



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Entangling Quantum Generative Adversarial Networks

Murphy Yuezhen Niu, Alexander Zlokapa, Michael Broughton, Sergio Boixo, Masoud Mohseni, Vadim Smelyanskyi, and Hartmut Neven

Phys. Rev. Lett. **128**, 220505 — Published 3 June 2022

DOI: [10.1103/PhysRevLett.128.220505](https://doi.org/10.1103/PhysRevLett.128.220505)

Entangling Quantum Generative Adversarial Networks

Murphy Yuezhen Niu,^{1,*} Alexander Zlokapa,^{1,2,3,*} Michael Broughton,¹
Sergio Boixo,¹ Masoud Mohseni,¹ Vadim Smelyanskiy,¹ and Hartmut Neven¹

¹Google AI Quantum, Venice, CA 90291, USA

²Division of Physics, Mathematics and Astronomy, Caltech, Pasadena, CA 91125, USA

³Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Generative adversarial networks (GANs) are one of the most widely adopted machine learning methods for data generation. In this work, we propose a new type of architecture for quantum generative adversarial networks (an *entangling* quantum GAN, EQ-GAN) that overcomes limitations of previously proposed quantum GANs. Leveraging the entangling power of quantum circuits, the EQ-GAN converges to the Nash equilibrium by performing entangling operations between *both* the generator output *and* true quantum data. In the first multi-qubit experimental demonstration of a fully quantum GAN with a provably optimal Nash equilibrium, we use of the EQ-GAN on a Google Sycamore superconducting quantum processor to mitigate uncharacterized errors, and we numerically confirm successful error mitigation with simulations up to 18 qubits. Finally, we present an application of the EQ-GAN to prepare an approximate quantum random access memory and for the training of quantum neural networks via variational datasets.

INTRODUCTION

Generative adversarial networks (GANs) [1] are one of the most widely adopted *generative machine learning* methods, achieving state-of-the-art performance in a variety of high-dimensional and complex tasks including photorealistic image generation [2], super-resolution [3], and molecular synthesis [4]. Given access only to a training dataset $S = \{x_i\}$ sampled from an underlying data distribution $p_{\text{data}}(x)$, a GAN can generate realistic examples outside S . Certain probability distributions generated by quantum computers are thought to be classically hard to sample from under plausible conjectures [5–7], and learning to generate these samples using a classical GAN can also be formidably hard [8]. In this work, we focus on developing a fully quantum mechanical GAN, where the true data is given by a quantum state; the task is then to learn a *generator* circuit that can reproduce the same quantum state. Following the framework of a GAN, a *discriminator* circuit is presented either with the true data or with fake data from the generator. The generator and discriminator are then trained adversarially [9]: the generator attempts to fool the discriminator, while the discriminator attempts to correctly distinguish true and fake data. Unlike hybrid quantum-classical GANs that are intrinsically restricted to learning classical distributions [10–13], we provide viable machine learning applications of the proposed quantum GAN for both quantum data and classical data.

Current quantum hardware suffers from both incoherent and coherent errors, which are often time-dependent and thus difficult to consistently characterize [14, 15]. Supervised quantum machine learning methods requiring exact implementation of certain procedures — e.g., a swap test between states — may consequently fail to converge to the correct optimum due to miscalibrated gate parameters. To maximize state overlap in the presence of

uncharacterized errors, we introduce an *entangling* quantum GAN (EQ-GAN) that takes a uniquely quantum approach compared to prior art: rather than providing the discriminator with *either* true *or* fake data, we allow the discriminator to entangle *both* true and fake data. Comparing to existing work, we show that a quantum GAN constructed as a direct analogy of the classical GAN architecture (QuGAN [16, 17]) may oscillate between a finite set of states due to mode collapse; the EQ-GAN is confirmed to properly converge on these problem instances.

We prove that the EQ-GAN architecture always contains a Nash equilibrium at the optimal point of accurately generating the true data. When initialized close to this equilibrium point, such as in the setting of suppressing small unknown errors, we find that the EQ-GAN converges to the optimal Nash equilibrium due to its stability. We provide the first multi-qubit experimental results of a fully quantum GAN in the literature with a provably optimal Nash equilibrium [11, 18, 19]. Experiments on the Google Sycamore quantum processor show that the EQ-GAN improves state overlap compared to an ideal swap test with uncharacterized errors. Moreover, numerical simulations of the EQ-GAN architecture up to 18 qubits provide evidence of the favorable error-mitigation properties for larger circuits. Since training quantum machine learning models can require extensive time to compute gradients on current quantum hardware, resilience to time-dependent gate errors during the training process is especially valuable in the noisy intermediate-scale quantum (NISQ) era of quantum computing.

Finally, we provide applications of the EQ-GAN in the broader context of quantum machine learning for *classical* data. Many of the most attractive quantum machine learning algorithms require a quantum random access memory (QRAM) [20]. By learning a shallow quantum circuit to generate a superposition of classical

data, an EQ-GAN can be used to create an approximate QRAM. We demonstrate an application of such a QRAM for quantum neural networks [21], improving the performance of a quantum neural network for a classification task due to the efficient access of the dataset in superposition.

PRIOR ART

To provide a pedagogical introduction to the quantum GAN, we begin with the direct analogy of a classical GAN from the prior work of Refs. [16, 17]. This architecture, despite its potential non-convergence (shown below and numerically verified in the supplementary material [22]), provides the clearest translation of GANs from a classical to a quantum setting. Several theoretical and experimental approaches to a fully quantum GAN do not necessarily have an optimal Nash equilibrium due to limited circuit expressivity [19, 23]; other approaches adopt quantum variants of the Wasserstein metric that may improve convergence [24, 25], but such works are ill-suited to the experimentally relevant regime we consider here, i.e. unsupervised optimization in the presence of unknown noise. The only experimental implementation of a fully quantum GAN with a provable Nash equilibrium adopts the original QuGAN architecture and demonstrates it on one qubit [18].

A GAN comprises of a parameterized generative network $G(\theta_g, z)$ and discriminator network $D(\theta_d, x)$. The generator maps a vector sampled from an input distribution $z \sim p_0(z)$ to a data example $G(\theta_g, z)$, thus transforming $p_0(z)$ to a new distribution $p_g(z)$ of fake data. The discriminator takes an input sample x and gives the probability $D(\theta_d, x)$ that the sample is real (from the data) or fake (from the generative network). The training corresponds to a minimax optimization problem, where we alternate between improving the discriminator’s ability to distinguish real/fake samples and improving the generator’s ability to fool the discriminator. Specifically, we solve $\min_{\theta_g} \max_{\theta_d} V(\theta_g, \theta_d)$ for a cost function V :

$$V(\theta_g, \theta_d) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(\theta_d, x)] + \mathbb{E}_{z \sim p_0(z)} [\log (1 - D(\theta_d, G(\theta_g, z)))] , \quad (1)$$

where $\mathbb{E}_{x \sim p_{\text{data}}(x)}$ represents the expectation over the distribution $p_{\text{data}}(x)$. If G and D have enough capacity, i.e. approach the space of arbitrary functions, then the global optimum of this minimax game exists and uniquely corresponds to $p_g(x) = p_{\text{data}}(x)$ [1]. Generalizing to the quantum setting, the classical data can be represented by a density matrix $\sigma = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ where $p_i \in [0, 1]$ are positive bounded real numbers and $|\psi_i\rangle$ are orthogonal basis states. In the first proposal of a quantum GAN (QuGAN) [16, 17], the generative network is defined by a quantum circuit U that outputs the quantum

state $\rho = U(\theta_g)\rho_0U^\dagger(\theta_g)$ from the initial state ρ_0 . The discriminator takes ρ_{in} as either the real data σ or the fake data ρ and performs a positive operator valued measurement defined by T whose outcome determines the probability of data being true:

$$D(\theta_d, \rho_{\text{in}}) = \text{Tr}[T\rho_{\text{in}}]. \quad (2)$$

Following Ref. [17], the QuGAN solves the minimax game

$$\min_{\theta_g} \max_T (\text{Tr}[T\sigma] - \text{Tr}[T\rho(\theta_g)]) . \quad (3)$$

Unfortunately, minimax optimization might not converge to a good Nash equilibrium. When ρ is close to σ , the optimal Helstrom measurement operator $T = P^+(\sigma - \rho)$ is close to orthogonal to the true quantum data σ and opposite to ρ . The next step of training will try align the generator state ρ with T to minimize the cost function in Eq. 3, perhaps overshooting σ . In the subsequent generator update, T will again be opposite to ρ . This leads to the oscillation of the generator and discriminator, possibly preventing convergence; we show a case of infinite oscillation in the supplementary material [22].

CONVERGENCE OF EQ-GAN

To ensure convergence to the optimal Nash equilibrium, we propose a new minimax optimization problem with a discriminator that is not directly analogous to the discriminator of a classical GAN. Rather than evaluating either fake or true data individually, the optimal discriminator is permitted to perform a measurement on the joint system of the true data σ and generated data $\rho(\theta_g)$ that, for appropriate parameters, gives the fidelity between the two inputs:

$$D_\sigma^{\text{fid}}(\rho(\theta_g)) = \left(\text{Tr} \sqrt{\sigma^{1/2} \rho(\theta_g) \sigma^{1/2}} \right)^2 . \quad (4)$$

Notice that in comparison Eq. 3 is a linear function of input states, which is not optimal in the state-certification problem [26] of evaluating quantum generative models. Let the discriminator $D_\sigma(\theta_d, \rho(\theta_g))$ represent the probability of measuring state $|0\rangle$ at the end of the discriminating circuit. If there exist parameters θ_d^{opt} that realize a perfect swap test, i.e. $D_\sigma(\theta_d^{\text{opt}}, \rho(\theta_g)) = \frac{1}{2} + \frac{1}{2} D_\sigma^{\text{fid}}(\rho(\theta_g))$, then D_σ is sufficiently expressive to reach the optimal discriminator during optimization. Since a traditional swap test across two n -qubit states requires two-qubit gates that span over $2n$ qubits, implementation on a quantum device with local connectivity incurs prohibitive overhead in circuit depth. Hence, we implement the discriminator with a parameterized destructive ancilla-free swap test [27]. The EQ-GAN architecture adversarially optimizes the generation of the state $\rho(\theta_g)$ and the learning of a fidelity measurement D_σ (Fig. 1).

We define a minimax cost function closer to that of the classical GAN in Eq. 1:

$$\min_{\theta_g} \max_{\theta_d} V(\theta_g, \theta_d) = \min_{\theta_g} \max_{\theta_d} [1 - D_\sigma(\theta_d, \rho(\theta_g))], \quad (5)$$

where $D_\sigma(\theta_d, \rho(\theta_g))$ is the parameterization of the swap-test result. We now show that a *Nash equilibrium exists*

$$HU(\theta_d)H|0\rangle_a|\psi\rangle|\zeta\rangle = \frac{i \sin \theta_d}{2} |1\rangle_a [|\zeta\rangle|\psi\rangle - |\psi\rangle|\zeta\rangle] + \frac{1}{2} |0\rangle_a [(e^{-i\theta_d} + \cos \theta_d)|\psi\rangle|\zeta\rangle - i \sin \theta_d |\zeta\rangle|\psi\rangle]. \quad (6)$$

Given the circuit ansatz defined above with the predefined range for the swap angle θ , the maximum value for distinguishing between two arbitrary states is uniquely achieved by perfect swap test angle $\theta = \pi/2$. More particularly, the probability of measuring state $|0\rangle$ at the end of the parameterized swap test depends on the swap angle θ according to

$$D_\sigma(\theta_d, \rho(\theta_g)) = \frac{1}{2} [1 + \cos^2 \theta_d + \sin^2 \theta_d D_\sigma^{\text{fid}}(\rho(\theta_g))]. \quad (7)$$

The discriminator aims to decrease the probability of measuring $|0\rangle$, and thus minimize Eq. 7 by getting close to $\theta_d = \pi/2$ which corresponds to the perfect swap test given $D_\sigma^{\text{fid}}(\rho(\theta_g)) \leq 1$. The next step is for the generator to maximize $D_\sigma^{\text{fid}}(\rho(\theta_g))$ by moving closer to the true data. Ultimately, the generator cannot improve when $\rho(\theta_g) = \sigma$. Since there is an explicit update process to train the EQ-GAN to approach the optimal Nash equilibrium, and since the equilibrium is stable (i.e. perturbations result in worse outcomes for both players), we conclude that initializations of the EQ-GAN close to the true state will converge to the Nash equilibrium. In the following section, we provide an example of such a scenario and provide experimental and simulated evidence of this behavior.

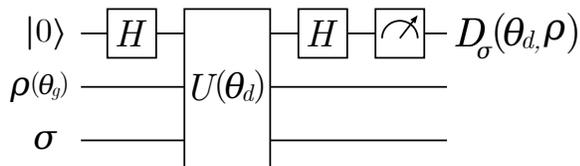


FIG. 1: EQ-GAN architecture. The generator varies θ_g to fool the discriminator; the discriminator varies θ_d to distinguish the state. Since an optimal discriminator performs a swap test, the global optimum of the EQ-GAN occurs when $\rho(\theta_g) = \sigma$. While we include an ancilla qubit in the figure for clarity, we implement a destructive ancilla-free swap test [27].

For simplicity, the example above does assume pure

states at the desired location. Consider a swap test circuit ansatz for the discriminator $U(\theta_d) = \exp[-i\theta_d \text{CSWAP}]$, which is the matrix exponentiation of a perfect controlled swap gate with angle θ_d . Under such ansatz, the input state $\rho_{\text{in}} = |\psi\rangle\langle\psi|$ and $\sigma = |\zeta\rangle\langle\zeta|$ will transform under the discriminator circuit into:

state input, although the cost function (Eq. 5) permits an EQ-GAN architecture for mixed states σ and $\rho(\theta_g)$ (see supplementary material [22]). In the experiments presented below, we use a hardware-efficient ansatz for the discriminator designed to correct dominant coherent gate errors.

LEARNING TO SUPPRESS ERRORS

Since the EQ-GAN architecture is agnostic to the precise parameterization of the discriminator, an appropriate ansatz can learn to correct coherent errors observed on near-term quantum hardware. In particular, the gate parameters of two-qubit entangling gates can drift and oscillate over the time scale of $O(10)$ minutes [14, 15], which can largely be mitigated by including additional single-qubit Z phase compensations [28]. Such unknown systematic and time-dependent coherent errors provides significant challenges for applications in quantum machine learning where gradient computation and update requires many measurements. Here, we demonstrate an application of the EQ-GAN to suppress uncharacterized errors in a two-qubit CZ gate. In experiments on the Google Sycamore quantum processor, we find that the adversarial approach reliably outperforms a fixed “perfect” swap test that fails to account for the unknown errors. We confirm this error-mitigation behavior up to 18 qubits in numerical simulation.

Suppose the adversarial discriminator unitary is given by $U(\theta_d)$, where $U(\theta_d^{\text{opt}})$ corresponds to a perfect swap test *in the absence of noise*. Given a trace-preserving completely positive noisy channel \mathcal{E} , the discriminator is replaced by a new unitary operation $\tilde{U}(\theta_d)$. While a supervised approach would apply an approximate swap test given by $\tilde{U}(\theta_d^{\text{opt}})$, the adversarial swap test will generically perform better if there exist parameters θ_d^* such that $\|\tilde{U}(\theta_d^*) - U(\theta_d^{\text{opt}})\|_2 < \|\tilde{U}(\theta_d^{\text{opt}}) - U(\theta_d^{\text{opt}})\|_2$. Because the discriminator defines the loss landscape optimized by the generator, the $\rho(\theta_g)$ produced by EQ-GAN may converge

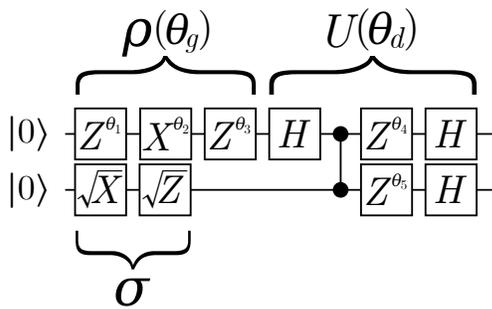


FIG. 2: EQ-GAN experiment for learning a single-qubit state. The discriminator $U(\theta_d)$ is constructed with free Z rotation angles to suppress errors due to the CZ gate (represented by connected black dots), allowing the generator $\rho(\theta_g)$ to converge closer to the true data state σ by varying X and Z rotation angles.

to a state closer to σ than possible by a supervised approach if the parameterization of the noisy unitary \tilde{U} is general enough to mitigate errors.

QML model	Minimum error in state fidelity
Perfect swap	$(2.4 \pm 0.5) \times 10^{-4}$
EQ-GAN	$(0.6 \pm 0.2) \times 10^{-4}$

TABLE I: Comparison of EQ-GAN and a perfect swap test on a Sycamore quantum device. The error of the EQ-GAN (i.e. $1 - \text{state fidelity}$) is significantly lower than that of the perfect swap test, demonstrating the successful adversarial training of an error-suppressed swap test. Uncertainties show two standard deviations.

As an example, we consider the task of learning the superposition $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ on a quantum device with noise (Fig. 2). To learn to correct gate errors, the discriminator is defined by an ideal swap test using a CZ gate, followed by adversarially learned angles of single-qubit Z rotations. The 2-qubit EQ-GAN model obtains a state overlap significantly better than that of the perfect swap test (Table I). Although both methods do not stay at the optimal point (Fig. 3), this is typical of noisy gradient measurements and minimax optimization: after convergence to the Nash equilibrium, discretization can induce perturbations while non-zero higher-order gradients lead the training to deviate from the global optimum [29]. Noisy simulations up to 18 qubits are seen to preserve the error-mitigating properties of the EQ-GAN in learning large GHZ states by producing states with near-maximal overlap despite coherent CZ gate errors consistent with experimental data (Fig. 4).

APPLICATION TO QRAM

Many quantum machine learning applications require a quantum random access memory (QRAM) to load *clas-*

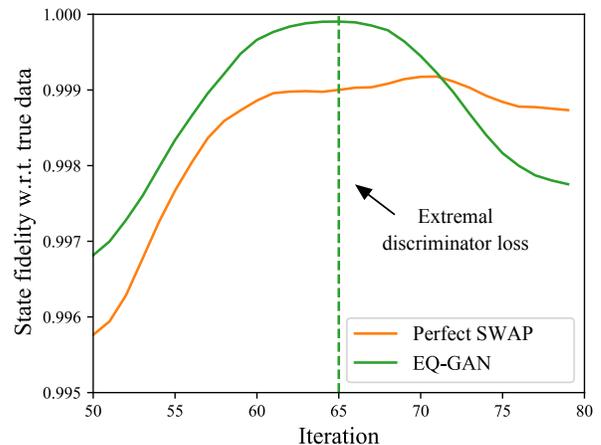


FIG. 3: Example of training an EQ-GAN and a perfect swap test on the Google Sycamore processor. We experimentally confirm that the EQ-GAN converges to a higher state overlap by learning to correct such errors with additional single-qubit rotations (see Table I for error bars). The “converged” EQ-GAN (dashed line) coincides with the iteration where the discriminator loss is minimized.

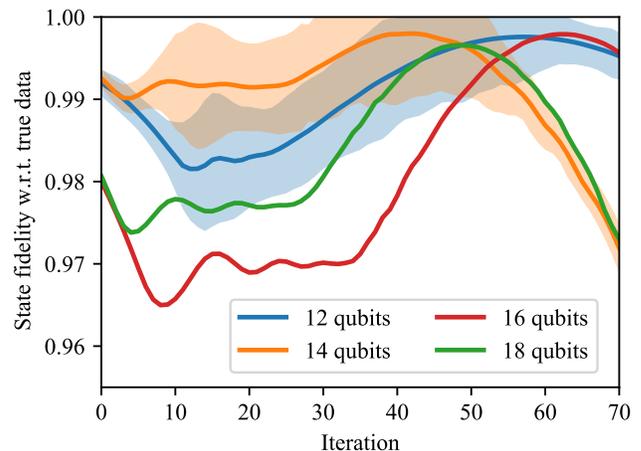


FIG. 4: Numerical simulation of EQ-GAN learning multi-qubit generalizations of the GHZ state, i.e. $(|0\dots 0\rangle + |1\dots 1\rangle)/\sqrt{2}$. Numerous simulations of a population of EQ-GAN models learning 6- and 7-qubit GHZ states (i.e. 12- and 14-qubit EQ-GAN) show the robustness of correcting Z phase errors in an ideal swap test; individual examples are shown for learning states with up to 9 qubits due to the computational cost of simulation. States are initialized and the EQ-GAN is trained with CZ gate errors consistent with experimentally observed noise [15]. Error bars show the 25th to 75th percentile range of state overlaps, centered around a single instance that obtains the mean improvement in state fidelity. See supplementary material for further analysis [22].

sical data in superposition [20]. More particularly, a set

of classical data can be described by the empirical distribution $\{P_i\}$ over all possible input data i . Most quantum machine learning algorithms require the conversion from $\{P_i\}$ into a quantum state $\sum_i \sqrt{P_i}|\psi_i\rangle$, i.e. a superposition of orthogonal basis states $|\psi_i\rangle$ representing each single classical data entry with an amplitude proportional to the square root of the classical probability P_i . Preparing such a superposition of an arbitrary set of n states takes $O(n)$ operations at best, which ruins the exponential speedup. Given a suitable ansatz, we may use an EQ-GAN to learn a state approximately equivalent to the superposition of data.

To demonstrate a variational QRAM, we consider a dataset of two peaks sampled from different Gaussian distributions. Exactly encoding the empirical probability density function requires a very deep circuit and multiple-control rotations; similarly, preparing a Gaussian distribution on a device with planar connectivity requires deep circuits. As described in the supplementary material [22], we adopt a double exponential peak ansatz [30] using 3 two-qubit gates on a planar architecture, whereas encoding the dataset in an exact superposition requires 57 two-qubit gates. Once trained to approximate the empirical data distribution, the variational QRAM closely reproduces the original dataset (Fig. 5).

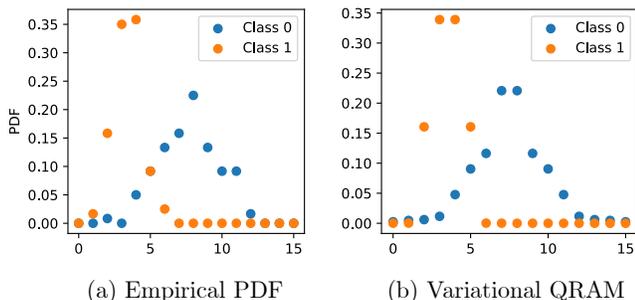


FIG. 5: Two-peak total dataset (sampled from normal distributions, $N = 120$) and variational QRAM of the training dataset ($N = 60$). The variational QRAM is obtained by training an EQ-GAN to generate a state ρ with the shallow peak ansatz to approximate an exact superposition of states σ . The training and test datasets (each $N = 60$) are both balanced between the two classes.

As a proof of principle for using such QRAM in a quantum machine learning context, we train a quantum neural network (QNN) [21] and compute hinge loss by considering each class in superposition encoded by the variational QRAM. The test accuracy on a balanced dataset of $N = 60$ examples is found to be $69\% \pm 2\%$ to two standard deviations. This demonstrates the capability of the EQ-GAN to prepare useful primitives in larger quantum machine learning frameworks, identifying shallow circuits that may replace subprocedures that may otherwise be deep and thus noisy.

CONCLUSION

Motivated by limitations of preexisting quantum GAN architectures in the literature, we propose the EQ-GAN architecture to overcome issues of non-convexity and mode collapse. We adopt a parameterization of Hilbert-Schmidt norm as the cost function as oppose to trace distance based on the optimality of Hilbert-Schmidt norm in state-certification problems. Similar advantages of Hilbert-Schmidt norm has been shown in quantum embedding designs of quantum kernel learning [31]. Other approaches to a quantum GAN may improve a quantum GAN’s convergence properties — notably, recent work suggests that certain cost functions such as the Wasserstein metric may provide more robust convergence [25] — but rely on accurate estimation of the discriminator, which may be difficult in the presence of unknown noise. We find that the EQ-GAN’s shallow discriminator is effective at suppressing device errors and ensures robust convergence in running laboratory quantum computers, making the EQ-GAN particularly relevant for near-term applications of quantum computing. Moreover, we demonstrate the first experimental application of EQ-GAN using Google’s cloud quantum computers in a machine learning application. This work opens up new directions in utilizing quantum generative models to achieve a quantum speedup in machine learning that necessitates the efficient and high-fidelity preparation of QRAM, as well as an additional setting to establish a theoretical understanding of the computational complexity of training the EQ-GAN and similar variational quantum algorithms [32].

An open source implementation of the EQ-GAN is available made available on GitHub with accompanying tutorials [36].

ACKNOWLEDGEMENTS

AZ acknowledges support from Caltech’s Intelligent Quantum Networks and Technologies (INQNET) research program and by the DOE/HEP QuantISED program grant, Quantum Machine Learning and Quantum Computation Frameworks (QMLQCF) for HEP, award number DE-SC0019227.

MYN and AZ contributed equally to this work.

-
- * These two authors contributed equally to this work.
- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in neural information processing systems* (2014), pp. 2672–2680.
 - [2] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, in *Advances in Neu-*

- ral Information Processing Systems*, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016), vol. 29, pp. 2234–2242, URL <https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf>.
- [3] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
 - [4] E. Putin, A. Asadulaev, Y. Ivanenkov, V. Aladinskiy, B. Sanchez-Lengeling, A. Aspuru-Guzik, and A. Zhavoronkov, *Journal of Chemical Information and Modeling* **58**, 1194 (2018), pMID: 29762023, <https://doi.org/10.1021/acs.jcim.7b00690>, URL <https://doi.org/10.1021/acs.jcim.7b00690>.
 - [5] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, J. M. Martinis, and H. Neven, arXiv preprint arXiv:1608.00263 (2016).
 - [6] S. Aaronson and A. Arkhipov, in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2011), STOC '11, p. 333342, ISBN 9781450306911, URL <https://doi.org/10.1145/1993636.1993682>.
 - [7] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, et al., *Nature* **574**, 505 (2019), URL <https://doi.org/10.1038/s41586-019-1666-5>.
 - [8] M. Y. Niu, A. M. Dai, L. Li, A. Odena, Z. Zhao, V. Smelyanskiy, H. Neven, and S. Boixo, arXiv preprint arXiv:2010.11983 (2020).
 - [9] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, arXiv preprint arXiv:1312.6199 (2013).
 - [10] J. Zeng, Y. Wu, J.-G. Liu, L. Wang, and J. Hu, *Phys. Rev. A* **99**, 052306 (2019), URL <https://link.aps.org/doi/10.1103/PhysRevA.99.052306>.
 - [11] C. Zoufal, A. Lucchi, and S. Woerner, *npj Quantum Information* **5**, 103 (2019), ISSN 2056-6387, URL <https://doi.org/10.1038/s41534-019-0223-2>.
 - [12] H. Situ, Z. He, Y. Wang, L. Li, and S. Zheng, *Information Sciences* **538**, 193 (2020), ISSN 0020-0255, URL <https://www.sciencedirect.com/science/article/pii/S0020025520305545>.
 - [13] J. Romero and A. Aspuru-Guzik, *Advanced Quantum Technologies* **4**, 2000003 (2021), URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qute.202000003>.
 - [14] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al., arXiv:1910.11333 (2019), URL <https://arxiv.org/abs/1910.11333>.
 - [15] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, M. Broughton, B. B. Buckley, et al., arXiv preprint arXiv:2010.07965 (2020).
 - [16] P.-L. Dallaire-Demers and N. Killoran, *Phys. Rev. A* **98**, 012324 (2018), URL <https://link.aps.org/doi/10.1103/PhysRevA.98.012324>.
 - [17] S. Lloyd and C. Weedbrook, *Phys. Rev. Lett.* **121**, 040502 (2018), URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.040502>.
 - [18] L. Hu, S.-H. Wu, W. Cai, Y. Ma, X. Mu, Y. Xu, H. Wang, Y. Song, D.-L. Deng, C.-L. Zou, et al., *Science Advances* **5**, eaav2761 (2019), <https://www.science.org/doi/pdf/10.1126/sciadv.aav2761>, URL <https://www.science.org/doi/abs/10.1126/sciadv.aav2761>.
 - [19] K. Huang, Z.-A. Wang, C. Song, K. Xu, H. Li, Z. Wang, Q. Guo, Z. Song, Z.-B. Liu, D. Zheng, et al., *npj Quantum Information* **7**, 165 (2021), ISSN 2056-6387, URL <https://doi.org/10.1038/s41534-021-00503-1>.
 - [20] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, arXiv:1611.09347 (2016).
 - [21] E. Farhi and H. Neven, arXiv preprint arXiv:1802.06002 (2018).
 - [22] See Supplemental Material [url] for additional numerical experiments and extensions of EQ-GAN, which includes Refs. [33–35].
 - [23] M. Benedetti, E. Grant, L. Wossnig, and S. Severini, *New Journal of Physics* **21**, 043023 (2019), URL <https://doi.org/10.1088/1367-2630/ab14b5>.
 - [24] S. Chakrabarti, H. Yiming, T. Li, S. Feizi, and X. Wu, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alche Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), vol. 32, URL <https://proceedings.neurips.cc/paper/2019/file/f35fd567065af297ae65b621e0a21ae9-Paper.pdf>.
 - [25] B. T. Kiani, G. D. Palma, M. Marvian, Z.-W. Liu, and S. Lloyd, *Quantum earth mover's distance: A new approach to learning quantum data* (2021), 2101.03037.
 - [26] C. Bădescu, R. O'Donnell, and J. Wright, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019), pp. 503–514.
 - [27] J. C. Garcia-Escartin and P. Chamorro-Posada, *Phys. Rev. A* **87**, 052330 (2013), URL <https://link.aps.org/doi/10.1103/PhysRevA.87.052330>.
 - [28] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, et al., *Science* **369**, 1084 (2020), ISSN 0036-8075, URL <https://science.sciencemag.org/content/369/6507/1084>.
 - [29] A. Brock, J. Donahue, and K. Simonyan, arXiv preprint arXiv:1809.11096 (2018).
 - [30] N. Klco and M. J. Savage, *Phys. Rev. A* **102**, 012612 (2020), URL <https://link.aps.org/doi/10.1103/PhysRevA.102.012612>.
 - [31] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, arXiv preprint arXiv:2001.03622 (2020).
 - [32] L. Bittel and M. Kliesch, *Physical Review Letters* **127**, 120502 (2021), URL <http://dx.doi.org/10.1103/PhysRevLett.127.120502>.
 - [33] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, et al., *Pennylane: Automatic differentiation of hybrid quantum-classical computations* (2020), 1811.04968.
 - [34] L. Mescheder, A. Geiger, and S. Nowozin, in *Proceedings of the 35th International Conference on Machine Learning*, edited by J. Dy and A. Krause (PMLR, Stockholmsmssan, Stockholm Sweden, 2018), vol. 80 of *Proceedings of Machine Learning Research*, pp. 3481–3490, URL <http://proceedings.mlr.press/v80/mescheder18a.html>.
 - [35] S. Aaronson, *The complexity of quantum states and transformations: From quantum money to black holes* (2016), 1607.05256.
 - [36] <https://github.com/tensorflow/quantum/tree/>

research/eq_gan