# Conditioned Quantum-Assisted Deep Generative Surrogate for ATLAS Particle-Calorimeter Interactions

D. Kraft[1,2], L. Zhu[2,3], J. Q. Toledo-Marin[2,4], I. Lu[5], H. Jia[2,3], C. Gay[3], R. G. Melko[4], E. Paquet[6], G. Fox[7], M.Swiatlowski[2], W. T. Fedorko[2]

1. University of Victoria 2. TRIUMF 3. UBC 4. Perimeter Institute for Theoretical Physics 5. University of Toronto 6. National Research Council Canada 7. University of Virginia

## ATLAS and the HL-LHC

- The ATLAS detector studies high-energy proton-proton collisions and the calorimeters measure energy.
- Simulating the calorimeters is CPU intensive, and the HL-LHC upgrade is estimated to require millions of CPU core-years annually.
- Simulations are needed to search for new physics and precision studies of the Higgs boson!
- We propose a deep generative model with a **variational autoencoder** and conditioned **restricted Boltzmann machine,** enabling sampling from D-Wave's Zephyr **quantum annealer**.

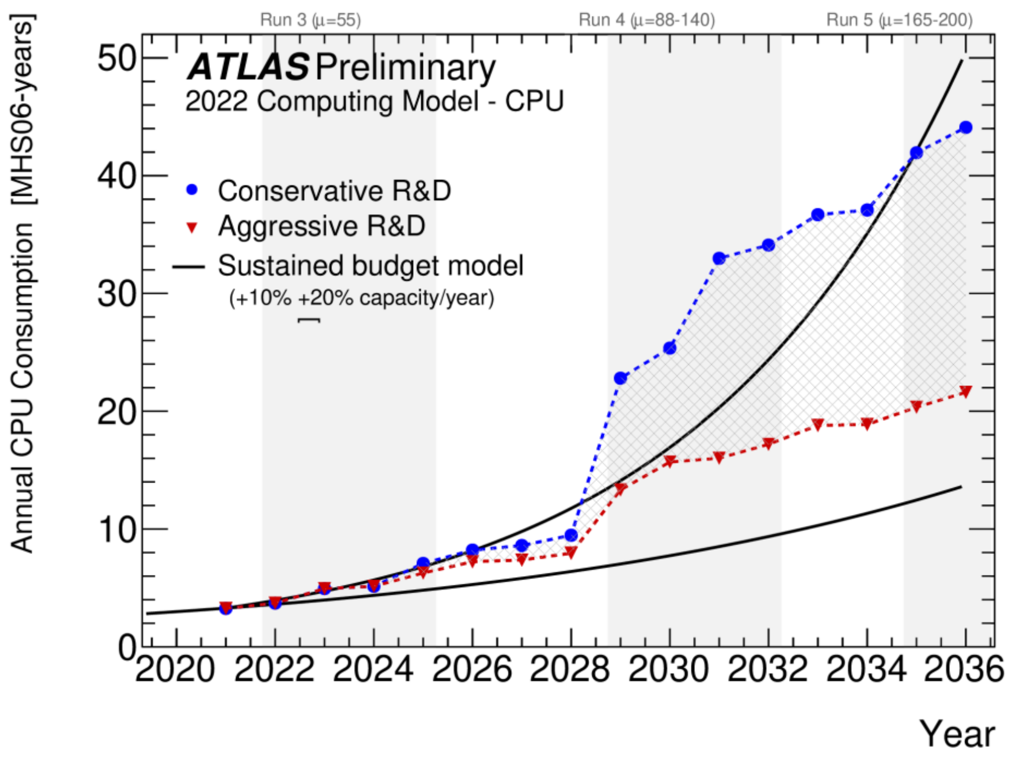



Figure 1. Projected CPU usage in ATLAS [1].

## Calorimeter Voxelization

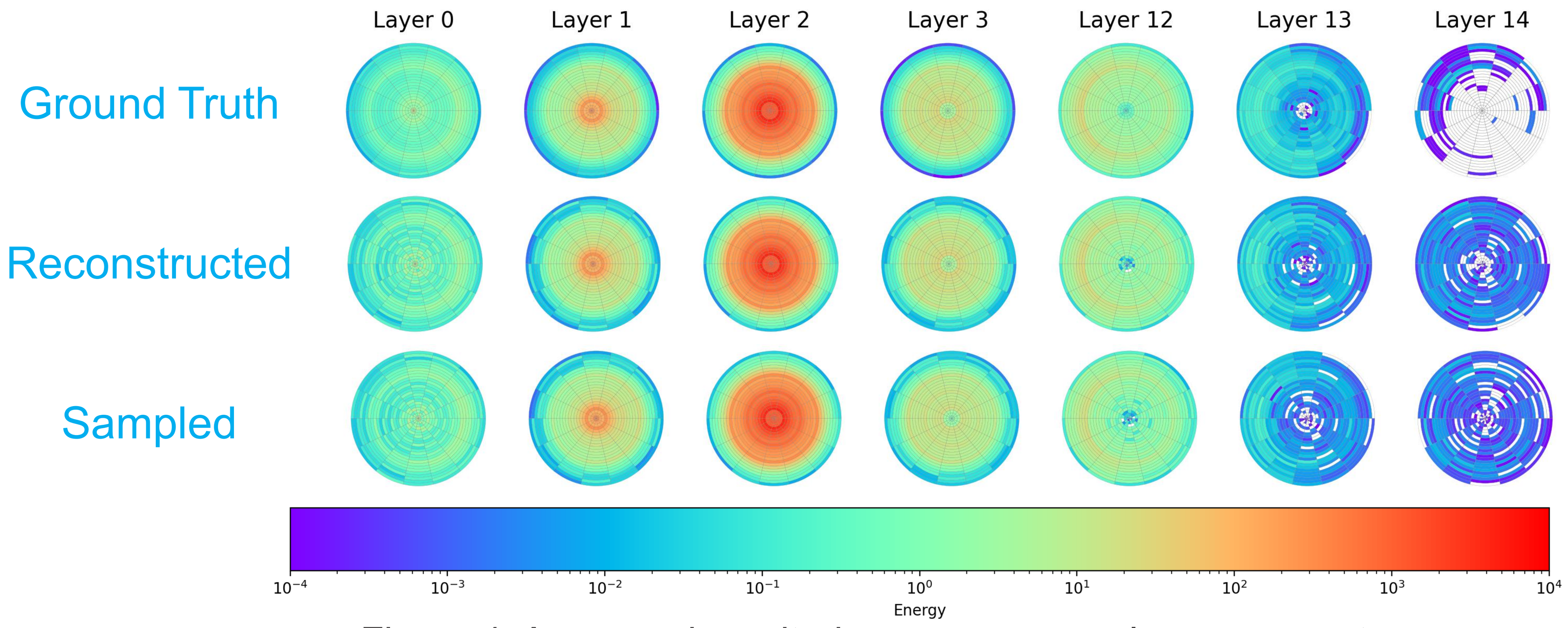



Figure 4. Average deposited energy per voxel across events.

## Energy Conditioning

- Run 3 of the LHC produces collisions at a center-of-mass energy of $\sqrt{s}$ = 13.6 TeV and the HL-LHC is expected to reach $\sqrt{s}$ = 14 TeV!
- Calorimeter showers have a wide range of energies.

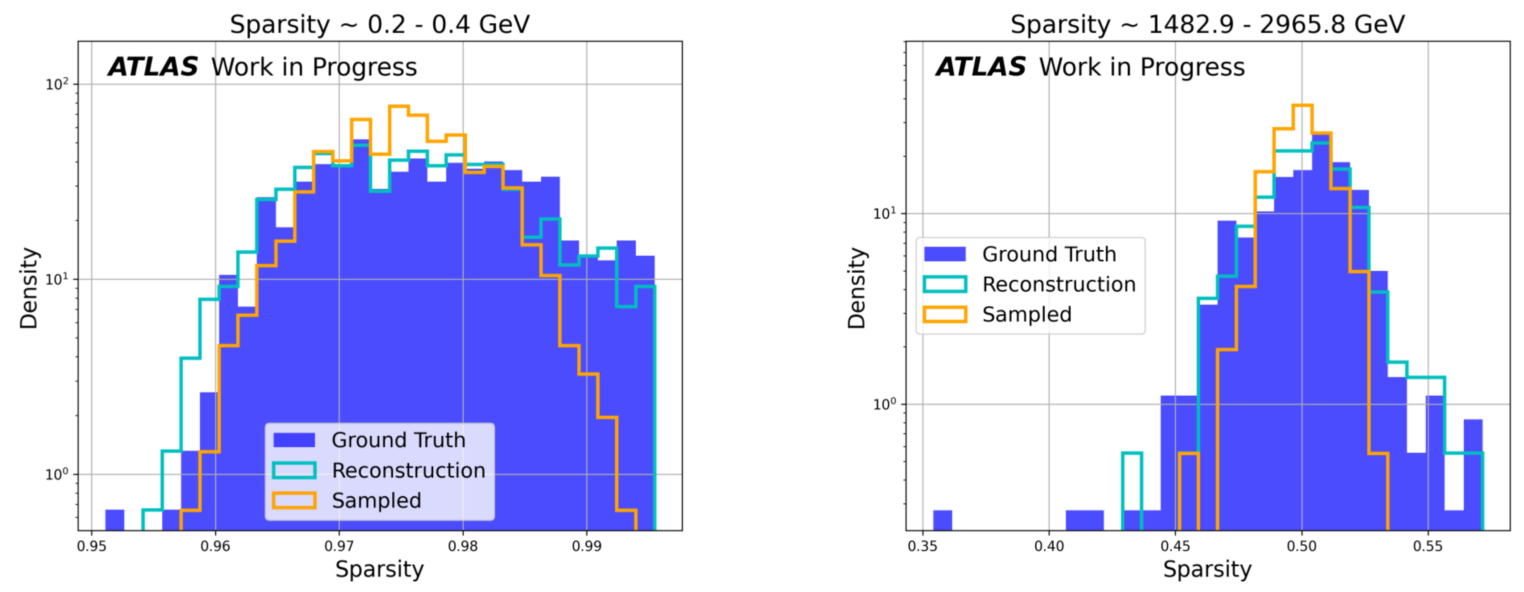



Figure 5. Energy conditioned sparsity.

## Detector Layer Evaluation

- The ATLAS calorimeters consist of many layers that each measure energy deposited by particles.

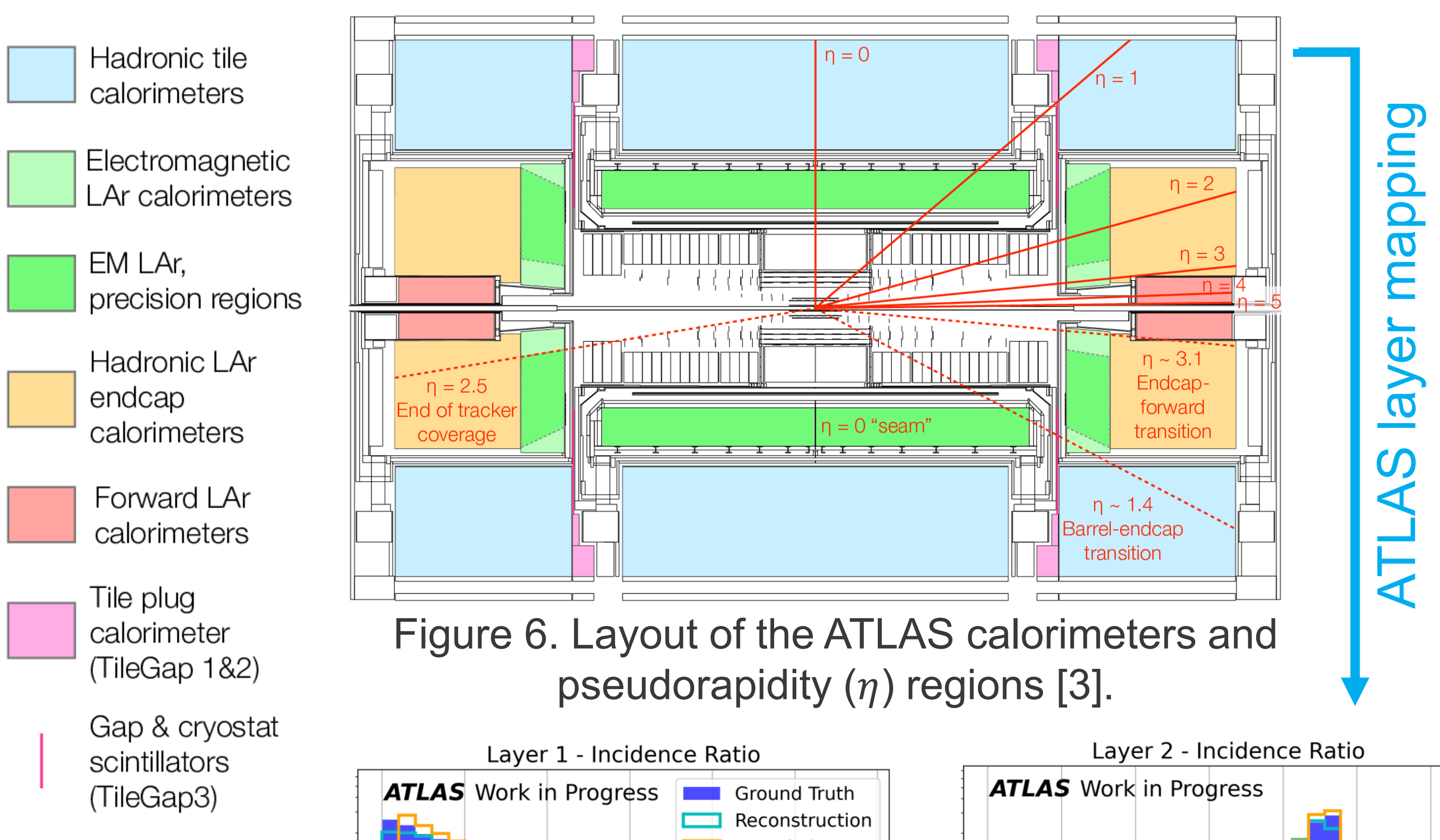



Figure 6. Layout of the ATLAS calorimeters and pseudorapidity ($\eta$) regions [3].

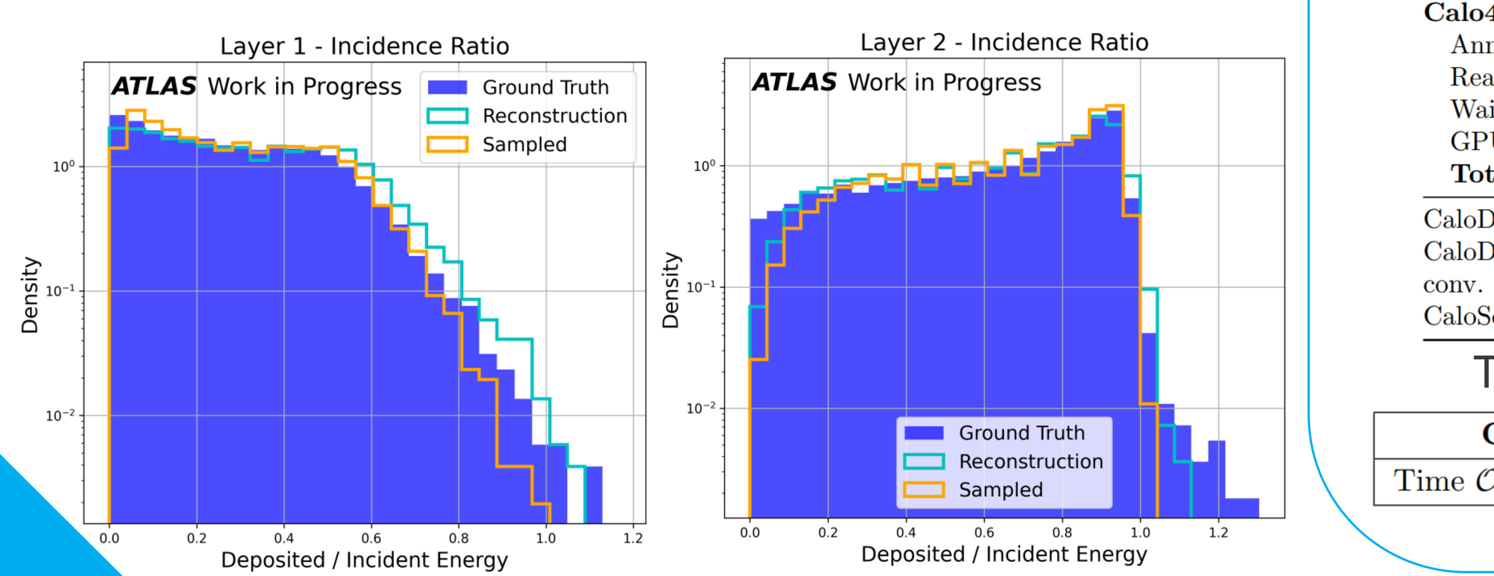



Figure 7. Ratio between deposited energy and incident energy.

## Model Architecture

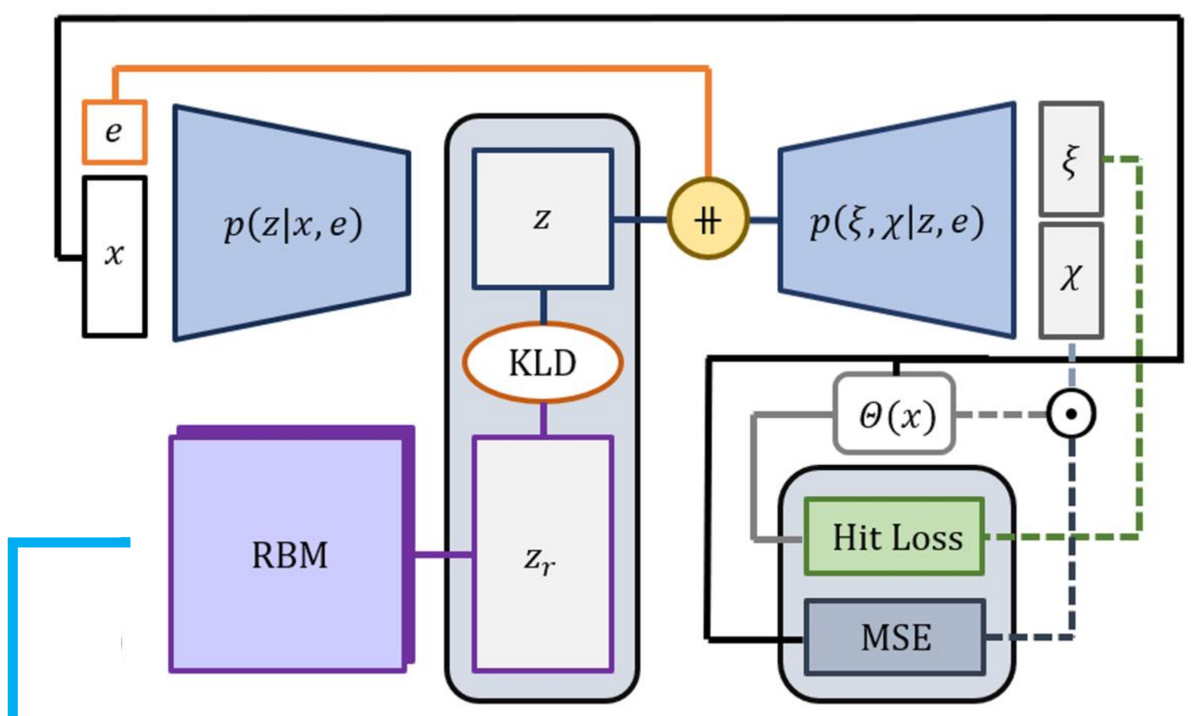



Figure 2. Training architecture flowchart [2].

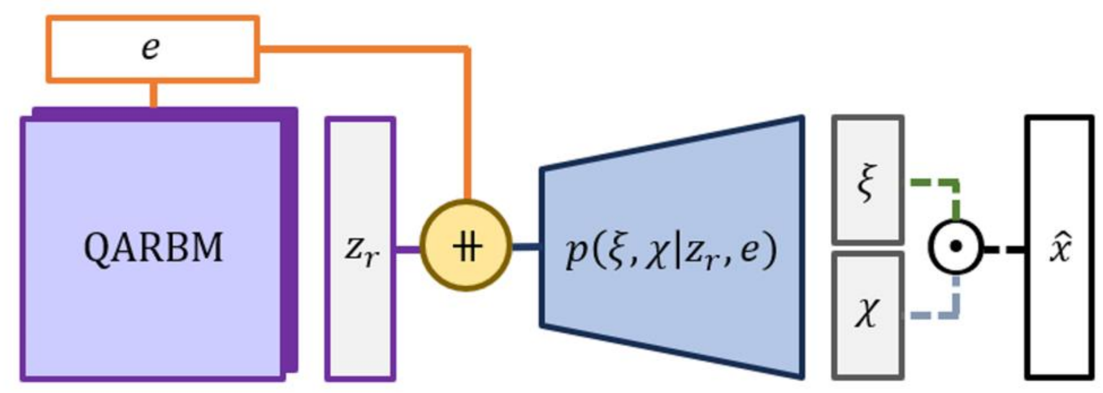



Figure 3. Sampling architecture flowchart [2].

## 4-Partite Restricted Boltzmann Machine

- Parameters updated to assign low energies to data.
- The RBM energy is given by:

$$E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{a},\boldsymbol{b},\boldsymbol{W}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i W_{ij} h_j$$

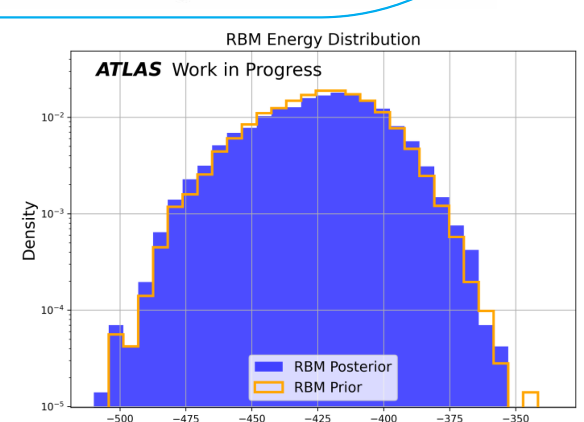
## Quantum Annealing

- Superconducting flux quantum bits with programmable spin-spin couplings.
- Aim to reach the ground state of the Hamiltonian $H_0$.

$$\mathcal{H}_{ising} = \underbrace{-\frac{A(s)}{2}\left(\sum_i \hat{\sigma}_x^{(i)}\right)}_{\substack{\text{Initial Hamiltonian} \\ H_1}} + \underbrace{\frac{B(s)}{2}\left(\sum_i h_i \hat{\sigma}_z^{(i)} + \sum_{i>j} J_{i,j}\hat{\sigma}_z^{(i)}\hat{\sigma}_z^{(j)}\right)}_{\substack{\text{Final Hamiltonian} \\ H_0}}$$

## Overall Metrics

- Model performance evaluated with distributions across layers and incident energies.



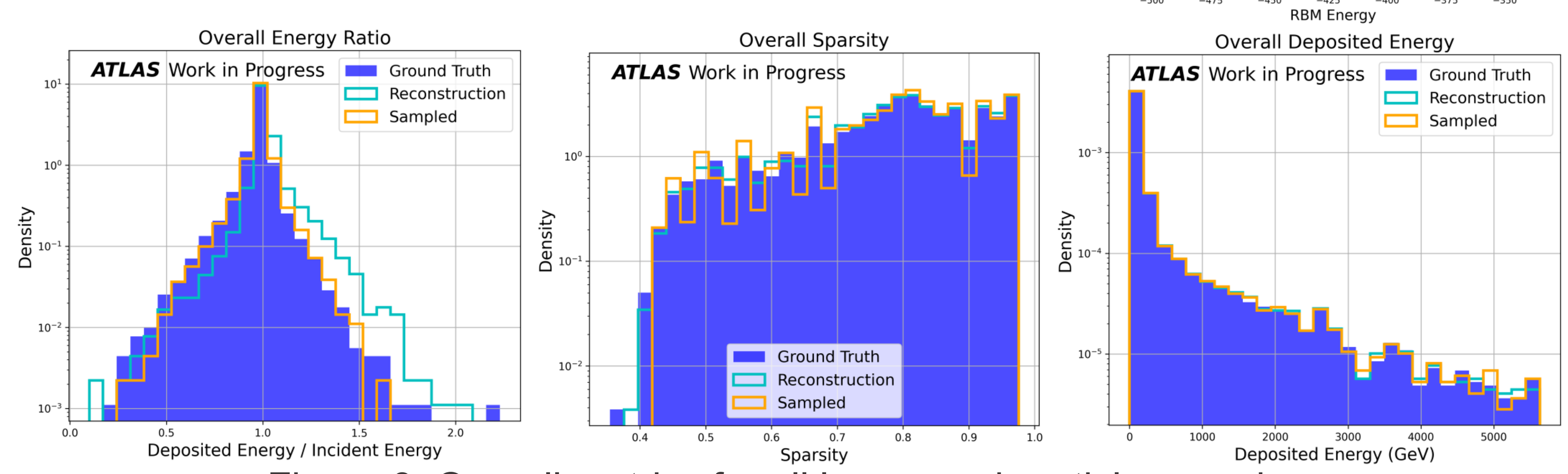




Figure 8. Overall metrics for all layers and particle energies.

## Model Comparison and Future Work

- **$10^3$-$10^6$ times faster** than traditional Geant4 simulations!
- Higher efficiency with comparable sample quality to SOTA deep generative models.
- Continue improving architecture to maximize performance and train with QPU.

| Model | Time / sample | Energy / sample [J] |
|---|---|---|
| **GEANT4** | 1 s | 8 |
| **Calo4pQVAE (QPU)** | | |
| Annealing | 20 µs | 0.3 |
| Readout | 87 µs | 2.2 |
| Wait | 20 µs | 0.3 |
| GPU postprocess | 54 µs | <0.1 |
| **Total** | **181 $\mu$s** | **2.0** |
| CaloDream | 74.3 ms | 30 |
| CaloDiffusion | 99.5 ms | 40 |
| conv. L2LFlows | 1.6 ms | 0.6 |
| CaloScore (single-shot) | 2.5 ms | 1 |

Table 1. CaloChallenge comparisons.

| GEANT4 | GPU (A100) | QPU | Anneal time |
|---|---|---|---|
| Time $\mathcal{O}(0.1) - \mathcal{O}(10^2)$ s | ~2 ms | ~0.2 ms | ~0.02 ms |

Table 2. Model timing.

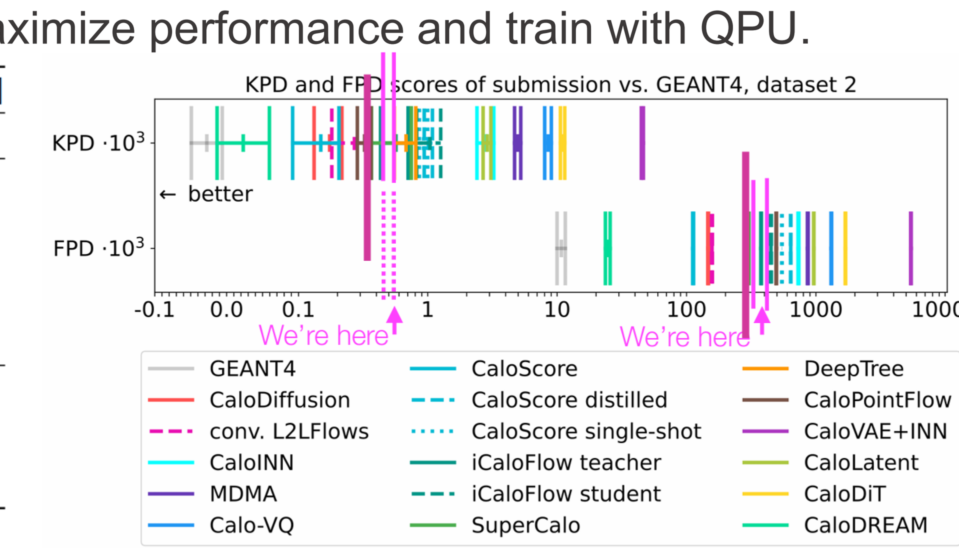



Figure 9. FPD and KPD metrics.

QaloSim/CaloQuVAE

[1]. ATL-SOFT-PROC-2024-009 [2]. Toledo-Marín, J.Q., Gonzalez, S., Jia, H. *et al.*, Conditioned quantum-assisted deep generative surrogate for particle-calorimeter interactions. *npj Quantum Inf* 11, 114 (2025). [3]. CERN-EP-2020-083