Evaluating Generative Adversarial Networks for particle hit generation in a cylindrical drift chamber

Authors: Irene Andreou[1], Noam Mouelle[2]

1: Imperial College London, 2: University of Cambridge (work carried out while at Imperial College London)

Motivation

Precisely observing the properties of known particles is one way of searching for New Physics. This approach requires precision experiments, like COMET, to produce a huge number of events. To prepare for such experiments, it is desirable to obtain real-size mock datasets, which Monte Carlo simulations, like GEANT4, cannot produce within reasonable times. Statistical models must be used instead. The COMET collaboration, for instance, will use Generative Adversarial Networks (GANs) to generate a real-size fake dataset equivalent to 10^{19} events (Phase-I). GAN are unsupervised learners, meaning that standardized GAN evaluation metrics are critical to their development in HEP, and to support searches for New Physics with precision experiment.

The COMET experiment

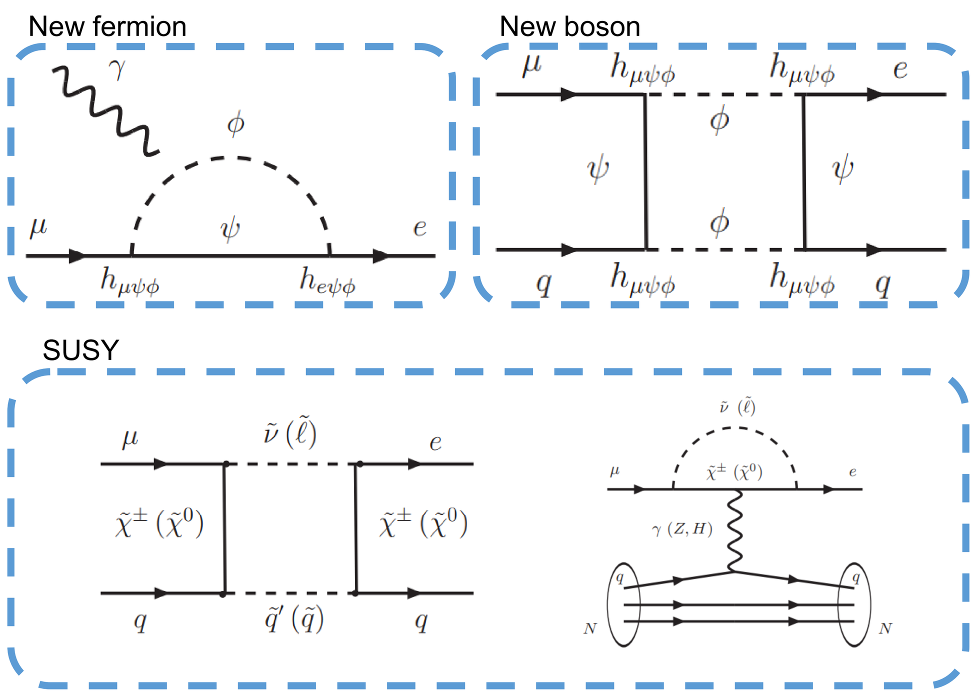
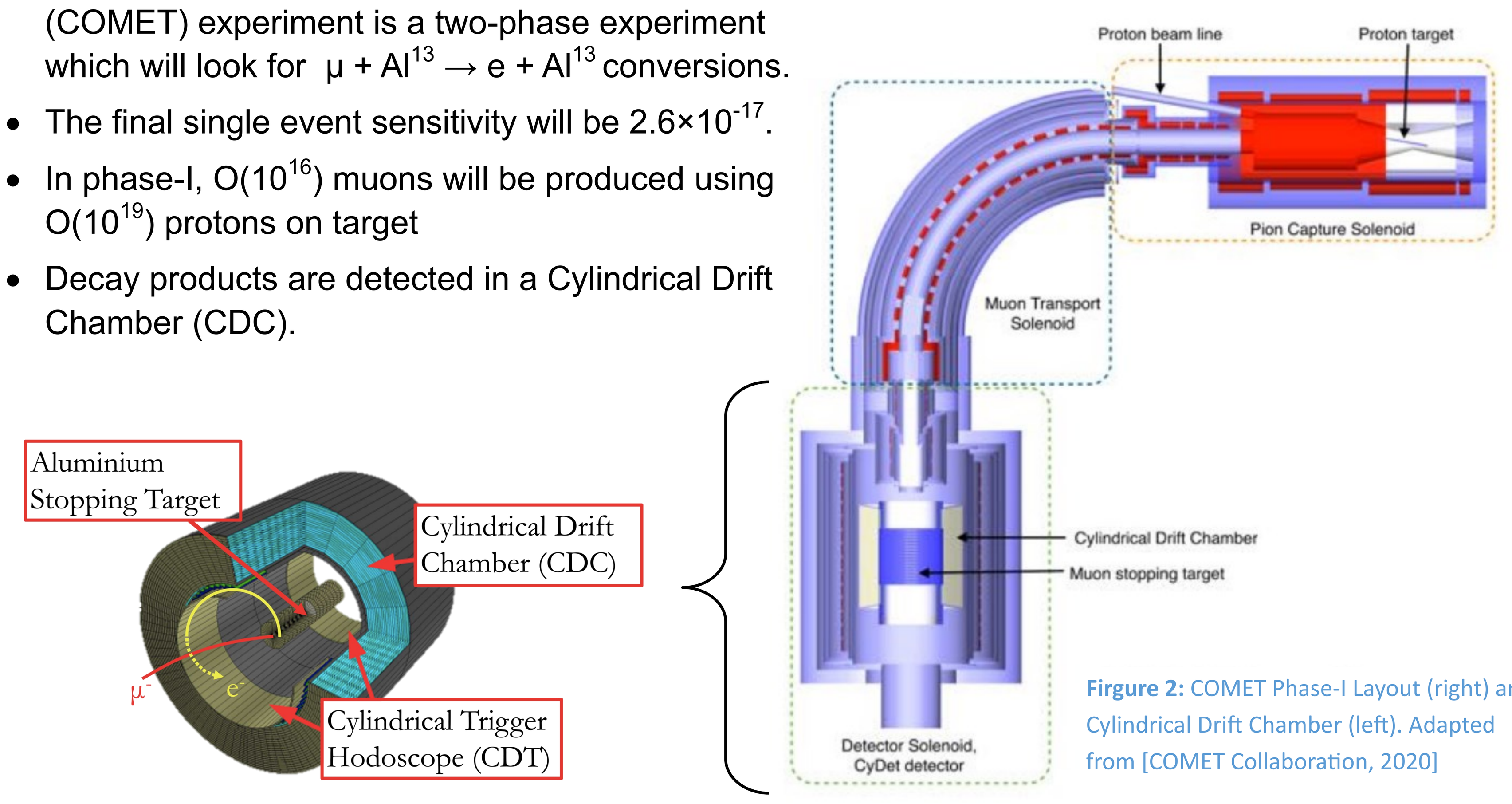


Figure 1: Possible μ→e conversion channels. Adapted from [COMET Collaboration, 2020]

- In the Standard Model, lepton flavour conservation results from an accidental symmetry.
- Charged Lepton Flavour Violation would be an indicator of new physics
- Neutrinoless muon to electron conversion in muonic atoms (μ+N→e+N) is a CLFV process with a clean signature: a single 105 MeV electron.

- The COherent Muon to Electron Transition (COMET) experiment is a two-phase experiment which will look for $\mu + Al^{13} \rightarrow e + Al^{13}$ conversions.
- The final single event sensitivity will be 2.6×10^{-17}.
- In phase-I, O(10^{16}) muons will be produced using O(10^{19}) protons on target
- Decay products are detected in a Cylindrical Drift Chamber (CDC).



Firgure 2: COMET Phase-I Layout (right) and Cylindrical Drift Chamber (left). Adapted from [COMET Collaboration, 2020]

Evaluation method

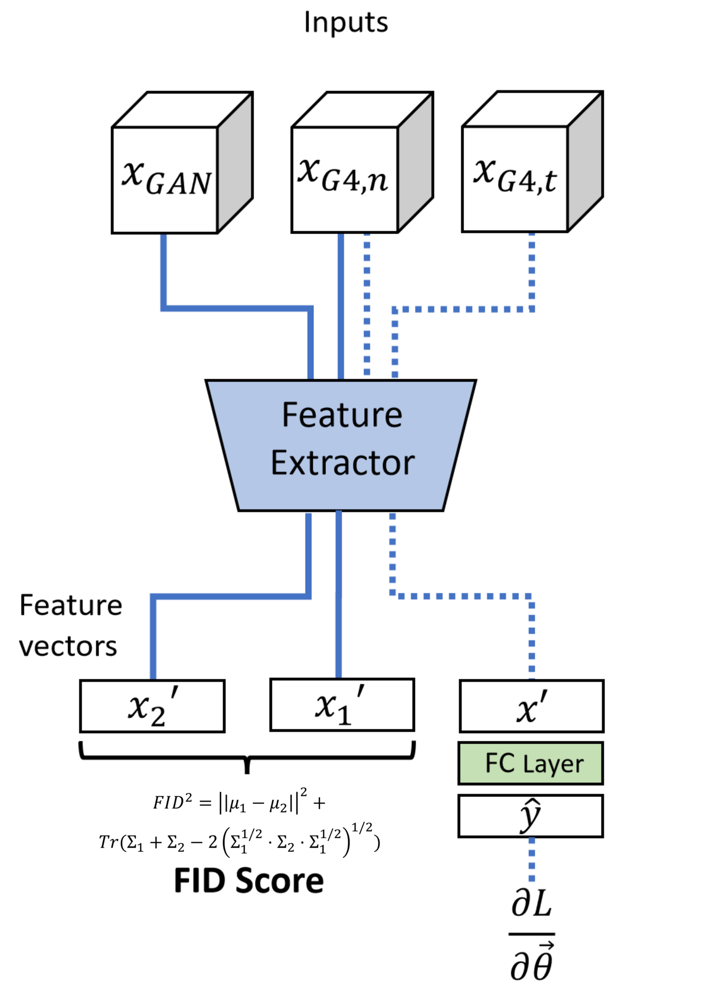


Figure 5: Training and using a feature extractor. During the training of the feature extractor, the dashed path is used. During FID computation, the solid path is used. $x_{G4,n}$ and $x_{G4,t}$ are our two classes: noise and track sequences.

- In computer vision, the standard GAN metric is Fréchet Inception Distance (FID):

$$FID^2 = ||\mu_1 - \mu_2||^2 + Tr(\Sigma_1 + \Sigma_2 - 2\left(\Sigma_1^{1/2}\cdot\Sigma_2\cdot\Sigma_1^{1/2}\right)^{1/2})$$

- We use three deep CNNs as feature extractors:
  - Inception-v3 (Iv3) [Szegedy, Christian and Vanhoucke, 2015], a 2D CNN pre-trained on ImageNet;
  - A Fine-tuned Iv3 (FTIv3), a 2D CNN pretrained on ImageNet and fine-tuned to classifiy noise vs track hit sequences;
  - A 3D CNN, only trained on MC hit sequences, for noise vs track classification.
- After training the extractor on MC noise and track sequences, we measure the FID between the distributions of feature vectors produced by GAN and MC –generated noise sequences.

- We wish to compare the space-time structure of MC and GAN hit sequences.
- MC and GAN samples are passed into the feature extractors as 3D (2 space dim, 1 time dim) images.
- For Iv3 and FTIv3, we take the 2D projections of the 3D images (for FTIv3, the projections are combined together using a Conv2D layer).

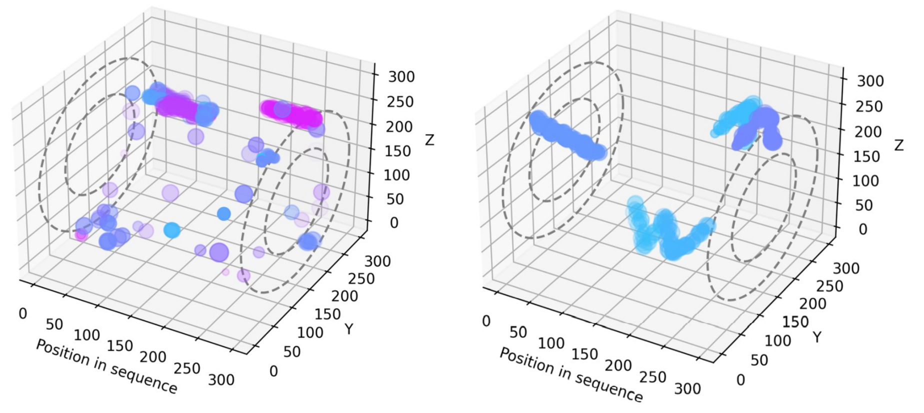


Figure 6: 3D space-time images of the CDC hit sequences. Left: noise-like hits. Right: hits belonging to reconstructible tracks.

Particle hits in the Cylindrical Drift Chamber

- 1 ms of detector data costs 30 weeks of computation using MC simulations.
- Most hits in the CDC are noise-like, while the number of hits in reconstructible tracks is relatively small (3:1 ratio).
- Definition of reconstructible track: p > 50MeV/c and a number of hits ≥ 4.
- **Noise-like hit generation is delegated to a Generative Adversarial Network (GAN).**

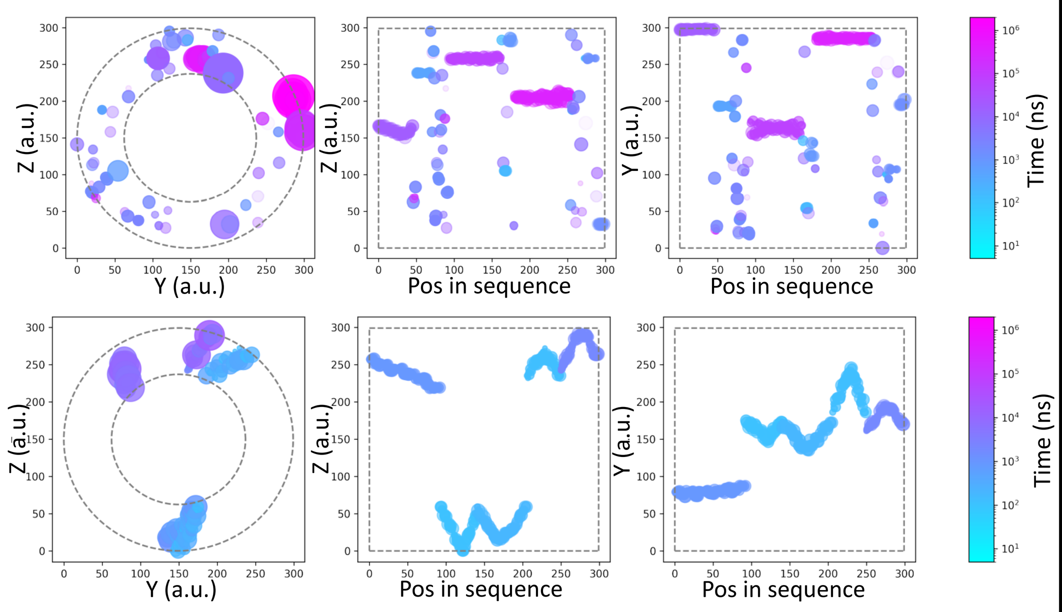


Figure 3: Examples of hit sequences labelled as noise-like (top) and as belonging to reconstructible tracks (bottom). Three projection planes of the 3D image of the CDC.

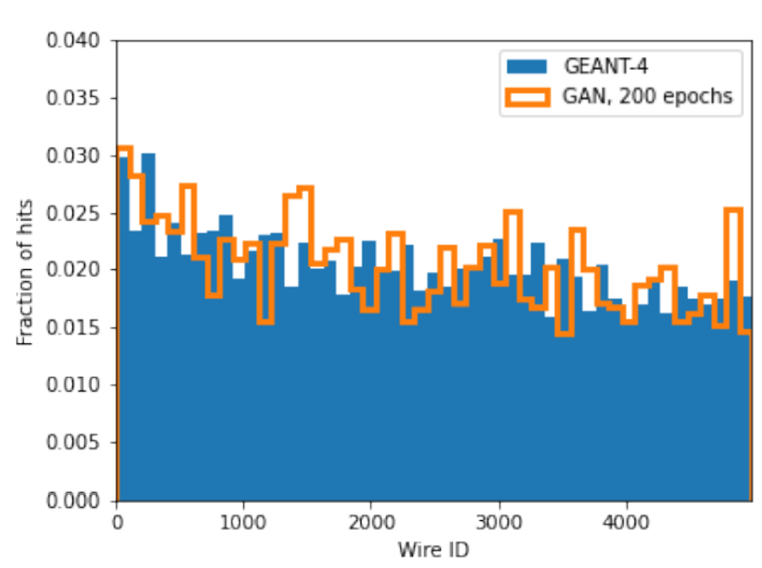


Figure 4: GAN and Geant4 (MC) generated distributions for one particle-level feature: wire ID.

- The GAN generates sequences of noise-like hits. Each hit has 4 features:
  - ⇒ Energy deposit (MeV)
  - ⇒ Hit time (ns)
  - ⇒ Distance of closest approach to the CDC wire (mm)
  - ⇒ Wire ID
- The distribution of particle-level features is well reproduced by the GAN. **What about higher level, space-time features?**

Results

- Two generative models were compared:
  - ⇒ GAN
  - ⇒ GAN with self-attention layers.
- We introduced relative FID to compare scores for different extractors.

- Effectively unbiased FIDs for Iv3 are obtained by extrapolating to an infinite number of samples
- Distributions in feature space can be visualized using t-SNE or UMAP

Feature extractor used	Inception v3			Fine-tuned Inception v3	3D CNN
Projection	**Y-Z**	**T-Y**	**T-Z**	**N/A**	**N/A**
FID, Geant4-Geant4	4.6	6.4	6.4	0.64	8,408
FID, GAN-Geant4	69.5	95.5	92.8	17.8	54,678
Relative FID, GAN-Geant4	**15.1**	**14.9**	**14.5**	**29.7**	**6.5**
FID, GAN+Self-Attention -Geant4	37.3	72.1	71.3	4.5	15,150
Relative FID, GAN+Self-Attention-Geant4	**8.1**	**11.3**	**11.1**	**7.1**	**1.8**

Table I: FID values. Relative FID corresponds to the GAN-Gean4 distance over the Geant4-Geant4 distance.

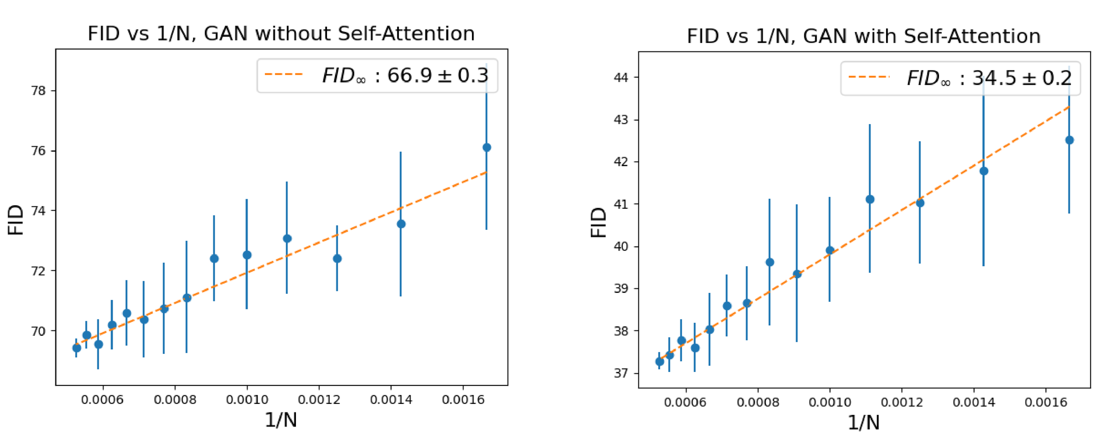


Figure 7: Effectively unbiased FID values obtained by extrapolating FID vs 1/N [Chong, 2019].

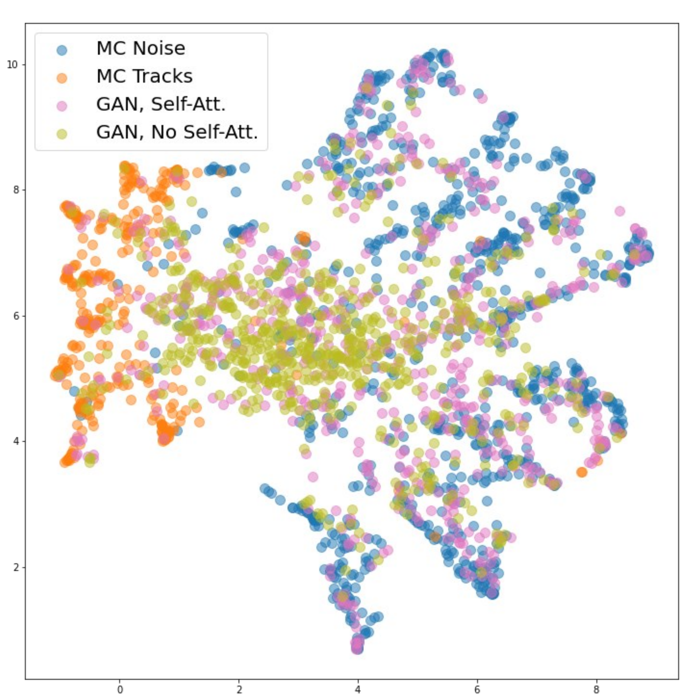


Figure 8: UMAP plot of the 267-dimensional feature vectors of hit MC and GAN hit sequences obtained by the 3D CNN.

Conclusions

- We measured FID with respect to features which are relevant to noise/track classification.
- StyleGAN-XL (image generation SOTA in 2022) has a relative FID of ~1 [Shmelkov, 2018], indicating that the CDC GAN performs well.
- The choice of feature extractor can be extended to any classification task, making the evaluation method generalizable.
- Our method demonstrates the superiority of the GAN with self-attention layers.