Differentiable Jet Clustering

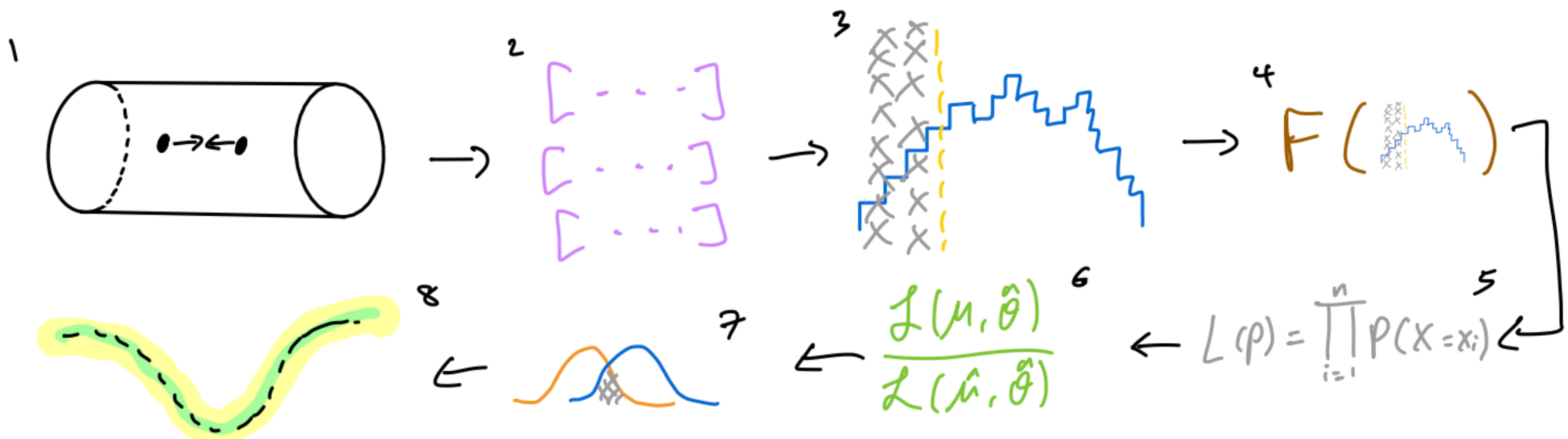
Nishank Gite (UC Berkeley)

Advisors:
Annalena Kofler (MPI-IS)
Nicole Hartman (TUM)
Lukas Heinrich (TUM)
Michael Kagan (SLAC)

10/16/2024

Why?

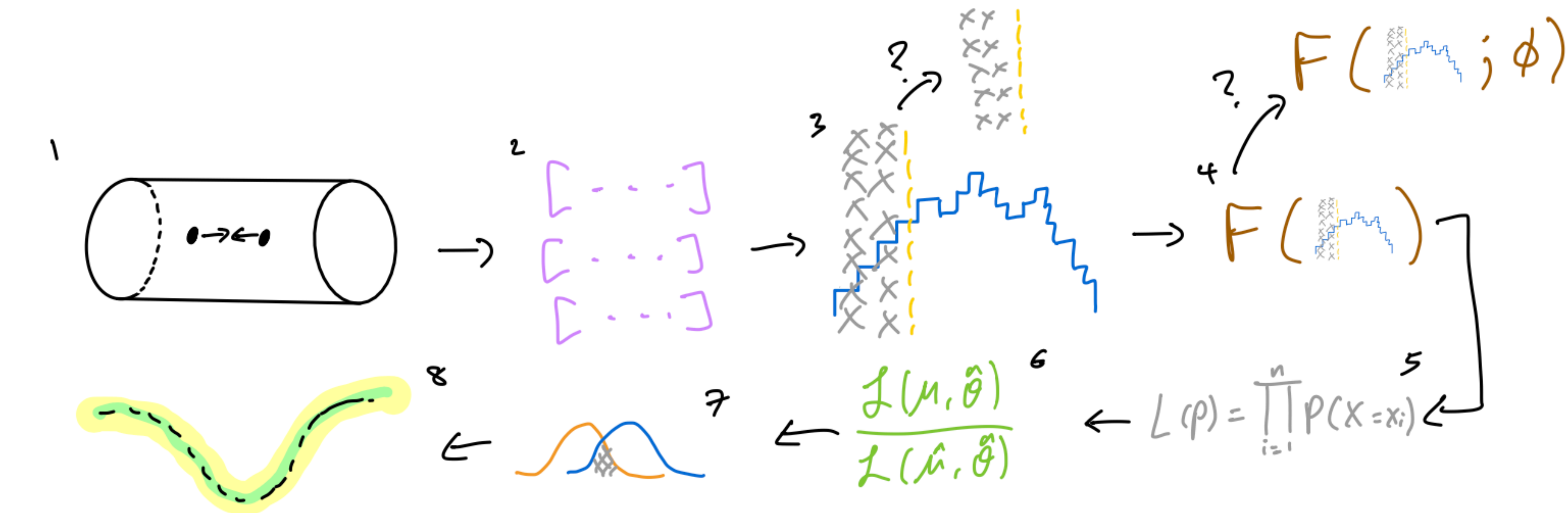
- **IRIS-HEP "Analysis Grand Challenge"**: create an end-to-end differentiable pipeline for HEP data analysis



Simulation ⟶ Data ⟶ Cuts ⟶ Observables ⟶ Model ⟶ Test Statistic ⟶ Hypothesis Test ⟶ p-vals

Why?

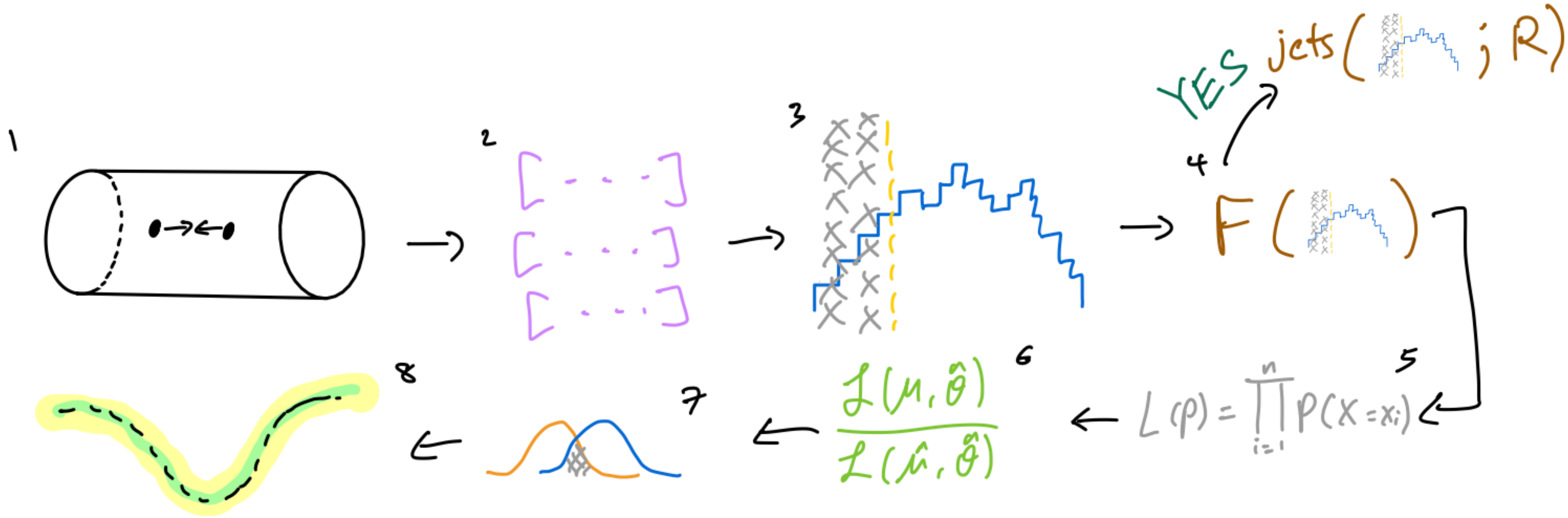
- **IRIS-HEP "Analysis Grand Challenge"**: create an end-to-end differentiable pipeline for HEP data analysis



Simulation ⟶ Data ⟶ Cuts ⟶ Observables ⟶ Model ⟶ Test Statistic ⟶ Hypothesis Test ⟶ p-vals

Why?

- **IRIS-HEP "Analysis Grand Challenge"**: create an end-to-end differentiable pipeline for HEP data analysis



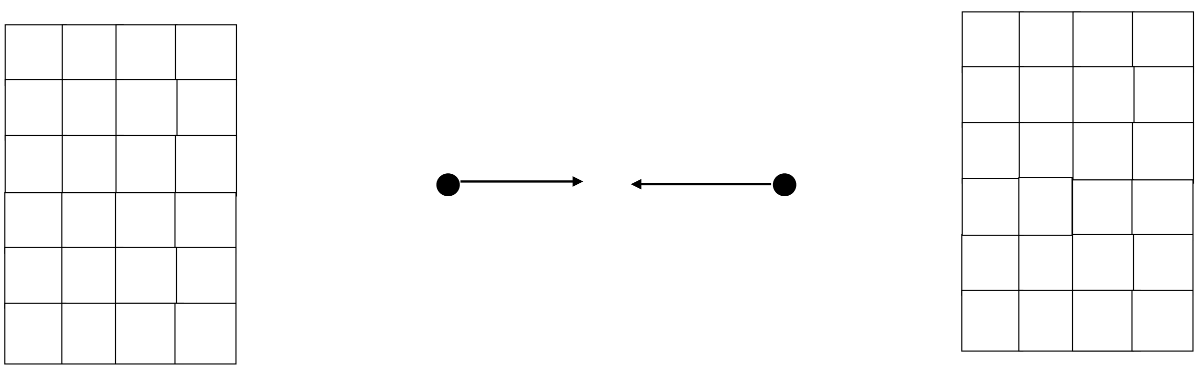
Simulation ⟶ Data ⟶ Cuts ⟶ Observables ⟶ Model ⟶ Test Statistic ⟶ Hypothesis Test ⟶ p-vals

Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?

Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
    - Particles move towards eachother

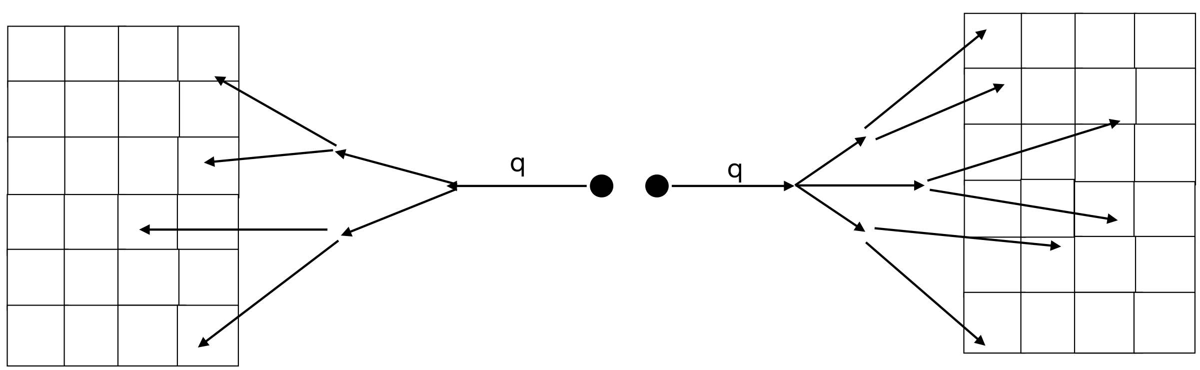


Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
    - They collide!

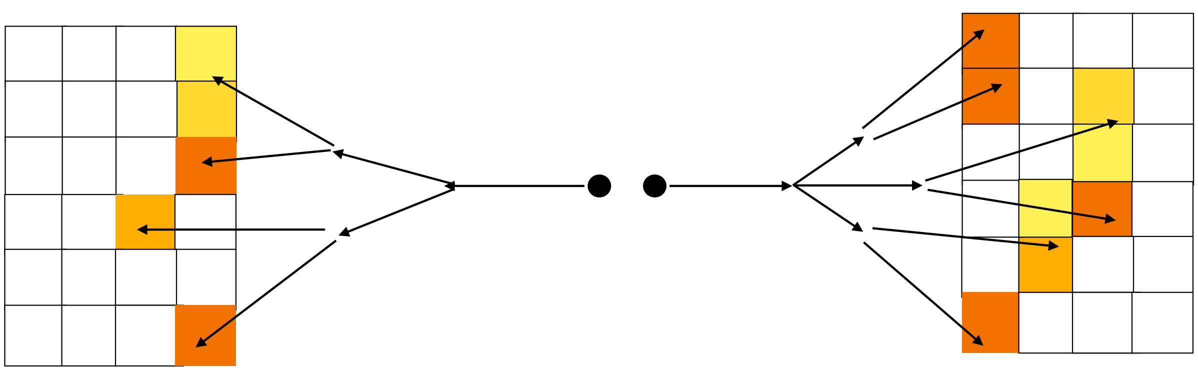
Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
    - Produces particles which can decay… then those can decay!



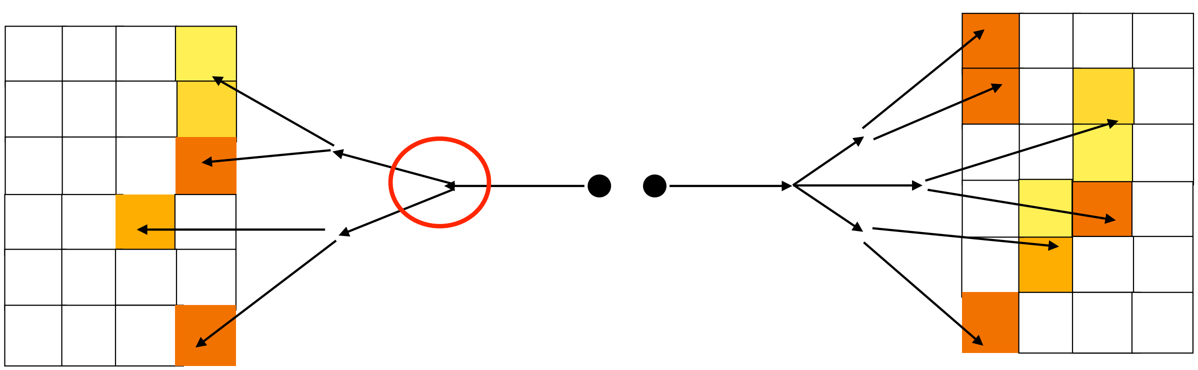
Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
    - Logged as hits in our detector



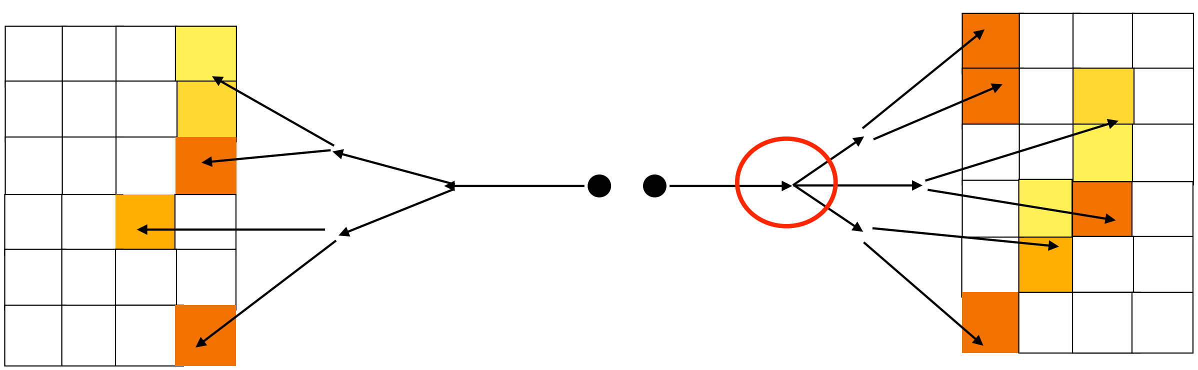
Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
    - Combine particles that decayed from the same original particle



Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
    - Combine particles that decayed from the same original particle



1

First, determine the distance between particles using the following formulas:

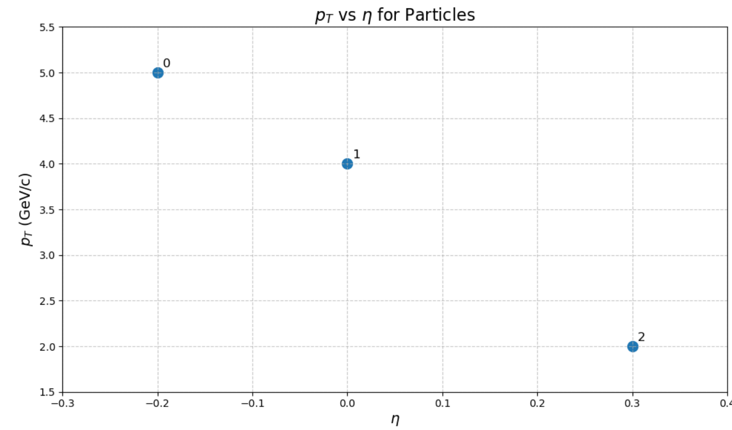
Off-Diagonal: $d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \cdot \frac{\Delta R_{ij}}{R^2}$

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$$

Diagonal: $d_{iB} = p_{T,i}^{-2}$

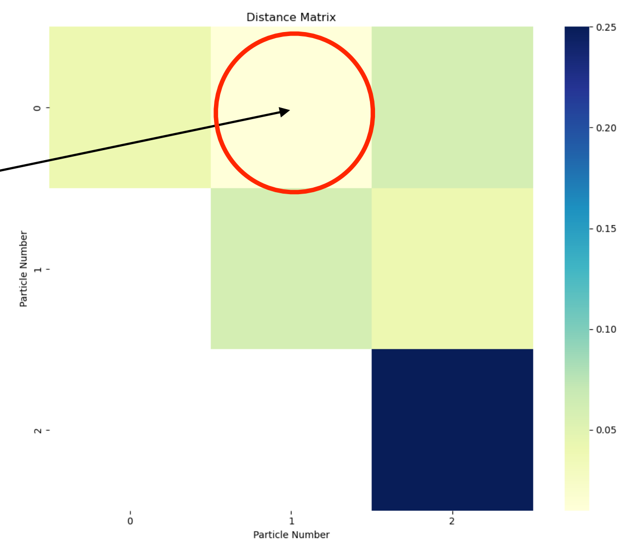
2

Consider the following 3 particles:



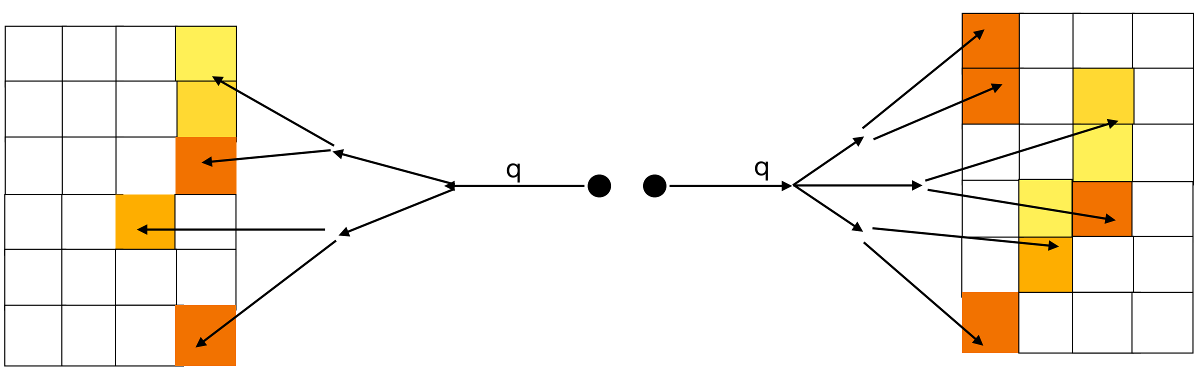
3

Merge particles that are "closest" together



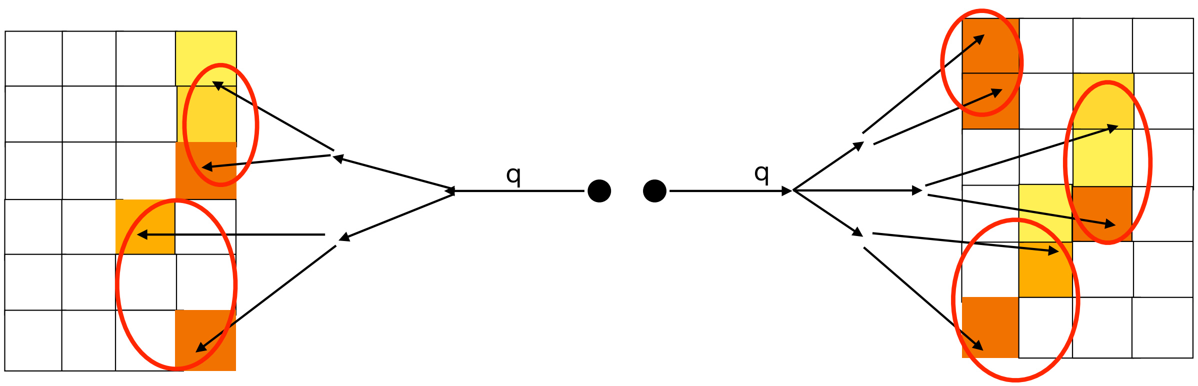
Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
    - Now let's cluster this!



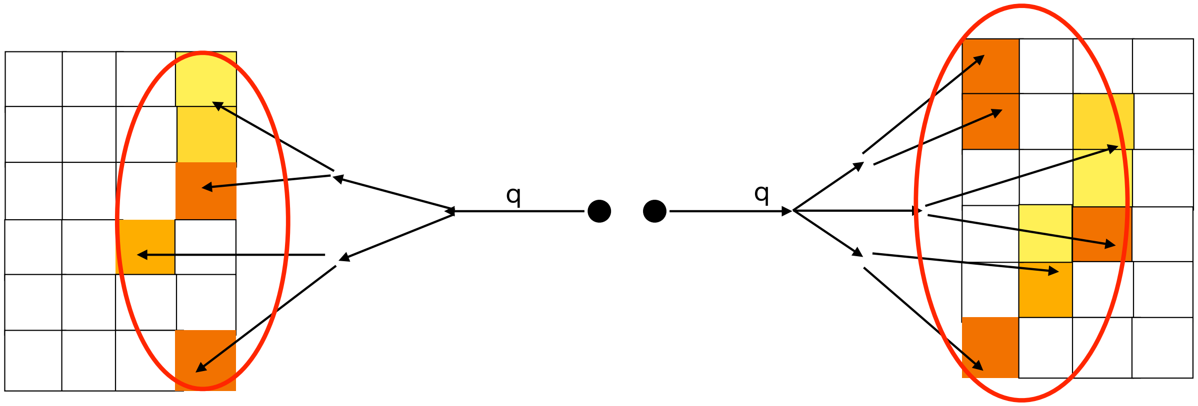
Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
    - Collimated spray of particles created by a cascade of particle decays
    - What does this look like?
    - Step-by-step the closest particles get clustered



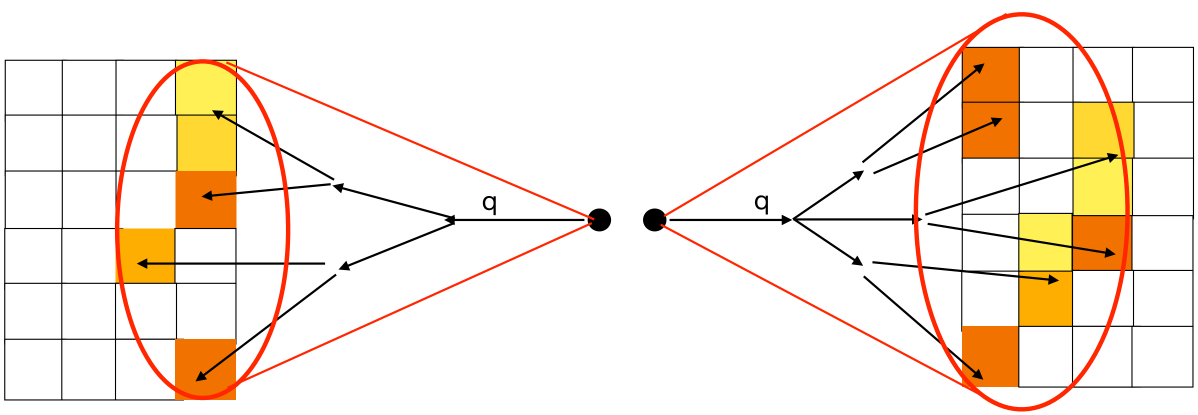
Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
  - Collimated spray of particles created by a cascade of particle decays
  - What does this look like?
  - Once this is done the clusters will then be clustered together



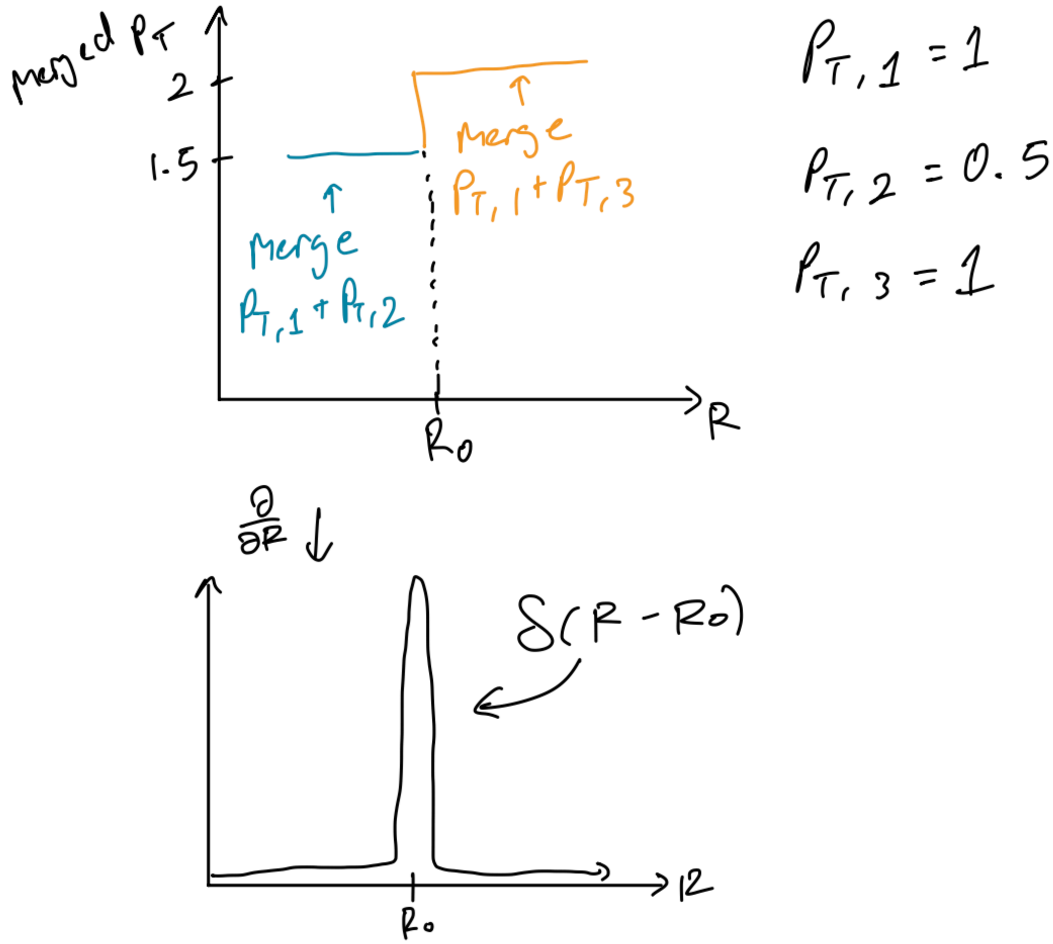
Anti-k_T Jet Clustering

- Quick Recap: What is a Jet?
    - Collimated spray of particles created by a cascade of particle decays
    - What does this look like?
    - Finally we have our jets!



Limitations?

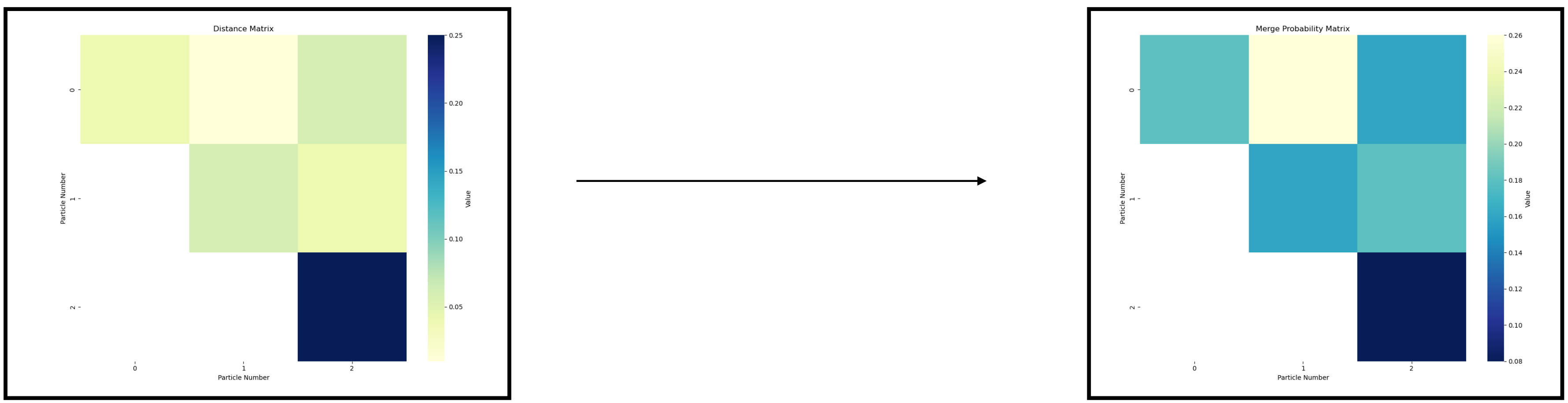
- Jet Clustering has one free parameter: Radius
    - Radius is optimized once (R = 0.7) then kept constant
- Small R changes can create **non-differentiability**

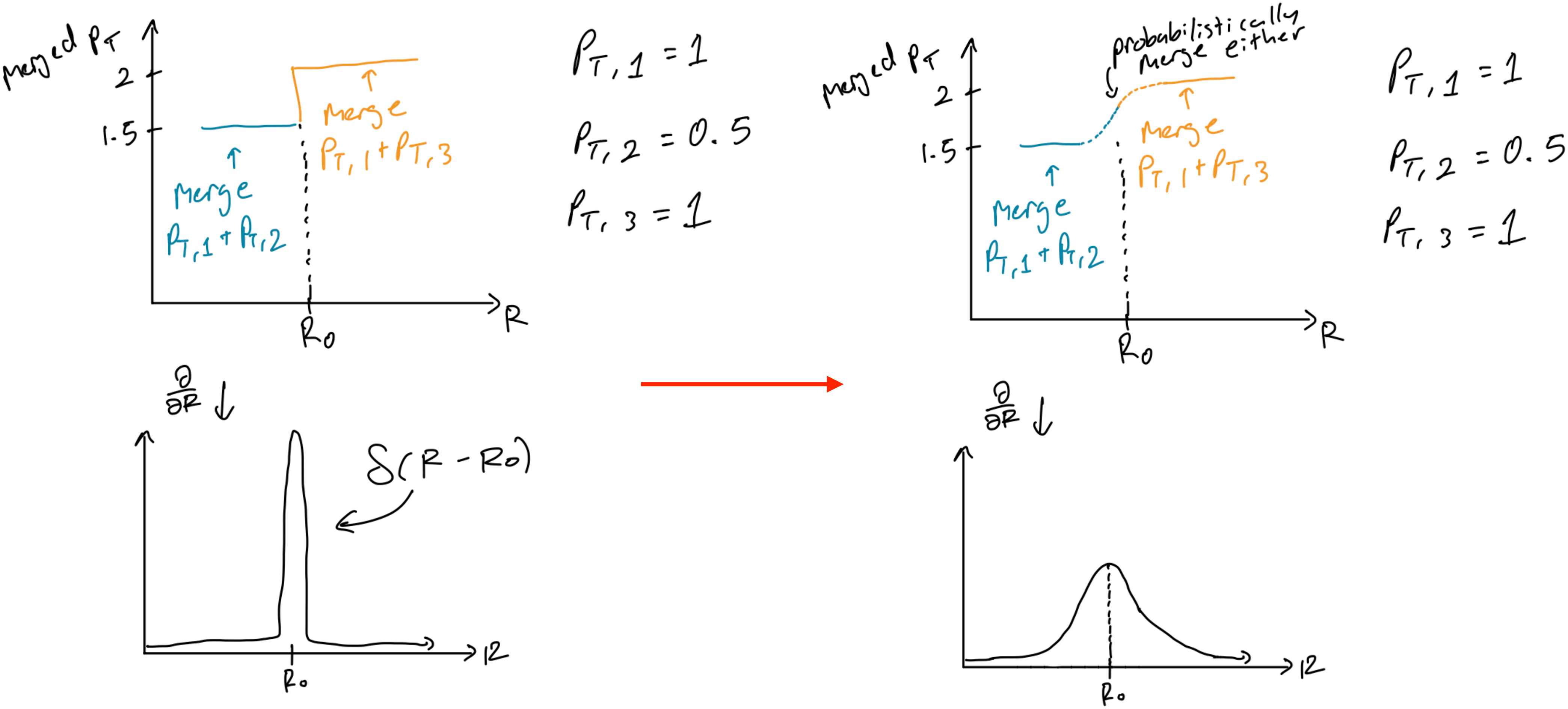


Our Fix?

- Make merge decisions probabilistic by assigning each potential merge a probability, so now our process looks like this:

- What does this look like for us?





merged p_T
2
1.5
merge $p_{T,1} + p_{T,2}$
merge $p_{T,1} + p_{T,3}$
R_0
R
$p_{T,1} = 1$
$p_{T,2} = 0.5$
$p_{T,3} = 1$
$\frac{\partial}{\partial R}$
$\delta(R - R_0)$
probabilistically merge either

Score-Based Gradient Estimator

- We are still making a random decision $\longrightarrow$ Not Directly Differentiable
- **Score-Estimator**: Sample-based approach to computing derivatives:

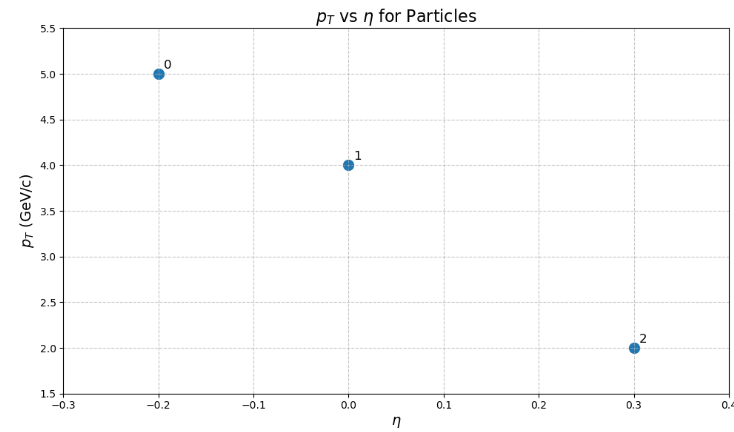
$$\frac{\partial}{\partial R}\mathbb{E}[L(R)] = \mathbb{E}\left[L(R)\frac{\partial}{\partial R}\log p(n \mid R)\right]$$

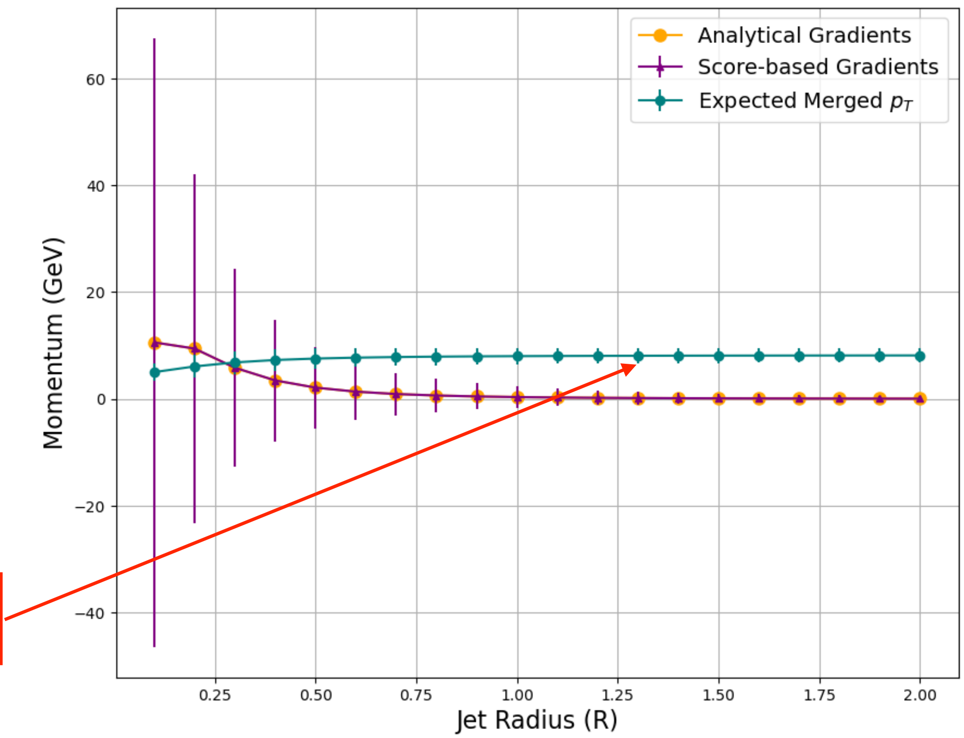
How "well" jet is clustered

Merge Probability

3 Particle Toy Example (2-D)

- 2-D Toy Example: 3 particles with known p_T and η
  - Aim: Study how **first** merge behavior varies with radius
- Use merged p_T (the combined p_T of the particles we merge) as a proxy for $L(R)$



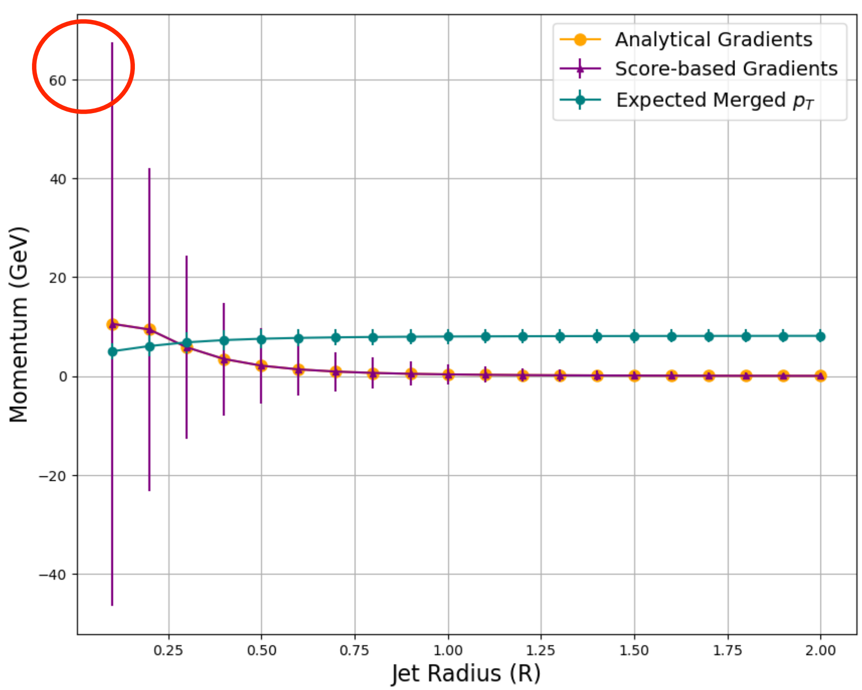


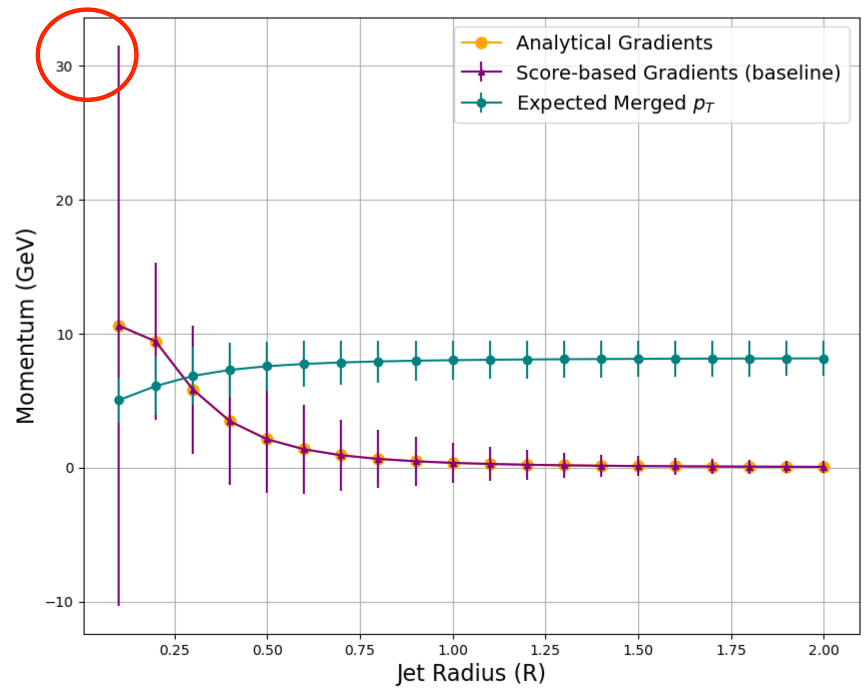
Merged p_T plateaus for large radius

Baseline-Variance Reduction

- Reduce variance by shifting the effective loss to 0:

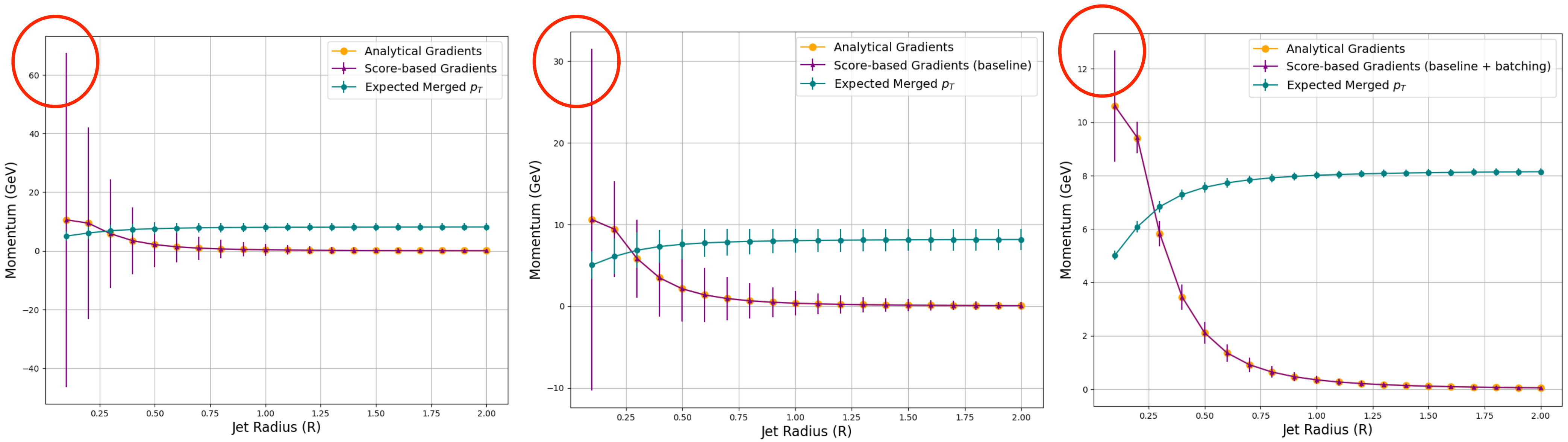
$$\nabla_R \mathbb{E}[L(R)] = \mathbb{E}\left[(L(R) - b) \cdot \nabla_R \log P(\text{decision} \mid R)\right]$$





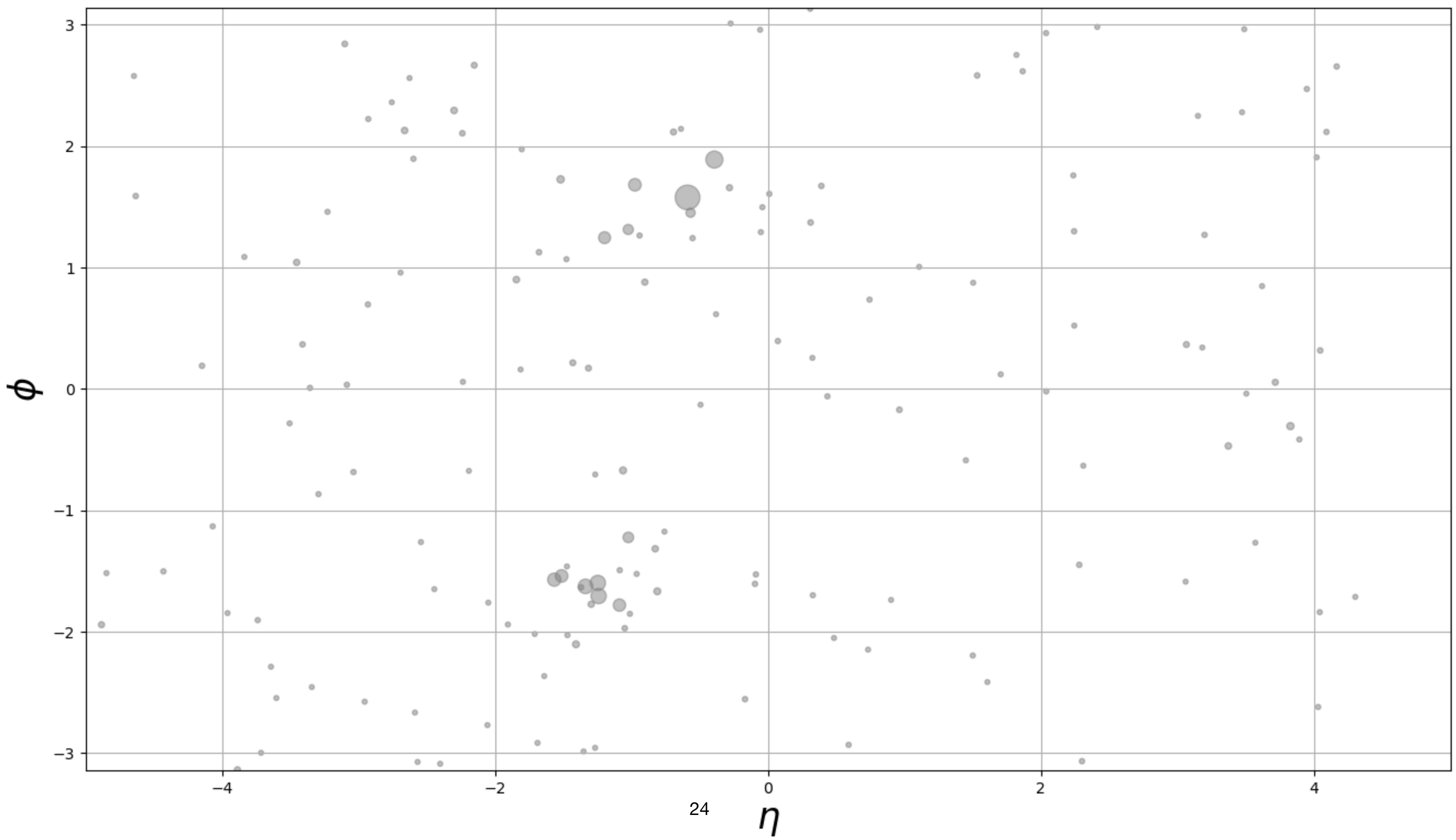
Variance Reduction via Batching

Averaging over multiple samples reduces random fluctuations



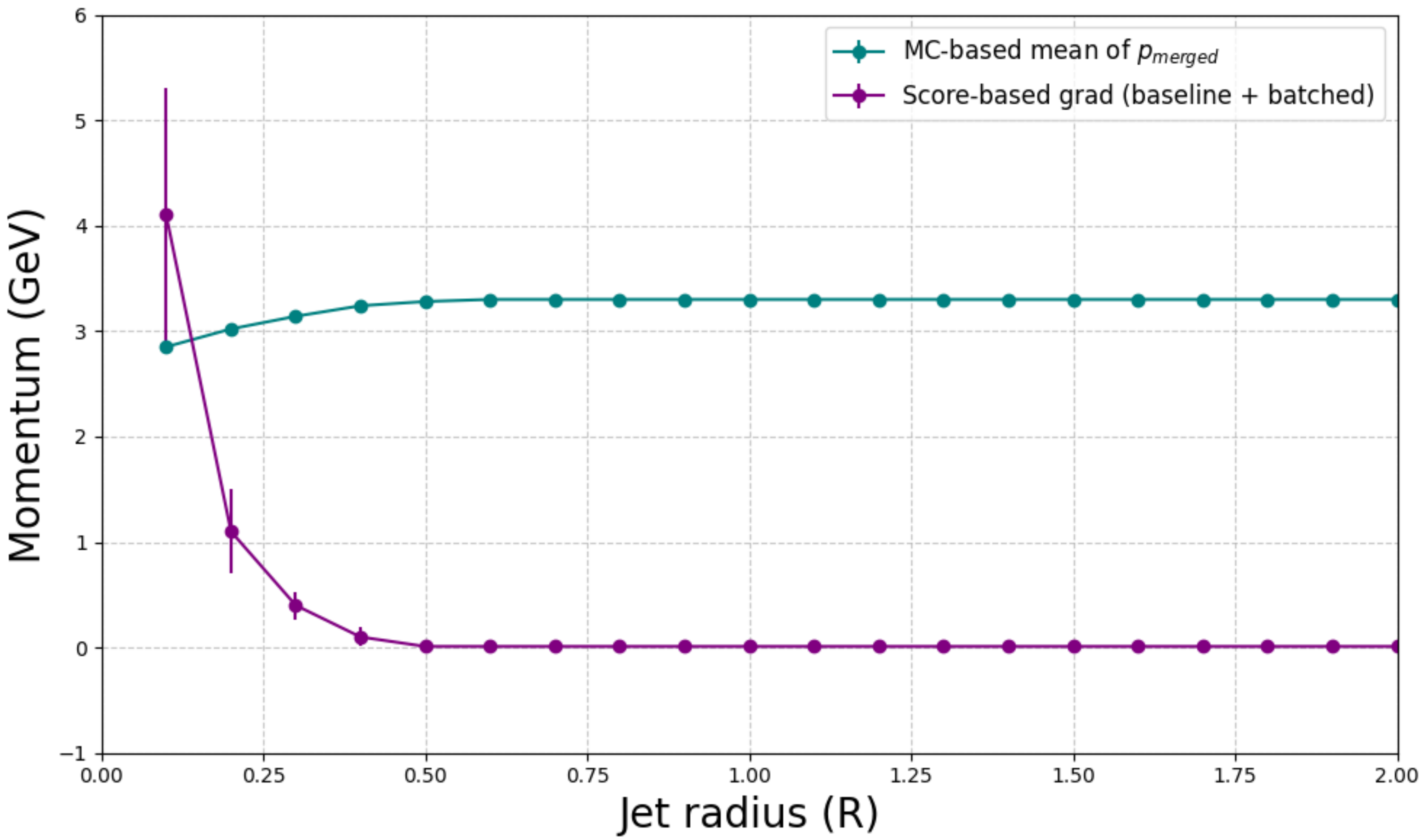
Real Particle Case

PYTHIA Generated Particles

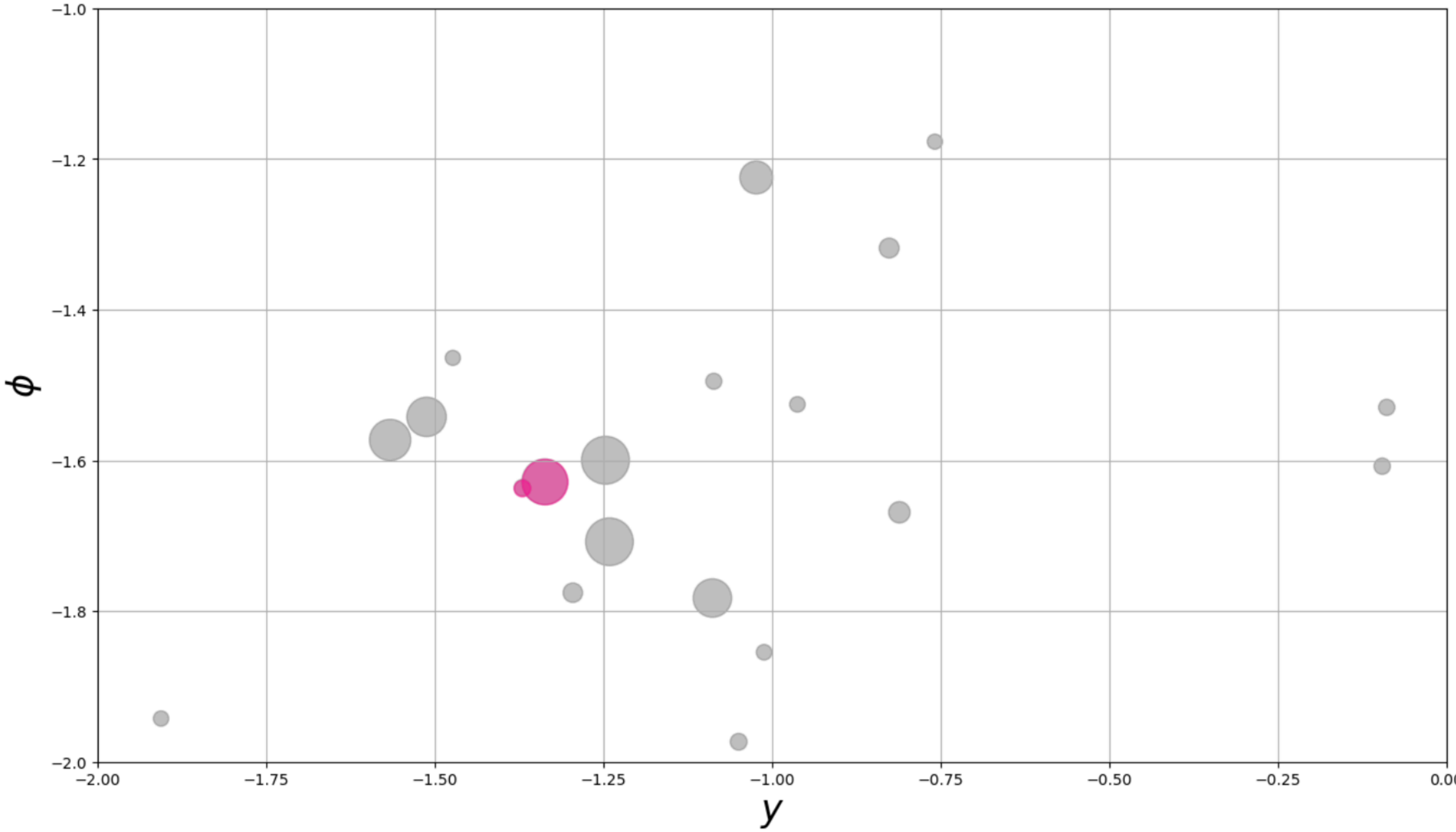


plot by Nicole Hartman

Gradients for Real Particles

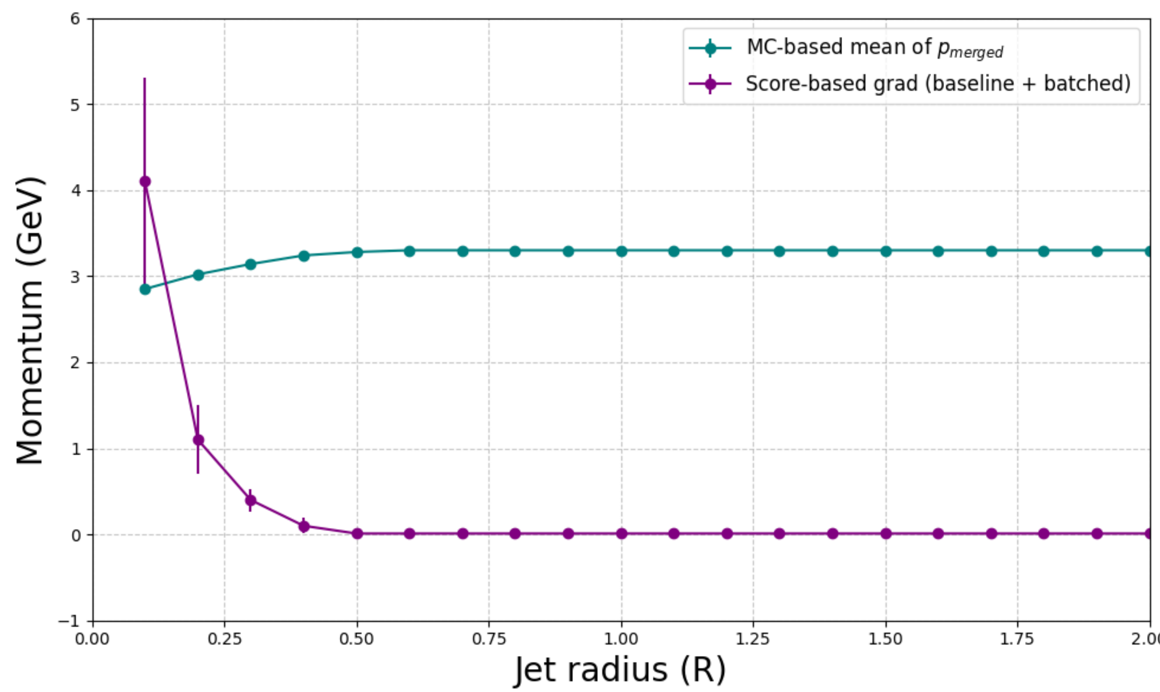


169: pt=0.44,eta=-1.37, phi=-1.64
44: pt=2.85,eta=-1.34, phi=-1.63



Summary

- Created differentiable anti-k_t
- Demonstrated effectiveness:
  - 3-particle toy example
  - Real Data
- **End-Goal**: Differentiable Jet Clustering is one part of optimization pipeline!



References

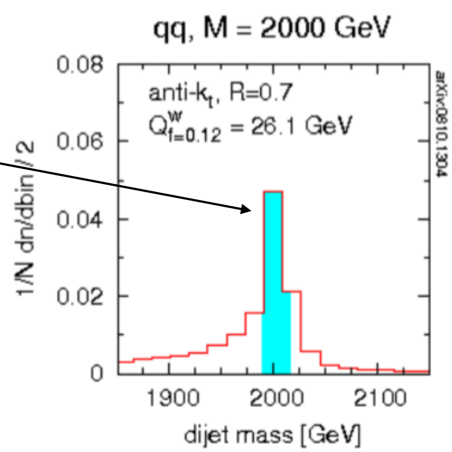
1. Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, "The anti-k_t jet clustering algorithm", *Journal of High Energy Physics*, JHEP 04 (2008) 063, DOI: 10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
2. Biswarup Mukhopadhyaya, Tousik Samui, and Ritesh K. Singh, "Dynamic Radius Jet Clustering Algorithm", *Journal of High Energy Physics*, JHEP 04 (2023) 074, DOI: 10.1007/JHEP04(2023)019, arXiv:2301.13074.
3. Dilani Kahawala, David Krohn, and Matthew D. Schwartz, "Jet Sampling: Improving Event Reconstruction through Multiple Interpretations", *Journal of High Energy Physics*, JHEP 06 (2013) 006, DOI: 10.1007/JHEP04(2023)006, arXiv:1304.2394.
4. John Schulman, Nicolas Heess, Theophane Weber, and Pieter Abbeel, "Gradient Estimation Using Stochastic Computation Graphs", *Proceedings of the 28th International Conference on Neural Information Processing Systems* (NIPS 2015), DOI: 10.48550/arXiv.1506.05254, arXiv:1506.05254.

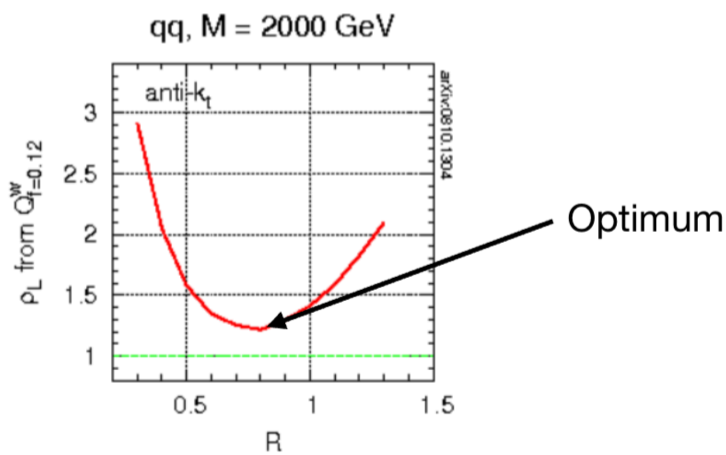
Next Steps

- Use **real** loss function: invariant mass of the 2 leading p_T jets (dijet)

Example:

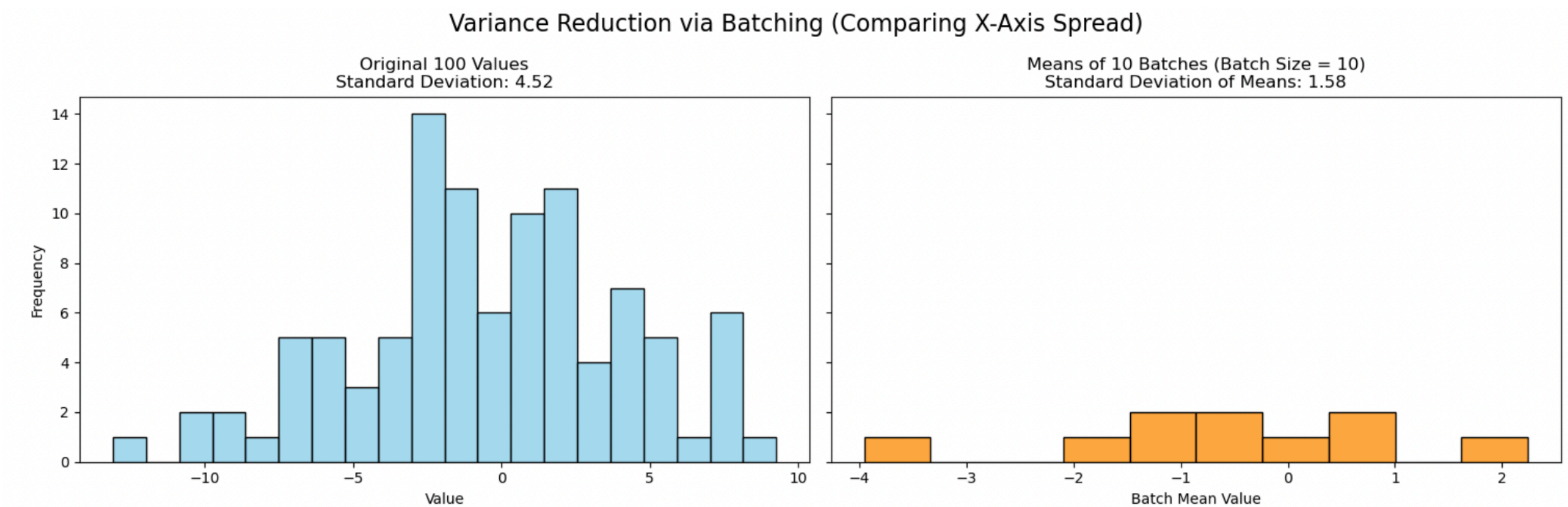
Resonance decay such as $pp \to Z'/H \to jj$, a correctly reconstructed dijet mass should peak around the mass of the resonance particle





Variance Reduction via Batching

- Average smaller groups of the full sample to minimize variance further:



- Distribution spread reduced!