# Particle Cloud Generation with Message Passing GANs

**Raghav Kansal**[1], Javier Duarte[1], Hao Su[1], Breno Orzari[2], Thiago Tomei[2], Maurizio Pierini[3], Mary Touranakou[3,4], Jean-Roch Vlimant[5], Dimitrios Gunopulos[4]

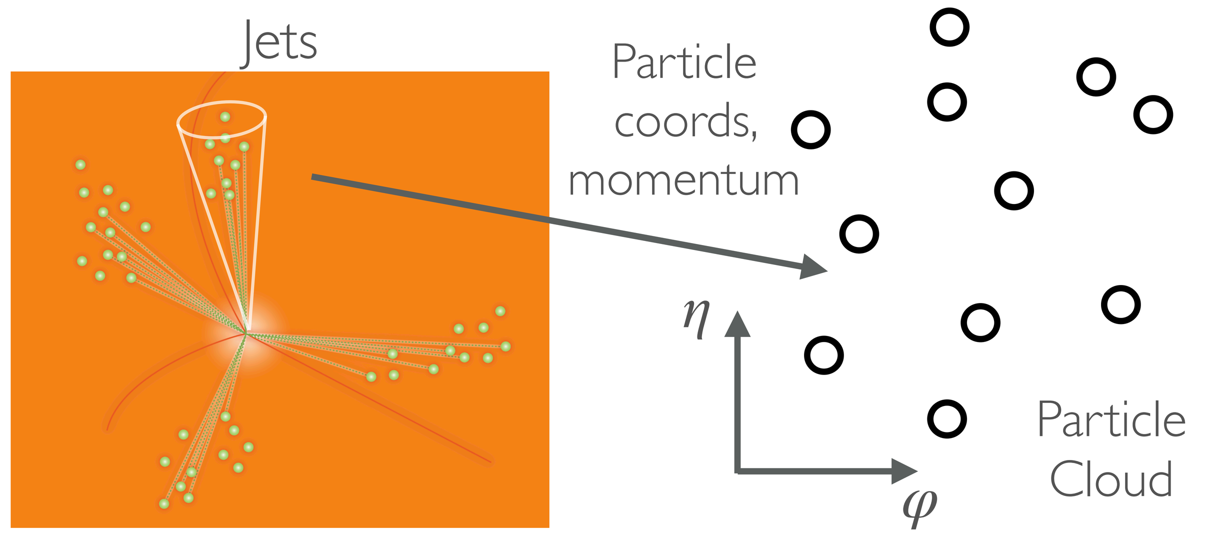
[1]UC San Diego, [2]Universidade Estadual Paulista, [3]CERN, [4]National and Kapodistrian University of Athens, [5]California Institute of Technology

## ML for CERN LHC Simulations

- Traditional computing techniques in high energy physics (HEP) can't keep up with data needs at the Large Hadron Collider (LHC)
- Significant opportunity to speed up tasks such as simulation using ML
- In this work, we:
  1. Release a new HEP dataset and package (**JetNet**) to facilitate research in this area,
  2. Test existing point cloud GANs on JetNet
  3. Develop a new physics-informed GAN which is significantly more performant

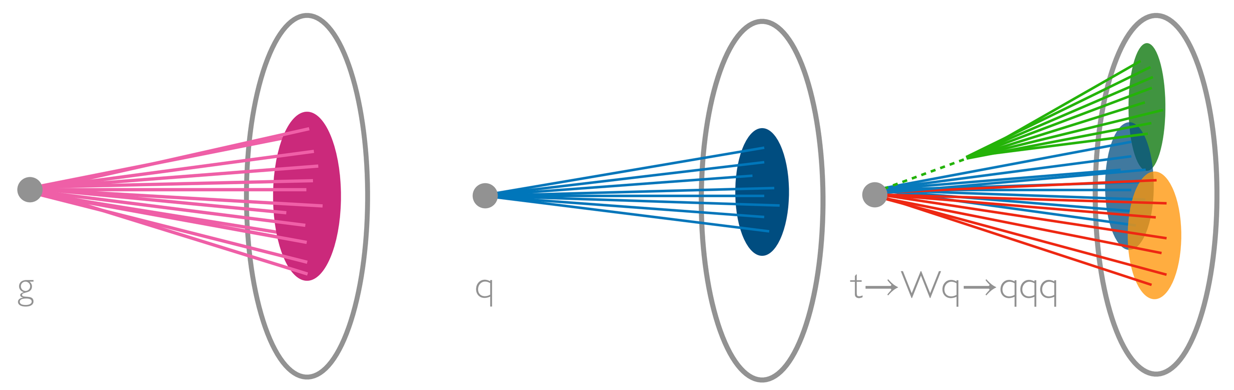
## Jets

- Jets, collimated sprays of high energy particles, are ubiquitous at the LHC
- Natural representation as a "particle cloud"
  - Particle angular coordinates $(\eta, \phi)$ and transverse momenta $p_T$ as node features

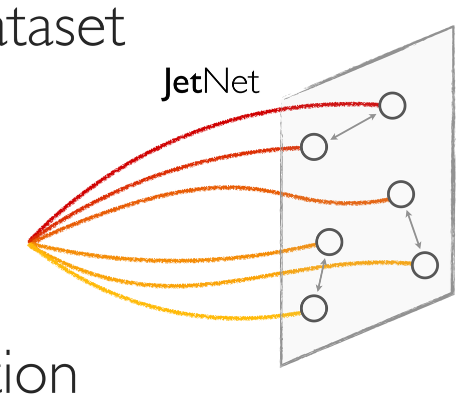



## Dataset: JetNet

- JetNet [1]: high $p_T$ jets of max 30 particles
- 3 classes/jet types:
  - Gluon: simple baseline generation test
  - Lighter quarks: fewer particles; test handling of variable-sized clouds
  - Top quark: complex topology




- We test r-GAN (fully-connected), GraphCNN-GAN, and TreeGAN generators on JetNet
- Results inadequate for physics applications
- Instead, our new MPGAN approach is significantly more performant
- We invite researchers to improve on this, and we release this dataset + **JetNet** package [2] with:
  - Accessible interfaces for ML+HEP datasets
  - Implementations for evaluation metrics
  - Conveniences to facilitate research in this area



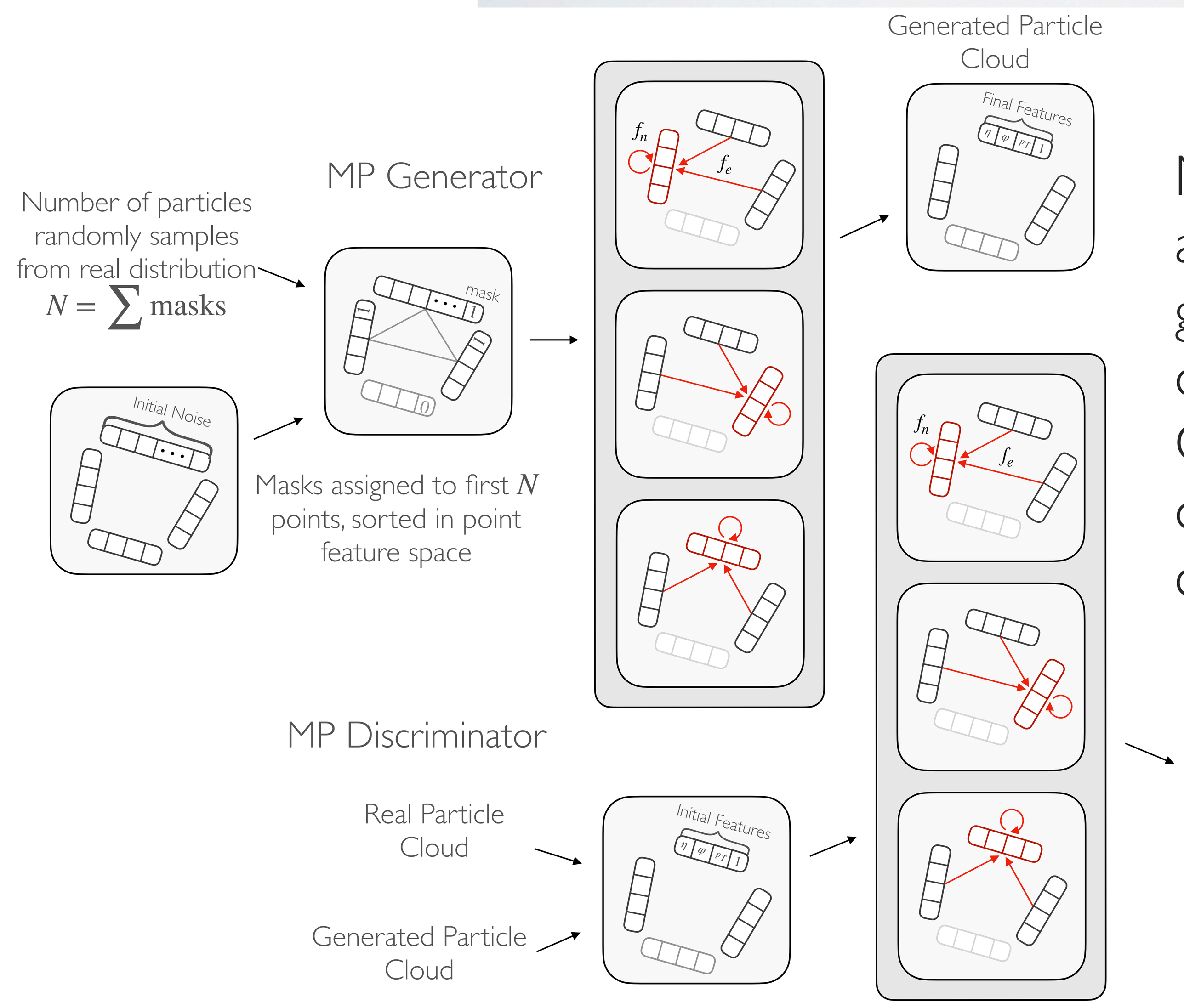

## Evaluation

- Want to evaluate **quantitatively** and in a **standardised** way key aspects of simulations
- We develop four physics- and computer-vision-inspired metrics:
  1. Minimum matching distance (MMD)
  2. Coverage
  3. Fréchet ParticleNet Distance (FPND)
  4. 1-Wasserstein (W1) distances between particle- and jet-level feature distributions, with bootstrapped real baselines

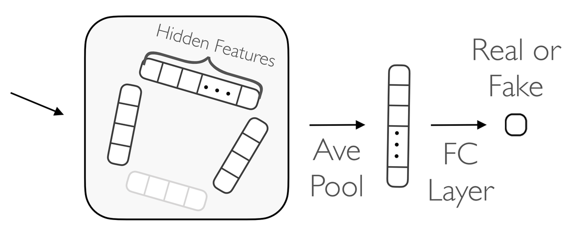
| Simulations Aspect | MMD | COV | FPND | W1 |
|---|---|---|---|---|
| Quality | ✓ | | ✓ | ✓ |
| Diversity | | ✓ | ✓ | ✓ |
| Physics Performance | | | | ✓ |

- We find them to be complementary:
  - MMD and coverage are focused tests of quality and diversity
  - FPND is the most discriminating, good for model selection
  - W1 and comparing with bootstrapped baselines gives interpretable validation

## Approach: MPGAN

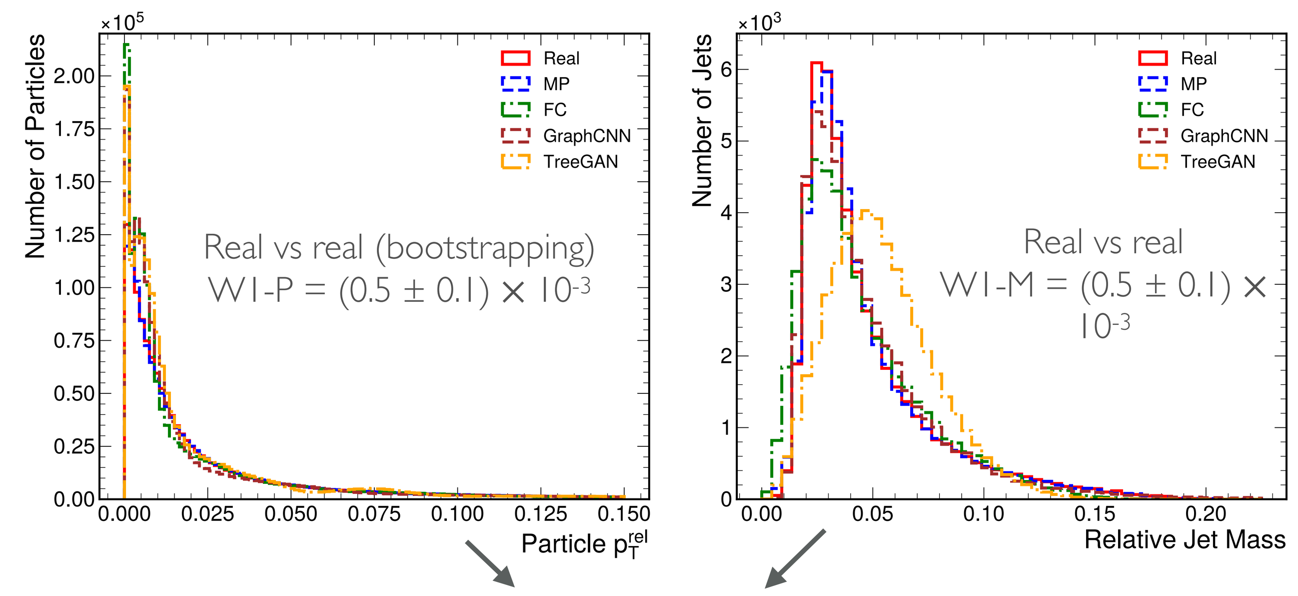



Message passing architecture for the generator and discriminator of a GAN to operate on the particle clouds



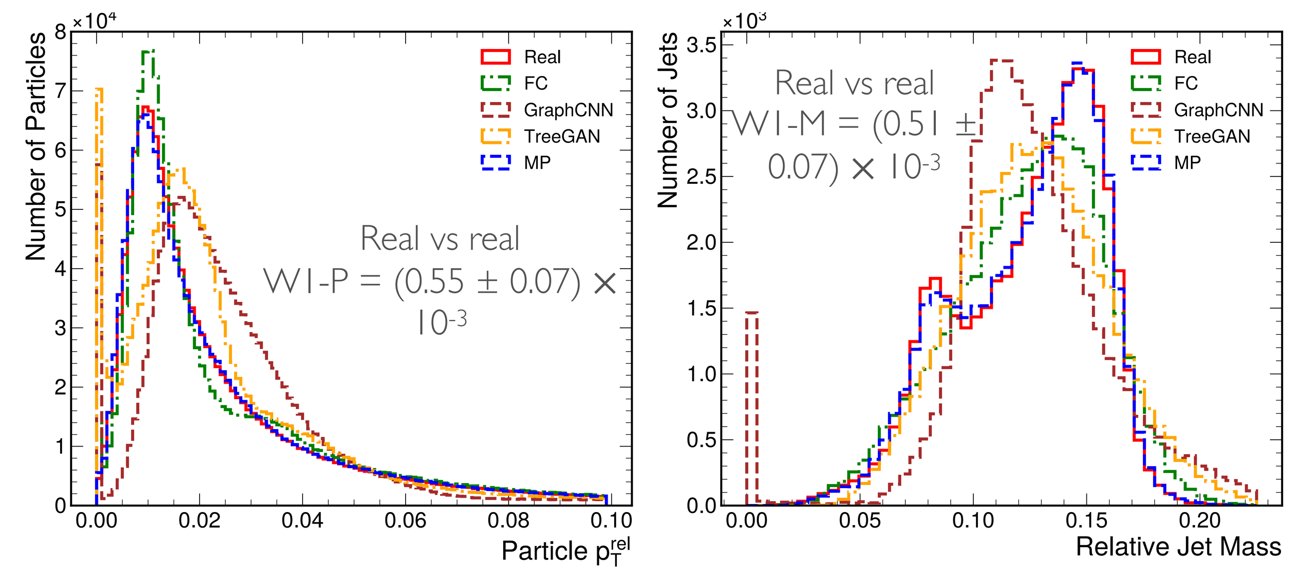

## Results

- Sample feature distributions, with our MPGAN compared to existing generators + PointNet discriminators for **light quark** jets:




| Generator | W1-P ($10^{-3}$) | W1-M ($10^{-3}$) | FPND | Coverage | MMD |
|---|---|---|---|---|---|
| Fully Connected (FC) | 4.5 ± 0.4 | 3.1 ± 0.2 | 17 | 0.37 | 0.028 |
| GraphCNN | 5.2 ± 0.5 | 4 ± 1 | 316 | 0.37 | 0.031 |
| TreeGAN | 5.7 ± 0.5 | 10.1 ± 0.1 | 11 | 0.47 | 0.031 |
| MP | 4.9 ± 0.5 | 0.6 ± 0.2 | 0.35 | 0.50 | 0.026 |

- And **top quark** jets:




| Generator | W1-P ($10^{-3}$) | W1-M ($10^{-3}$) | FPND | Coverage | MMD |
|---|---|---|---|---|---|
| Fully Connected (FC) | 1.6 ± 0.4 | 2.7 ± 0.1 | 3.9 | 0.56 | 0.075 |
| GraphCNN | 30 ± 10 | 11.3 ± 0.9 | 30k | 0.39 | 0.085 |
| TreeGAN | 9.1 ± 0.3 | 5.19 ± 0.08 | 17 | 0.53 | 0.079 |
| MPGAN | 2.3 ± 0.3 | 0.6 ± 0.2 | 0.37 | 0.57 | 0.071 |

- MPGAN best performing on nearly every metric
- Significantly outperforms on high level (jet kinematics, substructure) feature metrics i.e. W1-M, FPND…
- Mass and other substructure $W_1$ scores are within error of the real vs real baseline ⇒ learning jet substructure correctly
- Only one to learn bimodal top jet distributions

## Summary/Outlook

- We advocate for physics-motivated particle cloud representations for HEP data
- We propose four physics- and computer-vision-inspired metrics for evaluating particle cloud generative models
- Our MPGAN outperforms existing point cloud GANs on nearly all metrics
- Next: **conditional GAN**, **scaling up** to larger clouds, **dataset development**
- Contact us at rkansal@ucsd.edu if you're interested in collaborating!

Link to code and paper

[1] https://zenodo.org/record/5502543
[2] https://github.com/jet-net/JetNet