Learning Hard Distributions with Quantum-enhanced Variational Autoencoders*†

Anantha S Rao ‡
Department of Physics,
University of Maryland,
Maryland 20742, USA
`anantha@umd.edu`

Dhiraj Madan
IBM Research, India
`dmadan07@in.ibm.com`

Anupama Ray
IBM Research, India
`anupamar@in.ibm.com`

Dhinakaran Vinayagamurthy
IBM Research, India
`dvinaya1@in.ibm.com`

MS Santhanam
Department of Physics,
IISER Pune,
Maharastra, 411008, India
`santh@iiserpune.ac.in`

Abstract

Generative learning is an important task in classical machine learning with several models including generative adversarial networks (GANs) and variational autoencoders (VAEs) which are popular. In quantum machine learning an important task in this setting is that of modeling the distributions obtained by measuring quantum mechanical systems. Classical generative algorithms, including GANs and VAEs, can model the distributions of product states with high fidelity, but fail or require an exponential number of parameters to model entangled states. In this paper, we introduce a quantum-enhanced VAE (QeVAE), a generative quantum-classical hybrid model that uses quantum correlations to improve the fidelity over classical VAEs, while requiring only a linear number of parameters. We provide a closed form expression for the output distributions of the QeVAE. We also empirically show that the QeVAE outperforms classical models on several classes of quantum states, such as 4-qubit and 8-qubit quantum circuit states, haar random states, and quantum kicked rotor states, with a more than 2x increase in fidelity for some states. Finally, we find that the trained model outperforms the classical model when executed on the IBMq Manila quantum computer. As an application we show that our techniques can be also used for the task of circuit compilation. Our work paves the way for new applications of quantum generative learning algorithms and characterizing measurement distributions of high-dimensional quantum states.

1 Introduction

Research in quantum information science promises to enable the development of a fault-tolerant quantum computer that can perform certain tasks faster and more efficiently than any classical computer. To this end, several scientists have demonstrated that quantum algorithms can, in theory, outperform the best-known conventional algorithms when tackling specific problems and, in some

*A short video explaining the work can be seen at `https://www.youtube.com/playlist?list=PLMXUGj9FrociUhCTRh2cDqgcXZijDSQen`

†The source code can be found at `https://github.com/Anantha-Rao12/Quantum-enhanced-variational-autoencoder`

‡The author was a student at IISER Pune when the work was done

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

situations, deliver a 'quantum speedup' [1, 2]. For instance, certain quantum algorithms can take exponentially fewer resources for tasks such as factorization and eigenvalue decomposition, and quadratically fewer resources to search through unsorted databases [4,5,12]. This pursuit of 'quantum speedup' has motivated generations of physicists and engineers to discover novel algorithms that leverage the properties of superposition, entanglement and interference.

In this paper, we study the problem of modeling the measurement distributions obtained from unknown quantum states. This problem is fundamental in quantum information science, as it can reveal useful information about the properties and dynamics of quantum systems. Moreover, it can enable applications such as quantum state reconstruction, and entanglement quantification [3, 6, 7, 10]. These applications can help us understand and manipulate quantum systems in various fields, such as chemistry, materials science, and cybersecurity.

In this paper, we make four main contributions to the field of quantum generative learning.

1. We propose the Quantum-enhanced VAE (QeVAE), which can enhance the expressive power of classical VAEs and produce distributions that are classically intractable.
2. We provide a mathematical closed-form expression to theoretically analyze the class of models that QeVAE can model.
3. We demonstrate experimentally that our QeVAE outperforms classical VAEs on modeling measurement samples of different classes of quantum states, such as haar random states, quantum circuit states, and quantum kicked rotor states, and report an increase in fidelity by more than 2x for an 8-qubit quantum circuit state.
4. We show how to use our algorithm for the task of circuit compilation as well.

Our work is organized as follows: In Section 2, we discuss the background work on generative learning, variational autoencoders, and the problem of learning measurement distributions of quantum states. In Section 3, we propose the QeVAE and mathematically characterize the output distribution of the model. We describe our experiments and results in section 4. Finally, in Section 5, we discuss our conclusions and future outlook in the greater context of developing quantum algorithms for generative learning.

2 Background and Related work

2.1 Generative learning and Variational Auto Encoders

The task of generative modeling involves modeling a parameterized distribution $p_\theta(\mathbf{x})$, given independent and identically distributed (iid) samples from the distribution $\{\mathbf{x_i}\}_{i=1}^m$ where each $\mathbf{x_i} \sim p(\mathbf{x})$. Here the goal is to maximize the log likelihood of the data given by $L(\theta) = \sum_i \log p_\theta(\mathbf{x_i})$.

Variational learning involves modeling the distribution through a latent vector $\mathbf{z}$. One can define a prior on the latent variables $\mathbf{z} \sim p(\mathbf{z})$ and a parameterized likelihood function $p_\theta(\mathbf{x}|\mathbf{z})$. The joint distribution over the variables $\mathbf{x}$ and $\mathbf{z}$ can then be written as $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{z})p_\theta(\mathbf{x}|\mathbf{z})$.

Variational Autoencoders [9] seek to maximize the ELBO defined by ELBO $= \mathbb{E}_{q(\mathbf{z}|\mathbf{x})} (\log p(\mathbf{x}|\mathbf{z})) - KL(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})))$. The algorithm involves using the posterior network (encoder) to create the distribution $q_\phi(\mathbf{z}|\mathbf{x})$ and provides samples $\mathbf{z}$, which are then passed through the likelihood network (decoder) to produce a distribution over $\mathbf{x}$ conditioned on sampled vector $\mathbf{z}$. One then seeks to maximize the difference of log likelihood of generated distribution $p(\mathbf{x}|\mathbf{z})$ and the KL divergence between posterior ($q(\mathbf{z}|\mathbf{x})$) and prior distributions $p(\mathbf{z})$. The first term (reconstruction term) tries to maximize the likelihood of recovering back $\mathbf{x}$ from the latent variable. The second term is a regularization term that tries to ensure that posterior distribution is close to prior.

2.2 Quantum computation and quantum machine learning

Quantum computing is a model of computation with a potential to provide a speedup over its classical counterpart. The fundamental units in quantum computing framework consist of quantum bits or **qubits**. An n qubit state lies in a Hilbert space spanned by basis states corresponding to 2^n classical bit strings, $|0...0\rangle$ through $|1...1\rangle$. An arbitrary state can be a unit vector spanned by the 2^n basis strings as $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$, where $\sum_x |\alpha_x|^2 = 1$. The basic operations on qubits include quantum gates, which are unitary operators acting on one or more qubits. Common examples of single qubits gates include the Hadamard, Pauli Gates(X,Y,Z), and S gate. The exponentiated Pauli Gates also provide us a set of parameterized gates, $R_x(\theta) = \exp(-i\frac{\theta}{2}X)$, $R_y(\theta) = \exp(-i\frac{\theta}{2}Y)$ and $R_z(\theta) = \exp(-i\frac{\theta}{2}Z)$. A common 2 qubit gate is $CNOT$ which acts as $CNOT\,|x, y\rangle = |x, x \oplus y\rangle$ on basis states. A quantum circuit consists of a sequence of single qubit and two-qubit gates acting on an initial state (typically $|0...0\rangle$) followed by measurement of the final state. A measurement of a state $|\psi\rangle = \sum_x \alpha_x |x\rangle$ yields one of the bit strings x with probabilities $|\alpha_x|^2$.

A parameterized quantum circuit (PQC) is a quantum circuit that has some gates that depend on parameters θ. The unitary operator of the PQC is denoted by $\hat{U}(\theta)$, which can produce a state $|\psi(\theta)\rangle = \hat{U}(\theta)\,|0\rangle^{\otimes n}$. The PQC can also be conditioned on an input $\mathbf{x}$ as $\hat{U}(\mathbf{x}, \theta)$. This can be decomposed into a feature map $\hat{U}_\phi(\mathbf{x})$ and a trainable ansatz $\hat{V}(\theta)$. The feature map is a data encoding circuit that transforms input data $\bar{x} \in \mathbb{R}^n$ into a quantum state using single qubit and two qubit gates, potentially parameterized by the input variables. The ansatz is a parameterized circuit that consists of alternating rotation layers and entanglement layers. The rotation layers are single-qubit gates applied on all qubits. The entanglement layer uses two-qubit gates to entangle the qubits according to a predefined scheme. The parameters of the PQC can be optimized to achieve a certain goal, such as minimizing the energy of a quantum system or maximizing the accuracy of a machine learning task. The parameters can be updated iteratively using classical optimization methods such as gradient-based (ADAM), [8], SPSA [13]) or gradient-free (COBYLA) algorithms. The measurement distribution of $|\psi(\mathbf{x}, \theta)\rangle = \hat{U}(\mathbf{x}, \theta)\,|0\rangle$ in Z-basis defines the conditional distribution $p(\mathbf{y}|\mathbf{x}) = |\langle y|\psi(\mathbf{x}, \theta)\rangle|^2$ $\forall y \in \{0, 1\}^n$.

3 Quantum-enhanced Variational Autoencoders

In this section, we define a quantum-enhanced variational autoencoder (QeVAE) to model the measurement distribution of an unknown n-qubit quantum state. As a baseline, classical VAEs have already been used to model such a distribution [11]. The hybrid model that we propose consists of a feedforward classical encoder, a continuous latent space, and a parametrized quantum circuit as a decoder. We model the approximate posterior (encoder network) $Q_\phi(\mathbf{z}|\mathbf{x})$ through a classical feedforward neural network and the latent variable as $\mathbf{z} \sim \mathcal{N}(0, I)$. The likelihood (generator) distribution $p_\theta(\mathbf{x}|\mathbf{z})$ is defined via a quantum circuit i.e. $p_\theta(\mathbf{x}|\mathbf{z}) = |\langle x|\hat{U}(\theta, \mathbf{z})|0\rangle^{\otimes n}\rangle|^2$. Here, the input z is encoded using Z or ZZ feature map. This is followed by 2- Local ansatz. The Evidence Lower bound loss (ELBO) is optimized through a classical optimizer such as ADAM. The model is trained to mimic the given measurement distribution of states. During training, the parameters of the encoder and the rotation gates in the decoder (with a pre-selected entanglement type) are iteratively varied and learned. It will enable scientists to generate certain quantum states in different physical quantum computers just by knowing the set of rotation and entangling gates to perform. We now present a theorem that allows us to mathematically characterize the class of distributions that can be obtained via the above model(See Appendix A for proof.).

Theorem 1. *Consider a latent variable model with $\mathbf{z} \sim p(\mathbf{z})$ and $p(\mathbf{x}|\mathbf{z}) = |\langle \mathbf{x}|\, V_\theta U_\phi(\mathbf{z})\,|0^n\rangle|^2$, where $U_\phi(\mathbf{z})$ is a feature map and V_θ is a parameterized ansatz. Then*

1. *$\exists$ a density matrix ρ, such that $p(x) = \langle x|\, V_\theta \rho V_\theta^\dagger\,|x\rangle$. In other words the distribution of $\mathbf{x}$ can be obtained by evolving a density matrix ρ under unitary ansatz V_θ followed by measurement in standard basis.*

2. *Conversely, for each density matrix ρ, there exists a prior $p(\mathbf{z})$ and a feature map $U_\phi(\mathbf{z})$, such that $p(x) = \langle x|\, V_\theta \rho V_\theta^\dagger\,|x\rangle$.*

4 Results

We train our algorithm on multiple datasets of bits strings obtained by measuring random quantum circuit state, haar random state and kicked rotor states (obtained by Hamiltonian evolution under kicked rotor Hamiltonian). We compare the final fidelity between the target distribution and that produced by a random uniform guess, a classical variational autoencoder (CVAE), and a Quantum-enhanced variational autoencoder (QeVAE). For each type of state, we consider five different random seeds. QeVAE results include the best fidelity observed across different hyper-parameters.

Table 1: Fidelity for Quantum circuit states

No qubits	4						8					
Seed	12	16	27	44	102	**Mean** ↓	12	16	27	44	102	**Mean** ↓
Uniform	.477	.366	.315	.154	.431	.348	.158	.279	.080	.309	.107	.187
CVAE	.501	.667	.597	**.925**	.758	.690	.229	.423	.079	.304	.181	.243
QeVAE	**.981**	**.976**	**.950**	.873	**.912**	**.938**	**.665**	**.654**	**.388**	**.548**	**.591**	**.569**

Table 2: Fidelity for Haar random states

No qubits	4						8					
Seed	42	96	27	101	102	**Mean** ↓	12	43	16	27	2	**Mean** ↓
Uniform	.772	.776	.777	.771	.768	.773	.766	.770	.773	.781	.772	.772
CVAE	.795	.798	.800	.788	.788	.794	.754	.755	.757	.766	.763	.759
QeVAE	**.839**	**.983**	**.913**	**.988**	**.932**	**.931**	**.876**	**.878**	**.887**	**.887**	**.887**	**.883**

Table 3: Fidelity of Quantum-kicked rotor states

No qubits	4			8		
Type	Localized (k=6)	Diffusive (k=0.5)	**Mean** ↓	Localized (k=6)	Diffusive (k=0.5)	**Mean** ↓
Uniform	.175	.838	.506	.053	.418	.236
CVAE	.723	.908	.815	.061	.406	.233
QeVAE	**.991**	**.992**	**.991**	**.912**	**.616**	**.764**

From tables(1-3), we observe that across all quantum states we considered the final fidelity obtained from a QeVAE outperforms the classical VAE and a random guess. In addition, the number of learnable parameters in the classical VAE is typically of the order $2^{(n)}$ while those in a QeVAE is $4n + \epsilon$ where n is the number of qubits and ϵ is a constant ($\epsilon < 4$). To further validate our findings, we run the best QeVAE models on real quantum devices and see that the obtained fidelity is higher than those achieved by classical methods (Table 1 in Appendix C). Our proposed algorithm achieves the highest fidelity across all types of datasets.The inherent ability of our model to learn quantum correlations i.e we are able to produce entangled multi-qubit states through variational quantum circuits and tailor the rotation gates to reproduce a desired distribution allows them to outperform the classical model for the quantum circuit, haar random and kicked rotor states.

Hardware run: We execute the best model trained on the simulator on IBM Hardware by transpiling the decoder circuit. To generate the output distribution, random samples from $N(0, 1)$ are propagated through the preprocessor linear layer and then through the circuit (executed on hardware). An average over all initial random points yields the desired output distribution. Our results are depicted in Figure 2 in Appendix B and in Table 1 in Appendix C. We find that final fidelity is lesser than that on a simulator. Using error mitigation and suppression techniques, we are able to perform better than the classical VAE.

Circuit compilation: We now present a practical application of QeVAEs on the task of circuit compilation. Circuit compilation or circuit compression is an important area of focus in the NISQ era. Since the depth of circuits executable on hardware is limited, there is a need to transform deep circuits into shallow ones by altering the sequence of gates and reducing the overall size and complexity.

We simulate an unknown quantum state by considering a deep quantum circuit with twenty layers of rotation and entangling gates. In reality, the form of the circuit is unknown and one only has access to the measurement data. A projective measurement on such a state produces a measurement distribution as shown in Figure 1(e) in Appendix B. Note that our goal here is to reproduce the measurement distribution and not to learn the original state itself. Through the QeVAE learning approach, we can learn the measurement distribution with high fidelity. After training, we can discard the encoder part of the circuit, and the decoder provides a sequence of gates that can be implemented on hardware to generate the same distribution. We achieve a final fidelity of 0.956 (Figure 1(d) in Appendix B) and a multi-fold reduction in the number of gates as seen in Figure 1(b) in Appendix B.

The initial circuit contains: 80 Rx gates, 80 Ry gates, 63 CNOT gates which are reduced to 28, 20, and 6 respectively, yielding an overall 4x reduction in total number of gates and 3x reduction in the number of parameters.

5 Conclusion and Outlook

The major outcome of our work is a working hybrid quantum-classical machine learning model for generative learning .In addition, our models are particularly suitable for quantum devices with small number of noisy qubits with limited connectivity. With this algorithm, we have investigated the ability of quantum generative models to learn the distributions obtained from the measurement of quantum many-body systems. Such distributions are known to be demanding for classical generative models and we verify the same in our experiments. We further go on to show that our proposed hybrid-model can learn these distributions with a much higher final fidelity. In addition, we have shown that QeVAEs can be useful for the practical task of circuit compilation.

References

[1] Scott Aaronson. Read the fine print. *Nature Physics*, 11(4):291–293, April 2015.

[2] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, September 2017.

[3] Guillermo García-Pérez, Matteo A.C. Rossi, Boris Sokolov, Francesco Tacchino, Panagiotis Kl. Barkoutsos, Guglielmo Mazzola, Ivano Tavernelli, and Sabrina Maniscalco. Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms. *PRX Quantum*, 2:040342, Nov 2021.

[4] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.

[5] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical review letters*, 103(15):150502, 2009.

[6] Hsin-Yuan Huang. Learning quantum states from their classical shadows. *Nature Reviews Physics*, 4(2):81–81, January 2022.

[7] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, June 2020.

[8] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *stat*, 1050:1, 2014.

[10] Andrea Rocchetto, Scott Aaronson, Simone Severini, Gonzalo Carvacho, Davide Poderini, Iris Agresti, Marco Bentivegna, and Fabio Sciarrino. Experimental learning of quantum states. *Science Advances*, 5(3), March 2019.

[11] Andrea Rocchetto, Edward Grant, Sergii Strelchuk, Giuseppe Carleo, and Simone Severini. Learning hard quantum distributions with variational autoencoders. *npj Quantum Information*, 4(1):1–7, 2018.

[12] Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.

[13] James C Spall. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins apl technical digest*, 19(4):482–492, 1998.

A Proofs

Proof of Theorem 1.

Proof. 1.

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z}) d\mathbf{z}$$
$$= \int p(\mathbf{z}) | \langle \mathbf{x} | V_\theta U_\phi(\mathbf{z}) | 0^n \rangle |^2 d\mathbf{z}$$
$$= \int p(\mathbf{z}) \langle \mathbf{x} | V_\theta U_\phi(\mathbf{z}) | 0^n \rangle \langle 0^n | U_\phi(\mathbf{z})^\dagger V_\theta^\dagger | \mathbf{x} \rangle d\mathbf{z}$$
$$= \langle \mathbf{x} | V_\theta (\int p(\mathbf{z}) U_\phi(\mathbf{z}) | 0^n \rangle \langle 0^n | U_\phi(\mathbf{z})^\dagger d\mathbf{z}) V_\theta^\dagger | \mathbf{x} \rangle$$

(By linearity of inner product)

We define $\rho := \int p(\mathbf{z}) U_\phi(\mathbf{z}) |0^n\rangle \langle 0^n| U_\phi(\mathbf{z})^\dagger d\mathbf{z}$
Note

(a) ρ is a valid density matrix i.e. $\rho^\dagger = \rho$, $\rho \geq 0$ and $Tr(\rho) = I$
(b) ρ is independent of both $\mathbf{x}$ and $\mathbf{z}$.

Then,

$$p(\mathbf{x}) = \langle \mathbf{x} | V_\theta \rho V_\theta^\dagger | \mathbf{x} \rangle$$

Thus, such the $p(x)$ is equivalent to preparing a density state ρ, evolving under unitary V_θ, and performing a measurement in standard basis.

2. For the second part, given a density matrix ρ, consider the sepctral decomposition of ρ as $\rho = \sum_z \lambda_z |\psi_z\rangle \langle \psi_z|$. Since ρ is a valid density matrix, λ_z's define a distribution $p(\mathbf{z}) = \lambda_z$. Now, for each $\mathbf{z}$, choose a unitary such that $U_{\phi(\mathbf{z})} |0\rangle = |\psi_z\rangle$. Now, one can verify that such a feature map satisfies the required equation.

□

B Figures

C Tables

Table 4: Hardware results for a 4 qubit quantum circuit state

State	Fidelity	Simulator	Hardware	Suppression	Mitigation
Uniform	0.477	✓			
CVAE	0.501	✓			
QeVAE	0.981	✓			
QeVAE	0.658		✓		
QeVAE	0.642		✓	✓	✓

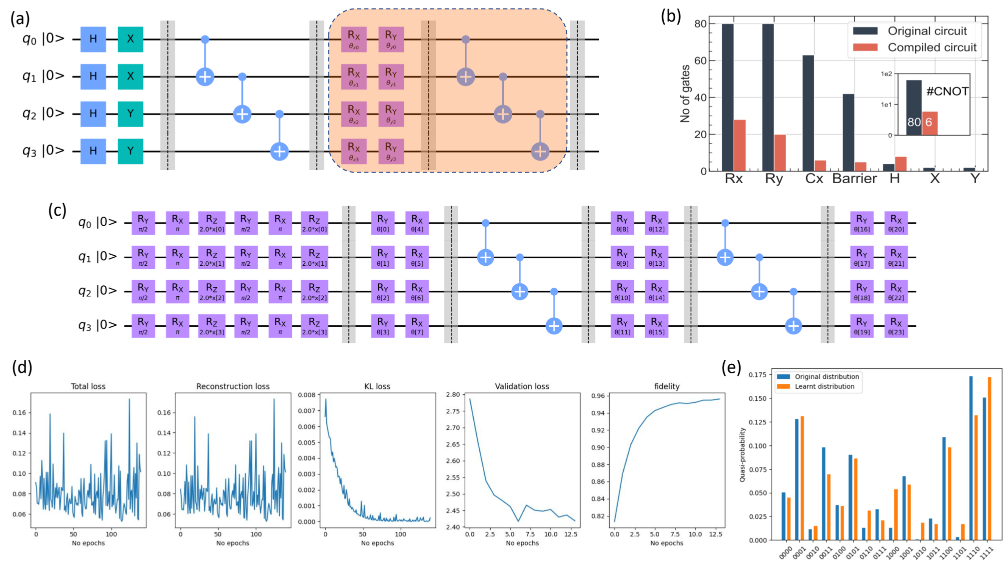

Figure 1: **Circuit compilation with QeVAEs**: (a) Original circuit structure that produces a measurement distribution. The orange box represents a single layer of rotation and entangling layer that is repeated 20 times. (b) The compiled circuits requires very few gates when compared to the original circuit that produces the state. Particularly, the number of CNOT gates is reduced by ∼14X (c) The variational quantum circuit containing the Pauli-Z feature Map that embeds data x from the latent space and the ansatz with only 16 parameters. (d) Results of training the ansatz to produce the measurement distribution. We achieve a final fidelity of 0.956 (e) Original and the distribution produced by the QeVAE after training.

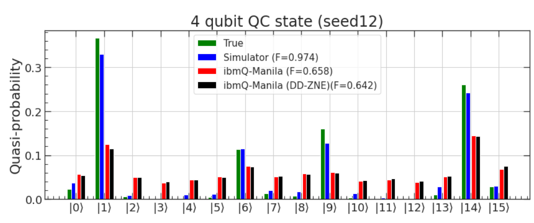

Figure 2: **Inference on IBM hardware**: The true distribution produced by an unknown 4-qubit system (green) is used to train a QeVAE on IBM's qasm-simulator (blue) that results in a final of 0.97. After training, the decoder is executed on the IBM-Manila. The measurement distribution produced from hardware has a total fidelity of 0.658 which changes to 0.642 with error-mitigation (Zero noise extrapolation) and error-suppression (Dynamic decoupling)