i \sum_j f_{ij} d_{ij}$$

Minimize!

Wasserstein or Earth mover's distance



The optimal transport plan is given by the following optimization problem:

$$\min_{\boldsymbol{f}} \sum_i \sum_j f_{ij} d_{ij}$$

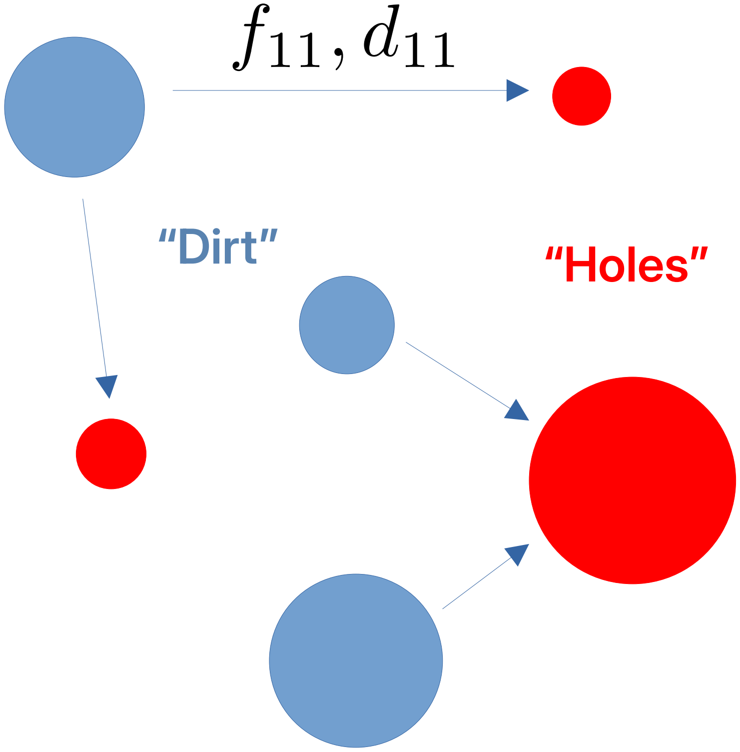
Subject to the following constraints:

$$\sum_i f_{ij} = a_i$$

$$\sum_j f_{ij} = b_j$$

$$f_{ij} \geq 0 \quad \forall \quad (i,j)$$

Wasserstein or Earth mover's distance



The optimal transport plan is given by the following optimization problem:

$$\min_{\boldsymbol{f}} \sum_i \sum_j f_{ij} d_{ij}$$

Subject to the following constraints:

$$\sum_i f_{ij} = a_i$$

$$\sum_j f_{ij} = b_j$$

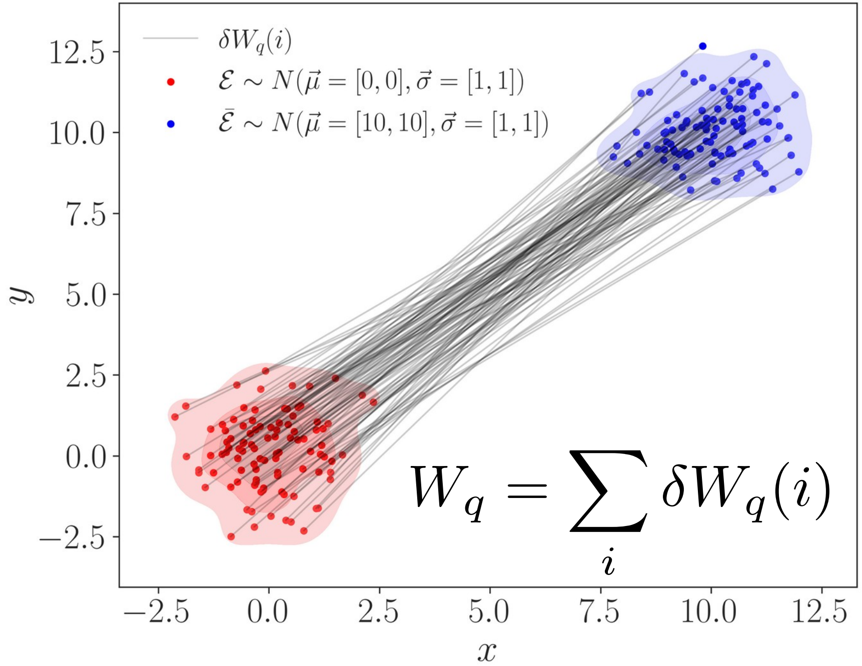
$$f_{ij} \geq 0 \quad \forall \quad (i,j)$$

Linear programming optimzation problem, many numerically efficient algorithms exist e.g. simplex, etc

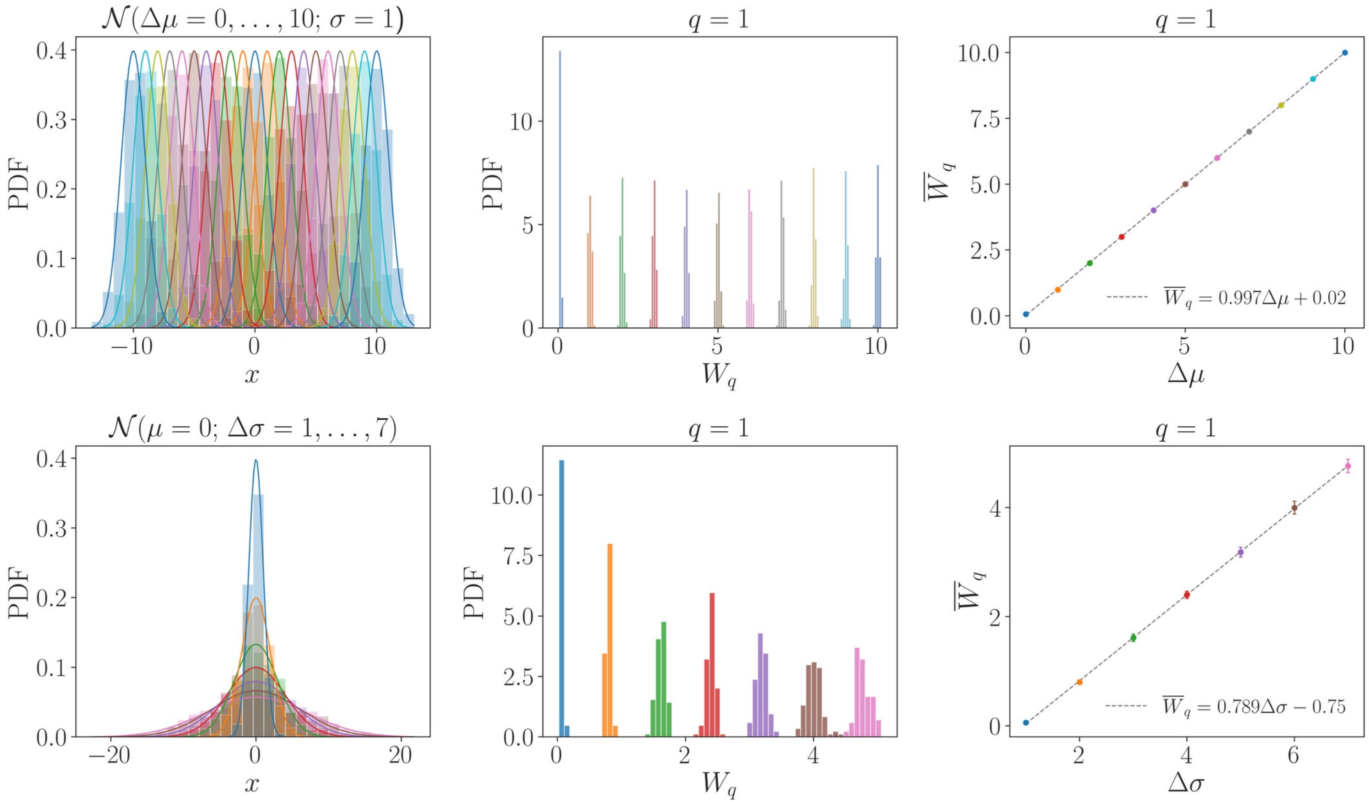
Earth mover's distance in practice

Compute the transport matrix and sum up all distances

$$W_q(\mathcal{E},\bar{\mathcal{E}}) = \left[\min_{\{f_{ij}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{\bar{N}} f_{ij}\left(\hat{d}_{ij}\right)^q\right]^{1/q}$$



EMD intuition

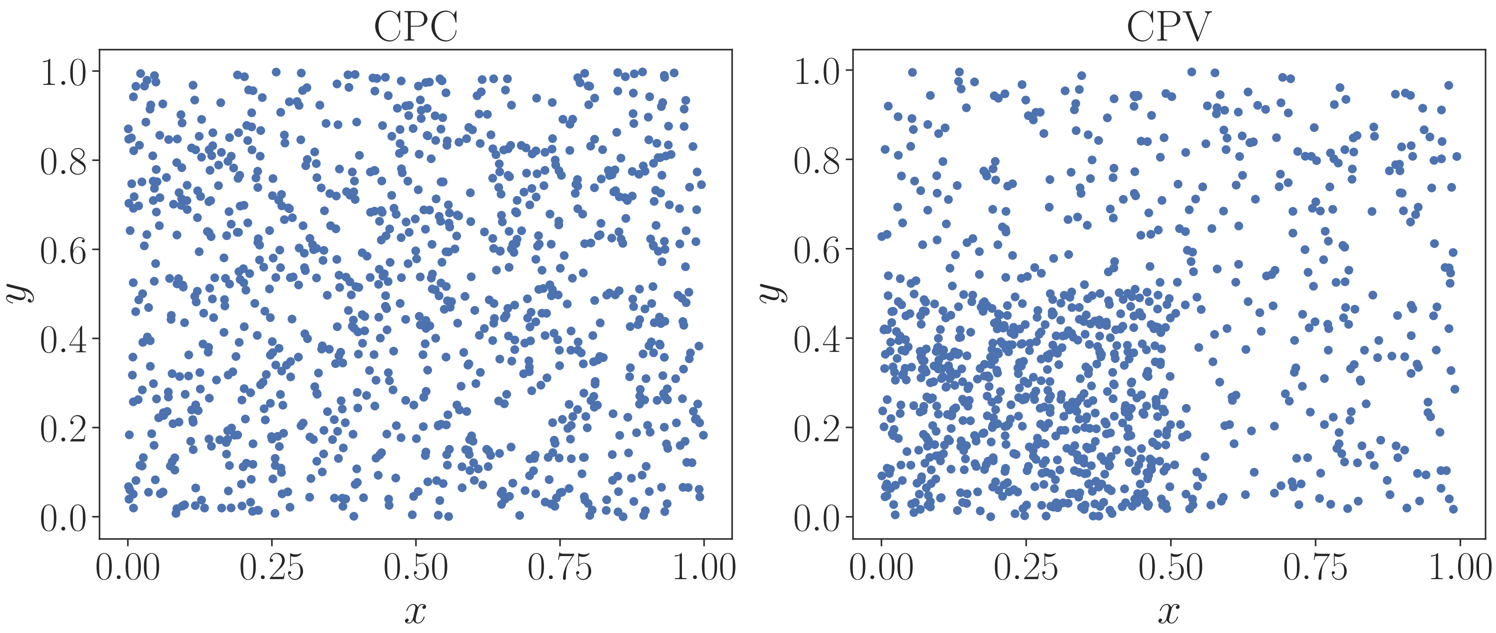


Construction of the test statistic

1) Model the CP conserving distribution of the statistic e.g. permutation method
2) Compute the statistic between conjugate amplitudes
3) Read off p-value based on CP conserving model

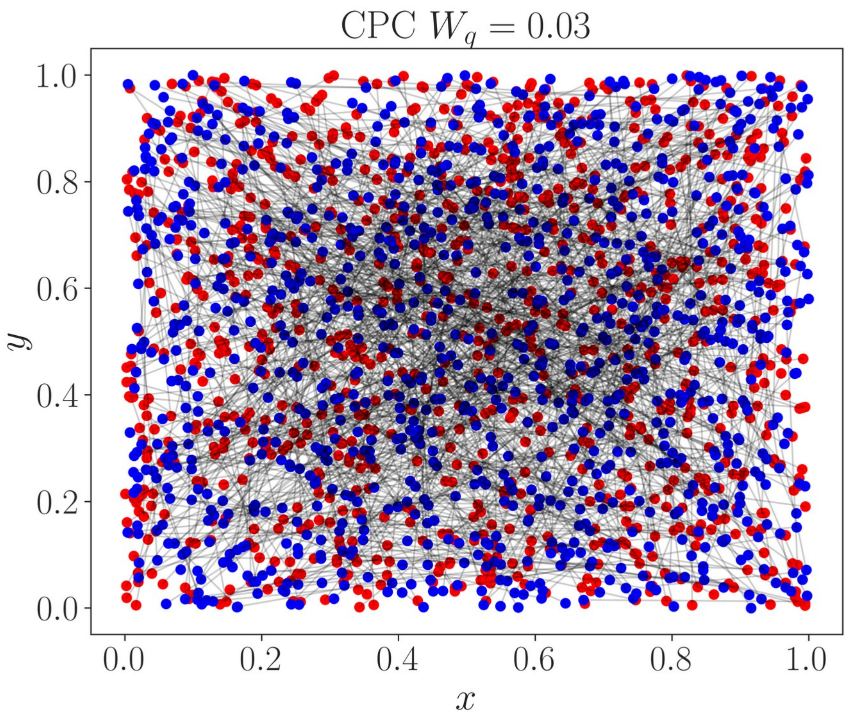
EMD and CPV

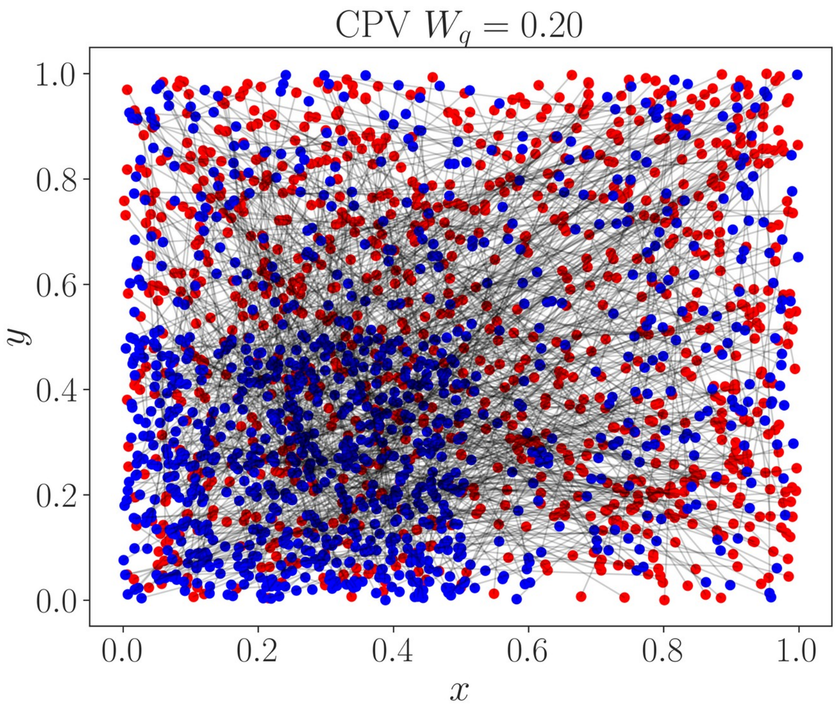
An exagerated example:



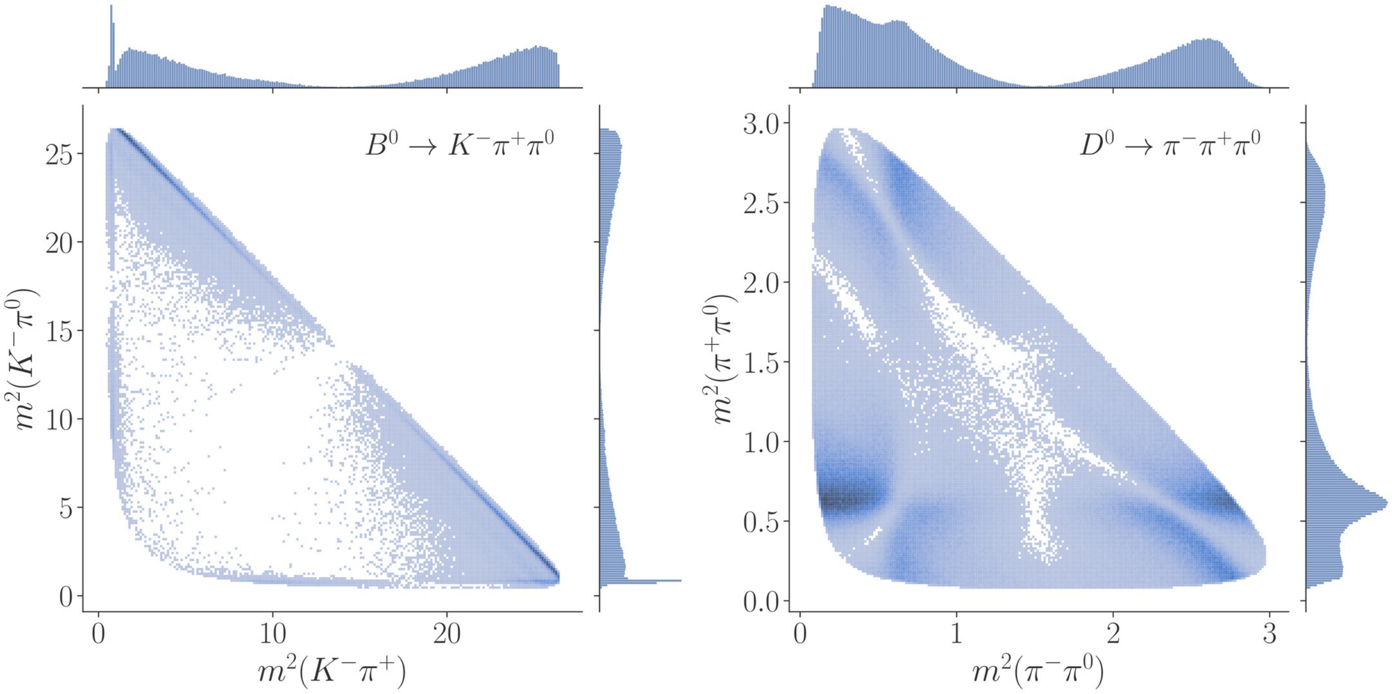
EMD and CPV

Direct CPV manifests as local density asymmetries between conjugate Dalitz plots.





Data

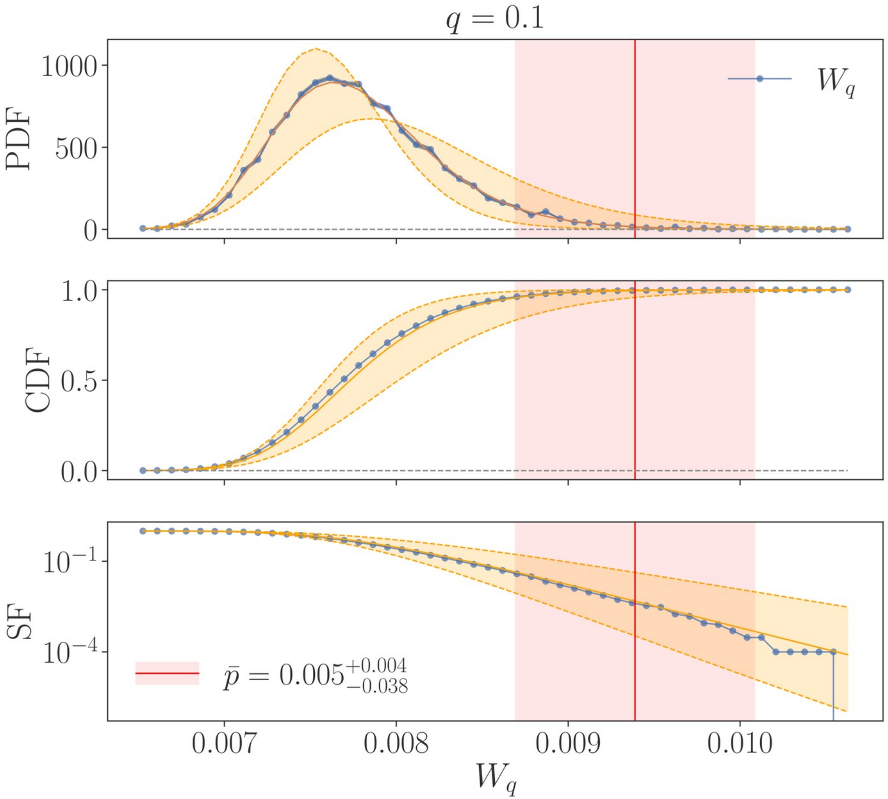


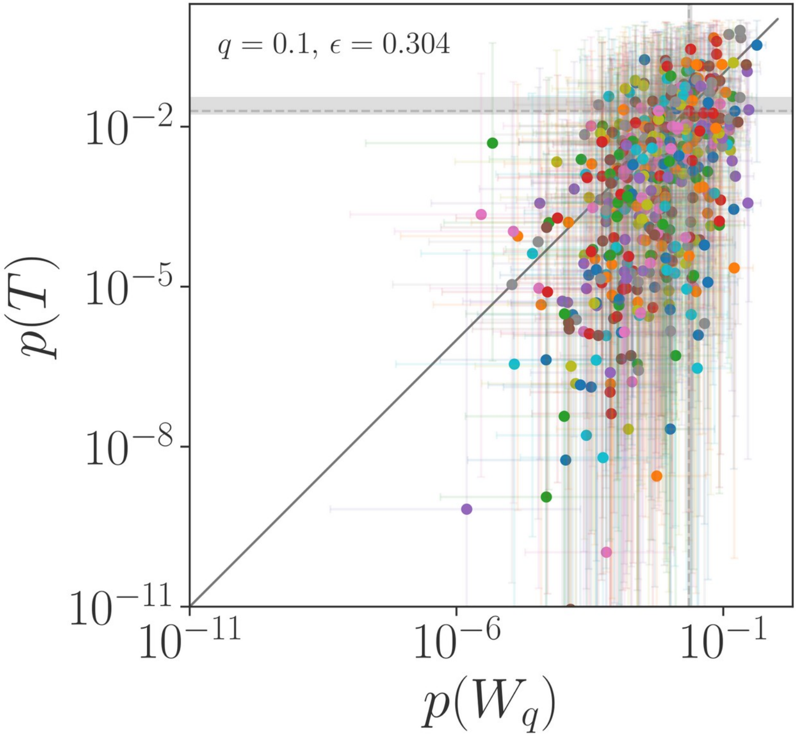
Two analyses: Low statistics (1000 event datasets) and large statistics (10^5 event datasets)

CP conserving model

Comparison between permuation and “master” models of CPV

Performance as a statistic



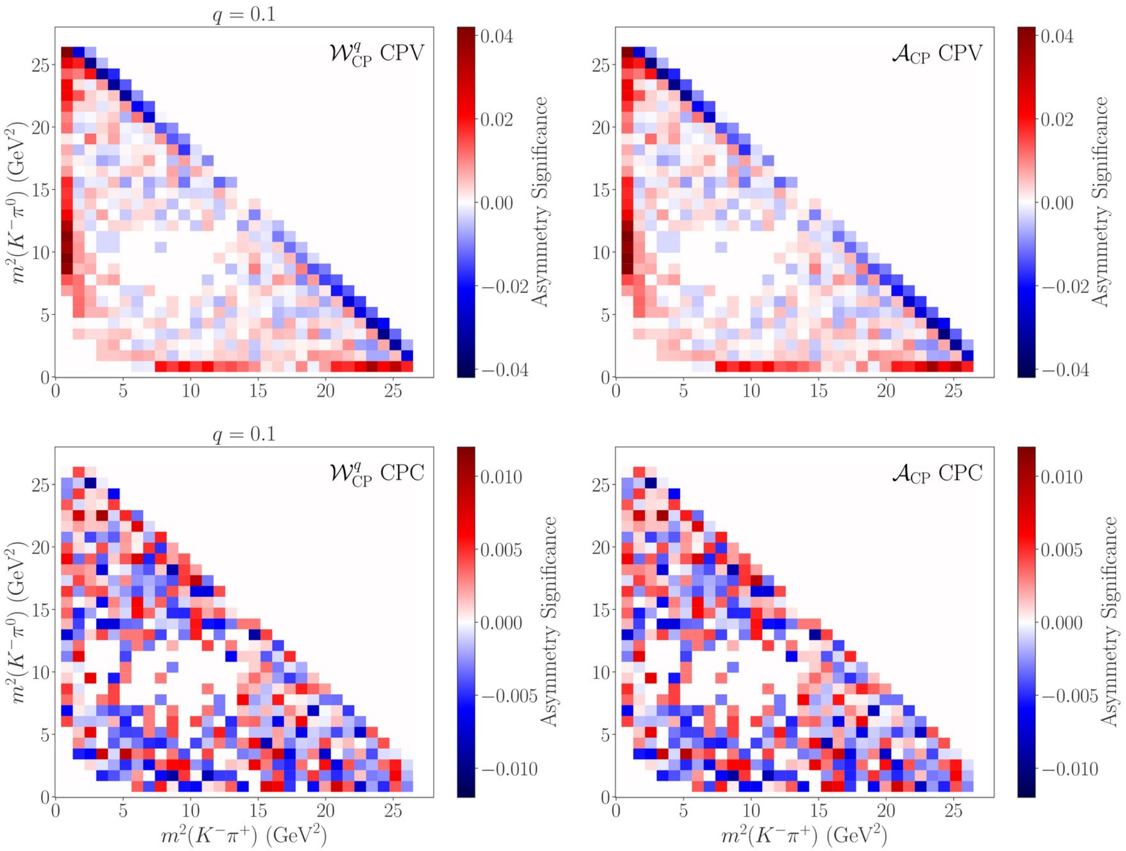


Visualization

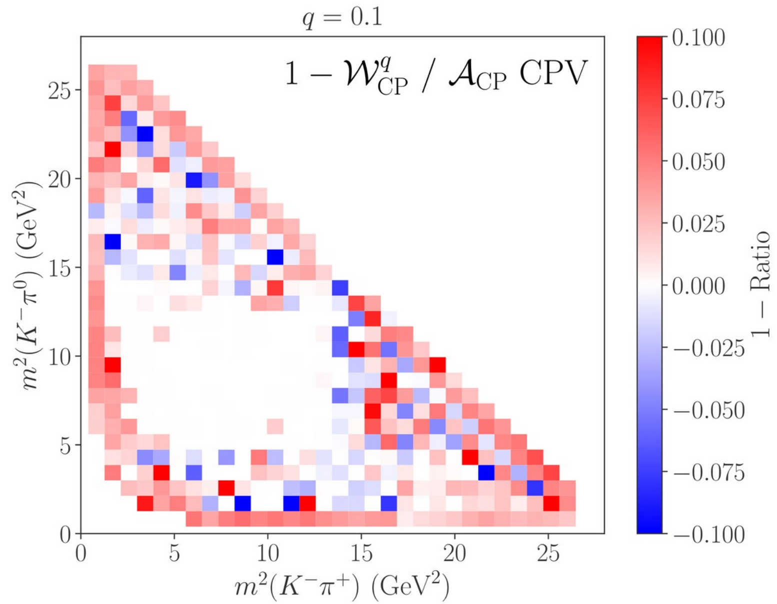
$$\mathcal{A}_{\mathrm{CP}}(s_{12}, s_{13}) = \frac{d\bar{\Gamma}(\bar{s}_{12}, \bar{s}_{13}) - d\Gamma(s_{12}, s_{13})}{d\bar{\Gamma}(\bar{s}_{12}, \bar{s}_{13}) + d\Gamma(s_{12}, s_{13})}$$

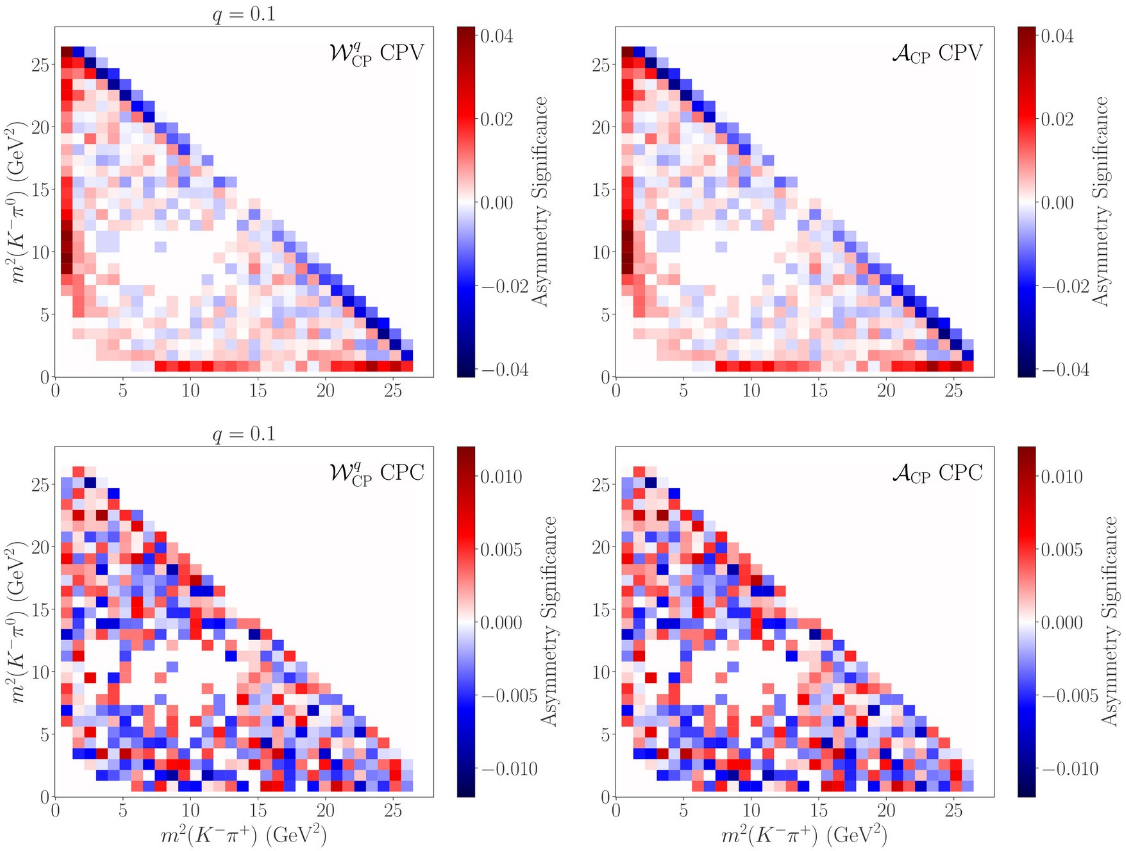
$$W_q = \sum_i \delta W_q(i)$$

$$\mathcal{W}^q_{\mathrm{CP}}(s_{12}, s_{13}) = \frac{\sum_i \delta \bar{W}_i - \sum_i \delta W_i}{\sum_i \delta \bar{W}_i + \sum_i \delta W_i}$$



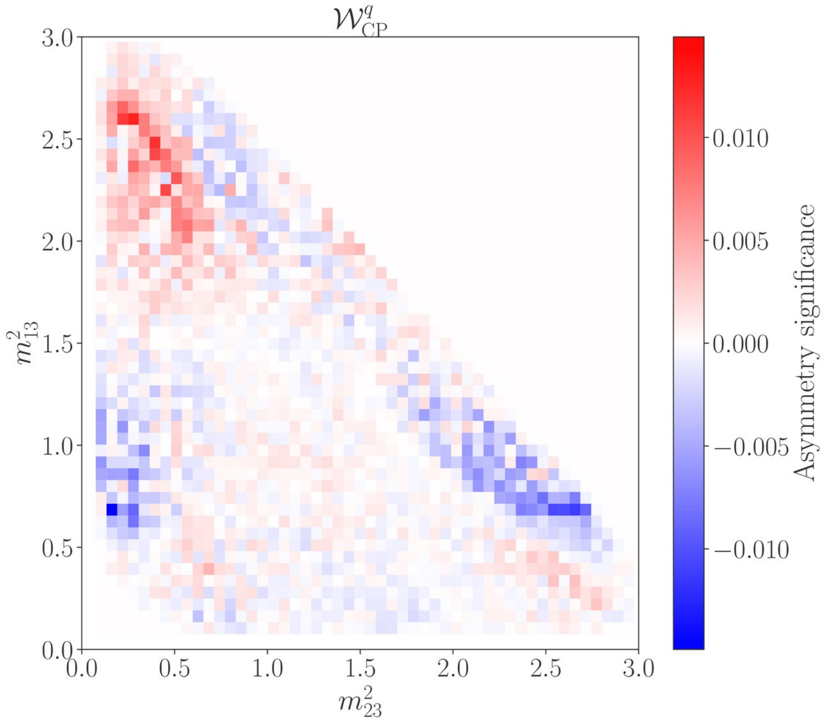
Visualization

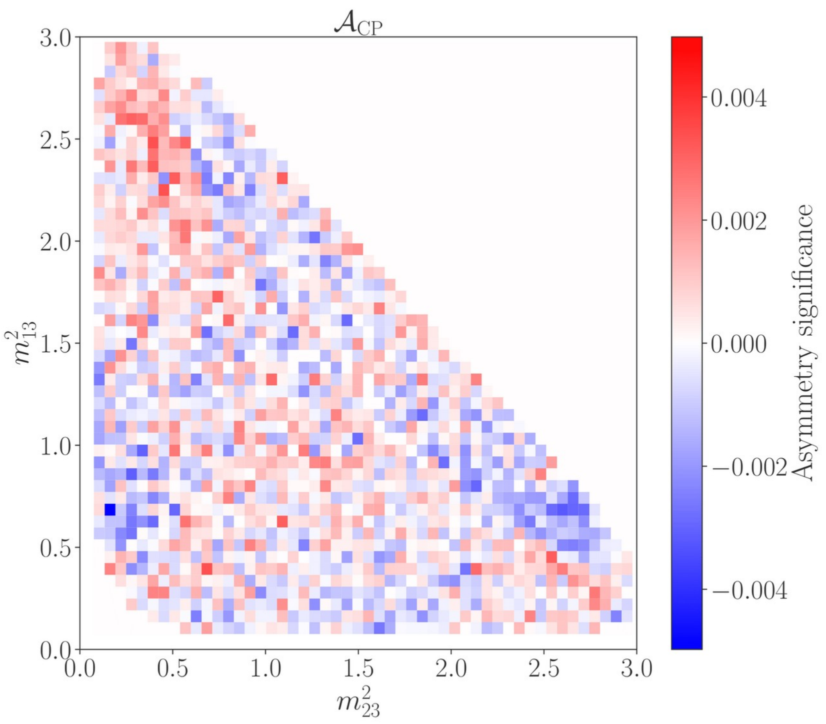




Visualization

Binned variation for large dataset:





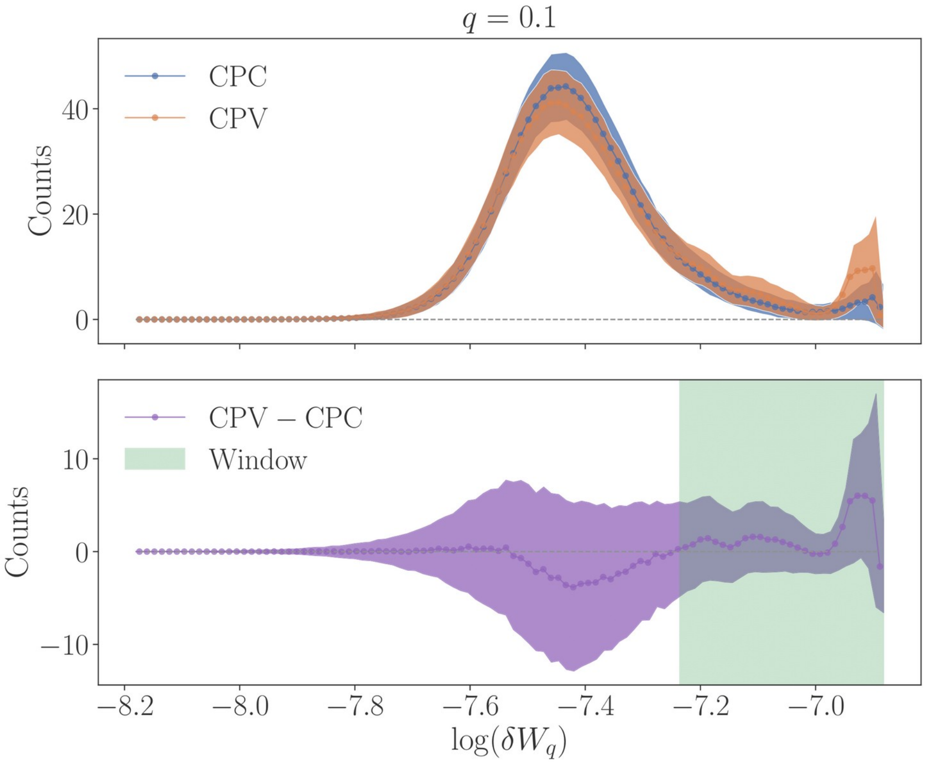
How can we increase sensitivity?

Important: Density asymmetries signify CP violation

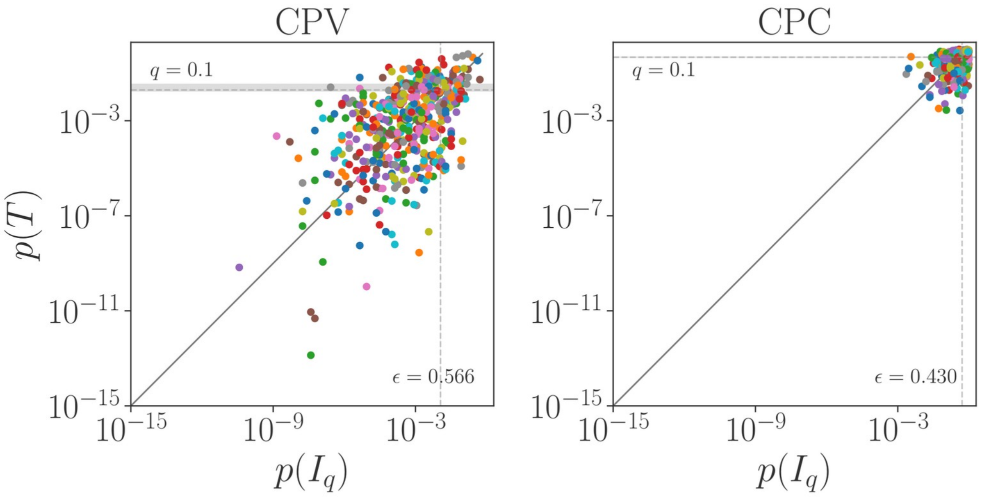
The only feature that is relevant for determining CPV are the frequency of relative distances (not the magnitude).

Can we get rid of the polluting CP conserving distances?

Windowed-Wasserstein statistic



$$w(x) = \begin{cases} +1 & x \in [\delta W^{\text{window}}_{\text{min}}, \delta W^{\text{window}}_{\text{max}}] \\ -1 & x \in [\overline{\delta W^{\text{window}}_{\text{min}}}, \overline{\delta W^{\text{window}}_{\text{max}}}] \\ 0 & \text{otherwise} \end{cases}$$

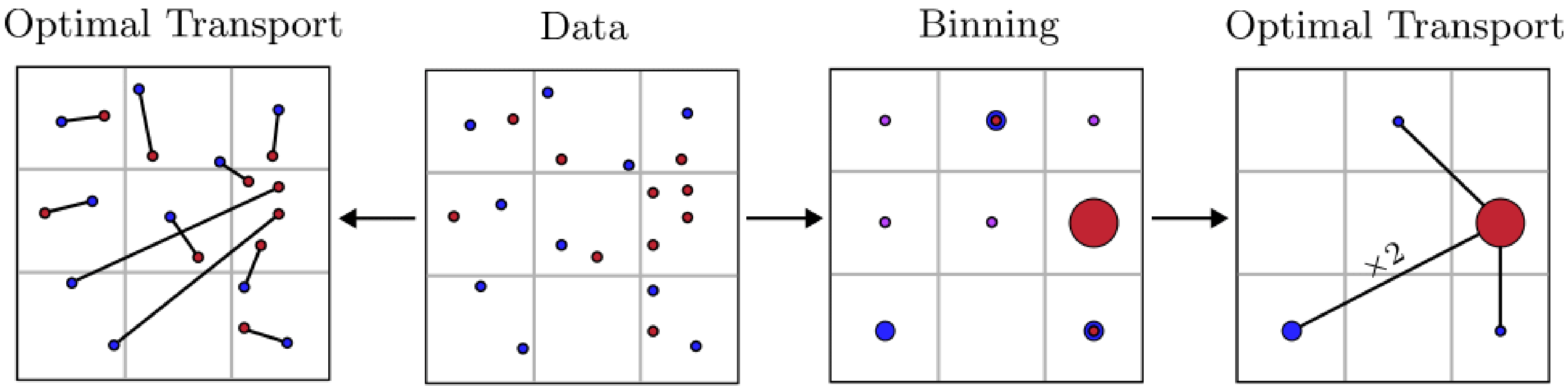


Windowed-Wasserstein statistic

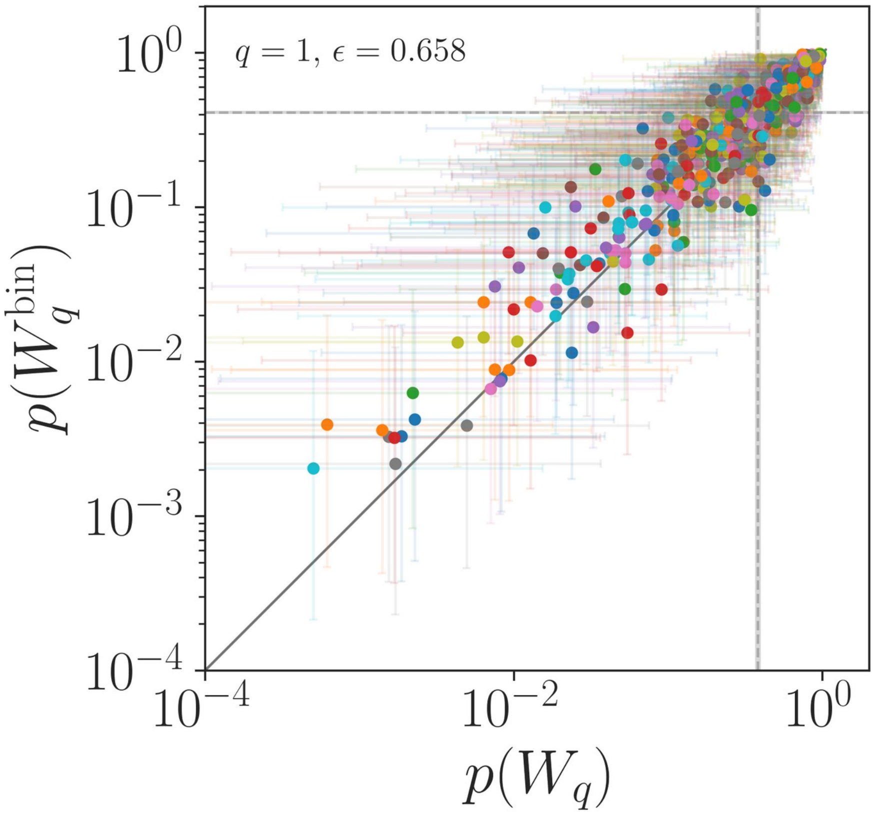
Dalitz mover's distance variants

Naively the EMD algorithm scales like $O(N^3)$

Two ways to improve: (1) Binned-EMD

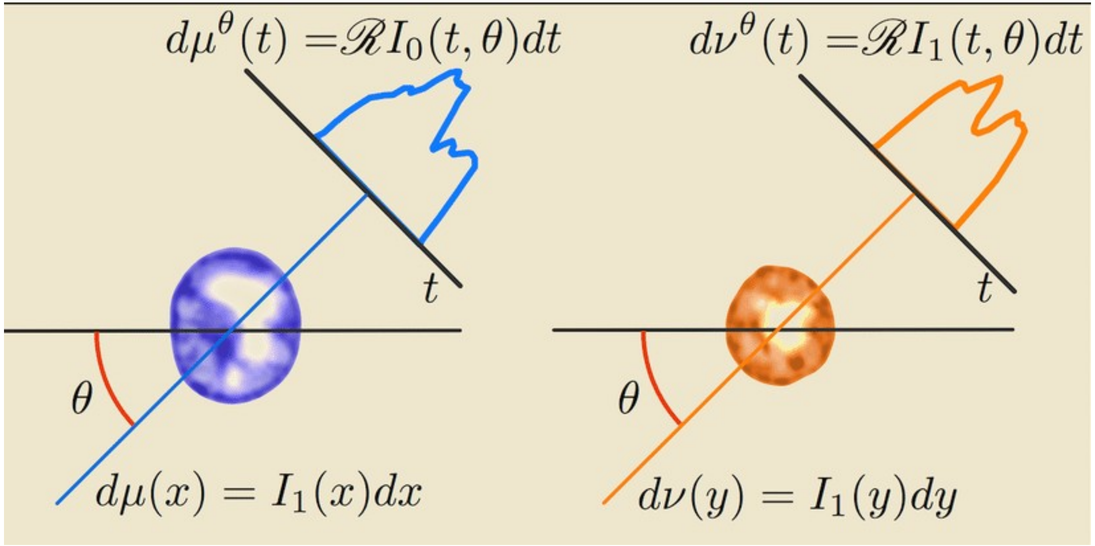


Dalitz mover's distance variants



Dalitz mover's distance variants

(2) Sliced EMD - scales like O(n $\log_2$n)



Dalitz mover’s distance variants

In practice

- In practice the Dailtz mover's distance is complementary to the direct CP asymmetry

What about other types of CPV?

Neutral meson oscillations introduce a richness and complexity into the phenomenology of CP violation.

$$\Gamma(X^0 \to f) \neq \Gamma(\overline{X^0} \to f)$$

$$\text{Prob}(X^0 \to \overline{X^0}) \neq \text{Prob}(\overline{X^0} \to X^0)$$

$$\text{Prob}(X^0_{\rightsquigarrow \bar{X}^0} \to f)(t) \neq \text{Prob}(\bar{X}^0_{\rightsquigarrow X^0} \to f)(t)$$

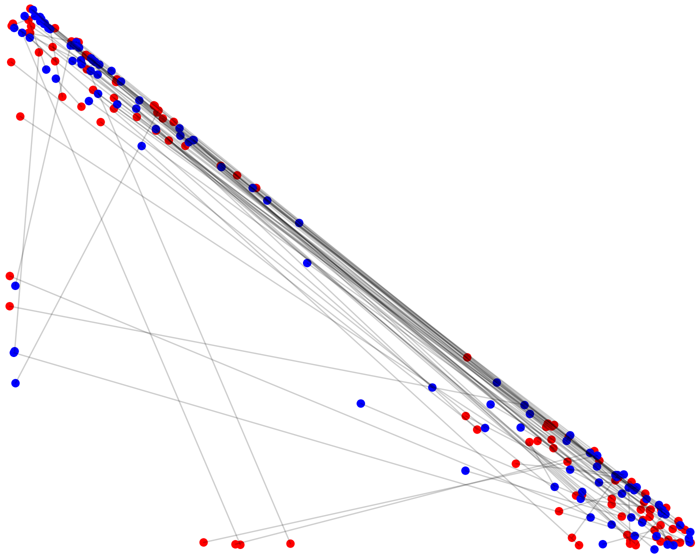
What about N>3?

The complexity of amplitude models increase significantly with N – it is still interesting to know if these modes exhibit CP violation.

Conclusion

The Wasserstein distance can be used as a robust, model independent, unbinned test of CP violation.

- **Future directions:**
  - Time dependent CPV
  - Flavor violation
  - Is there a better test?
  - Model independent way of classifying CPV resonances



BACK-UP

Unbinned tests for CPV

Energy Test: $O(n^2)$, tuned on σ

$$T = \sum_{i,j>i}^{N} \frac{\psi_{ij}}{N(N-1)} + \sum_{i,j>i}^{\bar{N}} \frac{\psi_{ij}}{\bar{N}(\bar{N}-1)} - \sum_{i,j}^{N,\bar{N}} \frac{\psi_{ij}}{N\bar{N}}, \qquad \psi_{ij} \equiv \psi(d_{ij};\sigma) = e^{-d_{ij}^2/2\sigma^2}$$

CPC

CPV