Application Quantum Optics 2026
22 - 28 Feb 2026, Obergurgl, Tyrol, Austria

Title:
Interpretable representation learning of quantum data

Abstract:
Interpretable machine-learning tools are playing an increasingly important role in quantum-science research. Our work presents a broadly applicable methodology that combines modern generative models with quantum many-body physics to learn compact and physically meaningful representations directly from raw, unlabelled snapshots of quantum experiments.

Among interpretable models, variational autoencoders (VAEs) have gained considerable attention due to their ability to autonomously extract minimal and meaningful representations from complex datasets. However, existing applications of VAEs to quantum data have largely overlooked its intrinsic stochasticity. This omission is critical: if a VAE cannot accurately capture the underlying probability distribution of the data, it cannot produce meaningful representations.

To address this, we propose two key adaptations that enable VAEs to properly account for the stochastic nature of quantum data: we reformulate VAE training as an explicit probability-distribution matching task and design an autoregressive decoder capable of faithfully representing quantum probability distributions.

Using benchmark quantum spin models, we identify regimes where standard methods fail while the representations learned by our approach remain meaningful and interpretable. Applied to experimental data from Rydberg atom arrays, the model autonomously uncovers the underlying phase structure without access to prior labels, Hamiltonian details, or knowledge of relevant order parameters, highlighting its potential as an unsupervised and interpretable tool for the study of quantum systems.