Fast simulation of the ILD electromagnetic calorimeter with Deep Learning

Peter McKeown

School of Physics and Astronomy

University of Nottingham

United Kingdom

September 4, 2019

Abstract

The ability to perform precise detector simulations is crucial in high energy physics. However, traditional techniques relying on Monte Carlo methods are computationally intensive and time consuming. This is particularly true in calorimeters where the large number of particle-material interactions require significant computational resources. This work evaluates the performance of a proposed fast simulation architecture, based on generative adversarial networks (GANs), for a simplified model of the International Large Detectors' calorimeter. Good qualitative and computational performance is obtained, in agreement with the original study. However, further work is required to obtain the precision needed for a simulation that could be used in physics studies.

Contents

1 Introduction **3**

2 Theory **4**
2.1 Calorimetry and the International Large Detector 4
2.2 Generative Adversarial Networks 6

3 Methods **6**
3.1 Training data 6
3.2 The CaloGAN architecture 8
3.3 Training strategy 9

4 Results **10**
4.1 Qualitative Review 10
4.2 Quantitative Review 11
4.3 Selecting a particular energy 13
4.4 Computational performance 15

5 Summary and Outlook **15**

6 Acknowledgements **16**

1 Introduction

The ability to perform fast and detailed simulations is of fundamental importance in modern high energy physics experiments. Not only is simulation crucial in the design, optimisation and characterisation of various detector elements, it is also a necessity for physics analyses to be able to compare experimental observations and theoretical models.

The nature of collisions in such experiments results in a rich and diverse range of physical processes as particles interact and decay. Traditionally, such physics processes are simulated using *Monte Carlo* (MC) methods in a multi-stage process. Firstly, the final state particles produced by hard scattering in the collider are generated from a matrix element and numerically solving the relevant quantum field theory equations. Examples of event generators include HERWIG [1], PYTHIA [2] and WHIZARD [3]. The resulting particles are then fed into a second stage that simulates their passage through the material in the detector. In the majority of experiments, the GEANT4 simulation toolkit is used for this stage [4] [5]. As a final step in the simulation, reconstruction and particle identification algorithms are run on the simulated data to allow for appropriate benchmarking and performance analysis [6].

While sufficiently accurate, the performance of Monte Carlo simulations is severely limited by the vast computing resources required. In recent years, more than 50% of computing time on the LHC Computing Grid (WLCG) was spent performing Monte Carlo simulations [7]. The high-luminosity phase of the LHC (HL-LHC), planned to start runs in 2025, will intensify the problem due to the increased levels of pile up (interactions per bunch crossing). In addition, analyses that study particularly rare processes require large MC statistics in order to sufficiently reduce errors, and therefore require many events to be simulated [8].

In order to combat these problems, many solutions have been proposed. Some attempt to reduce the amount of MC simulation required, for example by mixing simulated and experimental data [9]. Others attempt to optimise the existing MC simulation code for modern hardware, such as by using multi-threading [10]. However it is clear that to get significant performance improvements, a fundamentally different approach to MC simulation is required. The most computationally intensive part of a full detector simulation is emulating showers in the calorimeters, due to the large number of material-particle interactions. This is therefore a logical starting point for an alternative simulation technique.

In 2017, M. Paganini et al. proposed CaloGAN; a deep learning architecture that seeks to perform fast simulation of particle showers in an electromagnetic calorimeter [11]. In the original paper, the architecture was tested using a calorimeter designed to model the ATLAS electromagnetic calorimeter [12]. The goal of this paper is to replicate the analysis performed in [11], but using a calorimeter model which emulates that designed for the International Large Detector (ILD)- one of the detectors for the proposed International

Linear Collider (ILC) [13].

2 Theory

2.1 Calorimetry and the International Large Detector

Calorimetery is an integral part of any particle physics experiment, as it allows for the measurement of the energy and direction of travel of many types of particles that result from a collision. Furthermore, they are useful for particle identification techniques that seek to identify particles from their unique shower shapes [14]. Detectors for collider experiments contain two different types of calorimeter: electromagnetic and hadronic. These calorimeters exploit the different types of particle-material interactions and shower shapes to distinguish particles. Electromagnetic calorimeters seek to measure the energy of particles largely by electromagnetic interactions, whereas hadronic calorimeters seek to measure the energy of particles by inelastic nuclear processes. In this work we will focus on electromagnetic calorimeters, and so will be concerned with the showers of electrons, photons, positrons and pions.

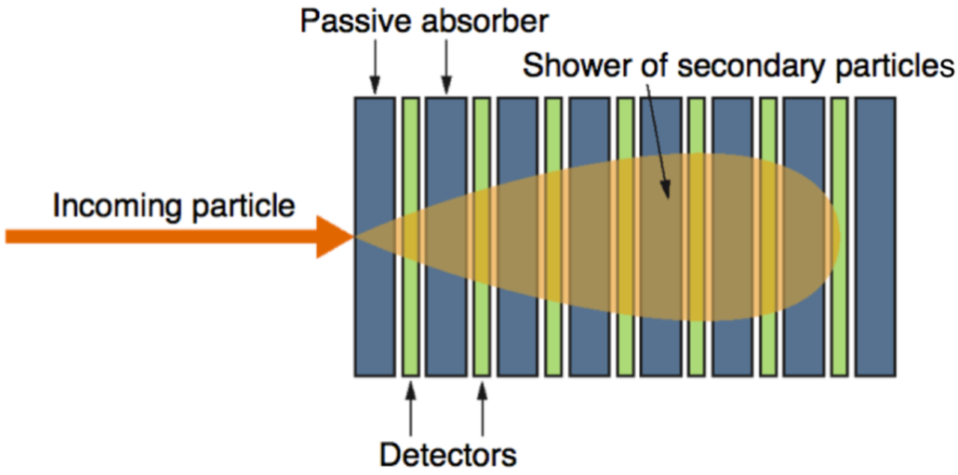

Figure 1: Diagram demonstrating a particle shower in a sampling calorimeter [15]

Calorimeters measure the energy of incident particles destructively, by causing a *shower* of secondary particles as shown in Figure 1. As these particles cascade, their energy is recorded and read out via electric signals. Due to the nature of particle showers, calorimetry is inherently statistical and each particle produces on average N secondary particles, with N being proportional to the energy of the initial particle. This continues until to a critical energy is reached, below which the particles can no longer shower. The performance of calorimeters is well known to improve with increasing energy of the incoming particle; in the case of an ideal homogeneous calorimeter the relative energy

resolution scales as:

$$\frac{\sigma(E)}{E} \sim \frac{1}{\sqrt{E}}. \tag{1}$$

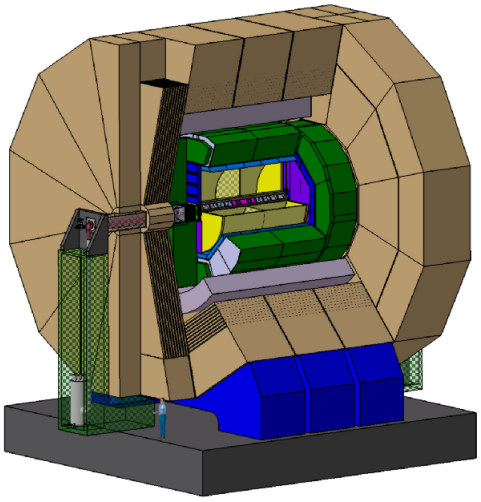

Figure 2: Artists' impression of the International Large Detector, one of the two detectors for the proposed International Linear Collider [13]

This work seeks to use a model of the electromagnetic calorimeter proposed for the International Large Detector (ILD) (see Figure 2), one of the two detectors planned for the International Linear Collider (ILC). This project aims to build a 20km long electron-positron collider with an initial center of mass energy of 250GeV[1] , in order to conduct precision measurements specifically targeted at studying the Higgs boson. The precision required for such physics measurements places tight constraints on the detector design. In particular, a high granularity calorimeter is required in order to allow the implementation of a Particle Flow algorithm (PFO), enabling the reconstruction of every particle individually in the detector. In order to satisfy these requirements, the ILD will use a sampling calorimeter- a type of calorimeter consisting of alternating layers of absorber and active material, as shown in Figure (1). This kind of calorimeter only measures a part of the total energy of the incident particle, since there is no read out of energy deposition in the absorber layers. It is therefore necessary to perform extensive calibration and testing in order to accurately understand the calorimeter and be able to determine the total initial energy. In total, the calorimeter consists of 30 layers, which are longitudinally segmented with absorber layers of tungsten (W), and active readout layers of silicon (Si). The segmentation in each layer is uniform, with one option which

[1]There is the potential to later upgrade to 500GeV and 1TeV center of mass energies, for an extended physics reach

uses scintillator tiles of 3×3 cm^2, and another which uses a gas-based readout and permits 1×1 cm^2 cells [13].

2.2 Generative Adversarial Networks

Much modern research in machine learning is focused on *generative models*. This broad term refers to an unsupervised learning method[2], in which a generator attempts to learn a target probability distribution. By performing random sampling of the learnt distribution, new data points from the target distribution can be produced. Deep neural networks are excellent candidates for such models, as they are capable of fitting complicated, non-linear functions. The pre-eminent framework among deep generative models is the generative adversarial network (GAN), originally proposed by I. Goodfellow et al. in 2014 [16]. This scheme pits a generator network G, against a discriminator network D, with both trying to achieve discordant goals. The generator G tries to map a latent space of random noise $z \sim p_z(z)$ to the target distribution of the training data, whereas D is a binary classifier that tries to return the probability of a given sample being real. The training strategy corresponds to D trying to maximise the probability that it correctly classifies real and fake data, and to G trying to minimise the probability that D performs the classification correctly, thereby fooling D. More formally, if $D(x)$ represents the probability that a sample x came from the true distribution rather than the generated distribution, and $G(z)$ is the map which the generator performs, then we can define the following two-player minimax game with the value function $V(D, G)$:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))], \tag{2}$$

which serves as the loss for the adversarial training, $\mathcal{L}_{adv}$. In fact this training method admits a unique solution. From the perspective of Game Theory, this corresponds to a Nash Equilibrium, developed by J. F. Nash Jr. [17]: for a non-cooperative two (or more) player game in which both players have a chosen strategy, no player can benefit from a change in strategy. In the case of GANs, this corresponds to G replicating the underlying distribution of the training data, and $D = 0.5$ i.e. the discriminator can no longer distinguish the real data from the generated data and essentially guesses.

3 Methods

3.1 Training data

In contrast to the detailed and highly granular electromagnetic calorimeter planned for the ILD outlined in Section 2.1, we use a simplified model inspired by the training data used in [11]. This setup is based on the GEANT4 B4 example, and consists of alternating layers of Si active and W passive absorber material that are 2mm and 4mm thick respectively. Three calorimeter layers are then created by summing the energies in

[2]This refers to a technique in which no data labels are provided

both the active and passive layers. This reduces the sparsity in the calorimeter images produced- a major problem that is faced by the network.

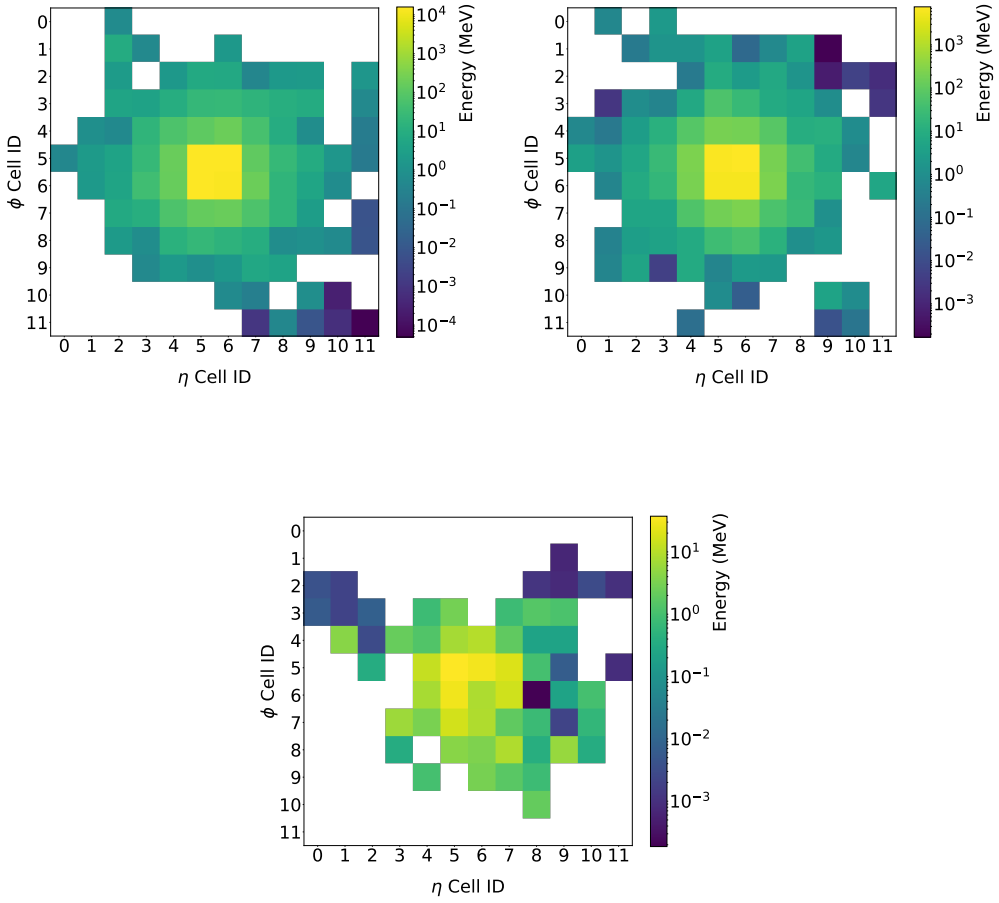

Figure 3: Example training data images of showers produced in the first, second and third layers of the GEANT4 calorimeter model.

The cells in each layer of the calorimeter have a uniform segmentation to reflect the uniformity of the ILD calorimeter. That said, our model has a significantly coarser granularity (of about one order of magnitude), with cells that are 40×40 mm. This results in images that consist of 12×12 pixels, as shown in Figure (3). These examples clearly illustrate some of the difficulties that are specific to the problem of simulating calorimeter showers. Firstly there is a high dynamic range, as the energy depositions span many orders of magnitude, particularly in the first layer of the calorimeter. Secondly, there are high levels of sparsity present in the images, particularly in the final layer of the calorimeter. These are both issues that a good network architecture must address.

3.2 The CaloGAN architecture

The CaloGAN architecture used in this study [11] is a custom architecture, designed specifically to deal with the challenges presented by our training data (see Section 3.1). CaloGAN consists of a Deep Convolutional GAN (DCGAN) [18] altered to replace some convolutional layers with locally connected layers, according to a previous paper by the same authors, to produce a Location Aware GAN (LAGAN) [19].

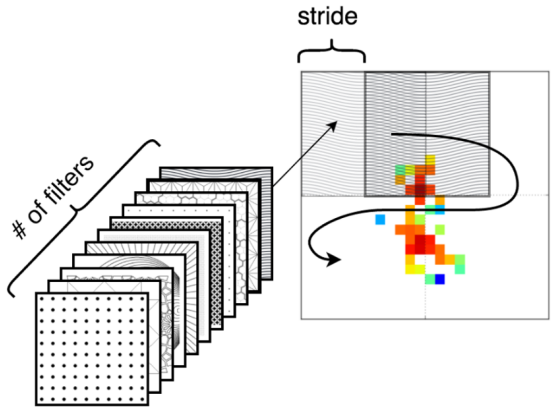

Figure 4: Diagram illustrating a traditional convolution filter that is applied to the entirety of an image [19].

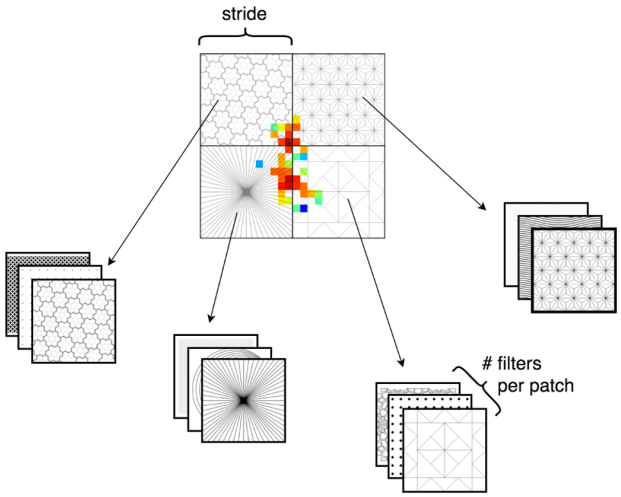

Figure 5: Diagram illustrating the filtering method used in the LAGAN architecture. A specific series of filters are learnt for each individual pixel location of the image [19].

Figures (4) and (5) demonstrate how filters are applied in the DCGAN and LAGAN respectively. In order to preserve the three-dimensional structure of particle showers in the generated images, an attentional mechanism is employed by the authors, similar to that in [20]. Such a mechanism allows the network to focus on specific features of

its input and means that information about the distribution and intensity of energy patterns can be transferred between layers. Due to the high variation of shower shape with particle type (pions in particular can shower in a large variety of patterns, as they also undergo nuclear interactions), one generative model is trained per particle type. Sparsity is induced in the generated images by using a ReLU activation function in the final layer of the generator. This function is defined as

$$f(x) = \max(0, x), \tag{3}$$

and so is zero for x less than or equal to zero, but increases linearly for x greater than zero. This increases the number of nodes in the final layer that are not activated, and therefore produce no output, leading to more empty calorimeter cells.

Mini-batch discrimination is used to increase sample entropy in each of the training batches (subsets of the training data set used during training). This helps to prevent a common failure mechanism of GANs called *mode collapse*, whereby G only learns a subset of the training distribution. If G learns this set well enough, D cannot discriminate it from real data, and so G never learns the entirety of the true distribution. Having high entropy in each training batch means that during each training cycle, the training distribution is well sampled, helping to prevent such failure. Perhaps the most fundamental requirement of such simulations is to enforce conservation of energy. This is done in the CaloGAN architecture by making an addition to the original loss term (Eq.(2)) using

$$\mathcal{L}_E = \mathbb{E}_{z\sim p_z(z)}[\delta(E, \hat{E}(G(z)))] + \mathbb{E}_{x\sim p_{data}}[\delta(E, \hat{E}(x))], \tag{4}$$

where $\delta(e, e') = |e - e'|$, E is the requested energy, and $\hat{E}$ is the reconstructed energy. It is not possible, however to enforce energy conservation as a hard cut- we can merely punish (and thereby reduce) instances of too much or too little energy deposition.

Combining this additional loss term with the traditional loss $\mathcal{L}_{adv}$ given in Eq.(2), the generator will maximise

$$\mathcal{L}_{generator} = \lambda_E \mathcal{L}_E - \mathcal{L}_{adv}, \tag{5}$$

and the discriminator will maximise

$$\mathcal{L}_{discriminator} = \lambda_E \mathcal{L}_E + \mathcal{L}_{adv}. \tag{6}$$

Here λ_E is a hyper-parameter set to 0.05 to down weight the importance of the additional loss term $\mathcal{L}_E$. [11]

3.3 Training strategy

The network was trained on $100,000$ events in which a photon was fired perpendicular to the face of the calorimeter, in a simulation performed with GEANT4. The photons have energies distributed in a continuous and uniform range between 10 and 100 GeV. As in

the original paper [11], training data batches of 256 are used, and both the discriminator and generator are trained for 50 epochs. Following the recommendations in the original work [11], a learning rate of 2×10^{-5} is chosen for the discriminator, and a learning rate of 2×10^{-4} is chosen for the generator.

4 Results

It is well known that reviewing the performance of GANs is fraught with difficulty- good performance quantified with one metric does not necessarily imply good performance with another [21]. The strategy from [11] is thereby employed, with a qualitative review followed by a quantitative review in terms of physical variables.

4.1 Qualitative Review

By plotting a sample of GEANT4 simulation images along with their nearest neighbours in the generated data set, it is possible to obtain an overview of how well the generator is performing.

Figure (6) shows a randomly selected set of GEANT4 images and their nearest neighbours in the generated data set. There is good qualitative correspondence between the samples from each data set in all layers, and no obvious signs of mode collapse (see Section 3.2). The generator is able to match the sparsity levels relatively well, even in the final layer. It is interesting to note that in some samples of the GEANT4 data in the second layer there is significant low energy deposition (of order 10^{-1} MeV, represented in dark blue), which the generator is not able to match. Although tentative due to the low number of samples presented, this highlights a shortcoming in this network architecture, which results from the high dynamic range present in the images- an area which requires further work.

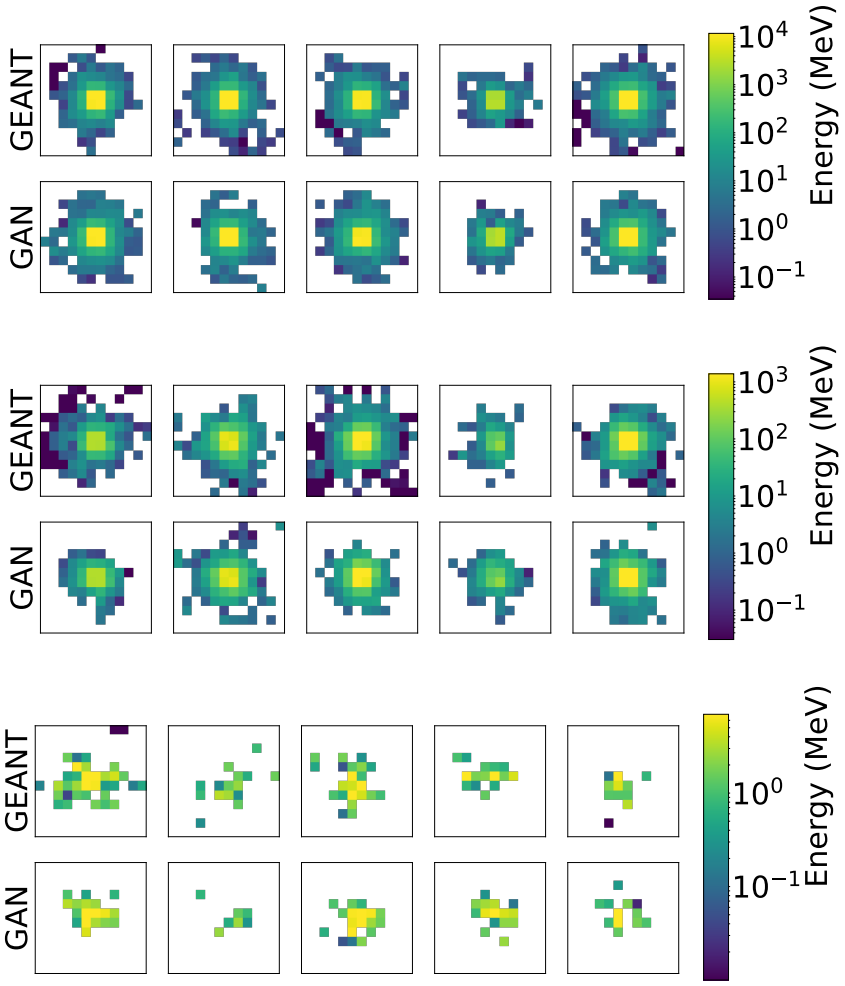

Figure 6: Calorimeter images of five randomly selected photon showers in the (GEANT4) training set, and their nearest neighbours in the generated (CaloGAN) data set. The euclidean distance was used as the nearest neighbour metric. The images are collected by calorimeter layer, going from the first layer (top) to the third layer (bottom). Note that the GEANT4 images in different layers are not necessarily from the same shower.

4.2 Quantitative Review

In order for a calorimeter simulation to be useful, it must be able to reproduce properties of the showers produced by the various particles. It is therefore informative to use these as a metric of how well the GAN performs. The most fundamental and simple quantity is the energy deposition in the calorimeter, both the total energy deposition in the calorimeter as a whole, and in each of the individual layers of the calorimeter.

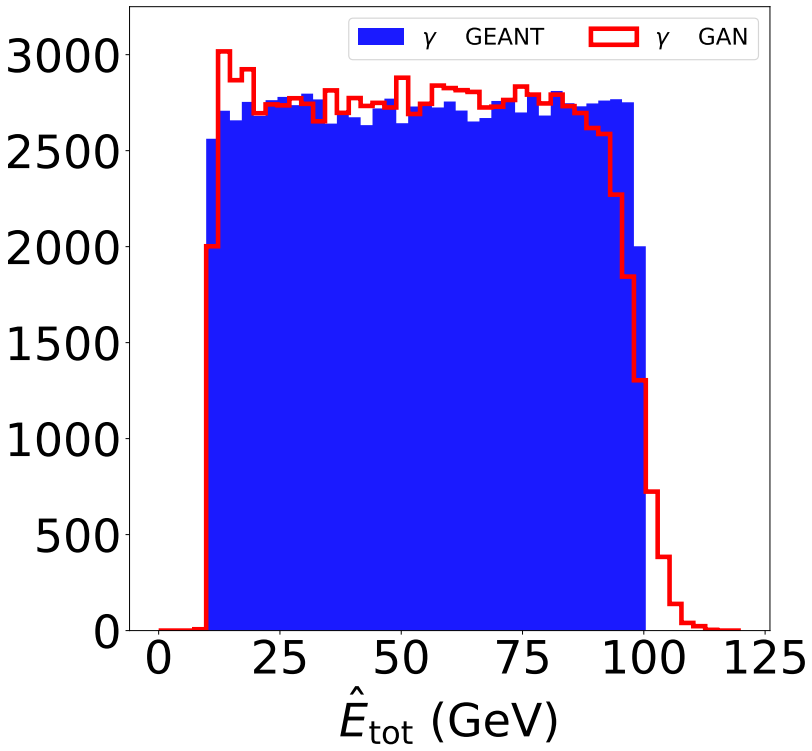

Figure 7: Histogram of the total energy deposited by photons in the calorimeter. The blue region is the total energy deposition in the GEANT4 calorimeter, with all $100,000$ events being included in the plot. The red line represents the total energy deposition in the calorimeter produced using the generator. $100,000$ photon showers were used to produce this plot, allowing comparison of the two distributions.

Figure (7) shows the total energy deposited in both the GEANT4 calorimeter, and the calorimeter simulated using the generator from the GAN. In both cases $100,000$ events are included, and the total energy is found by simply summing the energy across each calorimeter layer. The generator was requested to simulate photons in the continuous energy range 10 to 100 GeV used in the GEANT4 simulation. Since we find the total energy deposition in the calorimeter, the GEANT4 plot just returns the total energy of incident photons, giving abrupt cut-offs at 10 GeV and 100 GeV. However, since we are unable to enforce a hard physical notion of energy conservation (see Section 3.2) in the generation process, we see a suppressed but noticeable spill over of some events with more than 100 GeV in the generated data. Except for this discrepancy, there is a relatively good correspondence between the training and generated distributions.

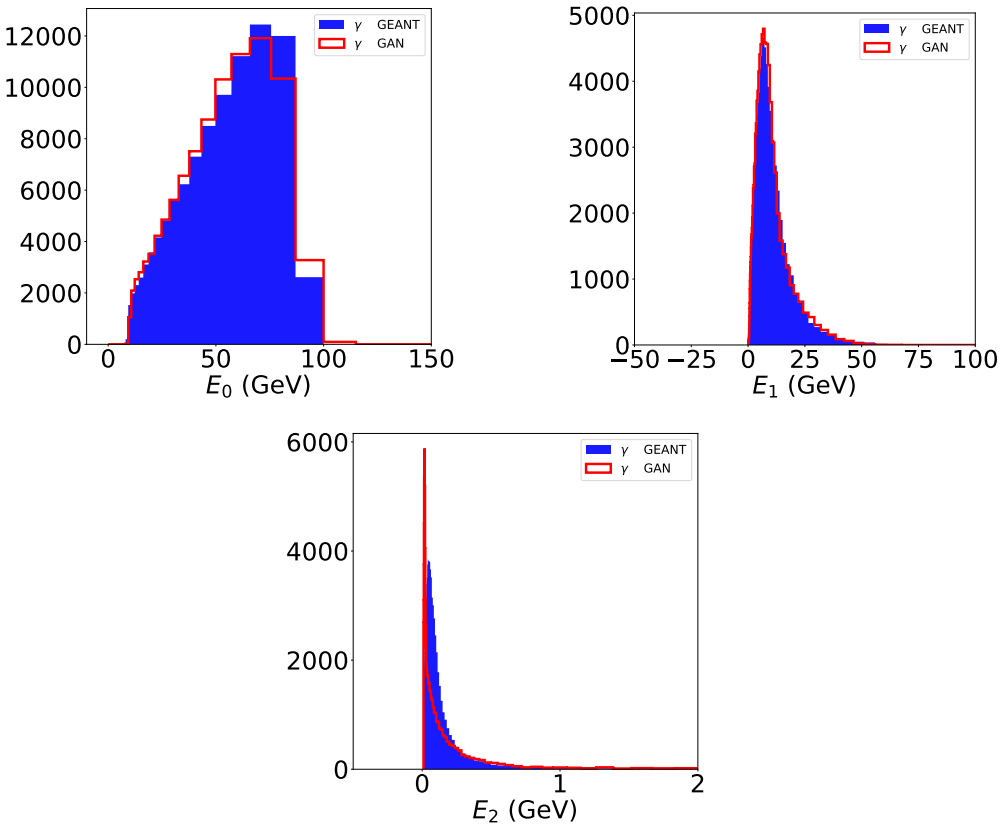

Figure 8: Histogram of the energy deposited by photons in each layer of the calorimeter. The blue region is the energy deposition in the GEANT4 calorimeter layer, with all 100,000 events being included in the plot. The red line represents the energy deposition in the same calorimeter layer, produced using the generator. 100,000 photon showers were used to produce each plot, allowing comparison of the two distributions.

Figure (8) shows the energy deposition in each layer of the calorimeter, using the same continuous distribution of energies as previously. The generator matches the training data well in the first two layers, with an excellent match being obtained in the second layer. However, in the third layer there is a noticeably narrower distribution (and as such a much higher peak at lower energies). This is likely due to heightened levels of sparsity in the third layer, as opposed to the first and second layers, which the generator struggles to learn. It is worth reminding ourselves at this point that the CaloGAN architecture does not incorporate shower shape variables explicitly in the training process, and it is suggested that their inclusion in the loss-function could improve simulation fidelity [11].

4.3 Selecting a particular energy

Note that so far we have only considered a continuous range of energies. This is of limited use in detector simulation, however, as the particles in any single event will have

a given energy, and we must be able to request a particle of such an energy. For this reason, the generator was requested to produce shower simulations of $10,000$ events at specific energies.

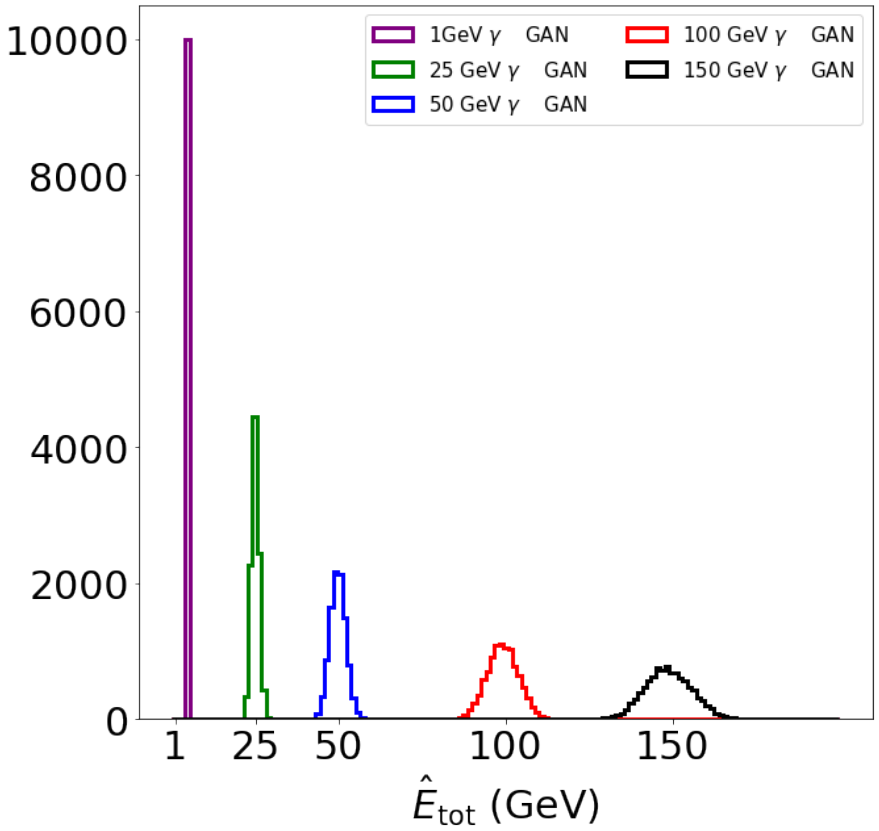

Figure 9: Response of the calorimeter to simulation of incident photons at a single energy. Note that the *total* energy deposited in the calorimeter is plotted. Events were generated with energies of 1, 25, 50, 100, and 150 GeV, and $10,000$ samples were produced in each case. Note that the energies 1 and 150 GeV were outside of the training range.

Figure (9) shows the response of the calorimeter when a simulation of photons incident with a single energy is requested. Note that this is the total energy deposited in the calorimeter (including both absorber and active layers). This means the equivalent GEANT4 plot would consist of a single peak with no spread at each of the requested energies. Notice that the energies requested at 1 GeV and 150 GeV lay outside of the training range, and thus produced a mode that is slightly shifted towards the domain of the training data. The broadening of the peak at higher energies was also studied, to see if it scaled like the relative energy resolution equation given in Eq. 1.

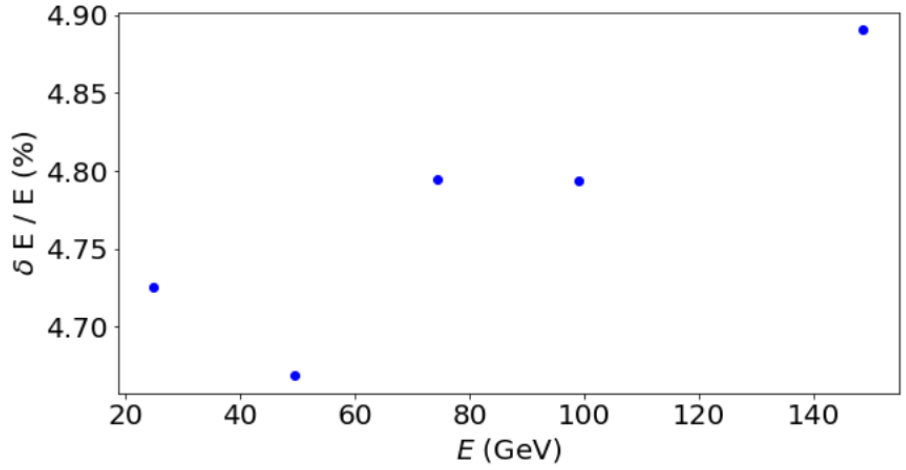

Figure 10: Plot of the relative energy resolution $\frac{\sigma(E)}{E}$ of each of the responses at energies of 25, 50, 100, and 150 GeV shown in Figure (9). Note the very fine gradations in the y scale.

The plot of how the relative energy scales with energy is shown in Figure (10). Note that the variations are on the sub-percent level, and as such $\frac{\sigma(E)}{E}$ is constant at about 4.8%. This means that the broadening of the peak is an error inherent in the generation process.

4.4 Computational performance

No rigorous performance test was undertaken as part of this study, as the GEANT4 simulation was performed on the Batch Infrastruktur Resource at DESY (BIRD) [22] (the weights of the generator were also obtained by training on the BIRD, but with a GPU), whereas the generation was performed on an Intel® Pentium(R) CPU G4400T @2.90GHz ×2. However, it was noted that while it took $\sim 3,000$ seconds to simulate $1,000$ showers with GEANT4, it only took ~ 3 seconds to simulate $1,000$ showers with the CaloGAN generator. This corresponds to a speed up of order $\sim 10^3$, in agreement with the results obtained in [11] for simulation on a CPU.

5 Summary and Outlook

This work has assessed the performance of an existing deep learning architecture designed to perform fast simulation of calorimeter showers, by producing photon showers in a multi-layer silicon-tungsten calorimeter, which models that of the International Large Detector. The architecture is found to be able to reproduce shower shape variables to a reasonable degree of accuracy. A simulation speed up of $\sim 10^3$ on dissimilar CPU hardware is obtained, agreeing with the more rigorous tests performed by the original authors. However, significant further work is required before such simulations will

be able to produce simulations of sufficient accuracy. This will involve altering the loss function, as well as more significant modifications to the network architecture. It will also require a more physical calorimeter model, consisting of more layers. Already further steps have been taken towards this goal, with studies undertaken on more modern GAN architectures. Examples include Wasserstein GAN architectures [23] and modified CaloGAN architectures [24] that condition on other physical parameters of particles, as well as energy (e.g. position and momentum). As a result, generative deep learning architectures show great promise for producing high energy physics simulations that are both fast and precise.

6 Acknowledgements

I would like to thank the FLC group for hosting my project this summer, in particular the Software sub-group in which I worked. Special thanks must go to Dr. Engin Eren for his dedication to helping me with any problems or questions I had, however menial (or stupid!), and to Dr. Frank Gaede for setting up this project and providing extensive guidance throughout. The PhD students and Postdocs in the FLC group were incredibly friendly, actively arranging external activities (thank you to Uli, who kindly agreed to show some of my visiting friends around HERA). A final thanks must go to the summer students, of whom there are too many to acknowledge. I am certain I have gained some close friends for life.

References

[1] *G. Marchesini et al.*, HERWIG 5.1- a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons, Comp. Phys. Comms. 67, pp. 465-508, (1992)

[2] *T. Sjöstrand*, High-energy physics event generation with PYTHIA 5.7 and JETSET 7.4, Comp. Phys. Comms. 82, pp. 74-90, (1994)

[3] *W. Kilian et al.*, WHIZARD- Simulating Multi-Particle Processes at LHC and ILC, J. Eur. Phys. J. C, 71: 1742, (2011)

[4] *S.Agostinelli et al.* (Geant4 Collaboration), GEANT4- a simulation toolkit, Nucl. Instrum. Meth. A, 506, pp. 250-303, (2003)

[5] *J. Allison et al.* (Geant4 Collaboration), Recent developments in GEANT4, Nucl. Instrum. Meth. A, 835, pp. 186-225, (2016)

[6] *V. Daniel Elvira*, Impact of Detector Simulation in Particle Physics Collider Experiments, Phys. Rep., 695, pp. 1-54, (2017)

[7] *V. Breton et al.*, Computing Resources Scrutiny Group, Tech. Rep. CERN-RRB-2015-014, (2015)

[8] *W. Lukas*, Fast Simulation for ATLAS: Altfast-2 and ISF, J. Phys.: Conf. Ser., 396, Part 2, (2012)

[9] *ATLAS Collaboration*, Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1, arXiv:1603.02934, (2016)

[10] *M. Clemencic et al.*, Gaudi components for concurrency: Concurrency for existing and future experiments, J. Phys.: Conf. Ser., 608, conference 1, (2015)

[11] *M. Paganini et al.*, CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks, arXiv:1712.10321, (2017)

[12] *ATLAS Collaboration*, ATLAS liquid-argon calorimeter: Technical Design Report, CERN, Geneva, (1996)

[13] *T. Behnke et al.*, The International Linear Collider Technical Design Report- Volume 4: Detectors, arXiv:1306.6329, (2013)

[14] *C. W. Fabjan et al.*, Calorimetry for particle physics, Rev. Mod. Phys., 75, 1243, (2003)

[15] *M. Krammer*, Detectors for Particle Physics: Calorimeters, `https://www.hephy.at/fileadmin/user_upload/VO-6-Calorimeters.pdf`, accessed 03.09.2019

[16] *I. J. Goodfellow et al.*, Generative Adversarial Networks, arXiv:1406.2661, (2014)

[17] *J. F. Nash Jr.*, Equilibrium points in n-person games, PNAS, 36 (1), pp. 48-49, (1950)

[18] *A. Radford et al.*, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434, (2015)

[19] *L. de Oliveira et al.*, Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis, arXiv:1701.05927, (2017)

[20] *H. Zhang et al.*, StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks, arXiv:1612.03242, (2016)

[21] *L. Theis et al.*, A note on the evaluation of generative models, arXiv:1511.01844, (2015)

[22] *Y. Kemp et al.*, Batch Infrastruktur Resource at DESY (BIRD), `https://confluence.desy.de/pages/viewpage.action?pageId=67639562`, accessed 04.09.2019

[23] *M. Erdmann et al.*, Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network, arXiv:1807.01954, (2018)

[24] *V. Chekalina et al.*, Generative Models for Fast Calorimeter Simulation: the LHCb case, arXiv:1812.01319, (2018)