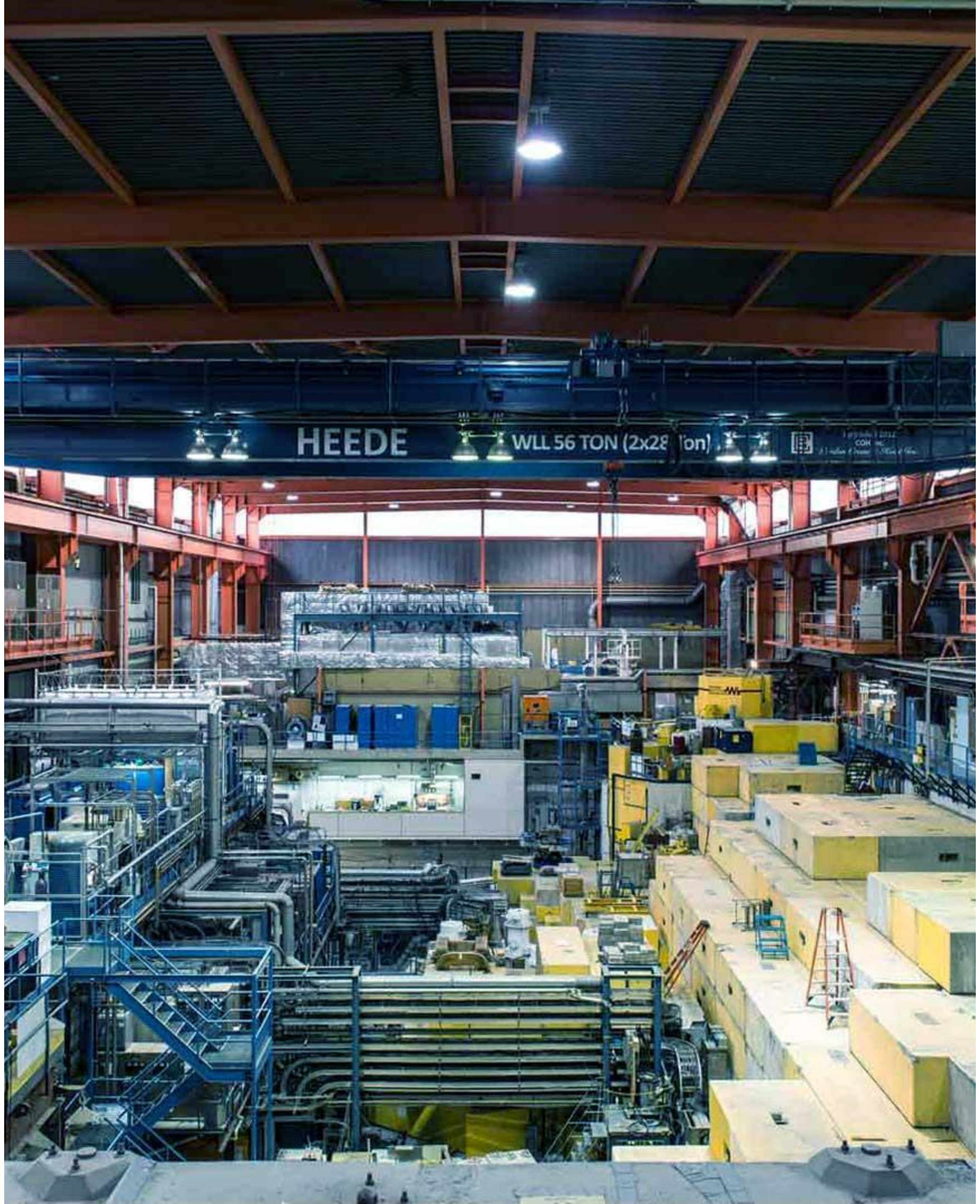
TRIUMF

CaloDVAE : Discrete VAEs for Fast Calorimeter Simulation

A. Abhishek, E. Drechsler, W. Fedorko, B. Stelzer

abhishek@myumanitoba.ca

2019-03-27

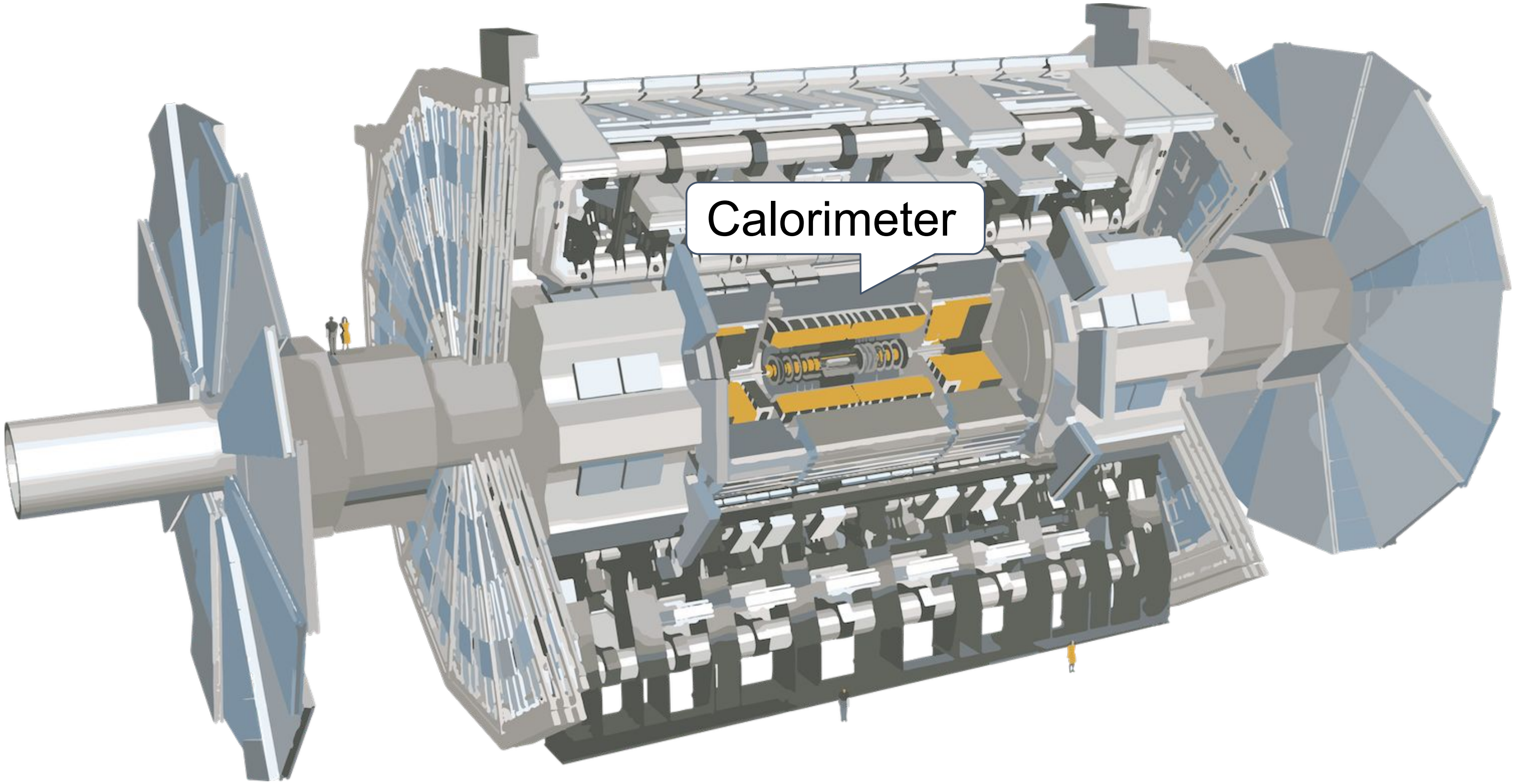


Discovery, accelerated

ATLAS at Large Hadron Collider

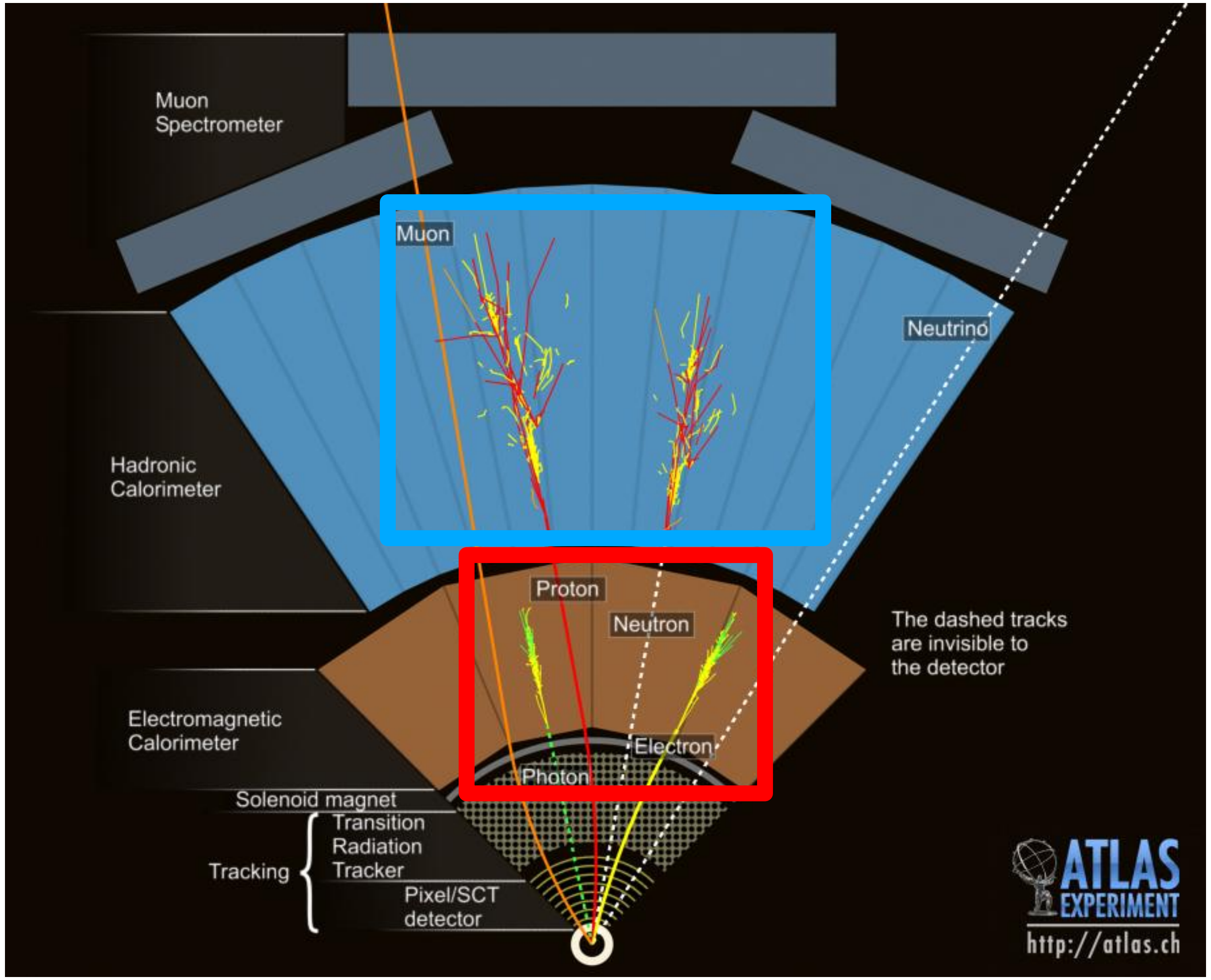
Physics goals :

1. Higgs Boson measurements
2. Physics beyond the Standard Model (BSM) e.g. Dark matter
3. Rare processes e.g. WWW production



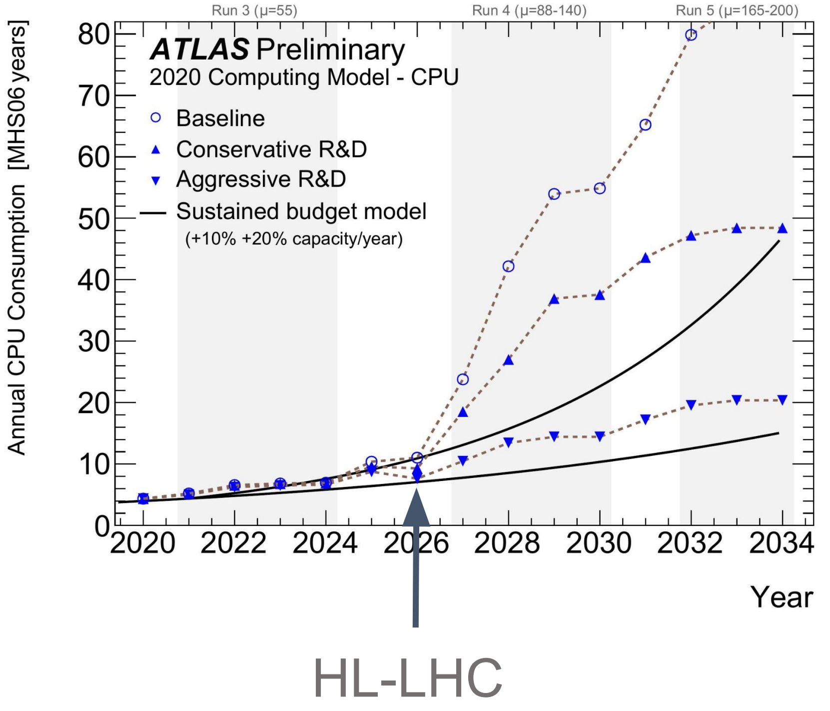
Calorimeters

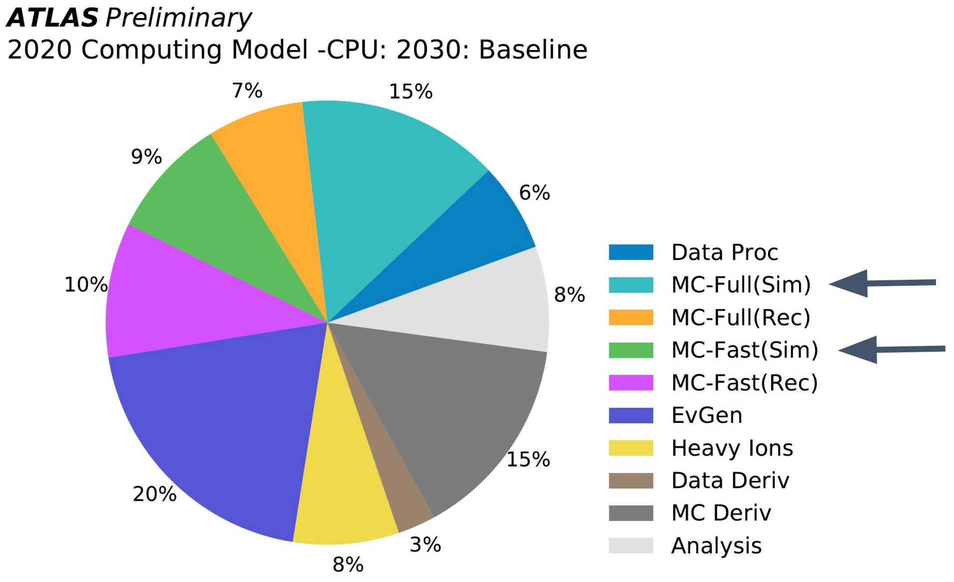
Measure energy of the particles through Electromagnetic or Hadronic interactions



Motivation

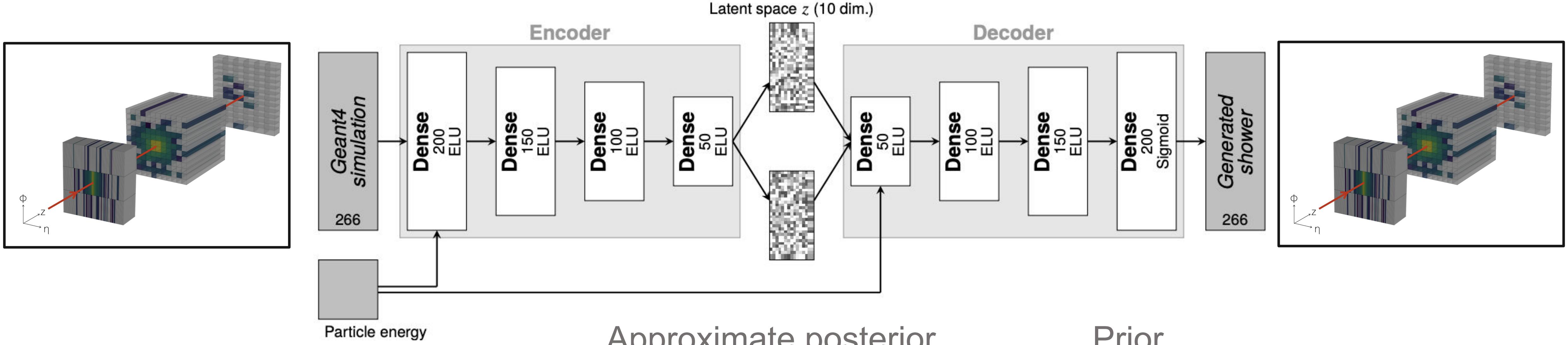
Current techniques for Calorimeter shower simulation are computationally expensive





[1] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults

Previous Work : DGMs for Fast Calorimeter Simulation



Approximate posterior

$$z_i \sim \mathcal{N}(\mu_i(\mathbf{x}, e), \sigma_i^2(\mathbf{x}, e))$$

Prior

$$\mathbf{z}_i \sim \mathcal{N}(0, 1)$$

$$\text{Loss}(\mathbf{x}, \mathbf{x'}), \text{e.g. } \frac{1}{n}\Sigma_{i=1}^{n}(\mathbf{x_i} - \mathbf{x'_i})^2 \quad + \quad D_{KL}[q_\phi(\mathbf{z}|\mathbf{x}, e)||p(\mathbf{z})]$$

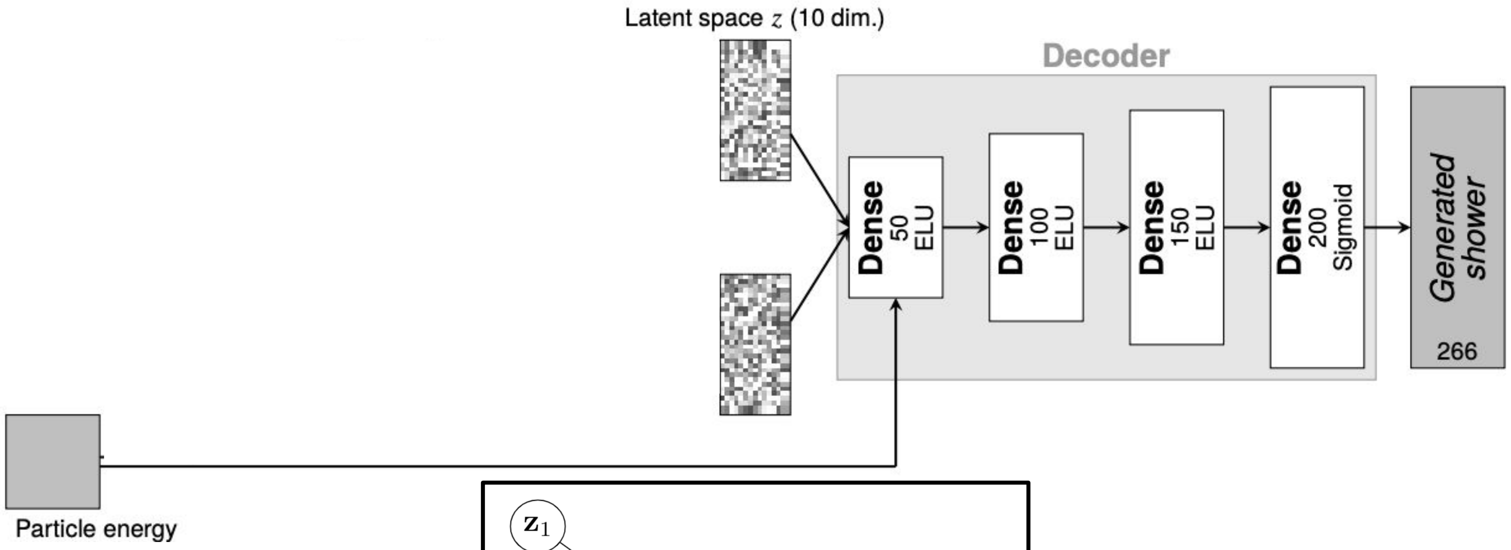
Autoencoding loss

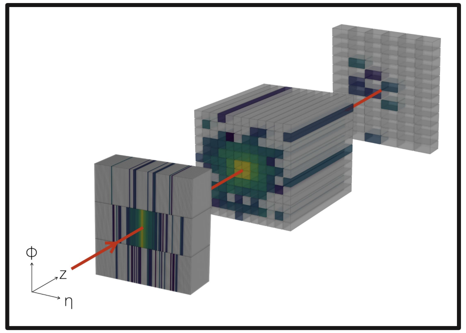
KL divergence loss

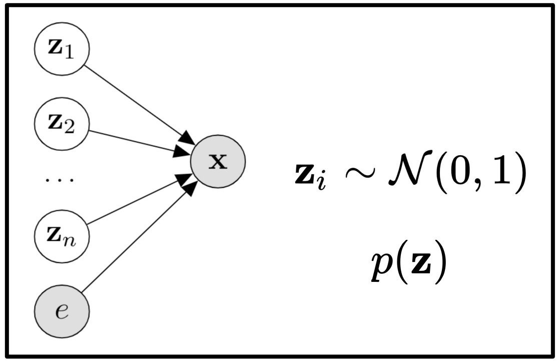
Upper bound to the Negative log-likelihood of the data under the model distribution

[1] Paganini, de Oliveira and Nachman (2018), arXiv:1712.10321 [2] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001.

Previous Work : DGMs for Fast Calorimeter Simulation





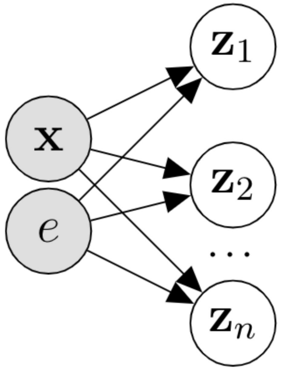


[1] Paganini, de Oliveira and Nachman (2018), arXiv:1712.10321 [2] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001.

VAEs to Discrete VAEs

Approximate posterior distribution using during autoencoding $q_\phi(\mathbf{z}|\mathbf{x}, e)$

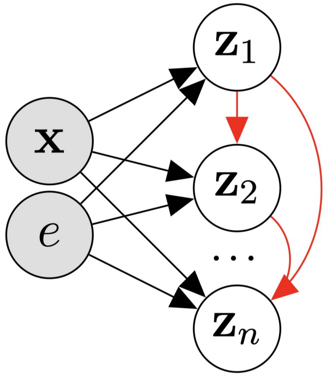
Continuous, Factorial Gaussian



$$z_i \sim \mathcal{N}(\mu_i(\mathbf{x}, e), \sigma_i^2(\mathbf{x}, e))$$

Independence assumption

Discrete, Hierarchical Bernoulli

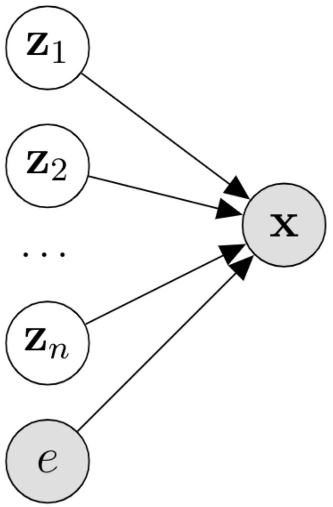


$$\mathbf{z}_i \sim q_{\phi_i}(\mathbf{z}_i | \mathbf{z}_{j<i}, \mathbf{x}, e)$$

No independence assumption

VAEs to Discrete VAEs

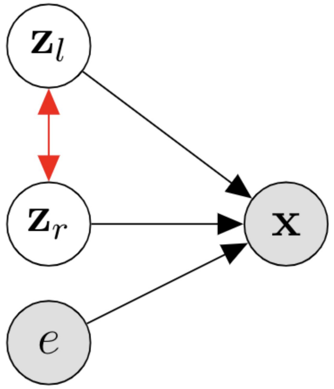
Factorial Normal Gaussian prior



$$\mathbf{z}_i \sim \mathcal{N}(0,1)$$

- Computationally efficient sampling
- Independence assumption

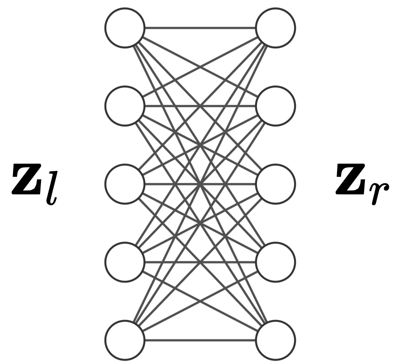
Trainable Restricted Boltzmann Machine prior



$$\mathbf{z}_i \sim p_{\theta_{RBM}}(\mathbf{z})$$

$$p_{\theta_{RBM}}(\mathbf{z}) = e^{-E_{\theta_{RBM}}(\mathbf{z})} / Z_\theta$$

$$= e^{\mathbf{a}_l^T \mathbf{z}_l + \mathbf{a}_r^T \mathbf{z}_r + \mathbf{z}_l^T \mathbf{W} \mathbf{z}_r} / Z_\theta$$

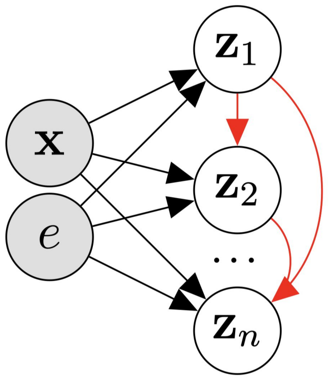


Parameters jointly trained with VAE parameters

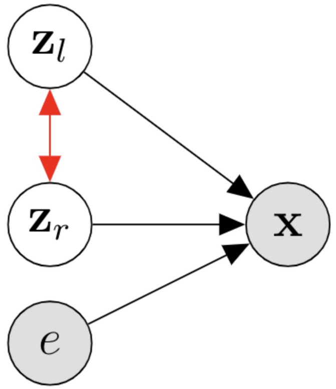
- Computationally expensive block Gibbs sampling

Why Discrete VAEs ?

1. More expressive latent space → Better performing generative model



$$\mathbf{z}_i \sim q_{\phi_i}\left(\mathbf{z}_i \middle| \mathbf{z}_{j<i}, \mathbf{x}, e\right)$$



$$\mathbf{z}_i \sim p_{\theta_{RBM}}(\mathbf{z})$$

$$p_{\theta_{RBM}}(\mathbf{z}) = e^{-E_{\theta_{RBM}}(\mathbf{z})} / Z_\theta$$

$$= e^{\mathbf{a}_\mathrm{l}^T \mathbf{z}_\mathrm{l} + \mathbf{a}_\mathrm{r}^T \mathbf{z}_\mathrm{r} + \mathbf{z}_\mathrm{l}^T \mathbf{W} \mathbf{z}_\mathrm{r}} / Z_\theta$$

2. Prospect of **Quantum VAEs**

- Replace the classical Restricted Boltzmann Machine (RBM) prior with a Quantum Boltzmann Machine (QBM) prior
- Use Quantum Annealing to generate latent variable samples instead of Block Gibbs sampling

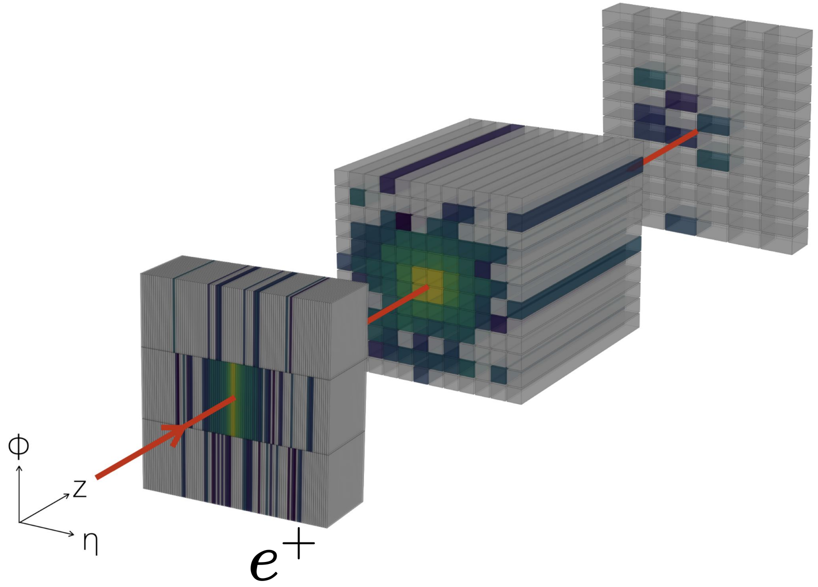
EM Calorimeter Shower Dataset

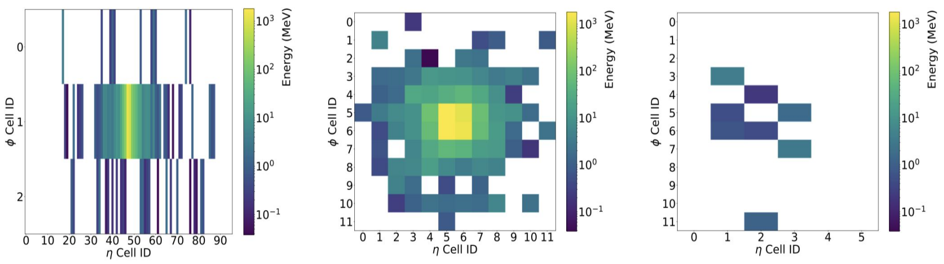
e^+, γ, π^+

$1 < E < 100$ GeV

$100,000$ events per particle type

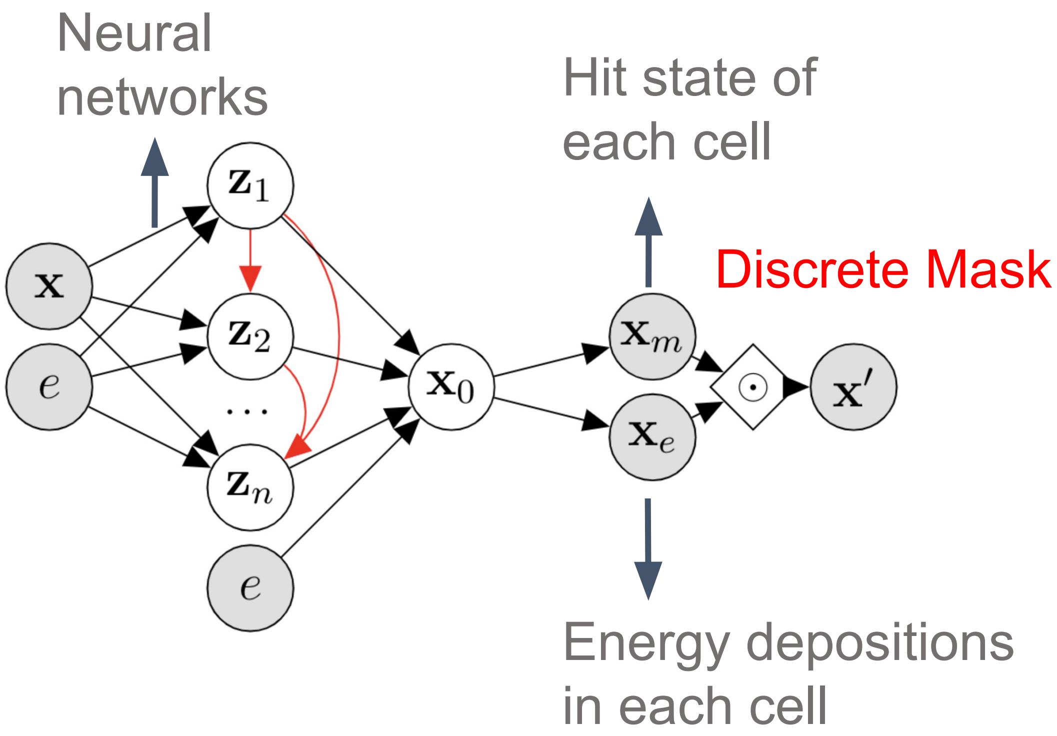
Inspired by ATLAS LAr EM Barrel Calorimeter Geometry



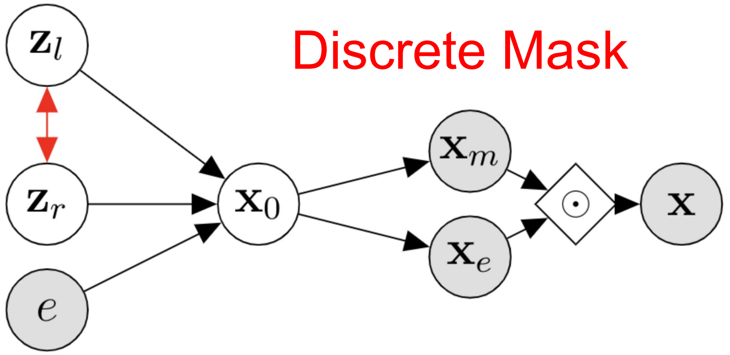


Nachman, de Oliveira and Paganini (2017), Mendeley Data, V1, doi: 10.17632/pvn3xc3wy5.1

DVAEs for Calorimeter Simulation



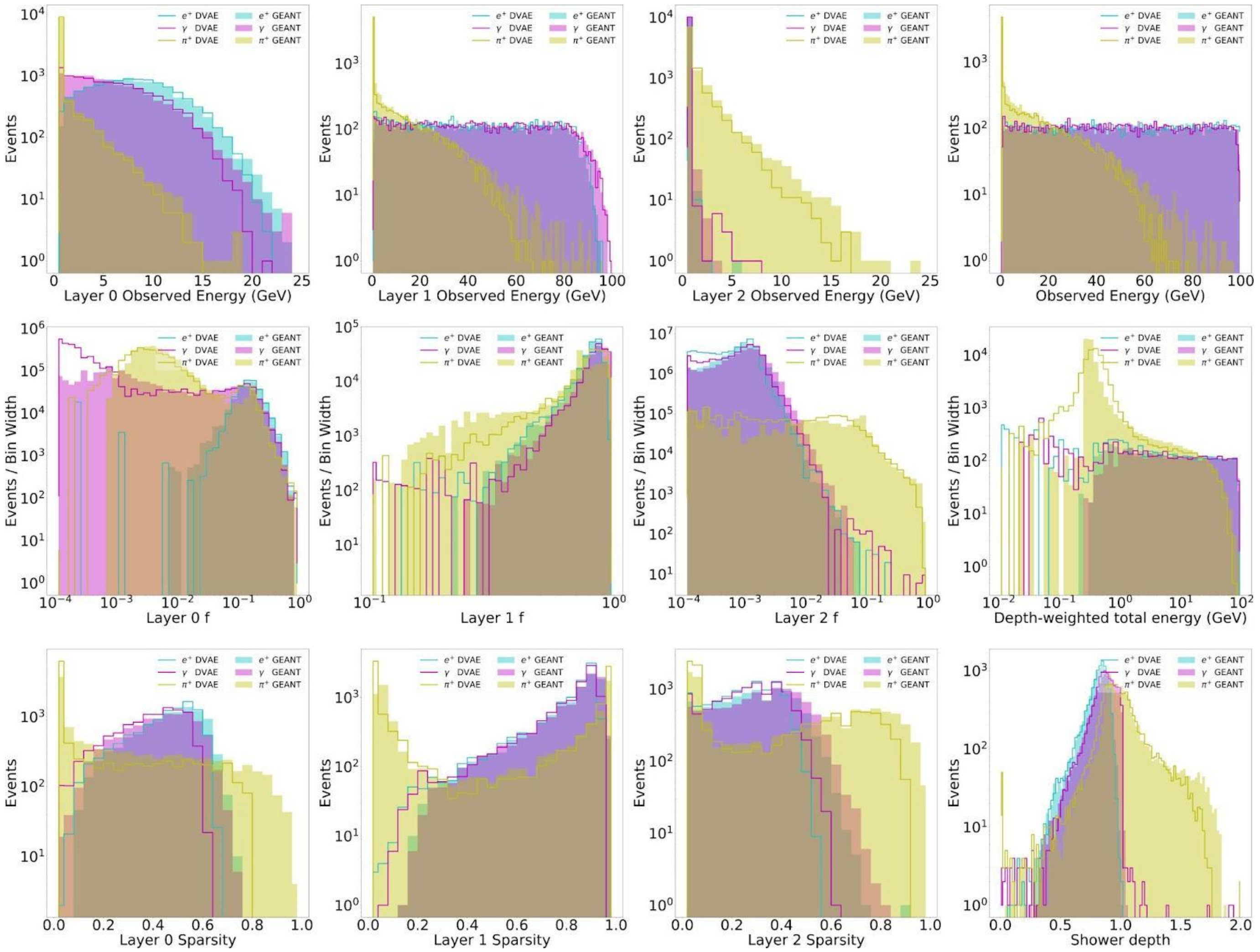
Autoencoding model



Generative model

Results : Shower shape variables

Shower shape variable	Notes
$E_i = \sum_{\text{pixels}} \mathcal{I}_i$	Energy deposited in the i^{th} layer of calorimeter
$E_{\text{tot}} = \sum_{i=0}^{2} E_i$	Total energy deposited in the electromagnetic calorimeter
$f_i = E_i / E_{\text{tot}}$	Fraction of measured energy deposited in the i^{th} layer of calorimeter
Depth-weighted total energy, $l_d = \sum_{i=0}^{2} i \cdot E_i$	The sum of the energy per layer, weighted by layer number
Shower Depth, $s_d = l_d / E_{\text{tot}}$	The energy-weighted depth in units of layer number

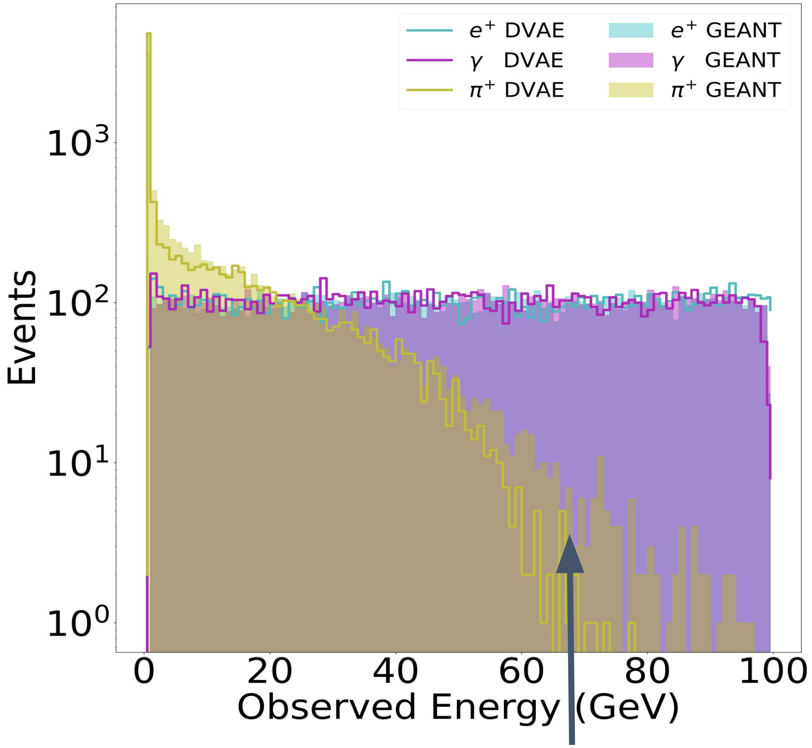


[1] Paganini, de Oliveira and Nachman (2018), arXiv:1712.10321

Energy spectra

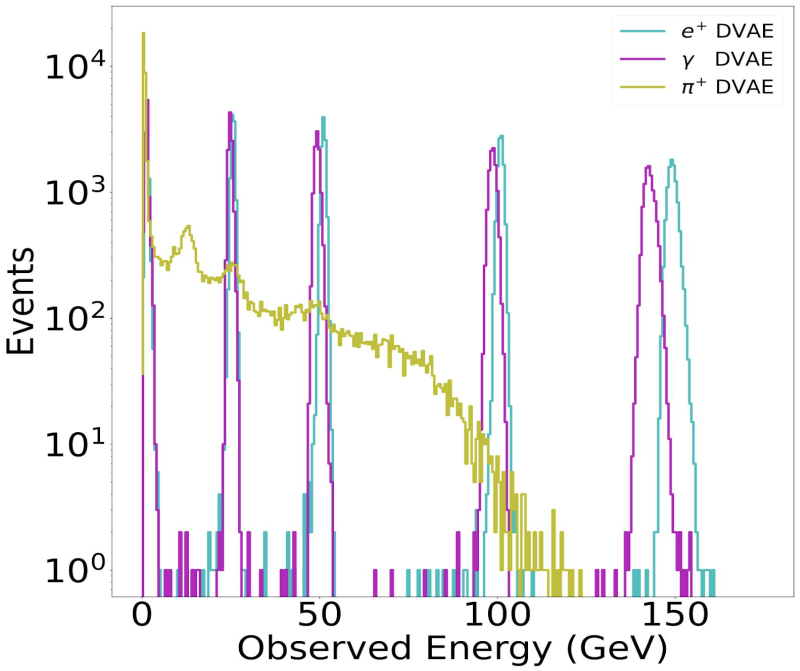
Energy conditioning

$$1 < E < 100 \text{ GeV}$$



Energy spectra of high energy charged pions underestimated

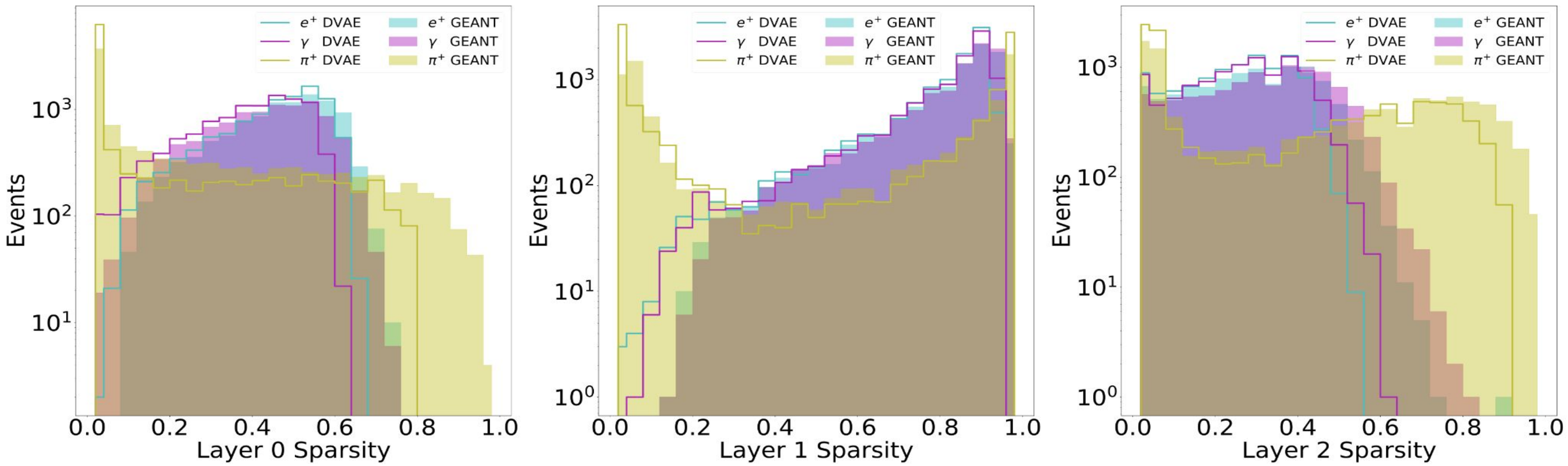
$$E \in [1, 25, 50, 100, 150] \text{ GeV}$$



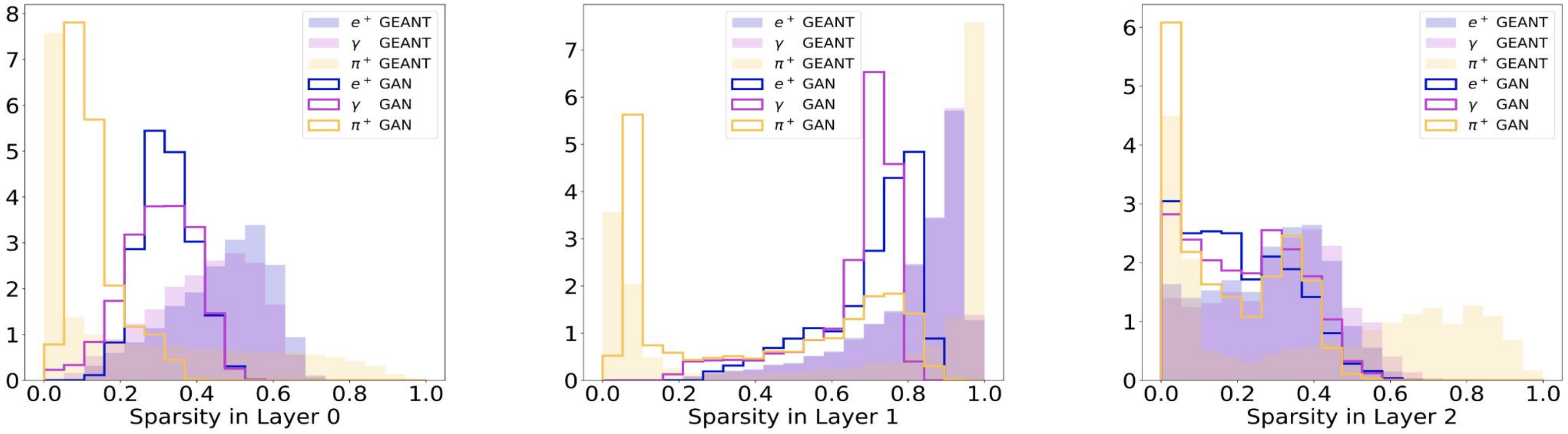
- Compelling peak at 150 GeV for positrons and photons
- Uncontained charged pions

Sparsity (Fraction of cells hit) per layer

CaloDVAE



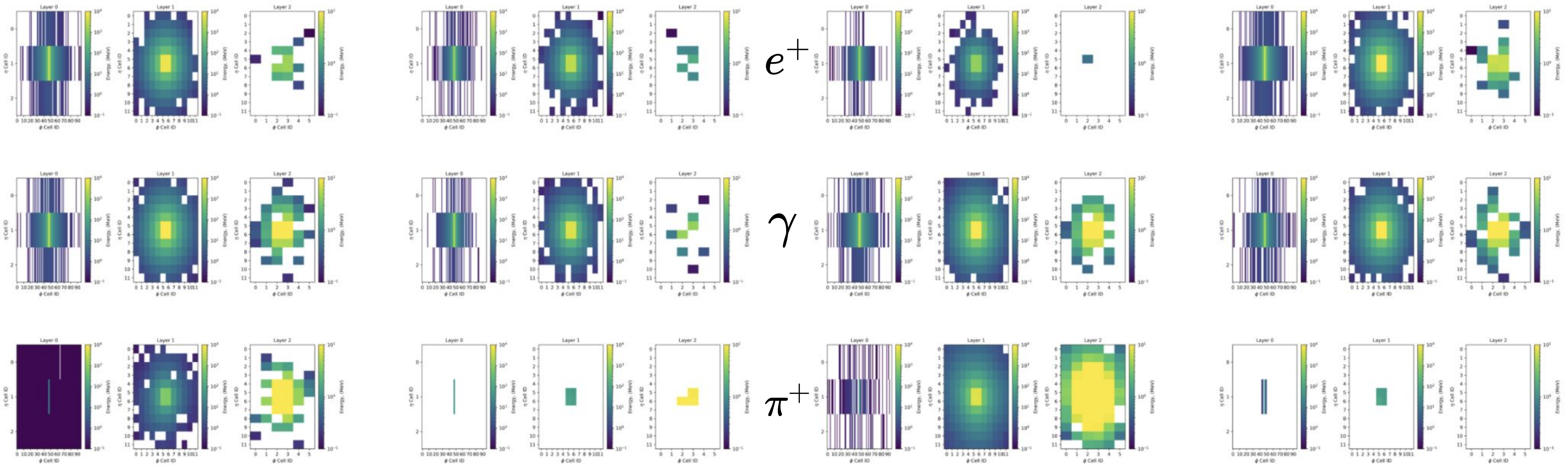
CaloGAN



- Previously observed to be challenging to reproduce
- Bimodal distribution for charged pions matches well

[1] Paganini, de Oliveira and Nachman (2018), arXiv:1712.10321

Shower Images



Generated samples recover :

- Wide variety of patterns of activated and non-activated cells
- **Centrality** and **lateral width** of the clusters
- Longitudinal behaviour of the shower - Most of the energy deposited in the middle layer

Summary and Future Outlook

- **DVAEs** show promising results for Calorimeter shower simulation
  - Able to match distributions of “shower shape” variables

- Latent generative process modelled by a **Restricted Boltzmann Machine (RBM)** is still computationally expensive

- Prospect of using **QVAEs** with latent generative process modelled by a **Quantum Boltzmann Machine (QBM)** allow us to use **quantum annealers** as sampling devices for the latent variables

TRIUMF

Thank you
Merci

www.triumf.ca

Follow us **@TRIUMFLab**

Discovery, accelerated

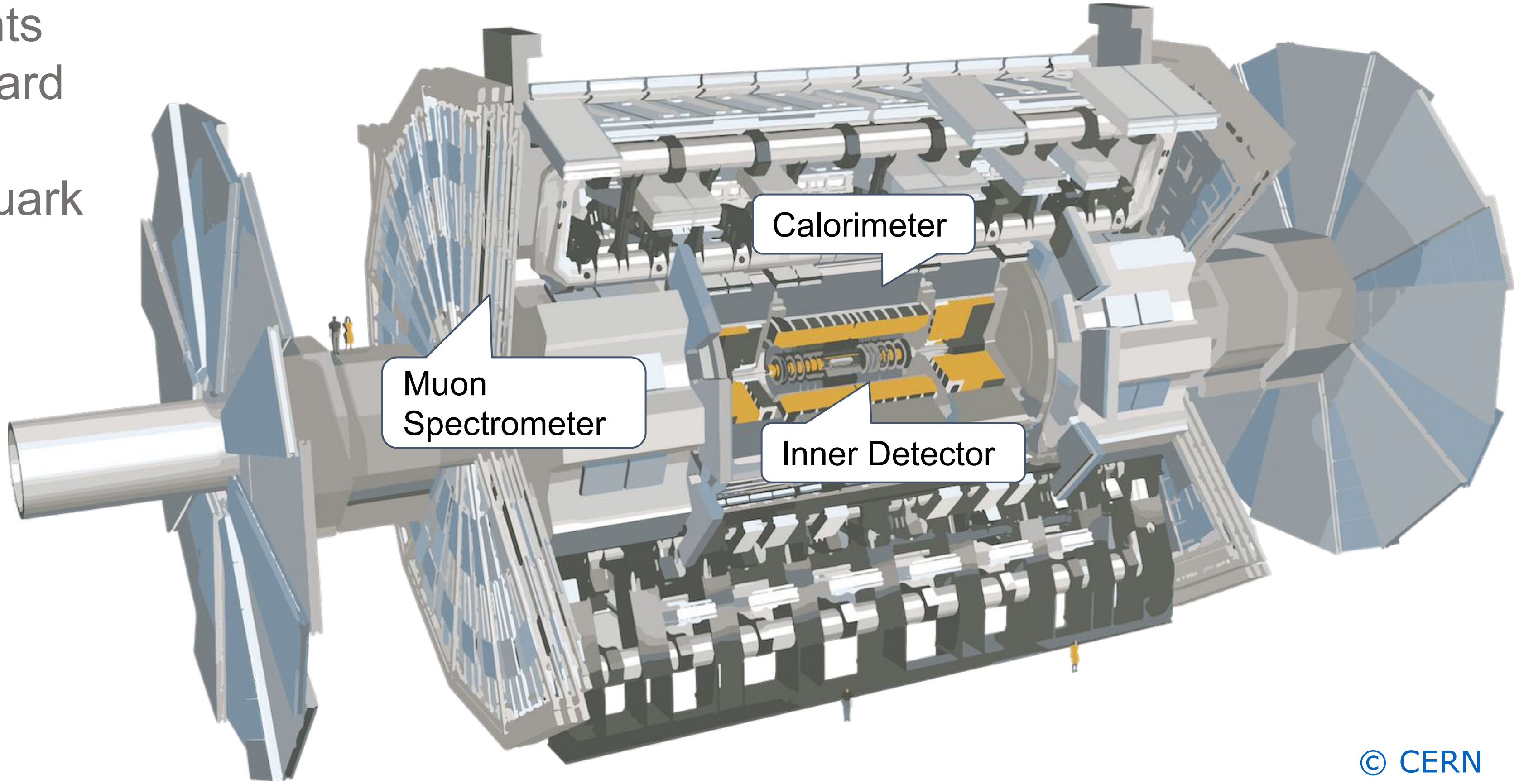
Backup

ATLAS at Large Hadron Collider

General-purpose detector

Scientific goals :

1. Higgs Boson Measurements
2. Physics beyond the Standard Model (BSM)
3. Rare processes e.g. top quark production

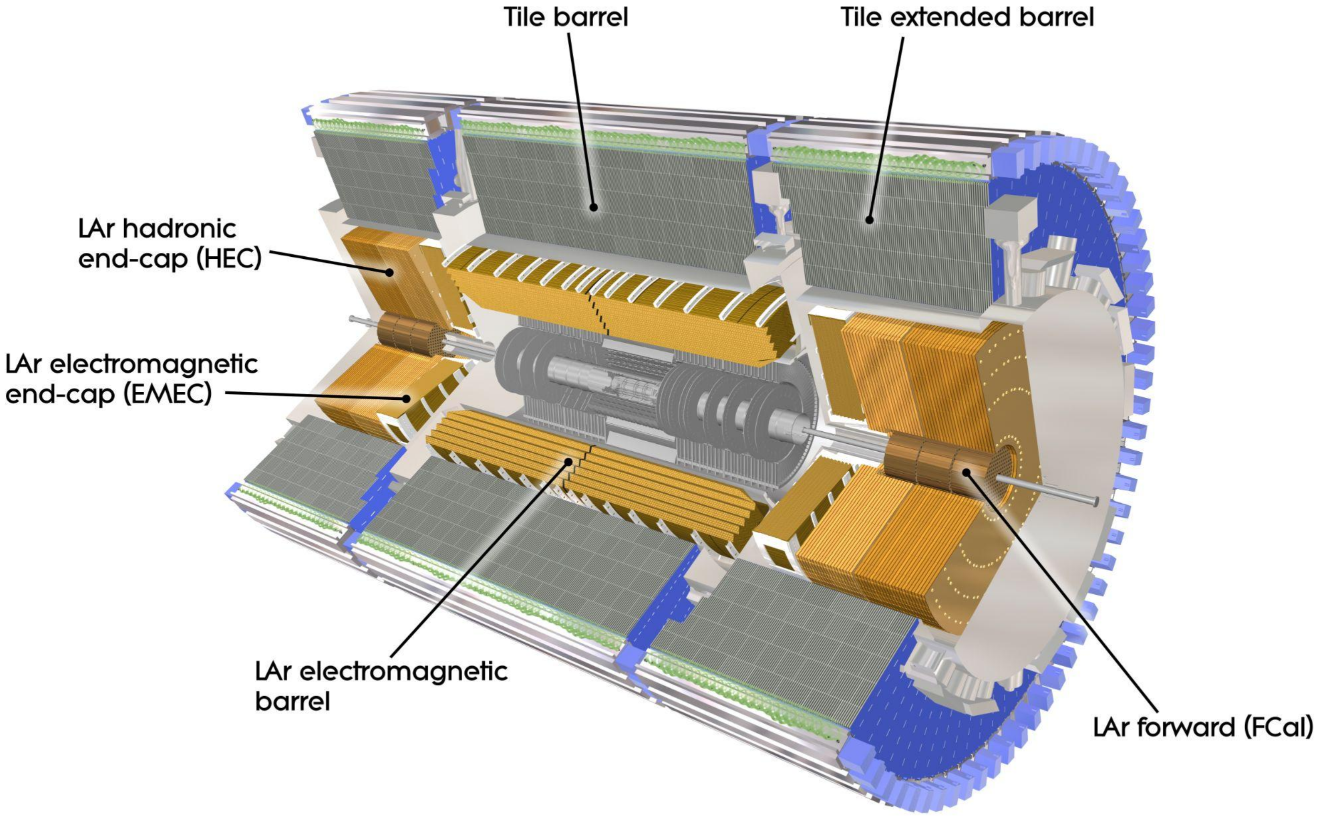


Calorimeters

Measure the energy of the particles through :

1. Electromagnetic showers
2. Hadronic showers

Cascading process



Discrete VAEs (in practice)

$$\frac{\partial}{\partial \phi}\mathbb{E}_{q(z|x,\phi)}\left[\log p(x|z,\theta)\right] \approx \frac{1}{N}\sum_{\rho\sim\mathcal{D}}\frac{\partial}{\partial\phi}\log p(x|f(x,\rho,\phi),\theta)$$

$$\mathbf{z}_i \sim \mathcal{N}(m(x,\phi), \sqrt{v(x,\phi)})$$

$$z = f(x,\phi,\rho) = m(x,\phi) + \sqrt{v(x,\phi)}.\,\rho$$
$$\rho \sim \mathcal{N}(0,1)$$

$$\mathbf{z}_i$$

$$p(z=1) = \bar{q}$$
$$\bar{q} = \sigma(l)$$

$$z = f(x,\phi,\rho) = \mathcal{H}(\rho - (1-\bar{q}(x,\phi))$$
$$= \mathcal{H}(l(x,\phi) + \sigma^{-1}(\rho)), \rho \sim \mathcal{U}(0,1)$$

Non-differentiable $\quad \frac{d\mathcal{H}(x)}{dx} = \delta(x)$

Gumbel-Softmax Trick $\quad z = f(x,\phi,\rho) = \sigma\left(\frac{l(x,\phi)+\sigma^{-1}(\rho)}{\tau}\right)$

Towards QVAE

Classical RBM

$$p_{\boldsymbol\theta}(\mathbf{z}) \equiv e^{-E_{\boldsymbol\theta}(\mathbf{z})}/Z_{\boldsymbol\theta}\,, \quad Z_{\boldsymbol\theta} \equiv \sum_{\mathbf{z}} e^{-E_{\boldsymbol\theta}(\mathbf{z})}\,,$$

$$E_{\boldsymbol\theta}(\mathbf{z}) = \sum_l z_l h_l + \sum_{l<m} W_{lm} z_l z_m, \quad \mathbf{h}, \mathbf{W} \in \{\boldsymbol\theta\}$$

Markov Chain Monte Carlo sampling

Quantum BM

$$p_{\boldsymbol\theta}(\mathbf{z}) \equiv \mathrm{Tr}[\Lambda_{\mathbf{z}} e^{-\mathcal{H}_{\boldsymbol\theta}}]/Z_{\boldsymbol\theta}\,, \quad Z_{\boldsymbol\theta} \equiv \mathrm{Tr}[e^{-\mathcal{H}_{\boldsymbol\theta}}]\,,$$

$$\mathcal{H}_{\boldsymbol\theta} = \sum_l \sigma_l^x \Gamma_l + \sum_l \sigma_l^z h_l + + \sum_{l<m} W_{lm} \sigma_l^z \sigma_m^z, \quad \boldsymbol\Gamma, \mathbf{h}, \mathbf{W} \in \{\boldsymbol\theta\}\,,$$

$$b_l = \beta^*_{eff} h_l, \quad W_{lm} = \beta^*_{eff} J_{lm}, \quad \Gamma_l = \beta^*_{eff}\Gamma^*,$$

$$\beta_{eff} \equiv B(s^*)/\beta_{phys}, \quad \Gamma^* \equiv A(s^*)/B(s^*)\,. \qquad (17)$$

Quantum Annealing

$$\mathcal{H}(s) = A(s)\sum_l \sigma_l^x + B(s)\left[\sum_l \sigma_l^z h_l + \sum_{l<m} J_{lm}\sigma_l^z\sigma_m^z\right]$$

DVAE to QVAE Challenges : Effective temperature estimation

Effective temperature β^*_{eff}

$$b_l = \beta^*_{eff} h_l, \quad W_{lm} = \beta^*_{eff} J_{lm}, \quad \Gamma_l = \beta^*_{eff} \Gamma^*,$$
$$\beta_{eff} \equiv B(s^*)/\beta_{phys}, \quad \Gamma^* \equiv A(s^*)/B(s^*)\,. \qquad (17)$$

Quantum annealer operates in a thermal environment

β^*_{eff} is not fixed

Need to perform real-time estimation

Quantum BM

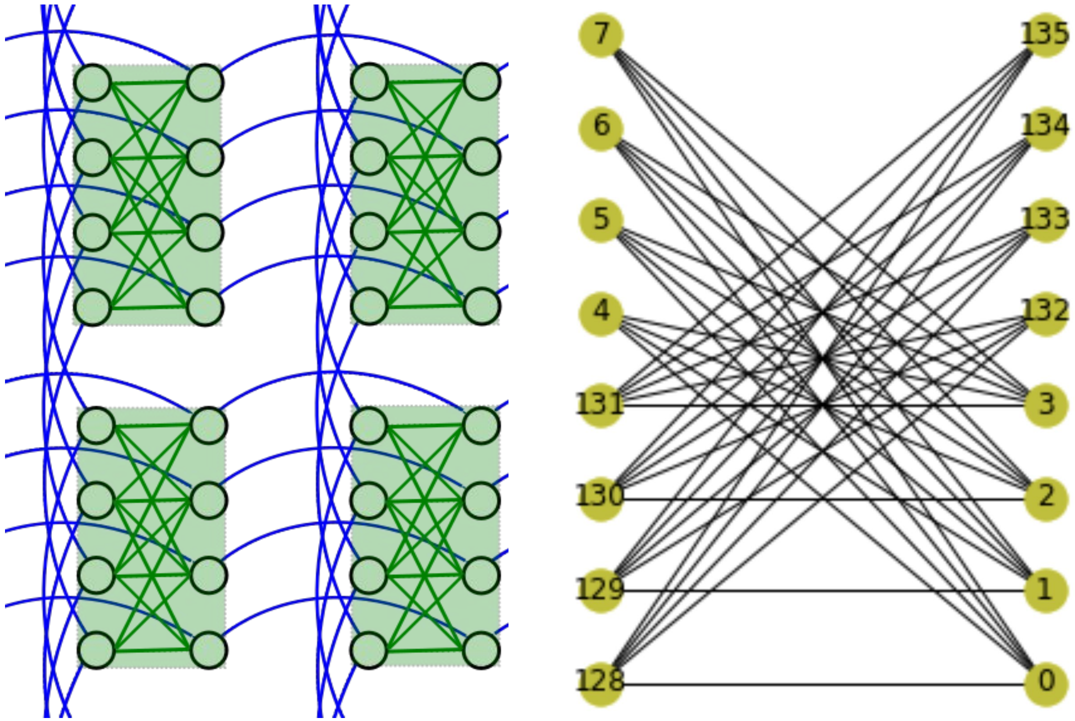
$$p_{\boldsymbol{\theta}}(\mathbf{z}) \equiv \mathrm{Tr}[\Lambda_{\mathbf{z}} e^{-\mathcal{H}_{\boldsymbol{\theta}}}]/Z_{\boldsymbol{\theta}}\,, \quad Z_{\boldsymbol{\theta}} \equiv \mathrm{Tr}[e^{-\mathcal{H}_{\boldsymbol{\theta}}}]\,,$$
$$\mathcal{H}_{\boldsymbol{\theta}} = \sum_l \sigma^x_l \Gamma_l + \sum_l \sigma^z_l h_l + \sum_{l<m} W_{lm} \sigma^z_l \sigma^z_m, \quad \boldsymbol{\Gamma}, \mathbf{h}, \mathbf{W} \in \{\boldsymbol{\theta}\}\,,$$

Quantum Annealing

$$\mathcal{H}(s) = A(s) \sum_l \sigma^x_l + B(s) \left[\sum_l \sigma^z_l h_l + \sum_{l<m} J_{lm} \sigma^z_l \sigma^z_m \right]$$

DVAE to QVAE Challenges : Qubit Connectivity

Chimera connectivity (D'Wave 2000Q)



Maximum qubit connectivity : 6

Quantum BM

$$p_{\boldsymbol{\theta}}(\mathbf{z}) \equiv \mathrm{Tr}[\Lambda_{\mathbf{z}} e^{-\mathcal{H}_{\boldsymbol{\theta}}}]/Z_{\boldsymbol{\theta}}\,, \quad Z_{\boldsymbol{\theta}} \equiv \mathrm{Tr}[e^{-\mathcal{H}_{\boldsymbol{\theta}}}]\,,$$

$$\mathcal{H}_{\boldsymbol{\theta}} = \sum_l \sigma_l^x \Gamma_l + \sum_l \sigma_l^z h_l + + \sum_{l<m} W_{lm} \sigma_l^z \sigma_m^z, \quad \boldsymbol{\Gamma}, \mathbf{h}, \mathbf{W} \in \{\boldsymbol{\theta}\}\,,$$

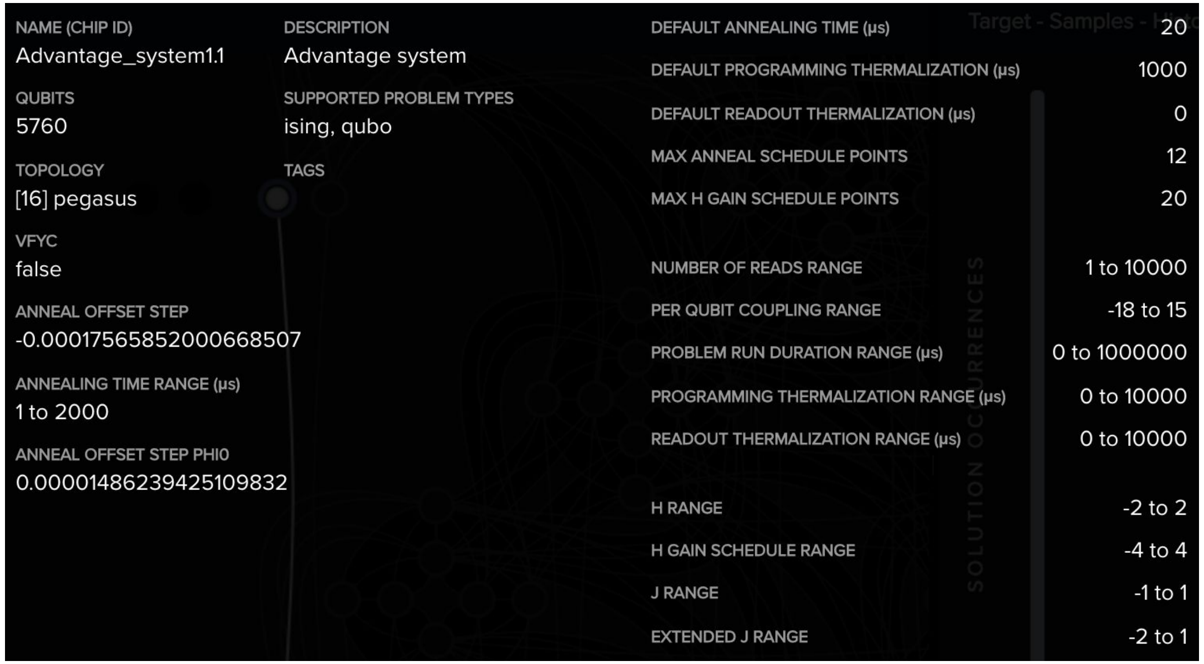
Quantum Annealing

$$\mathcal{H}(s) = A(s) \sum_l \sigma_l^x + B(s) \left[\sum_l \sigma_l^z h_l + \sum_{l<m} J_{lm} \sigma_l^z \sigma_m^z \right]$$

DVAE to QVAE Challenges : QPU Resources

	Positron e^+	**Photon γ**	**Pion π^+**
Model Type	Model II	Model IV	Model IV
Learning Rate	10^{-4}	0.5×10^{-4}	10^{-4}
Epochs	100	100	100
Batch Size	100	100	100
Gibbs Steps	50	60	50
Latent smoothing temperature, $\tau_{\mathbf{z}}$	$1/5$	$1/7$	$1/5$
Output mask smoothing temperature, $\tau_{\mathbf{x}_m}$	$1/5$	$1/5$	$1/9$

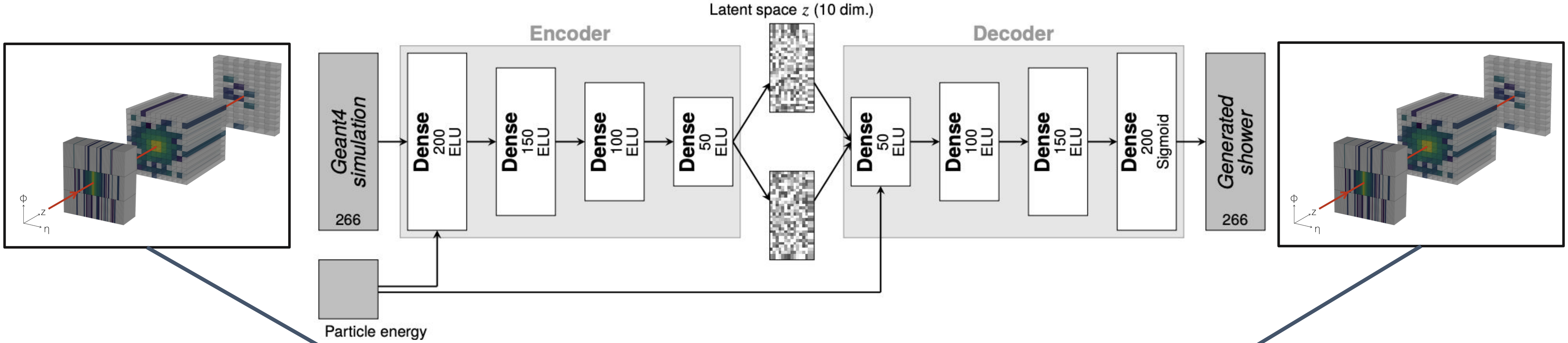
Parameter	Value
Model	Advantage
Graph Size	P16
Qubits	5436
Couplers	37440
Qubit Temperature (mK)	15.8 ± 0.5
M_{AFM}[1] (pH)	1.951
Average Single Qubit Thermal Width (Ising units)	0.196
FM Problem Freezeout (scaled time)	0.066
Single Qubit Freezeout (scaled time)	0.609
Annealing Time Range (μs)	1.0 to 2000.0
Readout Time Range[2] (μs)	18.0 to 131.0
Programming Time[3] (μs)	$\sim$ 25100
Readout Error Rate[4]	≤ 0.001



Per batch : 25.1 ms (Programming time) + [0.02 (Default annealing time) + 0.131 (Readout time)]*100 (Batchsize) ms = 40.2 ms (no delay)
100000 events / particle type = 1000 batches

Per epoch : 40.2 s (no delay)
Per particle type : 4020 s (no delay)

Previous Work : DGMs for Fast Calorimeter Simulation

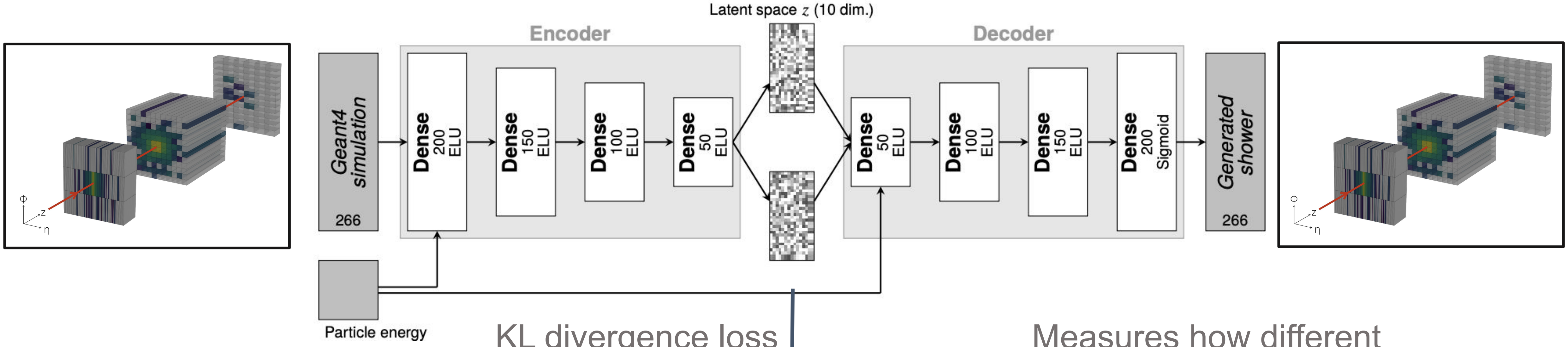


Autoencoding loss

$$\mathrm{Loss}(\mathbf{x}, \mathbf{x'}), \mathrm{e.g.}\ \frac{1}{n}\Sigma_{i=1}^{n}(\mathbf{x_i} - \mathbf{x'_i})^2$$

[1] Paganini, de Oliveira and Nachman (2018), arXiv:1712.10321 [2] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001.

Previous Work : DGMs for Fast Calorimeter Simulation



KL divergence loss

$$D_{KL}[q_\phi(\mathbf{z}|\mathbf{x}, e)||p(\mathbf{z})]$$

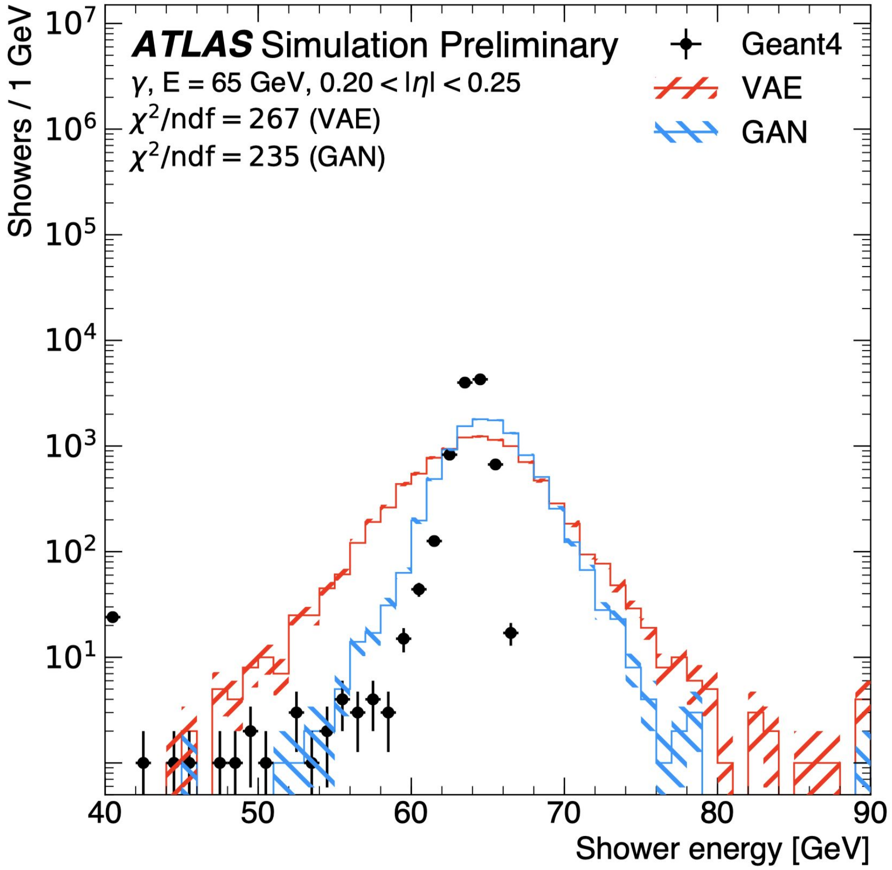
Measures how different is one probability distribution from another

$$z_i \sim \mathcal{N}(\mu_i(\mathbf{x}, e), \sigma_i^2(\mathbf{x}, e)) \qquad \mathbf{z}_i \sim \mathcal{N}(0, 1)$$

[1] Paganini, de Oliveira and Nachman (2018), arXiv:1712.10321 [2] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001.

Previous Work : DGMs for Fast Calorimeter Simulation

ATLAS-VAE/GAN [1]

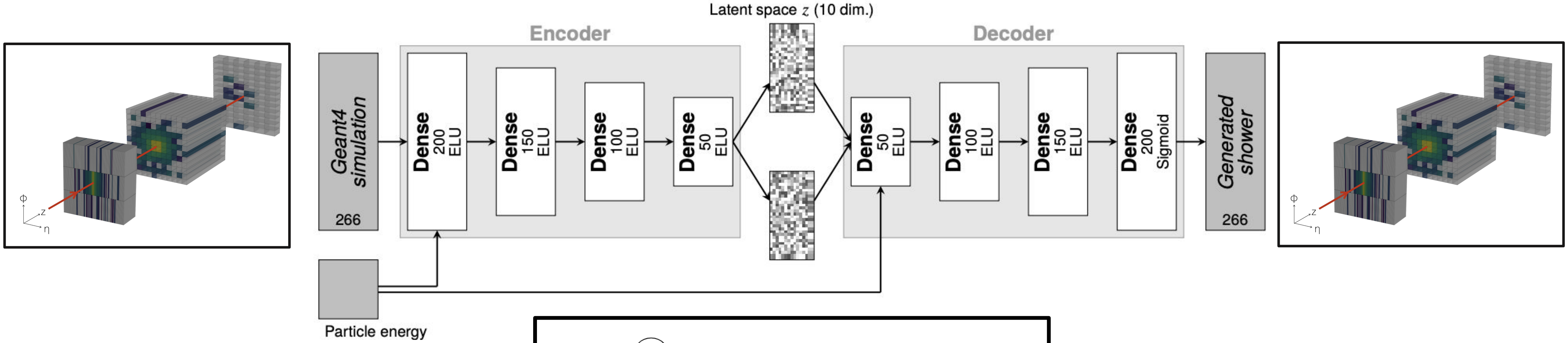


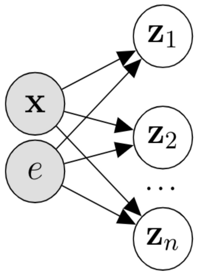
- Shape of the energy spectra of samples produced by DGMs matches
- No explicit energy conservation - DGMs produce showers with higher energy than incident particle energy

[1] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001.

Previous Work : DGMs for Fast Calorimeter Simulation

CaloGAN [1], ATLAS-VAE/GAN [2]

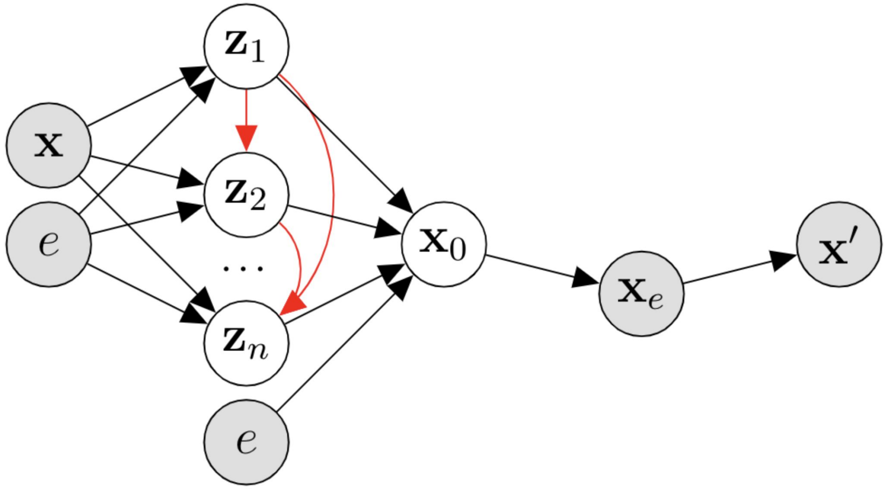




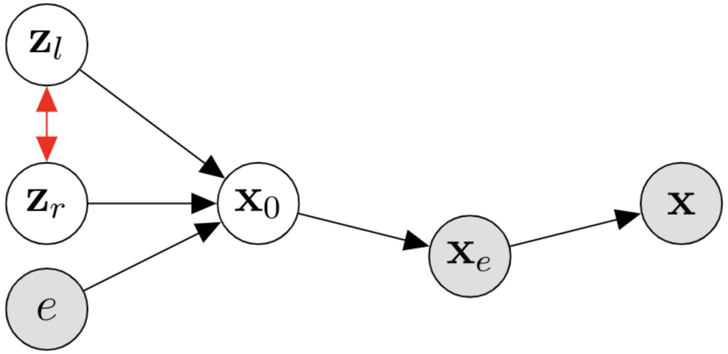
$$z_i \sim \mathcal{N}(\mu_i(\mathbf{x}, e), \sigma_i^2(\mathbf{x}, e))$$

[1] Paganini, de Oliveira and Nachman (2018), arXiv:1712.10321 [2] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2018-001.

DVAEs for Calorimeter Simulation : Preliminary

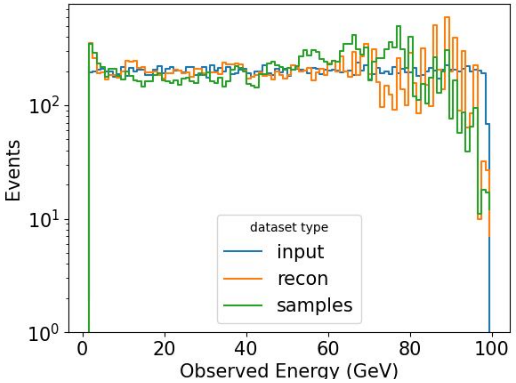


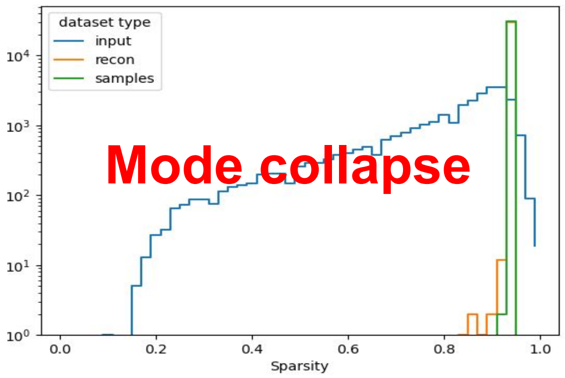
Autoencoding model

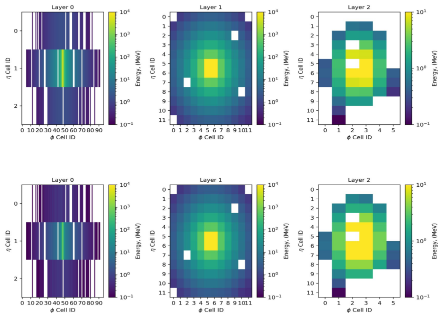


Generative model

e^+

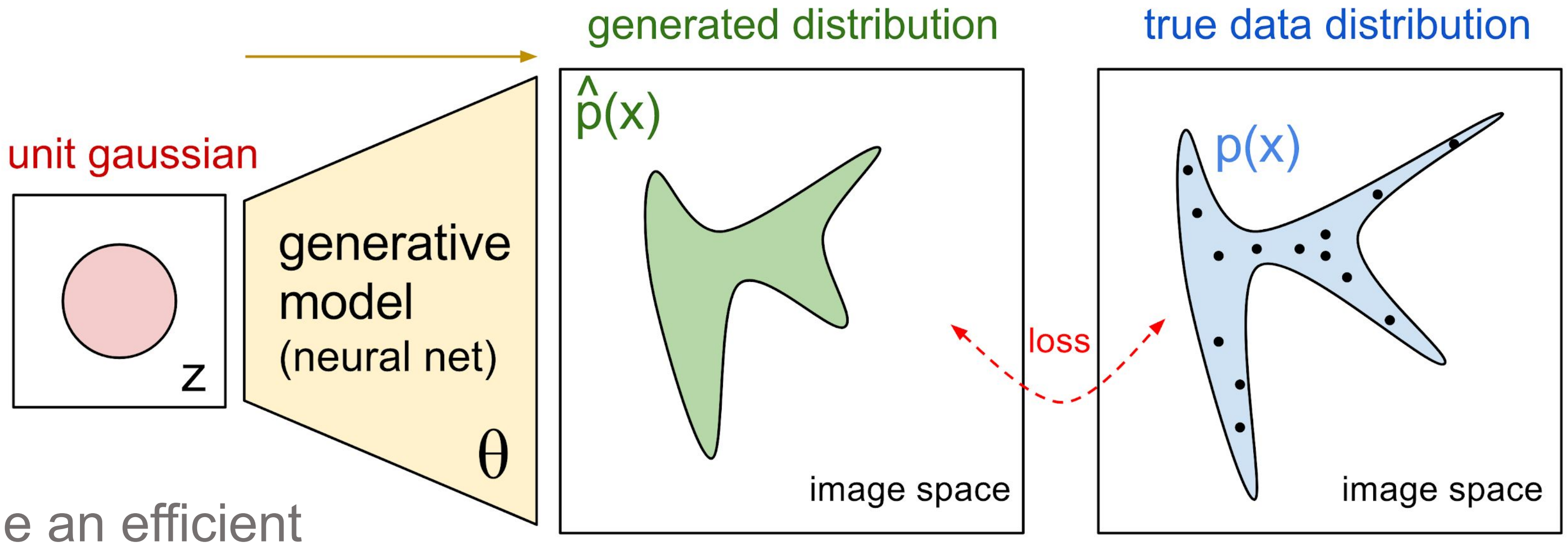






Generative Models

1. Approximate true data distribution in a high-dimensional space → Maximize the log-likelihood of the data under the model distribution



2. Provide an efficient sampling method to sample from the model distribution

[1] https://openai.com/blog/generative-models/