Update on 1st paper

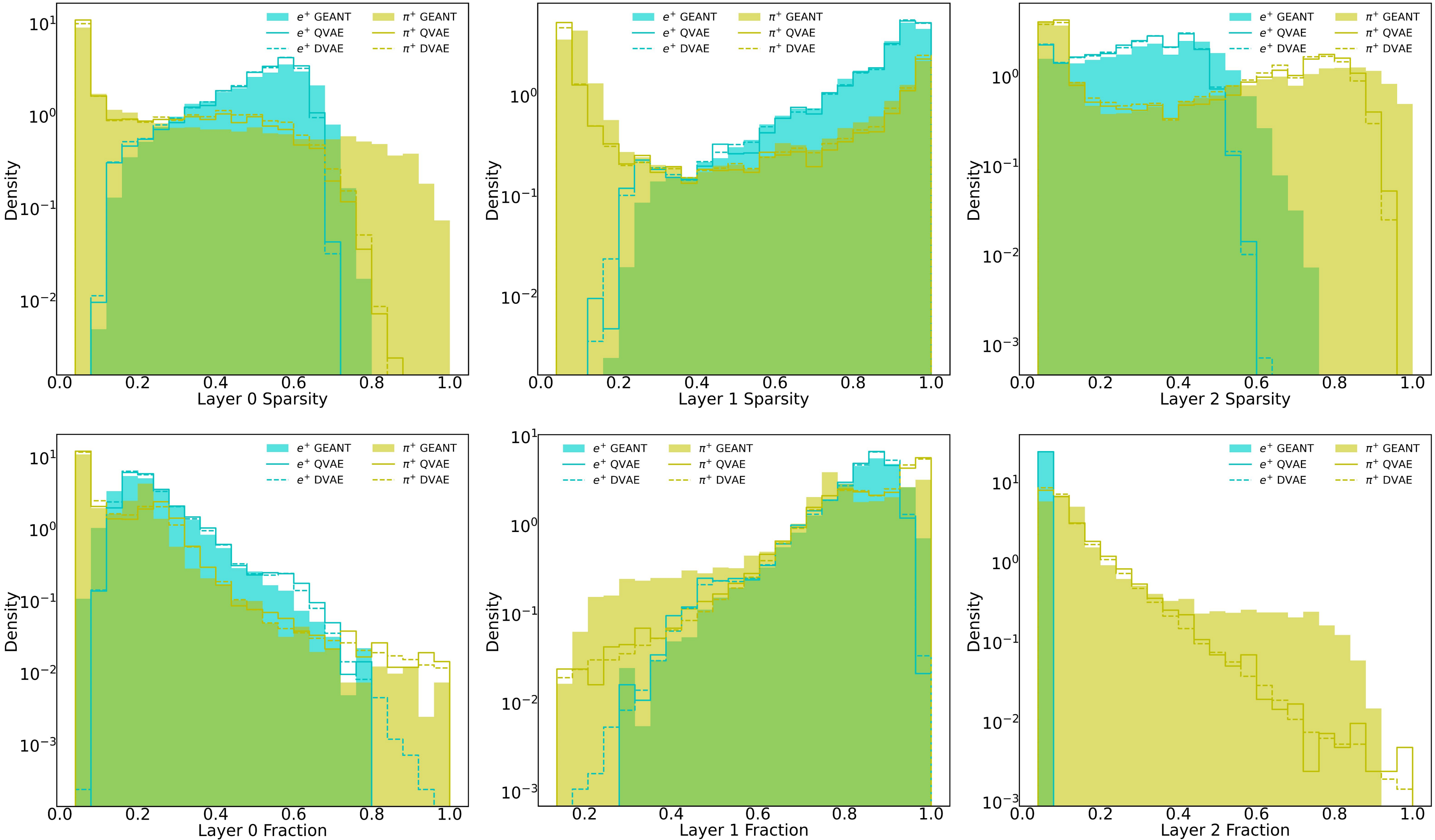
- Most typos/format changes have been done based on the comments from the PRL. reviewers.
  - Proper way to cite from D-WAVE?
  - Plots for single particle?
  - Rewrite the theory part with clearer setup and definitions.

evidence lower bound (ELBO) to the true log-likelihood:

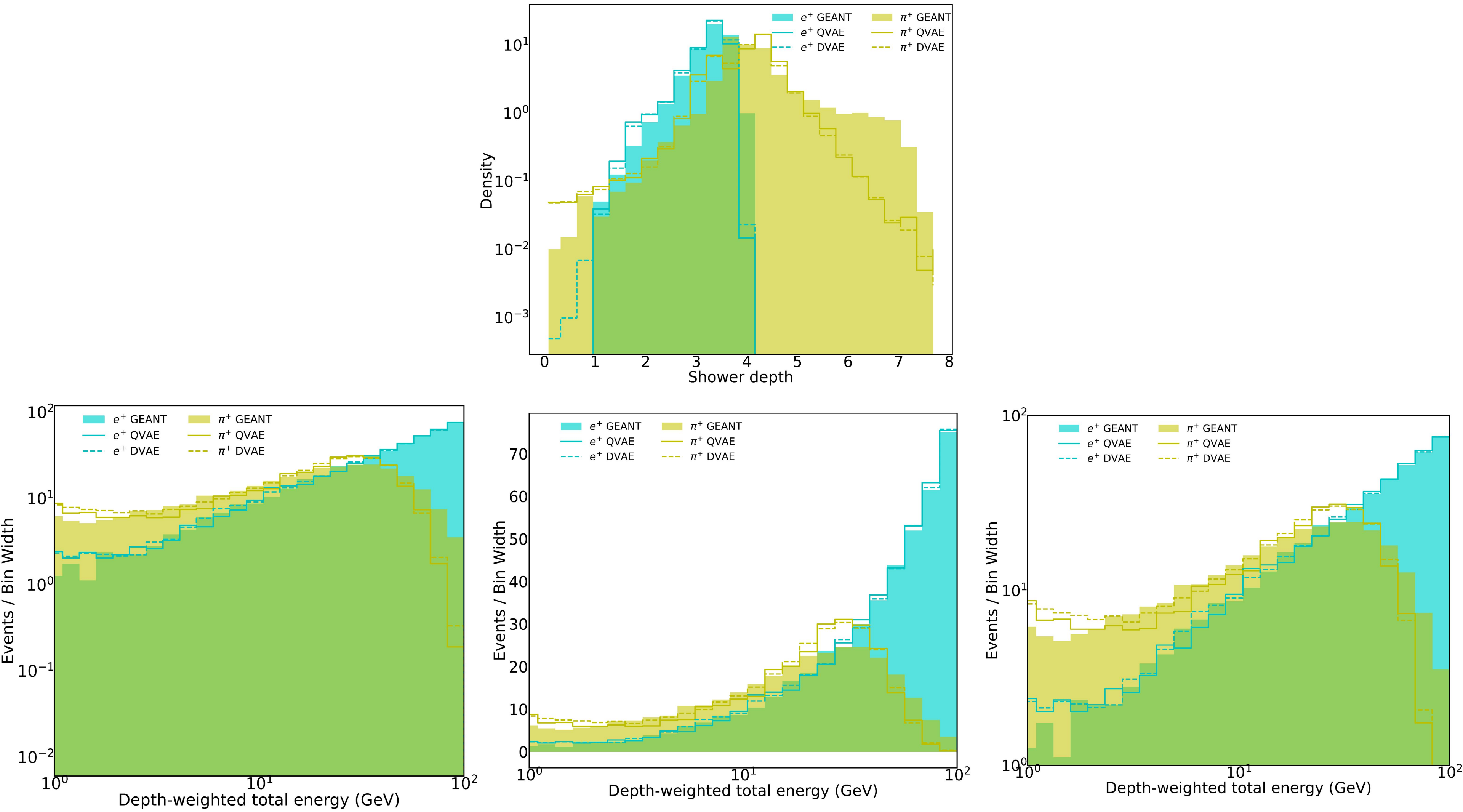
$$\log p_\theta(\mathbf{x}) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \mathrm{KL}[q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})] \tag{1}$$

$\mathrm{KL}[Q||P]$ is the Kullback-Liebler divergence between two probability distributions, Q and P while $\mathbb{E}_p$ denotes an expectation value over the distribution p. The variables $\mathbf{x}$ and $\mathbf{z}$ represent data (in this case a vector of calorimeter cell energies) and latent variable respectively. The approximating posterior, $q_\phi(\mathbf{z}|\mathbf{x})$ and the generative, $p_\theta(\mathbf{x}|\mathbf{z})$ distributions are often parameterized using neural networks. In the original VAE framework [6, 19], $q_\phi(\mathbf{z}|\mathbf{x})$ and $p_\theta(\mathbf{z})$ are assumed to be factorized Gaussian distributions, $\mathcal{N}(\boldsymbol{\mu}_\phi(\mathbf{x}), \boldsymbol{\Sigma}^2_\phi(\mathbf{x}))$ and $\mathcal{N}(\mathbf{0}, \mathbf{I})$ respectively.

CaloDVAE [7, 8] is a hierarchical discrete VAE which extends the VAE framework by introducing discrete latent variables $\mathbf{z}_i \in \{0, 1\}$ in the latent space. Both the encoder, $q_\phi(\mathbf{z}|\mathbf{x}, \mathbf{e})$, and decoder, $p_\theta(\mathbf{x}|\mathbf{z}, \mathbf{e})$, are conditioned on energy [18, 20] and are modelled by fully connected neural networks. The approximate posterior of the VAE has a hierarchical structure $q_\phi(\mathbf{z}|\mathbf{x}, \mathbf{e}) = \prod_i q_{\phi_i}(\mathbf{z}_i|\mathbf{z}_{j<i}, \mathbf{x}, \mathbf{e})$ with N latent groups and the latent distribution of the VAE is modelled by an RBM where $\mathbf{z} = [\mathbf{z}_1, ..., \mathbf{z}_N]$ are partitioned into 2 equal subsets to make the 2 sides of the RBM [18]. The RBM is modeled by probability distribution $p_{RBM} = e^{\mathbf{a}_l^T \mathbf{z}_l + \mathbf{a}_r^T \mathbf{z}_r + \mathbf{z}_l^T \mathbf{W} \mathbf{z}_r}/Z$ where Z is the partition function and $(\mathbf{a}_l, \mathbf{a}_r, \mathbf{W})$ are the RBM parameters which are trained along with the VAE parameters (θ, ϕ).



Density
Layer 0 Sparsity
Layer 1 Sparsity
Layer 2 Sparsity
Layer 0 Fraction
Layer 1 Fraction
Layer 2 Fraction
e^+ GEANT
e^+ QVAE
e^+ DVAE
π^+ GEANT
π^+ QVAE
π^+ DVAE



e^+ GEANT
e^+ QVAE
e^+ DVAE
π^+ GEANT
π^+ QVAE
π^+ DVAE
Density
Shower depth
Events / Bin Width
Depth-weighted total energy (GeV)