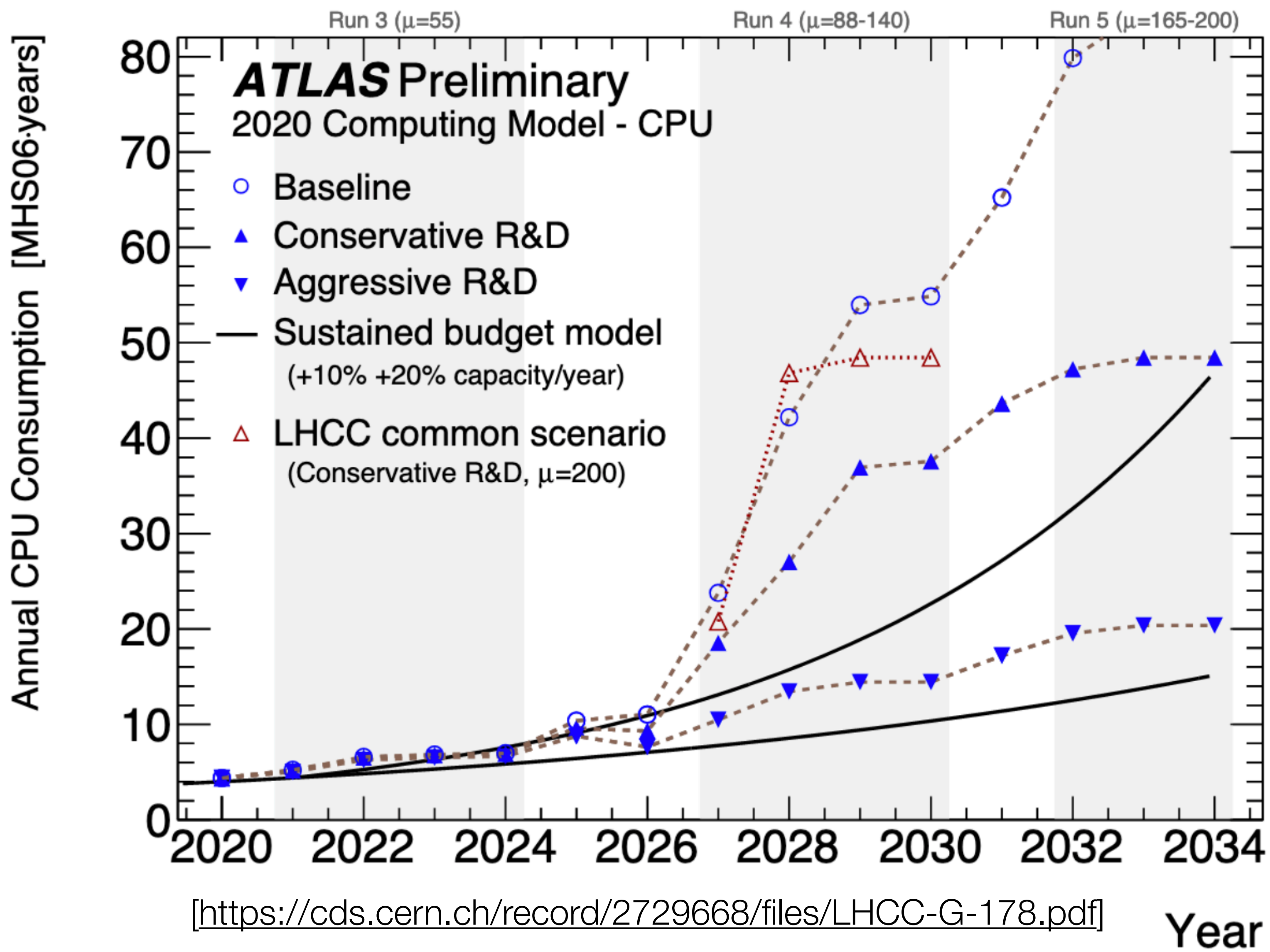
Quantum-assisted Machine Learning for ATLAS

Tiago Vale
TRIUMF Science Week
1st August 2023

Generative models: why bother

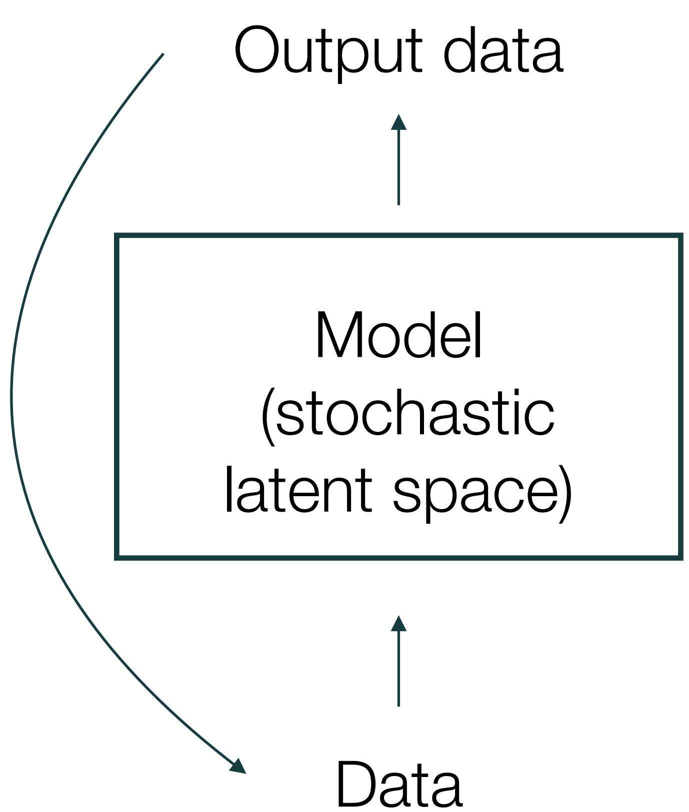
- Modelling of particle showers in **calorimeters** is the most demanding part of particle physics simulation
  - Can take minutes per event in state of the art platforms [1005.4568]
  - Not sustainable with increase in pileup



[https://cds.cern.ch/record/2729668/files/LHCC-G-178.pdf]

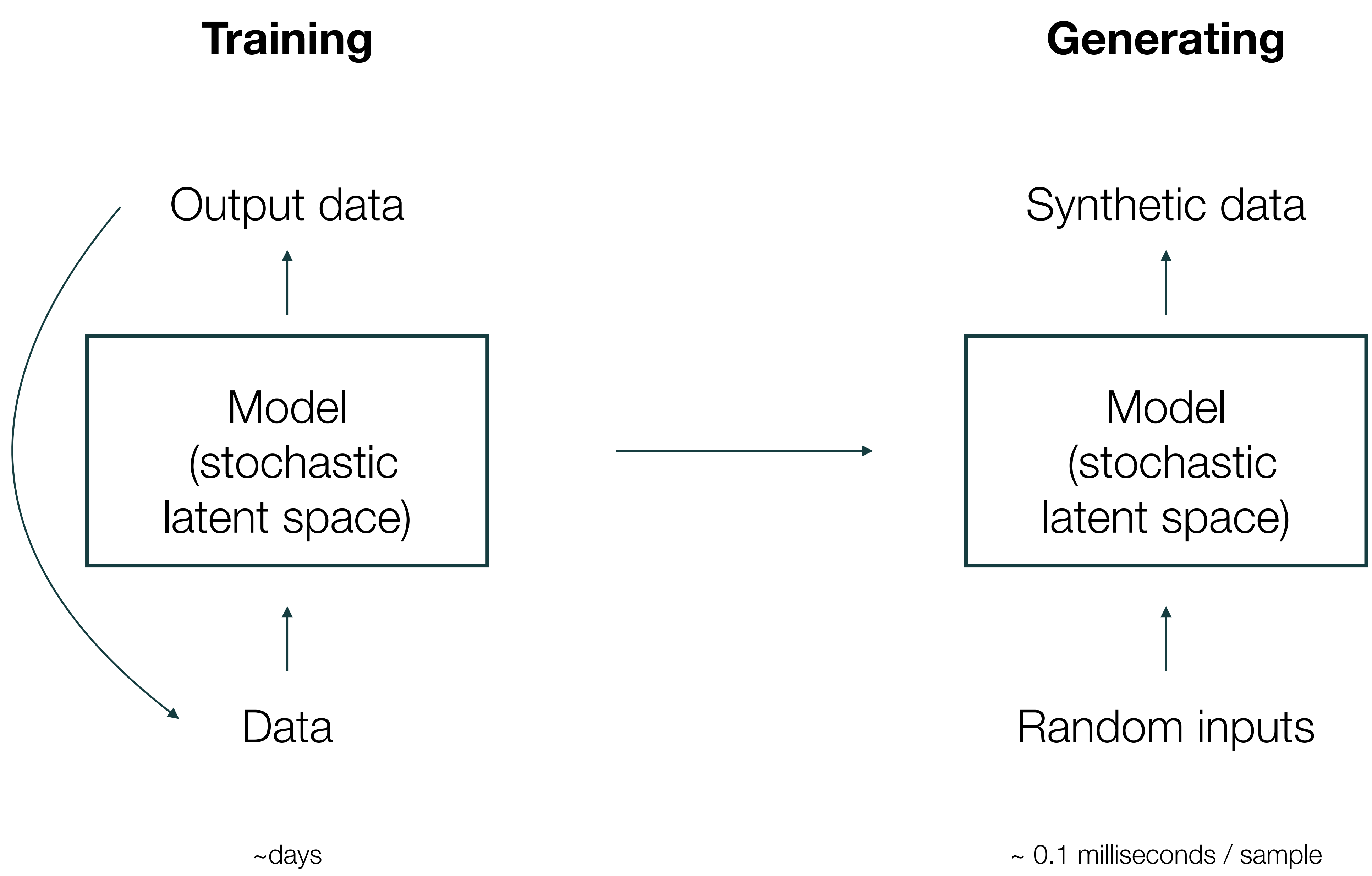
Generative models

Training



~days

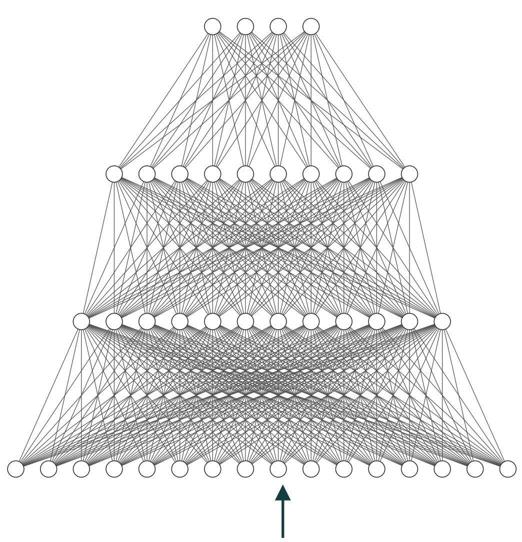
Generative models



Variational Autoencoders

Data (x)

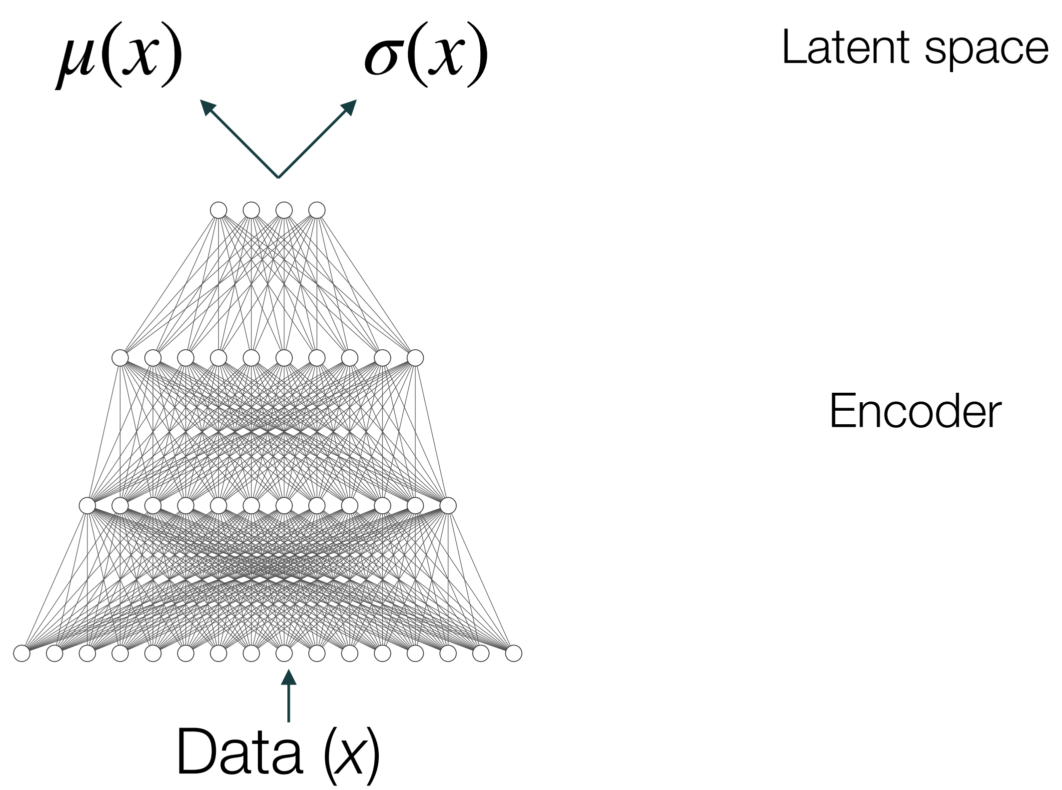
Variational Autoencoders



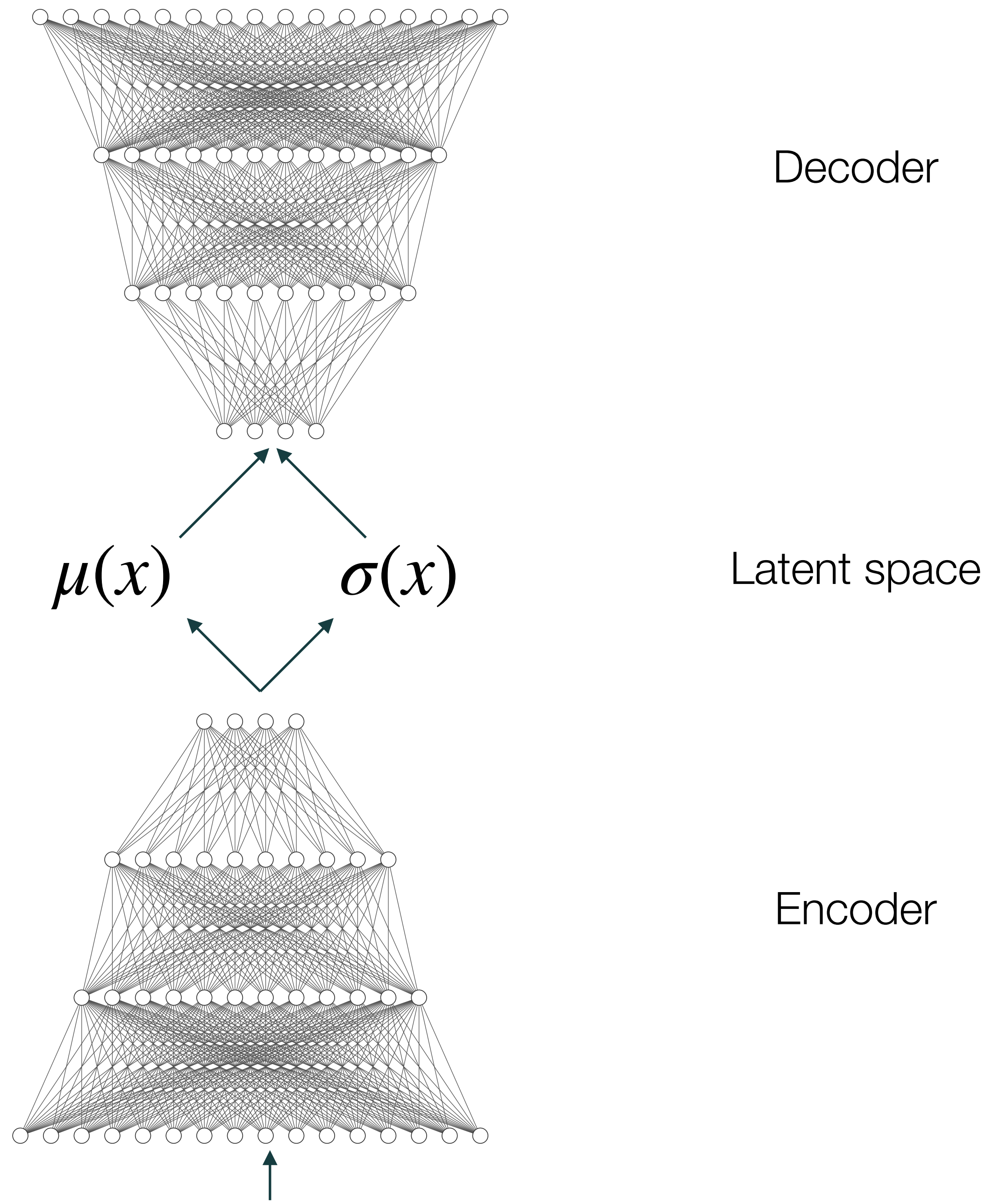
Encoder

Data (x)

Variational Autoencoders

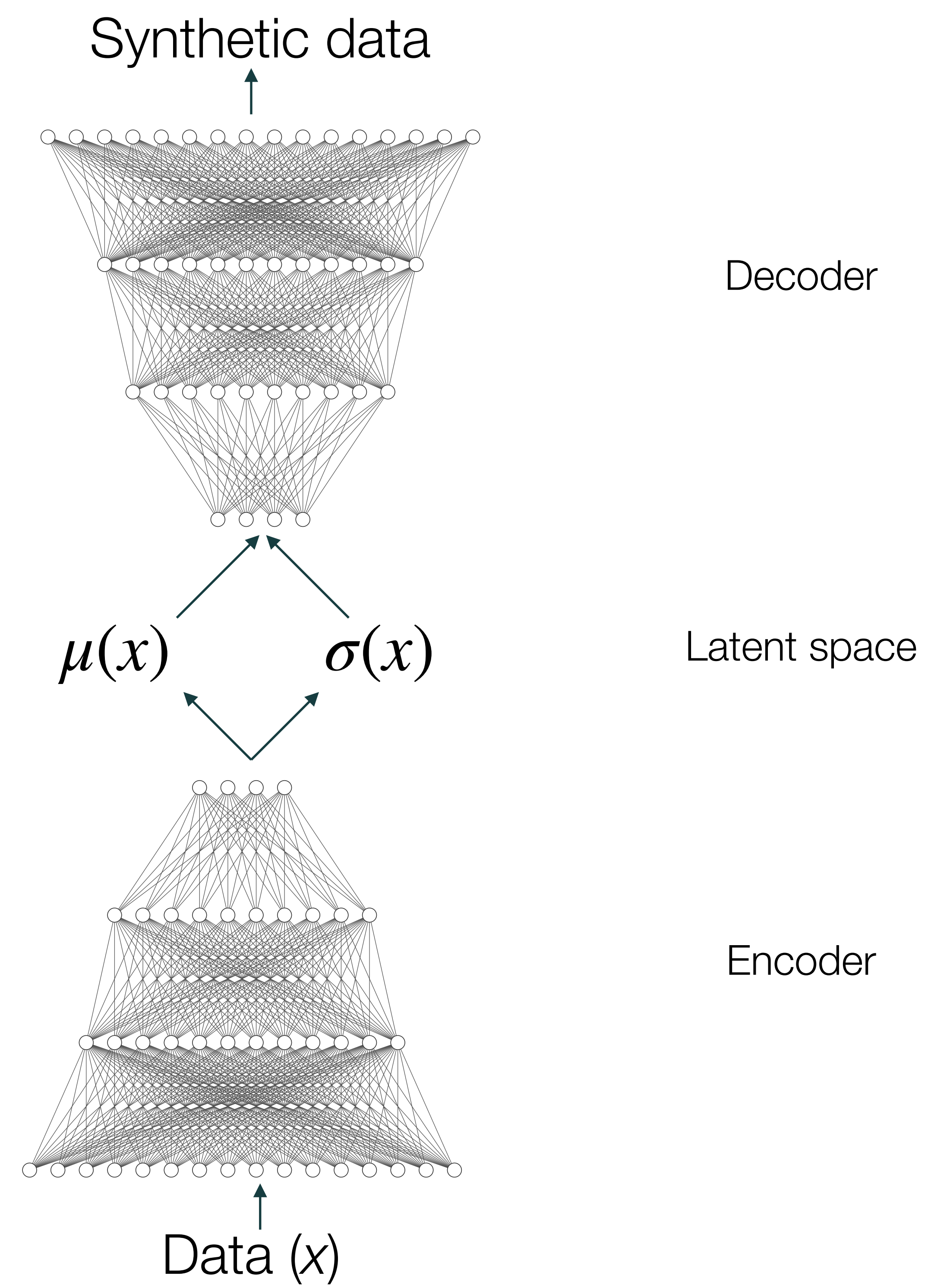


Variational Autoencoders



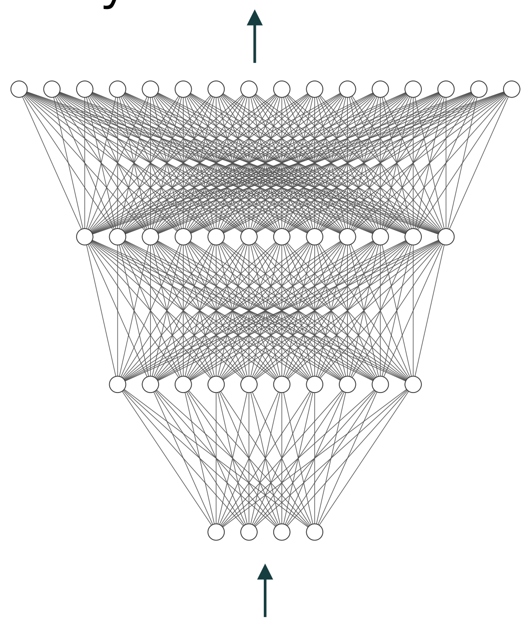
Data (x)

Variational Autoencoders



Variational Autoencoders

Synthetic data



$z \sim N(0, I)$

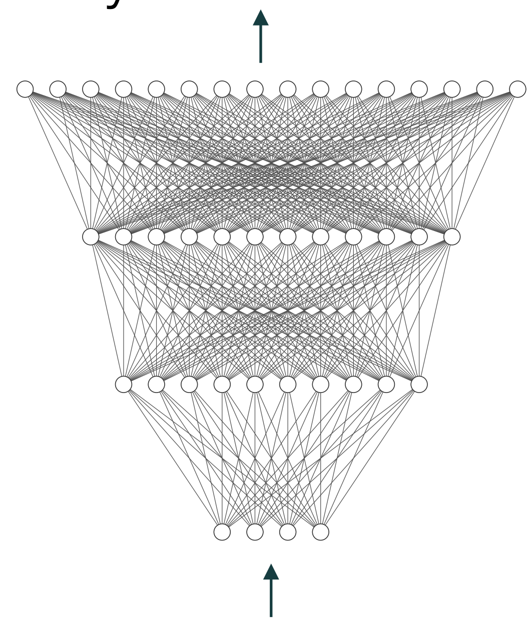
Decoder

Latent space

Factorized Gaussian

Variational Autoencoders

Synthetic data



Decoder

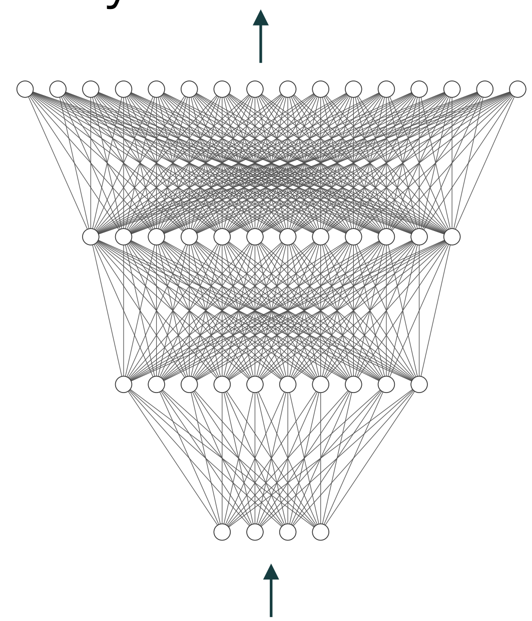
$z \sim N(0, I)$

Latent space

Limited expressivity ← Factorized Gaussian

Variational Autoencoders

Synthetic data



Decoder

$z \sim N(0,I)$

Latent space

Factorized Gaussian

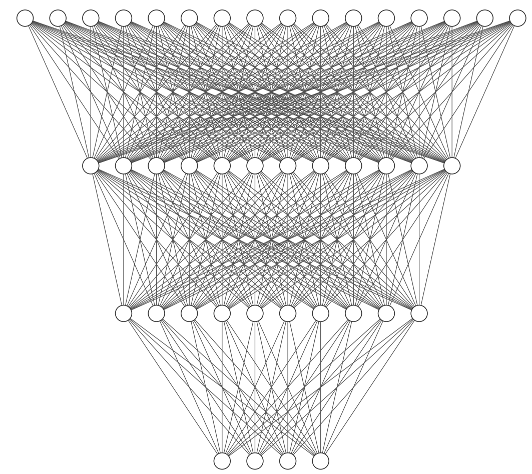
Limited expressivity ←

↓

Use a trainable model
to encapsulate latent space

Variational Autoencoders

Synthetic data



$z \sim N(0, I)$

Decoder

Latent space

Limited expressivity ← Factorized Gaussian

↓

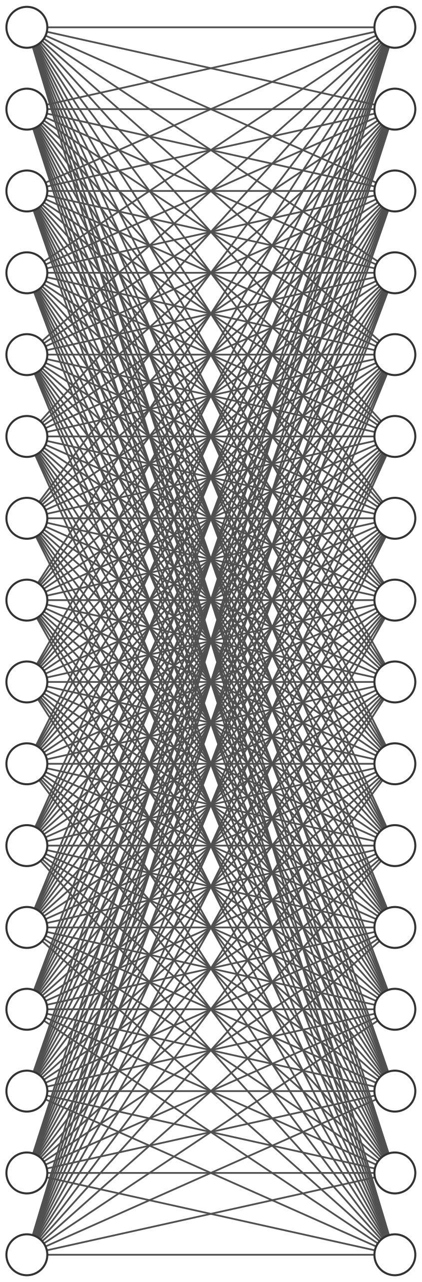
Use a trainable model
to encapsulate latent space

↓

Restricted Boltzmann Machine

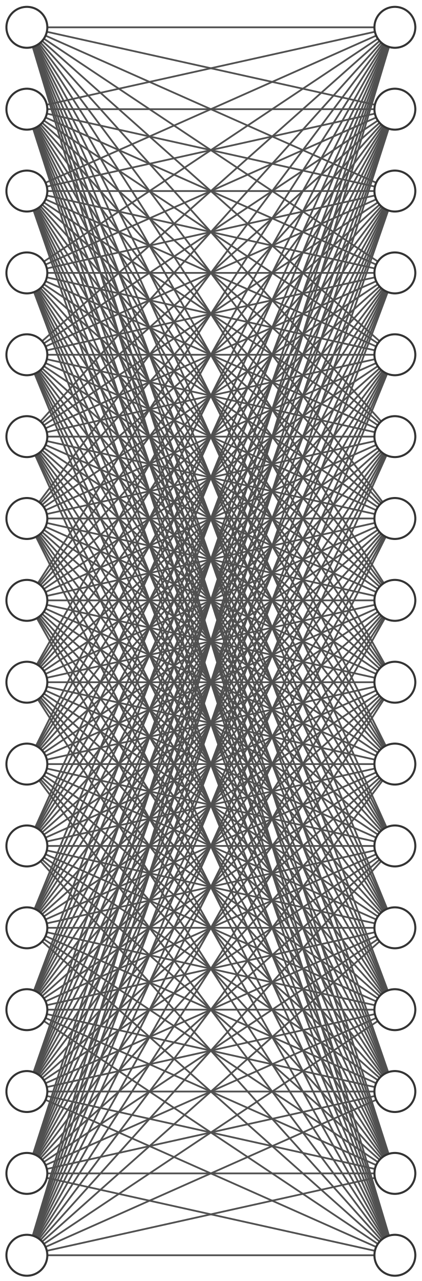
Restricted Boltzmann Machines

- Neural network that can learn a probability distribution
- **Boolean** inputs forming a bipartite graph (restricted)



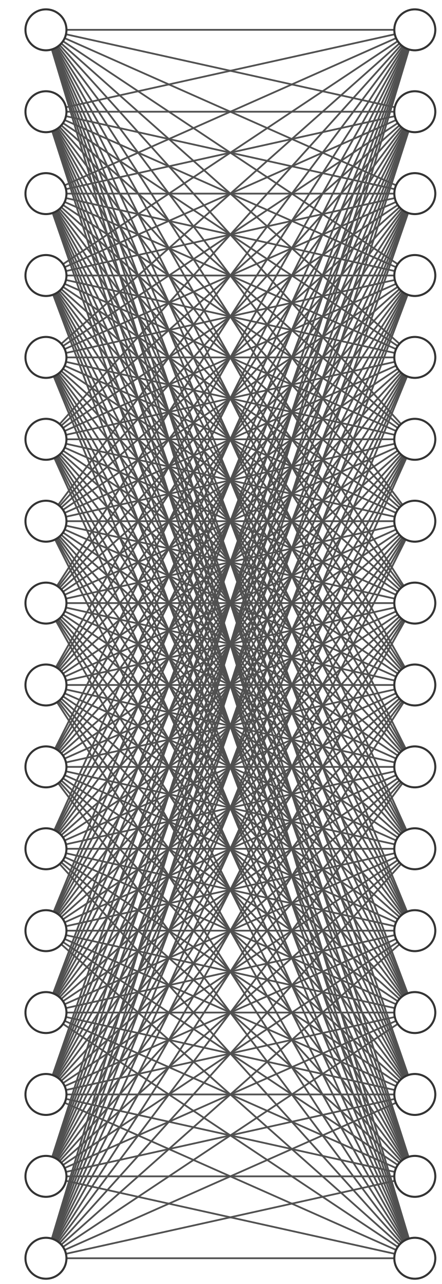
Restricted Boltzmann Machines

- Neural network that can learn a probability distribution
- **Boolean** inputs forming a bipartite graph (restricted)
- $$E(v, h) = -\sum_i \sum_j w_{ij} v_i h_j - \sum_i a_i v_i - \sum b_j h_j$$
- $$p(v, h) = \frac{1}{Z} e^{-E(v,h)}$$
- $$Z = \sum_{v,h} e^{-E(v,h)}$$



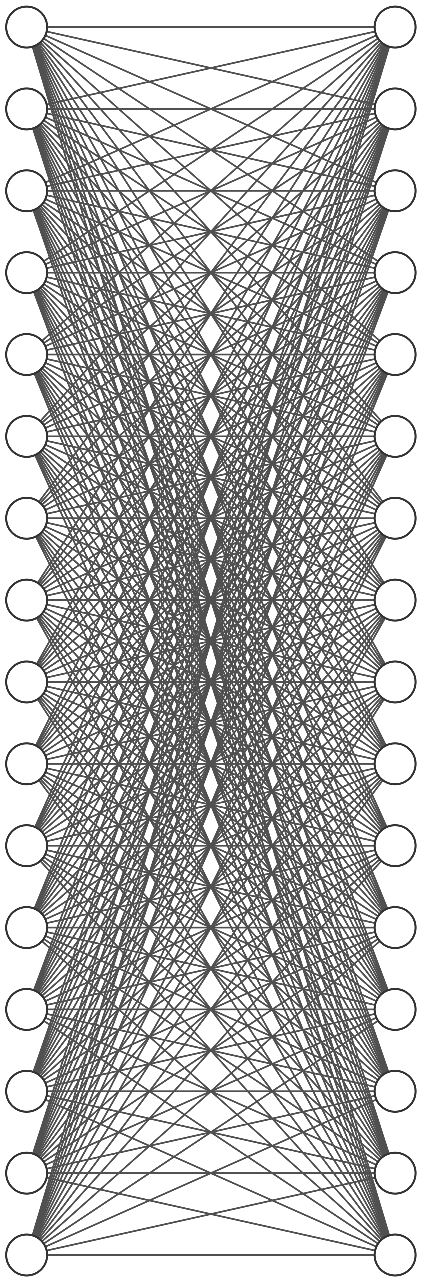
Restricted Boltzmann Machines

- Neural network that can learn a probability distribution
- **Boolean** inputs forming a bipartite graph (restricted)
- $E(v, h) = -\sum_i \sum_j w_{ij} v_i h_j - \sum_i a_i v_i - \sum b_j h_j$
- $p(v, h) = \frac{1}{Z} e^{-E(v,h)}$
- $Z = \sum_{v,h} e^{-E(v,h)}$
- Discrete input variables → latent space not differentiable



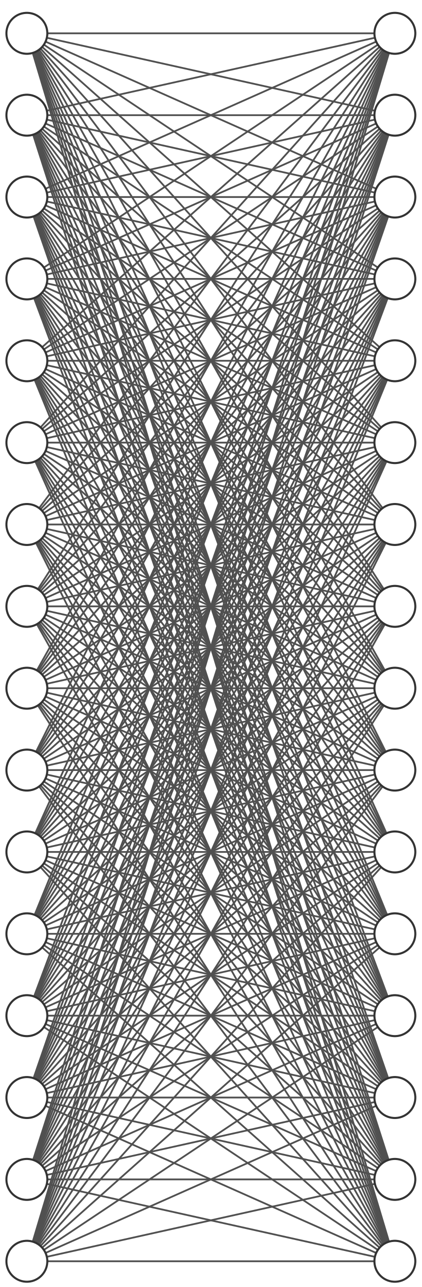
Restricted Boltzmann Machines

- Neural network that can learn a probability distribution
- **Boolean** inputs forming a bipartite graph (restricted)
- $$E(v, h) = -\sum_i \sum_j w_{ij} v_i h_j - \sum_i a_i v_i - \sum b_j h_j$$
- $$p(v, h) = \frac{1}{Z} e^{-E(v,h)}$$
- $$Z = \sum_{v,h} e^{-E(v,h)}$$
- Discrete input variables → latent space not differentiable
- $$z \to \zeta = \sigma\left(\frac{l + \sigma^{-1}(\rho)}{\tau}\right)$$ [1805.07349]



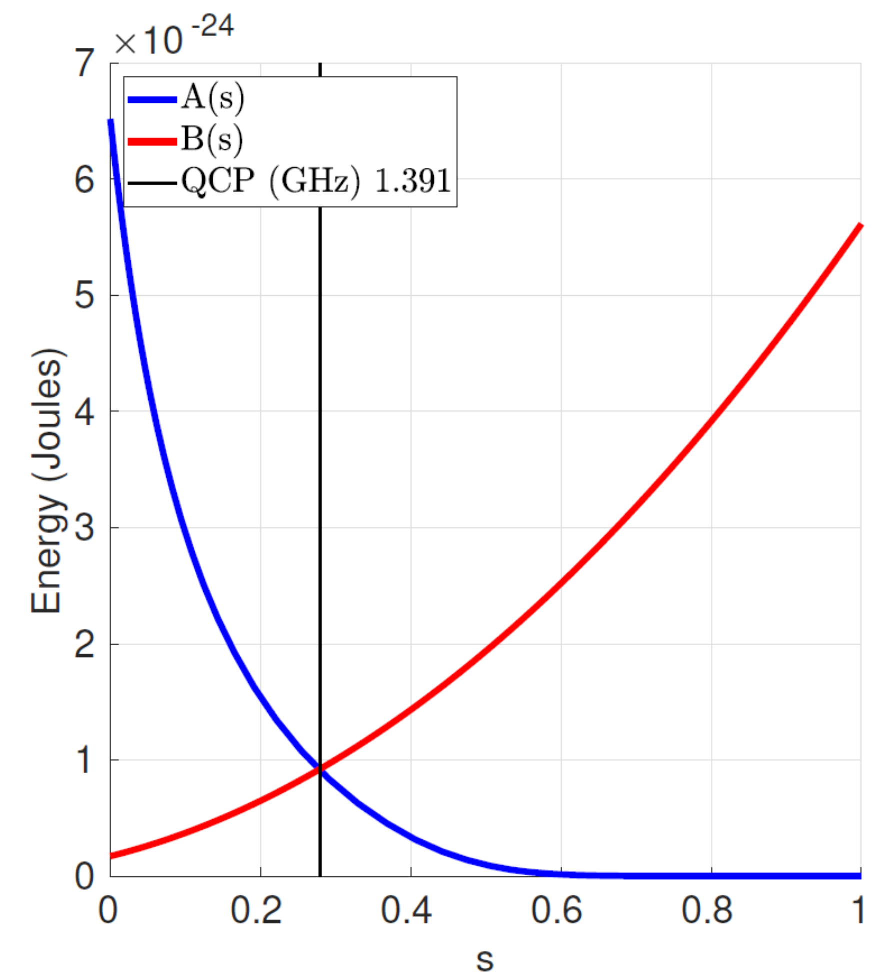
Restricted Boltzmann Machines

- Neural network that can learn a probability distribution
- **Boolean** inputs forming a bipartite graph (restricted)
- $E(v,h) = -\sum_i \sum_j w_{ij} v_i h_j - \sum_i a_i v_i - \sum b_j h_j$
- $p(v,h) = \frac{1}{Z} e^{-E(v,h)}$
- $Z = \sum_{v,h} e^{-E(v,h)}$
- Discrete input variables → latent space not differentiable
- $z \to \zeta = \sigma \left(\frac{l + \sigma^{-1}(\rho)}{\tau} \right)$ [1805.07349]
- Computationally expensive sampling → Quantum computers!



Quantum annealing

- Using D-Wave annealing processors
- Prepare qubits in a superposition state
- Control energy of each state via external magnetic field
- Configure coupling strength and sign via couplers
- $$H_{Ising} = -\frac{A(s)}{2}\left(\sum_i \sigma_x^{(i)}\right) + \frac{B(s)}{2}\left(\sum_i h_i \sigma_z^{(i)} + \sum_{i>j} J_{ij}\sigma_z^{(i)}\sigma_z^{(j)}\right)$$
- Start with $A(0) \gg B(0)$, slowly anneal to $A(1) \ll B(1)$



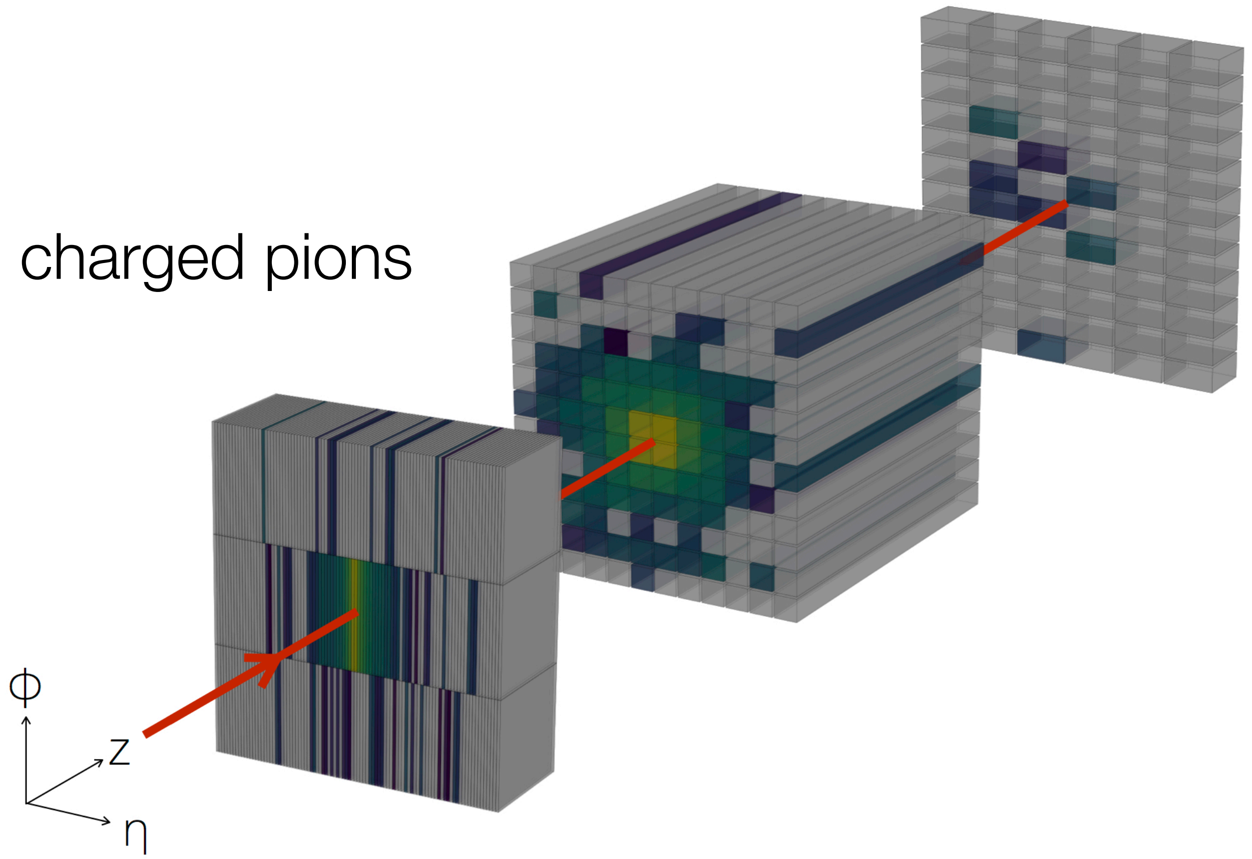
[https://docs.dwavesys.com/docs/latest/c_gs_2.html]

Quantum RBM

- Natural mapping of boolean nodes to quantum bits [1601.02036]
- Translate RBM nodes to an Ising Hamiltonian
- Slow RBM sampling from CPU becomes an annealing process [1912.02119]
- Sample each event in the order of milliseconds
- Quantum computer as the sampler!

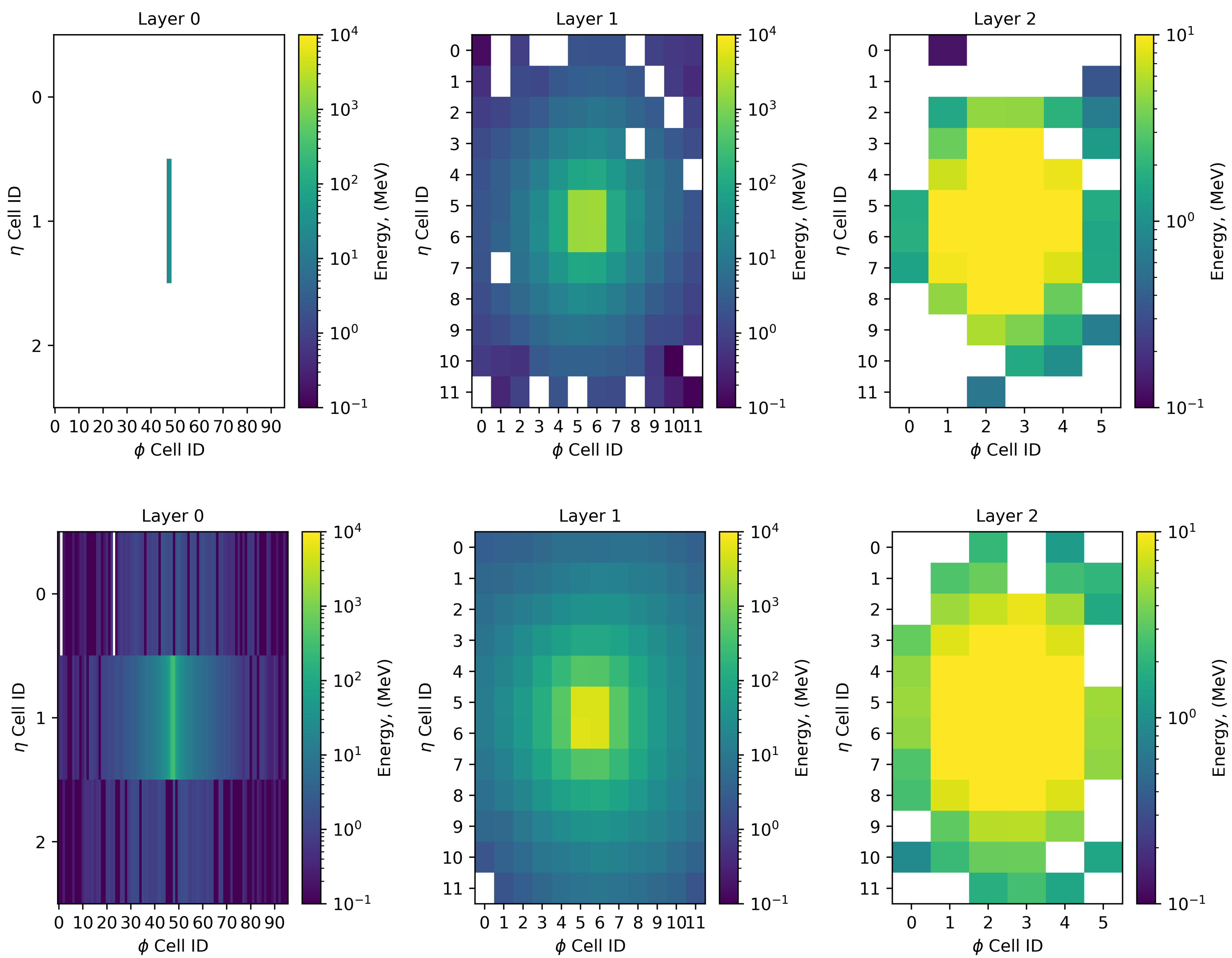
ATLAS like Calorimeter dataset

- Dataset from a pioneering paper that uses GANs (CaloGAN) [1712.10321]
- Simplified version of the ATLAS LAr calorimeter
- 3 calorimeter layers with different depth and granularity
- Generated with GEANT4
- 3 particle types: positrons, photons and charged pions
- 100k events per particle type



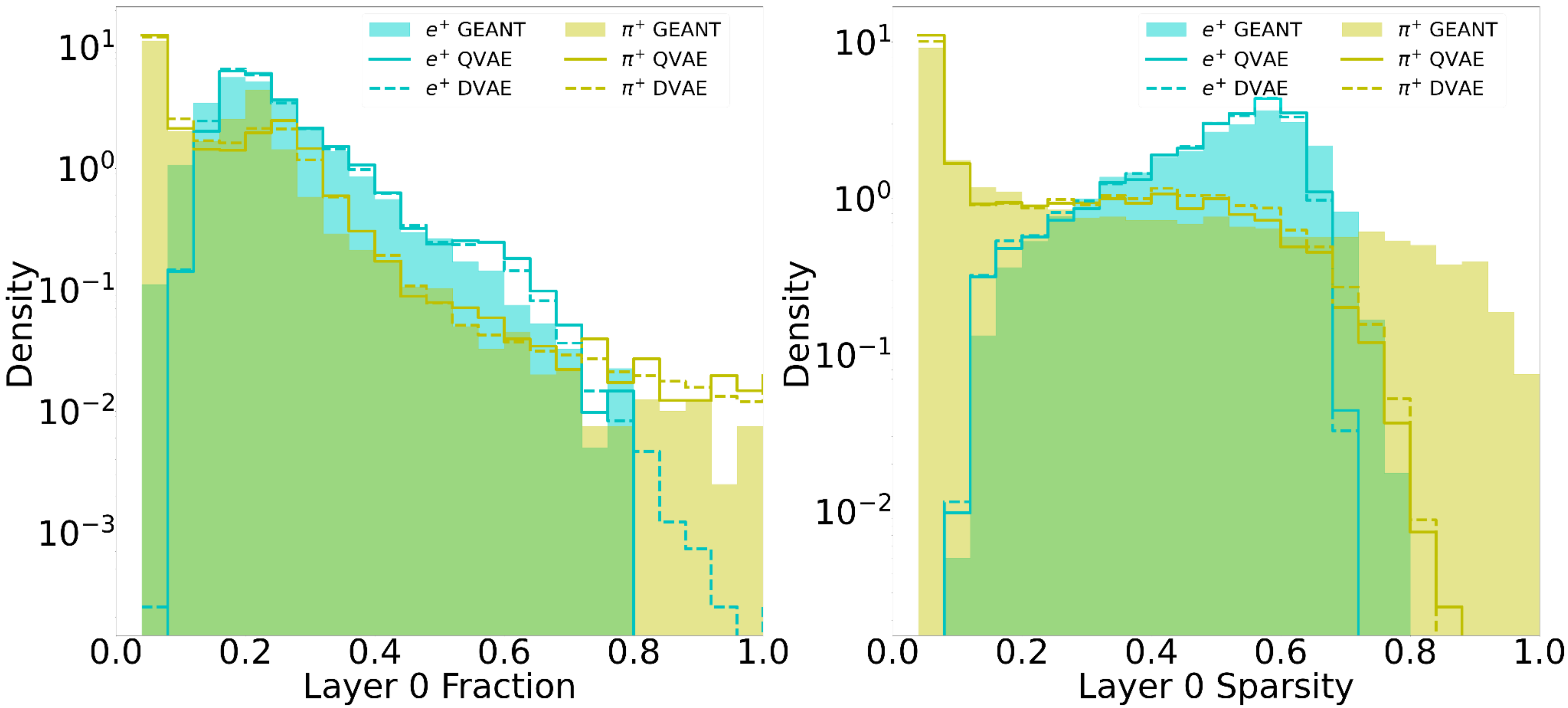
Shower images

Generated with QPU!



Pions

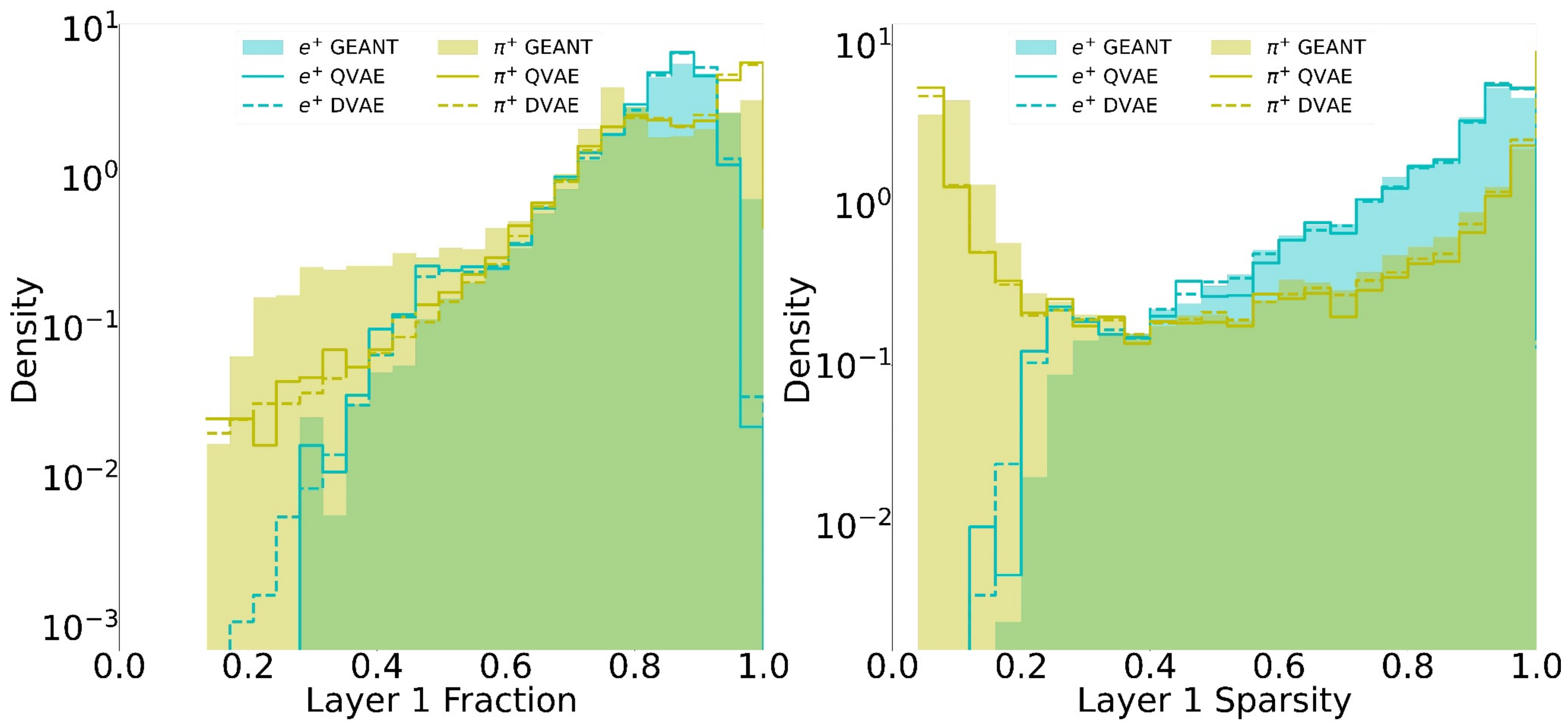
Shower distributions



$$f_i = \frac{E_i}{E_{total}}$$

$$S_i = \frac{N_{OccupiedCells}}{N_{total}}$$

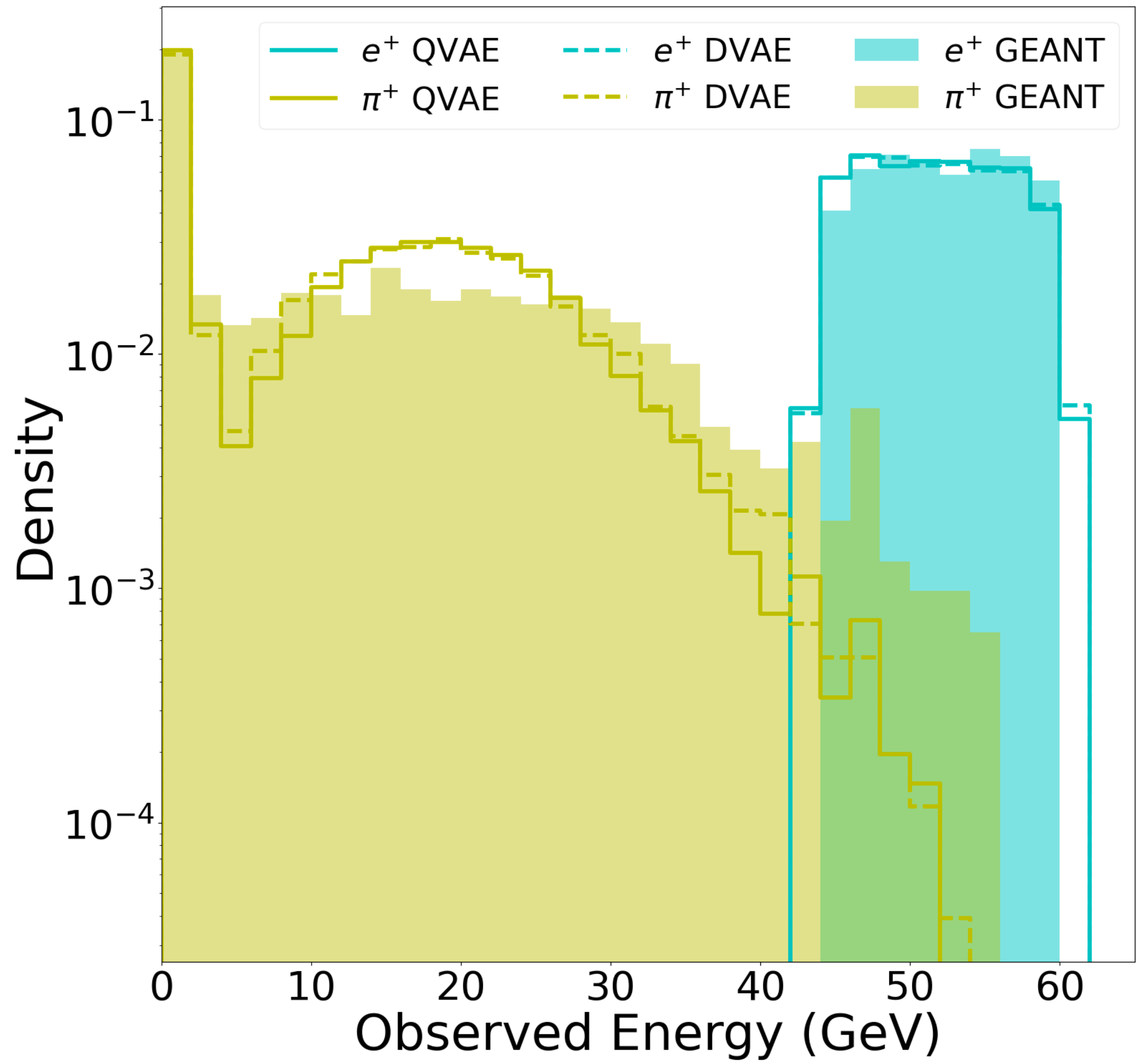
Shower distributions



$$f_i = \frac{E_i}{E_{total}}$$

$$S_i = \frac{N_{OccupiedCells}}{N_{total}}$$

Shower distributions



Energy conditioning

The future

- Train and sample on QPUs
- Use ATLAS data and MC
- Model improvements
- Move to jets/full events

Thank you!

Quantum Boltzmann Machine

- Translate RBM parameters to Ising parameters [1601.02036]

$$p(v, h) = \frac{1}{Z} e^{-E(v,h)}$$

$$\downarrow$$

$$p(z) = \frac{1}{Z} \mathrm{Tr}[\Lambda_z e^{-H}]$$

$$E(v, h) = -\sum_i \sum_j w_{ij} v_i h_j - \sum_i a_i v_i - \sum b_j h_j$$

$$\downarrow$$

$$H = \sum_{i<m} w_{im} \sigma_i^z \sigma_m^z + \sum_i \sigma_i^x \Gamma_i + \sum_i \sigma_i^z h_i$$

5.4 Run 4 R&D

- **Baseline**: Events would be simulated full GEANT4 Simulation or fast simulation (primarily parametrized calorimeter response). Most likely the GEANT4 version would be updated at the start of Run 4. Digitization would continue to be done using MC+MC Overlay, but high memory queues would be required to produce the pre-mixed pile-up RDO files.
- **Conservative R&D**: Fast Simulation would be the default simulation method. Static compilation of Athena code with GEANT4 dependencies against GEANT4 will be implemented. AthenaMT

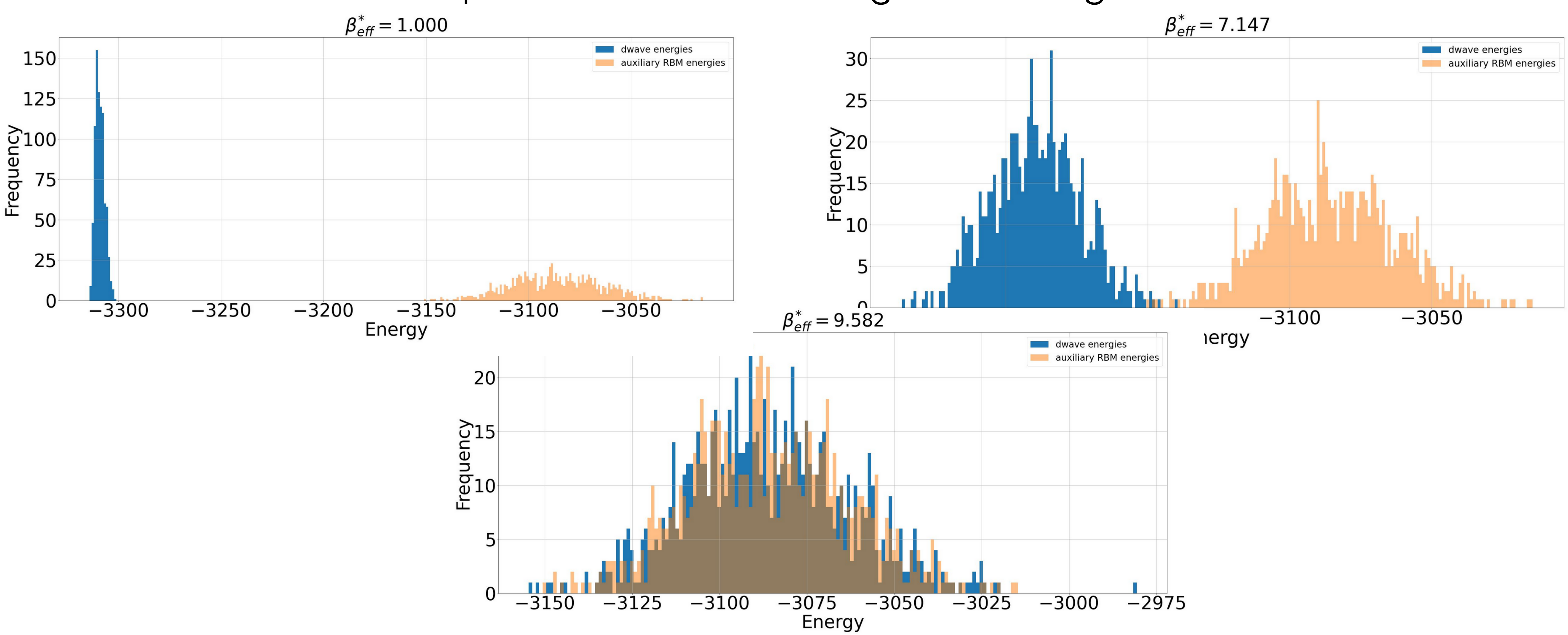
compatible pile-up digitization will allow grid resources to be used more efficiently for the pre-mixed pile-up RDO production step.
- **Aggressive R&D**: Substantial speed-ups in GEANT4 are found. FatRas would be available as an option to simulate particles in the inner tracker. Running simulation and digitization in a single Athena job and running EVNT to AOD in a single production step would be possible.

Run 4 analyses will need significantly more simulated events than in previous runs (see §12.2). ATLAS is undertaking a major simulation software R&D programme to speed up the MC production chain:

- A simulation based on the ACTS-based FatRas for the Inner Tracker and FastCaloSim for the calorimeter will make the simulation time small compared to the reconstruction time. Further speed-ups to the MC production workflow will require the reduction of the total digitization and reconstruction time.
- Using Trigger-like algorithms to filter events prior to reconstruction, so that events which will never be used in analyses are not reconstructed (or written out), could save CPU and disk space.
- Skipping reconstruction algorithms that are not needed for some MC sample production could also save CPU.
- Using MC generator information or simulation information augmented with parametrizations could speed up parts of digitization and reconstruction algorithms.
- Another idea involves using a strategy similar to MC Overlay, but reconstructing the pile-up tracks in separate job, to produce special RDO files containing pile-up tracks. These tracks could be copied through the overlay step. Reconstruction would then consist of running tracking for the hard-scatter, then combining the track collections and using the merged track collection as the input to the rest of the reconstruction.

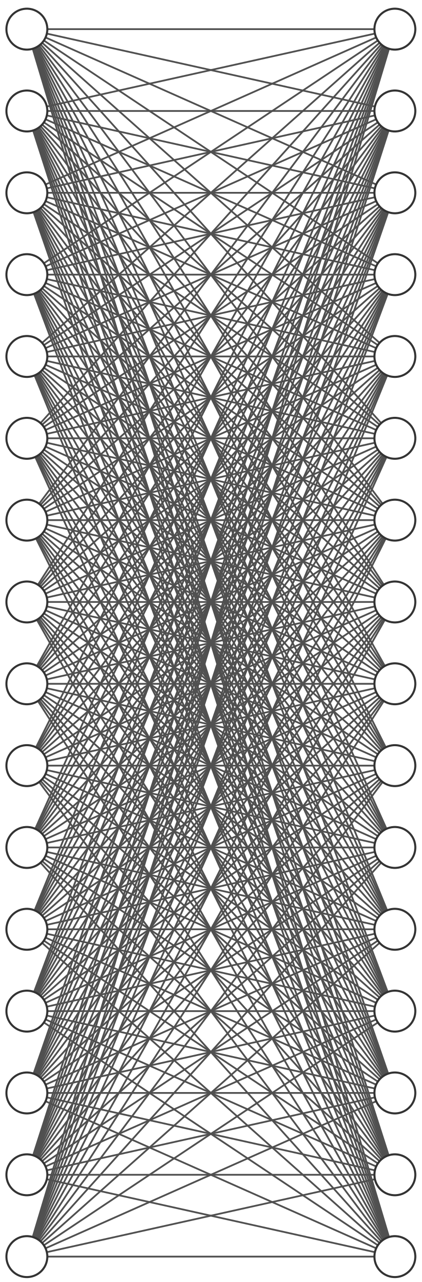
Freeze out

- Relaxation time can exceed annealing time [1912.02119]
  - State freeze
- Leads to an “effective temperature” that is configuration dependent
  - Factored to Ising parameters
- Iterate effective temperature until it converges with target distribution



Restricted Boltzmann Machines

- Neural network that can learn a probability distribution
- **Boolean** inputs forming a bipartite graph (restricted)
- $E(v, h) = -\sum_i \sum_j w_{ij} v_i h_j - \sum_i a_i v_i - \sum b_j h_j$
- $p(v, h) = \frac{1}{Z} e^{-E(v,h)}$
- $Z = \sum_{v,h} e^{-E(v,h)}$
- Discrete input variables → latent space not differentiable
- $\sigma(l)$ is monotic → $z = \mathcal{H}(l + \sigma^{-1}(\rho))$ [1805.07349]
- Still not differentiable! $z \to \zeta = \sigma\left(\frac{l + \sigma^{-1}(\rho)}{\tau}\right)$



Classifier

Testing accuracies **CaloQVAE**(CaloGAN [1])	e^+ vs π^+		e^+ vs λ	
	Test on	Test on	Test on	Test on
Train on	**99.83**	**99.85**	**66.78** (66.1)	**74.07** (70.6)
Train on	**94.26**	**100.0**	**53.73** (54.3)	**99.93**